**CUSTOMER PREFERENCE ANALYSIS**

**(Retail Industry)**

**DATA-296 Final Project**

Submitted by,

Megha R Rao (SJSU ID:013709488),

Rajasree Rajendran (SJSU ID:013774358)

Sai Chaitanya Tolem (SJSU ID:013008788)

M. S. Data Analytics, San Jose State University.

**Table of Contents**

[Abstract 3](#_Toc532597489)

[Introduction 4](#_Toc532597490)

[Literature Review 6](#_Toc532597491)

[Methods used for Customer Preference Analysis 7](#_Toc532597492)

[Any other IT skills needed to work in the area of the project 11](#_Toc532597493)

[DISCUSSION 12](#_Toc532597494)

[EMPLOYMENT PROSPECTS 15](#_Toc532597495)

[CONCLUSION 19](#_Toc532597496)

[References 22](#_Toc532597497)

# Abstract

The data science and analytics plays a pivotal role in steering retail sector towards greater productivity. Since information plays a crucial role in critical business decisions, customer preference analysis provides valuable insights that direct the course of business action. The project aimed to identify the scope, applications and methods used for this purpose. Customers, the primary focus of any business, are diverse and expect businesses to anticipate and cater to their needs. This goal can be accomplished using Big Data, Machine Learning and Data Mining methods such as cluster analysis, Support Vector Machine classification, K-means clustering and so on. In particular, cluster analysis seems very effective in preference analysis. Unsurprisingly, there is a dearth of skilled Data Practitioners in the industry having the required proficiency. Industry seeks candidates with statistical and mathematical acumen, domain knowledge along with substantial programming skills in Python, Julia, Big data, SQL, Tableau, MongoDB and Hadoop.

# 

# Introduction

*“Consumer data will be the biggest differentiator in the next two to three years. Whoever unlocks the reams of data and uses it strategically will win.”*

Above quote by Senior Vice President at Apple and former CEO at Burberry Angela Ahrendts, a leading figure in the retail and merchandising management, highlights the importance of data science and analytics in the retail industry. With every single click, data is growing at lightning pace. It is slated to become the most valuable commodity in the forthcoming century and organizations are expected to dive in and exploit data to thrive in today’s cutthroat business environment.

The industry focuses on selling goods and services to customers for their business, personal or household needs. Purchasing goods and services is largely influenced by the expectations and preferences of consumers. In this endeavor to cater to user needs and anticipate demand, businesses are obligated to enhance the overall user experience and personalize as per the individual needs. Customer retention can be improved by ensuring that customers view and receive their preferred products that meet their specifications and expectations. Let us not forget that retail industry has progressed and streamlined over the past decade due to modernization, increase in investment, access to superior tools along with global outreach. However, the growth in their customer base is threatened by an increase in the number of players in the industry. Countless possibilities presented to the customers resulted in a fastidious and demanding consumer base. Splurging customers are erratic and impatient, constantly experimenting with products. They demand constant attention and customization according to their individual needs.

Since the advent of advanced data analytics, organizations have employed machine learning and statistical algorithms to analyze their customer data and extract gainful insights. Retailers keep elaborate customer records, perform online sentiment analysis and data mine the user reviews to quench the data at hand and to understand their customer base. The algorithms learn from user’s past activity, reviews and purchases to generate useful insights. Based on preference, businesses can make decisions ranging from sourcing, product quality and quantity, tweaks in website and so on. In fact, retail industry is one of the methodical users of Machine Learning and Data Science algorithms.

Under this project, we intend to understand the various tools and methods used to analyze and understand customer preferences by the retail industry. Furthermore, we expect to learn the scope for data analytics and skills on demand in this ever - evolving business environment. For the analysis of customer preference, the retailer should have a recommendation system and filter results based on the collected data about the customer. The customer lifetime value is a factor that influences the need for recommendation for that particular customer. It is the analysis of the revenue a customer generates for the retailer. There are numerous parameters that help assess and gauge the customer behavior, currently utilized in the retail industry.

From our research we were able to comprehend the steps involved in customer preference analysis. It starts with data collection, data ingestion along with validation. This is where Big Data comes into play, since it has the capacity to process massive amounts of structured and unstructured data. The collection needs to be highly curated and well supervised as it can adversely influence the outcome. The data may not be all-encompassing or recorded in an accurate manner. The forthcoming data such as customer, product usage or promotion data, is stored in multiple repositories in varied formats based on the data architecture.

