Analysis Of Customer Activity on E-Commerce Platform

Sai Venkata Anudeep Sanchula
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
asgzq@umkc.edu

Sreekar Reddy Kasireddy
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
skmvk@umkc.edu

Riddhima Singh
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
rscdd@umkc.edu

Muhammad Zeshan
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
mzqbq@umkc.edu

Sai Venkata Giri Prasad Kanagala
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
sk55r@umkc.edu

Gokul Krishna Kanulla
School of Computing and Engineering
University of Missouri-Kansas City
Kansas City, Missouri
gk3hy@umkc.edu

***Abstract*—Big data analytics now play a more significant part in e-commerce than ever. Fifteen papers were reviewed to evaluate the impact of ample data research on e-commerce. The discussion section of this article analyzes BDA in the context of e-commerce using a thorough literature study. To better grasp cross-cutting analytical tools in e-commerce, the study's conclusions consider a variety of BDA principles (such as significant data categorization, categories, design, market relevance, and associated theories). One of the most challenging parts of e-commerce is the large amount of data that must be handled and evaluated to reap the information revolution's benefits—e-mailing and e-mailing. Based on the importance of analyzing big data and its benefits to e-commerce operations, this study investigates the benefits of employing BDA in e-commerce for vendors and customers. To summarize, while BDA improves customers' and sellers' electronic purchasing experiences, data production remains a challenge.**

I. Introduction

Big data refers to large amounts of data generated by a variety of sources, including social media platforms, internet logs, sensors, Internet of Things (IoT) devices, and many more. It might be unstructured, semi-structured (similar to XML files), or structured (similar to DBMS tables) (like audio, videos, and images). Traditional database management solutions cannot handle the volume of data. Businesses may develop meaningful data thanks to Big Data. Businesses utilize big data to improve operations, provide better customer service, create personalized marketing campaigns, and perform other jobs that will eventually increase sales and profits. Businesses that utilize it efficiently may have a competitive advantage over those that do not because they can respond more swiftly and with more information. Big data provides organizations with relevant consumer information that they can use to fine-tune their marketing, advertising, and promotions in order to increase customer engagement and conversion rates. Businesses may become more sensitive to customer requests and desires by examining historical and real-time data to evaluate changing consumer or corporate buying preferences.

An e-commerce platform is a content management system (CMS) and commerce engine for a website used to handle purchases, catalog items, and maintain user interactions with online shops. Your firm may prosper regardless of size, whether B2B or B2C, selling tangible things or providing remote services. Customers today want to engage and shop online more than ever before, so businesses must meet them where they are to increase sales. Many retailers depend on big data analytics to capture a commercial effect. Large-scale data eCommerce that uses Big Data and Analytics to improve consumer engagement, adapt the purchasing experience and increase revenue. Big data eCommerce engages customers in two ways: they receive tailored suggestions and have more access to product details. Big data tells you who your consumers are, what they want, and when they want it. Big Data eCommerce will also assist you in determining how pleased your consumers are and when and where they depart dissatisfied. Customer service teams can apply big data analytics to spot typical product faults, determine the source of a customer complaint and swiftly fix it, or identify the clients who are most likely to quit and design a strategy to keep them happy. Big data eCommerce may collect client feedback via chat, surveys, phone calls, and other real-time contact techniques. We may utilize Bigdata to learn about different elements, such as which items sell well, enhancing the business's income.

II. Methodology and approach

A. Data set Details

The E-Commerce Behavior for Multi-Category dataset is the one we've picked, and it can be found at https://www.kaggle.com/datasets/mkechinov/ecommerce-behavior-data-from-multi-category-store.

There are 412 million rows and nine columns. It offers extensive data on the that gives the analyst a clear insight of how sales are progressing during that specific time period.

Dataset shape: 411,709,736 × 9

Dataset size: 52.6GB

Dataset timestamp: Oct 2019 – Apr 2020

To summarize the dataset under consideration, the columns

contained in the dataset are:

Brand: Because the dataset we evaluated is connected to E-commerce websites, we have a wide range of items available on this platform, such as accessories, shoes, garments, electronics, and many more. A brand is a publicly distinguishable product, service, or concept that is used to ease communication and marketing. Nike, Samsung, Apple, and Puma are some well-known brands.

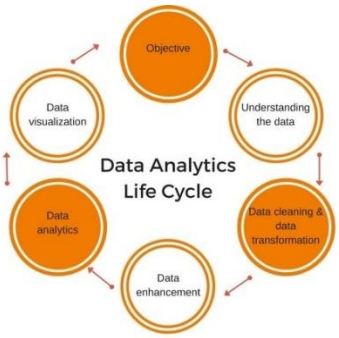
Price: The amount of money expected to be paid in exchange for purchasing a thing or product.

