

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

<u>ABSTRACT</u>	<u>3</u>
<u>INTRODUCTION</u>	<u>4</u>
<u>LITERATURE REVIEW</u>	<u>6</u>
<u>METHODS USED FOR CUSTOMER PREFERENCE ANALYSIS</u>	<u>7</u>
<u>ANY OTHER IT SKILLS NEEDED TO WORK IN THE AREA OF THE PROJECT</u>	<u>10</u>
<u>DISCUSSION</u>	<u>12</u>
<u>EMPLOYMENT PROSPECTS</u>	<u>15</u>
<u>CONCLUSION</u>	<u>19</u>
<u>REFERENCES</u>	<u>21</u>

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.

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

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.

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

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

Under Literature Review, we explored existing research from published literature online, primarily concentrating on studies on Retail data analytics. Springer Link, ScienceDirect, Scopus, Procedia and many international journals (JMIS) were perused to gain an understanding on 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 demand the use of advanced computational 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 has information to provide consolidations about home attributes and values and 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. Cooley 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.

Managing customer data in an e-commerce site ~GBs of data/day scale and subsequently analyzing them calls for an efficient big data approach. An essential component, for providing the right customers with the right services, products or advertisement, is to generate 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. After the dataset is ingested into the system (e.g. Hadoop ecosystem), The data needs to sort using various

techniques like placing the data-points together. Potential customers who are likely to buy products should be grouped and customers who are likely to just browse should be provided with appropriate options.

➔ DM - Cluster Analysis

Cluster is a collection of objects that belong to the same category. Meaning, similar objects are grouped under one cluster and dissimilar objects are grouped under a different cluster. In cluster analysis, we partition the data into groups based on data similarity and assign the labels for them. Clustering helps to discover distinct groups in their customer base. And they can characterize the customer groups based on the purchasing patterns.

➔ Partitioning Method

In this method, we are given a database of 'n' objects and the partitioning method constructs 'k' partition of data in such a way that each partition contains at least one object. which means, it will classify the data into k partitions. Here each partition is represented as a cluster and $k \leq n$. 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.

➔ ML – Classification

In machine learning, classification is a supervised learning model. In this method, the model is trained with input data and then classify new observation.

Typical examples of classification are: speech recognition, handwriting recognition, biometric identification, document classification etc.

➔ SVM Classification

SVM stands for Support Vector Machine. Support vector machines, in machine learning, are supervised learning models which analyze data used for classification and regression analysis. They work on large amounts of data to identify patterns in them. SVM generates parallel partitions by creating two parallel lines and divides those 2 lines as wide as possible. Flat and linear separations. This is done to separate the space to generate flat and linear partitions. This partitioning resembles a plane separation and is termed as hyperplane separation. Hyperplanes are objects which have the largest margin in a high-dimensional space. They can separate data into different classes. An SVM can create hyperplanes. The margin between the 2 classes represents the longest distance between the closest data points of those classes. In the classifier, larger margin results in lower generalization error. The algorithm is fed with training data for mapping the new data to appropriate partitioned categories. SVM is known to be the most flexible classifier. In a way, they are referred to as a probabilistic approach, but attribute dependencies are not considered.

Implementation: Clustering -> Classification -> Segmentation

➔ K-Means Clustering for Classification

Clustering is a machine learning algorithm widely used in applications like market segmentation wherein, we try to identify some patterns in the data. The idea here is to use the features generated

from clustering machine learning algorithm use the features in supervised machine learning model in a supervised machine learning algorithm

➔ **Classification Vs. Segmentation**

We apply segmentation to the results of classification. Segmentation can be done only after classification as segmentation require some characteristics and classification becomes irrelevant if data is not put into use. Customer classification is all about finding the common traits among a set of customers. Customer classification is done to know purchasing habits of the customer and similarities they share. Segmentation takes it a step further by subclassifying customers according to the identified similarities. Another way is to group those customers based on the commonalities found during the classification phase.

