

---

## SUPERSTORE DATA ANALYSIS USING PYTHON AND TABLEAU

**Komal Nagarkar<sup>\*1</sup>, Asmita Dhanawade<sup>\*2</sup>, Asmita Ulaswar<sup>\*3</sup>, Sneha Powar<sup>\*4</sup>,  
Supriya Kagale<sup>\*5</sup>, Amruta Bhoi<sup>\*6</sup>**

<sup>\*1,2,3,4,5,6</sup>Department Of Computer Engineering, Sharad Institute Of Technology Polytechnic,

Yadav (Ichalkaranji) Maharashtra, 416121, India

DOI : <https://www.doi.org/10.56726/IRJMETS45521>

---

### ABSTRACT

The project revolves around leveraging machine learning techniques to analyze the Superstore dataset, perform Exploratory Data Analysis (EDA), and implement clustering algorithms. The goal is to gain valuable insights into customer behavior and optimize the store's operations based on the findings.

The Superstore dataset contains information about sales transactions, customer demographics, product categories, and other relevant variables. The first step of the project involves conducting EDA to understand the dataset's structure, identify any data quality issues such as missing values or outliers, and visualize key features. This analysis provides a foundation for further exploration and helps uncover patterns and trends within the data.

After completing the EDA phase, the project focuses on customer segmentation. Clustering algorithms, such as K-means, DBSCAN, or hierarchical clustering, are applied to relevant features of the dataset. These algorithms group similar customers together based on their purchasing behavior, demographics, or geographic locations. By segmenting customers into distinct groups, the project aims to gain a deeper understanding of their needs, preferences, and profitability.

The final phase of the project involves extracting actionable insights from the clustering analysis. By analyzing the characteristics of each customer segment, the project aims to identify the most valuable customer groups and their specific purchasing patterns. This information enables the generation of data-driven recommendations to enhance the store's marketing strategies, product assortment, pricing, and supply chain management.

Overall, the project aims to provide the Superstore with valuable insights and recommendations that can drive improvements in various aspects of their operations. By leveraging machine learning techniques and conducting thorough EDA, the project contributes to optimizing customer satisfaction, increasing sales, and ultimately enhancing the overall success of the Superstore.

**Keywords:** Superstore Data Analysis Using Python & Tableau.

---

### I. INTRODUCTION

In this article, I am going to perform an Exploratory Data Analysis on the Sample Superstore dataset.

The link for the Dataset is: [www.kaggle.com](http://www.kaggle.com)

You can download it from the link.

In this dataset, we have many features like ship mode, Segment, country, City, State, Postal code, Region, category, sub-category, sales, Quantity, discount, and the Dependent variable is profit. The remaining are independent variables.

Here we will see for weak areas where there is less profit and will see how to overcome it.

### II. LITERATURE SURVEY

Following articles are studied during the literature survey phase of this work.

#### 2.1 Used Library

- **Numpy :**

NumPy is a Python library used for working with arrays.

It also has functions for working in domain of linear algebra, fourier transform, and matrices.

NumPy was created in 2005 by Travis Oliphant. It is an open source project and you can use it freely.

NumPy stands for Numerical Python.

- **Pandas :**

Pandas is a Python library.

Pandas is used to analyze data.

- **Matplotlib :**

Matplotlib is a low level graph plotting library in python that serves as a visualization utility.

Matplotlib was created by John D. Hunter.

Matplotlib is open source and we can use it freely.

Matplotlib is mostly written in python, a few segments are written in C, Objective-C and Javascript for Platform compatibility.

- **Plotly link :**

Plotly's Python graphing library makes interactive, publication-quality graphs. Examples of how to make line plots, scatter plots, area charts, bar charts, error bars, box plots, histograms, heatmaps, subplots, multiple-axes, polar charts, and bubble charts.