Event type and Event Time: This reflects the time at which the consumer accesses the e-commerce platform, and event type represents whether the client sees the product, adds it to the basket, or purchases it. This dataset's Event Type allows us to investigate consumer behavior.

Category ID: The category Id is used to anticipate which product belongs to which category. For example, if we examine electronics as a category, we can find several sorts of items in this category, such as smartphones, notebook computers, televisions, tablets, and so on. Other fields, such as **User ID** and User Session, let us determine when the user is viewing the website. Other columns include category code and **product ID.**

B. Challenges in Big Data Analysis

Because of the advantage of big data, it opens up new opportunities for ambitious academics to engage in knowledge-processing jobs. The solution rests in comprehending the many computing obstacles, information security, and ways for analyzing massive volumes of data. For instance, many statistical procedures that perform well with little data do not scale to significant volumes of data. Similar challenges develop when analyzing enormous volumes of data using various computer algorithms that excel at handling tiny data. Before formatting your paper, compose it and save it as a separate text file. Before formatting, do all content and organizational modifications.



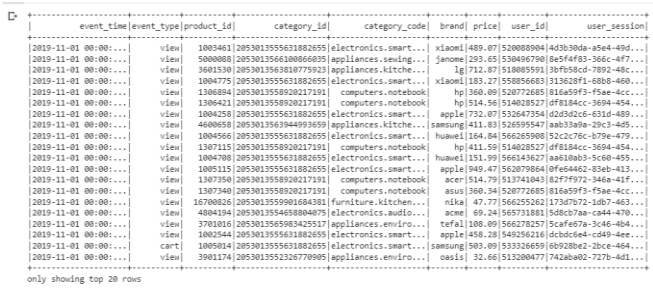
C. Tools Used In This Project:

The Tools that we have considered to use in this project are Python, Pandas, Matplotlib, Seaborn, Pyspark, SQL and Tableau.

Initial Process As the initial step for our project, the dataset that we have considered is for seven months. As the dataset is huge, we have considered to carry out the analysis on one month data i.e Oct 2019. The October month data is of 3.54 GB.

a. Cleaning Of the data:

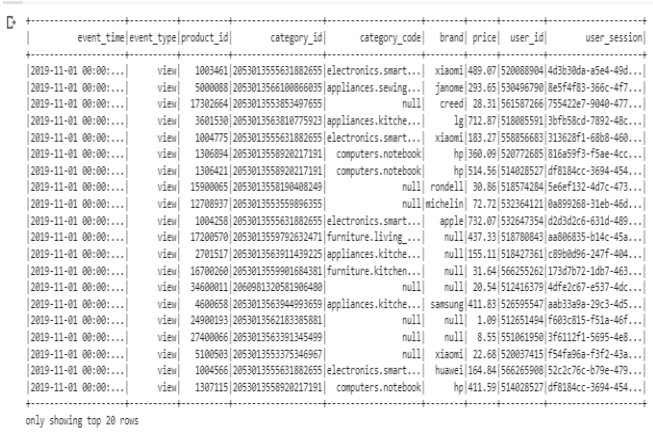
In the dataset we considered has certain null values that are present in the dataset. So we have run some queries using Python and SQL to remove the null values that are present in the dataset.



