HACK CHALLENGE 2021

CODE FOR A BETTER FUTURE

ANALYTICS TOOL FOR ECOMMERCE BUSINESSES

A PROJECT REPORT

Submitted by

KARTHIGA G

KEERTHANA S

PRIYADARSHINI M

PRIYADHARSHINI S

Of

3rd YEAR B.Tech

IN

Information Technology

From

SRI SAIRAM ENGINEERING COLLEGE

Sai Leo Nagar, West Tambaram, Tamil Nadu

Chennai-600044

Under the Guidance Of

DR. T.P RANI

1.INTRODUCTION

1.1 Overview

This fast-moving world had taught many of us what is life in the last two years. This situation demands new norms and crises. One such new norm of social distancing and crowd less shopping bloomed the e-commerce websites with much more profits, increased online retail sales share of total retail sales from 16% to 19% in 2020, according to estimates in a UNCTAD report published on 3 May. This digital transformation accelerated profits in a trillion for various big giants like Amazon, Walmart, Alibaba, eBay and more, ultimately increasing their market value also led to an increase in fierce competition in today's online marketplace. The e-commerce industry sells a diverse product line of grocery items and merchandise products, such as food, pharmaceuticals, apparel, games and toys, hobby items, furniture and appliances. The analysis of such an industry is of great importance as it gives insights into the sales and profits of various

products.

1.2 Purpose

Data analytics is the process of extracting and processing metrics. As part of it, we analyze data to find trends and answer various questions. It is particularly beneficial in the realm of online consumer behavior. The information helps eCommerce companies stay competitive in their niche markets. With these significant insights, businesses can identify bottlenecks in their selling processes, which provides an opportunity to refine strategies. The following factors drive us to give a detailed analysis on ecommerce businesses.

1. When shopping models emerge, we can consolidate better business strategies. Data analytics reveals how shoppers interact with the website, what their preferences are and their preferred brands.
2. Because data analytics intricates behind the scenes, we can know when spikes in demand take place. So, we can plan better for future sales.
3. Data analytics can lessen costs**.** Because the metrics give a greater awareness of what is profitable and what is not, the project budget concentrates on products that will gain higher profits.
4. Novel access open when data is analysed consumer's needs and wishes run behind hidden numbers, so we have tangible proof of what they want.
5. When we equip with the facts, transformation occurs in new product launches and building a brand around market demands, opening gates for new investors and wholesalers, moreover uniting all craftmanship across India under one roof as an online shopping application.
6. Measured data also reveals an apparent supply and demand formula so we can price items right.
7. **Firm purchasers** are often the result of a well-thought-out data analytics strategy.

These factors became the gravity holding our ideas on this topic. Also, to carry on the research that helps the e-commerce business to rise and shine.

2. LITERATURE SURVEY

2.1 Existing Problem

1. Analysis and optimization of online sales of products

Author: Zainab Pirani, Anuja Marewar, Zainab Bhavnagarwala, Madhuri Kamble

Published in: 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)

Date of publication: 01 February 2018
DOI: 10.1109/ICIIECS.2017.8276165
Publisher: IEEE

Abstract: This system aim is of developing a “Sales Analytics Tool” our paper describes the need
of a system to analyses the database transactions of e-commerce websites using various data mining techniques and algorithms such as affinity analysis, logistic regression and linear regression. The proposed idea through this research paper is to develop a system which takes input the database transactions of sold products, segments the data obtained, analyzes the graphs and extracts the market trends and product sales patterns. After this, the system optimizes this data on the basis of market requirements thereby improving sales and merchandise planning in a way that to increase the overall productivity and profits of the organization.

