

Louvain School of Management

How and to what extent are e-consumer perceptions of website success reflected in basic website performance metrics?

Author: Giulia Lopriore
Supervisor: Isabelle Schuiling
Academic year 2021-2022
Dissertation for the master of
CEMS Master in International Management
Daytime schedule

I would like to thank my supervisor, Isabelle Schuiling, who's guidance and feedback made this project possible. I also extend my warmest gratitude to my family for supporting me unconditionally throughout my personal and academic life, and my friends for always motivating and encouraging me.

Abstract:

Websites are constantly seeking to improve the profitability of online channels by attracting and retaining e-customers through successful web design. In the past, measuring the success of a site was primarily done through user and expert evaluations of the site. However, newer evaluation methods consider the use of web analytics data as an objective measure of web performance. This research project thus seeks to explore how and to what extent user perceptions of the site are reflected in key web analytics. Answering this question also entails addressing whether websites with better analytics performance are also perceived more positively by users, which web attributes appear to be most conducive to analytic performance, and how these different perspectives of web evaluation can be effectively combined to inform improvements to the site. To achieve these objectives, an existing heuristic evaluation model asked participants to score the interface, navigation and content of four different international travel & accommodation websites. Publicly accessible analytics were then gathered for these websites and qualitatively compared to the user evaluations to explore the relationships between the two data sets.

The analysis concludes that, as suggested by recent literature, analytics such as visit duration, pages per visit, and bounce rate were able to reflect attributes of the tested websites. However, these analytics do not replace the need for user evaluations but rather serve as tools to enhance the understanding of the online customer journey. In addition, graphic design principles, logical structure, and interactivity seemed to show the most relationships to analytics within each attribute and thus may be prioritised when attempting to design a successful website.

UNIVERSITÉ CATHOLIQUE DE LOUVAIN
Louvain School of Management

Place des Doyens, 1 bte L2.01.01, 1348 Louvain-la-Neuve
Boulevard Emile Devreux 6, 6000 Charleroi, Belgique
Chaussée de Binche 151, 7000 Mons, Belgique
www.uclouvain.be/lsm

Table of Contents

1.	Introduction.....	1
2.	Literature Review	4
2.1	Website evaluation models	4
	<i>Summary of Literature on Web Evaluation</i>	15
2.2	Common Web Attributes	17
	<i>Visual Design</i>	17
	<i>Content and Interactivity</i>	18
	<i>IT & Reliability</i>	18
2.3	Link to web analytics	18
	<i>Visit behaviour</i>	19
	<i>Traffic sources</i>	19
	<i>Bounce rate</i>	20
	<i>Speed & Technology</i>	20
	<i>Additional Metrics</i>	22
3.	Methodology.....	24
3.1	Website Evaluation.....	24
3.2	Analytics Data	27
3.3	Analysis	28
4.	Results	30
4.1	Interface Results	30
4.2	Navigation Results	37
4.3	Content Results.....	42
4.4	Aggregated Results	49
5.	Discussion.....	54
5.1	Interpretation of Results	54
	<i>Analytics as Predictors of Interface</i>	54
	<i>Performance Metrics Reflecting Perceptions of Navigation</i>	57
	<i>Analytics Describing Content Evaluation</i>	59
	<i>Analytics' Reflection of Aggregated Evaluation Scores</i>	61
5.2	Managerial Implications	63
5.3	Limitations and Recommendations for Future Research.....	64
6.	Conclusions.....	66
	Bibliography.....	68
	Appendix.....	70

1. Introduction

With approximately one fifth of all retail sales being made online, websites are an increasingly important revenue stream for companies. Having an online presence is now an essential tool for businesses to engage with their customers and have a stake in this \$4.3 trillion industry (Coppola, 2021). In order to maximise the profitability of their online channels, managers need to be able to optimise the online customer experience to encourage purchases and return visits. Traditional customer experience (CE) is defined by the subjective responses customers have to interactions with the company. The **online customer experience** (OCE) is the same concept with a few key differences (Rose, Hair, & Clark, 2011). Moving from face-to-face to an online environment removes the need for personal contact. Simultaneously, websites allow for more information to be displayed and transferred to the user, primarily through audio-visual media rather than showing physical product ranges. Lastly, in contrast to the offline CE where the company controls the time and place of interactions, e-commerce enables the user to determine when and where they interact with the company. As the user's impressions of the website are dependent on their personal experience on the site, OCE can be described as multifaceted and unique to each customer. It can, however, be influenced by optimising certain aspects of the online environment that are said to be antecedents of a positive OCE (Rose, Hair, & Clark, 2011).

One of these key aspects is **usability**. It is difficult to research the topic of e-commerce without coming across the term usability. Nielsen pioneered the study of usability in the 1990s and early 2000s, arguing that it is one of the key determinants of user experience and satisfaction on a site. The definition of usability encompasses different aspects of a website that make it easy to use and perform tasks in an efficient, effective, and satisfactory manner (Nielsen, 2012; ISO, 1998). Due to this broad and ambiguous definition, usability relates to many attributes of the online environment. for the purpose of this study, the term usability will strictly refer to the website attributes that help users achieve tasks (i.e. user-friendliness and navigation). Neglecting this critical aspect of the website can cut potential sales in half and reduce return visits by 40% (Cunliffe, 2000). These alarming figures have encouraged researchers to develop different ways of assessing website usability. however, other topics related to website quality and positive OCE have also been found and tested as e-commerce has grown. As the majority of these models ask for the judgement of users or web experts on

subjective attributes, they can be considered heuristic assessment/evaluation tools. The web assessment methods and related topics are further explored in section 2.1.

More recently, **web analytics** (WA) have provided a complimentary or alternative tool to the heuristic assessment of online channels. WA tools allow companies to easily gather, analyse, and communicate large amounts of website data. They are able to reveal hidden trends, patterns and relationships that allow managers to gain insights on how users are behaving on the website. To make this data more comparable to heuristic evaluation methods, a number of studies have linked different WA metrics to the critical website attributes determined in heuristic models (table 2) (Kumar & Hasteer, 2017; Hasan et al., 2013; Beri & Singh, 2013; Wiggins, 2018).

This research project focuses on the direction taken by more recent literature that aims to draw parallels between perceptions of web attributes and the analytic performance of the site. Exploring the links suggested in the literature will enhance the current knowledge base on this topic and build on it by proposing new analytics as measures of web attributes. Providing additional support and evidence on the topic will enable managers to make more informed decisions about aspects of the web they should prioritise and what effects these website improvements may have on the performance of the platform.

To summarise the scope of this research, the study aims to answer the question 'How and to what extent are e-consumer perceptions of website success reflected in basic website performance metrics?'. Answering this research question implies addressing the following aims and objectives as well:

- To understand whether/how heuristic website criteria are related to basic web analytics.
- To investigate whether websites that score well in heuristic evaluations generally perform better.
- To draw conclusions about which website attributes may lead to better performance in certain basic metrics.
- To provide a starting point for managerial recommendations on which parts of the website should be focused on to improve performance.

- To provide recommendations on how to use both web evaluation methods as a holistic measure of website effectiveness.

To answer the main research question and achieve the objectives of the project, the literature on the topic of web evaluation models and web analytics was first thoroughly reviewed (Chapter 2). The most important models from the literature were critically evaluated and finally the most suitable and viable model was chosen to make up the first half of the data collection process. From the literature on web analytics, five free online analytics websites were identified to gather the second half of the data. Chapter 3 of this report details how the websites for the study were chosen based on the requirements of the evaluation model and the availability of data on the analytics sites. Five participants were asked to take part in the evaluation and the analytics data was gathered in the same week. Chapter 3.3 explains how the data from both sources was combined into graphs facilitating the identification of trends and patterns. These graphs are shown and described in the results chapter and the implication of these findings are thoroughly discussed in chapter 5.1. The practical implications of the project are elaborated in chapter 5.2 where the methodology of this project is simplified into a hybrid evaluation model that managers could use for their platforms. This is followed by a critical reflection of the important limitations of the study including difficulties encountered with access to analytics related to aspects of website reliability and inability to measure the statistical significance or causality of the studied relationships (chapter 5.3). Finally, the conclusion provides the answer to the research question, explains to what extent the objectives have been reached and emphasises how the project has contributed to the literature on the topic of web evaluation and analytics (chapter 6).

2. Literature Review

2.1 Website evaluation models

As e-commerce has matured, the literature on web evaluation has grown to include many different models and methods for the measurement of web success. These models can largely be divided into qualitative models that use fewer but richer responses, and quantitative models that are able to draw statistical conclusions about different web dimensions. The models within these categories share similarities that will be further explored in this section by describing several important studies in depth.

Schubert & Selz, 1999

Schubert & Selz (1999) were among the first to formally suggest that traditional marketing paradigms cannot be directly transferred to the online environment and as such presented a new and adapted evaluation model for commercial websites. Schubert & Selz (1999) recognised that there are underlying characteristics of e-commerce that make it fundamentally different from traditional offline commerce. They thus developed the Web Assessment tool to adapt evaluation practices to the unique context created by the use of technology for commerce. The Web assessment methodology was then used as a basis for a software tool that enables users to evaluate websites.

The Web Assessment tool is constructed around several perspectives of the online environment. Firstly, it incorporates online transaction phases from the information phase that ensures initial satisfaction when users start browsing prices and conditions of different products, to the agreement phase where negotiations between buyer and seller take place and the final settlement phase that determines the details of the virtual or physical product delivery and potential after-sales contact. Additionally, the authors recognised that perhaps one of the most important components of e-commerce is the power of community. The communication made possible through internet commerce brings the consumer closer to the product and allows for interaction with other users that have similar interests to create greater trust in the company and product.

Beyond the consideration of these phases, the model bases criteria on media inherent characteristics including interface, 24-hour access, anonymity, and interactivity among others. Finally, emphasis is placed on performance marketing which measures the extent to which

the company is offering an overall experience and fine-tuned segmentation with the help of technology.

The resulting Web Assessment tool thus organises questions covering the media-inherent characteristics for each transaction phase. To keep the questionnaire as short as possible and encourage all participants to finish, only the most relevant media characteristics were included for each phase. Each criterion is phrased as a statement, for example "Good Access to Community", to which the respondent answers the extent to which they agree with this statement. The questionnaire uses a 4-point scale from strongly agree (+2) to strongly disagree (-2) with no neutral option forcing users to reveal their preferences. A "not applicable" option is provided for websites where the criterion is not relevant (for example scoring logistics for a virtual product or service). Each criterion includes a hyperlink to a detailed explanation to help users answer. After the respondent has scored criteria for each phase, they are asked to evaluate the general quality of the website. The model suggests that the success of the website is determined separately by combining the ability to fulfil customer needs and expectations (from the consumer's perspective measured using the Web Assessment tool) and the financial success of the company (from the vendor's perspective).

Table 1 shows the criteria used by Schubert & Selz (1999) for each transaction phase. These scores are then aggregated to produce either a single company profile or an inter-business comparison. In the case of a single company profile, the mean and deviation for each criterion can be measured to identify the average score and level of agreement among evaluators. The authors emphasise that the scoring system (from +2 to -2) is not a measurement of the overall success of the website but rather an indication of the aspects which the company deems necessary (0 to +2) or unimportant (0 to -2). Thus, when making an inter-business comparison, the scores for each criterion are compared between the websites to analyse how they focus their resources on different aspects and how that may be impacting their financial success.

The Web Assessment tool developed by Schubert & Selz (1999) is one of the first models to adapt marketing principles to the online environment. They have thus been able to develop a tool that is specifically relevant to online marketing channels. The scoring system provides a simple and accessible method for comparing different aspects of online channels. However, the model cannot be used to compare overall quality of websites and requires to be used in

combination with external data such as financial performance. Furthermore, the authors mention excluding certain media characteristics for different phases but fail to mention which ones and why they were excluded.

Olsina, La Fuente, & Rossi, 2001

Olsina, Lafuente, & Rossi (2001) contribute to the literature on website evaluation methods by developing a methodology for specifying key characteristics and attributes for website quality. In addition, the research organises these identified characteristics in a hierarchy to create an understanding of how these attributes can be measured and linked to one another. The research was carried out specifically for academic websites but is still able to provide insights into alternatives to Schubert & Selz's (1999) organisation of criteria in the early years of website evaluation methods. While Schubert & Selz (1999) organise web criteria along transaction phases, Olsina, Lafuente, & Rossi (2001) propose high level topics affecting all phases of the customer journey. These global quality characteristics are sub-divided into categories and then further split into detailed measurable attributes. While the lowest level attributes are specifically developed with the use in mind (in this case for example, criteria such as syllabus, scholarship information, student-oriented tour, etc.), the high-level topics are fundamental to all websites and can thus be used as a framework to build a customised evaluation model around. The high-level website aspects are usability, functionality, reliability and efficiency. It is clear from the literature summary in table 1 that these topics are still useful and relevant in website evaluation today.

Van der Merwe & Bekker, 2003

The model developed by van der Merwe & Bekker (2003) is an updated version of Schubert & Selz's Web Assessment tool. Like Schubert & Selz (1999), van der Merwe & Bekker's (2003) model is based on the phases of the online customer journey which they identify as need recognition, gathering information, evaluating information, and making the purchase. To each of these stages, they attribute different aspects of the website, namely, interface, navigation, content, reliability and finally the technical aspect that facilitates all stages. The authors then divide each web aspect into 4 groups, with each group including five criteria. The full list of 100 criteria creates an evaluation model that is able to comprehensively measure any website's ability to support and engage the customer in their online journey. Table 1 summarises the groups of criteria used.

Van der Merwe & Bekker's (2003) model uses the same scoring system as Schubert & Selz. The evaluator is asked to score each criterion from -2 (strongly disagree) to +2 (strongly agree) with no neutral (0) option. As described in the Web Assessment model, a "not applicable" option (with a weight of 0) is given for websites where the criterion is irrelevant. The model can only be used for websites within the same industry, thus if certain criteria are not applicable for the whole industry it will not affect the comparison between website scores. The authors suggest using a minimum of 3 evaluators to ensure objectivity. Once their scores are gathered, evaluators come together to discuss criteria that are scored differently, this results in a consensus score for each website. Criteria for each group are summed together to measure and compare the website's performance along different attributes. The data can be analysed in a contingency matrix that shows the attribute scores for each evaluated website. This can then be presented graphically on a matrix.

As an adapted version of Schubert & Selz's (1999) Web Assessment tool, van der Merwe and Bekker's (2003) model shares the same strengths with some additional advantages. Firstly, the criteria have been updated to better fit modern consumers' needs and expectation for online commerce channels. Furthermore, the authors provide a clear and easily understandable method for collecting, analysing, and presenting the findings from the model. This model allows for the comparison of websites along different attribute websites as well as their overall success. Van der Merwe & Bekker (2003) emphasise that the model has been developed for company managers as well as for market research and academic purposes. However, as the model aims to cover the full online customer journey, the depth and breadth could be a significant limiting factor in the replicability of the model. In particular, the Reliability attribute measured requires users to have stored customer profiles on each of the websites being evaluated as well as experience with the check-out and post-order processes. Additionally, the Technical attribute seeks to measure aspects of the online environment that require professional knowledge and experience within the field of IT, which can be measured more easily and accurately by online website evaluation tools. These criteria include the types of security systems used, loading speed and cacheing data, among others. The inclusion of these dimensions that could be replaced by online tools makes the survey extremely long, completing the evaluation for several websites would require several hours to complete.

Oppenheim & Ward, 2006

Another model using a similar methodology to Schubert & Selz (1999) and van der Merwe and Bekker (2003) is the one developed by Oppenheim and ward (2006). However, this model was specifically adapted to evaluate websites selling chocolate. The authors first identified general attributes that contribute to the success of a website by reviewing the literature (these attributes are summarised in table 1). They then interviewed volunteers to understand which factors are most important for this industry and combined this with knowledge from previous studies. Like Schubert & Selz (1999) and van der Merwe & Bekker (2003), Oppenheim & Ward used a scale from -2 (very poor) to +2 (very good) and included a "not applicable" option. Unlike the two models previously described, however, Oppenheim & Ward (2006) include a neutral option (0) for average performance in in the criterion. In addition, the factors that were repeatedly stressed throughout the literature are considered critical factors in this model and are thus waited from -4 to +4. When analysing the data to determine which evaluated site performed best, the authors assessed both the website with the highest frequency of top scores as well as the website with the highest average.

To select which websites to evaluate, the researchers looked up "chocolate" and "buy" in google, 50 relevant websites were chosen from the first 70 pages of search results. These 50 websites were then categorised under the following headings: poor, average, or good design, whether they are internet only, full or partial dedication to selling chocolate, and whether they are a supplier of handmade or recognised brand chocolate. The authors then picked 10 websites from a mix of the above listed category. The study used 2 panels of 4 volunteers (8 in total). Each panel was asked to evaluate 5 websites. Unlike in van der Merwe & Bekker's (2003) model, the evaluators were not asked to discuss scores, instead their individual scores were summed and averaged to achieve the final score for each site. The evaluations took place in the same week to account for any changes made to the website and the volunteers were asked to put items in their cart and abandon at checkout to collect as much information as possible about the transaction page. Once gathered, the websites were analysed qualitatively. The authors describe the overall quality of the website and strengths and weaknesses of each based on the gathered responses. Finally, the study is able to make generalisations about common shortcomings of web design in this particular industry

With this study, Oppenheim & Ward (2006) are able to provide an example of how a standardised model can be adapted to a specific industry thus adding value to similar models such as those proposed by van der Merwe & Bekker (2003) and Schubert & Selz (1999). The proposed use of average scores as opposed to a consensus score is arguably also a valuable adaptation as the creation of a single score through discussion may result in certain evaluators' opinions being overshadowed thus skewing the results. Instead using an average score may produce an evaluation that more accurately represents individual experiences of the websites.

Despite these strengths, there are a few significant methodological flaws in this study. While the primary purpose of the study is to tailor a model to the chocolate industry, the authors chose to double the weights of certain critical factors identified in the literature rather than the findings from their interviews with keen chocolate customers. This introduces the risk of placing emphasis on criteria that may not necessarily be critical to the success of websites within this industry, thus affecting the validity of the obtained scores. In addition, the research method introduces a significant bias when choosing websites under predetermined categories of design quality as there is already a preconceived notion of how they believe the results should look like. This issue could have been minimised by presenting the survey data graphically to allow readers to compare performance in different web dimensions and better support the subjective claims made in the qualitative descriptions of each website's evaluation.

Hung & McQueen, 2004

Hung & McQueen (2004) adopt a similar approach to web attribute categorisation as Schubert & Selz (1999) to create an instrument for website evaluation specifically from the perspective of first-time buyers. They use the following transaction phases; information, negotiation, settlement, and after-sale. However, Hung & McQueen (2004) use a unique method to the development of their web evaluation model by first identifying the process of satisfying customers and where potential failure points can occur in this process. The suggested satisfaction cycle begins with the user finding the function on the web, using this to conduct transaction activities which eventually lead to satisfaction. Between each of these steps there are failure points that dictate the level of satisfaction achieved. To identify and measure the occurrence of these failures the authors define 4 criteria that determine whether users are

able to surpass the failure points. Ease-of-identification is used to measure failures between finding the function and conducting the transaction. Ease-of-use identifies the failure point between conducting the transaction and producing satisfaction. Finally, usefulness and interactivity are used to measure potential failures between satisfaction and finding the function again. The authors use a scoring system from 0 to 10, with 0 indicating very poor and 10 being very good performance. Respondents are asked to conduct 14 transaction activities according to different transaction phases and score each of these along the outlined scale.

Hung & McQueen (2004) were able to provide an easy tool to measure any kind of website. Testing the tool with different kinds of respondents confirmed that participants do not need any prior knowledge or background in IT/e-commerce and the evaluation of individual websites can be completed in 20 minutes. However, the model only uses 4 ambiguous attributes of the web environment. These attributes could be measured more accurately by breaking them down into more detailed criteria as done in the other literature described thus far.

Chiou, Lin & Perng, 2010

By 2010, the literature on web design and evaluation methods had expanded massively with the growth of online commerce. Chiou, Lin, & Perng (2010) thus set out to summarise the findings made between 1995 and 2006 and use these to create a model for managers to use when assessing their online efforts. They began by scouring journals for relevant key words such as "website", "evaluation", "assessment", "measurement", "framework", and "model". The resulting articles were then filtered according to the journal importance and relevance. The authors were able to analyse 83 studies using different methodologies. The extensive search revealed several important findings. Firstly, the applied studies emphasise the use of transaction phases and thus support that frameworks should use transactional processes to identify key activities. For online commerce, information systems (IS) factors should be incorporated into the evaluation as they provide critical supporting factors for e-commerce. And finally, there is a lack of consideration for online strategy when developing website evaluation frameworks. Chiou, Lin, & Perng (2010) sought to address this final point when creating their framework. The topics covered throughout the literature were first categorised under unified factors (shown in Table 1), these criteria are then assigned to the 4PsC (Product, Promotion, Price, Place, and Customer Relations) in order to frame criteria to measure

strategic objectives. The authors finally provide a 5-step process for managers to use this pool of criteria to evaluate their online channels. First, web managers are interviewed to identify strategy and relevant criteria/weights. Then the evaluation instrument is developed in line with the strategy. A panel of experts conducts the evaluation, this is then converted to analysable numbers. The data can finally be presented in a 4PsC radar chart and matrix chart to draw conclusions on necessary improvements.

Chiou, Lin, & Perng (2010) thoroughly summarised the existing literature on web evaluation and added value by developing a practical method for managers. The model created is able to create a tailor-made evaluation framework that is both relevant to the company's objectives and grounded in theory. However, the suggested methodology does not provide a readily accessible model. In addition, the highly customised nature of the resulting model means it can only be used to assess a single website and no comparisons can be made between different domains. Nonetheless, for academics researching this field, this paper is still able to provide an overview of the main literature and different approaches to web evaluation.

Kim, Shaw, & Schneider, 2003

Kim, Shaw, & Schneider (2003) contribute to the literature on web evaluation by developing a tool for benchmarking website quality between industries. The model was used in a quantitative study measuring websites in 12 industries in Korea. University students were asked to select 20 websites in 3 to 4 different industries and to subsequently score these websites along the dimensions summarised in table 1. 2,020 responses were filtered down to 728 according to their quality and sincerity, and industries with 12 or less websites were excluded. This resulted in data from 245 websites in 12 different industries being analysed. The means of each industry were compared for each criterion. It was concluded that websites of online malls received the highest scores and could thus serve as a benchmark for good web design.

The quantitative nature of Kim, Shaw, & Schneider's (2003) study is able to provide statistical support for the quality of websites in different industries and provides a replicable way of comparing many different websites. The use of students as a sample, however, limits the type of websites explored as they will choose sites for their target group. Furthermore, the study

is focussed on Korean websites and should thus be replicated in other markets or for international sites to increase the value of inter-industry differences found.

Cao, Zhang, & Seydel, 2005

Like Kim, Shaw, & Schneider (2003), Cao, Zhang, & Seydel (2005) also opt for a quantitative approach to develop and support their evaluation instrument. However, their developed model is not for benchmarking but rather adds value to existing knowledge in the field by taking an IS perspective to identify factors contributing to website quality/effectiveness. The study surveyed a sample of students to test and verify which of the following are valid measures of web quality: system quality, information quality, service quality and attractiveness. Using the Technology Acceptance Model (a widely accepted method of predicting issues and system acceptance prior to user testing), these aspects of quality are then related to user's beliefs, attitudes, and intentions to come up with a framework that combines aspects of the site with the customer's perceptions.