event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session
2019-11-01 00:00:...	view	1003461	2053013555631882655	electronics.smart...	xiaomi	489.07	520088904	4d3b30da-a5e4-49d...
2019-11-01 00:00:...	view	5000088	2053013566100866035	appliances.sewing...	janome	293.65	530496790	8e5f4f83-366c-4f7...
2019-11-01 00:00:...	view	3601530	2053013563810775923	appliances.kitche...	lg	712.87	518085591	3bfb58cd-7892-48c...
2019-11-01 00:00:...	view	1004775	2053013555631882655	electronics.smart...	xiaomi	183.27	558856683	313628f1-68b8-460...
2019-11-01 00:00:...	view	1306894	2053013558920217191	computers.notebook	hp	360.09	520772685	816a59f3-f5ae-4cc...
2019-11-01 00:00:...	view	1306421	2053013558920217191	computers.notebook	hp	514.56	514028527	df8184cc-3694-454...
2019-11-01 00:00:...	view	1004258	2053013555631882655	electronics.smart...	apple	732.07	532647354	d2d3d2c6-631d-489...
2019-11-01 00:00:...	view	4600658	2053013563944993659	appliances.kitche...	samsung	411.83	526595547	aab33a9a-29c3-4d5...
2019-11-01 00:00:...	view	1004566	2053013555631882655	electronics.smart...	huawei	164.84	566265908	52c2c76c-b79e-479...
2019-11-01 00:00:...	view	1307115	2053013558920217191	computers.notebook	hp	411.59	514028527	df8184cc-3694-454...
2019-11-01 00:00:...	view	1004708	2053013555631882655	electronics.smart...	huawei	151.99	566143627	aa610ab3-5c60-455...
2019-11-01 00:00:...	view	1005115	2053013555631882655	electronics.smart...	apple	949.47	562079864	0fe64462-83eb-413...
2019-11-01 00:00:...	view	1307350	2053013558920217191	computers.notebook	acer	514.79	513741043	82f7f972-346a-41f...
2019-11-01 00:00:...	view	1307340	2053013558920217191	computers.notebook	asus	360.34	520772685	816a59f3-f5ae-4cc...
2019-11-01 00:00:...	view	16700826	2053013559901684381	furniture.kitchen...	nika	47.77	566255262	173d7b72-1db7-463...
2019-11-01 00:00:...	view	4804194	2053013554658804075	electronics.audio...	acme	69.24	565731881	5d8cb7aa-ca44-470...
2019-11-01 00:00:...	view	3701016	2053013565983425517	appliances.enviro...	tefal	108.09	566278257	5cafe67a-3c46-4b4...
2019-11-01 00:00:...	view	1002544	2053013555631882655	electronics.smart...	apple	458.28	549256216	dcbdc6e4-cd49-4ee...
2019-11-01 00:00:...	cart	1005014	2053013555631882655	electronics.smart...	samsung	503.09	533326659	6b928be2-2bce-464...
2019-11-01 00:00:...	view	3901174	2053013552326770905	appliances.enviro...	oasis	32.66	513200477	742aba02-727b-4d1...

only showing top 20 rows

The above image shows data with Null Values



event_time	event_type	product_id	category_id	category_code	brand	price	user_id	user_session
2019-11-01 00:00:...	view	1003461	2053013555631882655	electronics.smart...	xiaomi	489.07	520088904	4d3b30da-a5e4-49d...
2019-11-01 00:00:...	view	5000088	2053013566100866035	appliances.sewing...	janome	293.65	530496790	8e5f4f83-366c-4f7...
2019-11-01 00:00:...	view	17302664	2053013553853497655	null	creed	28.31	561587266	755422e7-9040-477...
2019-11-01 00:00:...	view	3601530	2053013563810775923	appliances.kitche...	lg	712.87	518085591	3bfb58cd-7892-48c...
2019-11-01 00:00:...	view	1004775	2053013555631882655	electronics.smart...	xiaomi	183.27	558856683	313628f1-68b8-460...
2019-11-01 00:00:...	view	1306894	2053013558920217191	computers.notebook	hp	360.09	520772685	816a59f3-f5ae-4cc...
2019-11-01 00:00:...	view	1306421	2053013558920217191	computers.notebook	hp	514.56	514028527	df8184cc-3694-454...
2019-11-01 00:00:...	view	15900065	2053013558190408249	null	rondell	30.86	518574284	5e6ef132-4d7c-473...
2019-11-01 00:00:...	view	12708937	2053013553559896355	null	michelin	72.72	532364121	0a899268-31eb-46d...
2019-11-01 00:00:...	view	1004258	2053013555631882655	electronics.smart...	apple	732.07	532647354	d2d3d2c6-631d-489...
2019-11-01 00:00:...	view	17200570	2053013559792632471	furniture.living_...	null	437.33	518780843	aa806835-b14c-45a...
2019-11-01 00:00:...	view	2701517	2053013563911439225	appliances.kitche...	null	155.11	518427361	c89b0d96-247f-404...
2019-11-01 00:00:...	view	16700260	2053013559901684381	furniture.kitchen...	null	31.64	566255262	173d7b72-1db7-463...
2019-11-01 00:00:...	view	34600011	2060981320581906480	null	null	20.54	512416379	4dfe2c67-e537-4dc...
2019-11-01 00:00:...	view	4600658	2053013563944993659	appliances.kitche...	samsung	411.83	526595547	aab33a9a-29c3-4d5...
2019-11-01 00:00:...	view	24900193	2053013562183385881	null	null	1.09	512651494	f603c815-f51a-46f...
2019-11-01 00:00:...	view	27400066	2053013563391345499	null	null	8.55	551061950	3f6112f1-5695-4e8...
2019-11-01 00:00:...	view	5100503	2053013553375346967	null	xiaomi	22.68	520037415	f54fa96a-f3f2-43a...
2019-11-01 00:00:...	view	1004566	2053013555631882655	electronics.smart...	huawei	164.84	566265908	52c2c76c-b79e-479...
2019-11-01 00:00:...	view	1307115	2053013558920217191	computers.notebook	hp	411.59	514028527	df8184cc-3694-454...