Disadvantage:

1. It involves more complicated techniques and algorithms which will be quite hefty.

2.2 Proposed Solution

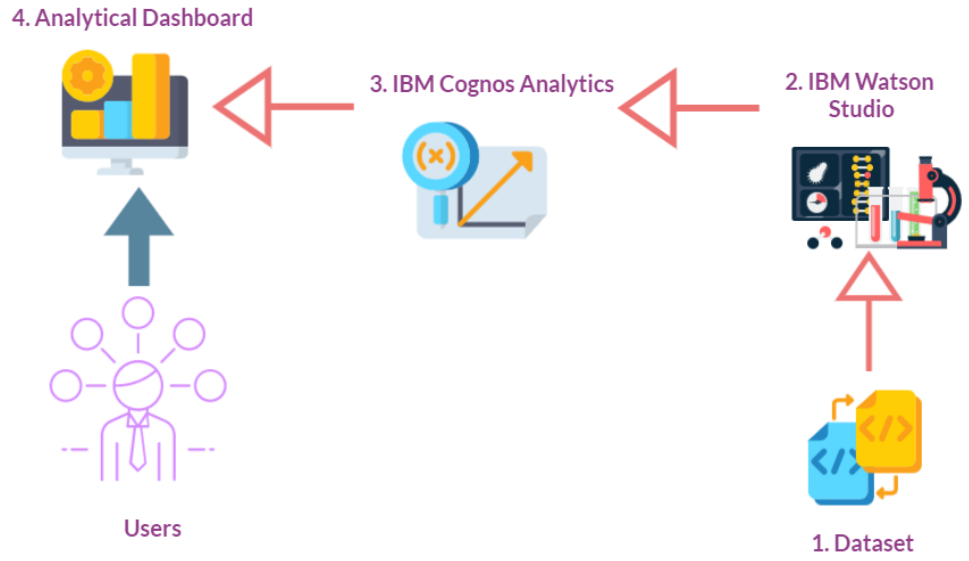
The object of this challenge is to develop an Analytical Dashboard where the owner/user will understand the growth/potential of the business in the market. So, he can find out statistics such as -

1. Region that accounts for greater number of orders.
2. Frequency distribution of quantity ordered.
3. Percentage sales by different product categories.
4. Profitable products or their sub products in last few years.
5. Products that incurred losses.
6. Product type that was ordered greater times.
7. Yearly sales for various states.
8. Forecasting future sales according to shipping date.
9. Trend in profit/sales over time (years/months/quarters).

With this analysis, the e-commerce business can identify various aspects of the shopping pattern and take measures if required.

3. THEORITICAL ANALYSIS

I. BLOCK DIAGRAM:



II. DATASET:

a. US Superstore
b. ComparisonAmzvsFli
c. Mobile Comparison AmzVsFlip
d. Food Comp Swiggy vs Zomato

4. EXPERIMENTAL INVESTIGATIONS

1. **Region that accounts for greater number of orders.**

   Considering the US Superstore dataset, we created a visualization that gives insight into the total number of orders under a particular region. The visualization goes like this: having regions as Central, East, South, West counting on the distinct user who placed their order, the west region tops the chart with 1611 order count.

2. **Frequency distribution of quantity ordered.**

   A frequency distribution is a representation available in a graphical or tabular format that displays the number of observations within a given interval. The goal of this analysis is to find the frequency distribution of the quantity over the state. The result reaches a higher value in California, New York, Texas ranking them the top three with higher frequency distribution.

3. **Percentage sales by different product categories.**

   The percentage of sales for different product categories: Technology, Office Supplies, Furniture values looked up for calculation. Segmenting the pie chart visualization with the value of Category along with the Sales as its size, we can complete the required

observation. The value obtained is as follows:

Technology: 36.4%

Furniture:32.3%

Office Supplies: 31.3%

4.Profitable products or their sub products in last few years.

Profit and loss are part of business similarly, in the US superstore the profitable product-subcategory for the years: 2014, 2015, 2016, 2017 are calculated. And it is found that the value of profit is unusually high in 2017 for the category Labels. Also, high in case of papers in 2015, 2014 for the Labels, and 2016 for the envelopes. The value of Profit Percent is most unusual when the Sub-Category is Labels, Paper, Envelopes and Chairs.

5.Product that incurred losses:

In this solution, we used loss percent as a size with multiple products as a segment to differentiate all products and their losses in percentage using pie chart, by calculating this we conclude that “office star “product has major loss [percentage (1%)].

