WORD COUNT: 2977

DATA ANALYSIS PROJECT

BS7602: Analytical tools for digital data

MEGAN SMITH (M.SMITH1.16)
1605351
MSc Digital Marketing & Analytics

Table of Contents

Table of figures

Introduction

Through the development of the Internet, data collection in marketing is essential in order to enhance marketers' understanding of consumers' purchase patterns and predict future market outcomes to assist with business decision-making (Luo, 2009: 196). Furthermore, companies are able to use technology to collect large amounts of data about consumers' interests and attitudes. Moreover, data collection provides valuable insights into consumers' purchase behaviours which marketers can translate into their marketing campaigns (Erevelles *et al.*, 2016; Bumblauskas *et al.*, 2017). The aspect that makes data interpretation challenging for marketers today is the exceptional volume and variety of data collected from consumers due to the nature of modern technologies, which is referred to as Big Data (Erevelles *et al.*, 2016: 897) which can be either structured or unstructured. It is important for marketers to develop the necessary skills to extract meaningful and relevant data and translate this into information to support decision-making, which awards the company a competitive advantage (Bumblauskas *et al.*, 2017: 703). However, often companies struggle to cope with analysing large sets of data due to the characteristics of the data itself, process challenges regarding how to capture and transform data, and management challenges including privacy and ethical issues (Sivarajah *et al.*, 2016: 265). Moreover, companies often overlook the benefits to data analytics in terms of a strategic advantage and instead focus on instinct and experience (Järvinen and Karjaluoto, 2015:7).

Since the rapid advancement of network and mobile technologies, it has become progressively convenient for customers to submit reviews for companies on their online platforms, where this review data is big, unstructured and hold useful customer feedback information (Zhou *et al.*, 2018:511). Although online consumer reviews (OCRs) are helpful to consumers in discovering strengths and weaknesses of different products and in discovering the most appropriate ones for their needs, they present a challenge for companies to analyse due to their 'volume, variety, velocity and veracity' (Salehan and Kim, 2016:30). In fact, industry experts have found that more than 80% of companies often struggle with interpreting their data (Reynolds, 2017, [online]). Since the nature of OCRs often act as a form of word of mouth through influencing other customers' purchasing decisions (Zhou *et al.*, 2018:512), it is critical for companies to effectively identify and respond to customers' demands, a concept known as customer agility (Zhou *et al.*, 2015, Roberts and Grover, 2012). Arguably, customer agility can be viewed as a significant indicator of big data analytical capability, and thus are better able to detect market opportunities (Zhou *et al.*, 2015:512, Roberts and Grover, 2012:232).

Research Aim/Objectives

The research aim of this data analysis project is to identify the most effective way to analyse customer reviews from a retailer. Using the dataset, the main research aim is to understand consumer's behaviours, patterns and opinions in order to recognise areas of improvement to develop/maintain customer satisfaction. Furthermore, using relevant tools to analyse sales patterns and present them visually will help assess consumer behaviour and predict where the main business opportunities lie.

The objectives of this report are to:

- Apply and evaluate relevant literature of quantitative and qualitative data analysis within the marketing context
- Define the business problem from chosen dataset
- Identify the most effective quantitative data tools to implement the most suitable solution
- Analyse and visualise data using the most appropriate tools and theories, against business metrics and performance
- Evaluate and measure the results of the analysis to identify patterns
- Refine the business problem and generate recommendations to influence business decisions, based upon new knowledge from the research findings

Analytical Approach and Process

The analytical approach to data is categorised by the type of source, the characteristics and what is wanted to achieve or discover from the data. This involves identifying a problem or a solution, or providing recommendations for the future, which is particularly significant in a marketing context for businesses to learn from, for example, their sales or consumers' behaviour. Data can be classified as being primary or secondary, and can be internal or external to the business, and either structured, semi-structured or unstructured. Structured data is already arranged into rows and columns and fixed fields, making it the easiest to search and organise, such as relational database spreadsheets (Tondak, 2020, [online]; Marr, 2019, [online]). Semi-structured data is information that is not completely structured or rigid, but still has some elements of structure to it, such as XML or HTML tagged text. Unstructured data is information that 'either does not organise in a pre-defined manner or not have a pre-defined data model' (Tondak, 2020, [online]), such as open-ended survey responses and social media content, making it more difficult to manage and analyse.

It is important for companies to understand how to efficiently analyse Big Data before using it to make decisions and influence strategies, since companies often still use data without providing the proper context and goals (Reynolds, 2017, [online]). The main types of data analytics are descriptive, diagnostic, predictive, and prescriptive. Descriptive analysis is the 'foundation of all data insight' (Gibson, 2021, online]) and is used to provide descriptions of the data, displaying what has happened, such as Google Analytics. Diagnostic analysis takes the descriptive analysis insights to find out why it happened – the causes of the outcomes to create connections between data and classify patterns of behaviour (Gibson, 2021, [online]). Predictive analysis seeks to predict future outcomes by devising models to help businesses plan ahead (Stevens, 2020, [online]), from which prescriptive analytics examines elements of all four types in order to establish what should be done next and for businesses to take advantage of the future outcomes predicted by adjusting business strategies appropriately (Stevens, 2020, [online]; Sivarajah *et al.*, 2017:277).