During data ingestion, the data is fetched from multiple sources and converted to a consistent format. For the initial steps, we narrowed down to tools available in Hadoop that facilitate efficient data handling. After data ingestion, Data mining helps to effectively utilize the data which has been collected. Cluster analysis uses a mathematical model to discover groups of similar customers based on small variations among the customers in each group. Although it is a relatively simple segmentation method, it efficiently generates and utilizes customer preference prototypes that are already available. Machine learning methods such as SVM classification and K-means clustering are used in further analysis of customer preferences. The various methods we explored are explained in the following in greater detail.

# Literature Review

In this section, we present the literature review by classifying

the section under the sub-titles-

In this section, we present the literature review by classifying

the section under the sub-titles-

In this section, we present the literature review by classifying

the section under the sub-titles-

In this section, we present the literature review by classifying

the section under the sub-titles-

In this section, we present the literature review for the customer preference analysis. In order to get possible required research from the published literature, a systematic literature search capturing big data with crowdsourcing work of the recent three years have been undertaken. Springer Link, ScienceDirect, Scopus, Procedia and many international journals (JMIS) were searched and we could understand the role of Data Analytics in the growth of retail industry.

Due to the exponential growth of Big Data, obtaining valuable customer experiences from large amounts of structured and unstructured data from different sources require the proper structures and tools. Organizations like Amazon and Netflix use big data in the form of recommendation engines that create value for customers by helping them search what they need. Zillow combines information to provide consolidated insight about home attributes and values, competitive properties, to buyers, sellers, and agents. Customer behavior is monitored in real time and analyzed to look carefully into processes to assess evidences and yield proper results. Web data mining is the process of applying data mining techniques to the discovery of behavior patterns based on internet data. Cooly et al. [1] presented a detailed description of data preparation methods for mining web browsing patterns. The pattern discovery tasks involve the discovery of association rules, sequential patterns, usage clusters, page clusters, user classifications or any other pattern discovery method [2]. The usage pattern extracted from internet data can be applied to a wide range of applications such as personalization, system improvement, UI modification, business intelligence discovery, etc. [3] Lee and Podlaseck [4] provided a detailed case study of the analysis of clickstream data from an online retail store. They focus on the online retailer which is also our consideration. They first analyzed the shopping behavior of customers according to the following four shopping steps: product impression, click-through, basket placement, and purchase. They then applied micro-conversion rates (e.g., click-to-buy rate), computed for each adjacent pair of these steps, to measure the effectiveness of efforts in merchandising.

# Methods used for Customer Preference Analysis

Big Data/Data Mining/Machine Learning methods used are as follows.

* **BD - Data Ingestion and Data Processing**

Managing customer data in an e-commerce site ~GBs of data/day scale and subsequently analyzing them calls for an efficient big data approach. A critical component, for providing the right customers with the right product, services or advertisement, is the ability to provide near real-time actions. The general mechanism is to be able to capture and ingest the vast amounts of data that describe the customer’s behavioral patterns, across the set of as many customers as possible. Once the large datasets (Big Data) has been ingested and acquired into the system (e.g. Hadoop ecosystem), the primary function of the engine is to sort through the data (using a variety of techniques) that can in the fastest ways possible, place like data-points together.

Customers who are more likely to purchase products need to be grouped correctly, as opposed to customers who are more likely to browse be presented with the appropriate options.

[Ref:<http://www.oracle.com/us/technologies/big-data/bda-customer-segmentation-engines-2045188.pdf>]

* **DM - Cluster Analysis**

Cluster is a group of objects that belongs to the same class. In other words, similar objects are grouped in one cluster and dissimilar objects are grouped in another cluster. While doing cluster analysis, we first partition the set of data into groups based on data similarity and then assign the labels to the groups. Clustering helps to discover distinct groups in their customer base. And they can characterize the customer groups based on the purchasing patterns.

* **Partitioning Method**

Suppose we are given a database of ‘n’ objects and the partitioning method constructs ‘k’ partition of data. Each partition will represent a cluster and k ≤ n. It means that it will classify the data into k groups, which satisfy the following requirements − Each group contains at least one object. Each object must belong to exactly one group. For a given number of partitions (say k), the partitioning method will create an initial partitioning. Then it uses the iterative relocation technique to improve the partitioning by moving objects from one group to other.

[Ref:<https://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm>]

* **ML - Classification**

In machine learning and statistics, classification is a supervised learning approach in which the computer program learns from the data input given to it and then uses this learning to classify new observation. Some examples of classification problems are: speech recognition, handwriting recognition, biometric identification, document classification etc.

