20XD46 – DATA ANALYTICS AND VISUALIZATION LAB

Package Submission

SUPERSTORE SALES ANALYSIS AND FORECASTING

Done By

20PD05 ANU RAMYA R
20PD24 SAMYUKTHA V

ABSTRACT

In this era of technology development, every business wants to equip its salesforce with a sustainable salesforce automation system to improve sales performance and customer relationship management capabilities.

All companies rely on sales figures to see how they are performing. That said, looking at the revenue generated cannot help pinpoint what's working (or not). Without metrics to guide you, it's hard to make educated and data-driven decisions that will benefit your company now and in the future. This is where sales analysis comes in.

A superstore with operations across the US aims to understand some certain features of the business and get vital information from its data to be able to plan or focus on what is selling with a full understanding of sales trend and forecasting for the future.

DATASET DESCRIPTION

The Super Store dataset contains data on order details of customers for orders of a superstore in the US. This includes the order details, customer details, product details, sales, profits etc.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer	Segment	Country	City	State	Postal Cod	Region
2	1	CA-2016-1	08-11-2016	11-11-2016	Second Cla	CG-12520	Claire Gut	Consumer	United Sta	Hendersor	Kentucky	42420	South
3	2	CA-2016-1	08-11-2016	11-11-2016	Second Cla	CG-12520	Claire Gut	Consumer	United Sta	Hendersor	Kentucky	42420	South
4	3	CA-2016-1	12-06-2016	16-06-2016	Second Cla	DV-13045	Darrin Van	Corporate	United Sta	Los Angele	California	90036	West
5	4	US-2015-1	11-10-2015	18-10-2015	Standard C	SO-20335	Sean O'Do	Consumer	United Sta	Fort Laude	Florida	33311	South
6	5	US-2015-1	11-10-2015	18-10-2015	Standard C	SO-20335	Sean O'Do	Consumer	United Sta	Fort Laude	Florida	33311	South

The Columns are:
Row ID, Order ID, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, Country, City, State, Postal Code, Region, Product ID, Category, Sub-Category, Product Name, Sales, Quantity, Discount and Profit.

The dataset comprises of the daily sales details for a period of four years (2014-2018).

Dataset link: https://www.kaggle.com/datasets/juhi1994/superstore

Database: Oracle SQL Developer

DATA VISUALIZATION TOOLS

- Tableau

  Interactive dashboards have been created using Tableau software to gain insights from the sales data. The dashboards have been published to Tableau Public.

- Matplotlib

  Matplotlib is used to visualize the time series data in Python. The plots help to visualize if data is stationary or not, which model best fits the data, etc.

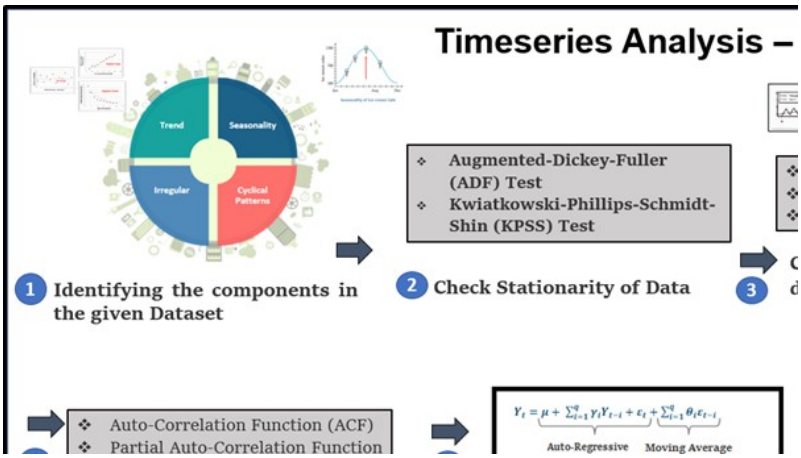
TOOLS USED

- **Oracle SQL Developer**
  The dataset is loaded into the Oracle database which is connected to Tableau.

- **Pandas**
  The pandas package in Python is used to perform data pre-processing and basic EDA to gather insights on a preliminary level.