only showing top 20 rows

The above data represents the data that is without the null values.

b. Analysis:

After cleaning the data, we move to the data analysis. What does data analysis mean? The cleansed data is analyzed, converted, and modeled in Data Analysis to disclose important information, inform conclusions, and improve decision-making. We attempted to run specific spark queries to gather dataset statistics such as user visits, number of brands sold, etc.

The below figures shows the daily visits of customers and number of customers visited and some statistics of the dataset.

```
Daily Visitor Statistics
--------------------------------------------------
count        31.00
mean    208,829.77
std      16,233.61
min     170,668.00
25%     195,260.50
50%     208,477.00
75%     223,321.00
max     231,849.00
Name: Number_of_daily_visitors, dtype: float64
--------------------------------------------------
Visitor Statistics by Dates
--------------------------------------------------
            count       mean       std        min        25%        50%  \
event_time
Friday       4.00 223,166.00  9,715.30 209,410.00 219,796.00 226,360.50
Monday       4.00 204,218.00 16,732.50 186,951.00 193,410.00 202,326.00
Saturday     4.00 205,851.25  9,731.79 194,958.00 200,368.50 205,263.50
Sunday       4.00 213,090.00 18,707.44 193,210.00 199,321.75 213,650.50
Thursday     5.00 197,994.80 15,736.89 170,668.00 198,571.00 205,321.00
Tuesday      5.00 211,615.80 18,526.04 190,188.00 193,168.00 221,204.00
Wednesday    5.00 208,073.80 18,982.96 184,965.00 191,965.00 214,140.00

                 75%        max
event_time
Friday     229,730.50 230,533.00
Monday     213,134.00 225,269.00
Saturday   210,746.25 217,920.00
Sunday     227,418.75 231,849.00
Thursday   206,937.00 208,477.00
Tuesday    223,384.00 230,135.00
Wednesday  219,100.00 230,199.00
```

```
Daily Visits Statistics
--------------------------------------------------
count         31.00
mean     299,510.55
std       25,531.55
min      241,086.00
25%      278,957.50
50%      301,219.00
75%      319,668.50
max      339,943.00
Name: Number_of_daily_visits, dtype: float64
--------------------------------------------------
Visit Statistics by Dates
--------------------------------------------------
             count       mean       std        min        25%        50%  \
event_time
Friday        4.00 321,772.50 16,017.26 300,651.00 313,704.00 325,731.00
Monday        4.00 293,353.00 26,033.15 263,437.00 276,610.00 294,346.50
Saturday      4.00 298,041.75 19,140.22 276,914.00 286,412.75 296,976.50
Sunday        4.00 309,080.00 32,756.35 273,034.00 285,634.75 311,671.50
Thursday      5.00 282,820.60 24,557.36 241,086.00 281,446.00 293,587.00
Tuesday       5.00 301,032.40 27,672.77 268,737.00 276,723.00 308,348.00
Wednesday     5.00 295,314.60 27,194.96 261,252.00 274,906.00 301,219.00

                  75%        max
event_time
Friday     333,799.50 334,977.00
Monday     311,089.50 321,282.00
Saturday   308,605.50 321,300.00
Sunday     335,116.75 339,943.00
Thursday   295,142.00 302,842.00
Tuesday    316,706.00 334,648.00
Wednesday  310,453.00 328,743.00
```

```
There are 347,118 customers, who purchased in October.
--------------------------------------------------
Distribution of Customer by Number of Purchases
1    0.62
2    0.19
3    0.07
4    0.04
5    0.02
6    0.01
7    0.01
8    0.01
9    0.00
10   0.00
Name: number_of_purchases, dtype: float64
--------------------------------------------------
There are 131,427 repeat customers, who bought more than once.
```

```
Total number of activity:42,448,764
--------------------------------------------------
Total number of visits:9,244,421
--------------------------------------------------
Total number of visitors:3,022,290
--------------------------------------------------
The number of categories:13
--------------------------------------------------
The number of subcategories:57
--------------------------------------------------
The number of brands:3,444
--------------------------------------------------
The number of products:166,794
```

c. Visualization:

The graphical display of various types of information or data using visual components such as charts, graphs, or maps is known as data visualization. Data visualization tools allow data trends, anomalies, and patterns to be and intuitively observed and understood. We attempted to display specific features of the data to learn about the best-selling goods in various categories and to visualize the user actions daily. Tableau and Power BI were the tools we utilized for this project.