[Ref:<https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>]

## 

* **SVM Classification**

In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. SVM stands for Support Vector Machine. It is a machine learning approach used for classification and regression analysis. It depends on supervised learning models and trained by learning algorithms. They analyze the large amount of data to identify patterns from them. An SVM generates parallel partitions by generating two parallel lines. For each category of data in a high-dimensional space and uses almost all attributes. It separates the space in a single pass to generate flat and linear partitions. Divide the 2 categories by a clear gap that should be as wide as possible. Do this partitioning by a plane called hyperplane. An SVM creates hyperplanes that have the largest margin in a high-dimensional space to separate given data into classes. The margin between the 2 classes represents the longest distance between closest data points of those classes. The larger the margin, the lower is the generalization error of the classifier. After training map the new data to same space to predict which category they belong to. Categorize the new data into different partitions and achieve it by training data. Of all the available classifiers, SVM provides the largest flexibility. SVMs are like probabilistic approaches but do not consider dependencies among attributes.

[Ref:<https://data-flair.training/blogs/svm-support-vector-machine-tutorial/>]

*Implementation:* Clustering->Classification->Segmentation

* **K-Means Clustering for Classification**

Clustering as a method of finding subgroups within observations is used widely in applications like market segmentation wherein, we try and find some structure in the data. Although an unsupervised machine learning technique, the clusters can be used as features in a supervised machine learning model.

[Ref:<https://towardsdatascience.com/kmeans-clustering-for-classification-74b992405d0a>]

* **Classification Vs. Segmentation**

In simple terms, segmentation is applied to the results of classification. Segmentation can’t happen without having some characteristics to use, and classification is pointless if the information is not put to use. Customer classification is the act of seeking out and identifying common traits in a group of customers. It answers a broad question: what is similar about these people and their purchasing habits? Segmentation takes that a step further by subdividing customers according to those similarities. It answers a more focused questions: what the most useful way is to group these people based on the commonalities found during classification.

[Ref:<https://conceptainc.com/blog/automatic-customer-classification-the-first-step-to-segmentation/>]

# 

# Any other IT skills needed to work in the area of the project

* **Data collection & Data Ingestion:**
* This project first requires the collection of relevant data, to do that we should know how to deal with tables, structured and unstructured data.
* ***Tools:*** SQL, Hadoop Components like Hive, Pig, Sqoop etc.
* **Data Cleaning and Mining (Data Processing):**
* After collecting the data, to clean the data before using it for that we should know the following tools.
* ***Tools****:* Python/R, Excel
* **Data Exploration:**
* Finally, when we have clean data, then we can explore it to understand the data, what insights we can draw from that data.
* ***Tools****:* Excel, Tableau
* **Data Analysis:**
* After successful exploration of data, data scientists perform in-depth analysis on that data to solve the business problem. However, this step also needs statistical and mathematical knowledge to an extent.
* ***Tools:*** Tableau

So, the best way to get started is to learn R, SQL, and Hadoop. Then have a passing understanding of Python and the tools that work with Hadoop like Hive, Pig, and others.

# DISCUSSION

Retail apocalypse of brick and mortar format is probably a hoax as per recent reports. Read about an above 5 trillion dollars in one of the recent years. Obviously, the industry outlook projects the proliferation of online marketplace and record evolution and transition in the industry itself. Next 5-year growth is expected to supersede the growth in last 100 years. The entities are faced with the mammoth task of managing copious amounts of data and integrating data in both online and store formats, prior to extracting valuable insights that can drive business decisions. This is where big data comes into play. The fragmented information in captured in different formats and databases need to be processed and the insights consolidated before making the final call on general customer preferences. For personalized customer relationship management, individual preferences, specifications and variations are to be recorded and processed in real-time before customizing the products and eradicating the issues encountered by the user.