The survey asked students to complete the same series of tasks on 3 different online bookstores. The path consisted of accessing the site and exploring, viewing books, selecting a book to add to the shopping cart, and finally viewing the content of the shopping cart. Respondents must score their perceptions of each website from 1 (strongly disagree) to 7 (strongly agree). The survey resulted in 213 usable responses. Based on the statistical analysis of the survey responses, Cao, Zhang, & Seydel (2005) can infer which aspects of the site are empirical to web success. The most important factors found were determined to be information accuracy, responsiveness, and search facility as well as trust and attractiveness.

The model is able to provide design guidelines that are statistically supported; however, the chosen criteria lack some important constructs defined throughout web evaluation literature such as user-friendliness and navigational aspects. Furthermore, the model has only been tested on small online shops within a single industry and uses respondents from one university thus limiting the potential replicability of results across different industries and types of respondents.

Loiacono, Waston, & Goodhue, 2007

Loiacono, Watson, & Goodhue (2007) also base their model on the Technology Acceptance model, along with the Theory of Reasoned Action (a theory suggesting individuals adapt their

actions based on the evaluated consequences of a certain behaviour), to develop and test a WebQual instrument allowing for consumer evaluation of websites. The instrument employs the 12 criteria summarised in table 1 and has been validated through several methods. First the dimensions were defined through an extensive literature review, exploratory surveys, and interviews with experts. Items were then developed for each dimension and these were refined by asking 2 samples to rate 12 different websites. The refined model includes 12 dimensions with 36 criteria in total (each scored on a 7-point likert scale). This model's validity was then confirmed using 2 new samples between the years 2000 and 2003. While Loiacono, Watson, & Goodhue (2007) have created an easy and accurate instrument for user-based web evaluation, the study focuses largely on testing the validity of the tool and lacks detail on how to use the model such as the number of respondents that should be used or analysis/interpretation of results.

Joia & Barbosa de Oliveira, 2008

The last study in this literature review to use a quantitative method is the model developed by Joia & Barbosa de Oliveira (2008). Joia & Barbosa de Oliveira (2008) aim to investigate the relationship between the interface of a website and customer behaviour. Similarly to the two previous studies, the authors draw on the Technology Acceptance Model to describe customer behaviour. However, they investigate how user style and orientation (external factors) influence the perception of a web site. 496 respondents filled a survey evaluating CD retail websites on a 5-point likert scale. Testing the model proved its validity but only for the selected industry (CD e-retailers) within Brazil.

Hasan, Morris, & Proberts, 2013

Finally, Hasan, Morris, & Proberts (2013) begin to make a connection between heuristic evaluation methods and web analytics to form a hybrid method for evaluating web usability with a particular focus on developing countries. The study evaluates 3 Jordanian websites using 3 distinct methods: user testing, heuristic evaluation, and web analytics. This resulted in a 4-step approach that takes the advantages and disadvantages of each method into account. The research made use of the 3 methods simultaneously. The Google Analytics (GA) script was inserted into the websites and monitored for 3 months. For the user testing, 20 users were chosen, and their session was tracked and measured using screen capture software while completing typical tasks on the website. Finally, a heuristic evaluation was devised through a

literature review. 5 web experts were asked to evaluate the usability of the site along 5 categories (architecture & navigation, content, accessibility & customer service, design, and order process) (see Table 1). For each category, the evaluators were asked to assess whether they thought potential usability problems would be minor or major.

Once all the methods had been applied, Hasan, Morris, & Probets (2013) analysed the data to determine which methods were able to assess different usability issues. This involved first evaluating the methods separately to consequently form a list of common usability problem areas. To enable comparison, the GA metrics were divided into 9 categories (general usability metrics, internal search metrics, top landing pages, top content pages, top exit pages, finding customer support info, purchasing process, visitors' metrics, and financial performance metrics).

The results from the GA method using the suggested 9 categories of metrics led to the categorisation of web metrics under the 5 areas covered in the heuristic method. This categorisation was supported by cross-referencing the analytics to comments made by the web experts and user testing subjects. The GA method further found that web analytics are able to uncover potential usability problem areas related to one or several sub-areas but cannot pinpoint specific issues. Uncovering specific issues requires supplementary user testing or heuristic methods. Finally, the other 2 methods revealed that heuristic evaluation is effective in identifying a larger number of minor problems while user testing is able to identify fewer major issues. As a result of the analysis of the strengths and weaknesses of each method, a 4-step approach was suggested to make the most of each. First, use Google Analytics to track website usage. Then use the gathered web metrics to uncover potential general usability issues. Combine this data with user testing, heuristic evaluations, or both to find specific usability issues. Finally redesign the site to mitigate the identified problems and keep track of the web metrics to see how users react.

Hasan, Morris, & Probets (2013) provide a thorough comparison of different website evaluation methods and show how new tools such as web metrics can be incorporated into more traditional methods like heuristic evaluation. However, the relationships proposed by this study between web metrics and subjective web dimensions could be further tested beyond the context of developing countries to improve validity and give managers a better

idea of how to read analytics. Furthermore, the design of the study makes it difficult to replicate as it requires access to the e-commerce website to acquire the full range of analytics data and detailed information on the (post-) order processes.

Summary of Literature on Web Evaluation

Table 1 shows a summary of the existing literature on heuristic evaluation models. Despite the exponential growth of e-commerce in the last two decades, comparing older models to more recent ones shows that the principles of web quality have remained the same, with certain attributes being clear stable factors of the online experience. The topics that are most prevalent are: **usability** (also referred to as ease of use, user friendliness, navigation), **visual design** (interface, aesthetics, attractiveness), **content** (information), and **interactivity** (involvement), as well as **reliability** and **IT** features (system, privacy/security). These topics have been highlighted in the table to show their frequency and are further explained in section 2.3.

Table 1: Summary of Web Evaluation Literature

Authors	Attributes/criteria used
Schubert & Selz, 1999	Information Phase: • Interface • Structure of content • Information quantity • Benefits from stored profiles • Good product/service combinations • Good system performance/availability • Cost benefits Agreement Phase: • Adjustable customer profile • Personalised services • Possibility of customised products • Integration of business rules • Security issues • Customer service Settlement Phase: • Easy settlement process • Integration of settlement services • Effective use of customer profile • Track & trace • After sales support Community component: • Access to community • Unique/original info • Adequate # of members • Collaborative filtering

	• Ability to choose avatar • Privacy • Perceived added value from membership • Customised push mechanism • Pull mechanisms
Olsina, Lafuente, & Rossi, 2001	• Usability (global understandability, feedback & help, interface, miscellaneous) • Functionality (search, navigation, domain-related features) • Reliability (non-deficiency) • Efficiency (Performance, Accessibility)
Kim, Shaw & Schneider, 2003	• Business function • Corporation credibility • Content reliability • Attractiveness • Systematic structure • Navigation
Van der Merwe & Bekker, 2003	• Interface (graphics, visual design, style & text, flexibility/compatibility) • Navigation (structure, ease of use, search engine, navigational necessities) • Content (product/service info, company info, information quality, interactivity) • Reliability (stored profile, order process, post-order, customer service) • Technical (speed, security, software & databases, system design)
Hung & McQueen, 2004	• Ease of identification • Ease of use • Usefulness of information • Interactivity
Cao, Zhang & Seydel, 2005	• System quality (search facility, responsiveness, multimedia capability) • Information quality (accuracy, relevance) • Service quality (trust, empathy) • Attractiveness (playfulness)
Oppenheim & Ward, 2006	• Presentation elements • Content • Accessibility • Navigation • Language • Transaction pages • Security, privacy & authority
Loiacano, Watson & Goodhue, 2007	• Ease-of-use (Intuitive operations, ease of understanding) • Information (Quality, tailored) • Transaction-related (response time, trust, online completeness, relative advantage, consistent image) • Entertainment value (visually appealing, innovative, emotionally appealing)
Joia & Barbosa de Oliveira, 2008	• Perceived utility • Perceived ease of use • Pleasure & attractiveness • Trust

Chiou, Lin, & Perng, 2010	• Ease of use • Responsiveness • Fulfilment • Security/privacy • Personalisation • Visual appearance • Information quality • Trust • Interactivity • Advertising/persuasion • Playfulness • Technology/integration
Davidavičienė & Tolvaišas, 2011	• Basic elements • Reliability • Easy to use • Information quality • Design • Interactivity • Added value • Order process
Hasan & Proberts, 2013	Heuristic: • Architecture & navigation • Content • Accessibility & customer service • Design & purchasing process Performance metrics: • See table 2

2.2 Common Web Attributes

Visual Design

Features of website design encompass all aspects of how the website looks and is an important contributor to the usability of the site. First and foremost, the design has to aid the user in completing the desired tasks by using legible text that is presented in a way that does not overwhelm the user. As e-commerce has matured over the years, clear design guidelines have emerged, these include for example, consistency in page layout, limited scrolling, avoidance of flashy or blinking graphics, and consistent use of colour (Gehrke & Turban, 1999). While design standards have developed across e-commerce to provide e-customers with a familiar experience, websites still have a lot of freedom to design the interface in a way that reflects company's branding and gives the customer an enjoyable experience on the site. The creativity of the web designing process and individual tastes of users makes it difficult to

objectively assess the success of websites in satisfying and entertaining users while making a purchase. Evaluation models have thus focussed on measuring the experienced effectiveness of design elements such as graphics, text, white space, and overall consistency.

Content and Interactivity

To convey information about the company and products sold, the site must include information that is both accurate and concise. The evaluation of website content thus depends both on the amount and quality of the provided information. If customers are not able to find the information they are seeking, they are likely to leave without making a purchase. Structure or organisation of information is a closely related concept that has been widely explored in literature as well as it helps users in navigating the website's content. Evaluation models summarise the success of information by assessing whether products are well described and pictured, whether the information is relevant and up to date, and finally whether the displayed information is tailored to the customer. This latter feature pertains to the interactivity of the site. Interactive features enable two-way communication where the user is able to express to the website which information they would like to see, supporting the customer along the purchasing journey. Interactivity can be heuristically evaluated by assessing the extent of possible customisation as well as the ability to access online communities through the site. (Chakraborty, Lala, & Warren, 2002)

IT & Reliability

The more technical side of the website is evaluated through features of IT and reliability. This includes aspects of infrastructure such as the ability to store customer profiles and provide a smooth and transparent order (and post-order) process, as well as the use of adequate systems for speed, security, and privacy (van der Merwe & Bekker, 2003). Unlike the more creative aspects of web design, IT and reliability is more easily measured with performance tools rather than heuristics as public tools can provide information about the loading speed and quality of technology used.

2.3 Link to web analytics

As many heuristic assessment models have now been developed and critical aspects of web design are converging, newer research has shifted more toward the inclusion of new tools such as web analytics. As more analytic tools emerge on the market, using heuristic models alone may set companies at a disadvantage compared to those who are also adjusting the site

based on performance data. Several studies have therefore researched the relationships between various attributes acknowledged in heuristic studies and WA metrics. However, these links are not yet well understood. Table 2 summarises which established website attributes could be measured with analytic performance metrics. Currently these suggested relationships are largely deduced from our knowledge of online customer behaviour rather than being evidence based. This could also mean that less obvious or logical relationships may have been overlooked. Testing these links could thus contribute to a better understanding of how to combine web metrics with heuristic evaluation and help managers to see which aspects of the site to change based on performance data.

Without direct access to a website, it is only possible to gather certain key metrics. Free tools like Similarweb.com and Semrush.com are able to provide a number of basic metrics (defined below) that can be tested against website attributes. The data that is publicly accessible is highlighted in bold in table 2, the free tools and accessible data are summarised in table 3.

Visit behaviour

The metrics that provide information about visitor's behaviour while on the site include visits, unique visits, visit duration, pages per visit. The total number of visits shows how many times the website has been accessed in a month; this includes users who have visited more than once. The number of unique visitors excludes users who have visited the site more than once to show only new users. These metrics can provide insights into the popularity and growth of the website. Instead, the amount of time users spend on the site (visit duration) and the number of pages that users explore before leaving (pages per visit) can give clues about the enjoyment users experience on the website (Bakaev et al., 2016; Kumar & Hasteer, 2017). Alternatively, studies have also argued that if the metric is not compared to other data such as page views, longer visit durations are an indication of usability issues as it takes a long time for customers to navigate the site (Beri & Singh, 2013). It is thus clear that visit behaviour is linked to different web attributes, however, the nature and extent of these relations requires further research.

Traffic sources

The way users reach the website can be identified by looking at the traffic sources. Organic traffic refers to users who find the website through unpaid search results. This type of traffic can be increased through Search Engine Optimisation (SEO) which helps website pages appear

in more searches across different browsers. The opposite of organic traffic is paid traffic; this is where the website buys advertising on browsers, social media, or other websites to drive traffic to the site. Paid traffic occurs primarily through pay per click (PPC) and cost per thousand (CPM). While advertisements help increase traffic, if the advertisements are not targeted properly it may lead to a higher bounce rate. Finally, users who reach the website by clicking on links on other websites are considered referral traffic. Looking at the sources of a site's referral traffic can provide insights into the interests of your target audience and help tailor the most popular referred pages to these interests.

Bounce rate

A website's bounce rate measures the number of single-page sessions as a percentage of all sessions. Unless the purpose of the site revolves around a single page (like a blog) a high bounce rate may be an indication that the website is not reaching the right audience, the interface is off-putting, or the website is difficult to navigate. The bounce rate can be improved by analysing which traffic sources and pages produce the highest bounce rates and adjusting those pages accordingly. Alternatively, if the problem is diffused across the website the design, language, and tagging of the overall website should be reconsidered.

Speed & Technology

Finally, there are certain website metrics that can be gathered through public analytics tools that describe critical functional aspects of the site. Although aspects of website infrastructure are not immediately visible to the user, they are essential to the online customer experience as they facilitate browsing and purchasing. With technological advancements, e-customers now have high expectations for the speed of a website and will get quickly frustrated if these expectations are not met. Websites like WebsiteGrader are able to analyse how fast the website is by assessing how long the website takes to load and the page size. They are also able to determine the level of security and potential remaining security issues on the site, and the overall quality of the technology used from basic to exceptional.

Table 2: Literature relating analytics to web attributes

Website Attributes	Related performance metrics
Usability/navigation	• Visits (Bakaev, Mamysheva, & Gaedke, 2016) • Return Visits (Kumar & Hasteer, 2017; Bakaev et al., 2016) • Visit duration (Kumar & Hasteer, 2017; Bakaev et al., 2016; Beri & Singh, 2013) • Internal searches (Beri & Singh, 2013)

	• Bounce rate (Hasan et al., 2013; Kumar & Hasteer, 2017; Bakaev et al., 2016) • Pages per visit (Hasan et al., 2013; Kumar & Hasteer, 2017; Beri & Singh 2013) • Click depth (Hasan, Morris, & Probets, 2013) • Searches per visit (Hasan, Morris, & Probets, 2013) • % visits using search (Hasan, Morris, & Probets, 2013) • Error pages (Wiggins, 2008) • Conversion rate (Wiggins, 2008) • Organic traffic (Bakaev, Mamysheva, & Gaedke, 2016) • User activity (Bakaev, Mamysheva, & Gaedke, 2016)
Design	• Pages per visit (Beri & Singh, 2013) • Click depth (Hasan, Morris, & Probets, 2013) • Visit duration (Hasan, Morris, & Probets, 2013) • Bounce rate (Hasan, Morris, & Probets, 2013) • Load speed (Wiggins, 2008)
Content	• Visit duration (Beri & Singh, 2013; Wiggins, 2008) • Link density (Wiggins, 2008) • Return visits (Wiggins, 2008) • Referral Traffic (Wiggins, 2008)
Structure	• Visit duration (Hasan, Morris, & Probets, 2013) • Searches per visit (Hasan, Morris, & Probets, 2013) • % visits using search (Hasan, Morris, & Probets, 2013) • Click depth (Hasan, Morris, & Probets, 2013) • Pages per visit (Hasan et al., 2013; Beri & Singh, 2013)
Interactivity	• Visit duration (Beri & Singh, 2013; Wiggins 2008) • Return visits (Wiggins, 2008)
Reliability/functionality	• Load speed (Kumar & Hasteer, 2017) • Experience (Kumar & Hasteer, 2017) • Technology (Kumar & Hasteer, 2017) • Compatibility (Kumar & Hasteer, 2017)

Table 3: Data accessible via public analytics websites

	SimilarWeb.com	Semrush.com	WebsiteGrader.org	Nibbler.silktide.com	Webrate.org
Visits	X	X	X		
Unique Visitors		X			X
Pageviews					X
Website Value					X
Global Rank	X				X
Category rank	X				
Market share		X			
Avg visit duration	X	X			
Pages/visit	X	X			X

Bounce rate	X	X			
Traffic sources	X	X			
Backlinks		X	X		X
Display Advertising		X			
Customer rating (trustpilot)		X			
Theme confidence			X		
Speed			X		X
Mobile			X	X	X
Security			X		X
Accessibility			X		
Search ranking			X		
Readability			X		
Technology			X		
Amount of content				X	
Headings				X	
Images				X	
Analytics				X	
Printability				X	
Popularity				X	

Additional Metrics

As seen in the table there are many insightful analytics that have not been mentioned in the literature but are publicly accessible and could give greater information about the success of a website. The most relevant of these additional analytics are briefly defined below.

Market Share

Semrush is able to provide insights into a website's competitive position within its industry. The market share captured is reported as the traffic obtained by the website as a percentage of total traffic to the market. Compiling this data with the traffic growth of players within the same market enables the mapping of the top competitors on a growth quadrant, seen in Figure 1.

Theme Confidence

WebsiteGrader attempts to quantify aspects of the website's interface by analysing different themes and how strongly they are represented through a measure called theme confidence. To do this, the tool takes a screenshot of the homepage and analyses how salient different visual themes are. These could be for example people, outdoors, animals as well as more

ambiguous themes like text or images. The strength of each of these themes is then rated as a percentage. In addition, the tool is able to detect a colour palette based on the main colours used on the site. The objective of this measure is to identify the strength of the website's branding and whether they have a clear message and strong visuals that are able to successfully capture the user's attention in the short amount of time they have before the user is bored and leaves the site.

Accessibility

The accessibility of a website is critical for reaching and retaining as many e-customers as possible, including those who might have visual, hearing, mobility, or cognitive impairments. Although there are many software that can be included on a website to specifically target these kinds of disabilities, there are many foundational design elements that help impaired users as well as improving the online experience of non-disabled users. These include multimedia to deliver content in multiple ways, avoiding jargon to improve readability, clearly discernible links, and consistent navigation (Brophy & Craven, 2007). Tools such as WebsiteGrader are able to identify potential barriers to accessibility and give a score out of 100.

Search Ranking Potential

Majority of visitors are likely to access a website by searching via browsers. It is thus essential to be at the top of the results page to attract as many users as possible to the site. Analytics tools are able to compile various SEO data to give a general score of the websites potential to rank highly on Google.

Given the gap in research in the relationship between website attributes and web analytics, this research will test findings from heuristic evaluation methods to data gathered from public analytics tools. This will allow for the qualitative exploration of the extent and nature of the proposed relationships and any potential new relationships

3. Methodology

This research sets out to answer the question how and to what extent are e-consumer perceptions of website success reflected in basic website performance metrics? A deductive approach is taken as a research strategy is designed to test the theorised relationships between heuristic website evaluation and web performance metrics. Despite using a survey to collect website evaluations, the nature of this research is qualitative. Limiting the number of participants allowed for the evaluation of more websites per participant and the ability to consult each participant on important terms that contribute to the validity of the gathered data.

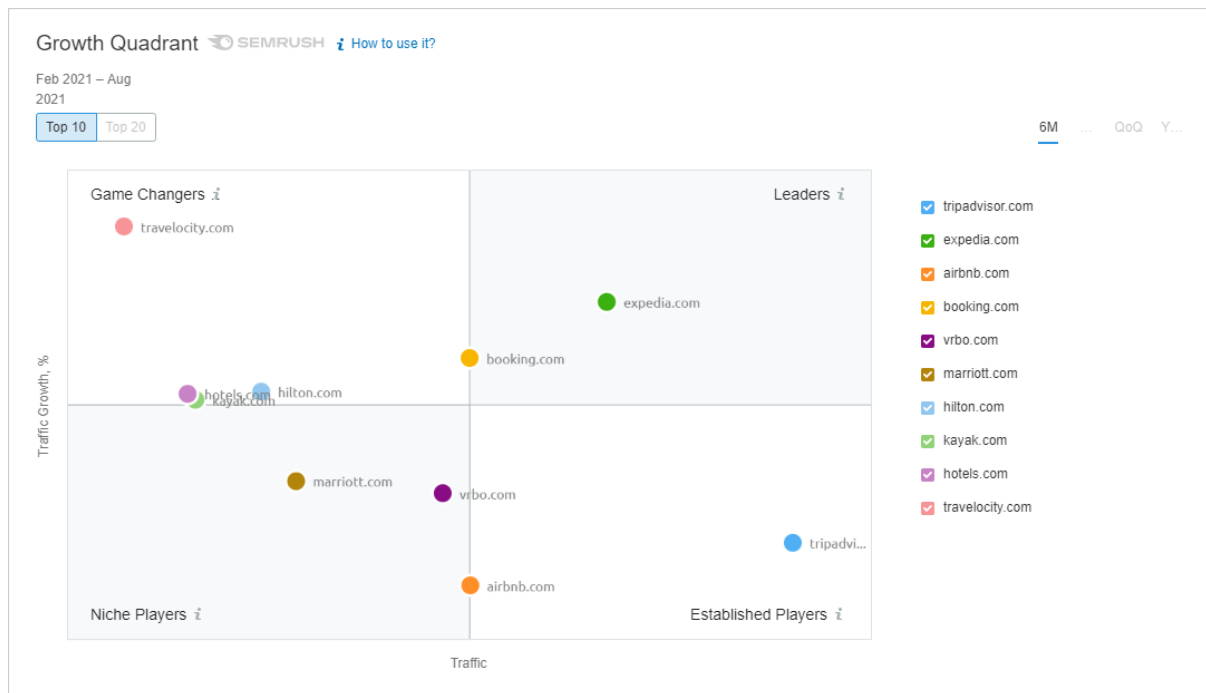
Answering this research question requires two main stages of data gathering. The first stage involves gathering consumer evaluations of several websites using a website evaluation model. The second stage requires gathering analytics data of the surveyed websites for comparison. This chapter describes in detail how this data was gathered to ensure the results of this research are both accurate and replicable and how the data was then analysed to draw meaningful conclusions.