## 2.2 Look Of The Data

	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Order Date	Ship Date	Ship Mode	Customer ID	Customer Na	Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Nam	Sales	Quantity	Discount	Profit
2	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1000	Furniture	Bookcases	Bush Somers	261.96	2	0	41.9136
3	08-11-2016	11-11-2016	Second Class	CG-12520	Claire Gute	Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1000	Furniture	Chairs	Hon Deluxe F	731.94	3	0	219.582
4	12-06-2016	16-06-2016	Second Class	DV-13045	Darrin Van Hi	Corporate	United States	Los Angeles	California	90036	West	OFF-LA-1000	Office Suppli	Labels	Self-Adhesive	14.62	2	0	6.8714
5	11-10-2015	18-10-2015	Standard Clai	SO-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-1000	Furniture	Tables	Bretford CRA4	957.5775	5	0.45	-383.031
6	11-10-2015	18-10-2015	Standard Clai	SO-20335	Sean O'Donn	Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-1000	Office Suppli	Storage	Eldon Fold 'N	22.368	2	0.2	2.5164
7	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1000	Furniture	Furnishings	Eldon Expre	48.86	7	0	14.1694
8	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1000	Office Suppli	Art	Newell 322	7.28	4	0	1.9656
9	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Mitel 5320 IP	907.152	6	0.2	90.7152
10	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	OFF-BI-1000	Office Suppli	Binders	DXL Angle-Vi	18.504	3	0.2	5.7825
11	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	OFF-AP-1000	Office Suppli	Appliances	Belkin FSC20	114.9	5	0	34.47
12	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	FUR-TA-1000	Furniture	Tables	Chromcraft R	1706.184	9	0.2	85.3092
13	09-06-2014	14-06-2014	Standard Clai	BH-11710	Brosina Hoffr	Consumer	United States	Los Angeles	California	90032	West	TEC-PH-1000	Technology	Phones	Konftel 250 C	911.424	4	0.2	68.3568
14	15-04-2017	20-04-2017	Standard Clai	AA-10480	Andrew Allen	Consumer	United States	Concord	North Carolin	28027	South	OFF-PA-1000	Office Suppli	Paper	Xerox 1967	15.552	3	0.2	5.4432
15	05-12-2016	10-12-2016	Standard Clai	IM-15070	Irene Maddo	Consumer	United States	Seattle	Washington	98103	West	OFF-BI-1000	Office Suppli	Binders	Fellowes PB2	407.976	3	0.2	132.5922
16	22-11-2015	26-11-2015	Standard Clai	HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-1000	Office Suppli	Appliances	Holmes Repli	68.81	5	0.8	-123.858
17	22-11-2015	26-11-2015	Standard Clai	HP-14815	Harold Pawla	Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-1000	Office Suppli	Binders	Storex DuraTr	2.544	3	0.8	-3.816
18	11-11-2014	18-11-2014	Standard Clai	PK-19075	Pete Kriz	Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-1000	Office Suppli	Storage	Stur-D-Stor S	665.88	6	0	13.3176