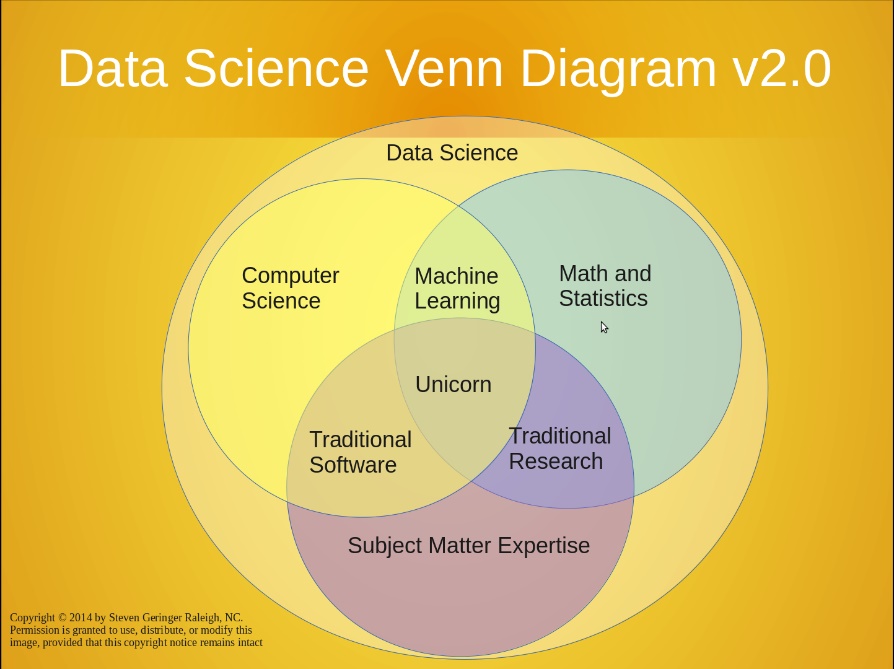
We discussed the dearth of desired skills and the need to practice, cultivate and upgrade skills to remain competitive in the industry. As per an interpretation of Drew Conway’s Venn Diagram (in the figure), the skills of an all-encompassing data practitioners are as rare as a unicorn. Thankfully, our curriculum addresses most of our concerns and incorporates core subjects with emphasis on hands-on practice experience. Henceforth, we are planning to devote more time, up to 80% commitment on weekends, towards self-exploration and exercises that hone our practical skills.

Fig.1 – Venn diagram of Data Science. Image taken from https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html

Powerful use of analytics by **Amazon** was discussed as a **case study**. They employ a host of machine learning algorithms to process customer preferences and provide swift recommendations. Furthermore, they have “anticipatory shipping” that uses predictive modeling to anticipate orders before they are placed. As per predictions, goods are shipping to local hubs. Even if the targeted customers do not buy, they offer deals and discounts to prospective buyers in the locality. No wonder Amazon is taking over the marketplace. Amazon Web Services (AWS) has recently announced free access to Amazon’s inherent e-learning portal with analytics tutorials. Over 10,000 educators and 1500 institutions already use their resources as per the website. The case study exemplifies the ubiquitous influence of data analytics in decision-making, customer satisfaction and market dominance.

Apart from discussing amongst ourselves, we sought advice from working professionals and acquaintances, who shall remain anonymous as requested. It took no time to realize how ‘hot’ Data Analytics is. Almost 80% of them talked about the occupation’s demand forecast which runs in millions and encouraged us to focus on procuring and refining the relevant skills. Nearly, more than 95% websites affirm the propositions. Data science is slated to become the sexiest job of this century, probably until automation takes over.

We conducted *an* **informational interview** with Mr. Mahesh R, working as a product manager in *Amadeus, a travel technology company* based out of Madrid, Spain. He had the following insights about the application of data analytics in the retail industry.

* Big data and machine learning applications in retail -

1. *Demand forecast* – name says all
2. *Market basket analysis* - aids in cross selling, up selling
3. *Customer segmentation* - based on purchase behavior
4. *Targeted marketing campaigns* - tailor made campaigns for each customer segments as per their preferences. Identify the stage of customer journey or purchase funnel and decide on campaigns to make them loyal customers or brand advocates. Make the results and optimize the campaigns to increase the return on ad spend utilizing individual customer preferences and behavior.
5. *Product improvements and decide on product mix*.
6. *New product ideas* - from search keywords, CRM

* Emerging technologies like virtual reality, augmented reality, 3D printing, increased use of digital channel, robotic process automation will increase the data points available to retailers. These data points can be leveraged using big data, DM and ML and AI to make decisions for the following purposes.

1. *Inventory management*
2. *Product mix*
3. *Product features*
4. *Marketing mix*
5. *CRM*

* Prudent decisions on these areas can translate to increased customer engagement, customer satisfaction, brand loyalty and advocacy. Below are the outcomes.