6.Product type that was ordered greater times

US superstore with various product categories offers its customer wide variety of products, in such pool of category few gained more attractions from the customers, so we were eager to calculate the product that count for greater order, to make this possible we used no of ordered counts (count distinct) as “x” axis and product type as “y” axis. The total number of results for **Order ID** for all values of **Sub-Category** is 9994.

After calculation, here we conclude that Binders (15.2 %) and Paper (13.7 %) are the most frequently occurring Product type with a combined count of 2893 items (28.9 % of the total), whereas these two was ordered greater times compared to others.

7. Yearly sales for various states.:

Sales over a state accounts for the people’s necessity in that particular region, so we planned to use sales and states data as “x” and “y” axis to compare accordingly to the given four years [2014,2015,2016,2017]by comparing them we conclude that For Sales, 2017 is the most important category of Year with a total value of 412,980 (33 % of the total).

8. Forecasting future sales according to shipping date.

As part of a business model, forecasting future sales is essential to top on business trends in the market, calculating future deals with the historical value (2014,2015,2016,2017) value of sales for 2018.

Considering average sales for each month in a year, the variation between each month proves linear. With this observation, we processed the data for calculating future purchases with the trend forecasting method. The trend forecasting method works on the linear regression technique of time series forecasting. Trend forecasting gives the best forecasting reliability when the driving factors of your business linearly affect your measures.

9. Trend in profit/sales over time (years/months/quarters).

Here we used no of sales as a column and year as a line in both “x” and “y” axis to calculate the trend in sales over year [2014,2015,2016,2017], by this we conclude that 2017th year has huge sales (733215.255) and profit (93439.27).

We didn't stop there, our exploration of IBM Cognos analytics craved more of our interest and started digging more and found interesting analysis on trending topics. The most popular e-commerce giants in India include Amazon, Flipkart, Myntra, and so on. Counting on these sellers the, most known is possibly amazon and Flipkart. Their humongous offers include Big billion days, Great Freedom Sales, Amazon Super Value, Book Bazar, Flipkart super sales, and more on the rover. As a user, we often have confusion choosing which product to order on which site.

We often compare their discounts and rely upon the reviews and ratings from other customers. To put it together: we need cost-effective quality products, this: lands up confusion about choosing between e-commerce websites. We are in a similar situation, where choosing Mobile phones and books at minimum cost with the best quality without any insights seems challenging. So we collected data regarding books, book author, price, review and ratings for Flipkart and Amazon as the dataset.

The result concedes as:

- Flipkart_Ratings slightly drives Flipkart__Books_Price by 20%.
- The average of Flipkart_Price for all values of Flipkart_Ratings and Book_Title is 248.6.
- The average values of Flipkart_Books_Price over all combinations of the inputs range from a minimum of 107 to a maximum of 563.
- The 5 AM Club: Own Your Morning- Elevate Your Life is the most frequently occurring book of Flipkart_Book_Title with a count of 9 items (14.5 % of the total).
- Amazon_Rating strongly drives Amazon_Books_Price by 81%.
- The value of Amazon_Price is unusually high when Amazon_Rating is 4.8.
- The 5 AM Club: Own Your Morning, Elevate Your Life is the most frequently occurring Book of Amazon_BookTitle with a count of 9 items (14.5 % of the total).

This analysis provided insight: books on both these sites have an affordable price and exceptional quality, most of the books outstrip their reviews. Moreover, the user reviews and

ratings drive the book purchases. Also, books from famous authors occur in a frequent order.

We continued the research for Various Mobile Brands in Amazon and Flipkart and suggesting which one has higher mobile brands with trending specifications. To bring this comparison, we used rating and Mobile brands for visualizations and, we obtained the result as:

- The sum of Stars for all values of Mobile Name is 135.2.
- The summed values of Ratings range from a minimum of 4.1 (when Mobile Name is GIONEE Max (Black, 32 GB)) to a maximum of 32 (when Mobile Name is Snexian Rock) from the Flipkart dataset.