3.1 Website Evaluation

When considering how to collect website evaluations in a reliable and valid way within the timeframe of this project, I decided to use one of the existing models found in the literature review. Using a model that has already been tested and cited by other studies meant I could save time improving the design of the model and focus on using the results to help answer the main research question. After critically evaluating the current literature on website evaluation models, I chose to use and adapt van der Merwe and Bekker's (2003) model (Appendix 1). This model covers all the key attributes identified throughout the literature and provides the most complete and ready-to-use assessment template. Furthermore, the article has been cited over 250 times and has been used as the basis for other assessment models (Hung & McQueen, 2004; Hasan, Morris, & Proberts, 2013; Oppenheim & Ward, 2006). The study proposes using e-commerce sites within the same industry, gathering evaluation data on each of these sites from a minimum of 3 evaluators (scoring each criterion between -2 and 2), and presenting the data in a contingency matrix. As the model is designed to compare different attributes for different websites, it is easily comparable to website data gathered through public analytics tools.

To maintain the strengths of the original model I aimed to follow it as closely as possible while editing certain aspects to mitigate the weaknesses and adapt it to the purpose of this study. One of the major weaknesses of this model identified in the literature review is the length of the survey. The first adaptation therefore involved removing Reliability as a criteria group. This group includes stored customer profiles, order process, after-order to order receipt, and customer service. These are all criteria that I could not realistically simulate in an online survey without relying on respondents' memory and past experiences (if any) with the websites. Furthermore, accessing analytics data related to reliability of the site would require access to several e-commerce websites which I do not have, thus making the reliability dimension altogether inappropriate for this study. I then presented the rest of the criteria to 5 subjects that would be willing to participate in the research and have significant experience navigating and purchasing online. These discussions revealed that all potential respondents would have significant difficulties scoring criteria related to technical aspects of the site. The strength of heuristic evaluation models is the ability to gain insights into the users experience and opinions of the site, while technical aspects can be more easily and accurately measured using online tools. The Technical criteria group was therefore also left out in favour of a shorter evaluation survey that focuses on the most subjective measures of web success (Wiggins, 2008). Of the remaining criteria there were a number of terms that participants were not sure how to identify in the context of online shopping, I thus created a glossary (Appendix 2) that respondents could refer to while going through the survey to ensure that all questions were interpreted correctly and answered in the most accurate way possible.

When choosing which and how many websites to evaluate, I again followed the constraints of van der Merwe and Bekker's (2003) model. I explored industries that have several large important players that allow for fair comparison, that any potential participants would normally make use of, and that may increase the generalisability of findings to other large international sites. The travel & accommodation industry was found to be a good fit as it has many large online players that are internationally recognised. I used Semrush, an online marketing and SEO tool, to identify important competitors within the industry (Figure 1).

Figure 1: Growth quadrant for leading Travel & Accommodation websites

Source: <https://www.semrush.com/market-explorer/overview/?q=booking.com&searchType=domain>

Enough websites had to be included to be able to identify potential trends in relationships between analytics and user evaluations but still limiting the length of the survey by not including too many. I therefore chose to include 4 websites in the study from different parts of the growth quadrant (figure 1). The 4 chosen websites are booking.com, tripadvisor.com, expedia.com and kayak.com (Referred to by domain name from this point on, i.e. Booking). I then created an online survey using van der Merwe and Bekker's (2003) questions for each criterion pertaining to the interface, navigation, and content of all 4 websites. Each aspect contains 4 groups each with 5 criteria, resulting in a list of 60 criteria per website in an evaluation survey lasting approximately 1 hour. As the original model suggested at least 3 evaluators, 5 respondents were selected. Previous literature has also suggested using between 3-5 evaluators (Nielson & Mack, 1994; Pickard, 2007). To select the respondents, I sought contacts with different backgrounds and ages that have extensive experience in making online purchases. I ensured that they were willing and available to participate in a research survey lasting approximately 1 hour. The result is a sample of 5 participants between the ages of 20 and 30 from 4 different nationalities. The online survey was sent out to each participant on the same day along with the glossary of terms. They were asked to complete the evaluation within the same week to avoid any significant changes on the websites

between responses. Once all 5 responses were successfully completed, the data was exported into excel for analysis (Appendix 3). When designing the online survey, the criteria were coded as prescribed by van der Merwe & Bekker (2003), with strongly disagree corresponding to -2, to strongly agree corresponding to 2, each website could therefore score a maximum of 120 points in the evaluation.

3.2 Analytics Data

Gathering web analytics data of the 4 chosen websites presented significant challenges. I started by looking at the tools used in the literature. Google Analytics is one of the most diffused and well-studied web analytics tools. There are several other similar free software available, however, majority of these require access to the website's code in order to track detailed website data. This makes it difficult to access and impractical for comparing industry competitors. Instead, Kumar & Hasteer (2017) suggest several publicly accessible websites that allow for the collection of data of any website. Of the suggest tools, 3 of the 5 are still active and publicly accessible (Qualidator, Readable.io, Nibbler), however due to the exclusion of Google Analytics which now requires website access, these tools lack important metrics that could be found through other channels (for example visitors, visit duration, pager per visits, etc.). Thus, I used this list as a starting point to find other open analytics websites by searching the following keywords in Google; "free", "public", "analytics", "tool". I read the descriptions on the first 5 pages of google results and tested relevant results. Approximately 20 free accessible tools were found in this search. Majority of these only measured specific aspects of the website such as speed, readability, or SEO performance while a minority covered a wide range of web performance metrics, from the traditional SEO metrics to more abstract measures of web experience. A shortlist of 5 websites was thus created that cover the broadest possible range of website performance measurements (Table 3). Using multiple analytics tools allowed me to access the main metrics that have already been researched in the literature but also access new, broader metrics of website usability that have not garnered as much attention in research, such as the amount of content on a site, the confidence of different themes used, and the overall experience.

The data from the 5 tools was gathered on the same day for all 4 websites to maximise validity. Some of the analytics tools have overlapping metrics with conflicting figures, when comparing the survey results to these metrics, I first identified whether the online analytics tool was able

to provide an accuracy rating for the site (Semrush for example shows an accuracy rating of 3/3 for all 4 websites and was thus preferred to other tools for overlapping metrics). The Analytics data of each site was exported to Excel and compared to the evaluation data as described in the following section.

3.3 Analysis

When analysing the data from both of the sources described above, a thematic approach was used. Thematic analysis is a generic approach to qualitative analysis where themes or patterns are identified throughout the data to guide further analysis (Saunders, Lewis, & Thornhill, 2016). As the objective of this project is to explore the nature and depth of relationships between themes of the web environment and web performance metrics, the gathered data can naturally be categorised into the themes of web design identified in the literature (i.e. Interface, Navigation, and Content). The relationships proposed in the literature (Table 2) set the direction for the analysis; the data sets were compared by creating combination graphs in excel where the bars show the evaluation survey data while the line shows the analytics data. Presenting the data in this way facilitates the recognition of trends and patterns between the two types of data, as can be seen in Figure 3. This method was inspired by Hasan, Morris & Proberts (2013), who use a clustered column chart to compare data from different evaluation methods. Where possible, the 2 measures are compared on a single vertical axis (see Figure 3) so that the nature of the trends are not distorted by differences in scales. However, for many analytic measures such as visit duration, visit volumes and metrics measured as percentages, a second vertical axis has to be introduced in order to make the graphs readable (See Figure 6). For these comparisons, emphasis was placed on percentage differences between different websites to provide the reader with an understanding of the true extent of the relationships depicted in the graphs rather than relying on visual comparisons that depend on scale choices. After identifying the graphs that initially appeared to show significant relationships, the evaluation scores were put in order from highest to lowest for these graphs to further improve the readability of the data. This step revealed that some of the graphs that initially appeared to show significant trends in fact did not show patterns when put in order.

Having familiarised myself with the trends in the data, I was able to first focus on the themes that appeared to be highly related to the analytics data (such as Interactivity) to search for additional metrics that could further explain, support, or even contradict the established

connections. Once these were analysed, the themes with little to no patterns were compared to additional metrics to attempt to gain a better understanding of why the suggested relationships may not be present. The findings on significant patterns recognised in the data are discussed thoroughly in section 4. For each web dimension, the overall score and individual sub-categories were analysed so that all potential relationships are accounted for. Finally, all the identified relationships per web attribute are summarised in Appendix 4, visualising these in a single table helps draw conclusions about how and to what extent the analytics are able to describe the different aspects of the website.

4. Results

This chapter describes the important results found throughout the analysis of the data gathered during this research. As the data is divided and tested along the web dimensions described in van der Merwe & Bekker's (2003) model, the results are also presented in this order. For each theme, the evaluation scores for each website and sub-category are shown and then compared to the gathered web performance metrics to see whether these relationships support those found in recent literature or if there are other measures that have not yet been discussed in the literature. The tables in Appendix 5 summarise all the data mentioned in this chapter.

4.1 Interface Results

Figure 2 summarises the user evaluations of all Interface categories for the four chosen websites. The total scores for the category have a relatively narrow range with the top website (Booking) scoring only three points higher than the bottom website (Tripadvisor). However, when looking at the separate categories there are clearly aspects of the site that vary significantly. Graphic Design Principles and Flexibility & Compatibility show the greatest variations between websites with standard deviations of 1.78 and 1.43, respectively. Graphic Design Principles and Style & Text are the categories with the overall highest average scores (7.1 and 7.7 out of 10) across all websites. Comparing this data to different web metrics gives further insights and potential explanations for the mentioned disparities.

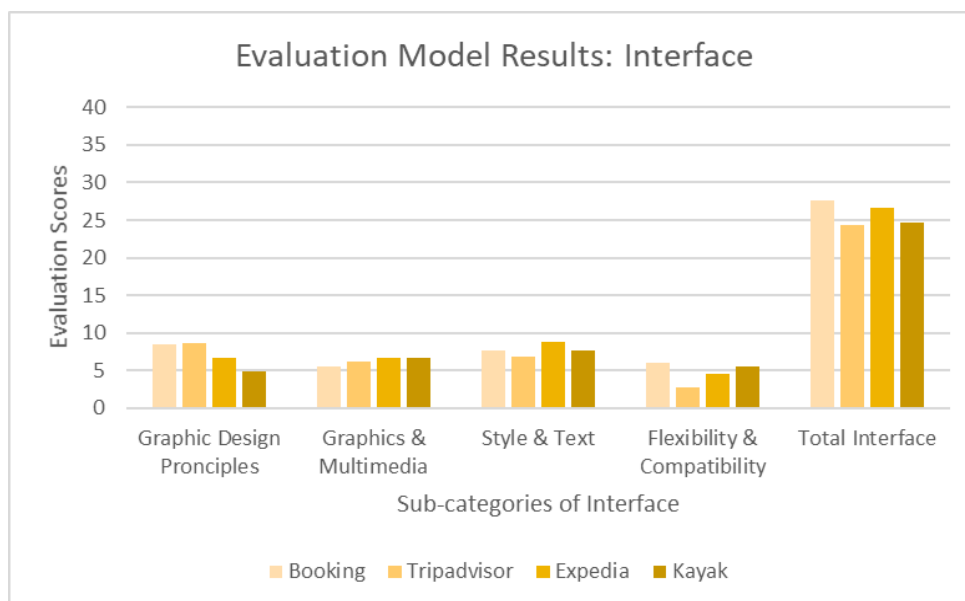


Figure 2

The literature reviewed in chapter 2.3 suggest Interface can be measured with the following metrics: pages per visit, visit duration, bounce rate, and load speed. All of these metrics revealed connections to the evaluation data in at least one sub-category or overall interface scores.

Pages per Visit Decrease with Interface Evaluation Scores

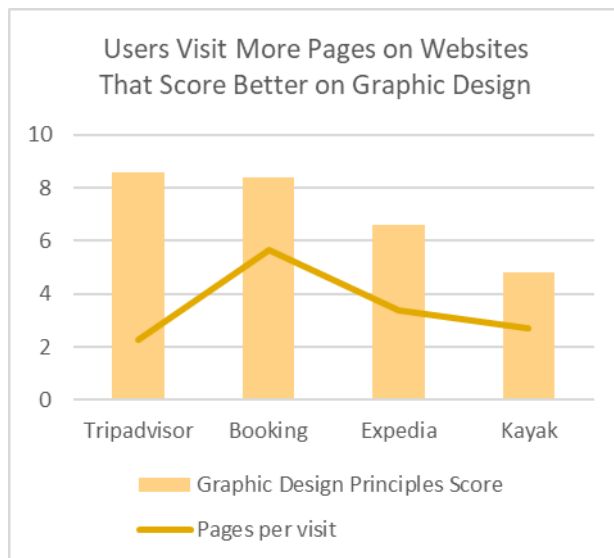


Figure 3

The average number of pages visited (dark yellow line in Figure 3) generally correspond to the evaluators' perceptions of the effectiveness of the websites' use of graphic design (light yellow bars in Figure 3). This trend would support the notion that websites with a higher number of pages per visit are perceived more positively by users in terms of graphic design. Tripadvisor is the only exception to this rule having the lowest number of pages per visit (2.26) yet the highest evaluation score (8.6).

The average number of pages per visit also corresponds to the perceived Flexibility & Compatibility of the website (Figure 4), although again, showing an exception to the trend (Kayak). In this sub-category Tripadvisor's low number of pages per visit is reflected in a low evaluation score. However, Kayak scored relatively high in Flexibility & Compatibility compared to the recorded pages per visit.

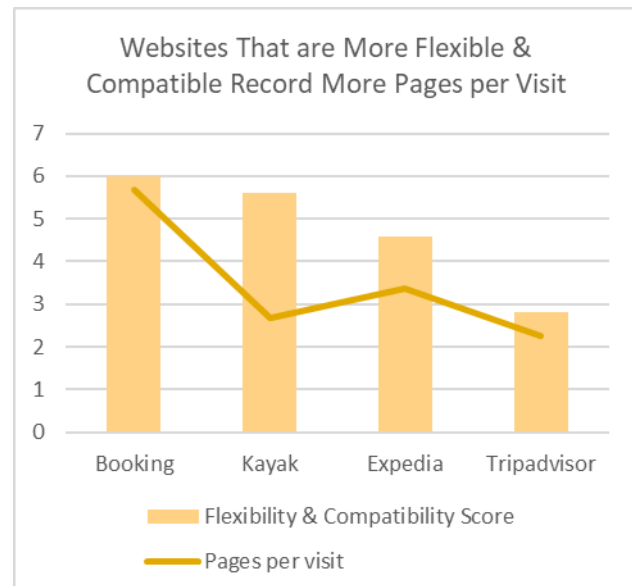


Figure 4

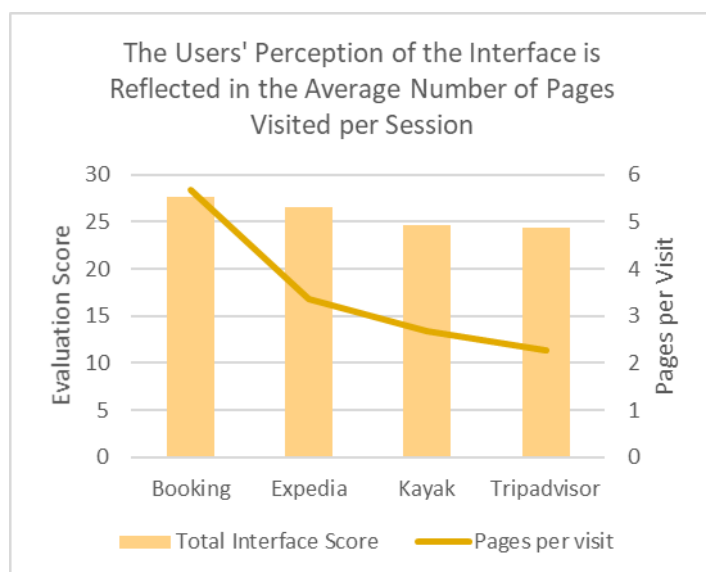


Figure 5

Despite this performance metric not showing a significant relation to two of the four sub-categories of Interface, the pages per visit follow the total Interface scores for all four websites tested (Figure 5). However, the analytics do not appear to change proportionately to the evaluation scores. This is most apparent for Booking and Expedia, where a difference in evaluation score of 3.7% is reflected in a 51.2%

difference in average number of pages per visit. Similarly, a minute difference in scores between Kayak & Tripadvisor of 0.82% is paired by a 17% difference in number of pages per visit.

The Time Spent on a Site is Positively Related to Interface

Comparing the Interface categories to the average amount of time spent on the site yields similar results to those of the number of pages visited. Like the pages per visit, the visit

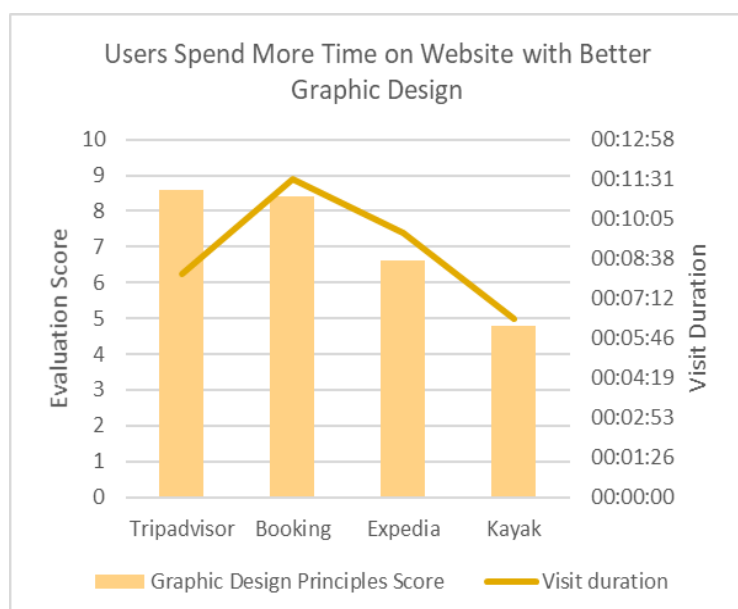


Figure 6

duration also showed connections to the Graphic Design Principles and Flexibility & Compatibility scores of the evaluation model, with both showing the same exceptions for the corresponding sub-categories (Tripadvisor for Graphic Design Principles, and Kayak for Flexibility & Compatibility). In particular, Figure 6 shows visit duration to seemingly be able to accurately predict

perceptions of design, with duration being proportionate to three out of four evaluation scores. For these websites conforming to the general trend, the total difference in score

between Booking and Expedia of 56.3% is mirrored with a difference in visit duration of 54.5%. In addition, the visit duration changes relatively linearly to match the almost linear decrease in evaluation scores for Graphic Design Principles.

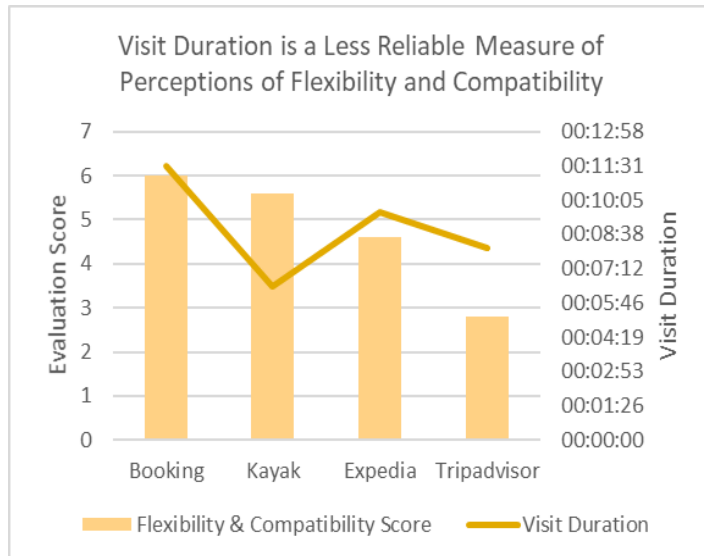


Figure 7

more dramatically between websites than the visit duration. The difference between the top and bottom rated websites from the evaluation of 72.7% is reflected in a proportionately smaller difference in visit duration of 35.4%.

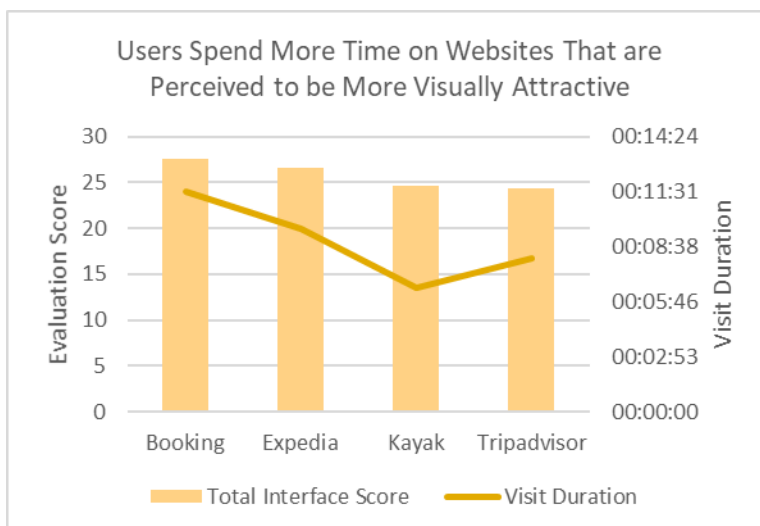


Figure 8

while shorter durations corresponded to lower evaluation scores. Kayak contradicts this trend, however, scoring better in the evaluation than Tripadvisor despite having shorter visits. Contrary to the findings for Flexibility & Compatibility, In the case of the total Interface score,

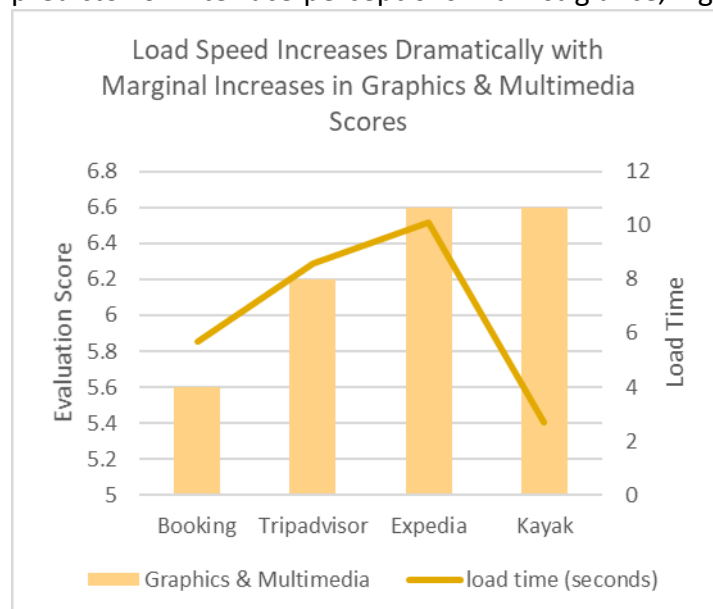
In contrast, the visit duration appears to be a slightly less reliable measure of Flexibility & Compatibility where only two of the four websites measure proportionately while Tripadvisor scores relatively lower in the evaluation compared to the average visit duration. From Figure 7 it appears the evaluation scores for Flexibility & Compatibility change

The average visit duration was not found to be meaningfully related to Graphics & Multimedia and Style & Text choices. Nonetheless, the metric appears to reflect the overall perceptions Interface to some extent (Figure 8). Websites with longer average visit durations generally scored better in the user evaluations

differences in visit duration appear to exaggerate the differences in evaluation scores. Booking's average visit duration is 35.4% longer than that of Tripadvisor while the score is only 12.3% higher.