1. *Increase in sales /revenue*
2. *Reduction in cost of marketing and operations*
3. *Increase in profitability*
4. *Increase in brand equity*

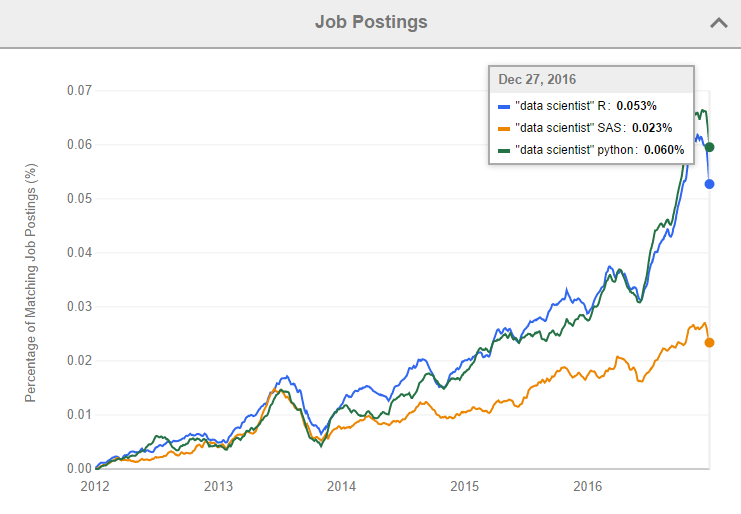
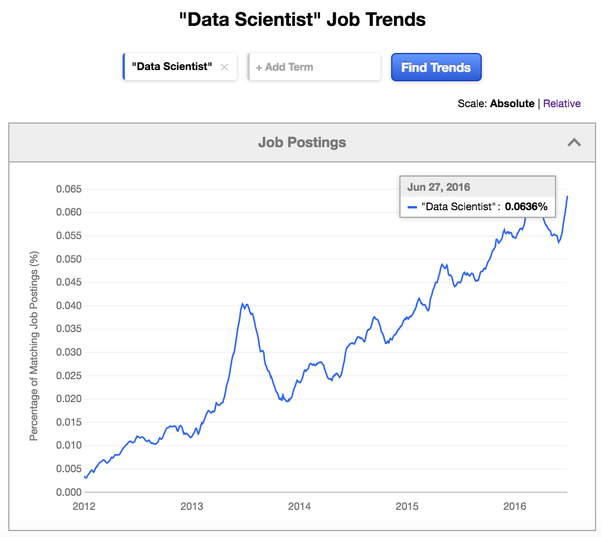
According to Mr. Mahesh R, the average annual salary for the cream of the industry in India is around Rs.70 lakhs (97k USD), which is about 6 times higher than the rest of the tech job salaries.

# EMPLOYMENT PROSPECTS

With the advent of world wide web, data has been growing at astronomical rate. Companies mine and dissect customer preferences of their clientele to unearth actionable insights and require the assistance of a relatively new breed of data practitioners who go by various names (such as data analyst, business analyst, data scientist, data engineer and so on). Equipped with a powerful and rare combination of skills in statistics, programming duly supplemented by domain knowledge, aspirants must act as a hybrid of data hacker, miner, business analyst, communicator, and trusted adviser. A typical day is way more challenging and tedious than envisioned by most beginners and demands serious contemplation and usage of analytical skills. As per Glassdoor (refer figure), Data Scientist has been named the best job in America with a median base salary of $110K. There are over 16,000 openings on Glassdoor and the median salary of analytics jobs at >$94K. We can compare the data with the extract from indeed in the coming pages.

*Figure:* Job postings for Data scientist with R, SAS and Python skills.

[Ref: <https://blogs-images.forbes.com/louiscolumbus/files/2018/01/top-50-best-jobs-in-america-2018.jpg>]



*Figure:* Job trend graph for Data scientist from 2012 to 2016

[Ref:[https: [Data Scientist" Job Trends](http://www.indeed.com/jobtrends/q-%22Data-Scientist%22.html) on [Indeed.com](http://indeed.com/)](https://blog.revolutionanalytics.com/2017/02/job-trends-for-r-and-python.html)]

Job trend graph shows the increased demand for data practitioners. Some of the major players are Google, Amazon, Microsoft, Walmart, eBay, LinkedIn, and Twitter. Below is a job search extract from indeed.com, a popular job search engine. Number of search results / hits are in brackets.