- **statsmodels.tsa**

  This library provides functions which help to implement the various stages involved in building a Time Series Model:

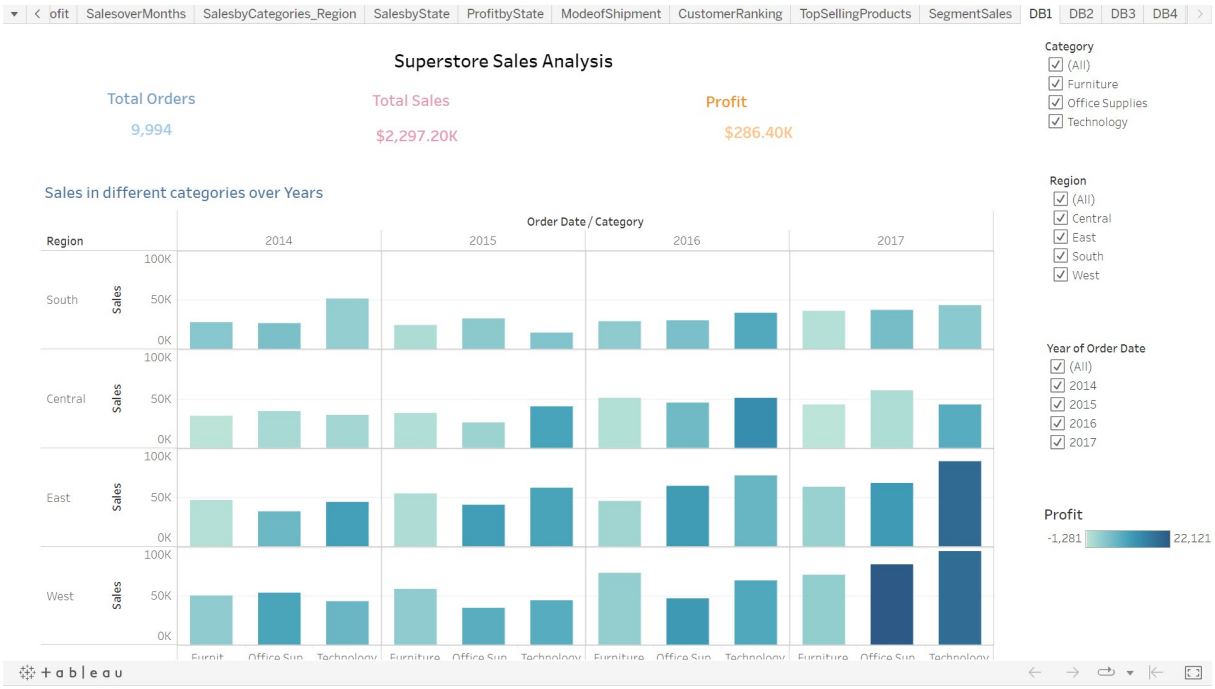


- **Streamlit**
  Streamlit platform has been used to create webpage for this project which has been deployed on Streamlit Cloud.