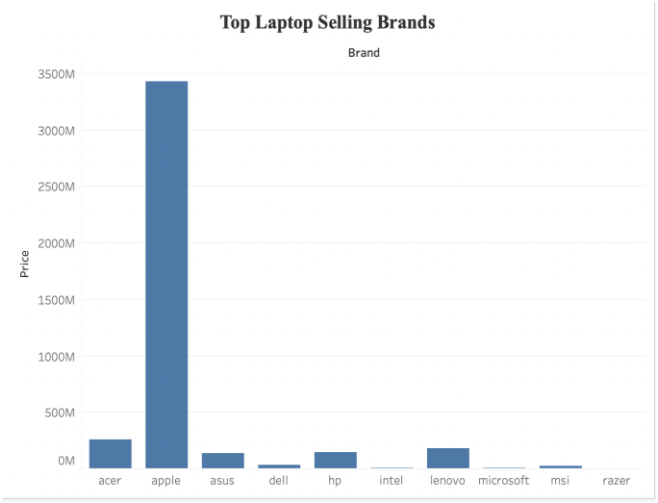


Fig 1: Top Selling Laptop Brands

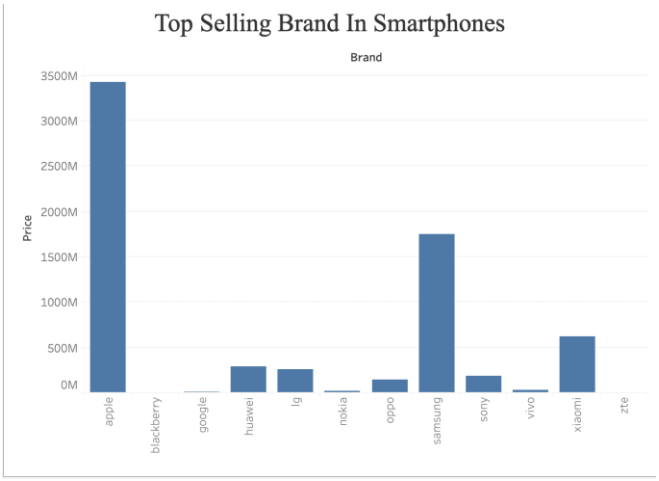


Fig 2: Top Selling Brands in Smartphones

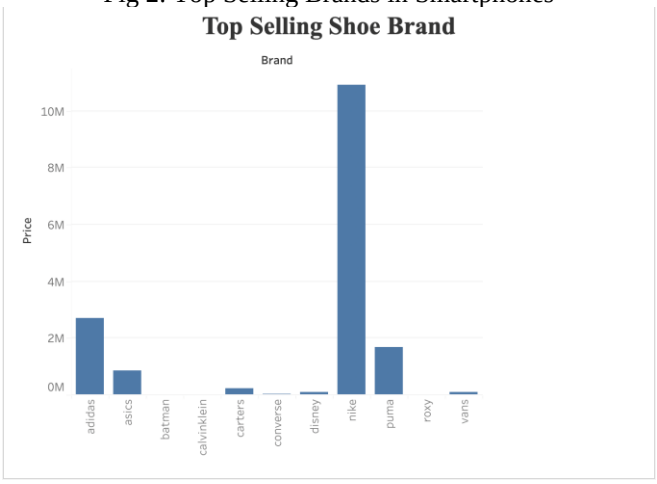
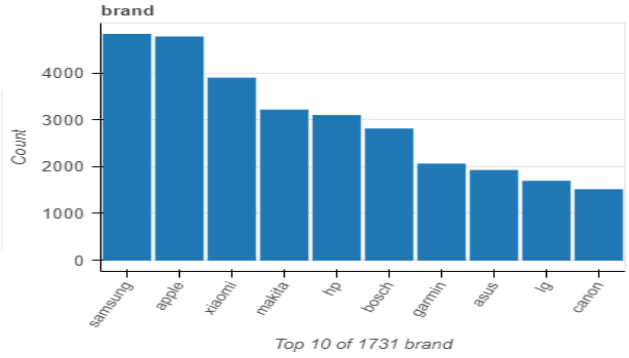
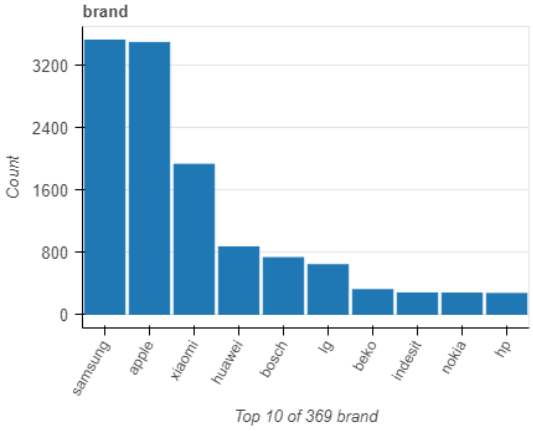
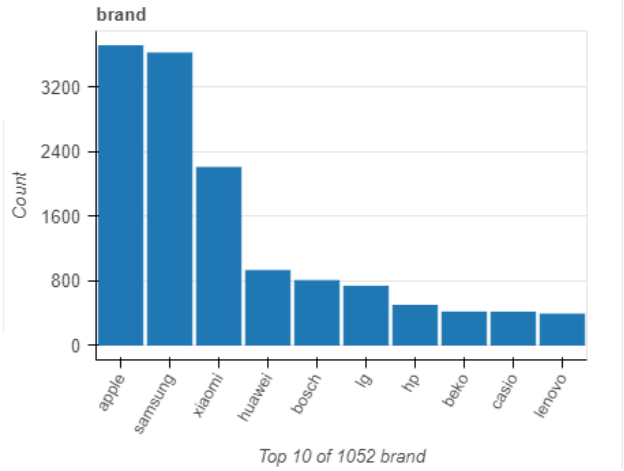


