ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# CUSTOMER SEGMENTATION IN THE ONLINE RETAIL INDUSTRY USING BIG DATA ANALYTICS

**RONALD S. CORDOVA**

Assistant Professor, Department of Computing Sciences, Gulf College, Oman

E-mail: ronald@gulfcollege.edu.om

**ABSTRACT**

Nowadays, customer data is abundant due to the growth of the online retail industry. It allows effective customer segmentation using big data analytics. This study examines how big data analytics segments online customers. Using segmentation algorithms and data processing to optimise marketing, sales, and customer experiences. Different customer segments can be identified by traits and behaviours. This research examines customer segmentation theory in the online retail industry. Data collection and preprocessing methods are discussed to ensure data quality and segmentation relevance. This study will also demonstrate how customer segmentation strategies can be used to drive digital marketing and sales campaigns and improve customer experiences using big data analytics.

Furthermore, this research will demonstrate big data analysis output using simulations. It will illustrate how big data analytics can segment customers. This will help online retailers tailor their marketing and sales to each cluster segment's preferences and behaviour. Results will emphasise data quality and algorithm choice. This research will conclude with insights on how online retailers can improve customer satisfaction and business performance through customer segmentation and big data analytics. The study found better segmentation methods that allow online retailers to use big data analytics to segment their customers more precisely. Customer insights are better with behavioral, demographic, and real-time data. This research shows that big data analytics can transform online retail by allowing businesses to switch from static, one-size-fits-all segmentation models to dynamic, data-driven approaches that better meet digital consumers' needs.

***Keywords:*** *Customer Segmentation, Big Data Analytics, Online Retail Industry, Machine Learning, Clustering Algorithms*

## 1. INTRODUCTION

The exponential growth of e-commerce has generated large amounts of consumer engagement data [1], which could improve decision-making through big data analytics customer segmentation. Businesses, especially online retailers, are using ML and BDA more [2]. Tech helps segment customers. Companies segment customers by similar traits or behaviors. This helps companies understand customers and tailor marketing and sales. A cluster-based customer segmentation uses ML/BDA.

This paper addresses a major gap in online retail: the need for a more dynamic and detailed approach to segmenting customers based on complex behavioral data. Traditional customer segmentation uses demographic or transactional data, which doesn't capture real-time customer behavior. This study addresses the following research questions:

1. Big data analytics can improve customer segmentation accuracy and depth in online retail.
2. Which behavioral, transactional, and demographic data are most effective for creating customer segments in digital retail?
3. Can big data-driven segmentation models improve customer engagement, retention, and personalization?

The paper shows that big data analytics can improve customer segmentation by making it more responsive, detailed, and aligned with dynamic online consumer behavior. It gives businesses actionable insights to tailor marketing, pricing, and product strategies to specific customer segments, improving competitiveness and customer satisfaction.

Companies can segment customers as data grows with big data analytics [3]. The study analyzes big data analytics' online retail customer segments. To improve marketing, sales, and customer experience with modern segmentation, data processing, and analysis. E-commerce platforms can target behavior. This study segments online retail customers using big data. Personalizing marketing, sales, and promotions and improving customer experiences are possible by segmenting customers by common traits and behaviors.

This study examines customer segmentation theories, data collection, preprocessing, and analysis. Big data analytics customer segmentation improves targeting, predicts future behavior, and allows immediate marketing and sales changes. Machine learning and big data analytics-driven segmentation advances are discussed. This research helps online vendors with marketing, sales, customer engagement, and business success. This study applies theory to online retail to demonstrate customer segmentation's impact. The study shows that big data analytics for consumer segmentation transforms e-commerce.

This study demonstrates how customer segmentation changes e-commerce academically. Online retail segmentation may not work well with traditional methods, so big data analytics can help. Online retail customer segmentation trends and opportunities are covered. Machine learning, exploratory data analysis, and descriptive statistics are examined. Thus, more research is needed to determine if big data analytics improves segmentation strategy accuracy and relevance.

Big data analytics in online retail customer segmentation is the focus of this study. The study seeks to:

- Understand customer segmentation theory and its relevance to online retail.
- Discover big data analytics methods for online retail customer segmentation.
- Use big data analytics to demonstrate customer segmentation's practical applications and benefits on real-world datasets.

There is a need for this study because current segmentation methods often overlook online customer behavior's dynamic nature. Traditional methods using demographic or transactional data are often too rigid to capture complex patterns in real time. Big data analytics will be used to create a more nuanced segmentation model using behavioral, transactional, and demographic data in this study. Better understanding online retail customers helps businesses grow and satisfy customers. Big data analytics can outperform demographic and transactional customer segmentation methods, according to research. Traditional methods can't capture online retail customers' complex and changing behaviors. Big data analyzes massive amounts of structured and unstructured data to help businesses understand customer behavior, preferences, and purchasing patterns. Advanced segmentation improves customer experiences through personalized marketing, product recommendations, and targeted engagement, the study found. Our insights help businesses make strategic decisions that retain customers and increase profits.

Industry relevance makes this paper strong. Big data analytics to improve customer segmentation in online retail is a major challenge. The study is feasible because many e-commerce companies use big data for customer insights. The paper clearly shows that traditional segmentation methods are flawed and recommends big data analytics. This context highlights the study's innovation and methodology. This study uses big data analytics, which is superior to previous methods. This focus puts the research in a high-demand technological space, increasing its academic and industry value.

## 2. LITERATURE REVIEW

Online retail requires customer segmentation for marketing. Big data analytics helps retailers segment customers for personalized marketing and satisfaction. This chapter describes big data analytics-based online retail customer segmentation. The literature review emphasizes online retail customer segmentation as a strategy.

E-commerce competitiveness and customer retention depend on personalized marketing, product recommendations, and customer engagement, according to research. Companies can tailor marketing, predict customer behavior, and optimize inventory and pricing by segmenting their customer base. Recent studies show that big data segmentation can improve customer experience and loyalty by enabling targeted interventions.