19	13-05-2014	15-05-2014	Second Class	AG-10270	Alejandro Gr	Consumer	United States	West Jordan	Utah	84084	West	OFF-ST-1000	Office Suppli	Storage	Fellowes Sup	55.5	2	0	9.99
20	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United States	San Francisc	California	94109	West	OFF-AR-1000	Office Suppli	Art	Newell 341	8.56	2	0	2.4824
21	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United States	San Francisc	California	94109	West	TEC-PH-1000	Technology	Phones	Cisco SPA 50	213.48	3	0.2	16.011
22	27-08-2014	01-09-2014	Second Class	ZD-21925	Zuschuss Do	Consumer	United States	San Francisc	California	94109	West	OFF-BI-1000	Office Suppli	Binders	Wilson Jones	22.72	4	0.2	7.384
23	09-12-2016	13-12-2016	Standard Clai	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-1000	Office Suppli	Art	Newell 318	19.46	7	0	5.0596
24	09-12-2016	13-12-2016	Standard Clai	KB-16585	Ken Black	Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-1000	Office Suppli	Appliances	Acco Six-Outl	60.34	7	0	15.6884
25	16-07-2017	18-07-2017	Second Class	SF-20065	Sandra Flana	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-1000	Furniture	Chairs	Global Delux	71.372	2	0.3	-1.0196
26	25-09-2015	30-09-2015	Standard Clai	EB-13870	Emily Burns	Consumer	United States	Orem	Utah	84057	West	FUR-TA-1000	Furniture	Tables	Bretford CRA4	1044.63	3	0	240.2649
27	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffman	Consumer	United States	Los Angeles	California	90049	West	OFF-BI-1000	Office Suppli	Binders	Wilson Jones	11.648	2	0.2	4.2224
28	16-01-2016	20-01-2016	Second Class	EH-13945	Eric Hoffman	Consumer	United States	Los Angeles	California	90049	West	TEC-AC-1000	Technology	Accessories	Imation 8GB	90.57	3	0	11.7741
29	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-BO-1000	Furniture	Bookcases	Riverside Pal	3083.43	7	0.5	-1665.0522
30	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-1000	Office Suppli	Binders	Avery Recycl	9.618	2	0.7	-7.0532
31	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-FU-1000	Furniture	Furnishings	Howard Mille	124.2	3	0.2	15.525
32	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-EN-1000	Office Suppli	Envelopes	Poly String Ti	3.264	2	0.2	1.1016
33	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-1000	Office Suppli	Art	BOSTON Moc	86.304	6	0.2	9.7092
34	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-BI-1000	Office Suppli	Binders	Acco Pressbo	6.898	6	0.7	-5.715
35	17-09-2015	21-09-2015	Standard Clai	TB-21520	Tracy Blumst	Consumer	United States	Philadelphia	Pennsylvania	19140	East	OFF-AR-1000	Office Suppli	Art	Lumber Crayc	15.76	2	0.2	3.546
36	19-10-2017	23-10-2017	Second Class	MA-17560	Matt Abelma	Home Office	United States	Houston	Texas	77095	Central	OFF-PA-1000	Office Suppli	Paper	Easy-staple p	29.472	3	0.2	9.9468