And for Amazon:

- The average values of Rating range from a minimum of 2.571 (when Product Name is Blackberry Pearl Flip 8220 GSM Cell Phone) to a maximum of 4.273 (when Product Name is BLU Jenny II T177 Unlocked GSM Dual-SIM Cell Phone - Black/Red).
- The value of Rating is most unusual when the values of Product Name are BLU Studio 5.0 C HD Unlocked Cell phone, White, Apple iPhone 4 32GB (Black) - Verizon, BLU Studio C 5.0-Inch Android Smartphone with Lollipop OS - Unlocked (Green), BLU Studio C 5.0-Inch Android Smartphone with Lollipop OS - Unlocked (Orange) and Apple iPhone 6 128GB Factory Unlocked GSM 4G LTE Smartphone, Space Gray (Certified Refurbished).

From this analysis, we obtained the result as Amazon offers its users a 332 variety of mobile brands, the product specifications and features attract its customer. Therefore, pushing amazon mobile category sales higher than that of Flipkart's.

Besides Groceries, Fashion, Mobiles, Electronics, Furniture and Home Accessories are available online, even food adds on the list. Swiggy, Zomato are the tech-giants who makes our lifestyle different, at times of covid they helped us stay home and eat desired food. So our study extends to comparing various cuisine available in Zomato and Swiggy, among which edibles are at an economical price. For this, we processed nearly 13k of data and obtained surprising results.

The average price for all values of cuisine is 336.

- Cuisine slightly drives price by 16% in ZOMATO.
- In ZOMATO, the price is unusual when the cuisines are Asian, Beverages, Cafe, Arabian and Chinese.
- According to Zomato's insight, Biryani is the most frequently occurring food of every cuisine, with 299 items (27.7 % total).
- Most of the customers liked to order North Indian Food items from Swiggy's stores.
- The summed values of Cost range from a minimum of 150 (when Cuisine is Ice Cream) to a maximum of 4102 (when Cuisine is North Indian).

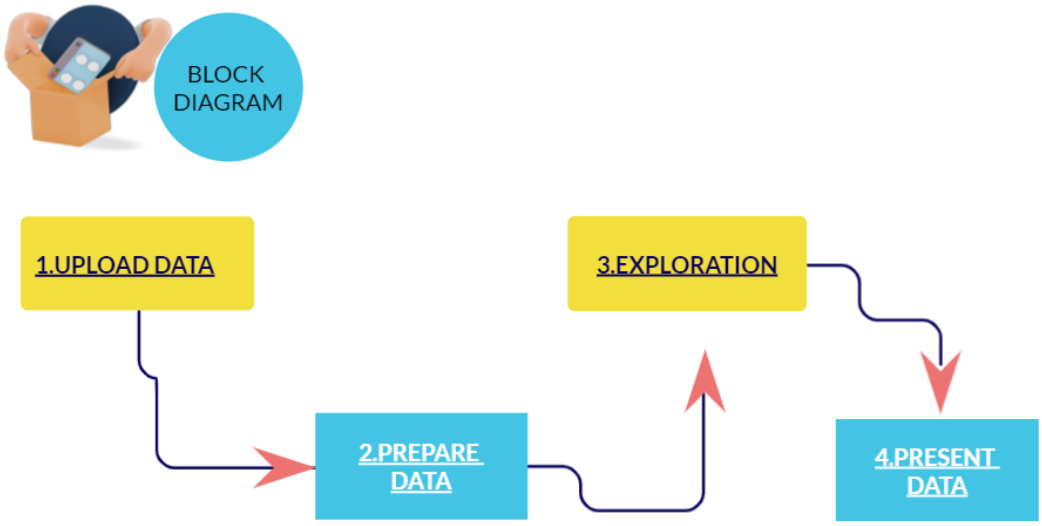
We concluded that: people's choices over food are unpredictable than their clothing choice, as