Load Speed is Not Related to Aspects of Interface

Contrary to what the literature suggests, the speed of the site did not appear to be a good predictor of Interface perceptions. At first glance, Figure 9 may look like websites that load



faster scored better in the evaluation, however, the left vertical axis shows the difference in scores between websites to be marginal compared to the load times. With only one point difference between the top and bottom scores for this sub-category, it is thus difficult to conclude whether potential trends identified are significant or arbitrary.

Figure 9

The Bounce Rate is Inversely Related to Total Interface Scores

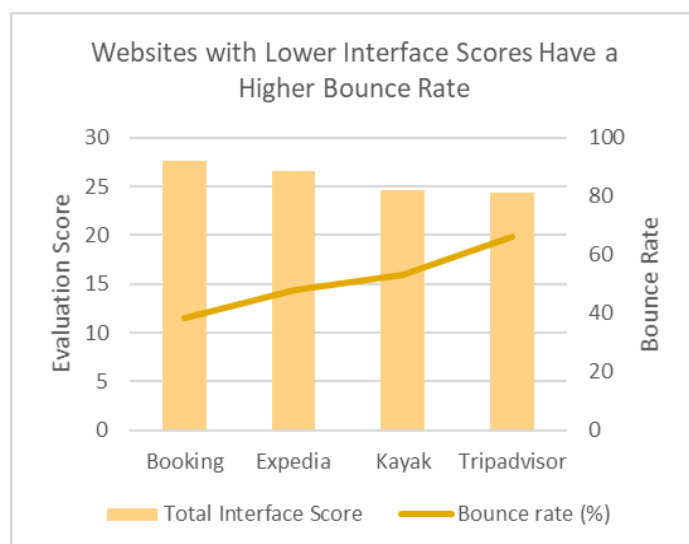


Figure 10

The final performance metric suggested in the literature as a measure for the successful design of the web Interface is the bounce rate. Bounce rate is the percentage of visitors leaving as soon as they land on the website and is thus desired to be as low as possible. Although the bounce rate did not show significant trends with any sub-categories of Interface, Figure 10 shows that the

performance metric moves opposite to the total interface score from the evaluation, suggesting an inverse relationship. The bounce rate is not, however, directly proportionate to the evaluation scores. Small differences in scores such as those between Booking and

Expedia (3.7%), and Kayak and Tripadvisor (0.82%) exhibit large differences in bounce rate (9.6% and 12.7%, respectively).

Additional Predictors of Interface Success

Beyond the web analytics suggested in the literature, additional metrics that could be gathered online were tested to propose new, unexplored links.

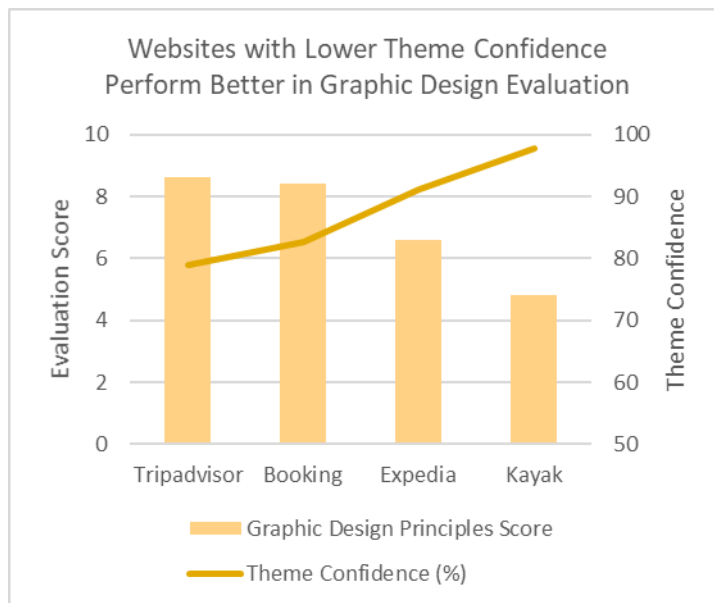


Figure 11

The theme confidence of each website was found to be related to three of the four sub-categories of Interface evaluation. The metric is negatively related to scores for Graphic Design Principles (Figure 11); websites with lower theme confidence scored higher in the evaluation. However, the differences in website scores are not proportionately reflected in the measure of theme confidence. The

2.4% difference in score between Tripadvisor and booking closely corresponds to the 4.5% difference in theme confidence but the larger 32% difference in evaluation scores between Expedia and Kayak is matched with a 7% difference in theme confidence.

Instead, the metric is positively related to the evaluation scores for Style & Text (Figure 12), and Flexibility & Compatibility (Figure 13) where higher evaluation scores correspond to higher theme confidence. For Style & Text, Kayak contradicts the trend recording the highest theme confidence while scoring the same as Booking in the Evaluation. Furthermore, for the

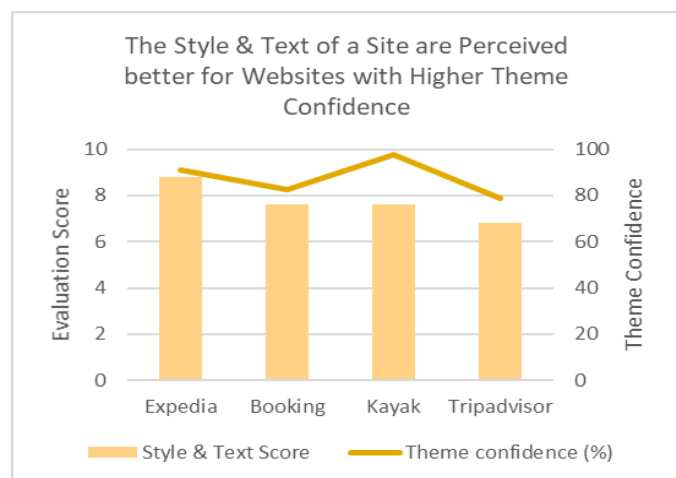


Figure 12

Websites that do fit the trend, the differences in evaluation scores are slightly underestimated

by changes in theme confidence. In the evaluation, Expedia scores 25.6% higher than Tripadvisor, yet the difference in theme confidence is only 12.2%.

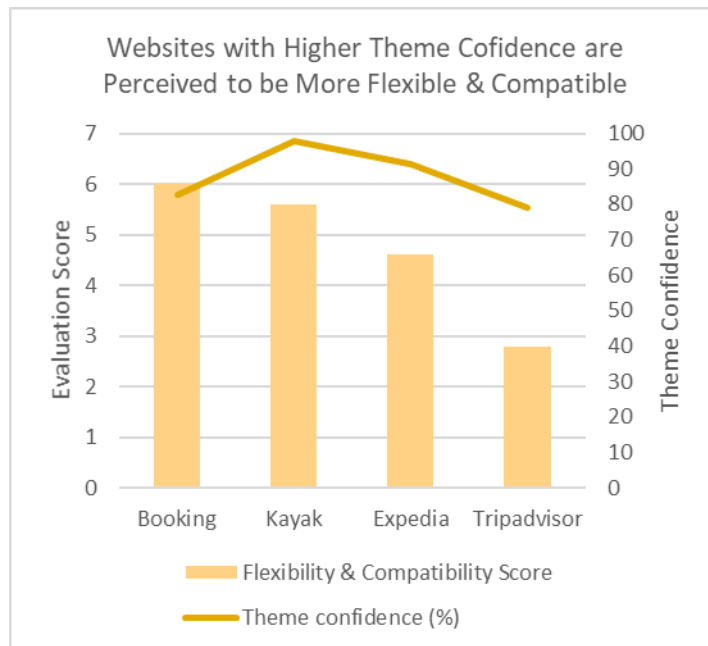


Figure 13

Instead, in the case of Flexibility & Compatibility, Booking exhibits the second lowest theme confidence despite having the highest evaluation score. Similarly to the relation to Style & Text the differences in evaluation scores are greater than those of theme confidence. The percentage change between Kayak and Tripadvisor is 3.5 times greater for evaluation scores than for theme confidence. Theme confidence was not found to be

related to the total interface evaluation scores despite being somewhat related to most sub-categories.

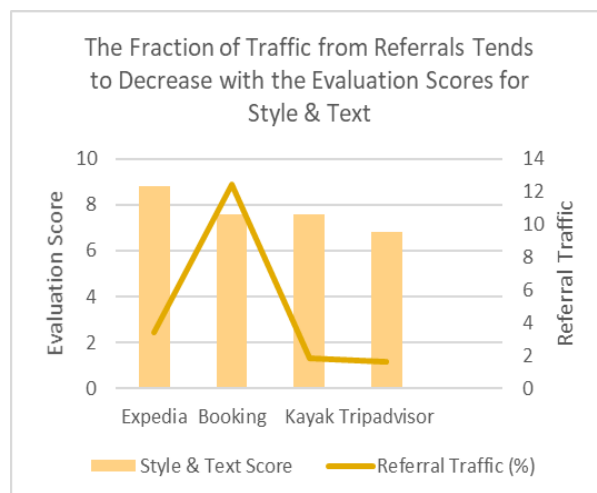


Figure 14

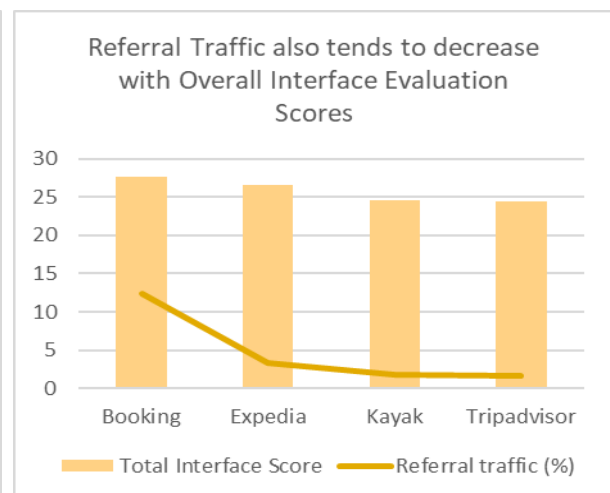
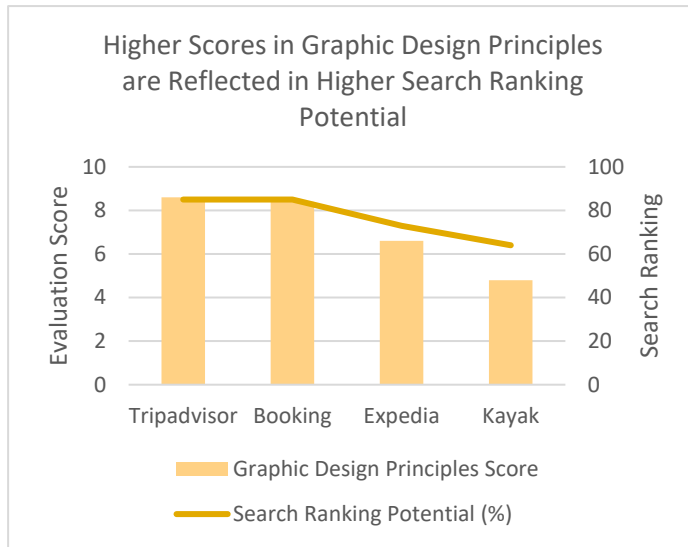


Figure 15

The amount of referral traffic as a percentage of all traffic sources was found to relate to the evaluation scores for Style & Text (Figure 14) as well as the total evaluation scores for Interface (Figure 15). Websites scoring higher in the evaluation exhibit greater proportions of referral

traffic. In both cases, the referral traffic of Booking is grossly exaggerated compared to its rivals.



Finally, the search ranking potential is closely related one sub-category of Interface for all websites. As seen in Figure 16, the analytics metric decreases relatively parallel to the evaluation scores for Graphic Design Principles.

Figure 16

4.2 Navigation Results

The evaluation scores from Navigation are more levelled than those of Interface (Figure 17). The greatest source of variance between total Navigation scores comes from the Logical Structure evaluation scores. The literature discussed in section 2.3 suggests six basic web analytics to measure the success of a website's navigation; these are visits, return visits, visit duration, pages per visit, bounce rate, and organic traffic. This section will discuss how these analytics are reflected in the evaluation results and potential new metrics that could be used.

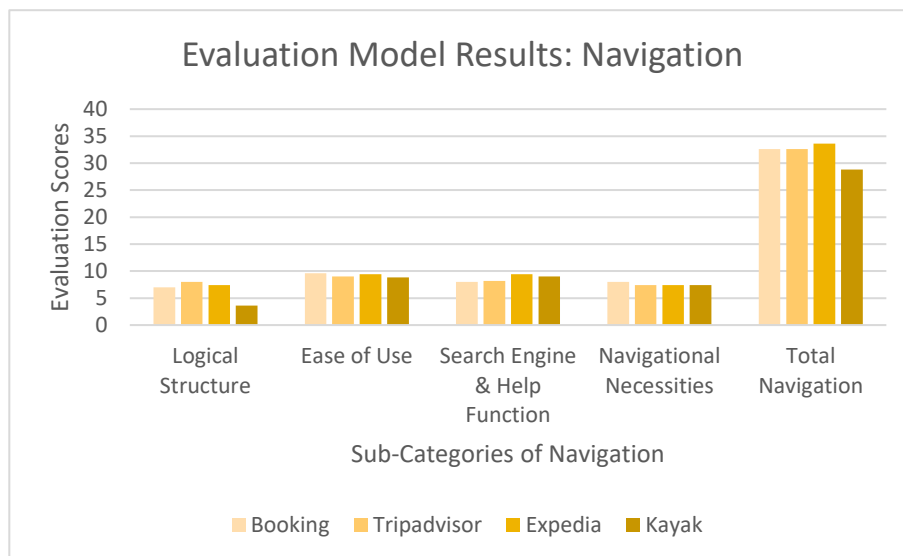


Figure 17

Visits and Return Visits Change with Logical Structure Scores

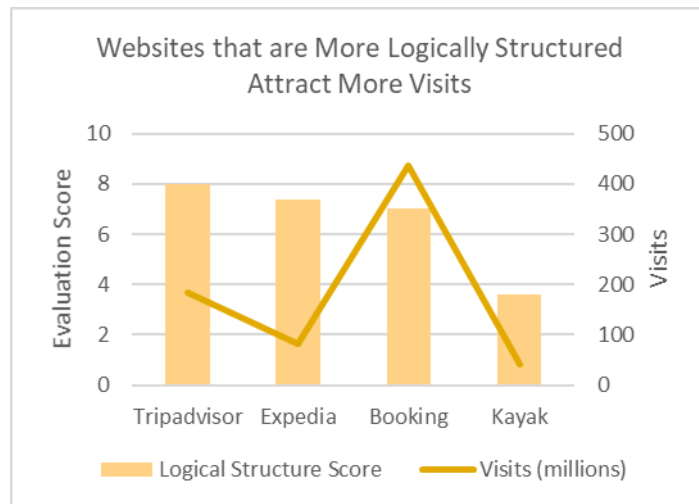


Figure 18

second lowest score in the evaluation. Additionally, both metrics change relatively proportionally to the change in scores between Expedia and Kayak. Expedia scores 69.1% higher than Kayak in the evaluation for Logical Structure and achieves 63.8% more total visits and 74.3% more return visits. However, both measures overestimate Tripadvisor's evaluation

score recording 76.7% more visits and 74.3% more return visits than Expedia, which only scored 7.79% lower than Tripadvisor in the evaluation.

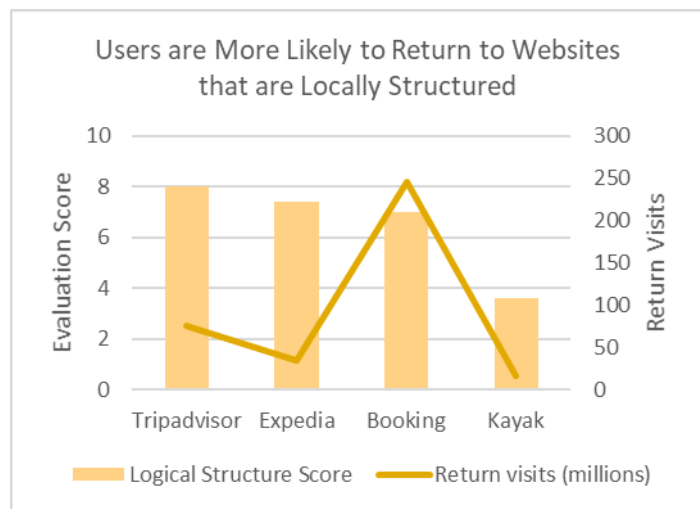
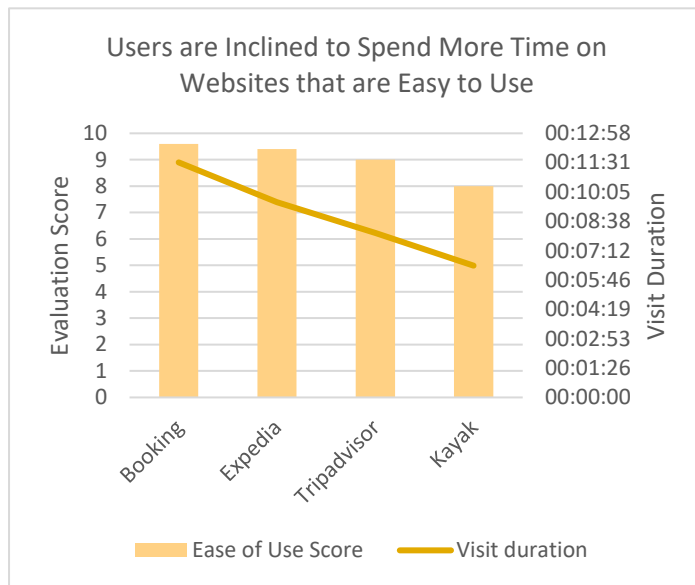


Figure 19

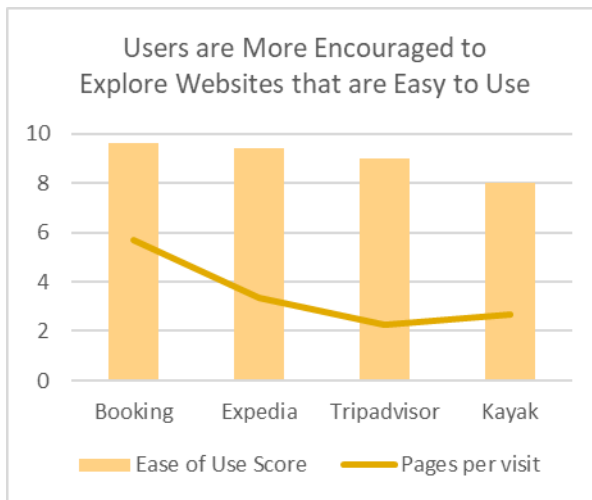
The Visit Duration Decreases with Ease-of-Use Scores

The average time spent on the tested websites corresponds to the perceived Ease of Use from the evaluation. Booking and Expedia received higher scores and measure longer visit durations while Tripadvisor and Kayak received lower scores and experience shorter visit durations (Figure 20). It should be noted that the range in scores for Ease of use is only 0.8 points, limiting the replicability of results. In comparison there are 5:04 minutes between the longest average visit duration (Booking) and the shortest (Kayak).

Figure 20

Pages per Visit Increase with Ease of Use

The average number of pages per visit was only found to relate to the user friendliness of the site (Figure 21). Kayak stands out as contradicting the trend with a higher number of pages per visit than Tripadvisor despite obtaining the lowest score in the evaluation. In addition, as the measures can be shown on the same axis, it is clear that the changes in pages per visit are relatively large compared to small changes in evaluation scores for Ease of Use.

**Figure 21**

The Bounce Rate Inversely Related to Logical Structure Scores

The evaluation scores for logical structure also reflect the measured bounce rate for three of the four websites (Figure 22). The data suggests that websites that are perceived to be structured more logically have lower bounce rates. Tripadvisor is the only website contradicting this trend with a disproportionately high bounce rate despite receiving the best evaluation score. For the websites that do conform the the trend, the bounce rate appears to

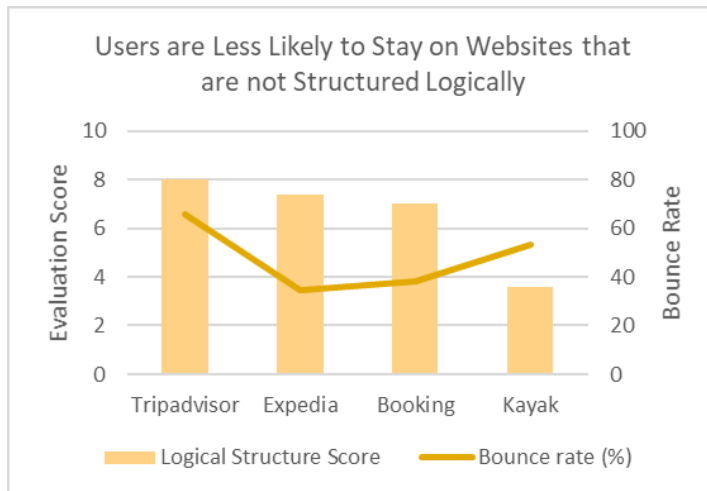


Figure 22

change less dramatically than the evaluation scores; the 75.9% difference in evaluation scores between Expedia and Kayak is paired with an 18.6% difference in measured bounce rate.

Organic Traffic Decreases with Logical Structure Scores

Lastly, the data suggests the organic traffic to a website is related to the site's structure.

Websites scoring higher on Logical Structure receive higher volumes of organic traffic. Expedia

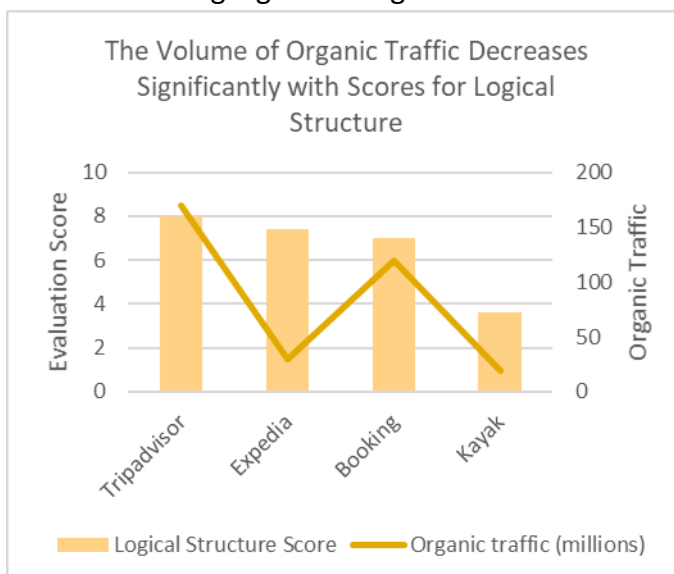


Figure 23

is the only exception to this rule with a relatively low organic traffic volume compared to its evaluation score (Figure 23). Unlike bounce rate, the volume of organic traffic changes more dramatically compared to the evaluation scores where the overall 75.9% difference between top and bottom scores corresponds to a 158.9% difference in volume of organic traffic.