### 2.1 Customer Segmentation

Wassouf et al.'s 2020 case study [4] shows how big data and predictive analytics increase telecom customer loyalty. In a competitive company, data analytics can boost customer loyalty. The study says telecommunications' economic success depends on consumer loyalty. The case study shows how big data-enabled predictive analytics can analyze call logs, service consumption, and customer feedback. This study covers data prep, predictive modeling, and customer segmentation. The method predicts customer attrition, finds targeted marketing and sales opportunities, and improves customer experiences. Tabianan, Velu, and Ravi's 2022 study [5] "Utilising the K-Means Clustering Algorithm for Intelligent Customer Segmentation Based on Customer Purchase Behaviour Data" examines purchase behavior. This study says modern data analytics improves customer segmentation. In marketing and sales strategy, the study emphasizes customer segmentation.

Effective segmentation lets companies target customers with specific products and messages. Client engagement and satisfaction rise. Customer buying behavior is classified using K-means clustering, a popular unsupervised machine learning method. This study analyzes customer data with K-Means. The study examines data preparation, feature extraction, and cluster evaluation.

According to Choi et al. [6], big data analytics improves operations management, product ideas, marketing, sales, and user experiences. The project may advance knowledge by investigating how massive and complex data sets can improve industry operations, procedures, and decision-making. The report highlights big data analytics' operations management transformational potential. Advanced analytics help companies analyze massive operational data. Supply chain, manufacturing, and quality assurance may benefit from these insights.

Big data analytics in operations management has pros and cons. Examine data-driven decision-making technology, methods, and organizational best practices. This research may emphasize data-driven tactics in today's competitive corporate environment. Big data analytics may help companies make faster, better decisions, giving them an edge.

## 2.2 Big Data Analytics

Complex dataset insights come from big data analytics. Big data helps online retailers segment customers by demographics, browsing, and purchases. Data mining and machine learning reveal trends [7]. Zineb, Najat, and Jaafar [8] say big data influences e-commerce decisions. The study shows how advanced data analytics helps e-commerce. The literature emphasizes data analysis in e-commerce, where daily data volumes are large. Case study covers machine learning and AI data analysis. This study includes data preparation, predictive modeling, and possibly recommendation systems. Their approach may enhance customer service, product recommendations, and inventory management. The study found that data-driven e-commerce decision-making may boost revenue, customer satisfaction, and competitiveness.

Sun et al. [9] published a banking big data analytics framework. This study advances big data's banking consumer analytics revolution. The study emphasizes banking consumer analytics for customer satisfaction, risk management, and profit. Big data examines client transactions, digital, and demographics. The study examines iCARE's data collection, storage, processing, and analytics. Also examined were machine learning predictive modeling implementation methods and technology. Their method detects fraud, customizes client suggestions, and provides fast customer service, according to research.

Shirazi and Mohammadi [10] investigated whether big data analytics can predict retiree client turnover. For proactive client retention, the authors emphasize demographic targeting and big data analytics. In many fields, firms should predict client churn, says the study. Companies can implement retention programs by identifying high-risk customers. Study retiree behavior with big data. The authors say this method improves attrition forecasts by providing more accurate data. The study covers analytics model-building and related topics. This project may require retiree-specific data preparation, feature selection, and machine-learning. Discussions include how their data can improve marketing, sales, retired client experience, and retention.

The 2015 Arora & Malik [11] report emphasized big data business value creation through analytics. This project may examine how analytics affect large databases. Literature says analytics turns data into insights. Only good analytics can unlock the potential of large, complex big data. Scholars say analytics improves decision-making, operational efficiency, customer behavior, and trend predictions. Personalising marketing, sales, and customer experiences boosts profits and competitiveness.

## 2.3 Online Retail Industry

Online retail has grown due to technology and consumer behavior. E-commerce platforms provide massive customer data from websites, mobile apps, and social media. Retailers need this data to compete and satisfy customers. Big data analytics helps Wang, Tsai, and Ciou study integrated supply chain consumer behavior [12].

Authors demonstrate how big data improves supply chain and consumer behavior. Study emphasizes supply chain consumer behavior and customer satisfaction. The hybrid method uses big data analytics to analyze massive amounts of consumer-centric data from sales records, customer feedback, and social media, the study found.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Data preparation, predictive modeling, and machine learning are examined in hybrid analytic methods. Writing about their study's methodology may aid inventory, production, and distribution. This method helps companies implement data-driven strategies by understanding customer trends, preferences, and habits. The study claims their method boosts demand forecasts, supply chain costs, and customer satisfaction.

Mariani and Fosso Wamba [2] say big data analytics affects consumer product innovation. This study found that big data analytics firms help consumer goods companies innovate online. The report says firms must innovate to compete in changing markets.

Consumer goods innovation is needed for new products, supply chain efficiency, and customer satisfaction. The study examined consumer goods companies' big data analytics. The study claims these relationships foster innovation through data-driven decision-making, market insights, and product development. The study examines data privacy and analytics competence opportunities and challenges. RFM categorized online shoppers by Kadir and Achyar [13].

The study shows how big data analytics, especially RFM analysis, can improve e-commerce customer segmentation. This study found that online purchasing customer segmentation is essential for marketing and selling to certain consumer groups. RFM groups users by frequency, value, and recent purchases. RFM examined a massive Bukku.id shopping portal dataset in 2019. This study suggests this method may produce more accurate and valuable consumer decision-making categories.

Sales, marketing, suggestions, and customer experiences benefit from RFM analysis. This study may examine RFM analysis's pros and cons in online shopping, proving its value to marketing and sales-optimizing companies.

### 2.4 Application of Big Data Analytics in Customer Segmentation

Large data analytics helps online retailers classify customers. Advanced analytics lets retailers customise marketing and experiences. Customer satisfaction improves business performance [14].

According to Yoseph et al. [15], data mining and clustering may change big data market segmentation. Advanced data analytics enhances market segmentation and knowledge. Market segmentation for companies targeting specific customer groups with products and marketing is stressed in the report. The study discusses how data mining and clustering can help firms find market segments in massive datasets. The study's data preparation, feature selection, and clustering are described. The research also discusses how the method targets niche markets, tailors marketing and sales, and improves product development based on customer preferences. The study also suggests data-driven market segmentation may increase customer engagement, conversion, and company success.

Perera, Dilini, and Kulawansa studied CRM strategy and big data analytics [16]. User studies aid big data analytics, business consumer management, and engagement research. The study emphasizes CRM in a competitive business environment. Good CRM boosts customer retention, growth, and engagement. Big data analytics uses massive consumer data. CRM big data analytics app review. Customer behavior, preferences, and feedback affect marketing, sales, needs, and personalization. The study addresses CRM big data analytics issues. This environment needs privacy, data integration, and experts. Grishikashvili et al. [17] examine how big data has changed digital sales and marketing. This study suggests digital data growth may impact marketing, sales, and customer interaction.