Fig 3: Top selling Shoe Brands

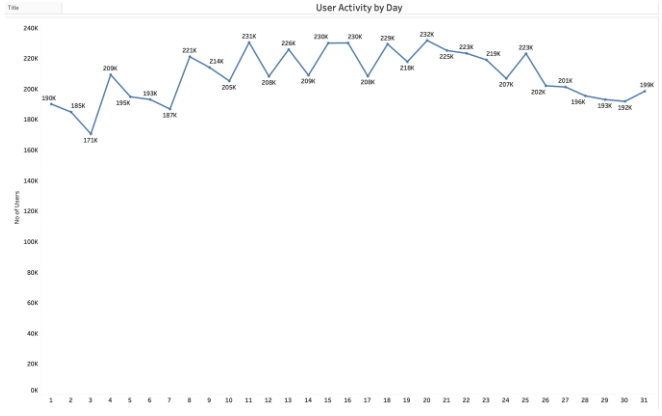
The visualizations shown above are connected to brand vs. price. We attempted to compare different shoe brands and brands linked to gadgets such as computers and cell phones.
We also attempted to visualize the top goods that have recently been seen, added to the basket, and purchased.



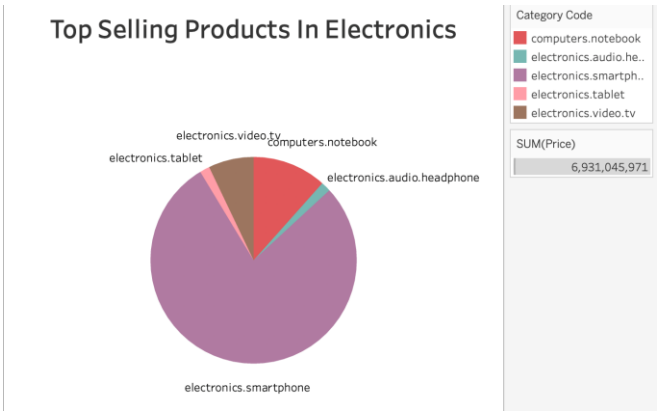




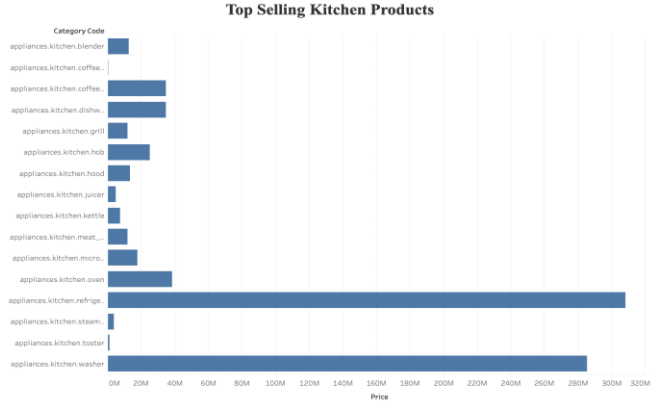
We attempted to depict user activity in a day for October. We tried to show how many people use the e-commerce website daily and which months have the most user activity.