|  |  |  |  |
| --- | --- | --- | --- |
| **Data science**  **in retail** | **Data engineer**  **in retail** | **Data analyst**  **in retail** | **Business analyst**  **in retail** |
| Salary Estimate   * [$60,000](https://www.indeed.com/q-Data-Science-Retail-$60,000-jobs.html)(35) * [$95,000](https://www.indeed.com/q-Data-Science-Retail-$95,000-jobs.html)(27) * [$100,000](https://www.indeed.com/q-Data-Science-Retail-$100,000-jobs.html)(23) * [$111,000](https://www.indeed.com/q-Data-Science-Retail-$111,000-jobs.html)(15) * [$130,000](https://www.indeed.com/q-Data-Science-Retail-$130,000-jobs.html)(9) | Salary Estimate   * [$30,000](https://www.indeed.com/q-Data-engineer-Retail-$30,000-jobs.html)(190) * [$40,000](https://www.indeed.com/q-Data-engineer-Retail-$40,000-jobs.html)(146) * [$46,100](https://www.indeed.com/q-Data-engineer-Retail-$46,100-jobs.html)(102) * [$65,000](https://www.indeed.com/q-Data-engineer-Retail-$65,000-jobs.html)(71) * [$100,000](https://www.indeed.com/q-Data-engineer-Retail-$100,000-jobs.html)(43) | Salary Estimate   * [$30,000](https://www.indeed.com/q-Data-Analytics-Retail-$30,000-jobs.html)(380) * [$35,000](https://www.indeed.com/q-Data-Analytics-Retail-$35,000-jobs.html)(332) * [$45,000](https://www.indeed.com/q-Data-Analytics-Retail-$45,000-jobs.html)(240) * [$60,000](https://www.indeed.com/q-Data-Analytics-Retail-$60,000-jobs.html)(166) * [$85,000](https://www.indeed.com/q-Data-Analytics-Retail-$85,000-jobs.html)(87) | Salary Estimate   * [$45,000](https://www.indeed.com/q-Business-analyst-Retail-$45,000-jobs.html)(107) * [$60,000](https://www.indeed.com/q-Business-analyst-Retail-$60,000-jobs.html)(88) * [$80,000](https://www.indeed.com/q-Business-analyst-Retail-$80,000-jobs.html)(64) * [$85,000](https://www.indeed.com/q-Business-analyst-Retail-$85,000-jobs.html)(54) * [$100,000](https://www.indeed.com/q-Business-analyst-Retail-$100,000-jobs.html)(28) |
| Job Type   * [Full-time](https://www.indeed.com/jobs?q=Data+Science+Retail&jt=fulltime)(27) * [Contract](https://www.indeed.com/jobs?q=Data+Science+Retail&jt=contract)(9) * [Part-time](https://www.indeed.com/jobs?q=Data+Science+Retail&jt=parttime)(5) * [Internship](https://www.indeed.com/jobs?q=Data+Science+Retail&jt=internship)(3) | Job Type   * [Full-time](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=fulltime)(156) * [Contract](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=contract)(72) * [Part-time](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=parttime)(65) * [Internship](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=internship)(20) * [Temporary](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=temporary)(8) * [Commission](https://www.indeed.com/jobs?q=Data+engineer+Retail&jt=commission)(4) | Job Type   * [Full-time](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=fulltime)(392) * [Part-time](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=parttime)(66) * [Contract](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=contract)(27) * [Internship](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=internship)(19) * [Commission](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=commission)(15) * [Temporary](https://www.indeed.com/jobs?q=Data+Analytics+Retail&jt=temporary)(14) | Job Type   * [Full-time](https://www.indeed.com/jobs?q=Business+analyst+Retail&jt=fulltime)(85) * [Contract](https://www.indeed.com/jobs?q=Business+analyst+Retail&jt=contract)(54) * [Part-time](https://www.indeed.com/jobs?q=Business+analyst+Retail&jt=parttime)(5) * [Commission](https://www.indeed.com/jobs?q=Business+analyst+Retail&jt=commission)(3) * [Internship](https://www.indeed.com/jobs?q=Business+analyst+Retail&jt=internship)(3) |
| Location   * [New York, NY](https://www.indeed.com/jobs?q=Data+Science+Retail&rbl=New+York,+NY&jlid=45f6c4ded55c00bf)(9) * [Chicago, IL](https://www.indeed.com/jobs?q=Data+Science+Retail&rbl=Chicago,+IL&jlid=402d6ad50e16c894)(4) * [SF, CA](https://www.indeed.com/jobs?q=Data+Science+Retail&rbl=San+Francisco,+CA&jlid=6cf5e6d389fd6d6b)(4) * [Palo Alto, CA](https://www.indeed.com/jobs?q=Data+Science+Retail&rbl=Palo+Alto,+CA&jlid=417f86aed2946c16)(2) * [Portland, OR](https://www.indeed.com/jobs?q=Data+Science+Retail&rbl=Portland,+OR&jlid=55fe4d923dbe0a35)(2) | Location   * [New York, NY](https://www.indeed.com/jobs?q=Data+engineer+Retail&rbl=New+York,+NY&jlid=45f6c4ded55c00bf)(13) * [Atlanta, GA](https://www.indeed.com/jobs?q=Data+engineer+Retail&rbl=Atlanta,+GA&jlid=966e6327a98f7e81)(5) * [Austin, TX](https://www.indeed.com/jobs?q=Data+engineer+Retail&rbl=Austin,+TX&jlid=d2a39b6d57d82344)(4) * [SF, CA](https://www.indeed.com/jobs?q=Data+engineer+Retail&rbl=San+Francisco,+CA&jlid=6cf5e6d389fd6d6b)(4) * [Seattle, WA](https://www.indeed.com/jobs?q=Data+engineer+Retail&rbl=Seattle,+WA&jlid=1e8a7dce52945215)(3) | Location   * [New York, NY](https://www.indeed.com/jobs?q=Data+Analytics+Retail&rbl=New+York,+NY&jlid=45f6c4ded55c00bf)(33) * [SF, CA](https://www.indeed.com/jobs?q=Data+Analytics+Retail&rbl=San+Francisco,+CA&jlid=6cf5e6d389fd6d6b)(24) * [Dallas, TX](https://www.indeed.com/jobs?q=Data+Analytics+Retail&rbl=Dallas,+TX&jlid=c9b29a6e6a9f190c)(13) * [Chicago, IL](https://www.indeed.com/jobs?q=Data+Analytics+Retail&rbl=Chicago,+IL&jlid=402d6ad50e16c894)(12) * [LA, CA](https://www.indeed.com/jobs?q=Data+Analytics+Retail&rbl=Los+Angeles,+CA&jlid=d05a4fe50c5af0a8)(10) | Location   * [New York, NY](https://www.indeed.com/jobs?q=Business+analyst+Retail&rbl=New+York,+NY&jlid=45f6c4ded55c00bf)(20) * [Palo Alto, CA](https://www.indeed.com/jobs?q=Business+analyst+Retail&rbl=Palo+Alto,+CA&jlid=417f86aed2946c16)(4) * [San Antonio, TX](https://www.indeed.com/jobs?q=Business+analyst+Retail&rbl=San+Antonio,+TX&jlid=dce72acfd1e9aa3f)(4) * [Flushing, NY](https://www.indeed.com/jobs?q=Business+analyst+Retail&rbl=Flushing,+NY&jlid=531a2dda0ec026fa)(4) * [Boston, MA](https://www.indeed.com/jobs?q=Business+analyst+Retail&rbl=Boston,+MA&jlid=e167aeb8a259bcac)(3) |
| Experience Level   * [Entry Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=entry_level)(26) * [Mid Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=mid_level)(8) * Senior Level (1) | Experience Level   * [Entry Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=entry_level)(145) * [Mid Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=mid_level)(35) * Senior Level (9) | Experience Level   * [Entry Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=entry_level)(293) * [Mid Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=mid_level)(102) * Senior Level (15) | Experience Level   * [Entry Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=entry_level)(81) * [Mid Level](https://www.indeed.com/jobs?q=Business+analyst+Retail&explvl=mid_level)(29) * Senior Level (4) |