Additional Measures of Navigation

As with Interface, several additional analytics were found to relate to the evaluation data. The level of accessibility for each site follows the evaluation scores for Ease of use (Figure 24). Expedia is the clear outlier as its accessibility level peaks above its rivals despite scoring lower than Booking in the evaluation. For the remaining websites the overall percentage difference in evaluation scores (18.2%) is triple the difference in accessibility levels (6%).

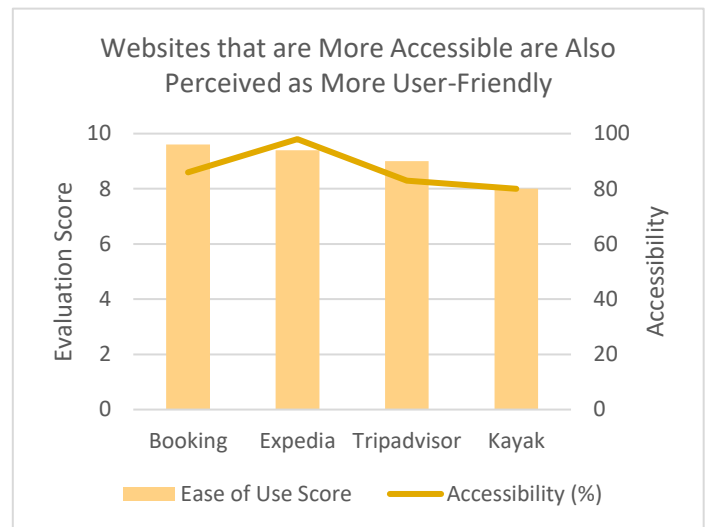


Figure 24

The market share appears to relate to both the Logical Structure of the website and the total navigation (Figures 25 and 26). The websites receiving higher scores in the evaluation are matched with higher market shares. Tripadvisor, however, contradicts this trend for both categories, achieving high scores in the evaluation while recording the lowest market share.

Figure 25

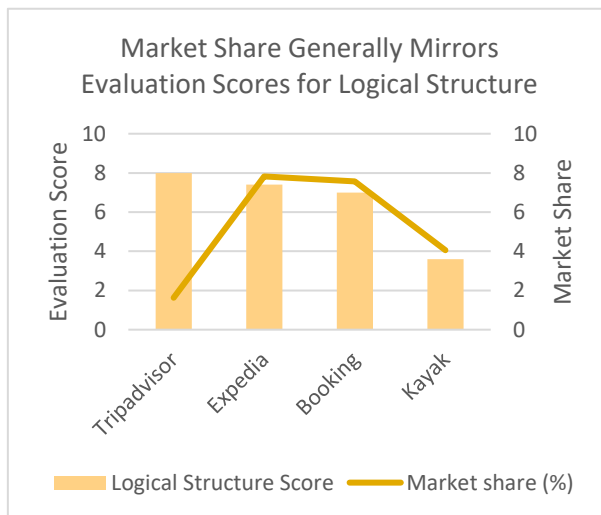
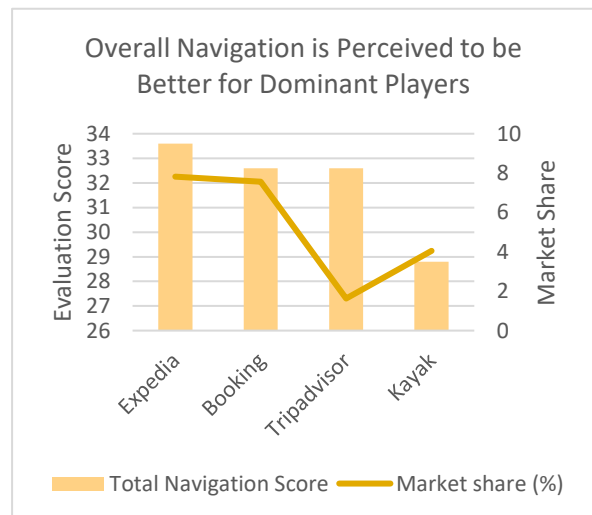


Figure 26



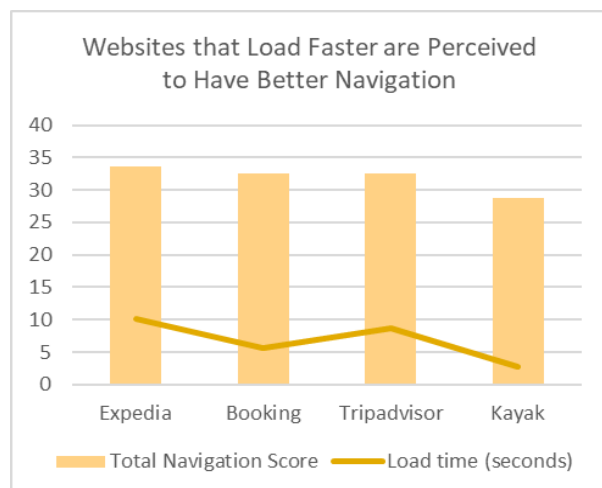


Figure 27

Within the theme of Navigation, the Logical Structure contributed to the majority of the connections found between the evaluation model and corresponding analytics. Trends among other sub-categories were difficult to identify due to the close proximity of scores between websites.

4.3 Content Results

Content is the theme with the greatest variance among websites for each sub-category (Figure 28). Aside from Information Quality which only has a range of 1.6 points, the scores for each sub-category have a range of at least 4 points, allowing for trends to be recognised more easily. Expedia and Kayak's significantly lower scores for Company & Contact information and Interactivity led to low scores in the overall perception of the websites' content. In this section, we again explore if and how these evaluation scores are reflected in the analytics suggested by the literature and which other metrics could potentially be used to measure user perceptions of content.

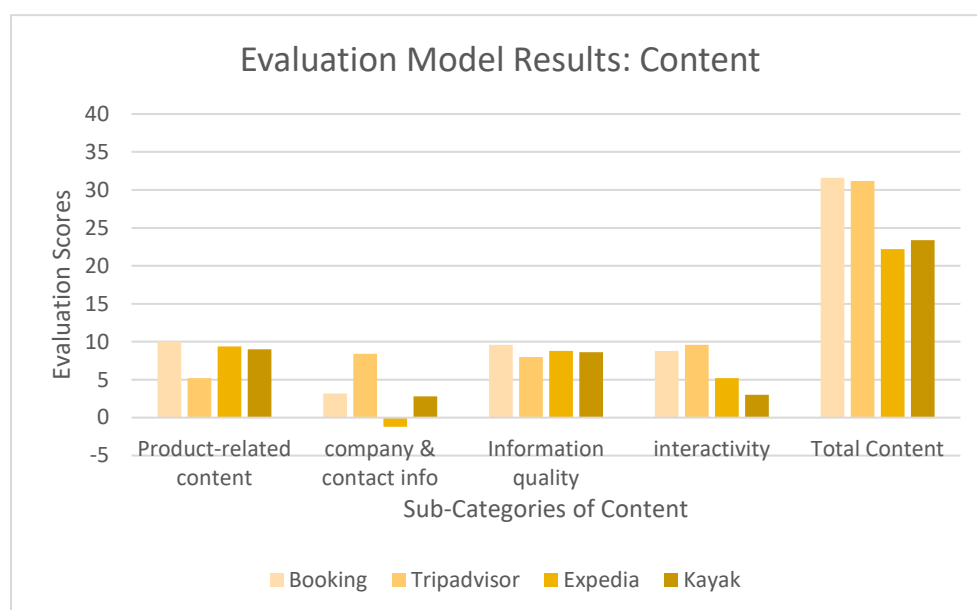


Figure 28

Visit Duration is Related to Majority of the Sub-Categories of Content

The average time spent on a website generally seemed to be a good predictor of perceptions of website content, matching trends in scores for three out of four sub-categories of content in the evaluation model. For all of these categories, majority of the websites scoring higher in the evaluation also recorded longer visit durations.

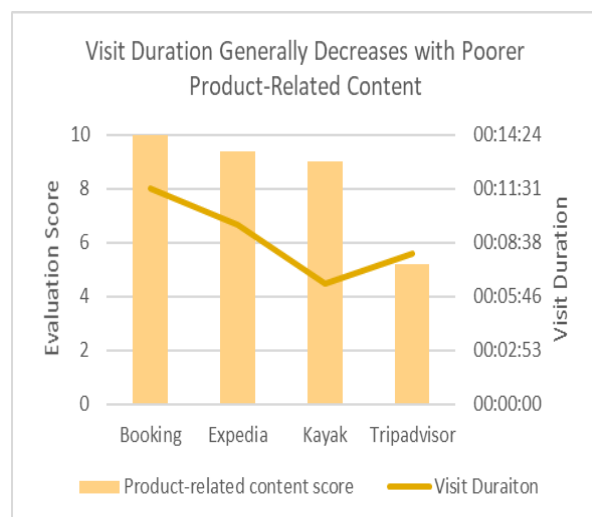


Figure 29

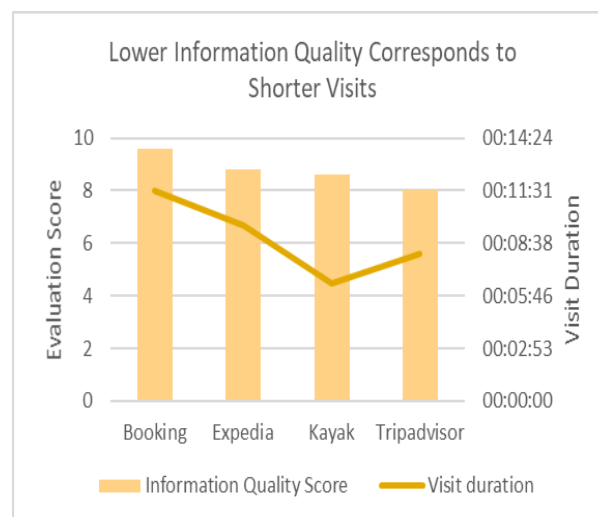


Figure 30

In terms of Product-Related Content (Figure 29), however, Kayak's evaluation score is relatively high compared to its short visit duration of just 06:28 minutes, while Tripadvisor's score is proportionately low compared to its longer average visit of 08:04 minutes. For Information Quality (Figure 30) Kayak again contradicts the trend with the shortest visit duration despite scoring slightly higher than Tripadvisor in the evaluation. Besides Kayak,

however, Figure 30 shows the visit duration to decrease in proportion with the evaluation scores for Information Quality.

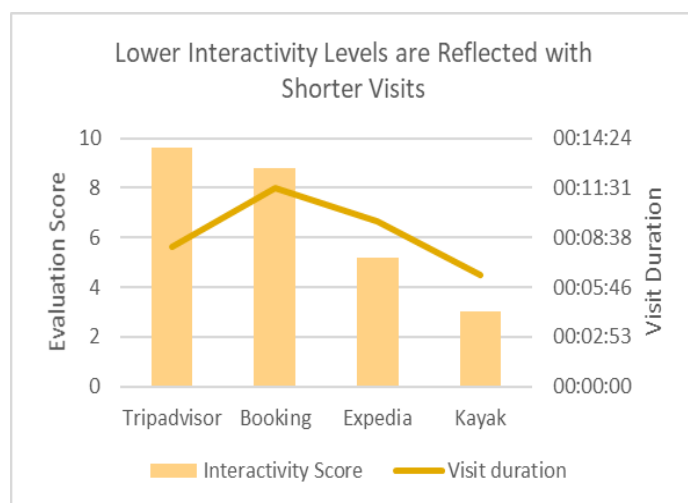


Figure 31

Instead, in the case of Interactivity, Tripadvisor opposes the relationship with a disproportionately high evaluation score despite the relatively short average visit duration. Furthermore, the Interactivity scores (the bars in Figure 31) conforming to the trend decrease more radically than the corresponding visit durations.

Despite visit duration being related to some extent to majority of the sub-categories of Content, there was no apparent trend found between the overall perceptions of a website's content and the time spent on the site.

Return Visits Decrease with the Perceived Interactivity of the Site

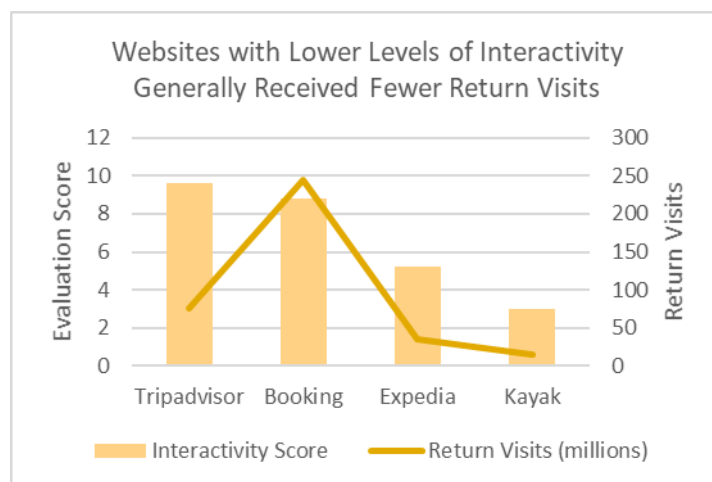


Figure 32

Return visits were only found to relate to one sub-category of content. As seen in Figure 32, the return visits tend to decrease with decreasing Interactivity scores with the exception of Booking, which attracts over three times the amount return visits than Tripadvisor despite scoring slightly lower in the evaluation. In addition,

the differences in recorded return visits appear to be more pronounced than the corresponding differences in interactivity scores; The number of return visits between Tripadvisor and Expedia, and Expedia and Kayak both show a decrease of 74% while the evaluations only decrease by 54% and 60% respectively.

Referral Traffic is Greater for Websites Scoring Higher

The referral traffic as a fraction of total traffic shows similar trends to visit duration when compared to users' perceptions of content. Figures 33 and 34 show how the percentage of traffic originating from referrals decreases with the evaluation scores for both Product-Related Content and Information Quality. In both cases however, Booking exhibits a disproportionately higher level of referral traffic of 12.44% while only scoring 0.6 points higher (for Product-Related Content) and 0.8 points higher (for Information Quality) than the next best website. In Addition, although both sub-categories show evaluation scores and referral traffic decreasing together, they are not proportionately related. Both graphs show greater differences in the percentage of referral traffic between Expedia and Kayak than between Kayak and Tripadvisor despite the difference in score being greater for the latter two websites.

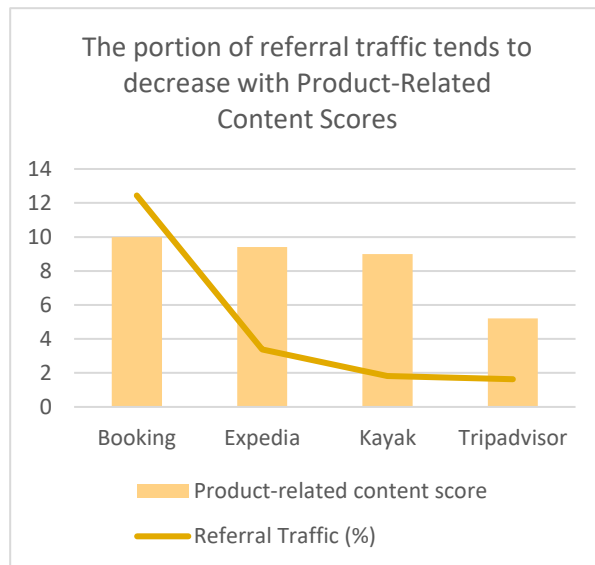


Figure 33

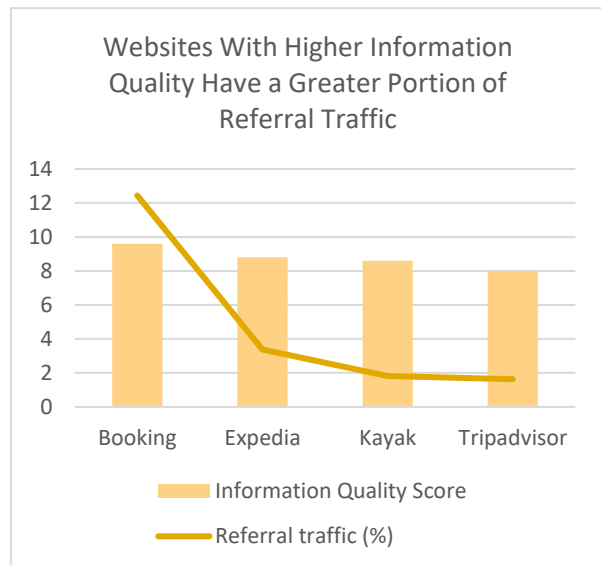


Figure 34

The results for Interactivity again illustrate referral traffic to decrease with the evaluation scores and show Booking's score to be exaggerated compared to users' perceptions of the site's interactivity (Figure 35). However, for this sub-category, Tripadvisor contradicts the trend by achieving the highest score in the evaluation while measuring the lowest level of referral traffic. Specifically in the case of Interactivity, comparing the evaluation scores to all the analytics suggested by the literature reveals that in each case, Tripadvisor is the only website that does not conform to the theorised relationships.

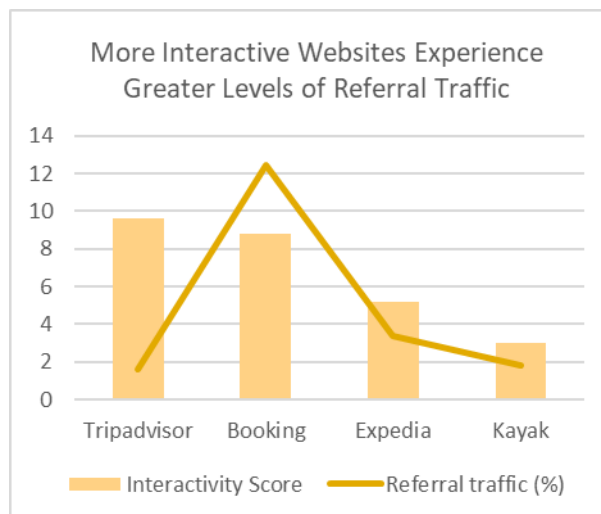


Figure 35

Additional Metrics for Content

Similarly to Navigation, there are recurring metrics that were found to apply to content as well, despite not having been suggested in the literature as of yet. These include pages per visit (Figures 36, 37, and 38), bounce rate (Figure 40), and organic traffic (Figure 41). Beyond these, the additional metric of page size was found to be uniquely related to the overall content scores (Figure 48).

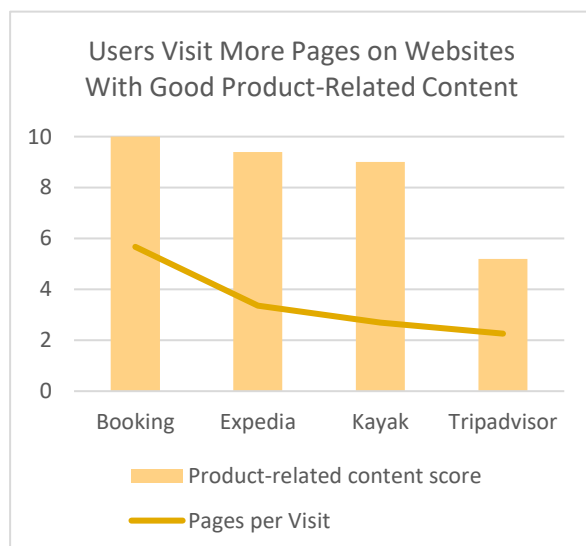


Figure 36

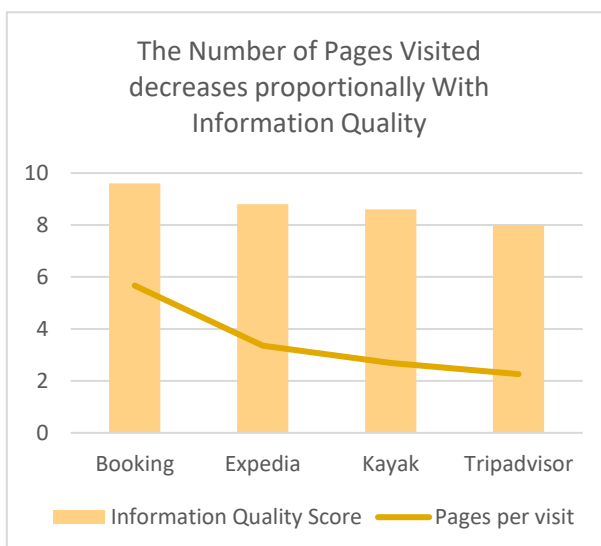


Figure 37

The average number of pages visited per session are seen to decrease in tandem to the evaluation scores for both product-related content and information quality. Looking closely

at the results for the product-related content scores, the average number of pages per visit for Tripadvisor is not proportionate to the evaluation score; The decrease in product-related content score between Kayak and Tripadvisor is far steeper than the difference in the number of pages visited, while for booking this difference appears to be overestimated with a steeper dark yellow line relative to the small difference in evaluation score to its next best competitor (Figure 36). Regarding the Information Quality, the same observation can be made for Booking exhibiting a relatively high number of pages visited, however, the average number of pages visited for the remaining websites appear to decrease relatively parallel to the evaluation Scores in Figure 37.

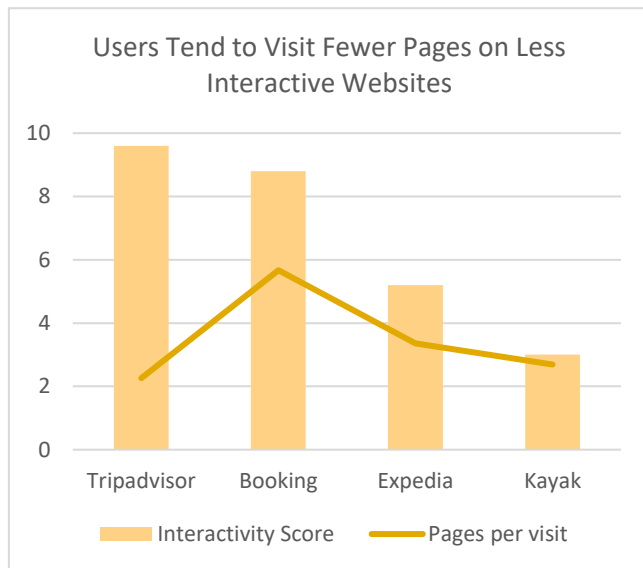


Figure 38

As for the measures for interactivity suggested in the literature, the average pages visited reveal Tripadvisor to be a contradiction to the trend for this sub-category achieving the highest score in the evaluation while recording the lowest number of pages visited (Figure 38). The remaining websites exhibit a decreasing number of pages per visit with decreasing evaluation scores. Between Booking and Expedia, the difference in both measures is proportionate, both falling by 51%,

while for Expedia and Kayak the decline in evaluation score is steeper (53%) than that of the pages visited (22%).

Figure 39 illustrates how the bounce rate is inversely related to the users' perception of the quality of information on a website. The difference in information quality scores between websites is reflected in a relatively similar difference in bounce rate. Booking scores 18.18% higher in the evaluation than Tripadvisor, this is reflected in a difference in bounce rate of 27.79%.