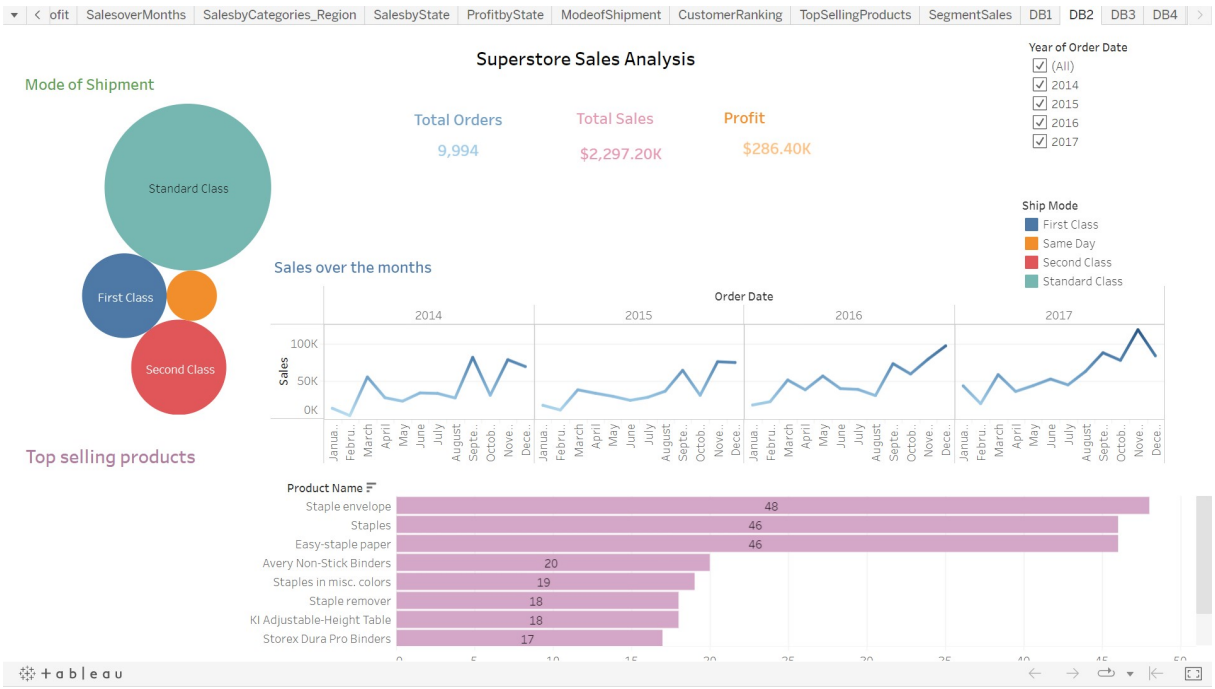
RESULTS

Snapshots of the dashboards created in Tableau are as follows:

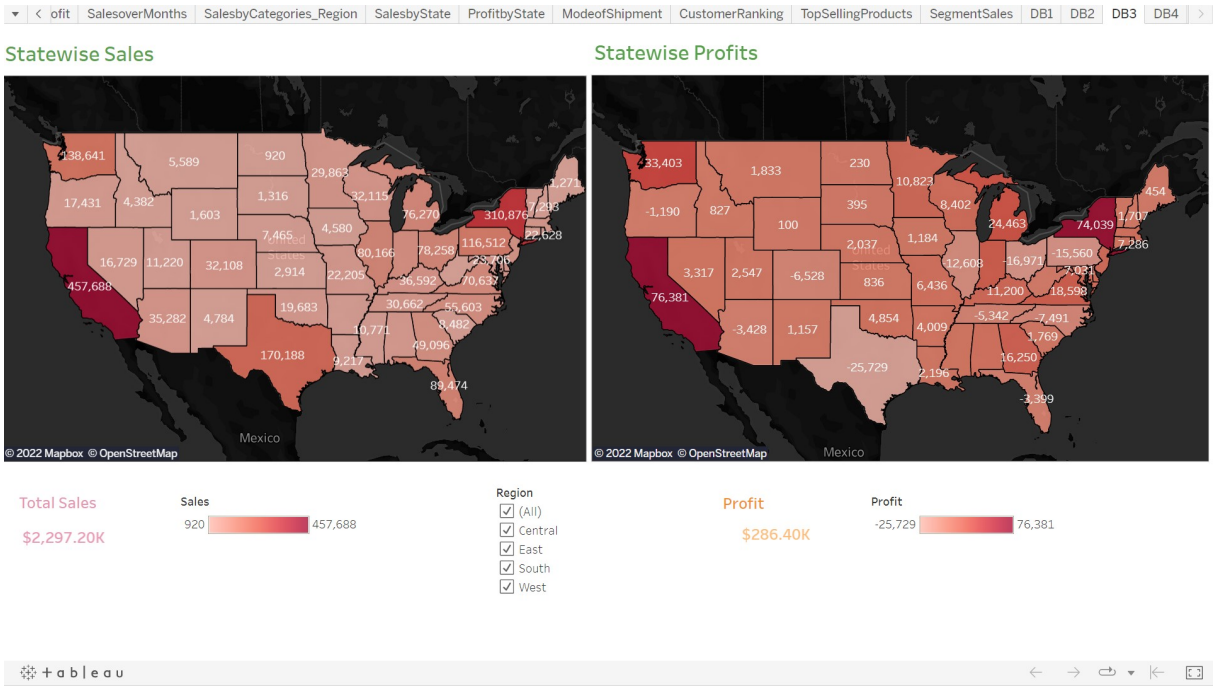
Dashboard 1:



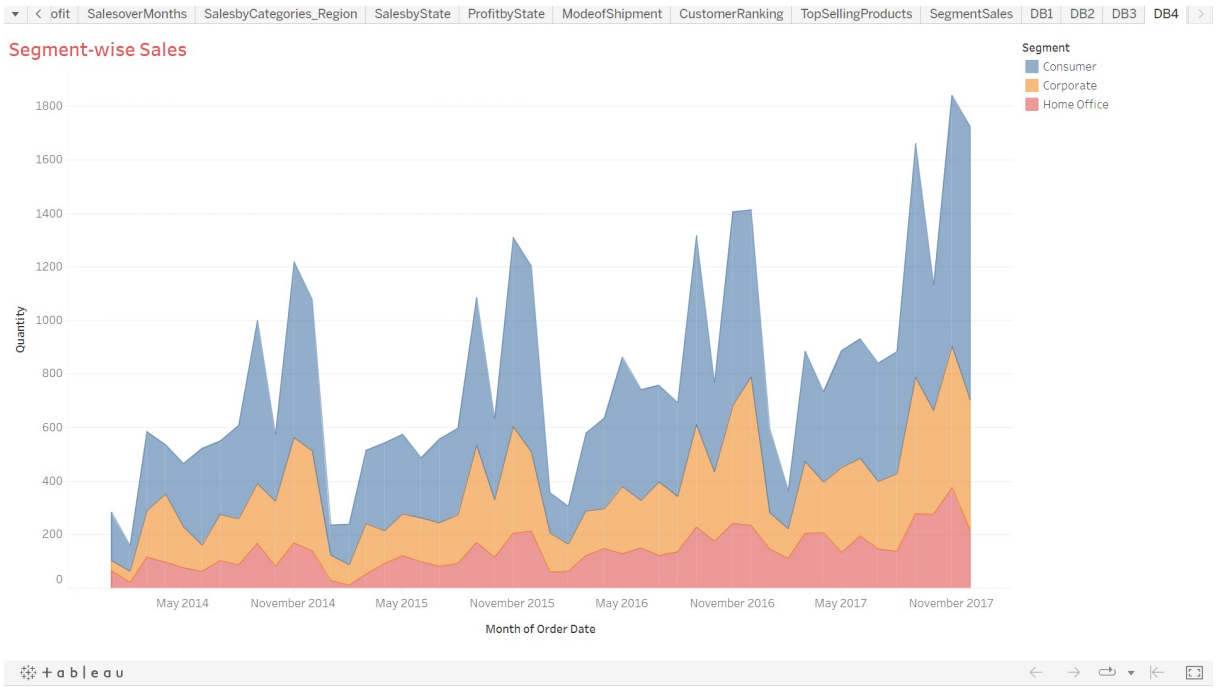
Dashboard 2:



Dashboard 3:



Dashboard 4:



Following insights can be gathered from the viualizaations:

- Total number of orders : 9994
- Total Sales : $2297.20K
- Total Profit : $286.40K
- The superstore has had a general increase in sales over the years.
- Products which belong to the category of Technology bring the maximum profits.
- The store has been performing poorly in the South region comparatively.
- Office Supplies have been the highest purchased items.
- High majority of customers have preferred the standard mode of delivery.
- The state of California has been the top performing state both in terms of sales and profits.
- The state of Texas has done well in sales, but has not brought any profits.
- A high population of the customers are general consumers, followed by the Corporate.

Results of Time Series Analysis:

Different models are built to fit the data and the model with the least RMSE (Root mean squared error) has been used for future sales prediction.

Model	RMSE
AR(3)	0.272
MA(2)	0.261
ARIMA	0.258
SARIMAX	0.241

The SARIMAX model (lowest RMSE) has been used to predict future sales.

LINKS

- **Webpage link**
  https://share.streamlit.io/anu4902/superstoresales/main/str.py

- **Link to Colab file**
  https://drive.google.com/file/d/1cnw7JpJgUeCqVC7Doi4yI0zK75LlNMe6/view?usp=sharing

- **Link to Tableau dashboard:**
  https://public.tableau.com/views/SuperstoreSalesAnalysis_16550189883770/SegmentSales?:language=en-US&:display_count=n&:origin=viz_share_link

REFERENCES

- Time Series Analysis: https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-to-time-series-analysis/

- Streamlit tutorial : https://youtube.com/playlist?list=PLuU3eVwK0I9PT48ZBYAHdKPFazhXg76h5