

**Task 3: Exploratory Data Analysis Retail**

**GRIP@The Sparks Foundation**

**Name: Ashutosh Kumar**

**#GRIPMAY21 #TSF #DATASCIENCE #GRIP**

**Dataset:** [**https://bit.ly/3i4rbWl**](https://bit.ly/3i4rbWl)

**DATASET : SAMPLESUPERSTORE.CSV**

Problem Statement:

Perform Exploratory Data Analysis on dataset Sample Superstore

As a business manager, try to find out the weak areas where you can work to make more profit.

GRIP @The Spark Foundation Data Science

Business Analytics Intern Task : Exploratory Data Analysis on Retail.

I have done the task on Jupyter Notebook by including the relevant libraries.

I performed statistics and analysis to raw conclusion on profit, sales and region wise avenue, which

performed well and which didn't and what were the reasons.

The task is for the month of MAY 2021. Thank You The Spark Foundation for giving this

Hello Everyone!

I just began my Data Science journey with The Sparks Foundation and completed the first task of Data Science and Business Analytics internship which was based on applying in statistics and analysis to the given dataset and conclusion on profit, sales and region wise avenue to predict future profits. Here I have created the analysis model of a Super market datasheet given by the company and have deployed the parameters successfully.

Language: Python

IDE: VS CODE / Jupyter notebook

Dataset: https://bit.ly/3i4rbWl

Libraries: numpy, pandas, scikit-learn, matplotlib

You can check it out on my GitHub repository:

Youtube :

Google drive :

Linkdin :

Leave your thoughts ,suggestion and new ideas here

# Exploratory-Data-Analysis-on-Retail-Dataset

Exploratory Data Analysis on Retail Dataset- SampleSuperstore

**Objective of the Task:**

Perform ‘EDA on dataset ‘SampleSuperstore’:

● Here, as a business manager, we will try to find out the weak areas where we can work to make more profit.

● We will find out business problems that can derived by exploring the data.

Language: Python

IDE: Jupyter Notebook /visual studio code

**Introduction**

**preface**

**Procedure:**

Step 1: Import libraries and load dataset

Step 2: Dataset Analysis and Data Cleaning

Step 3: Data Visualization

Step 4: relation between given attributes

Step 5 : Conclusion

**Step in working Code :**

# Importing the libraries

#Loading the given dataset

#Checking the dataset

#Checking the no of rows and columns of the dataset

# Checking summary of the dataset

#Checking the missing values in dataset

#Checking the datatypes of columns

#Checking Statistics of dataset

#Checking unique values in each columns

# Deleting the unwanted columns

#postal code is not required for analysis

#Checking the dataset after removing the columns

# Drawing Heatmap to identify correlation between numerical columns:

# Lineplot between Profit and Discount

# State wise Profit of United States of America

# State wise sales of United States of America

# Analyzing Category wise sales, Discount, Profit:

# Pie Charts of Category wise sales, Discount, Profit

# Analyzing Ship Mode wise sales, Discount, Profit

# Checking the profits and sales of 'Sub- Category'

Conclusions of the Task

1) Top 3 profit giving states: 'California', 'New York', and 'Washington' and . We can Increase sales in these states for increasing profit.

2) Top 3 loss giving states: 'Illinois', 'Pennsylvania' and, 'Ohio' and 'Texas'. We have to pay attention to these states for reducing the loss and making these states on a positive cash flow.

3) Top 3 States with highest Sales : 'California',' New York',' Texas' and 'Washington' and Top 4 states with lowest Sales: 'North Dakota', 'Maine', 'West Virginia 'and ' South Dakota'. We have to increase the sales of low sales states.

4) Texas, Pennsylvania, and Ohio have high Sales but negative Profit i.e. loss. This is becuase of the high discounts given to customer. We need to decrease discount percentage at Texas, Pennsylvania, and Ohio.

5) Illinois has the highest Avg\_discount\_percent among the states of USA but profit and sales are low here. So we can reduce discount % in Illinois

6) Highest profits for each region are as follows: South= Consumer Segment, West= Corporate Segment, Central= Home Office Segment and East= Home Office Segment. We can increase the sales of respective highest profit making segments in the respective region.

7) Lowest profits for each region are as follows: South= Home Office Segment, West= Home Office Segment , Central= Consumer Segment and East= Corporate Segment. We have to focus on these states for increasing their profits.

8) Technology Category has the maximum Sales and Profit. This is because we have given lowest discounts for Technology category. Thus here we can say that If we give more Discount on products our sales goes increases but Profit get down.

9) Office Supplies category has minimum sales and Furniture category has lowest Profit. We have to focus on these categories.

10) Same Day Ship Mode has the highest sales and First Class Ship mode has the highest Profit and Discount. So, we have to use Same Day Ship Mode and First Class Ship in greater frequency.

11) Standard Class Ship mode has lowest sales and lowest profit. So, we can avoid this ship mode.

12) Among 'Sub- Category' items, Copiers has the highest profits in unitedStates. Tables 'Sub- Category' is in loss. So, we have to focus on Tables 'Sub- Category'.

13) Among 'Sub- Category' items, Copiers and machines have the highest Sales in united States .Here we see that copiers had highest sale and Machines had second highest sales .In above two graph we see that 'Machines' had second highest sale but due to large discount it has less profit.

14) Low Sales 'Sub- Category' items: Fasteners, Labels, Paper, Art. We have to concentrate on these 'Sub- Category' items.

**Summary:**

* Profit is more than that of sale but there are some areas where profit could be increased.
* Profit and Discount is high in First Class
* Sales is high for Same day ship
* Sub-category: Copier: High Profit & sales
* Sub-category: Binders , Machines and then tables have high Discount.
* Category: Maximun sales and Profit obtain in Technology.
* Category: Minimun profit obtain in Furniture
* State: Vermont: Highest Profit
* State: Ohio: Lowest Profit
* Segment: Home-office: High Profit & sales
* Here is top 3 city where deals are Highest.
  1. New York City
  2. Los Angeles
  3. Philadelphia
* Sales and Profit are Moderately Correlated.
* Quantity and Profit are less Moderately Correlated.
* Discount and Profit are Negatively Correlated
* Here is top 3 state where deals are Highest.
  1. Califonia
  2. New York
  3. Texas
* **Wyoming** : Lowest Number of deal,Highest amount of sales= Wyoming(11.8%)
* Lowest amount of sales= **South Dakota(0.8%)**

**THANK YOU!**

Refrences :