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, NoSQL, 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 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

➔ 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. This step needs mathematical knowledge and statistical to an extent.
- *Tools:* Tableau, MongoDB

So, the best way to get started is to have a good understanding of python and learn R, Hive, Tableau, NoSQL, SQL, and Hadoop.

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.

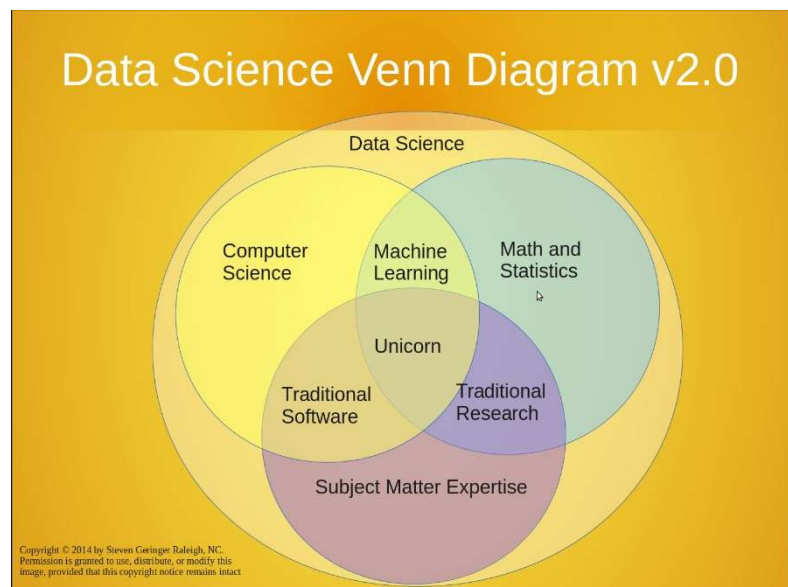


Figure 1 - Venn diagram of Data Science. Image taken from <https://www.kdnuggets.com/2016/10/battle-data-science-venn-diagrams.html>

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

Fortunately, 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.

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.

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

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, surveys and 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 annual salary for the cream or the experts of the industry in India can go up to 70 lakhs INR (97k USD), which is about 5 times higher than the rest of the tech job salaries (for experts with experience).

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.

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

As per Glassdoor (refer figure), Data Scientist job has been named one of the best occupations in America (year 2018) 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 2 – Top 50 Jobs in America 2018;
Source :Glassdoor, Re-printed from
<https://www.forbes.com/sites/louiscolumbus/2018/01/29/data-scientist-is-the-best-job-in-america-according-glassdoors-2018-rankings/#6f4428c85535>