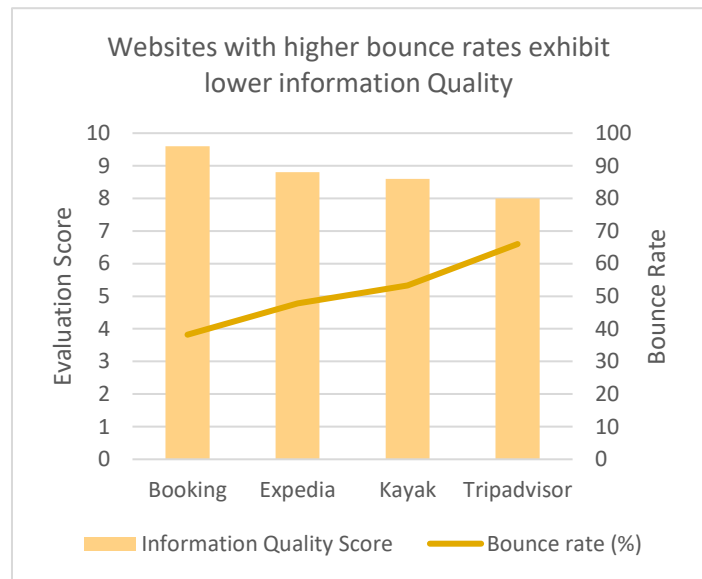


Figure 39

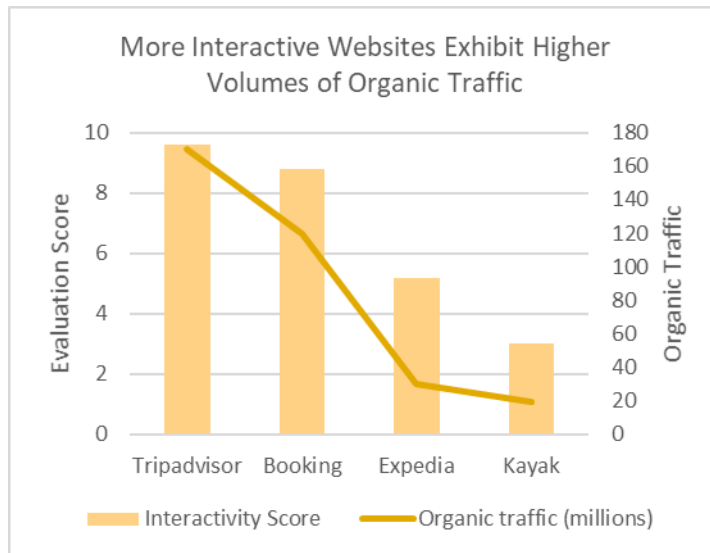


Figure 40

Interactivity was found to be the category with the most relationships to different analytic performance measures within the theme of web content. Figure 40 illustrates how the volume of organic traffic in millions decreases with the interactivity scores of the different websites from the evaluation model. While the changes seem relatively proportionate in the graph, the organic traffic changes dramatically with relatively small changes in evaluation score. The 18.18% difference between the highest and lowest evaluation score is far smaller than the 158.86% difference in organic traffic.

In contrast, very few metrics were found to be related to the total content scores in the evaluation and none were found to reflect trends in company and contact information available on the site.

Figure 41 illustrates the only additional metric found to be somewhat related to the total content scores. Websites with larger page sizes (Booking and Tripadvisor) generally received higher total content scores. Expedia contradicts this trend, however, with the lowest score despite recording the largest page size.

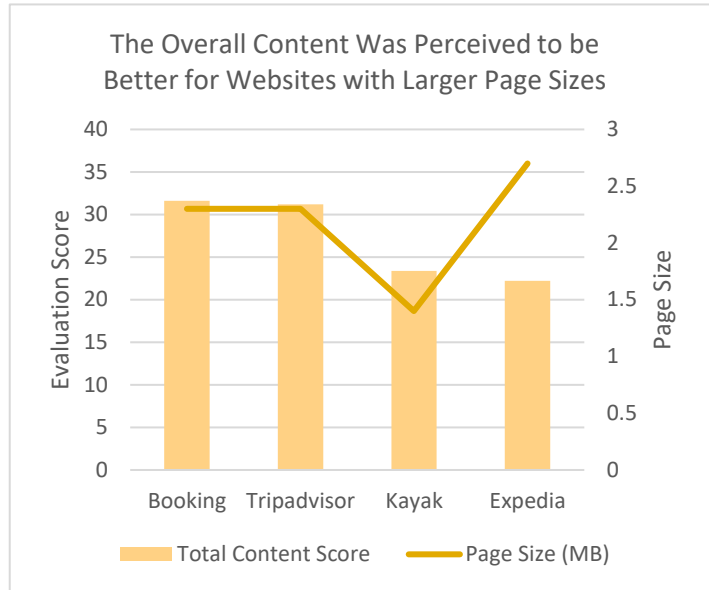


Figure 41

4.4 Aggregated Results

Finally, the scores can be aggregated to see whether each web performance metric is effective at recognising potential flaws in specific parts of the website or whether they are rather a measure of the overall performance of the website. Looking at the aggregated survey scores (Figure 42), navigation has the highest average score (31.9), while Interface scores the lowest on average (25.8). In addition, Interface and navigation both show relatively small variations in scores between websites, measuring standard deviations of 1.56 and 2.12, respectively. In contrast, Content records a greater range of scores between websites with a standard deviation of 4.99. The full contingency matrix is found in Appendix 6.

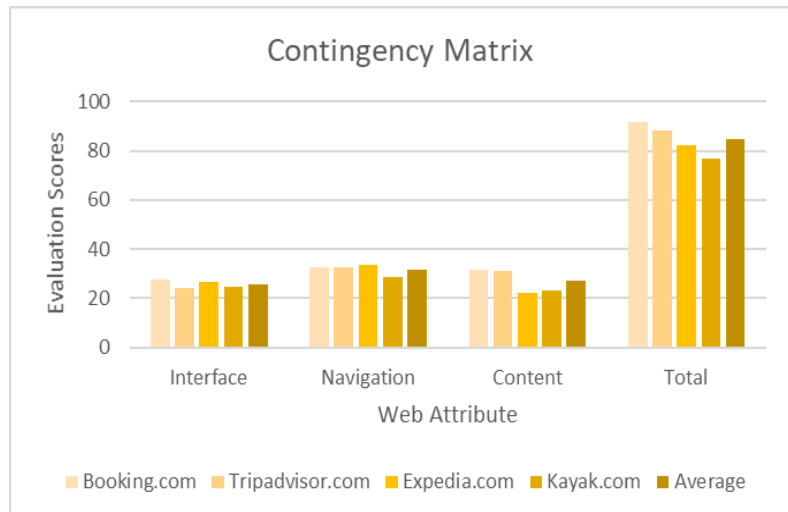


Figure 42

Several accessible analytics suggested in the literature (highlighted in Table 2) that show trends for themes of the web environment also relate to the aggregated evaluation scores. The volume of visits and return visits, for example, which are related to logical structure (Navigation) and interactivity (Content) can also be seen decreasing along with the overall evaluation scores (Figures 43 and 44). In both graphs, however, the volume of visits (dark yellow curve) appears to decrease at a decreasing rate, meaning the curve becomes shallower, while the difference in evaluation scores is more linear.



Figure 43

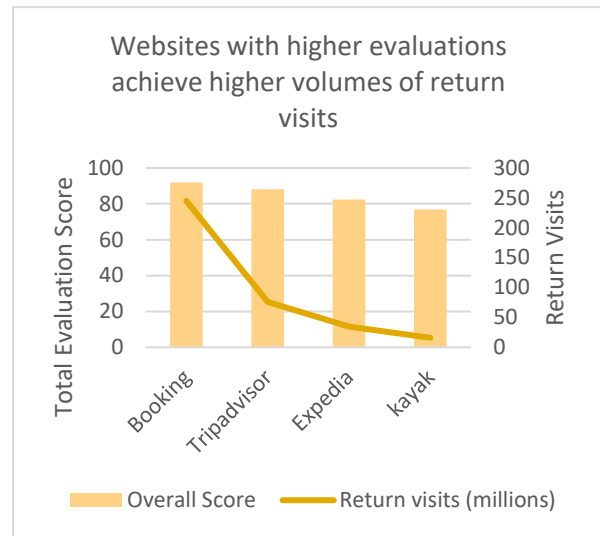


Figure 44

The visit duration was determined to be linked to some extent to all three aspects of the website measured in the evaluation and thus unsurprisingly also links to the overall evaluation scores. Figure 45 illustrates how the visit duration becomes shorter with lower evaluation scores, with the exception of Tripadvisor whose second highest evaluation score corresponds to the second shortest visit duration.

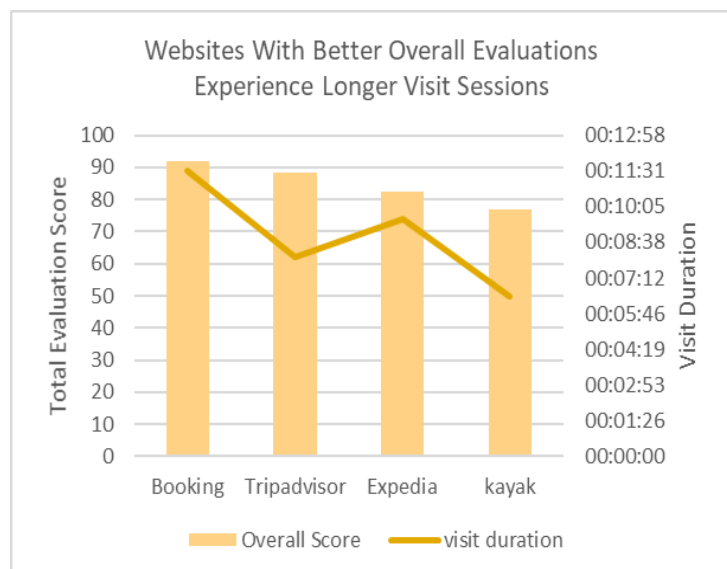


Figure 45

The pages per visit, organic traffic, referral traffic and bounce rate are also metrics from the literature that appear to be related to the overall scores. For all measures other than visits, Tripadvisor contradicts the trends. The figures for the following metrics thus exclude the data for Tripadvisor to show the relationship more clearly, the same results graphs with Tripadvisor's data can be found in Appendix 7.

Figure 46 exhibits the relationship between the overall score from the evaluation and the average number of pages visited. As with the visit duration, the analytics metric decreases dramatically in proportion to changes in evaluation scores; while the difference in evaluation scores range from 7-10% the change in pages per visit range from 22-52%.

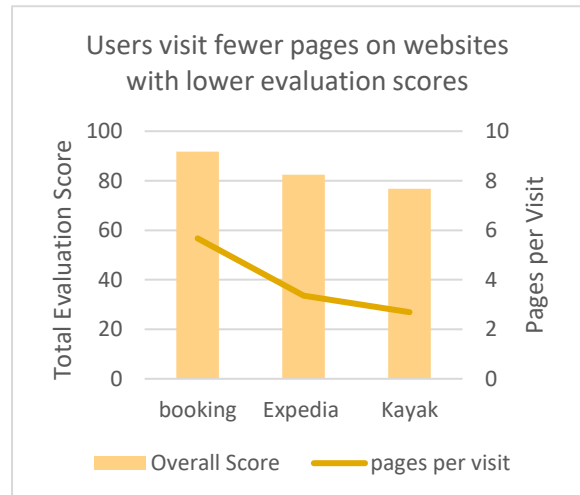


Figure 46

In regard to organic traffic, even though the majority of the websites show this metric to decrease with decreasing evaluation scores, the relationship does not seem to be strong as booking also exhibits a disproportionately high volume of organic traffic compared to its competitors (Figure 47).

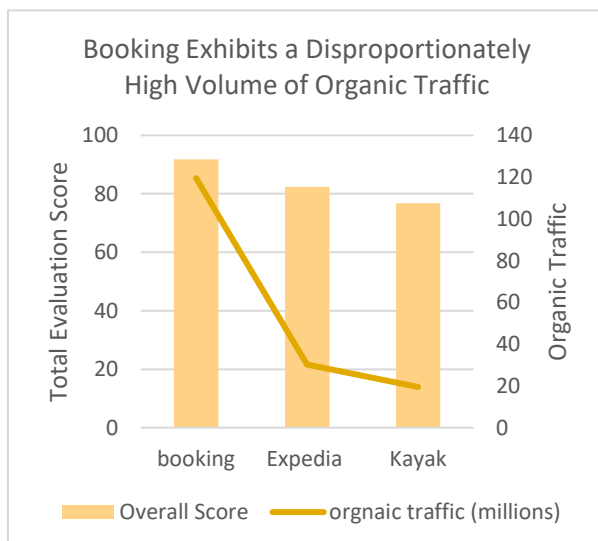


Figure 47

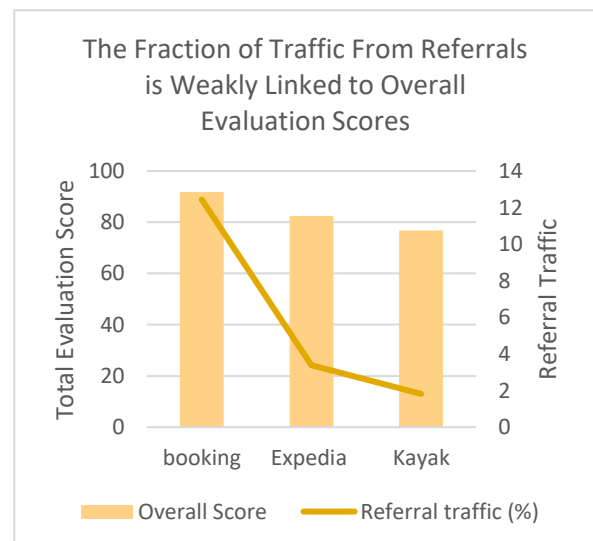


Figure 48

A similar observation can be made for the percentage of referral traffic (Figure 48), where booking records almost 4 times the percentage of the next highest website (Expedia) while only scoring 11% higher in the evaluation.

The strongest predictor of the overall score is perhaps the bounce rate which increases proportionally to decreases in evaluation scores (Figure 49). The overall percentage difference in evaluation scores (between booking and kayak) is 17.8% while the difference in bounce rate is 15% between the highest and the lowest scoring websites.

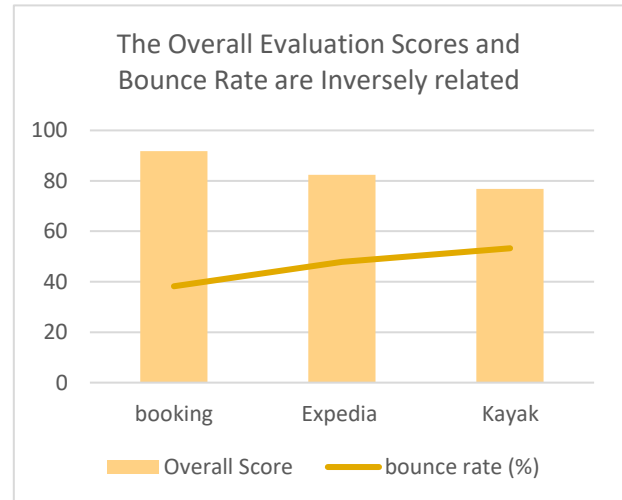


Figure 49

Unlike some of the attributes and sub-categories explored in the previous sections, the aggregated evaluation scores do not appear to be strongly related to the additional web analytics used in this research. The estimated monetary value of the websites and number of pageviews exhibit similar trends in regard to the overall evaluation scores (Figures 50 and 51). However, like the volume of visits, the performance metrics change more dramatically than the evaluation scores and begin to plateau in contrast to the linear decrease of the evaluation scores.

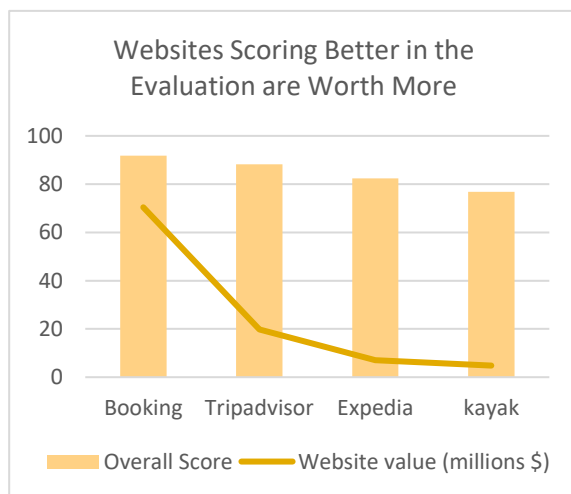


Figure 50

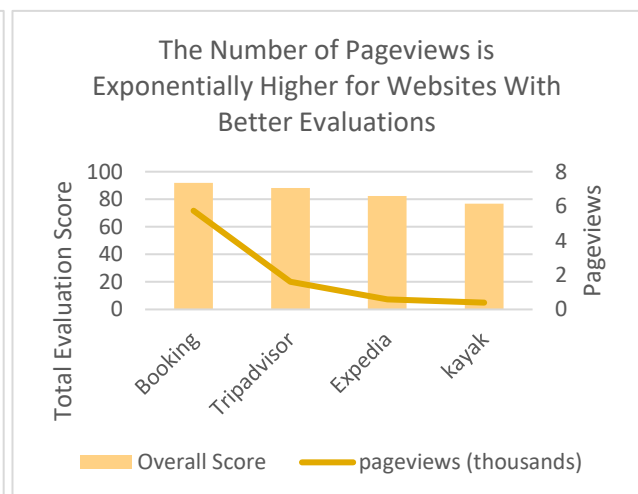


Figure 51

Looking at a summary of all the identified relationships (Appendix 4) several observations can be made regarding both the attributes and analytics. Firstly, all three attributes had at least one sub-category relating to the analytics suggested in the literature. However, the proposed analytics for Interface and Content were seemingly more supported, sometimes showing relationships with three sub-categories (Content with visit duration and referral traffic) or two sub-categories and the total attribute score (Interface with pages per visit and visit duration). While Navigation, on the other hand, only exhibited one related sub-category per analytic suggested by the literature. Additionally, while Navigation showed the most relationships to different analytics, all of these were linked to just one sub-category with a few also showing trends to the total Navigation scores. Instead, Interface and Content exhibit fewer relationships to different analytics but the analytics that do show patterns are related to more sub-categories.

In terms of Interface, the sub-category of graphic design principles and the total Interface scores showed the greatest number of relationships to different analytics, with each linking to four performance metrics. Logical Structure was the most salient sub-category in Navigation, making up six of the ten relationships found for the attribute. Finally, Interactivity was the most dominant category within Content exhibiting relationships to five different analytics.

Concerning the analytics, pages per visit, visit duration, and bounce rate appeared to be the most common predictors of evaluation scores with relationships to at least one sub-category of each attribute. This is followed by referral traffic, organic traffic, return visits and Load time, which are all related to at least one sub-category of two attributes. Finally, Navigation appears to be uniquely described by visits, accessibility, and Market share; average theme confidence and search ranking potential relate exclusively to aspects of interface; and page size is the only predictor of content that does not share relationships with other attributes.

5. Discussion

The primary objective of this research is to assess the nature and extent of the relationship between heuristic web evaluations and basic web analytics. While the results chapter describes the observed relationships, this chapter dives deeper into the meaning and implications of these results as well as the limitations of the study. To make the discussion as thorough as possible, this chapter can be organised in the same way as the results, interpreting the results for each web attribute (Interface, Navigation, and Content) as well as the overall results. The relationships suggested in the literature correspond to the general attributes tested in the evaluation, however, further insights can be drawn by interpreting relationships to the sub-categories of the different web attributes as well.

5.1 Interpretation of Results

Analytics as Predictors of Interface

The literature reviewed in chapter 2 suggests four accessible analytics as possible predictors of successful interface design, these being pages per visit, visit duration, bounce rate and load speed. The study found all of the suggested analytics to be somewhat related to the evaluation scores. In general, the evaluation of the use of graphics & multimedia could not easily be compared to any analytics due to the highly uniform results. The evaluation asked users whether the site is visually attractive, the media used contributes to the understanding of the site, the icons are easily understandable, not used excessively, and do not negatively impact loading times. The results indicate that standards for the use of media are consistent across important industry players. However, the category average of 6.25 out of 10 suggests there are still improvements to be made in the use of graphics and multimedia. In addition, the results showed graphics & multimedia to be the only sub-category with some relation to load speed. The websites that take longer to load generally scored slightly better in the evaluation. This may suggest that websites employing more media on the site (and thus taking longer to load) are perceived as more attractive in terms of visual design. However, this is unlikely to be the case for load times beyond a certain threshold where users will start to become impatient and frustrated, but this could not be seen within this sample of websites.

The evaluation of graphic design principles consists of more general use of design elements such as the effective and consistent use of white space, colour, and backgrounds, combinations of graphics, typeface and colour, and the quality of the homepage. This sub-category's relation to the pages per visit and visit duration on the site support the notion that

this aspect of Interface plays a role in the enjoyment a user experiences and encourages e-customers to explore and spend more time on the site. For both analytics related to graphic design principles, Tripadvisor contradicts this trend. Despite users assessing the use of design principles as successful, the analytics for the number of pages visited and time spent for this site did not reflect this high evaluation. When analysing why this might be, Tripadvisor stands out as a website that many users visit primarily to find information via the travel forum. Visitors search their questions via Google, land directly on the forum page with the answer and leave without exploring the rest of the site. The disproportionately high search traffic (80.4%) and bounce rate (66%) recorded for Tripadvisor supports this theory as well as analytics for top keyword searches such as “restaurants near me”, “things to do in [city]”, and “best restaurant in [city]” (Similarweb, n.d.). This outlier shows that while the pages per visit and visit duration may give clues about the perception of the site’s design choices, low numbers in these metrics are not a direct reflection of poor design. The analytics should thus be supplemented with heuristic data and additional analytics that may explain why the site appears to be underperforming and inform necessary actions to take to improve these analytics. In this way Tripadvisor, for example, is able to understand that although users find the site attractive, they need additional motivation to continue exploring the site once they have completed their initial objective of finding travel information.

The flexibility & compatibility of the websites was found to be related to the same analytics as graphic design principles but showed Kayak as the exception to the rule rather than Tripadvisor. Despite receiving high evaluation scores for this sub-category, users do not spend long on the site. Given that the analytics were found to relate to all web attributes in some way, it is difficult to pin-point which issue is most likely to affect the website’s performance. Kayak’s poor evaluation of graphic design principles and navigation are both likely explanations as users will not be stimulated to explore websites that are difficult to navigate or are unattractive. As with the discussion for graphic design principles, the results prove it is insufficient to rely solely on analytics to identify possible pain-points on the site and raises the additional issue that the analytics proposed in the literature appear to be related to more aspects than theorised. Managers relying on these theorised relationships may thus be adapting the website in ways that will not effectively improve performance on these analytics. As the visit duration and pages per visit relate to most of the sub-categories as well as the

overall evaluation results, these metrics may be used rather as indicators of the global success of the site rather than relying on them as measures of individual attribute success.

