COURSE: ADS ASSIGNMENT (STATISTICS AND TRENDS)
COURSE CODE: 7PAM2000 STUDENT NUMBER: 23013682
GITHUB LINK: Link to github

GLOBAL SUPERSTORE SALES DATASET

PROJECT OVERVIEW (INTRODUCTION) :

Global Superstore, a well-known online retailer with an extensive product selection, is based in New York. Dedicated to providing for its international clientele, the business serves clients in 147 nations by providing access to over 10,000 unique products. These products fall into three primary categories: furniture (including chairs), office supplies (such staples), and technology items (like smartphones). The project's goal is to gather insightful information by analyzing the Superstore dataset. These understandings will enable management to take well-informed decisions that raise profitability and performance levels overall.

DATA SOURCE:

The link to the dataset which was gotten from Kaggle is provided below;

GlobalSuperstore - Capstone.xlsx - Google Sheets
Click here

TOOLS USED

Excel: Data loading and cleaning
Python (Jupyter Notebook): Coding and visualization
Github: Documentation

PROJECT STRUCTURE

Data Collection: In the first stage of the project, we focus on obtaining the Global Superstore dataset which was secondary dataset obtained from Kaggle.

Data Cleaning: This process involves checking the data for mistakes, removing copies, handling any missing information, and fixing any problems found. I did this using Power Query in Excel. Also, I used the code **print(Stores_df.isnull().sum())** to see how many missing pieces of information there are in each part of the data (Stores_df). When I found missing info, I used **Stores_df.fillna(0, inplace=True)** to fill in those gaps with the number 0. This change happened right in the original data. Next, I used **print (Stores_df.duplicated().sum())** to see if there were any identical entries in the data. In this case, there weren't any, so everything is good.

Methodology/ Data Visualization:

Reading files: To read the csv file, I saved the file in my file path in my desktop folder using the function **pd.read_csv(file_path, encoding='latin1')**

Descriptive Statistics: I used the **method .describe()** to get the descriptive statistics for the column variables. To get my skewness and Kurtosis, I used the function **.skew()** and .kurtosis() from the python library scipy.stats

Relational Graph: I used a line chart for this to compute the trend of sales, profit and shipping cost over time.

Categorical Graph: I used a pie and bar chart to get the top 3 countries by total profit in 2014, subcategories with highest average shipping cost, Top 10 customer by profit, Profitability of tables in southeast Asia by country, Average profit by product subcategory in Australia.

Statistical Graph: I used a heatmap to compute the correlation matrix among the numerical variables Sales, quantity, Discount, Profit Shipping cost and boxplot to generate the top 3 product sub-categories.

Data Documentation: For this purpose, I've established an extensive documentation repository on GitHub.

EXPLORATORY DATA ANALYSIS (EDA):

(1) Which three nations brought in the most money overall for Global Superstore in 2014?

(2) Determine which three subcategories in the US have the greatest average shipping costs.

(3) Evaluating the profitability (i.e., total profit) of the United Kingdom in 2014. In what way does it differ from other nations?

(4) Determine which product subcategory in Southeast Asia is the least profitable. Is there a particular Southeast Asian

nation where Global Superstore ought to cease carrying the subcategory?

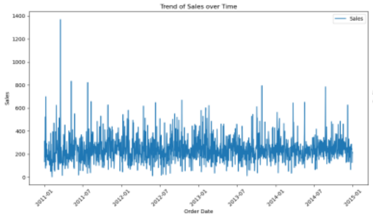
(5) Evaluate the long-term trends in sales, profit, and transportation expenses. In different storylines

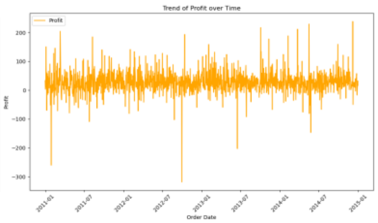
(6) Determine the correlation between the sales dataset's numerical variables.

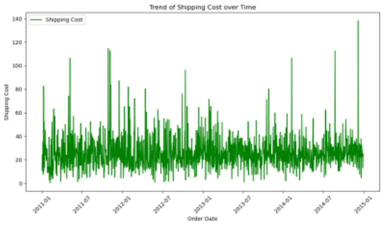
(7) In terms of average profit, which American city is the least profitable? Eliminate the cities with fewer than ten Orders for this study. b) Why is the average profit in this city so low?

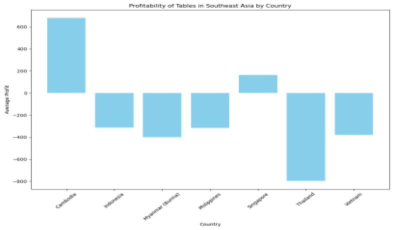
(8) What do the most valuable clients buy, and who are they?