Ranking	Job	Median Base Salary	Job Score (5.0 scale)	Job Satisfaction (5.0 scale)	Job Openings
1	Data Scientist	\$ 110,000	4.8	4.2	4,524
2	DevOps Engineer	\$ 105,000	4.6	4	3,369
3	Marketing Manager	\$ 85,000	4.6	4	6,436
4	Occupational Therapist	\$ 74,000	4.5	4	11,903
5	HR Manager	\$ 85,000	4.5	3.9	4,458
6	Electrical Engineer	\$ 76,000	4.5	3.9	5,839
7	Strategy Manager	\$ 135,000	4.5	4.2	1,195
8	Mobile Developer	\$ 90,000	4.5	4.1	1,809
9	Product Manager	\$ 113,000	4.4	3.7	7,531
10	Manufacturing Engineer	\$ 72,000	4.4	4	4,241
11	Compliance Manager	\$ 96,000	4.4	4.3	1,222
12	Finance Manager	\$ 116,000	4.4	3.8	2,998
13	Risk Manager	\$ 97,000	4.4	4.2	1,209
14	Business Development Manager	\$ 75,000	4.4	3.9	4,060
15	Front End Engineer	\$ 100,000	4.4	4.2	1,222
16	Site Reliability Engineer	\$ 120,000	4.4	4.1	1,064
17	Mechanical Engineer	\$ 75,000	4.4	3.8	5,079
18	Analytics Manager	\$ 115,000	4.4	3.9	1,381
19	Tax Manager	\$ 110,000	4.4	3.7	3,309
20	Creative Manager	\$ 110,000	4.3	4.3	824
21	Software Engineer	\$ 102,500	4.3	3.6	29,187
22	Hardware Engineer	\$ 115,000	4.3	4.2	806
23	Corporate Recruiter	\$ 65,000	4.3	4.3	2,330
24	QA Manager	\$ 92,000	4.3	3.8	1,741
25	Physician Assistant	\$ 104,000	4.3	3.6	5,517
26	Database Administrator	\$ 94,000	4.3	3.8	2,370
27	UX Designer	\$ 90,000	4.3	3.8	1,963
28	Nursing Manager	\$ 84,660	4.3	3.7	4,209
29	Engagement Manager	\$ 115,000	4.3	3.7	2,169
30	Solutions Architect	\$ 125,000	4.2	3.6	3,325
31	Process Engineer	\$ 78,000	4.2	3.8	3,033
32	Reliability Engineer	\$ 92,000	4.2	4.3	747
33	Data Engineer	\$ 100,000	4.2	3.7	2,816
34	Operations Manager	\$ 65,000	4.2	3.8	13,706
35	Speech Language Pathologist	\$ 72,000	4.2	3.7	11,573
36	Communications Manager	\$ 80,000	4.2	3.9	1,380
37	Audit Manager	\$ 100,000	4.2	3.7	1,951
38	Data Analyst	\$ 60,000	4.2	3.9	4,729
39	Systems Analyst	\$ 75,000	4.2	3.7	2,710
40	Facilities Manager	\$ 75,000	4.2	3.8	2,139
41	Strategic Account Manager	\$ 85,000	4.2	4.1	808
42	Business Intelligence Developer	\$ 86,000	4.1	3.9	882
43	Business Analyst	\$ 71,000	4.1	3.6	9,603
44	Accounting Manager	\$ 82,000	4.1	3.6	3,273
45	UI Developer	\$ 95,000	4.1	3.8	1,004
46	Executive Assistant	\$ 55,000	4.1	3.9	4,684
47	Management Consultant	\$ 110,000	4.1	3.8	1,024
48	Project Manager	\$ 80,000	4.1	3.5	23,274
49	Nurse Practitioner	\$ 100,000	4.1	3.5	8,510
50	HR Generalist	\$ 57,210	4.1	3.9	2,705

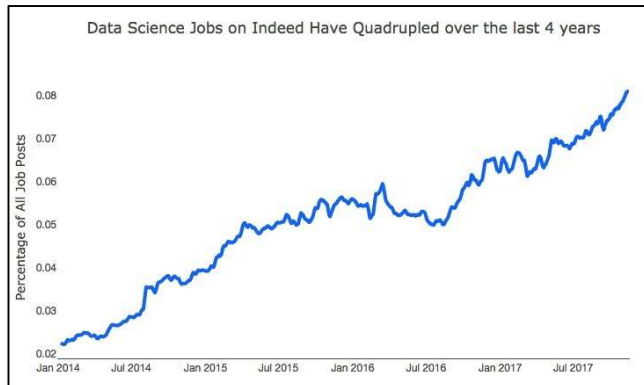
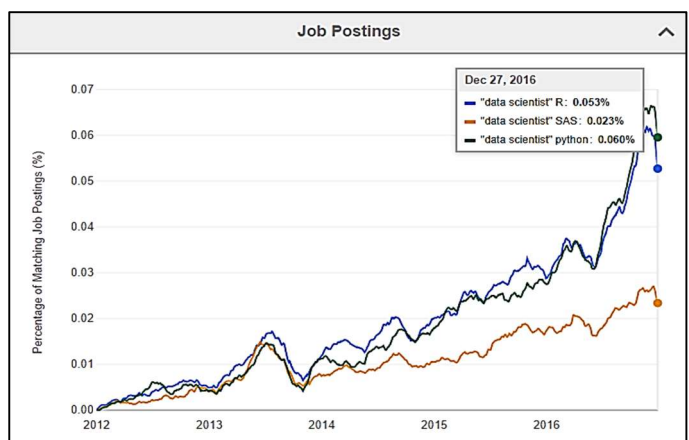