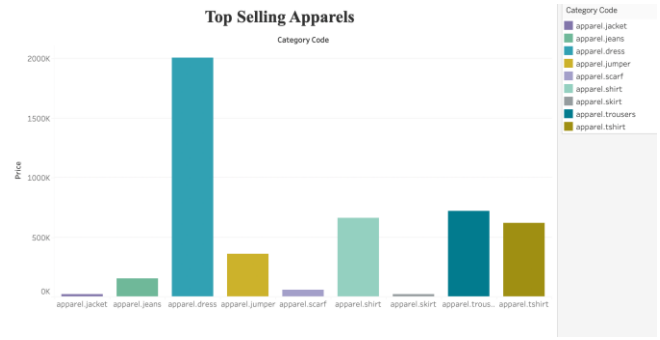
The pie chart in the graphic below depicts the top-selling electrical items in various hues. As you can see, the smartphone is the top selling device as opposed to TV, tablets, etc.



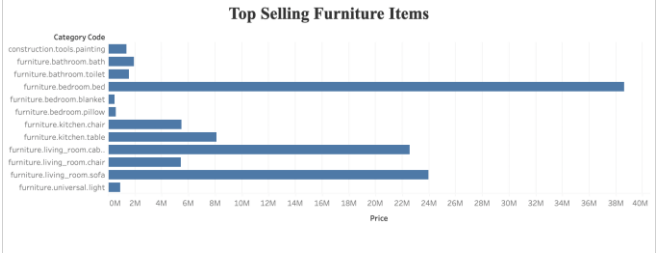
Similarly, the chart below depicts the top-selling kitchen equipment such as ovens, juicers, dishwashers, etc. As you can see, refrigerators and washers are the most popular appliances.



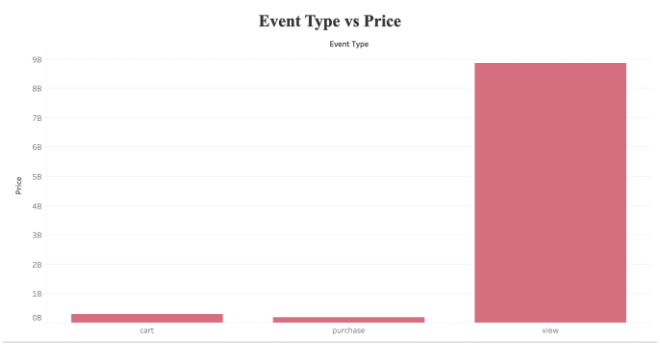
As seen in the bar chart, clothing dresses have the most significant client demand compared to t-shirts, skirts, pants, jeans, and so on.



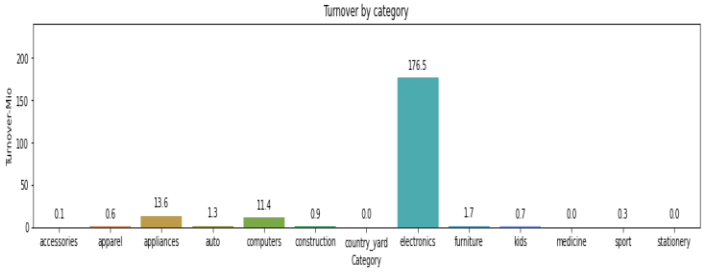
The image below is about Furniture, and the category code displays the name of the objects with the code. As you can see, most buyers are interested in bedroom beds rather than sofas, tables, chairs, blankets, and so on.



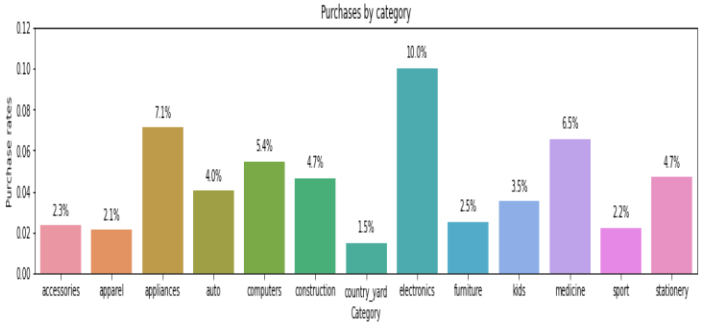
Below bar chart shows the event type of customer behavior how many people are interested to purchase and how many just come for window shopping only. As you can see statistics tell mostly people only view the product and few of them cart as well as the least number of people interested to buy the products.