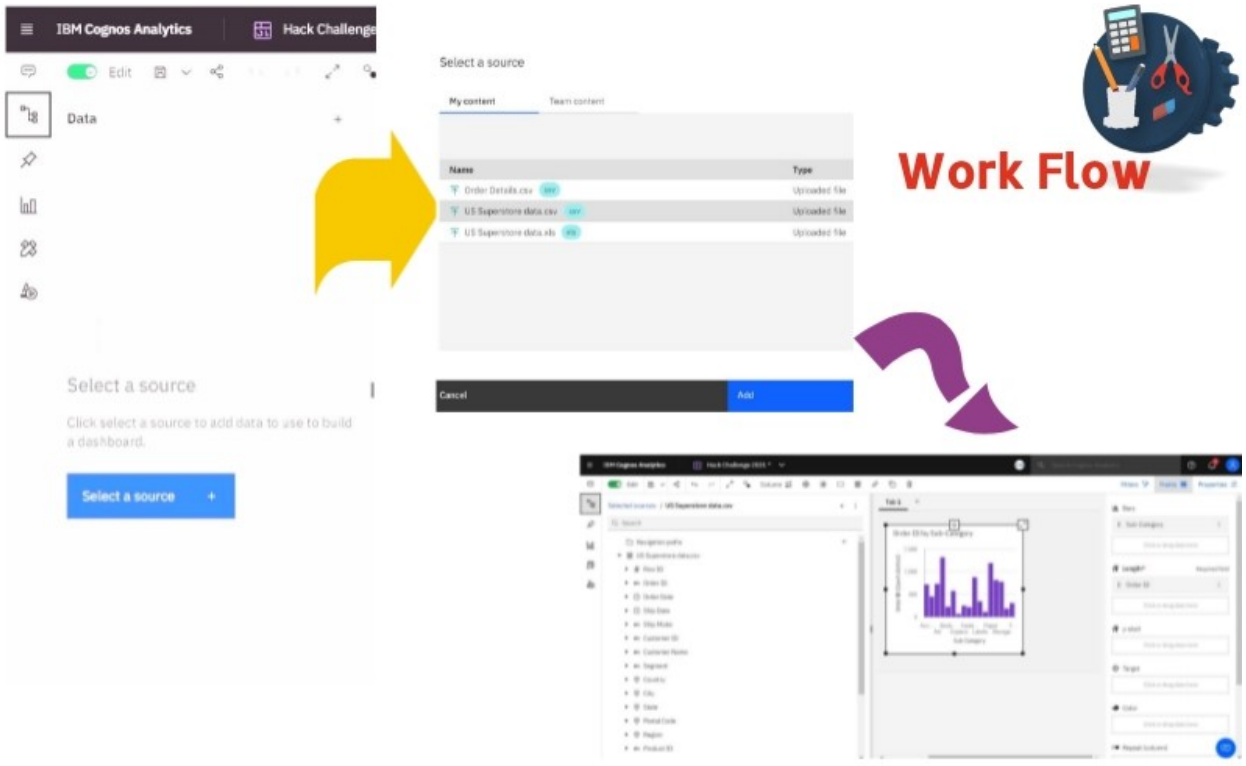
many cuisines are available in both tomato and swiggy people crave over food makes each cuisine top their rank among the other.

5.FLOW CHART

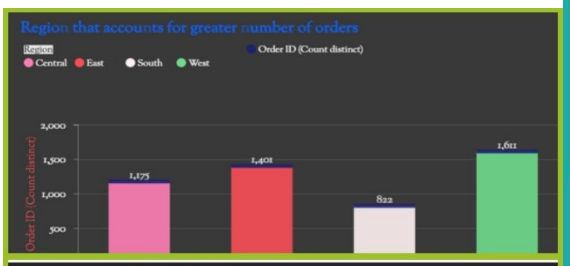
I. SYSTEM ARCHITECTURE:

CONTROL FLOW:

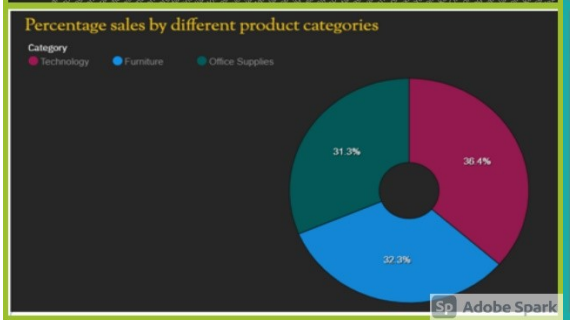



IBM Cognos Analytics
Hack Challenge
Edit
Data
Select a source
Click select a source to add data to use to build a dashboard.
Select a source
Select a source
My content
Team content
Name
Type
Order Details.csv
US Superstore data.csv
US Superstore data.xls
Uploaded file
Cancel
Add
Work Flow