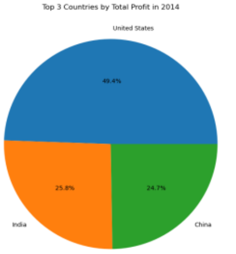
CONCLUSION: Strategic direction for Global Store can be informed by the analysis, which focuses on important markets, product performance, and operational effectiveness. By utilizing these data to inform targeted strategies in product diversification, market expansion, and logistics optimization, Global Store can strengthen its position as a global leader in e-commerce, providing value to customers while optimizing profitability and operational effectiveness. This is merely an overview of the complete dataset according to the assignment guidelines. For Python code documentation, further information is available on Jupiter and github. **REFERRENCE:**1, 2, https://pandas.pydata.org/docs/, https://matplotlib.org/stable/index.html, Bootcamp

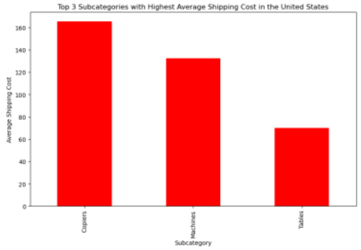

Trend of Sales over Time
Sales
Order Date

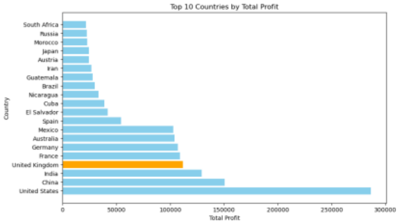

Trend of Profit over Time
Profit
Order Date

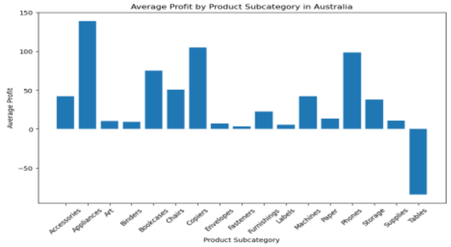

Trend of Shipping Cost over Time
Shipping Cost
Order Date

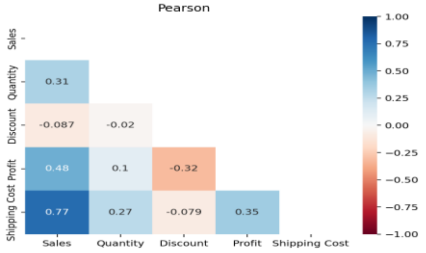

Profitability of Tables in Southeast Asia by Country
Average Profit
Cambodia
Indonesia
Myanmar (Burma)
Philippines
Singapore
Thailand
Vietnam
Country

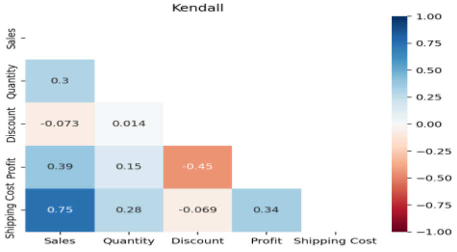

Top 3 Countries by Total Profit in 2014
United States
India
China

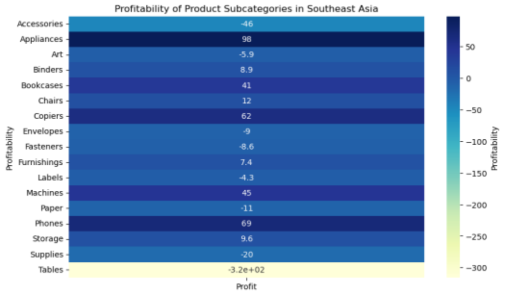

Top 3 Subcategories with Highest Average Shipping Cost in the United States
Average Shipping Cost
Copiers
Machines
Tables
Subcategory


Top 10 Countries by Total Profit
Country
South Africa
Russia
Morocco
Japan
Austria
Iran
Guatemala
Brazil
Nicaragua
Cuba
El Salvador
Spain
Mexico
Australia
Germany
France
United Kingdom
India
China
United States
Total Profit


Average Profit by Product Subcategory in Australia
Average Profit
Accessories
Appliances
Art
Binders
Bookcases
Chairs
Copiers
Envelopes
Fasteners
Furnishings
Labels
Machines
Paper
Phones
Storage
Supplies
Tables
Product Subcategory


Pearson
Sales
Quantity
Discount
Profit
Shipping Cost
Sales
Quantity
Discount
Profit
Shipping Cost


Kendall
Sales
Quantity
Discount
Profit
Shipping Cost
Sales
Quantity
Discount
Profit
Shipping Cost


Profitability of Product Subcategories in Southeast Asia
Profitability
Accessories
Appliances
Art
Binders
Bookcases
Chairs
Copiers
Envelopes
Fasteners
Furnishings
Labels
Machines
Paper
Phones
Storage
Supplies
Tables
-3.2e+02
Profit
Profitability