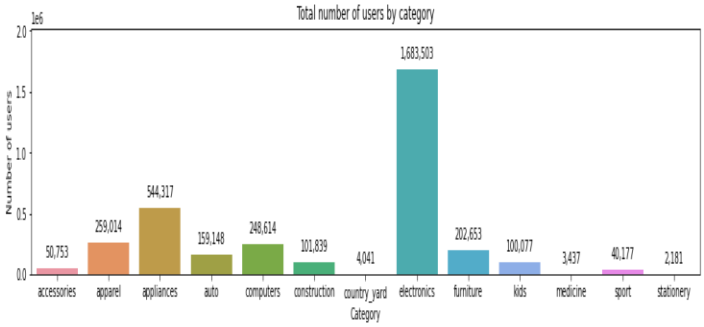
This figure shows the turnover by items category, In which electronics items gave a huge turnover as compared to other products like medicine, furniture, and stationery. Etc.



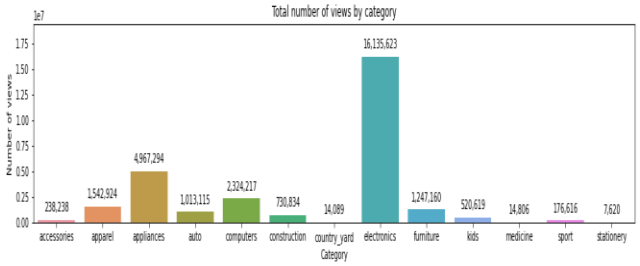
Below figure represents the conversion rates by categories



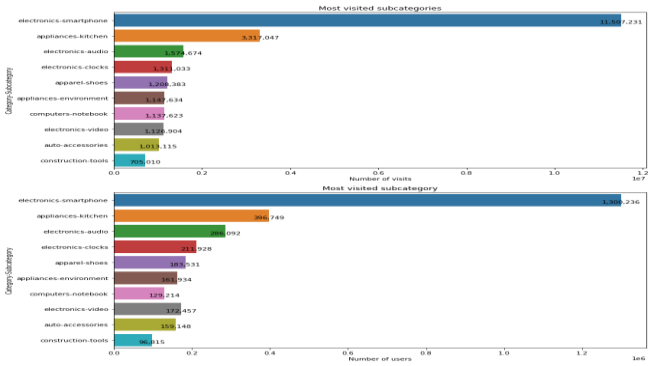
Below plot illustrates number of visitors in each category during that period



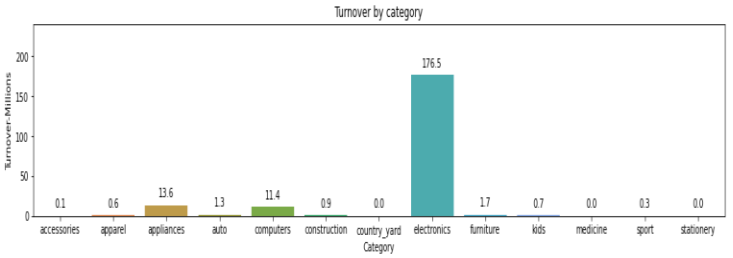
Below figure illustrates the number of visits in each category during the period



Below plot shows the most popular subcategories and number of visits and visitors during the period



The below figure shows the category turnover for that period



CONCLUSION

We may analyze consumer behavior by utilizing big data techniques in the sales dataset to see if individuals prefer online platforms just to compare brand rates and offers to those offered in-store. The utilization of big data also assists organizations in determining which projects are profitable and which are not. From a business standpoint, we can learn which products are contributing to an increase in revenue and GDP.

REFERENCES

[1] https://www.webofproceedings.org/proceedings_series/ESS P/IWASS%202019/SS06029.pdf

[2] https://www.researchgate.net/publication/349741121_Implementing_Big_Data_Analytics_in_ECommerce_Vendor_and_Customer_View

[3] https://zenodo.org/record/1491581/files/Contribution_11339_fullpaper.pdf?download=1

[4] https://iaeme.com/MasterAdmin/Journal_uploads/IJARET/ VOLUME_11_ISSUE_12/IJARET_11_12_166.pdf

[5] https://labfiles.linuxacademy.com/aws/aws-procourse/whitepapers/big-data-analytics-options-on-aws.pdf

[6] https://piermick.files.wordpress.com/2018/11/awsgsgemr.pdf

[7] https://www.researchgate.net/publication/347506371_Big_ Data_Analytics_Adoption_in_Cloud_Amazon_EMR

[8] https://cloud.netapp.com/blog/aws-cvo-blg-aws-big-data-6optionsyoushouldconsiderhttps://www.alluxio.io/learn/emr/aws/