[Ref: <https://www.indeed.com/jobs>?]

Please note that experience level is an optional field and many employers do not specify the same.

# CONCLUSION

In a world filled with countless choices, customers are expecting a superior personalized experience. No two customers are same and when brands consider their preferences and execute on them, engagement, purchases and customer satisfaction soar. Further, brands can forge a long – term relationship and retain customers, along with the quintessential capacity to anticipate demands, tweak the product mix, upgrade inventory, expedite shipping and tailor targeted campaigns, thereby enabling companies to maintain a competitive edge over their rivals.

For full disclosure, all statistics are referenced from v12data.com and thebalance.com. More details in References section.

Undoubtedly, the industry is in transition owing to digital transformation, and trends reveal a dramatic shift to elevated customer-centric experiences in the coming years, rendering retail analytics key to thriving in business. Quoting an Accenture study - 75% of consumers are more likely to buy from a retailer that recognizes them by name, recommends options based on past purchases, knows their purchase history. As a result, entities focus 54% on customer experience, 14% on data-driven marketing, 11% on mobile applications, 4% on programmatic buying/optimization and 16% on cross-channel marketing according to a survey. Particularly for omni-channel retailers amassing and processing humongous portions of data, if trailing behind, it is high time to recalibrate, upgrade or institute methods to curate products and honor customer preferences. In fact, one third of retailers cited “Targeting and personalization” among their tactical practices for the years ahead. Customer analytics is an important component of retail analytics that enables curating experience and building predictive models. According to a survey by future Zebra, 68% respondents plan on investing in machine learning and cognitive computing by 2025. 73[% of retailers deem big data vital to survival while 78% plan to device an omni-channel integration, streamlined to customer preference and capable of a satisfactory, complete and holistic experience.](https://twitter.com/intent/tweet?url=http%3A%2F%2Fwww.forbes.com%2Fsites%2Flouiscolumbus%2F2017%2F03%2F19%2Finternet-of-things-will-revolutionize-retail%2F&text=73%25%20of%20retailers%20rate%20managing%20big%20data%20as%20important%20or%20business-critical%20to%20their%20operations.%20%23bigdata)