Research emphasizes big data's impact on digital sales and marketing. Large internet datasets help marketers analyse customer preferences, behaviour, and trends. Marketing and sales strategies can be customized with this intelligence. Research shows big data analytics in digital sales and marketing. Online advertising, customer segmentation, and digital sales campaigns benefit from analytics. Big data ethics in digital marketing and sales include consent and security.

Sentiment analysis influences business decisions, according to Maata et al. [18]. Many studies have used text mining and sentiment analysis to assess social media customer sentiment. Unstructured text benefits from sentiment analysis. Converting data to knowledge impacts digital marketing and sales. It helps companies determine consumer preferences and tailor sales and marketing. Studies show happy customers boost company success. Positive reviews boost brand loyalty, reputation, and retention. Big data analytics helps online retailers segment and target diverse customers.

Marketing, sales, client experiences, and corporate performance benefit from data organization. The framework began with purchase data. Consolidated data is preprocessed for accuracy and consistency. Select demographic and behavioral segmentation criteria carefully to create meaningful customer categories.

This paper emphasizes clustering and machine learning. Retail marketing and sales software are similar. These segments are constantly verified for accuracy, consistency, and suitability.

Using segmentation knowledge requires implementation. Stores can target segments with marketing, sales, and products. Market and customer behavior changes require segmentation monitoring and adjustment. In competitive markets, big data analytics for customer segmentation helps online retailers. The framework enhances data, personalization, and digital retail.

## 3. METHODOLOGY

The research methodology of the study uses descriptive analytics [19] to find historical data patterns and trends. This method informs decisions, optimises strategies, and provides research insights. This section describes data customer segmentation. Includes study methods and procedures.

This research as shown in figure 1 covers data collection, preparation, analysis, and interpretation [7]. This section discusses online retail customer segmentation using big data analytics. It describes this research's methods and will help segment customers for better sales and marketing.

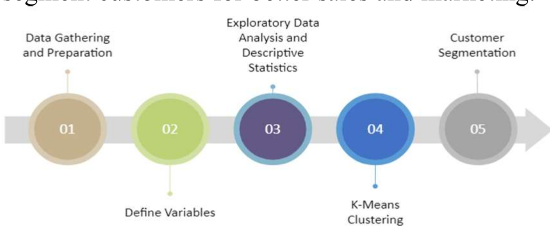



*Figure 1. The Five-Step Customer Segmentation Process*

Segmentation helps companies of all sizes understand their customers and target marketing and sales. Python's data analysis, machine learning, and visualisation libraries make it popular for consumer segmentation research. Python was used for research.

### 3.1 Data Gathering and Preparation

Multiple-source data quality, accuracy, and consistency are difficult to ensure. Data from multiple platforms and systems must be cleaned for segmentation. The Online Retail Ecommerce Dataset [20] collected customer transactions online. It contains key data from a popular online platform. The customer data was thoroughly prepared to ensure its quality and relevance for segmentation. To reduce outliers' impact on segmentation results, data cleaning, missing value deletion, and feature engineering were done. These operations compare customer relationships and data. The methods methodically analyse data to find patterns.

This study used online retail data. All transactions for a UK-based, registered non-store internet retailer between December 1, 2009, and December 9, 2011 are included. Many of this company's products are unique gifts for any occasion. Wholesalers and retailers are its main customers.

### 3.2 Define Variables

Big data analytics helps companies categorise customers by many factors to better understand and target them. Segmentation criteria adapt product and service marketing strategies. You can use demographic (age, gender), geographic (place, zip code), psychographic (interests, lifestyle), behavioural (buy history, loyalty), and other characteristics. Additionally, segmentation can consider customer lifecycle stages, purchase intent, communication medium, habits, social demographics, feedback, and reviews.

The machine learning and predictive analytics algorithms can find segmentation traits. Segmentation may use custom variables based on business goals and data sources. Environmental and economic factors may also provide data. Season and time of day improve segmentation accuracy. To customise, improve customer satisfaction, and increase profits, businesses must segment customers using big data analytics. Data-driven decision-making helps companies target the right customers with the right products.

### 3.3 Exploratory Data Analysis (EDA) and Descriptive Statistics (DS)

In consumer segmentation, the scikit-learn (sklearn) library [21] simulated and segmented customers by demographics, purchasing history, and other factors. Sklearn trained and tested consumer segmentation machine learning models [22]. Consumer segmentation analysis graphs were created using scikit-learn [23] and matplotlib [24]. Python is very useful for customer segmentation. Data analysis, machine learning, and visualization tools thrive using Python.

Python's sample code groups customers by purchase history and demographics. Reliable academic journals, industry reports, and conference proceedings provide secondary data. Secondary data on customer segmentation, big data analytics, and online retail approaches, patterns, and trends is collected through a comprehensive literature review. EDA finds correlations and patterns. Data visualisation, descriptive statistics, and correlation

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

analysis reveal client preferences [25]. EDA examines patterns and relationships. Explore Data Analysis (EDA) finds data patterns and relationships.

Descriptive statistics, data visualizations, and correlation analysis show client preferences. We find patterns, correlations, and anomalies with EDA. EDA helps consumer segmentation researchers identify customer behavior factors. Descriptive statistics calculate mean, median, mode, and variability for any dataset variable. EDA for consumer segmentation analyzes variable distribution using histograms and other representations. The correlation coefficient measures variable associations. In descriptive statistics, mean, median, and mode are calculated. We calculate standard deviation and range for each variable.

Cross-tabulates contingency table variables. Many ways EDA and descriptive statistics help consumer segmentation. The EDA identifies consumer behavior drivers. Make better consumer groups with this data. Statistical descriptions of client segments. Data improves segmented sales and marketing.

Big data aids innovation, decision-making, and insight across industries. This data-driven revolution uses EDA and descriptive statistics. Take these ideas to big data analytics. Analysis begins with exploratory data analysis. Analyze and interpret dataset trends and structures. EDA helps data scientists and analysts find outliers, missing values, and insights in massive datasets. EDA summarizes data descriptively. Data dispersion, central patterns, and variances are shown by mean, median, SD, and percentiles. Tools or libraries scale big data statistics.