Framework

Data is considered significant once it has been analysed in a way that positively influences changes in decision making in companies, such as improved customer service, personalised products and services, and the use of predictive analytics to drive action (Bumblauskas, 2017:7). Therefore, using an appropriate strategic model is critical in understanding and analysing data against the wider business problem.

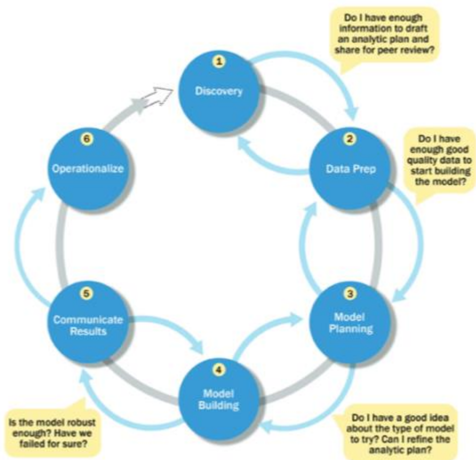

Figure 1 – Big Data Analytics Lifecycle Overview (EMC Education Services, 2015)

There are a number of frameworks used to uncover valuable knowledge from data to solve business problems, such as Cross Industry Standard Process for Data Mining (CRISP-DM), SMART, Big Data Analytics Lifecycle and the PPDAC model (Problem, Plan, Data, Analysis, Conclusions). Although these models offer differing characteristics in their process, they share similar systematic stages: determining the business problem, collecting and preparing data, data analysis, communicating results and finally interpretation and deployment. CRISP-DM, PPDAC and Data Analytics Lifecycle are similar in their structure of the process as a circle, whereby stages can be repeated if new circumstances surface during the data analysis phase which helps to ensure believable results. The Data Analytics Lifecycle was created specifically for problems with Big Data, with six phases where project work can move either forward or backward as new information is discovered and more is learnt about each stage, effectively representing a real project (EMC Education Services, 2015:26). On the other hand, the PPDAC model is concerned with ‘abstracting’ and ‘solving’ problems in data within a larger ‘real’ problem that needs to change (Wild and Pfannkuch, 1999:225).

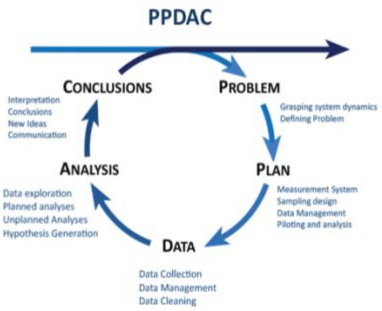

Figure 2 - PPDAC model (Wild and Pfannkuch, 1999)

The most appropriate frameworks for this project are the Data Analytics Lifecycle and PPDAC. Thus, a new framework has been proposed using elements of both previously mentioned models (shown in Figure 3), but with more emphasis on the business problem being integral to the whole process, and that every stage can be repeated if the results and evaluation stages suggest that further analysis is needed, or if the data needs to be prepped in a different way to answer a specific research question. Ultimately, if any element of the process generates new information, the process can start from the beginning. However, the inclusion of stage 3 aims to avoid this by taking the previous stage one step further by thoroughly developing the analytical techniques, tools and methods before deciding if they are most appropriate for the process.

Figure 3 - Data analysis model for project

Business problem	Define and investigate the business problem with acknowledgement of the wider marketing context, including the business domain.
Data preparation	Collection of relevant data helpful in examination to answer the business problem (structured or unstructured; internal or external; qualitative or quantitative), pre-processing, data cleaning, and selection and creation of attributes/variables.
Plan & build	Determine significant methods, techniques and tools intended for analysis. This is a detailed step in-between preparing the data and before it is equipped for analysis, where different methods may be needed for data analysis.
Analysis	Apply relevant analytics tools and software to analyse the data and to visualise the data in a creative way that the reader can easily understand.
Communications	Evaluation and assessment of the results of the analysis to be reported with potential solutions for the business problem.
Operationalise	On the basis of the evaluation being efficient, the results can be presented to the decision makers with the aim to be operationalised in the company. If not, then repeating previous stages or redefining the business problem may be necessary.

Figure 4 - Phases of data analysis model for project

Ethical consideration

There are certain ethical issues associated with secondary data analysis which should be considered before managing such data. Since secondary data works with large scale surveys or information collected through research, ethical issues can arise regarding the sharing of such results (Prasad, 2013:1478). These concerns about the secondary use of data include potential harm to individuals and issues of consent, but these tend to only arise through data collected as part of personal research when there is a risk of participants being identified within the data if it is not appropriately coded (Prasad, 2013:1478). Since the dataset used in this project was taken as part of a survey where individuals voluntarily submitted reviews and was available online where permission for further use was implied, such ethical issues should not be of concern.

Research Findings

This section will use the model above to analyse and visualise the data, present the findings and insights and critically review them.

Business problem

Since the model designed for this project is most concerned with the business problem, it is important for this to first be identified. For this project, the analysis of this dataset is aimed at identifying potential areas of improvement for product development. It also seeks to analyse customer satisfaction of varying clothing categories in order to identify the best performing departments and potential improvements in low performing departments based upon rating.