A 2025 forecast predicted a 10-fold explosion in data. It is no mystery that data is growing at meteoric pace, making the ability to make sense of the data central to maintaining customer affinity. Brands rely on solid data, metrics and facts to derive insights and support business decisions. However, customer preferences are not static and vary with age, seasonality, income, demographic or fashion trends, making the endeavor exasperatingly complex to execute manually. Data science and analytics comes to their rescue with a toolkit of myriad methods. Big data, machine learning and data mining techniques are employed for this purpose. Therefore, the corresponding skills (in Python, R, Julia, Tableau, Excel, SQL and NoSQL, Hadoop etc.) are in demand and the industry is seeking to absorb well-rounded candidates, with proficiency in statistics, programming along with domain knowledge, in the workplace.

*Figure: Retail Analytics.*

*Ref: https://dzone.com/articles/coffee-with-a-data-scientist-tuhin*

# References

[1] R. Cooley, B. Mobasher, J. Srivastava. Data preparation for mining world wide Web browsing patterns, Journal of Knowledge and Information Systems, 1 (1999).

[2] B. Mobasher, R. Cooley, J. Srivastava. Automatic personalization based on Web usage mining, Communications of the ACM, 43 (8) (2000), pp. 142-151.

[3] J. Srivastava, R. Cooley, M. Deshpande, P. Tan. Web usage mining: discovery and applications of usage patterns from Web data, SIGKDD Explorations, 1 (2) (2000), pp. 1-12.

[4] J. Lee, M. Podlaseck, E. Schonberg, R. Hoch. Visualization and analysis of clickstream data of online stores for understanding Web merchandising, Data Mining and Knowledge Discovery, 5 (1/2) (2001), pp. 59-84.

[5] Conceptainc website, <https://conceptainc.com/blog/automatic-customer-classification-the-first-step-to-segmentation/>]

[6] Quora.com, <https://www.quora.com/What-tools-do-data-scientists-use>]

[7] Data flair website, <https://data-flair.training/blogs/svm-support-vector-machine-tutorial/>]

[8] Oracle Customer segmentation engines. <http://www.oracle.com/us/technologies/big-data/bda-customer-segmentation-engines-2045188.pdf>]

[9] Data mining on Tutorialspoint. <https://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm>]

[10] Towards Data Science website. <https://towardsdatascience.com/kmeans-clustering-for-classification-74b992405d0a>]

[11] KDNuggets. <https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html>

[12] <https://www.indeed.com/jobs>?], <https://www.indeed.com/>

[13] <https://www.retailtouchpoints.com/resources/type/infographics/brave-new-world-evolving-retail-in-the-21st-century-infographic>

[14] <https://www.dmnews.com/data/article/13036702/the-importance-of-customer-preference>

[15] <https://www.kdnuggets.com/2018/02/data-science-improve-retail.html>

[16] <http://paidcontent.mysmallbusiness.com.au/amex/customers/article/importance-giving-customers-choice/>

[17] <https://www.smartdatacollective.com/amazon-wants-predictive-analytics-offer-anticipatory-shipping/>

[18] <http://www.intelligencenode.com/knowledge/retail-analytics-big-data/>

[19] <https://www2.deloitte.com/us/en/pages/consumer-business/articles/retail-distribution-industry-outlook.html>

[20] <https://aws.amazon.com/education/awseducate/>

[21] <https://www.thebalance.com/u-s-retail-sales-statistics-and-trends-3305717>

[22] <https://www.v12data.com/blog/50-statistics-about-retail-marketing-and-consumer-shopping-trends/>

[23] <https://www.forbes.com/sites/louiscolumbus/2018/01/29/data-scientist-is-the-best-job-in-america-according-glassdoors-2018-rankings/#60a74d155357>;

[24] <https://blogs-images.forbes.com/louiscolumbus/files/2018/01/top-50-best-jobs-in-america-2018.jpg>

[25] <https://www.forbes.com/sites/louiscolumbus/2017/03/19/internet-of-things-will-revolutionize-retail/#462df6f65e58>

[26] <https://blog.revolutionanalytics.com/2017/02/job-trends-for-r-and-python.html>