**REPORT ON CAPSTONE PROJECTS**

**PROJECT-1 : E-COMMERCE DATASET ANALYSIS**

**INTRODUCTION**

**Problem Statement** -We are a data analyst at an e-commerce firm (just like Amazon) where we on board sellers/retailers to sell their products on your website.

* After a customer purchases a product from your website, the seller gets notified to fulfill that order. Once the customer receives the product, the customer gets a survey email where he/she can share his/her shopping experience.

We have been asked to present a complete summary of the performance of the firm. Include all possible metrics and graph visualizations​ from delivery, pricing, products, payments, etc. ​to help the CEO make rational business decisions.

**Dataset**:

The dataset consists of 7 .csv files:

1. **customers.csv**​ - This dataset has information about the customer and his/her location.

● Each order is assigned to a unique customer \_id. This means that the same customer will be assigned different customer\_ id for different orders.

● There is another column, customer \_unique\_ id, which will remain unique to a customer forever, and you can use this to identify customers that made repurchases on the website.

2. **orders.csv**​ - This is the core dataset.

● order\_purchase\_timestamp: ​ Timestamp at which order was placed ● order\_ approved\_ at: ​ Timestamp at which payment was successful and order was placed/approved ● order\_delivered\_to\_carrier\_date: ​ Date on which the product was delivered to the carrier/logistic service partner

● order\_estimated\_delivery\_date: ​ Estimated delivery date shown to the customer at the time of purchase

● order\_delivered\_to\_customer\_date: ​ Actual date on which the order was delivered to the customer

3. **order\_items.csv**​ - This dataset includes data about the items purchased within each order. ​An order can be of many types, some of which are explained below: a) Orders with a single item a) Orders with multiple items of the same product.

4. **order\_payments.csv**​ - This dataset includes data about the order payment options. A customer pays for the entire order with only one payment method.

5. **products.csv**​ - This dataset includes data about the products sold by your company.

6. **sellers.csv**​ - ​This dataset includes data about the sellers. Use it to identify which seller fulfilled which product.

7. **user\_feedback.csv**​ - This dataset includes data about customer review score (1-5) for every order

. ● survey\_sent\_date - The date on which the feedback survey was sent to the customer

● survey\_answer\_timestamp - The date on which the customer responded to survey.

**TOOLS\SKILLS USED**

* Python Programming
* Tableau Desktop
* Python Libraries and Packages Numpy, Pandas, Matloplib, Seaborn
* Data Visualisation
* Exploratory Data Analysis

**OBJECTIVE**

We have to present a complete summary of the performance of the firm. Include all possible metrics and graph visualizations​ from delivery, pricing, products, payments, etc. ​We have to do extensive EDA to find insights from this data.

**METHODOLOGY**

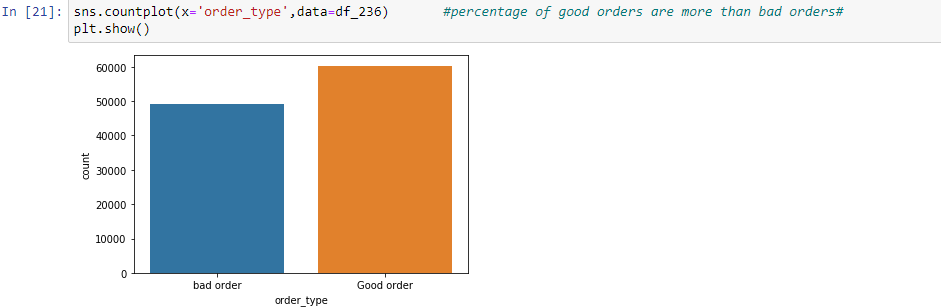
* Data Description
* EDA and Visualisation of the Data
* Visualisation of the Data
* Finding Insights