Data preparation

The dataset was captured from a data science website, from which the chosen dataset for this project is secondary, structured and internal (although the company the data is from is anonymous). See Appendix A for the full description of the dataset. The dataset represents the anatomy of Big Data, with it being large in volume and messy in its nature to hold veracity – the review texts consisting of textual errors and colloquial speech (Marr, 2015:80). Figure 5 shows the dataset in Excel before it was suitably prepped for analysis.

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Department Name	Class Name
0	767	33		Absolutely w	4	1	0	Initmates	Intimate	Intimates
1	1080	34		Love this dre	5	1	4	General	Dresses	Dresses
2	1077	60	Some major design flaws	I had such hi	3	0	0	General	Dresses	Dresses
3	1049	50	My favorite buy!	I love, love, l	5	1	0	General Petite	Bottoms	Pants
4	847	47	Flattering shirt	This shirt is v	5	1	6	General	Tops	Blouses
5	1080	49	Not for the very petite	I love tracy r	2	0	4	General	Dresses	Dresses
6	858	39	Cagrcoal shimmer fun	I aded this in	5	1	1	General Petite	Tops	Knits
7	858	39	Shimmer, surprisingly goes with lots	I ordered thi	4	1	4	General Petite	Tops	Knits
8	1077	24	Flattering	I love this dr	5	1	0	General	Dresses	Dresses
9	1077	34	Such a fun dress!	I'm 5"5' and	5	1	0	General	Dresses	Dresses
10	1077	53	Dress looks like it's made of cheap material	Dress runs sr	3	0	14	General	Dresses	Dresses
11	1095	39		This dress is	5	1	2	General Petite	Dresses	Dresses
12	1095	53	Perfect!!!	More and m	5	1	2	General Petite	Dresses	Dresses
13	767	44	Runs big	Bought the	5	1	0	Initmates	Intimate	Intimates
14	1077	50	Pretty party dress with some issues	This is a nice	3	1	1	General	Dresses	Dresses
15	1065	47	Nice, but not for my body	I took these	4	1	3	General	Bottoms	Pants
16	1065	34	You need to be at least average height, or taller	Material and	3	1	2	General	Bottoms	Pants
17	853	41	Looks great with white pants	Took a chanc	5	1	0	General	Tops	Blouses
18	1120	32	Super cute and cozy	A flattering,	5	1	0	General	Jackets	Outerwear
19	1077	47	Stylish and comfortable	I love the loo	5	1	0	General	Dresses	Dresses
20	847	33	Cute, crisp shirt	If this	4	1	2	General	Tops	Blouses
21	1080	55	I'm torn!	I'm upset be	4	1	14	General	Dresses	Dresses
22	1077	31	Not what it looks like	First of all,	2	0	7	General	Dresses	Dresses
23	1077	34	Like it, but don't love it.	Cute little dr	3	1	0	General	Dresses	Dresses
24	847	55	Versatile	I love this sh	5	1	0	General	Tops	Blouses
25	697	31	Falls flat	Loved the ma	3	0	0	Initmates	Intimate	Lounge
26	949	33	Huge disappointment	I have been v	2	0	0	General	Tops	Sweaters
27	1003	31	Loved, but returned	The colors w	4	1	0	General	Bottoms	Skirts
28	684	53	Great shirt!!!	I have severa	5	1	2	Initmates	Intimate	Lounge
29	4	28	Great layering piece	This sweater	5	1	0	General	Tops	Sweaters
30	1060	33		Beautifully m	5	1	0	General Petite	Bottoms	Pants
31	1060	46	Cuter in oerson!	I never woul	5	1	7	General Petite	Bottoms	Pants

Figure 5 - Dataset in Excel

Preparing the data is indeed a time-consuming process, but it is critical in ensuring that the data is suitably formatted and ready for analysis, including data cleaning, creating new variables and formatting variables (Grace-Martin, 2015, [online]). To prepare and check the data, Excel was used to add appropriate columns for the string variables that SPSS would not be able to analyse into numeric data to make it suitable for input into the software later (see Figure 6 for the prepped version). Moreover, it is important to select the most relevant metrics in order to be able to execute the model and perform the next steps effectively (EMC Education Services, 2015:36).