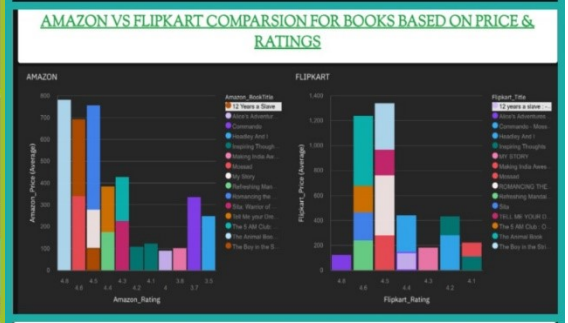
6. RESULTS

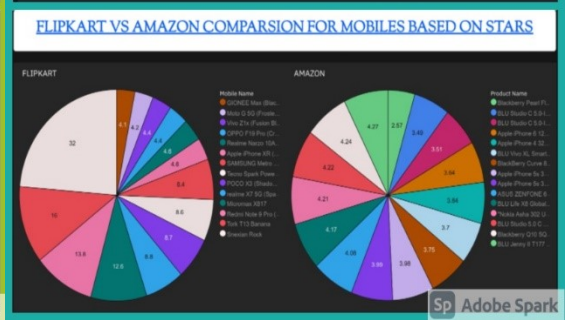

Region that accounts for greater number of orders
Region
Central
East
South
West
Order ID (Count distinct)
2,000
1,500
1,000
500
1,175
1,401
822
1,611


Frequency distribution of quantity ordered
Quantity (Sum)
Quantity Ordered
8000
6000
4000
2000
0

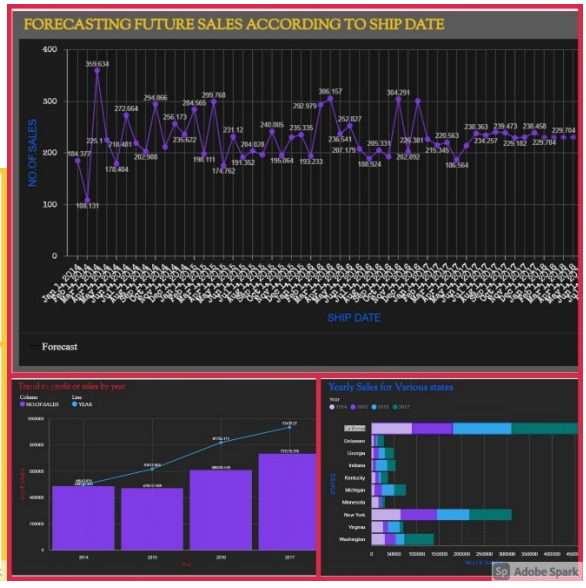

Percentage sales by different product categories
Category
Technology
Furniture
Office Supplies
31.3%
36.4%
32.3%
Adobe Spark


PRICE COMPARISON BETWEEN ZOMATO & SWIGGY BASED ON CUISINE
ZOMATO
SWIGGY
Price(Average)
Cuisine


AMAZON VS FLIPKART COMPARSION FOR BOOKS BASED ON PRICE & RATINGS
AMAZON
FLIPKART
Amazon_Rating
Flipkart_Rating


FLIPKART VS AMAZON COMPARSION FOR MOBILES BASED ON STARS
FLIPKART
AMAZON
Adobe Spark


Product Type That was Ordered Greater Times
NO.OF ORDERS
1,400
1,200
1,000
800
600
400
200
Adobe Spark


FORECASTING FUTURE SALES ACCORDING TO SHIP DATE
NO.OF SALES
SHIP DATE
Forecast
Trend in profit or sales by year
Yearly Sales for Various states
Adobe Spark

7.ADVANTAGES & DISADVANTAGE

Advantages:

Helps to build a robust supply chain

Customers shop online for two reasons – convenience and better prices. This is why the supply chain needs to be robust. Having too many units in stock will take up space in the inventory and increases costs. Striking this balance between supply and demand is tricky, and this is why using analytics to forecast future sales is vital.