The last performance metric from the literature for this attribute is the bounce rate. Although the bounce rate also appears to predict sub-categories of navigation and content it only showed patterns with the total attribute score of Interface. This is in line with the existing theory that the initial impressions are vital to the intention to stay on the site.

Finally, it could be argued based on the results of the study that some of the analytics that have not yet been explored in the literature may be used as alternative predictors of interface success. In particular, the theme confidence and search ranking potential were found to be uniquely related to aspects of interface. Interestingly, websites with stronger themes were consistently evaluated more poorly in terms of graphic design principles but were evaluated better in terms of style & text, and flexibility & compatibility. Interpreting these results may suggest websites should tone down the strength of graphics, backgrounds, or colour combinations that may put users off, but maintain strong branding and consistency in the style of pages and text as well as the ability of the site to accommodate users' needs by creating accessible and flexible websites. In terms of search ranking potential, looking more closely at Google's current ranking factors explains why this relationship exists. Google's current ranking factors include aspects related to interface such as the mobile friendliness of the site as well as loading speed and attractive layout of content (Crowe, 2021). Developers should, however, keep in mind that ranking factors are constantly evolving, and that this relationship is thus likely to change overtime. Monitoring ranking factors is essential to keep the website at the top of the search results. Lastly, the amount of referral traffic was only found to be linked to the total score for interface despite its relationship to sub-categories of content. Although the website's interface cannot directly be seen in referrals, choices in colours, branding, text, and typeface are likely to be deciding factors in third parties' choice to send the user to another site. Booking, for example, has been able to create a strong brand with recognisable colours that makes it attractive for websites to link through their sites, this is reflected in its high fraction of referral traffic of 12.4%.

To answer the research question in regard to the ability of certain web analytics to measure users' perceptions of the site's interface, the bounce rate appears to be the most reliable

measure along with the new metrics discussed in this chapter (Referral traffic, theme confidence and search ranking). These metrics are most likely to give clues specifically about the overall successful design of the interface rather than describing the general quality of the website. Despite being suggested in the literature, interpreting the results for pages per visit and visit duration revealed that these may not be reliable measures for issues pertaining to interface as they appear to also be related to navigation and content and may thus lead to inaccurate conclusions about the role of interface in improving these performance metrics.

Performance Metrics Reflecting Perceptions of Navigation

As one of the most researched topics within web design, navigation recorded the greatest number of proposed relationships throughout the literature. In reality, the study revealed that while analytics vary significantly between websites, the perceived navigability (or usability) is very uniform making it challenging to identify significant patterns in the data. Researchers have suggested navigation to be related to visits, return visits, bounce rate, pages per visit, visit duration, and organic traffic. All of these proposed relationships were found to be true for just one sub-category of navigation.

The sub-category of logical structure was found to be related to four of the six proposed analytics. This category asks users to evaluate the organising scheme and structure of content, including the effective use of menus and sitemaps, as well as the consistency of navigation throughout the site. Elements of this category clearly overlap with certain aspects of interface such as the use of whitespace and other design elements that affect the way content is organised on the page, the use of multimedia as a contributor to navigability, and the consistency in the style of different pages. It is thus not surprising that the bounce rate is related to both the total interface scores and the logical structure as this is one of the elements that users are likely to perceive as soon as they land on the site and contribute to their intention to stay. The structure of the site was additionally related to visits and return visits. The results would suggest that websites that are well structured are more likely to attract new and return visitors. However, it is important to consider that (return) visits are a direct reflection of the site's size and popularity (Similarweb, n.d). While the logical structure of the site is likely to contribute to some extent to the amount of traffic received, it would be difficult to directly attribute the success of a site to this dimension of navigation as it is more likely a combination of many factors such as marketing efforts, website maturity, type of

products/services sold etc. While it may be useful to track the number of visits on a site overtime as an indication of growth, traffic as a competitive measure has limited validity; websites that are well designed may have fewer visits due to being newer or having less marketing budget, for example. This would explain why Booking contradicts the pattern found in the results of this study. The website records more than double the volume of traffic than the next highest (Tripadvisor), the company is owned by a conglomerate achieving revenues of \$6.8 billion (Booking Holdings Inc., 2020) and thus has more resources to develop and grow its platform. When comparing sites, it is therefore preferable to look at different traffic sources as fractions of total traffic so that more valid comparisons can be made between sites of different sizes. This was possible for certain types of traffic such as referral, direct, and search traffic, however, other types of traffic such as organic and paid could only be gathered as absolute numbers. Data for the organic traffic shows that websites scoring high in logical structure achieved greater volumes, with the exception of Expedia which underperforms compared to its evaluation score. Looking at the total traffic, Expedia has the second lowest visits, it is thus clear that the website appears to contradict the trend because it is not able to achieve visit volumes comparable to its competitors despite receiving high evaluation scores.

As mentioned in the discussion for interface, the pages per visit and visit duration were also found to be related to some extent to navigation. Ease of use encompasses the user-friendliness of the website, including how easy it is to reach and explore specific topics/information, accessing the product range, and returning to the homepage. It is the only sub-category of navigation that exhibited a significant relationship to these analytics. These results would suggest that, as theorised in the literature, users are more inclined to explore and spend time on websites that are easy to use. It could thus be argued that each attribute plays a distinct role in stimulating the user to stay on the site where ease of use enables visitors to find the website and navigate the content.

Again, the additional metrics explored in the study uniquely describe the users' perceptions of a website's navigation. Beyond the analytics from the literature, accessibility was also a predictor of evaluation scores for ease of use. As described in Chapter 2, improving the accessibility of the site can significantly benefit the user experience for all visitors by making it easier to navigate. In addition, the websites that take less time to load received better evaluations for total navigation suggesting that speed is an important component to a user's

experience while exploring a site. Minimising loading times helps create a smoother journey for customers, even if the navigation of a site is well-designed, ignoring IT aspects like speed will leave visitors frustrated. Finally, the logical structure and total navigation scores mirror analytics data for market share for majority of the websites studied. As the market share is measured as the percentage of total market traffic, this would support the conclusions drawn from analytics on visits in suggesting that navigation is vital for a website's success.

Overall, the analytics proposed in the literature as measures for the navigability of a site were to some extent supported by the study. The pages per visit and visit duration may give indications about the user friendliness of the site while high bounce rates are more likely to indicate structural issues that stop users from exploring the site. However, given that these analytics are related to various attributes of the site, these conclusions should not be used in isolation but rather compared to potential interface or content issues that may be responsible for short or single-page visits. The remaining analytics from the literature also appeared to be related, however, the use of absolute numbers revealed these patterns have limited validity, particularly when comparing sites of different sizes. Finally, accessibility was found to be an additional metric that could be used to evaluate how easy a site is to use. Increasing the accessibility of a website should thus not be an afterthought but rather a necessary component to improve the online experience for all users.

Analytics Describing Content Evaluation

The content of a website covers multiple aspects including information on products/services, the quality of information, company & contact details, and interactivity. Existing literature suggest visit duration, return visits, and referral traffic to be reliable predictors of content quality. All sub-categories except company & contact information showed that websites with better content evaluations recorded longer visits. In addition, despite not being suggested in previous literature, these categories also related to the average pages per visit. For all attributes the pages per visit and visit duration exhibit relationships to the same categories, as users who visit more pages naturally spend more time on the website. In the context of content, websites with quality information and a high degree of interactivity were able to retain users for longer periods of time. Thus, while the usability of the site likely helps users navigate the site's content, the quality of information and ability for users to interact with the website and other users is also important to keep visitors entertained and stimulated.

Company & contact information did not show significant trends for these analytics; users likely only search for this information in specific cases where they need direct contact with the company and thus do not play an important role in the overall experience for most visitors.

Product-related information, information quality, and interactivity were additionally related to referral traffic. This supports the existing research based on the theory that third parties are more likely to refer to websites with quality content. The user's experience on the referred website is to some extent an extension of their experience on the referring website as users trust the website to lead them to reliable information. In addition, one of the main strategies for increasing referral traffic is to be published on review websites (Kiel, 2019); websites that scored high in the evaluation for interactivity enable users to publish reviews on valuable and lively communities and are thus more likely to appear in these directories, explaining why they are able to achieve high volumes of referral traffic. Again, the referral traffic was not found to be related to scores for company & contact information. Users coming from other sites are unlikely to be looking for specific information on the company and thus shows no relation to the amount of referral traffic received.

Lastly, the study showed return visits to decrease with the interactivity of the site, supporting the notion that customers are likely to return to websites that are stimulating and enable quality interactions to take place. As discussed in the interpretation of results for navigation, however, the return visits are recorded in absolute numbers and may therefore show different results if tested on a different set of websites. The issues with the reliability of this metric are supported by the fact that the results for Booking again contradict the trend. As Booking receives a greater volume of total traffic, the number of return visits is also greater for this site, despite scoring lower on interactivity than Tripadvisor. The volume of organic traffic appears to be a better predictor of interactivity than return visits; despite also being measured in absolute numbers, the analytic appears to change more proportionately to changes in evaluation scores without any contradictions to the pattern. Nonetheless, using this metric to compare websites of significantly different sizes will likely be biased towards websites with greater market shares.

Beyond the analytics suggested by the literature, the bounce rate and page size also exhibited trends with aspects of content. Websites with better information quality exhibited lower

bounce rates. These results indicate that it is not sufficient to have an attractive interface that is well structured, if users perceive the information to be of poor quality they will not be encouraged to stay and purchase from the site. The results for the page size revealed users generally evaluated larger pages more positively implying visitors appreciate more content on the website. This was not found to be true for Expedia, however, which exhibited the lowest evaluation score and largest page size. This could be an indication that the trend is valid up to a certain threshold until the page size has significantly negative effects on the load times as would be supported by Expedia's long load time of almost 10 seconds. Beyond this threshold users may get impatient and struggle to efficiently navigate the content.

In general, the suggested analytics were able to reflect users' perceptions of different aspects of content. Like interface and navigation, the content of the site plays its role in convincing users to spend time on the site. In addition, the amount of referral traffic is able to give clues about the success of the websites content. Tracking return visits for a single site may be able to give indications about improvements in the interactivity of the site, however, it may not be the most practical measure for comparing websites, particularly when they vary greatly in traffic volumes. Additional analytics that may be useful measures for perceptions of content are bounce rate and page size. Websites need to combine attractive designs with sensible structures and high-quality information to encourage users to visit more pages, high bounce rates could indicate issues in any of these attributes. Furthermore, developers should take into consideration that users may like to consume more content but not at the expense of long loading times. Finally, the company & contact information did not exhibit relations to any analytics; particularly with large international companies such as those used in this study, the average user will likely not enter with the objective to directly contact the company. This attribute may be more important for smaller websites or start-ups that are easier to interact with.

Analytics' Reflection of Aggregated Evaluation Scores

Given that this study excludes aspects of IT and Reliability, the aggregated evaluation scores from this research cannot be used to draw conclusions about the overall success of the studied websites. However, they can still contribute to the understanding of the ability of different analytics to accurately describe attributes of the web environment. Almost all of the analytics suggested in the literature to describe individual attributes also mirrored the aggregated

evaluation scores, with the exception of speed. This indicates that these metrics may be influenced by a variety of web elements, particularly for measures such as pages per visit, visit duration, and bounce rate which show relations to sub-categories of all three tested attributes. Comparing these analytics to users' opinions of the site is essential to avoid wasting resources on changing elements of the website that are not the source of the issue.

The aggregated scores further reflect absolute measures of traffic (visits, return visits, and organic traffic), supporting the theory that websites that are well perceived along all three tested attributes are more successful. This is additionally supported by the overall scores' relation to website value and the number of pageviews. However, as mentioned throughout the discussion these absolute values are not able to consider relative performance and should thus not be taken at face value.

Beyond looking at the relations to general attributes, the nature of the chosen evaluation model also allowed for the exploration of sub-categories of web attributes and their relation to the proposed analytics on a more detailed level. This revealed that the graphic design principles applied by the site appear to be most important aspect of interface in terms of analytics performance, suggesting companies should invest resources in developing websites that effectively combine colours, graphics, and backgrounds that appeal to their audience and encourage exploratory behaviour. The logical structure of the site was the most conducive element of navigation, exhibiting majority of the relationships to analytics for this attribute. Thus, it may be implied that the effective organisation of graphic design elements and user-friendly menu systems are an important enabler of exploratory behaviour on the site that contribute to its overall success. Finally, the interactivity of the site appeared to be the most important aspect of content. Even websites that are visually appealing and easy to navigate may appear to underperform if they are not able to let the customer interact with the content in a way that keeps them entertained and stimulated throughout their online journey. The analytics' apparent ability to predict multiple aspects of different attributes highlights the interdependence between them. Trying to measure these attributes in isolation using only key web analytics such as those suggested in the current literature is challenging. Thus, some of the new analytics proposed in this study were seemingly able to predict perceptions of individual attributes, for example the theme confidence, accessibility, and page size.

Nonetheless, cross-referencing these analytics with user evaluations is critical to identify core problems on the site.

5.2 Managerial Implications

Interpreting the results of the study reveals that analytics can be useful measures of a website's success in different attributes. However, managers should not use analytics as a replacement for traditional evaluations by users. Relying on existing literature's suggestions for different metrics as measures for different aspects of the web environment may not guarantee increased analytic performance. Managers should understand how the attributes are interrelated to create a smooth and satisfactory customer journey encompassing the most important aspects of each attribute, these being graphic design principles, logical structure, and interactivity. In addition, the study and subsequent critical discussion propose a hybrid method for managers to identify key issues of the site more reliably, and effectively improve analytic performance. Based on the methodology employed by this study managers can assess and improve their website performance according to the process suggested in Figure 52.

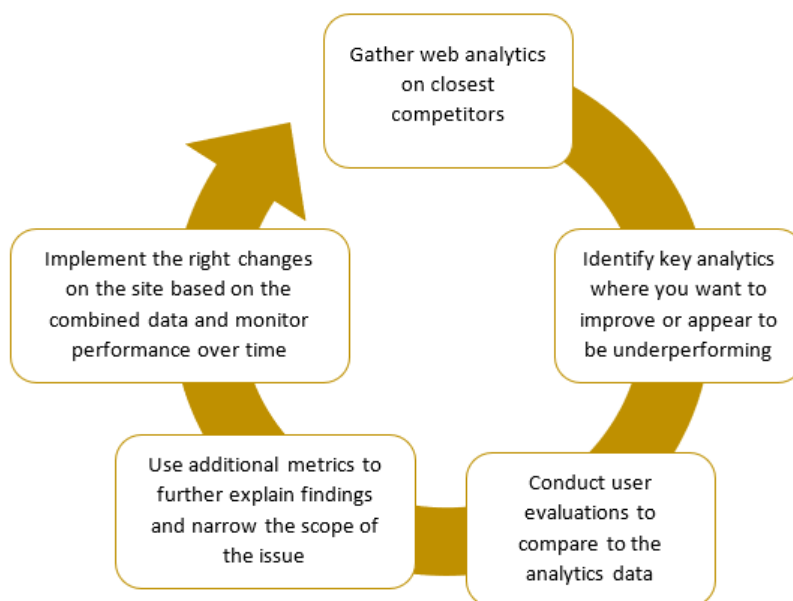


Figure 52: Suggested hybrid model for managerial use

Managers or developers in charge of a company's website can follow a similar process to the one used in this study as a hybrid model to pinpoint design issues and improve the online customer experience. Firstly, they can use the public analytics sites suggested in this research to gather key data on their own and close competitors'

websites. These can then be analysed to see which metrics the website is performing well in and which could be improved. The company can then conduct user evaluations on the site and its competitors using an existing model like van der Merwe & Bekker's (2003). Comparing the user evaluations to the analytics provides insights about how users perceive the websites versus the actual performance of the site. Gathering additional metrics (such as traffic sources

and top keywords) as we did for Tripadvisor in the discussion, enables managers to explain any incongruences between heuristic and analytics data and focus on the root issues rather than using trial and error to see which attributes need changing to improve analytic performance. Based on the combined findings of evaluations and analytics, managers can implement changes to the site more confidently, knowing which attributes users perceive less positively that may be leading to poorer analytic performance. After having implemented these changes, the analytics can be monitored over time to assess whether they have had the desired effect. The process can be repeated regularly as search engines change ranking criteria or when the website foresees dramatic changes due to new strategic objectives.

Managers should take into consideration that underperforming in certain key metrics does not necessarily mean the website needs to be improved. When evaluating the site's performance, the purpose should always be considered as this is likely to affect how users navigate the site and may lead to a significantly different strategy that impacts analytic performance. As with many business decisions, there are trade-offs in analytics that likely create a unique profile compared to competitors such as the distribution of traffic sources. Managers need to decide how resources can best be allocated to target their customer base and improve web performance on metrics that are in line with the company's objectives.

5.3 Limitations and Recommendations for Future Research

The chosen research design allows for the exploration of potential relationships between user evaluations and web analytics, however the qualitative nature of the study leaves room for further investigation. Conducting a similar design in a quantitative study would allow for statistical inferences to be made about the strength and causality of the tested relationships. For example, where analytics appear to relate to various web attributes, testing hypotheses about the direction and correlation of these links may give further insights into which attributes should be prioritised and whether changing an attribute has a ripple effect on perceptions of other attributes or performance in other analytic measures.

In addition, this research focussed on four industry leaders to ensure public analytics sites had enough data to accurately portray their performance. However, having access to different websites would provide richer data about other performance metrics such as order processes and shopping cart information that are vital to certain web attributes such as reliability which

could not be tested in this study. Having greater access would allow for the exploration of sites of different sizes and popularity that might result in less uniform results and thus easier identification of trends in the data. Due to the constraints of the evaluation model, this research was only able to focus on one industry (travel & accommodation); repeating this research for other types of websites from other industries would improve the generalisability of results. Alternatively, further research could also place greater focus on the potential thresholds found in the data about the suitable amount of content or information that users can handle before getting frustrated with load times.

Finally, an important limitation of this research is that the findings are not static. Even though e-commerce has matured, web analytics and users' perceptions and expectations of the online customer journey are still constantly evolving. This study serves primarily as a starting point for further exploration into the relationships between web evaluations and analytic performance and how to combine data from these distinct sources. Companies trying to replicate this research should repeat it regularly for the results to remain accurate and relevant and should edit the methodology based on their objectives and level of access to different analytics.

6. Conclusions

This research primarily aimed to explore how perceptions of a site are reflected in key analytic measures. Combining qualitative user evaluations and analytic data allowed for the identification of trends among analytics and attributes of the web based on relationships suggested in previous literature. The links previously suggested were tested and largely found to show patterns with web evaluations. In general, websites receiving higher evaluation scores also demonstrated better analytic performance. However, the key analytics from the literature were found to relate to multiple attributes and should thus not be used in isolation as a direct replacement for user evaluations of the site. Some of the additional analytics tested (such as theme confidence, accessibility, and page size) in the study appeared to be more unique descriptors of individual attributes and could thus be used to provide supplementary insights into the success of certain web dimensions like interface, ease of use, and content.

Additional objectives of the project were to provide recommendations on which parts of the website should be prioritised to improve performance and how to use evaluations and analytics as a holistic measure of website effectiveness. One sub-category of each attribute studied stood out as having the most relations to different analytics, thus suggesting that these elements may have the most impact on the analytic performance of a website. These were graphic design principles for interface, logical structure for navigation, and interactivity for content. In section 5.2 on managerial implications, the methodology of this study was summarised and simplified as a process that managers could use to assess and compare evaluation methods to inform practical and actionable improvements to the site's design.

To obtain these results, the existing literature in the field of both heuristic web evaluation and web analytics was thoroughly reviewed. After critically evaluating the most important models, van der Merwe & Bekker's (2003) model was selected and followed as closely as possible to maximise the validity of the study and several relationships to key web analytics were identified. When selecting the websites and participants for the evaluation, several considerations were made to improve the generalisability of the findings. Large international sites were used, and participants of different nationalities were chosen so that results are not constricted to a specific region. The analytic data for each website was gathered using a variety of public tools found online measuring the more common analytics mentioned in the literature as well as additional measures that contribute new theories on the link between

analytics and web attributes. The two sources of data were combined in graphs and analysed qualitatively to help identify significant trends, supplementary analytics were able to provide potential explanations for these patterns in a similar way to what managers could do to evaluate the performance of their own platforms.

There are still certain questions that the design of this research project was not able to answer, as highlighted in section 5.3. Researchers with greater resources and access to websites can add to the findings of this study by expanding on the attributes that could not be studied in this project (IT and Reliability). Future research could also focus on quantitatively testing hypotheses about the statistical significance and strength of the different relationships tested in this study to provide managers with more evidence about key web attributes to focus on and potential consequences on performance that any changes to web attributes may cause. Other related research may instead focus on different types of websites or industries to explore whether these results can be applied beyond leading websites in travel & accommodation. Finally, additional studies could be conducted to explore in what ways the web environment is still expected to change and how this may affect the relationships tested in this research.

Nonetheless, this research presents important contributions to the literature by building on the existing theories on the emerging hybrid evaluation models and relationships between analytics and attributes of the web. Even though majority of the theorised relationships were to some degree supported by the study, the interdependence of the different web attributes makes these suggested links unreliable as isolated measures. This study not only provides evidence for some of the theorised relationships in existing literature but also proposes new analytics and provides practical methods and recommendations on how to combine data for user evaluations and analytics.

Bibliography

- Albuquerque, A. B., & Belchior, A. D. (2002). E-commerce Website Quality Evaluation. *Proceedings of the 28 th Euromicro Conference*. Fortaleza: IEEE.
- Bakaev, M., Mamysheva, T., & Gaedke, M. (2016). Current trends in automating usability evaluation of websites: Can you manage what you can't measure? *2016 11th International Forum on Strategic Technology (IFOST)* (pp. 510-514). Novosibirsk: IEEE.
- Beri, B., & Singh, P. (2013). Web Analytics: Increasing Website's Usability and Conversion Rate. *International Journal of Computer Applications*, 72(6), 35-38.
- Cao, M., Zhang, Q., & Seydel, J. (2005). B2C e-commerce web site quality: an empirical evaluation. *Industrial Management & Data*, 105(5), 645-661.
- Chakraborty, G., Lala, V., & Warren, D. (2002). An empirical investigation of antecedents of B2B websites' effectiveness. *Journal of Interactive Marketing*, 16(4), 51-72.
- Chiou, W.-C., Lin, C. C., & Perng, C. (2010). A strategic framework for website evaluation based on a review of the literature from 1995-2006. *Information & Management*, 282-290.
- Crowe, A. (2021, September 15). *Top 8 Google Ranking Factors: What REALLY Matters for SEO*. Retrieved December 08, 2021, from Search Engine Journal: <https://www.searchenginejournal.com/ranking-factors/top-ranking-factors/#close>
- Cunliffe, D. (2000). Developing usable Web sites - a review and model. *Internet Research: Electronic Networking Application and Policy*, 10(4), 295-307.
- Davidavičienė, V., & Tolvaišas, J. (2011). MEASURING QUALITY OF E-COMMERCE WEB SITES: CASE OF. *Economics & Management*, 723-729.
- Gehrke, D., & Turban, E. (1999). Determinants of Successful Website Design: Relative Importance and Recommendations for Effectiveness. *Proceedings of the 32nd Hawaii International Conference on System Sciences*. IEEE.
- Hasan, L., Morris, A., & Probets, S. (2013). E-commerce websites for developing countries - a usability evaluation framework. *Online Information Review*, 37(2), 231-251.
- Hung, W.-H., & McQueen, R. J. (2004). Developing an Evaluation Instrument for e-Commerce Web Sites from the First-Tiem Buyer's Viewpoint. *Electronic Journal of Information Systems Evaluation*, 7(1), 31-42.
- ISO. (1998). ISO 9241-11, international standard first edition. Ergonomic requirements for office. Retrieved from www.it.uu.se/edu/course/homepage/acsd/vt09/ISO9241part11.pdf
- Joia, L. A., & Barbosa de Oliveira, L. C. (2008). Development and Testing of an E-Commerce Web Site Evaluation Model. *Journal of Electronic Commerce in Organizations*, 6(3), 37-53.
- Kim, S.-E., Shaw, T., & Schneider, H. (2003). Website design benchmarking within industry groups. *Internet Research: Electronic Networking Applications and Policy*, 13(1), 17-26.
- Kumar, R., & Hasteer, N. (2017). Evaluating Usability of a Web Application: A comparative analysis of open-source tools. *Proceedings of the 2nd International Conference on Communication and Electronics Systems* (pp. 350-354). Coimbatore: IEEE.