	Clothing ID	Age	Title	Review Text	Rating	Recommended IND	Positive Feedback Count	Division Name	Division name SPSS	Department Name	Department name SPSS (	Class Name	Class name SPSS (
0	767	33		Absolutely wonderful - silky and	4	1	0	Initmates	3	Intimate	3	Intimates	6
1	1080	34		Love this dress! it's sooo pretty.	5	1	4	General	1	Dresses	2	Dresses	4
2	1077	60	Some major design fla	I had such high hopes for this dr	3	0	0	General	1	Dresses	2	Dresses	4
3	1049	50	My favorite buy!	I love, love, love this jumpsuit. it	5	1	0	General Petite	2	Bottoms	1	Pants	14
4	847	47	Flattering shirt	This shirt is very flattering to all	5	1	6	General	1	Tops	5	Blouses	1
5	1080	49	Not for the very petite	I love tracy reese dresses, but th	2	0	4	General	1	Dresses	2	Dresses	4
6	858	39	Cagrcoal shimmer fun	I aded this in my basket at hte la	5	1	1	General Petite	2	Tops	5	Knits	9
7	858	39	Shimmer, surprisingly	I ordered this in carbon for store	4	1	4	General Petite	2	Tops	5	Knits	9
8	1077	24	Flattering	I love this dress. i usually get an	5	1	0	General	1	Dresses	2	Dresses	4
9	1077	34	Such a fun dress!	I'm 5"5' and 125 lbs. i ordered tl	5	1	0	General	1	Dresses	2	Dresses	4
10	1077	53	Dress looks like it's ma	Dress runs small esp where the	3	0	14	General	1	Dresses	2	Dresses	4
11	1095	39		This dress is perfection! so prett	5	1	2	General Petite	2	Dresses	2	Dresses	4
12	1095	53	Perfect!!!	More and more i find myself reli	5	1	2	General Petite	2	Dresses	2	Dresses	4
13	767	44	Runs big	Bought the black xs to go under the larkspur midi dress because they didn't bother lining the skirt portion (grrrrrrrrrrr). my stats are 34a-28/29-36 and the xs fit very smoothly around the chest and was flowy around my lower half, so i would say it's running big. the straps are very pretty and it could easily be nightwear too. i'm 5'6" and it came to just	5	1	0	Initmates	3	Intimate	3	Intimates	6
14	1077	50	Pretty party dress with	This is a nice choice for holiday	3	1	1	General	1	Dresses	2	Dresses	4
15	1065	47	Nice, but not for my bc	I took these out of the package a	4	1	3	General	1	Bottoms	1	Pants	14
16	1065	34	You need to be at least	Material and color is nice. the le	3	1	2	General	1	Bottoms	1	Pants	14
17	853	41	Looks great with white	Took a chance on this blouse and	5	1	0	General	1	Tops	5	Blouses	1
18	1120	32	Super cute and cozy	A flattering, super cozy coat. wi	5	1	0	General	1	Jackets	4	Outerwear	13
19	1077	47	Stylish and comfortabl	I love the look and feel of this tu	5	1	0	General	1	Dresses	2	Dresses	4
				if this product was in petite, i would get the petite. the regular is a little long on me									

Figure 6 - Dataset prepped in Excel ready for SPSS

Plan and build

Before analysis, this phase is important in understanding the possible relationships between variables and to fully understand the business sphere in order to solve the problem (EMC Education Services, 2015:44). This stage aims to determine the analytical methods, techniques and tools and if the current tools are efficient enough for analysis (EMC Education Services, 2015:30), and also if the analysis will suitably inform the research objectives. It is also important to identify the variables of particular significance for data analysis (see Appendix B).

Analysis

The analysis of data needs to be conducted with the business problem and the research objectives in mind to allow the identification of possible relationships between variables and therefore isolate relevant and suitable analysis to be performed. The more appropriate type of analysis to be performed for this dataset is descriptive. This is because the dataset is based on reviews and ratings, which is influential for understanding the current existing sales patterns and customer behaviour (David, 2019 [online]). The reviews also reveal levels of customer satisfaction, through the text review itself and the rating given by the customer, so descriptive analysis will enable the identification of such variable and sample characteristics that influence insights (Thompson, 2009:57).

The first part of the analysis of the dataset is a simple bar graph (see Figure 7) showing the percentage ratio of each review rating from 1 (worst) to 5 (best) awarded by customers reviewers. Over half the reviews were rated the highest score, with the lowest category being a review rating of 1. This gives an overall sense of positive customer feedback and a high level of customer satisfaction.

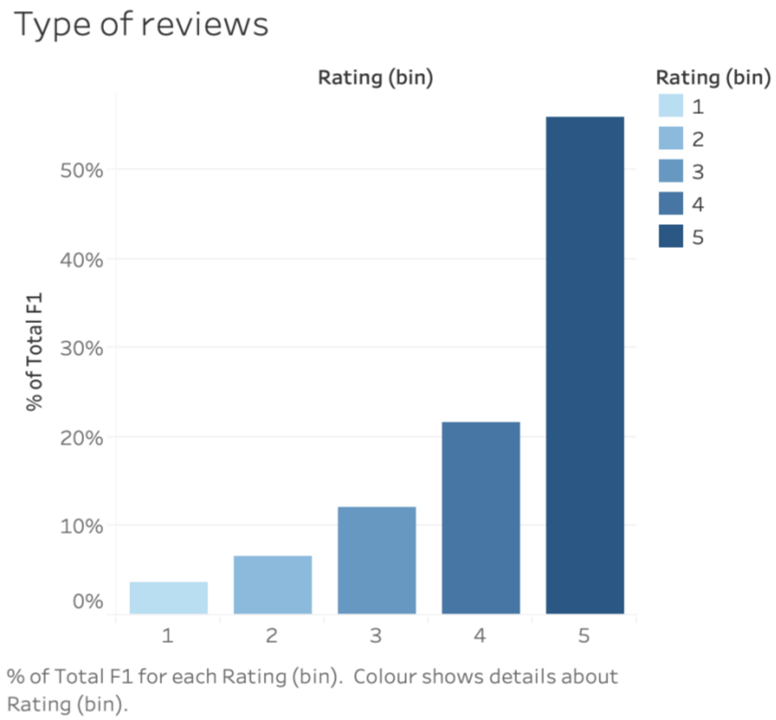

Figure 7 - Type of reviews (rating)

Furthermore, it would be useful to investigate the effect of age (by decade) on rating given (from 1 to 5). Figure 8 shows a graph breakdown of each rating category and within each one how many customers rated it that score from each age decade, and this is measured against percentage of total count of reviews. The graph clearly shows that the 30-year-old category provided more reviews than any other age group across all rating categories.