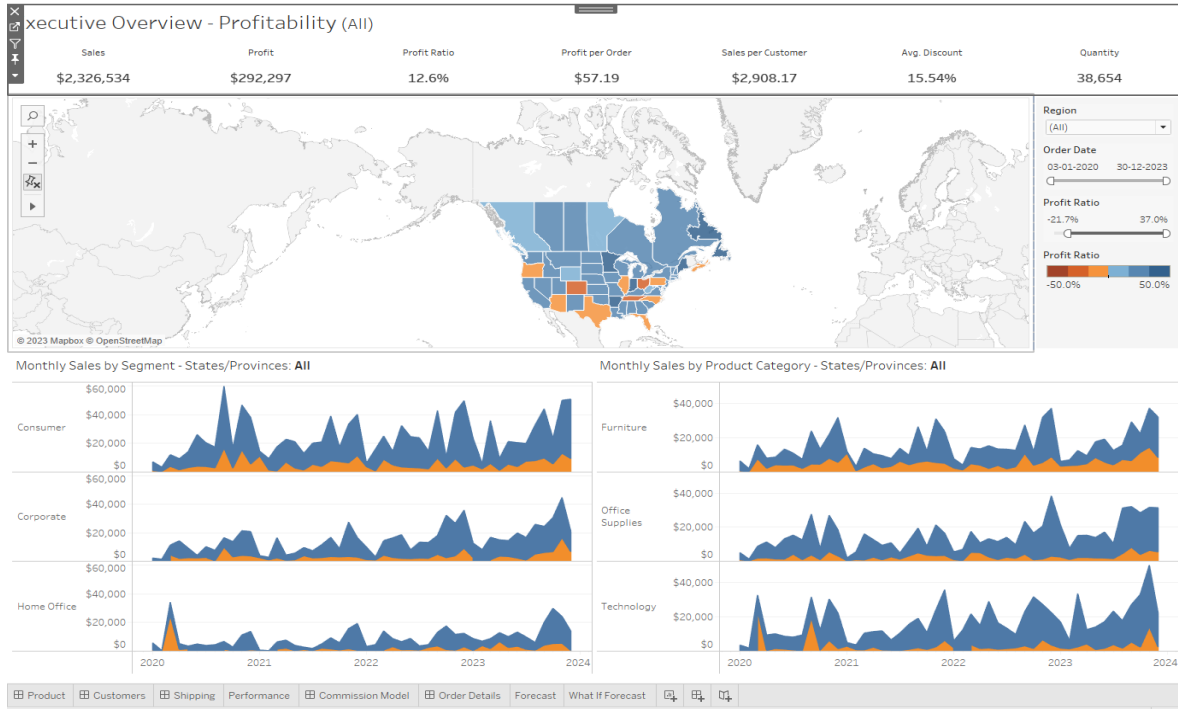
## 2.3 Data Cleaning

Data Cleansing is the process of detecting and changing raw data by identifying incomplete, wrong, repeated, or irrelevant parts of the data. For example, when one takes a data set one needs to remove null values, remove that part of data we need based on application, etc. Besides this, there are a lot of applications where we need to handle the obtained information

## 2.4 Data Visualization

Data Visualization is a technique of presenting data graphically or in a pictorial format which helps to understand large quantities of data very easily. This allows decision-makers to make better decisions and also allows identifying new trends, patterns in a more efficient way.