Figure 3 - Job postings for Data scientist with R, SAS and Python skills.
[Ref: <https://blogs-images.forbes.com/louiscolumbus/files/2018/01/top-50-best-jobs-in-america-2018.jpg>]

Figure 4 – Data science Jobs on Indeed in last 4 years quadrupled, Data Source: Indeed. Reprinted from -

Ref: <https://medium.com/indeed-engineering/transitioning-from-academia-to-industry-perspectives-from-indeeds-data-scientists-de890acd1bfc>



DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

Data science in retail	Data engineer in retail	Data analyst in retail	Business analyst in retail
Salary Estimate <ul style="list-style-type: none"> • \$60,000(35) • \$95,000(27) • \$100,000(23) • \$111,000(15) • \$130,000(9) 	Salary Estimate <ul style="list-style-type: none"> • \$30,000(190) • \$40,000(146) • \$46,100(102) • \$65,000(71) • \$100,000(43) 	Salary Estimate <ul style="list-style-type: none"> • \$30,000(380) • \$35,000(332) • \$45,000(240) • \$60,000(166) • \$85,000(87) 	Salary Estimate <ul style="list-style-type: none"> • \$45,000(107) • \$60,000(88) • \$80,000(64) • \$85,000(54) • \$100,000(28)
Job Type <ul style="list-style-type: none"> • Full-time(27) • Contract(9) • Part-time(5) • Internship(3) 	Job Type <ul style="list-style-type: none"> • Full-time(156) • Contract(72) • Part-time(65) • Internship(20) • Temporary(8) • Commission(4) 	Job Type <ul style="list-style-type: none"> • Full-time(392) • Part-time(66) • Contract(27) • Internship(19) • Commission(15) • Temporary(14) 	Job Type <ul style="list-style-type: none"> • Full-time(85) • Contract(54) • Part-time(5) • Commission(3) • Internship(3)
Location <ul style="list-style-type: none"> • New York, NY(9) • Chicago, IL(4) • SF, CA(4) • Palo Alto, CA(2) • Portland, OR(2) 	Location <ul style="list-style-type: none"> • New York, NY(13) • Atlanta, GA(5) • Austin, TX(4) • SF, CA(4) • Seattle, WA(3) 	Location <ul style="list-style-type: none"> • New York, NY(33) • SF, CA(24) • Dallas, TX(13) • Chicago, IL(12) • LA, CA(10) 	Location <ul style="list-style-type: none"> • New York, NY(20) • Palo Alto, CA(4) • San Antonio, TX(4) • Flushing, NY(4) • Boston, MA(3)
Experience Level <ul style="list-style-type: none"> • Entry Level(26) • Mid Level(8) • Senior Level (1) 	Experience Level <ul style="list-style-type: none"> • Entry Level(145) • Mid Level(35) • Senior Level (9) 	Experience Level <ul style="list-style-type: none"> • Entry Level(293) • Mid Level(102) • Senior Level (15) 	Experience Level <ul style="list-style-type: none"> • Entry Level(81) • Mid Level(29) • Senior Level (4)

[Table 1 – Job search results from Indeed.com; Ref: <https://www.indeed.com/jobs?>]

Please note that experience level is an optional field and many employers do not specify the same. Number of search results / hits are in brackets. Job trend graph shows the increased demand for data practitioners. Some of the major players are Google, Amazon, Microsoft,

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

Walmart, eBay, LinkedIn, and Twitter. Above is a job search extract from indeed.com and below is a salary search from Glassdoor, both popular job search engine.