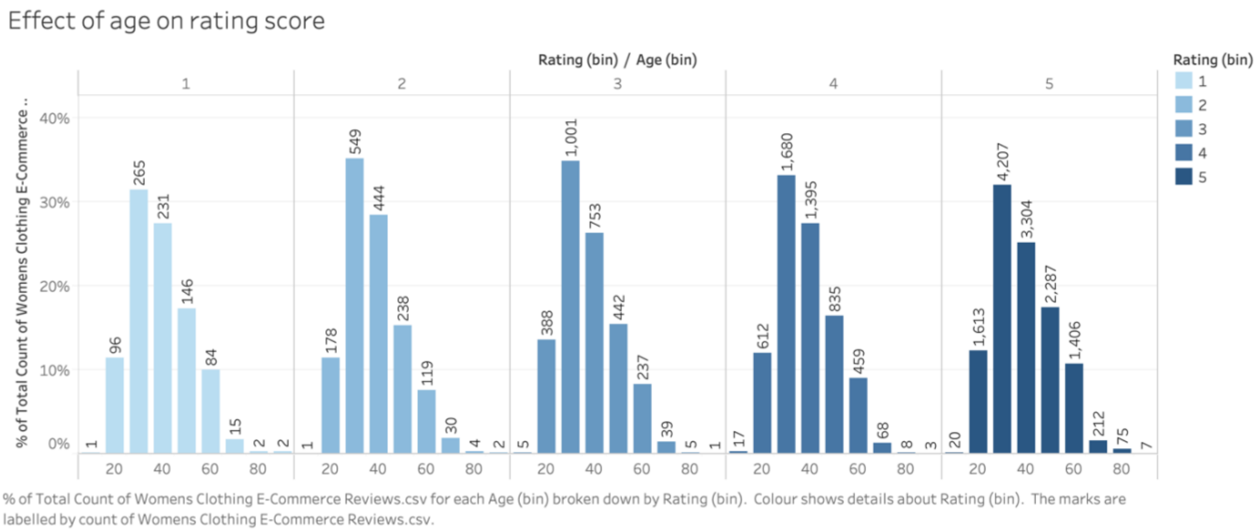

Figure 8 - Age of customer and review rating given

The review rating is an important variable to measure relationships since it is similar to the value of brands' star ratings, which are the most important review factors to customers (Murphy, 2020, [online]).

To help the company understand their audience better, it would be useful to identify the average age of reviewers. Figure 9 shows the number of reviews by age group, visualised in a pie chart. Since there was such a diverse range of reviewer ages which made an unappealing bar chart (see Figure 10), the ages were coded into groups by decade. Through a univariate analysis, Figure 11 shows a histogram of number of reviews given by each age decade. The mean age is 38.51 and knowing this allows the company to develop their target marketing strategy accordingly, through their product range and prioritisation of resources (*Chron,* 2021, [online]).

Age (decade)

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	10	44	.2	.2	.2
	20	2887	12.3	12.3	12.5
	30	7702	32.8	32.8	45.3
	40	6127	26.1	26.1	71.4
	50	3948	16.8	16.8	88.2
	60	2305	9.8	9.8	98.0
	70	364	1.5	1.5	99.5
	80	94	.4	.4	99.9
	90	15	.1	.1	100.0
	Total	23486	100.0	100.0	