### 2.5 Executive Overview-Profitability(All)



### 2.6 Sales By Product Category

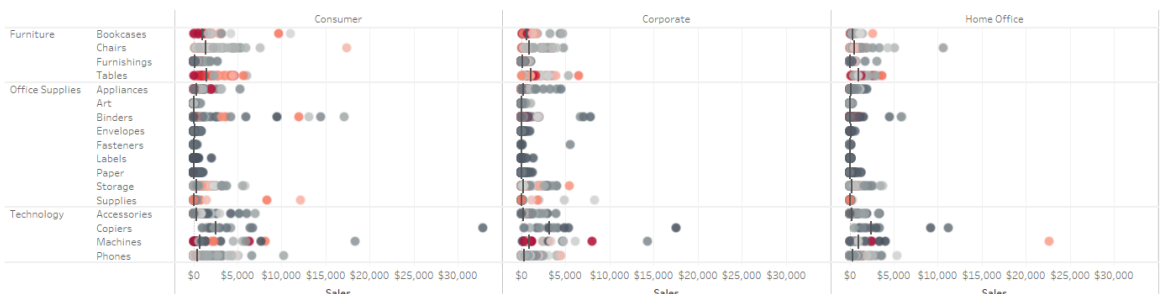
Product Drilldown

Sales by Product Category

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sept	Oct	Nov	Dec
<b>Furniture</b>												
2020	\$6,308	\$1,840	\$15,817	\$7,945	\$8,505	\$13,206	\$10,821	\$7,320	\$23,816	\$13,136	\$21,763	\$31,648
2021	\$11,805	\$3,134	\$13,743	\$10,476	\$9,419	\$7,714	\$13,674	\$9,639	\$26,273	\$12,027	\$30,881	\$23,546
2022	\$7,688	\$3,926	\$14,116	\$13,291	\$15,160	\$13,231	\$13,116	\$12,483	\$27,327	\$11,889	\$31,784	\$36,970
2023	\$6,029	\$6,866	\$12,369	\$9,323	\$17,483	\$19,009	\$12,476	\$15,442	\$29,097	\$22,246	\$37,057	\$31,917
<b>Office Supplies</b>												
2020	\$4,858	\$1,072	\$8,606	\$11,155	\$7,872	\$13,027	\$15,121	\$12,388	\$27,487	\$7,298	\$26,862	\$18,267
2021	\$1,822	\$5,368	\$15,892	\$12,559	\$9,185	\$10,648	\$4,720	\$11,894	\$19,337	\$8,673	\$21,218	\$16,355
2022	\$5,313	\$6,794	\$17,347	\$10,647	\$13,050	\$11,334	\$13,915	\$9,561	\$23,311	\$16,403	\$20,487	\$38,210
2023	\$21,287	\$7,408	\$14,931	\$15,072	\$13,763	\$16,986	\$10,304	\$21,069	\$31,946	\$28,334	\$31,479	\$31,256
<b>Technology</b>												
2020	\$3,353	\$1,609	\$32,511	\$9,195	\$9,943	\$8,436	\$8,004	\$9,210	\$31,367	\$11,979	\$30,201	\$22,093
2021	\$4,835	\$3,449	\$10,344	\$11,161	\$11,643	\$6,435	\$10,371	\$15,525	\$19,017	\$10,707	\$23,874	\$35,632
2022	\$5,830	\$12,259	\$21,568	\$14,891	\$28,833	\$16,372	\$13,269	\$9,672	\$22,883	\$31,540	\$27,141	\$22,323
2023	\$16,943	\$6,027	\$33,429	\$12,383	\$13,910	\$17,061	\$23,210	\$17,619	\$27,022	\$32,895	\$49,919	\$22,002

Sales and Profit by Product Names

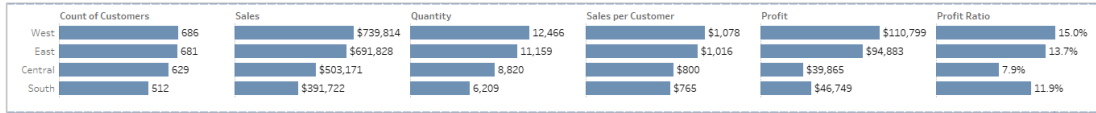
Year: All, Month: All, Product Category: All



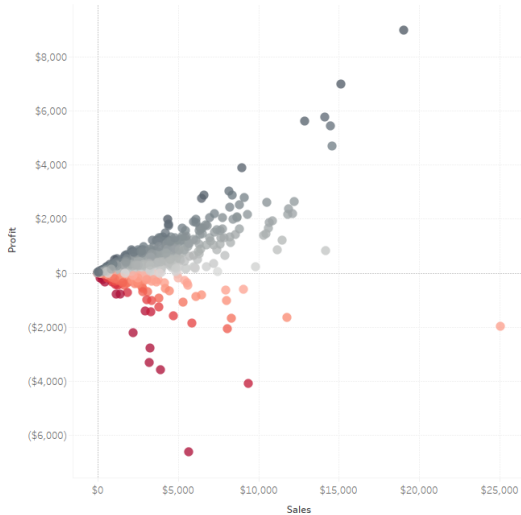
The figure above show us a comparison between sales and profit by the Product names, the most profitable category is copiers and binders, and most non profitable is Machine, chairs and tables. So from this figure. we can find out what make them non profitably.