Exploratory and descriptive statistics must work together in big data. Data analysis with descriptive statistics and EDA is accurate, interpretable, and actionable. Firms can streamline, make data-driven decisions, and compete. Big data improves industry innovation, decision-making, and insight. This data-driven revolution uses EDA, Customer Segmentation, and Descriptive Statistics. Use big data to analyze these ideas. Begin exploratory data analysis. Understanding data trends and structures. EDA helps data scientists and analysts find outliers, missing values, and insights in massive datasets. EDA summarizes numbers with descriptive statistics. Mean, median, standard deviation, and percentiles show data dispersion, central patterns, and variances. Dedicated tools or libraries scale big data statistics.

### 3.4 K-Means Clustering

Customer preferences and interactions change. Monitor and adjust segmentation strategies for dynamic changes. ML improves consumer segmentation. Complex customer data patterns and correlations improve segmentation precision and sophistication with ML. ML may segment customers better. Finding complex customer data patterns and relationships improves segmentation. K clusters segmented consumers. This algorithm uses behavior and demographics. Computing uses machine-learning algorithms. Computers learn and anticipate without instruction with these algorithms. Machine learning and preprocessed data segment consumers. K-means clustering classified customers by demographics and behavior. Use descriptive statistics to evaluate K-Means Clusters. Cluster mean and standard deviation show data centrality and dispersion. Statistics inform cluster significance and decision-making [26].

### 3.5 Customer Segmentation

Customers are segmented by shared traits or activities. This helps marketing and business planning. Companies can segment customers for custom products, services, and marketing. We use descriptive statistics to evaluate all Customer Segmentation sections. Each group's average purchase amount, transaction frequency, and demographics show customer preferences. Customer service and marketing personalization improve with this data. Finally, Exploratory Data Analysis, Customer Segmentation, and Descriptive Statistics analyse massive datasets. EDA and Descriptive Statistics help firms interpret data early, while Customer Segmentation optimizes processes, makes data-driven decisions, and competes. Sales increase with better customer service and targeted marketing.

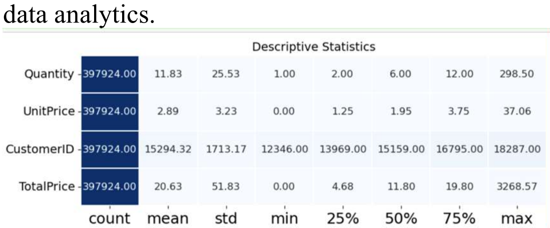
## 4. RESULTS AND DISCUSSIONS

Case studies and applications of online retail customer segmentation are here. These case studies demonstrate how big data analytics boosts online retail value and customer engagement. Segmenting customers by traits, habits, and preferences is crucial. Segmentation targets specific customer groups with marketing, sales, and services. Results from large datasets and customer trait categorization are shown here. This article discusses segmentation simulation and visualisation. Clustered consumer segment visuals show diverse customer preferences and behaviours.

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

For accurate segmentation, data quality and algorithm selection are essential. Online merchants can use big data analytics for customer segmentation with this section's theoretical and practical examples. Here's how to simulate and visualise segmentation results. Clustered client segment visualisations show consumer habits. Meaningful segmentation requires data quality and algorithm selection. This section combines theory and practice to help online retailers segment consumers using big data analytics.



Descriptive Statistics

| | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| Quantity | 397924.00 | 11.83 | 25.53 | 1.00 | 2.00 | 6.00 | 12.00 | 298.50 |
| UnitPrice | 397924.00 | 2.89 | 3.23 | 0.00 | 1.25 | 1.95 | 3.75 | 37.06 |
| CustomerID | 397924.00 | 15294.32 | 1713.17 | 12346.00 | 13969.00 | 15159.00 | 16795.00 | 18287.00 |
| TotalPrice | 397924.00 | 20.63 | 51.83 | 0.00 | 4.68 | 11.80 | 19.80 | 3268.57 |

*Figure 2. Descriptive Statistics of the Dataset before data cleaning*

Data for descriptive statistics before cleaning or preparation are in Figure 2. This is essential for understanding the dataset's start and finding issues. Quantity and UnitPrice outliers are unusual values. Manage outliers for analysis and model performance. Outlier removal, data transformation, and strong statistical metrics may be needed. Cancelled orders and returns lower UnitPrice and Quantity. These should be removed or converted into positive numbers if they indicate separate returns.

Customer analysis and product identification are hampered by missing Customer ID and Description values. Filling gaps requires data imputation/purification. Sales analysis multiplies Unit Price and Quantity to "Total Price" to calculate transaction income. Early descriptive statistics find outliers, negative values, and missing data. Well-cleansed data ensures accurate modeling and analysis..

```
# Summary of data set.
ORC.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 541909 entries, 0 to 541908
Data columns (total 8 columns):
 #   Column       Non-Null Count   Dtype
---  ------       --------------   -----
 0   InvoiceNo    541909 non-null  object
 1   StockCode    541909 non-null  object
 2   Description  540455 non-null  object
 3   Quantity     541909 non-null  int64
 4   InvoiceDate  541909 non-null  datetime64[ns]
 5   UnitPrice    541909 non-null  float64
 6   CustomerID   406829 non-null  float64
 7   Country      541909 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(4)
memory usage: 33.1+ MB
```

*Figure 3. Summary of Dataset*

Key variables from the study are shown in Figure 3. Data structure and component information are crucial to this summary. The data set has many important columns:

- Invoice number: Provides a unique invoice number for tracking and recording sales transactions.
- Stock Code: This field identifies products to organise inventories.
- Description: This column lists goods sold.
- Tracking sales volume requires knowing how many items were sold per transaction.
- Invoice Date field: records transaction dates and times for time-based analysis.
- Unit Price: displays the price of a single product, aiding in sales calculations.
- Consumer ID: Unique IDs enable segmentation and tracking.
- Country: indicates sales location, which helps market analysis and international business.