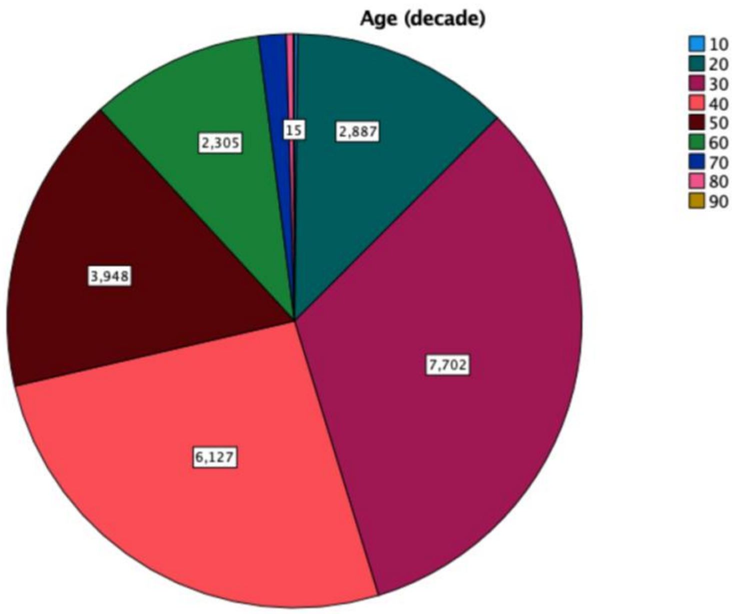

Figure 9 - Most reviews by age group

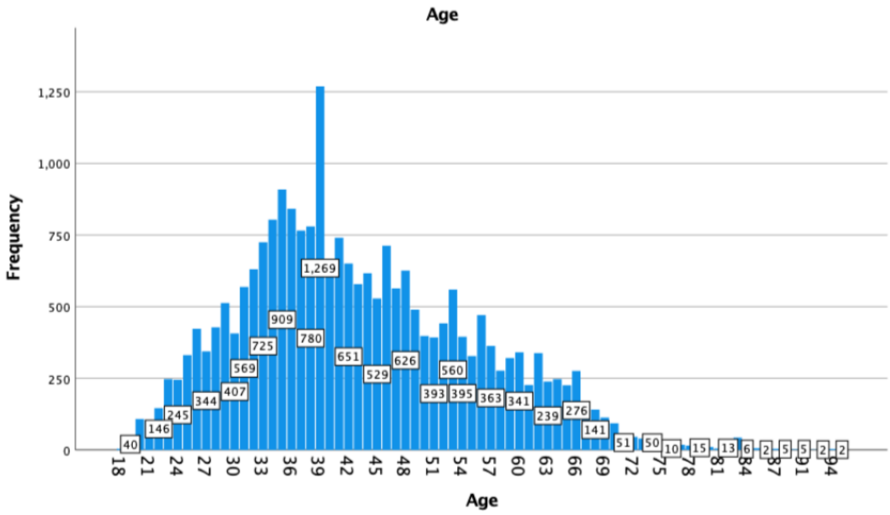

Figure 10 - Age of reviewers

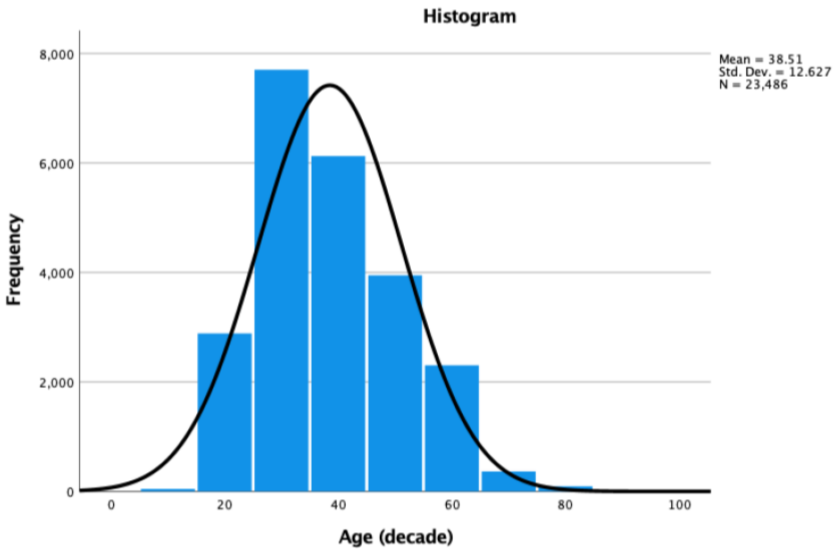

Figure 11 - Histogram of age of reviewers

Since a study found that reviews are important factors in customer's decision to purchase, with less reviews delaying customers in buying (*Fan and Fuel,* 2016, [online]), the company's review feature must remain an element in the customer journey.

To understand clothing categories that are well-received by customers, a univariate analysis investigates the number of reviews of each class name to discover the top 3; first is (4) Dresses = 6319; second is (9) Knits = 4843; third is (1) Blouses = 3097. The data is visually represented in the pie chart in Figure 12. Since over half of the reviews were rated positive, it is feasible to state that a high number of reviews suggest that these three clothing categories were purchased more. These findings can be used to establish best-selling products within the company, which can help identify products for development from the lower rated departments.

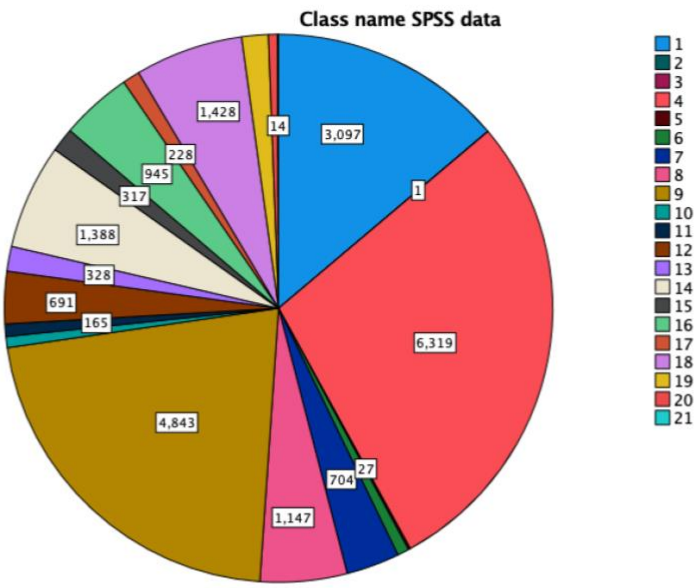

Figure 12 - Number of reviews against class name

Communications

The communications phase aims to understand if the business problem has been *fully* tackled and unravelled If not, the process needs to re-start from the analysis stage. The communications stage will also see the results concisely and clearly outlined in a report to present to the managing team, highlighting problems, findings and possible solutions. Presenting and interpreting the results must be done in a meaningful way to ensure they 'inform decision-making and improve performance' (Marr, 2015:155). If this is not demonstrated, even having a technically accurate data analysis will not be enough, since the value is in the manner in which the results and communications are demonstrated (EMC Education Services, 2015:30; Marr, 2015).