### 2.7 Customer Analysis

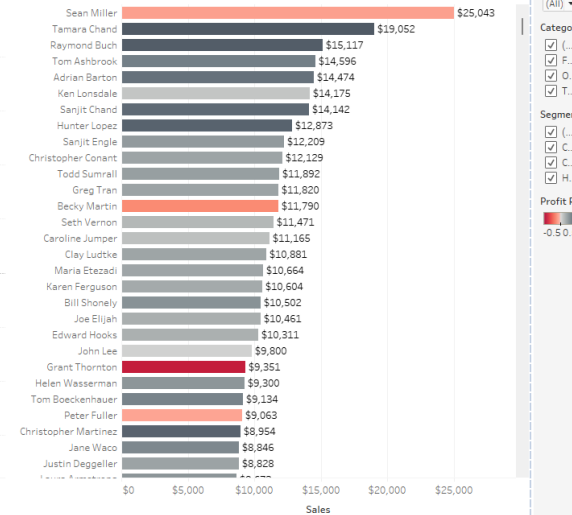
Customer Analysis



Sales and Profit by Customer

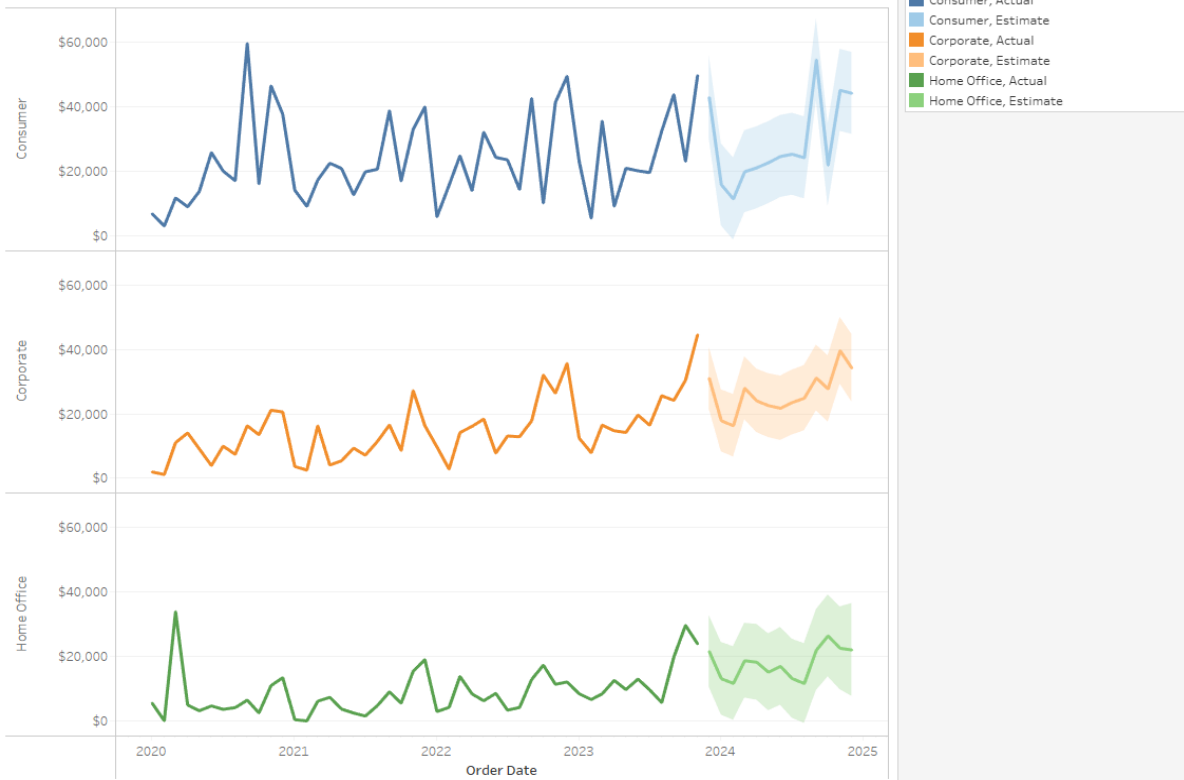


Customer Ranking



### 2.8 Sales Forecast

Sales Forecast



**Strengths:**

Tableau is a powerful data visualization tool used in the Data Analytics and Business Intelligence Industry. It helps in simplifying raw data which in turn can be easily turned into an understandable format., to gain meaningful insights out of the data.