Figure 3 summarises the dataset's structure and variables, laying the groundwork for the study's analysis and findings.

| | index | Country |
|---|---|---|
| 0 | United Kingdom | 495478 |
| 1 | Germany | 9495 |
| 2 | France | 8557 |
| 3 | EIRE | 8196 |
| 4 | Spain | 2533 |
| 5 | Netherlands | 2371 |
| 6 | Belgium | 2069 |
| 7 | Switzerland | 2002 |
| 8 | Portugal | 1519 |
| 9 | Australia | 1259 |

*Figure 4. Top 10 countries having the highest orders from the online retailer.*

Modern e-commerce is worldwide, as seen in figure 4, the top 10 nations with the biggest order volumes from the online shop. These nations have adopted online buying and provided digital platforms strong trust among consumers. The UK, Germany, France and Ireland, Spain, Netherlands, Belgium, Switzerland and Portugal rank in the top 10, demonstrating Europe's strong e-commerce history. Finally, Australia shows that internet buying is popular worldwide. These rankings show that

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

customers globally are increasingly shopping online for convenience, variety, and low prices.

```
In [10]: # Descriptive statistics for StockCode.
         ORC['StockCode'].describe()

Out[10]: count      536641
         unique       4070
         top        85123A
         freq         2301
         Name: StockCode, dtype: object
```

*Figure 5. Descriptive statistics for StockCode*

Product codes are shown by StockCode descriptive statistics. As shown in figure 5, the 4,070 product numbers show a variety. The most common product code, "85123A," appears 2,301 times. This dataset shows how popular this product or item is. The prevalence of "85123A" suggests it is a best-selling, frequently purchased, or default transaction code. Understanding product code distribution aids inventory, sales, and demand forecasting. Though "85123A" is the most popular, other product codes are important. Specialty or rare purchase. Products can be optimized, tactics tailored, and hot items advertised by analyzing all product codes. StockCode descriptive statistics show commodity diversity and popularity, with "85123A" being the most popular code. Data can help companies understand their products and customers.

```
# top 10 selling products by their counts in data set.
ORC['Description'].value_counts().sort_values(ascending=False)[:10]

WHITE HANGING HEART T-LIGHT HOLDER    2357
REGENCY CAKESTAND 3 TIER              2189
JUMBO BAG RED RETROSPOT               2156
PARTY BUNTING                         1720
LUNCH BAG RED RETROSPOT               1625
ASSORTED COLOUR BIRD ORNAMENT         1488
SET OF 3 CAKE TINS PANTRY DESIGN      1465
PACK OF 72 RETROSPOT CAKE CASES       1367
LUNCH BAG  BLACK SKULL.               1323
NATURAL SLATE HEART CHALKBOARD        1272
Name: Description, dtype: int64
```

*Figure 6. Top 10 selling products by their counts in dataset*

As shown in figure 6, a company's or market's top 10 selling products by count show its most popular items. This data helps with inventory, marketing, and customer preferences. Top-selling items reveal customer preferences. Businesses can meet demand with enough of these. High-demand products can be priced higher or bundled with complementary items to optimise pricing. This data can be marketed. Promoting top-selling products may boost sales, and knowing why can help tailor marketing to similar customer preferences.

Knowing what works can inform product development and procurement. Diversifying product lines or investing in best-seller manufacturing or sourcing may result. Any company must identify and evaluate its top 10 selling products for data-driven decision-making. Income, operations, and customer satisfaction increase.

```
# Descriptive Statistics for Country
ORC['Country'].describe()

count              536641
unique                 38
top        United Kingdom
freq               490300
Name: Country, dtype: object
```

*Figure 7. Descriptive Statistics for a Particular Country*

As shown in figure 7, a dataset's customers' geographic distribution can be determined by analysing UK descriptive data. Where it says "The majority of customers reside in the United Kingdom," customers are concentrated. For better marketing, shipping, and products, companies must understand customer geography. Customer concentration lets UK marketers tailor marketing to local audiences and cultures. Inventory management can also be affected, ensuring the supply chain efficiently serves this dominant market.

This specialisation may reflect the retailer's heritage or market focus. Enterprises must invest where customers are most numerous. It may encourage outside growth or collaboration. Finally, descriptive data on a large UK client base can inform marketing, inventory, and expansion decisions, improving efficiency and customer focus.

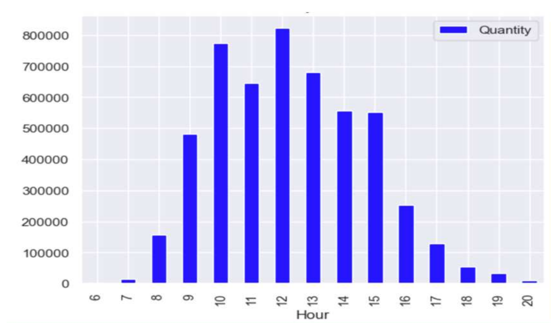



*Figure 8. Hourly Sales*

As shown in figure 8, a hourly sales data can improve corporate operations by revealing customer behaviour. "The majority of sales happen between 10 am to 3 pm" indicates peak sales. Know sales peak times to schedule staff. Businesses may increase resources during these hours to improve customer service and sales. This information can affect inventory management because popular goods may need extra restocking before peak hours. Marketing could target peak hours. Timed promotions, ads, and events may attract customers during peak hours. Sales peak knowledge aids demand forecasting and supply chain management.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

It aids production forecasting and supply chain logistics.

Knowing that most transactions happen between 10 am and 3 pm helps firms allocate resources, manage inventory, promote, and operate efficiently. It prepares the firm for peak client demand.

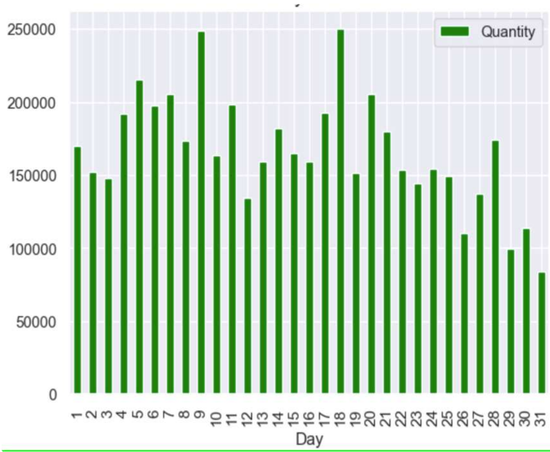



*Figure 9. Daily Sales*

As shown in figure 9, companies can use daily sales data to understand customer buying habits and make strategic decisions. The line "the majority of sales happened on the 5th, 7th, 9th, 18th, and 20th days of every month" indicates sales increases. Understanding peak sales days has many benefits. Inventory management is informed first. On critical days, businesses may stock up on product and people to meet demand. Schedule promotions or marketing around these days to maximize impact.