To help evidence decision-making, data should be presented clearly and simply so that the reader can easily understand. This can be accomplished through data visualisation, including graphs and charts, which make the data more 'accessible and meaningful' and also more effectively illuminates relationships between the data (Marr, 2015:157). A more creative and efficient way to communicate results is through infographics, using an amalgamation of text and graphics to report data (Marr, 2015:178).

Operationalise

In this final phase of data analysis, once results have been interpreted, they can be converted into steps to be operationalised by the business. This is an important stage as it is based upon the crucial actionable knowledge from the analysis to form final recommendations that evidence decision-making (Bumblauskas *et al.*, 2017:19). Rather than deploying the new recommendations immediately, this stage suggests implementing them on a pilot study before wide-scale rollout to minimise risk of error (EMC Education Services, 2015:50-51).

Conclusions and Recommendations

To conclude, the analysis of consumer review data is critical to digital marketers to gain valuable insights and facilitate customer agility to make better business decisions (Zhou *et al.*, 2018; Roberts and Grover, 2014).

Using different analytical tools and software allows more meanings and insights to be extracted from the data. Although simple descriptive text and figures show results, it is not able to measure relationships between variables and present data in a creative and well-defined way like tools of data visualisation presented in this report can.

As presented in this project, the findings from customer reviews are important in understanding consumer purchase patterns, behaviours, and levels of customer satisfaction. The company should focus on product development, and since the data analysis revealed insights into the ages of customers, the company should use this information to effectively target this demographic by offering products that reflect their interests and needs.

With the importance of reviews influencing customers to purchase, the company would benefit from using the positive reviews as customer testimonials during lead acquisition to share with potential customers (Cox, 2021, [online]). With regards to the small number of negative reviews, the company should follow up with these customers through personalised responses to make the show that the company is interested in their customers' fulfilment, with the aim to lower numbers of negative reviews even more and increase customer retention. A follow-up analysis after this recommendation has been deployed within the company would be valuable to measure extent of improvements and to see if this goal has been achieved.

Even though over half of the reviews in the dataset reflect positive opinions, a more thorough analysis of the lower rated reviews would be valuable to identify precisely what products are not as highly rated in the eyes of consumers. This could be investigated through further tests.

The analysis of the dataset for the company evidenced consumer purchase patterns in regard to age group and popular clothing categories. However, further analysis could be carried out to consider other elements, such as conducting an in-depth sentiment analysis of the review texts to understand more about consumer attitudes (both positive and negative) towards products, and to understand emotions and opinions in a more qualitative way.

Appendices

Appendix A – Description of Dataset

	Dataset
Name of the Dataset	Women's E-Commerce Clothing Reviews
Describe Type of Data	The dataset represents reviews of women's e-commerce clothing, the department of the item reviewed, the review rating, the age of the customer, and recommended IND
Location and/or Ownership (*Internal*/*External*)	The dataset was retrieved from www.kaggle.com and was published by Github of an anonymised retailer as it is real commercial data External from the market, i.e., customers
Data Format (*Structured*/*Unstructured*)	Structured
Data Collection Method	Secondary data – downloaded from www.kaggle.com
Data Volume/Scale	Rows: 23485 of clothing reviews Columns: 11 variables referring to customer review categories (10 feature variables)
Data Quality	Complete to allow for multivariate analysis

Appendix B – Description of attributes/variables

The variables in grey reflect the ones that were not used for analysis.

Attribute/Variable	Description	Variable Type
ID	ID number	Numerical
Clothing ID	Integer referring to the specific clothing piece being reviewed	Categorical
Age	Positive integer of the age of reviewer	Numerical
Title	Title of the review	Nominal
Review Text	Review body	Nominal
Rating	Positive ordinal integer for the product score given by the customer (from 1 Worst, to 5 Best)	Numerical (ordinal)
Recommended IND	Binary variable of whether the customer recommends the product (1 is recommended, 0 is not recommended)	Numerical
Positive Feedback Count	Positive integer showing the number of other customers who found the review positive	Numerical
Division Name	Name of product division	Categorical
Department Name	Name of the product department name	Categorical
Class Name	Name of the product class name	Categorical

References

Bumblauskas, D., Nold, H., Bumblauskas, P. and Igou, A. (2017) Big data analytics: transforming data to action. *Business Process Management Journal*, 23, (3) 703-720. Available at: https://www.emerald.com/insight/content/doi/10.1108/BPMJ-03-2016-0056/full/pdf?title=big-data-analytics-transforming-data-to-action. [Accessed 2 April 2021].

Chron. (2021) The Importance of a Target Audience of Consumers. Available at: https://smallbusiness.chron.com/significance-market-research-small-business-owners-41749.html. [Accessed 22 May 2021].