**VISUALISATION OF THE DATA**

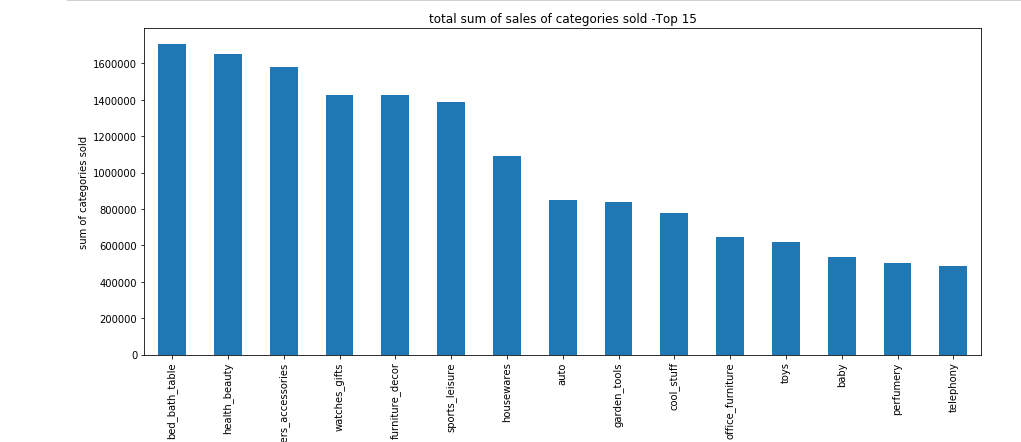
**Visualisation Snippets:**



* + From above analysis we can conclude that No of orders were almost equal in 3-6 month, lower in 9th, 10th,12th month.
  + But it was highest in 7th month maybe this can be due to a sale in that period of time. There was a significant downfall after 8th & 11th month.

****

* From the above analysis we can see that percentages of ‘good orders’ is higher, Which means delivery time of 45% orders are more than 10 days and 55% products are delivered in less than 10 days.



1. From above analysis we can see that the Quantities of ‘bed\_ bath\_ table’ sold is higher and also its sales is higher as compared to other categories.
2. bed\_ bath\_ table, health\_ beauty , computer\_ accessories ,watch\_ gifts, furniture\_ decor are have highest sum of sales.

**RESULT AND INSIGHTS**

**Sales**

* The total sum of sales of year 2016 is significantly lower than 2017 & 2018.From this analysis we can hypothesize that business was in growing stage in 2016,it has taken significant growth in next 2 years.
* No of orders were almost equal in 3-6 month, lower in 9th, 10th, 12th month. But it was highest in 7th month maybe this can be due to a sale in that period of time. There was a significant downfall after 8th & 11th month.

**Orders**

* -97% of orders have ‘order\_ status’ ‘delivered’ ',1% have order\_ status ‘shipped’,0.62% are ‘cancelled’'.
* This states that performance of firm is good in fulfilling the orders. -SP is the state with maximum no of orders & Maximum % of orders are from first 7 states(SP,RJ,MG,RS,PR,SC,BA).

**Products & Payments**

* The Quantities of ‘bed\_bath\_table’ sold is higher and also its sales is higher as compared to other categories. –
* ‘bed\_bath\_table’, ‘health\_beauty’, ‘computer \_accessories’ , ‘watch\_gifts’, ‘furniture\_ decor’ are have highest sum of sales. –
* ‘Average\_ Payment’\_ amount’ is similar for debit card, boleto , credit\_ card but lower for voucher mode
* No of payments & total sum of payments is highest for credit card. This shows us that most of the customers prefers credit card as compared to other modes

**Delivery**

* The average days required to deliver a product to a customer is 12 days. But ,The delivery time of most of orders were between "6-9 Days".
* From analysis we can find that AM is the state with highest Average delivery time & RS is the state with lowest Average delivery time. In other states delivery time is almost equal.
* Firm has to work on delivery time of some product categories for better customer service. -Also percentages of ‘good orders’ is higher ,Which means delivery time of 45% of orders are more than 10 days and 55% orders are delivered in less than 10 days.

**User\_ feedback**

Users have mostly given review score 5 which states us that most of the customers are satisfied with service provided by the firm.