Patterns show customer behavior. These dates may match paydays or other spending triggers. Knowing why these peaks occur may help companies boost sales on other days. Cash flow and financial planning are affected by daily sales. Knowing when monthly revenue will rise or fall may aid budgeting. Knowing that sales peak on the 5th, 7th, 9th, 18th, and 20th of each month optimizes inventory, marketing, and financial planning, making firms more efficient and profitable.

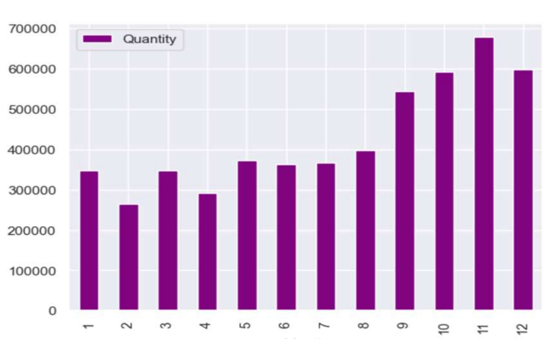



*Figure 10. Monthly Sales*

To determine product seasonality and make decisions, companies must analyse monthly sales patterns as it is shown in figure 10. September and December are the months with the most sales and peak client activity. Sales peak in September and December for several reasons. Summer's end and back-to-school shopping can boost September spending. To attract customers, many stores offer discounts and promotions. Christmas and New Year's Eve are in December. Gift-giving, Christmas shopping, and end-of-year clearance sales boost sales.

Trend analysis helps inventory management. Businesses may stock up during peak months to meet demand. It helps companies plan Christmas ads and events. These findings affect financial planning. Companies can adjust budgets and cash flow predictions to account for these months' sales swings, improving resource allocation and financial stability. For maximum profits, businesses must recognise September and December as strong sales months and manage inventory, marketing, and finance.

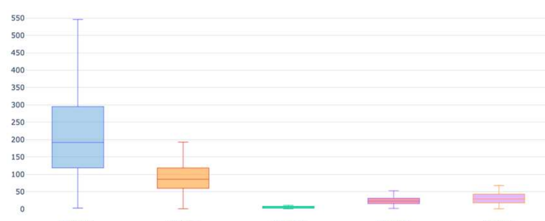



*Figure 11. Difference in Sales Frequency from Cluster to Cluster*

The difference in sales frequency from cluster to cluster helps to understand consumer behavior is shown in figure 11. The 5-cluster graphic shows category-specific consumer purchase frequency. More sales in Clusters 1 and 2 indicate segment buying. Commercial effects of realisation:

- High sales frequency may make Clusters 1 and 2 the most valuable customers. Promote these clusters for revenue.
- Inventory management: Knowing which clusters buy more may help. Company stocks serve active customers. Companies target high-selling clusters to engage customers. Personalisation and loyalty boost sales.
- Product Development: This data aids repeat-customer product development.

Journal of Theoretical and Applied Information Technology
30th November 2024. Vol.102. No. 22

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195


*Figure 12. Sales Variation Between Clusters*

Companies analyse cluster sales differences to understand customer buying habits and develop strategy as shown in figure 12. This figure with five sales amount clusters shows Clusters 1 and 2 sell most. Sales growth may make Clusters 1 and 2 the most profitable consumers. Marketers and salespeople can target and retain high-spending clients. These greatly affects business:

- Pricing Strategies: Knowing which clusters sell best helps set prices. To maximize cluster income, businesses may upsell, cross-sell, or offer premium products.
- Product Focus: Cluster spending impacts inventory and product development. Businesses may want to stock up on expensive items.
- Customer Retention: High-spending clusters maximize customer lifetime. Customer satisfaction, loyalty, and targeted marketing keep customers.

Focusing on Clusters 1 and 2's high sales will help businesses improve their customer-centric strategies, increase revenue, and boost profitability. Furthermore, data-driven decisions require cluster sales frequency differentiation. Optimising resource allocation, customer engagement, and income by targeting consumer behaviour.

This paper has several implications for academics and businesses:

1. Better Customer Experience and Personalization - Big data analytics may help online retailers understand customer behavior beyond demographics, according to the study. This deeper insight lets businesses offer personalized recommendations, targeted marketing, and customized promotions, increasing customer satisfaction and loyalty. Personalisation retains customers because they return to platforms that understand and value them.
2. Better Retail Strategy Choices - Businesses use big data segmentation to make inventory, pricing, and marketing budget decisions. Retailers can optimize inventory for predicted demand and reduce overstocking and stockout costs by understanding customer segment preferences and purchasing behaviors. Specific segment pricing can boost conversions and revenue.
3. E-commerce Competitive Edge – The research suggests that advanced customer segmentation can help online retailers compete in a saturated market. Big data analytics improves customer understanding and responsiveness, differentiating companies. Because they respond faster to customer preferences and market trends, advanced segmentation methods may perform better.
4. Inter-industry Use - This study discusses online retail, but its findings apply to industries like finance, telecommunications, and healthcare that collect significant customer data. These sectors use big data analytics to segment customers and users for personalized services and higher satisfaction. Applicability across industries emphasizes big data-driven segmentation's universal value.
5. Implementation Problems - According to the study, big data analytics for customer segmentation is difficult. These include the need for large data infrastructure investments, complex dataset management and analysis, and skilled data professionals. To satisfy customers and comply with regulations, businesses must address data privacy and ethics.
6. Future Research Implications - This research advances customer segmentation algorithms and machine learning models. Applying real-time data and predictive analytics to customer behavior could improve these models' segmentation. It also recommends using social media and mobile data to understand customer preferences.

The implications of this research show how big data analytics can transform online retail customer segmentation. It gives businesses personalization, strategic decision-making, and competitive differentiation tools. It requires a balanced approach to big data's technical, ethical, and operational challenges. This study lays the groundwork for industry and academic research in big data-driven segmentation models.

Big data analytics rely on accurate, consistent, and complete data for advanced segmentation. When data sources are inconsistent, suboptimal segmentation can lead to insights that don't fully represent customer behavior. Many organizations struggle to access third-party behavioral data due to privacy concerns or data-sharing restrictions.