Glassdoor paints a pretty picture with high average remuneration for data practitioners.

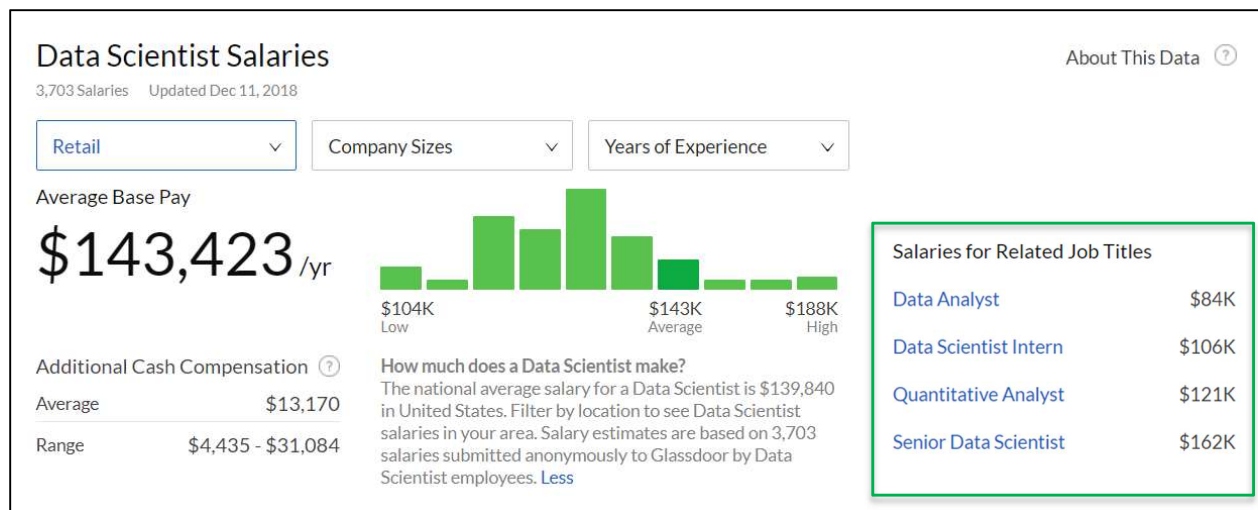


Figure 5 – Salary for Data Practitioners with varied experience levels
Ref: https://www.glassdoor.com/Salaries/data-scientist-salary-SRCH_KO0,14.htm#



Figure 6 - Salary for entry – level Data Practitioners
Ref: https://www.glassdoor.com/Salaries/data-scientist-salary-SRCH_KO0,14.htm#

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.

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

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.

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 7: 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, <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/jobs?;); <https://www.indeed.com/>; <https://www.glassdoor.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/>

DATA 296 PROJECT - CUSTOMER PREFERENCE ANALYSIS

- [17] <https://www.smartdatacollective.com/amazon-wants-predictive-analytics-offer-anticipatory-shipping/>
- [18] <http://www.intelligencecode.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/#6f4428c85535>; <https://blogs-images.forbes.com/louiscolumbus/files/2018/01/top-50-best-jobs-in-america-2018.jpg>
- [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>
- [27] <https://www.edvancer.in/50-amazing-big-data-and-data-science-quotes-to-inspire-you/>
- [28] <https://dzone.com/articles/coffee-with-a-data-scientist-tuhin>
- [29] <https://medium.com/indeed-engineering/transitioning-from-academia-to-industry-perspectives-from-indeeds-data-scientists-de890acd1bfc>;
- [30] https://www.glassdoor.com/Salaries/data-scientist-salary-SRCH_KO0,14.htm
- [31] <https://medium.com/@sifium/machine-learning-types-of-classification-9497bd4f2e14>