Cox, L.K. (2021) 31 Customer Review Sites for Collecting Business & Product Reviews. *HubSpot.* Available at: https://blog.hubspot.com/service/customer-review-sites. [Accessed 22 May 2021].

David, D. (2019) Understand Data Analytics Framework with a case study in the business world. *Becoming Human: Artificial Intelligence Magazine.* Available at: https://becominghuman.ai/understand-data-analytics-framework-with-a-case-study-in-the-business-world-15bfb421028d. [Accessed 2 April 2021].

EMC Education Services (2015) *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data.* Indianapolis: Wiley.

Erevellles, S., Fukawa, N. and Swayne, L. (2016) Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69, (2) 897-904. Available at: https://doi.org/10.1016/j.jbusres.2015.07.001. [Accessed 2 April 2021].

Fan and Fuel. (2016) No online customer reviews means BIG problems in 2017. Available at: https://fanandfuel.com/no-online-customer-reviews-means-big-problems-2017/. [Accessed 22 May 2021].

Grace-Martin, K. (2015) Preparing Data for Analysis is (more than) Half the Battle. *The Analysis Factor*. Available at: https://www.theanalysisfactor.com/preparing-data-analysis/. [Accessed 21 May 2021].

Järvinen, J. and Karjaluoto, H. (2015) The use of Web analytics for digital marketing performance measurement. *Industrial Marketing Management*, 50, 117-127. Available at: https://jyx.jyu.fi/bitstream/handle/123456789/47504/imm13334rrrevisedmanuscript.pdf?sequence=1. [Accessed 21 May 2021].

Luo, Y. (2009) Using Internet Data Collection in Marketing Research. *International Business Research*, 2, (1) 196-202. Available at: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.668.4640&rep=rep1&type=pdf. [Accessed 2 April 2021].

Marr, B. (2015) *Big Data: Using SMART Big Data, Analytics and Metrics to Make Better Decisions and Improve Performance.* Chichester: Wiley & Sons Ltd.

Marr, B. (2019) What's The Difference Between Structured, Semi-Structured And Unstructured Data? *Forbes.* Available at: https://www.forbes.com/sites/bernardmarr/2019/10/18/whats-the-difference-between-structured-semi-structured-and-unstructured-data/?sh=1ed33af92b4d/. [Accessed 21 May 2021].

Murphy, R. (2020) Local Consumer Review Survey 2020. *Bright Local.* Available at: https://www.brightlocal.com/research/local-consumer-review-survey/. [Accessed 21 May 2021].

Prasad, J. (2013) Secondary Data Analysis: Ethical Issues and Challenges. *Iranian J Publ Health,* 42, (12) 1478-1479. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4441947/pdf/IJPH-42-1478.pdf. [Accessed 21 May 2021].

Reynolds, C. (2017) Using Online Reviews and Big Data for Positive Impact. *Datafloq.* Available at: https://datafloq.com/read/using-online-reviews-big-data-for-positive-impact/3125. [Accessed 21 May 2021].

Roberts, N. and Grover, V. (2012) Leveraging Information Technology Infrastructure to Facilitate a Firm's Customer Agility and Competitive Activity: An Empirical Investigation. *Journal of Management Information Systems*, 28, (4) 231-270. Available at: https://doi.org/10.2753/MIS0742-1222280409. [Accessed 21 May 2021].

Salehan, M. and Kim, D.J. (2016) Predicting the performance of online customer reviews: A sentiment mining approach to big data analytics. *Decision Support Systems*, 81, 30-40. Available at: https://doi.org/10.1016/j.dss.2015.10.006. [Accessed 21 May 2021].

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017) Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286. Available at: https://doi.org/10.1016/j.jbusres.2016.08.001. [Accessed 2 April 2021].

Sivarajah, U., Kamal, M.M., Irani, Z. and Weerakkody, V. (2017) Critical analysis of Big Data challenges and analytics methods. *Journal of Business Research*, 70, 263-286. Available at: https://doi.org/10.1016/j.jbusres.2016.08.001. [Accessed 21 May 2021].

Stevens, E. (2020) What Are the Different Types of Data Analysis? *Career Foundry.* Available at: https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/. [Accessed 21 May 2021].

Thompson, C.B. (2009) Descriptive Data Analysis. *Air Medical Journal*, 28, (2) 56-59. Available at: https://doi.org/10.1016/j.amj.2008.12.001. [Accessed 22 May 2021].

Tondak, A. (2020) Data Types: Structured Data Vs Unstructured Data Vs Semi-Structured Data. *K21Academy.* Available at: https://k21academy.com/microsoft-azure/dp-900/structured-data-vs-unstructured-data-vs-semi-structured-data/. [Accessed 21 May 2021].

Wild, C.J. and Pfannkuch, M. (1999) Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67, (3) 223-265. Available at: http://iase-web.org/documents/intstatreview/99.Wild.Pfannkuch.pdf. [Accessed 21 May 2021].

Zhou, S., Qiao, Z., Du, Q., Wang, G.A., Fan, W. and Yan, X. (2018) Measuring Customer Agility from Online Reviews Using Big Data Text Analytics. *Journal of Management Information Systems*, 35, (2) 510-539. Available at: https://doi.org/10.1080/07421222.2018.1451956. [Accessed 21 May 2021].