Big data analytics necessitates significant infrastructure investments and technical expertise in data science and machine learning. These resource requirements may prevent smaller retailers or those with limited budgets from using these segmentation methods. Real-time data can improve segmentation, but processing it can be difficult, according to the study. Advanced processing and robust systems are needed due to data speed and volume, but they are hard to maintain. This limitation may delay the study's practical application, especially in fast-paced retail.

Big data analytics for customer segmentation raises privacy and misuse concerns. The study recognizes these issues but does not propose solutions for responsible customer segmentation data handling. They can damage customer trust and long-term relationships if ignored.

Targeted marketing is possible with predictive analytics in segmentation models, according to this study. Previous research focused on descriptive segmentation. Future behavior predictions help retailers adjust marketing and inventory. Segmentation becomes more practical with this shift from descriptive to predictive insights.

This study offers a segmentation model for online retail that can be applied to other data-intensive industries, enhancing existing literature. This adaptability expands big data segmentation beyond retail, laying the groundwork for future research in finance, healthcare, telecommunications, and other fields that value customer behavior insights.

This paper provides a flexible segmentation framework that can be applied to industries like finance, healthcare, and telecommunications that collect large amounts of customer data, in addition to online retail. The model was designed for online retail, but customer segmentation is needed in all industries that need to understand customer behavior. The study creates a flexible model that can be adapted and customized to accommodate diverse data types and segmentation goals, expanding its impact into other sectors. The model's modular structure allowed businesses to add or remove data types based on industry needs. While transactional data is important in retail, usage data may be more important in telecom. This flexible design makes the model useful for academic and industry audiences across sectors.

Traditional segmentation methods are cheaper and simpler, but they lack the depth, adaptability, and real-time capabilities needed in online retail. Big data-driven segmentation meets modern e-commerce needs with improved precision, personalization, and scalability. However, costs, complexity, and privacy issues arise. This balanced comparison shows the value of big data analytics and clarifies the trade-offs, helping businesses decide whether to adopt advanced segmentation methods.

The framework in this paper improves customer segmentation by addressing model limitations. Traditional segmentation uses demographic or transactional data, which limits complex online behavior capture. This study creates realistic customer segments using demographic, transactional, and behavioral data.

This study also introduces real-time and predictive segmentation updates based on customer behavior. This adaptability allows timely, targeted marketing strategies that respond to changing preferences and trends, making it ideal for dynamic e-commerce environments. This model optimizes customer experience, engagement, and loyalty by personalizing recommendations beyond broad segments. The model is valuable in retail, healthcare, finance, and telecommunications due to its scalability and cross-industry adaptability. The study concludes with confidentiality and compliance with the General Data Protection Regulation (GDPR) for ethical data use. This comprehensive, privacy-focused approach redefines customer segmentation, closing critical gaps in traditional models and enabling advanced, ethical segmentation in online retail as well as other industries.

This study offers a comprehensive, practical, and effective big data customer segmentation method. Data variety, computational demands, ethical concerns, and cross-industry applicability were considered in each contribution to online retail challenges. The study raises the bar for customer segmentation research in academia and industry.

## 5. CONCLUSION

Big data analytics for online retail customer segmentation has benefits and drawbacks. It explained data quality, algorithmic selection, and changing customer behaviour. Artificial intelligence, real-time data processing, predictive analytics, and ethical considerations shape consumer segmentation's future, so its potential is high. As technology and consumer behaviour evolve, customer segmentation techniques have limitless potential.

Big data analytics can better segment by combining behavioral, transactional, and demographic data, according to the study. This method provides a more complete picture of customer preferences, purchasing patterns, and engagement. Online retail segmentation requires behavioral, transactional, and demographic data.

Each data type provides unique insights for a robust segmentation model. Behavioral data shows customer interactions and preferences, transactional data shows purchase history and spending habits, and demographic data provides background information for comprehensive segmentation. Businesses can improve customer engagement with big data-driven marketing strategies and recommendations. The study shows that these models improve personalization, allowing for customer behavior prediction and timely, relevant recommendations. This proactive approach boosts customer satisfaction and retention and gives businesses an edge in the fast-paced online retail industry.

This study adds several knowledge points. The research has clarified online retail customer segmentation theory. Big data analytics improves segmentation accuracy and relevance. The study showed how online retailers use customer segmentation to improve customer engagement, optimise marketing and sales, and grow their businesses. This study offers insights for online retail professionals.

Segmentation and big data analytics help online merchants understand customer preferences and interactions. Thus, targeted marketing, sales, customised suggestions, and better customer experiences are possible. Academicians and researchers can study advanced machine learning algorithms, real-time segmentation, and predictive analytics to predict client behaviour.

Big data analytics for online retail customer segmentation is promising and growing. This research shows that segmentation-driven tactics improve customer interaction, marketing, sales, and business success. Online merchants can lead by adopting new technologies to solve their business issues. They can customise digital customer experiences to their needs.

Finally, this research advances online retail customer segmentation. It answers the introduction's key questions and provides a practical, actionable framework to help retailers move from static, generalized segmentation models to dynamic, personalized, and customer-centric ones. The transformation boosts customer satisfaction and long-term business growth. The study provides valuable insights into big data analytics in e-commerce, laying the groundwork for academic research and customer segmentation applications.

**REFERENCES:**

[1] S. Vanaja and M. Belwal, "Aspect-Level Sentiment Analysis on E-Commerce Data," *IEEE Xplore*, Jul. 01, 2018. https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8597286

[2] M. M. Mariani and S. Fosso Wamba, "Exploring how consumer goods companies innovate in the digital age: The role of big data analytics companies," Journal of Business Research, vol. 121, pp. 338–352, Dec. 2020, Available: https://www.sciencedirect.com/science/article/pii/S0148296320305956

[3] G. Wang, A. Gunasekaran, E. W. T. Ngai, and T. Papadopoulos, "Big Data Analytics in Logistics and Supply Chain management: Certain Investigations for Research and Applications," International Journal of Production Economics, vol. 176, no. 2, pp. 98–110, Jun. 2016, doi: https://doi.org/10.1016/j.ijpe.2016.03.014.

[4] W. N. Wassouf, R. Alkhatib, K. Salloum, and S. Balloul, "Predictive analytics using big data for increased customer loyalty: Syriatel Telecom Company case study," Journal of Big Data, vol. 7, no. 1, Apr. 2020, doi: https://doi.org/10.1186/s40537-020-00290-0.