The great thing about tableau is its user-friendliness it offers to the users for various analysis. By using Tableau, even a non-technical user can create a customized dashboard. The best feature Tableau are

- Data Blending
- Real time analysis
- Collaboration of data

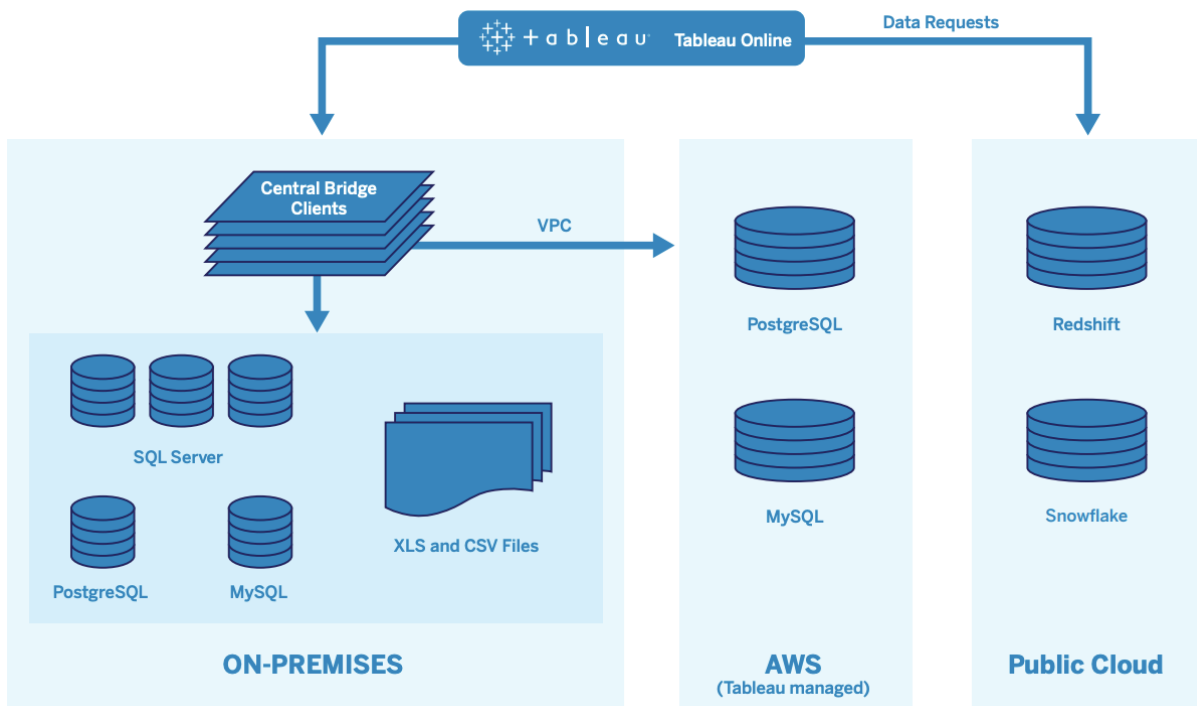
**Use:**

The use of this project is that if we can see how many sales have been sold.

Can withdraw the profit of the month.

You can see when the shipping is on which day.

**III. METHODOLOGY**



**System Architecture**

Tableau manages many different data sources, including SQL Server, PostgreSQL, MySQL, Redshift, Snowflake, Google Big Query, and flat files across various networks. Public Cloud data connections are generally made directly. For private connections, a pool of Tableau Bridge services facilitates the live on-premises or VPC database query capabilities, and scheduled extract refreshes.

**IV. CONCLUSION**

Here we can easily view that there is a huge gap Between the Central and the south region The Sales in the Central region is the highest in all the Years under consideration.

Moreover, as the Slope of the central region is upward sloping, so we can conclude that sales in the Central region is increasing overtime. But as far as the other regions are concerned, the sales is either increasing at a very slow pace or either stagnated, particularly for the Caribbean and Canada Region

#### V. REFERENCES

- [1] Data Src : <https://community.tableau.com/docs/DOC-1236> Hafizhan Aliady | Data Mining | 26 Oct 2017
- [2] Tableau Server /Tableau Cloud
- [3] <https://www.realcode4you.com/post/sample-superstore-sales-data-analysis-using-tableau-tableau-data-analysis-help-realcode4you>
- [4] Use Tableau Public
- [5] Use Jupyter Notebook
- [6] Kaggle.com
- [7] GeeksforGeeks.com