Analyse information to detect fraud

Analytics plays a significant role in the detection of fraud. Since analytics can detect patterns based on customer behaviour, it can warn: when an anomalous transaction is detected. As a correction to an e-commerce business, they can send a notification to the customer asking if it is them: who are trying to conduct a particular transaction and await their express approval before processing the transaction.

Measures Marketing

Information related to online marketing analytics is crucial to a retailer. Businesses may pour in hundreds or thousands of dollars in online marketing campaigns, but are not able to determine their effectiveness, will never be able to figure out whether or not the marketing strategy is working.

Predicts what is in store

Analytics for eCommerce merchants helps them determine future trends based on several factors. Like, transactions made, the sales season, section of most ordered products. Such analysis is meaningful in determining what future sales will look like, what products they need to focus on, whether they need to readjust their inventory, what marketing tactics they need to deploy to promote reliable products and what promotions to offer on different products.

Optimize pricing of the products

In the eCommerce industry, the price of a product is determined based on: the demand for a product, its availability in the market and how competitors are pricing the same product.

Let’s to know the customers better

In e-commerce businesses, retailers must understand the types of customers, their behavior as well as the churn cases.

All such analysis is essential in determining the product strategy and any realignment needed to meet revenue and business goals. Such analysis can help in deciding: What promotions to launch to maximize conversions of visitors who are visiting your website.

Disadvantages:

Using predictive analysis for future prediction makes the entire process data driven due to this availability of data act's major role in deciding forecast of sales, so it is necessary to check on the faithfulness of data. And data needs to cleaned and processed before analyzing, that making the process quite longer and hectic.

So far, on the analytics side of eCommerce, their disadvantages are very low and yet to define. Moreover, Data analysis meets its standard only in the big giants like Amazon, eBay, Walmart. Their utilization by other retailers or sellers online had marked remarkably low in percentage. People's choices over products are vast and vibrant their shopping experience not just stops with Amazon or Flipkart but also with boutiques and other shopping stores available online.

Adding to the point current situation needs safe and secured delivery online, further analysis to be carried concentrating more on security and detecting anomalies.

8. REAL TIME APPLICATIONS OF E-COMMERCE BUSINESSES

Amazon is a leader in using a comprehensive, collaborative filtering engine (CFE). The company follows the concept of behavioral analytics. It analyzes the purchasing patterns of the customers from the previously purchased items, items in the shopping cart or on their wish list, the products reviewed and rated by them, to most searched products.

In Flipkart, the aggregation on the 'Click' data provides a 'Confidence Measure' showing the number of times that users clicked on a product for the same search query and hence the store mapping happens almost instantly.

This approach largely depends on the data and works well with categories that are common, popular, and well-known among the users.

9.CONCLUSION

E-commerce has bloomed over the years and is one of the fastest-growing domains in the online world. Though it took some time for this to be accepted by the end-users, today we are at a point where the majority of the people love to shop online. There were numerous concerns revolving around online shopping at its launch, but over years people tend to have started trusting E-commerce for all their shopping needs. In other words, the future of ecommerce will be driven by increasing convenience for both merchants and their shoppers, providing rich, compelling shopping experiences, and enabling experiences across channels in a consistent, customer-friendly way.

10. FUTURE SCOPES

We have planned to enhance this project with updated privacy and security policy for an eCommerce website by assuring the users their details and screen activity will not be shared or used unwantedly. Then the process is to reduce the data-driven dependency in our project. Also we will be concentrating on:

1. Reducing friction in the buying (and selling) journey.
2. Creating rich brand experiences.
3. Selling anywhere and everywhere.

11. BIBILOGRAPHY

1.https://www.researchgate.net/publication/335952834_Ecommerce_and_Business_Analytics_A_Literature_Review

2.https://www.researchgate.net/publication/307946921_Service_Quality_Evaluation_by_ExplorinE_Social_Users'_Contextual_Information

3.https://www.researchgate.net/publication/326453374_Application_of_data_analytics_for_product_design_Sentiment_analysis_of_online_product_reviews

4. https://ieeexplore.ieee.org/document/8276165