[5] K. Tabianan, S. Velu, and V. Ravi, "K-Means Clustering Approach for Intelligent Customer Segmentation Using Customer Purchase Behavior Data," Sustainability, vol. 14, no. 12, p. 7243, Jun. 2022, doi: https://doi.org/10.3390/su14127243.

[6] T.-M. Choi, S. W. Wallace, and Y. Wang, "Big Data Analytics in Operations Management," Production and Operations Management, vol. 27, no. 10, pp. 1868–1883, Feb. 2018.

[7] L. Luo, B. Li, I. Koprinska, Shlomo Berkovsky, and F. Chen, "Tracking the Evolution of Customer Purchase Behavior Segmentation via a Fragmentation-Coagulation Process," Aug. 2017, doi: https://doi.org/10.24963/ijcai.2017/336.

[8] E. F. Zineb, R. Najat, and A. Jaafar, "An Intelligent Approach for Data Analysis and Decision Making in Big Data: A Case Study on E-commerce Industry," International Journal of Advanced Computer Science and Applications, vol. 12, no. 7, 2021, doi: https://doi.org/10.14569/ijacsa.2021.0120783.

[9] N. Sun, J. G. Morris, J. Xu, X. Zhu, and M. Xie, "iCARE: A framework for big data-based banking customer analytics," IBM Journal of Research and Development, vol. 58, no. 5/6, pp. 4:1–4:9, Sep. 2014, doi: https://doi.org/10.1147/jrd.2014.2337118.

[10] F. Shirazi and M. Mohammadi, "A big data analytics model for customer churn prediction in the retiree segment," International Journal of

Information Management, vol. 48, pp. 238–253, Oct. 2018, doi: https://doi.org/10.1016/j.ijinfomgt.2018.10.005.

[11] D. Arora and P. Malik, "Analytics: Key to Go from Generating Big Data to Deriving Business Value," International Conference on Big Data, Mar. 2015, doi: https://doi.org/10.1109/bigdataservice.2015.62.

[12] S.-C. Wang, Y.-T. Tsai, and Y.-S. Ciou, "A hybrid big data analytical approach for analyzing customer patterns through an integrated supply chain network," Journal of Industrial Information Integration, vol. 20, p. 100177, Dec. 2020, doi: https://doi.org/10.1016/j.jii.2020.100177.

[13] M. Kadir and A. Achyar, "Customer Segmentation on Online Retail using RFM Analysis: Big Data Case of Bukku.id," Proceedings of the International Conference on Environmental Awareness for Sustainable Development in conjunction with International Conference on Challenge and Opportunities Sustainable Environmental Development, ICEASD & ICCOSED 2019, 1-2 April 2019, Kendari, Indonesia, 2019, doi: https://doi.org/10.4108/eai.1-4-2019.2287279.

[14] Y. Zhao, X. Xu, and M. Wang, "Predicting overall customer satisfaction: Big data evidence from hotel online textual reviews," International Journal of Hospitality Management, vol. 76, no. 0278-4319, pp. 111–121, Jan. 2019, doi: https://doi.org/10.1016/j.ijhm.2018.03.017.

[15] F. Yoseph, N. H. Ahamed Hassain Malim, M. Heikkilä, A. Brezulianu, O. Geman, and N. A. Paskhal Rostam, "The impact of big data market segmentation using data mining and clustering techniques," Journal of Intelligent & Fuzzy Systems, vol. 38, no. 5, pp. 6159–6173, May 2020, doi: https://doi.org/10.3233/jifs-179698.

[16] W. K. R. Perera, K. A. Dilini, and T. Kulawansa, "A Review of Big Data Analytics for Customer Relationship Management," 2018 3rd International Conference on Information Technology Research (ICITR), Dec. 2018, doi: https://doi.org/10.1109/icitr.2018.8736131.

[17] K. Grishikashvili, S. Dibb, and M. Meadows, "Investigation into Big Data Impact on Digital Marketing," Online Journal of Communication and Media Technologies, vol. 4, no. October 2014 - Special Issue, pp. 26–37, Oct. 2014, doi: https://doi.org/10.30935/ojcmt/5702.

[18] R. L. Maata, F. Epoc, A. Pineda, and R. Cordova, "Global Business and Management Research," An International Journal, vol. 13, no. 3, 2021, Available: http://www.gbmrjournal.com/pdf/v13n3/V13N3-1.pdf

[19] Y. Wang, Q. Chen, T. Hong, and C. Kang, "Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges," IEEE Transactions on Smart Grid, vol. 10, no. 3, pp. 3125–3148, May 2019, doi: https://doi.org/10.1109/tsg.2018.2818167.

[20] "Online Retail Ecommerce Dataset," www.kaggle.com. https://www.kaggle.com/datasets/ineubytes/online-retail-ecommerce-dataset

[21] B. Bengfort and R. Bilbro, "Yellowbrick: Visualizing the Scikit-Learn Model Selection Process," Journal of Open Source Software, vol. 4, no. 35, p. 1075, Mar. 2019, doi: https://doi.org/10.21105/joss.01075.

[22] N. I. Kulin and S. B. Muravyov, "A meta-feature selection method based on the Auto-sklearn framework," Scientific and Technical Journal of Information Technologies, Mechanics and Optics, vol. 21, no. 5, pp. 702–708, Oct. 2021, doi: https://doi.org/10.17586/2226-1494-2021-21-5-702-708.

[23] Scikit-learn, "scikit-learn: machine learning in Python," Scikit-learn.org, 2019. https://scikit-learn.org/stable/

[24] Matplotlib, "Matplotlib: Python plotting — Matplotlib 3.1.1 documentation," Matplotlib.org, 2012. https://matplotlib.org/

[25] A. Alvarez P., "Exploratory Data Analysis with MATLAB, Second Edition by Wendy L. Martinez, Angel R. Martinez, Jeffrey L. Solka," International Statistical Review, vol. 79, no. 3, pp. 492–492, Nov. 2011, doi: https://doi.org/10.1111/j.1751-5823.2011.00159_13.x.

[26] L.-L. (Luke) Chiang and C.-S. Yang, "Does country-of-origin brand personality generate retail customer lifetime value? A Big Data analytics approach," Technological Forecasting and Social Change, vol. 130, pp. 177–187, May 2018, doi: https://doi.org/10.1016/j.techfore.2017.06.034.