- Loiacano, E. T., Watson, R. T., & Goodhue, D. L. (2007). WebQual: An Instrument for Consumer Evaluation. *International Journal of Electronic Commerce*, 11(3), 51-87.
- Nielsen, J. (2012, January 3). *Usability 101: Introduction to Usability*. Retrieved from Nielsen Norman Group: <https://www.nngroup.com/articles/usability-101-introduction-to-usability/>
- Nielsen, J., & RL, M. (1994). *Usability Inspection Methods*. New York, NY: John Wiley & Sons.
- Olsina, L., Lafuente, G., & Rossi, G. (2001). Specifying Quality Characteristics and Attributes for Websites. In S. Murugesan, & Y. Deshpande, *Web Engineering. Lecture Notes in Computer Science*, vol 2016 (pp. 266-278). Berlin: Springer.
- Oppenheim, C., & Ward, L. (2006). Evaluation of web sites for B2C e-commerce. *Aslib Proceedings*, 58(3), 237-260.
- Pickard, A. a. (2007). *Research Methods in Information*. London: Facet.
- Rose, S., Hair, N., & Clark, M. (2011). Online Customer Experience: A Review of the Business-to-Consumer Online Purchase Context. *International Journal of Management Reviews*, 13, 24-39.
- Saunders, M., Lewis, P., & Thornhill, A. (2016). *Research Methods for Business Students* (Vol. Seventh Edition). Pearson.
- Schubert, P., & Selz, D. (1999). Web Assessment - Measuring the Effectiveness of Electronic Commerce Sites Going Beyond Traditional Marketing Paradigms. *Proceedings of the 32nd Hawaii International Conference on System Sciences* (pp. 1-10). Maui: IEEE.
- van der Merwe, R., & Bekker, J. (2003). A framework and methodology for evaluating e-commerce Web sites. *Internet Research*, 13(5), 330-341.
- Wiggins, A. (2008). Data-driven design: Using web analytics to validate heuristics system. *Bulletin of the American Society for Information Science and Technology*, 33(5), 20-24.

Appendix

Appendix 1: Website Evaluation Questionnaire

Thank you for participating in this research!

For my thesis, I am exploring how user-based evaluations relate to a website's analytic performance. I will thus be asking you to evaluate 4 websites along 3 dimensions. Your scores will then be compared to various web analytics metrics.

Before scoring each website, please read the instructions carefully. The first website you will be evaluating is Booking.com

Open the website by searching booking.com in your standard browser. Once the website is open, answer the survey questions while browsing.

You will be asked to evaluate the websites one by one. Once you are introduced to the next website you can close Booking.com.

Interface:

1. Graphic Design Principles

	strongly disagree	disagree	agree	Strongly agree	N/A
The homepage is concise and clear	☐	☐	☐	☐	☐
Whitespace is used effectively throughout the site	☐	☐	☐	☐	☐
Colours are used consistently and effectively throughout the site	☐	☐	☐	☐	☐
Backgrounds are used consistently and effectively throughout the site	☐	☐	☐	☐	☐
The combinations of graphics, typeface and colours used are effective	☐	☐	☐	☐	☐

2. Graphics & Multimedia

	strongly disagree	disagree	agree	Strongly agree	N/A
The site is visually attractive	☐	☐	☐	☐	☐
The graphics and multimedia contribute to the understanding and navigation of the site	☐	☐	☐	☐	☐
The icons used are easy to understand	☐	☐	☐	☐	☐
Graphics & multimedia re not excessively used	☐	☐	☐	☐	☐
The size of the media do not have a negative impact on loading times	☐	☐	☐	☐	☐

3. Style & Text

	strongly disagree	disagree	agree	Strongly agree	N/A
The style of pages is consistent throughout the site	☐	☐	☐	☐	☐
The typefaces are consistent and easy to read	☐	☐	☐	☐	☐
Spelling and grammar is correct throughout the site	☐	☐	☐	☐	☐
The text is concise is relevant	☐	☐	☐	☐	☐
The purpose of the site is made clear on the homepage	☐	☐	☐	☐	☐

4. Flexibility & Compatibility

	strongly disagree	disagree	agree	Strongly agree	N/A
Pages are sized to fit the browser window	☐	☐	☐	☐	☐
There are printable versions of different pages	☐	☐	☐	☐	☐
Pages are available in text-only versions	☐	☐	☐	☐	☐
The website is available in multiple languages	☐	☐	☐	☐	☐
There are accommodations for disabled users	☐	☐	☐	☐	☐

Navigation:

5. Logical Structure

	strongly disagree	disagree	agree	Strongly agree	N/A
The organising scheme is intelligible and straightforward	☐	☐	☐	☐	☐
The content is logically structured in different sections and levels	☐	☐	☐	☐	☐
The menus are understandable and straightforward	☐	☐	☐	☐	☐
There is a sitemap/table of contents available	☐	☐	☐	☐	☐
The navigation is consistent throughout the site	☐	☐	☐	☐	☐

6. Ease of Use

	strongly disagree	disagree	agree	Strongly agree	N/A
The website is easy to find	☐	☐	☐	☐	☐
It is easy to explore a specific idea/subject	☐	☐	☐	☐	☐
It is easy to return to the main page	☐	☐	☐	☐	☐
It is easy to find specific information	☐	☐	☐	☐	☐
It is easy to access the complete product/service range	☐	☐	☐	☐	☐

7. Search Engine & Help Function

	strongly disagree	disagree	agree	Strongly agree	N/A
The website search engine is easy to use	☐	☐	☐	☐	☐
The search engine is accurate	☐	☐	☐	☐	☐
The search engine findings are well described	☐	☐	☐	☐	☐
There are no search engine errors	☐	☐	☐	☐	☐
The help function is easy to use	☐	☐	☐	☐	☐

8. Navigational Necessities

	strongly disagree	disagree	agree	Strongly agree	N/A
There are no broken links	☐	☐	☐	☐	☐
There are no pages "under construction"	☐	☐	☐	☐	☐
Links are clearly identifiable, well labelled and defined	☐	☐	☐	☐	☐
Your current position on the	☐	☐	☐	☐	☐

site is clearly
labelled

Frames are used
effectively, and
non-frames
versions are
available

○

○

○

○

○

Content:

9. Product/Service -Related Content

	strongly disagree	disagree	agree	Strongly agree	N/A
There is extensive information available on products/services offered	○	○	○	○	○
Information on price and cost benefits are communicated	○	○	○	○	○
The breadth of product range is appropriate	○	○	○	○	○
The website adequately advertises its own products	○	○	○	○	○
The website adequately advertises other companies	○	○	○	○	○

10. Company & Contact Information

	strongly disagree	disagree	agree	Strongly agree	N/A
The full company information is available	○	○	○	○	○
Terms and conditions are easy to access	○	○	○	○	○
An e-mail address is provided on the site	○	○	○	○	○
A telephone number is provided on the site	○	○	○	○	○

The postal or physical address is provided on the site	☐	☐	☐	☐	☐
--	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------

11. Information Quality

	strongly disagree	disagree	agree	Strongly agree	N/A
The displayed content is current and updated	☐	☐	☐	☐	☐
The content is relevant to the purpose of the site	☐	☐	☐	☐	☐
The content is concise and non-repetitive	☐	☐	☐	☐	☐
I perceive the products/services to be of high quality	☐	☐	☐	☐	☐
I perceive the advertisements to be of high quality	☐	☐	☐	☐	☐

12. Interactivity

	strongly disagree	disagree	agree	Strongly agree	N/A
I am able to customise the content to fit my needs	☐	☐	☐	☐	☐
There is a large amount of personalisation possible	☐	☐	☐	☐	☐
It is easy to access the online community	☐	☐	☐	☐	☐
I perceive the value of the online community to be high	☐	☐	☐	☐	☐
There is a high degree of interactivity in the community	☐	☐	☐	☐	☐

*This set of questions (1 through 12) are repeated for Tripadvisor.com, Expedia.com, and Kayak.com.

Appendix 2: Glossary of Evaluation Terms

White Space	refers to the space that is left blank i.e. not taken up by any text, images, or design elements on the site
Graphics & Multimedia	Graphics refer to any visual design elements on the interface while multimedia is any combination of content such as audio, visual, animation, text, etc.
Typeface	refers to the style/design of the lettering or font used on the site
Navigation	refers to how the website helps or hinders your exploration of its content for example through the organisation scheme, use of menus, etc.
Printable version	Whether a website translates to a printable version can be seen by typing ctrl+P, websites that are still useful and legible in this format may be considered printable
Search engine	the search bar on a website where you are able to explore specific themes or topics
Help function	also referred to as the help center, is where users can go to find frequently asked questions or get in contact with customer service
Broken links	a broken link links to a page that is empty or does not exist, clicking a broken link will lead to an error message being displayed on the Screen
Cost benefits	assesses whether the website communicates the advantage of making a purchase at a certain price
Personalisation	the extent to which the website is able to adapt to your preferences and thus create a unique experience for you on the site
Online Community	any form of communication you are able to have with other users of the site for example through ratings, reviews or forums
Interactivity	any 2-way communication between you and the website where you are able to make choices about the content you are shown

Appendix 3: Website Evaluation Data

Phase 1 - Interface																									
Criteria		Booking.com					Tripadvisor					Expedia					Kayak								
	Resp1	Resp2	Resp3	Resp4	Resp5	AVG	Resp1	Resp2	Resp3	Resp4	Resp5	AVG	Resp1	Resp2	Resp3	Resp4	Resp5	AVG	Resp1	Resp2	Resp3	Resp4	Resp5	AVG	
Group 1: Graphic design principles	Home page concise & clear	1	1	1	2	1	1.2	2	2	2	2	2	2	2	2	2	1	1	1.6	1	1	1	-1	-1	0.2
	Effective use of white space	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2	1.6	1	2	1	1	2	1.4
	Effective & consistent use of colour	2	2	2	2	2	2	2	1	1	2	1	1.4	1	1	1	2	2	1.4	1	1	1	1	1	1
	Effective & consistent use of backgrounds	2	2	2	1	2	1.8	1	2	2	1	2	1.6	1	1	1	1	1	1	1	1	1	2	1	1.2
	Effective graphics/typography	1	2	1	1	2	1.4	2	2	1	2	1	1.6	1	1	1	1	1	1	1	1	1	1	1	1
Group 2: Graphics & Multimedia	total	8	9	8	8	9	8.4	9	9	8	9	8	8.6	6	7	6	7	7	6.6	5	6	5	4	4	4.8
	Site visually attractive	-1	1	1	1	-1	0.2	2	2	2	1	2	1.8	-1	1	1	1	1	0.6	-1	1	1	1	1	0.6
	graphics and multimedia contribution to the understanding and navigation	1	1	1	1	2	1.2	2	1	1	2	1	1.4	1	1	1	1	1	1	1	1	1	1	1	1
	icons easy to understand	2	1	1	2	2	1.6	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1	2	1	1.6
	not excessively used	2	1	2	1	2	1.6	1	1	1	1	1	1	1	2	2	2	2	2	2	2	1	2	2	1.8
Group 3: Style & text	size of media have no negative impact on loading times	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	2	1	1.6
	total	5	5	6	6	6	5.6	7	6	6	6	6	6.2	5	7	7	7	7	6.6	6	7	6	7	7	6.6
	Style of pages consistent	1	1	1	1	1	1	2	1	1	1	1	1.2	1	2	1	2	2	1.6	1	1	1	1	1	1
	typefaces consistent and easy to read	1	2	2	2	2	1.8	2	2	2	1	1	1.6	2	2	2	2	2	2	2	2	1	1	2	1.6
	correct spelling & grammar	1	2	2	2	2	1.8	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Group 4: Flexibility & compatibility	text concise & relevant	1	2	1	1	1	1.2	1	1	1	1	1	1	1	1	1	2	1	1.2	1	1	1	1	1	1
	purpose of site made clear on homepage	1	2	2	2	2	1.8	1	1	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2
	total	5	9	8	8	8	7.6	8	7	7	6	6	6.8	8	9	9	9	9	8.8	8	8	7	7	8	7.6
	pages sized to fit in browser window	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
	printable versions of certain pages available	2	2	2	2	2	2	-1	-1	-1	-1	-2	-1.2	1	1	1	1	2	1.4	1	2	2	2	1	1.6
Group 5: Accessibility	text-only version available	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	foreign language support	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	1	1	1.2	2	2	2	2	2	2
	accommodation for disabled	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	total	6	6	6	6	6	6	3	3	3	3	3	2.8	4	5	4	5	5	4.6	5	6	6	6	5	5.6
		24	29	28	28	29	27.6	27	25	24	24	22	24.4	23	28	26	28	28	26.6	24	27	24	24	24	24.6

Phase 2 - Navigation																					
Group 1: Logical structure	Resp1	Resp2	Resp3	Resp4	Resp5	Resp1	Resp2	Resp3	Resp4	Resp5	Resp1	Resp2	Resp3	Resp4	Resp5	Resp1	Resp2	Resp3	Resp4	Resp5	
	1	1	1	2	1	12	2	2	2	1	2	18	1	2	1	1	12	1	1	1	1
	1	1	1	1	1	1	1	1	1	2	12	2	2	2	2	2	2	1	1	1	1
	1	2	1	1	2	14	1	1	1	1	1	1	1	1	2	1	12	1	2	1	2
	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	1	1	-1	-1	-1	-1
1	1	2	2	1	14	2	2	2	2	2	2	2	2	2	2	2	2	1	1	1	12
6	7	7	7	7	7	8	8	8	7	9	8	7	8	7	8	7.4	4	3	3	4	36
Group 2: Ease of use																					
easy to find site	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
easy to explore specific ideas or subject	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
easy to return to main page	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
easy to find specific	2	2	2	1	1	16	1	1	1	2	14	2	1	2	2	18	1	2	1	1	12
easy to access complete product/service range	2	2	2	2	2	2	1	2	1	2	16	2	2	1	1	2	2	1	1	2	16
10	10	9	9	9	9.6	8	9	8	10	10	9	10	9	9	10	9.4	9	8	9	9	8.8
Group 3: search engine & help function																					
easy to use search engine	2	2	2	2	2	2	2	1	1	2	16	2	2	2	2	2	1	2	2	1	14
search engine is accurate	1	2	2	2	2	18	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
good description of search engine findings	1	2	1	1	1	12	2	2	1	2	1	16	1	1	2	1	14	2	2	2	2
no search engine errors	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
help function easy to use	1	1	1	1	1	1	1	1	1	1	1	1	2	2	2	2	2	1	1	2	2
7	9	8	8	8	8	8	9	8	7	9	8.2	9	9	10	9	9.4	9	9	9	9	9
Group 4: Navigational																					
No broken links	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
no "under construction" pages	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
links clearly discernible, well labelled & defined	2	2	2	2	2	2	1	1	1	2	2	14	2	1	2	1	16	2	2	2	2
clear label of current position on the site	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2	2	18	1	2	1	14
effective use of frames, non-frames version available	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	8	8	8	8	8	8	7	7	7	8	7.4	8	7	7	7	8	7.4	7	8	7	7.4
31	34	33	33	32	32.6	32	32	30	34	35	32.6	34	33	33	34	34	33.6	29	30	28	29

Phase 3 - Content																									
Group 1: Product/service related content																									
extensive product/service information available	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
price information and cost benefits communicated	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
adequate breadth of product	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
adequate amount of advertising own products	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
adequate amount of advertising by other companies	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
total	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10	10
Group 2: Company & contact information																									
full company info available	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
terms & conditions easily	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
e-mail addresses available	-1	-2	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
telephone numbers available	-1	-2	-2	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
postal & physical addresses available	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
total	4	2	2	2	4	4	3	2	8	9	9	8	8	8	8	8	8	8	8	8	8	8	8	8	8
Group 3: information																									
content is current & updated	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
content is relevant to purpose of site	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
content is concise and non-repetitive	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
high perceived quality of product/service	1	1	2	2	2	2	2	2	1	1	1	-1	1	0.6	1	2	2	2	2	2	2	2	2	2	2
quality of advertisements is	2	2	2	2	2	2	2	2	2	2	2	2	2	1	14	1	2	2	2	2	2	2	2	2	2
total	9	9	10	10	10	10	10	9	9	8	8	7	8	8	8	9	9	9	9	9	9	9	9	9	9
Group 4: Interactivity																									
user able to customise content to fit needs	2	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
large amount of personalisation possible	2	2	2	1	2	2	2	2	2	2	2	2	2	1	16	1	1	1	1	1	1	1	1	1	1
easy to access online	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
high perceived value of online community	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
high degree of interactivity in community	1	2	2	1	1	1	1	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
total	9	9	9	8	8	9	8	8	10	9	10	10	9	9.6	4	6	6	5	5	5	5	5	5	5	5
total	32	30	31	32	33	31	32	31	33	32	31	29	31	31.2	21	22	22	23	23	23	23	23	23	23	23
Group 5: Booking.com scores																									
Resp1	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp2	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp3	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp4	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp5	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Group 6: TripAdvisor scores																									
Resp1	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp2	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp3	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp4	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp5	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Group 7: Expedia scores																									
Resp1	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp2	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp3	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp4	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp5	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Group 8: Kayak scores																									
Resp1	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp2	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp3	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp4	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
Resp5	87	93	92	93	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94	94
total	32	30	31	32	33	31	32	31	33	32	31	29	31	31.2	21	22	22	23	23	23	23	23	23	23	23
Avg score	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4	82.4

Appendix 4: Summary of Identified Relationships

Analytics Metric	Identified Patterns per Attribute			
	Interface	Navigation	Content	Overall
Pages per Visit	GDP, F&C, Total	EoU	PRC, IQ, Int	X
Visit Duration	GDP, F&C, Total	EoU	PRC, IQ, Int	X
Bounce Rate	Total	LS	IQ	X
Visits	-	LS	-	X
Return Visits	-	LS	Int	X
Organic Traffic	-	LS	Int	X
Referral Traffic	S&T, Total	-	PRC, IQ, Int	X
Load Time (Speed)	G&M	Total	-	
Theme Confidence*	GDP, S&T, F&C	-	-	
Search Ranking Potential*	GDP	-	-	
Accessibility*	-	EoU	-	
Market Share*	-	LS, Total	-	
Page Size*	-	-	Total	
Website Value*				X
Pageviews*				X

*Have not been explored in previous literature

Interface: Graphic Design Principles (**GDP**), Flexibility & Compatibility (**F&C**), Style & Text (**S&T**), Graphics & Multimedia (**G&M**), Total Interface (**Total**)

Navigation: Logical Structure (**LS**), Ease of Use (**EoU**), Total Navigation (**Total**)

Content: Product-Related Content (**PRC**), Information Quality (**IQ**), Interactivity (**Int**), Total Content (**Total**)

Appendix 5: Results Data Tables

Interface:

	Evaluation Model Data: Interface					Literature Analytics				Additional Metrics		
	Graphic design principles (GDP)	Graphics & Multimedia (G&M)	Style & text (S&T)	Flexibility & Compatibility (F&C)	Total Interface score (Total)	Pages per visit	Visit duration (minutes)	Bounce rate (%)	Load speed (Seconds)	Theme Confidence (%)	Referral traffic (%)	Search (%)
Booking	8.4	5.6	7.6	6	27.6	5.67	11:32	38.21	5.7	82.6	12.44	85
Tripadvisor	8.6	6.2	6.8	2.8	24.4	2.26	08:04	66	8.6	79	1.63	85
Expedia	6.6	6.6	8.8	4.6	26.6	3.36	09:35	47.85	10.1	91.2	3.39	73
Kayak	4.8	6.6	7.6	5.6	24.6	2.69	06:28	53.27	2.7	97.8	1.82	64

Navigation:

	Evaluation Model Data: Navigation				Literature Analytics				Additional Metrics					
	Logical structure (LS)	Ease of use (EoU)	Search engine & help (SEH)	Navigational necessities (NN)	Total Navigation score (Total)	Visits (millions)	Return visits (millions)	Visit duration (minutes)	Bounce rate (%)	Organic Traffic	Pages per visit	Accessibility (%)	Market share (%)	Load time (seconds)
Booking	7	9.6	8	8	32.6	436.6	244.9	11:32	38.21	119.5	5.67	86	7.57	5.7
Tripadvisor	8	9	8.2	7.4	32.6	182.8	75.7	08:04	66	170.1	2.26	83	1.56	8.6
Expedia	7.4	9.4	9.4	7.4	33.6	81.5	34.7	09:35	47.85	30.2	3.36	98	7.82	10.1
Kayak	3.6	8.8	9	7.4	28.8	42.1	15.9	06:28	53.27	19.5	2.69	80	4.05	2.7

Content:

	Evaluation Model Data: Content					Literature Analytics			Additional Metrics			
	Product-related content (PRC)	Company & contact info (C&C)	Information quality (IQ)	Interactivity (Int)	Total Content score (Total)	Return visits (millions)	Visit duration	Referral traffic (%)	Pages per visit	Bounce rate (%)	Organic traffic (millions)	Page size (MB)
Booking	10	3.2	9.6	8.8	31.6	244.9	11:32	12.44	5.67	38.21	119.5	2.3
Tripadvisor	5.2	8.4	8	9.6	31.2	75.7	08:04	1.63	2.26	66	170.1	2.3
Expedia	9.4	-1.2	8.8	5.2	22.2	34.7	09:35	3.39	3.36	47.85	30.2	2.7
Kayak	9	2.8	8.6	83	23.4	15.9	06:28	1.82	2.69	53.27	19.5	1.4

Appendix 6: Contingency Matrix

Website	Interface	Navigation	Content	Total
Booking.com	27.6	32.6	31.6	91.8
Tripadvisor.com	24.4	32.6	31.2	88.2
Expedia.com	26.6	33.6	22.2	82.4
Kayak.com	24.6	28.8	23.4	76.8

Appendix 7: Overall Score Graphs Including Tripadvisor