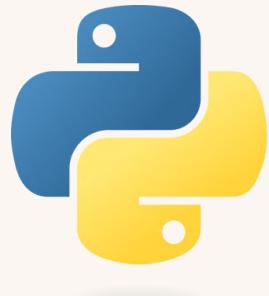
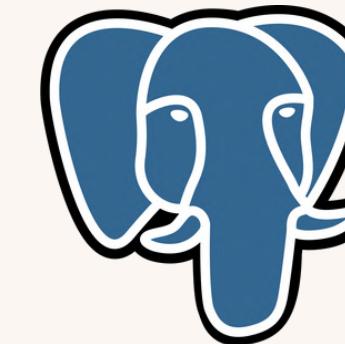


FAISAL MALEKZADA

Data Analyst

Email: Faisalmalikzada1@gmail.com



Power BI



Portfolio Overview

A Showcase strong analytical skills by transforming complex datasets into actionable insights.

Skills used in analyzing the datasets include but not limited to SQL, Python, Excel, and data visualization tools like Tableau, Power BI and Powerpoint.

Projects

Influenza Season

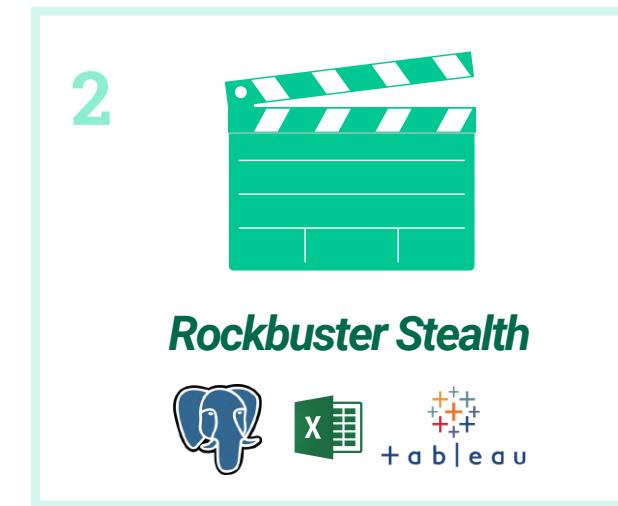
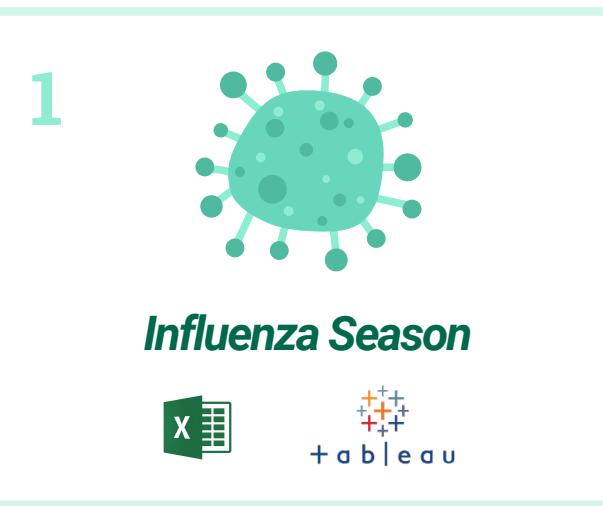
Rockbuster Stealth

Instacart Grocery Basket

Pig E. Bank

Looker E-Commerce

Project Overview



Project Contents

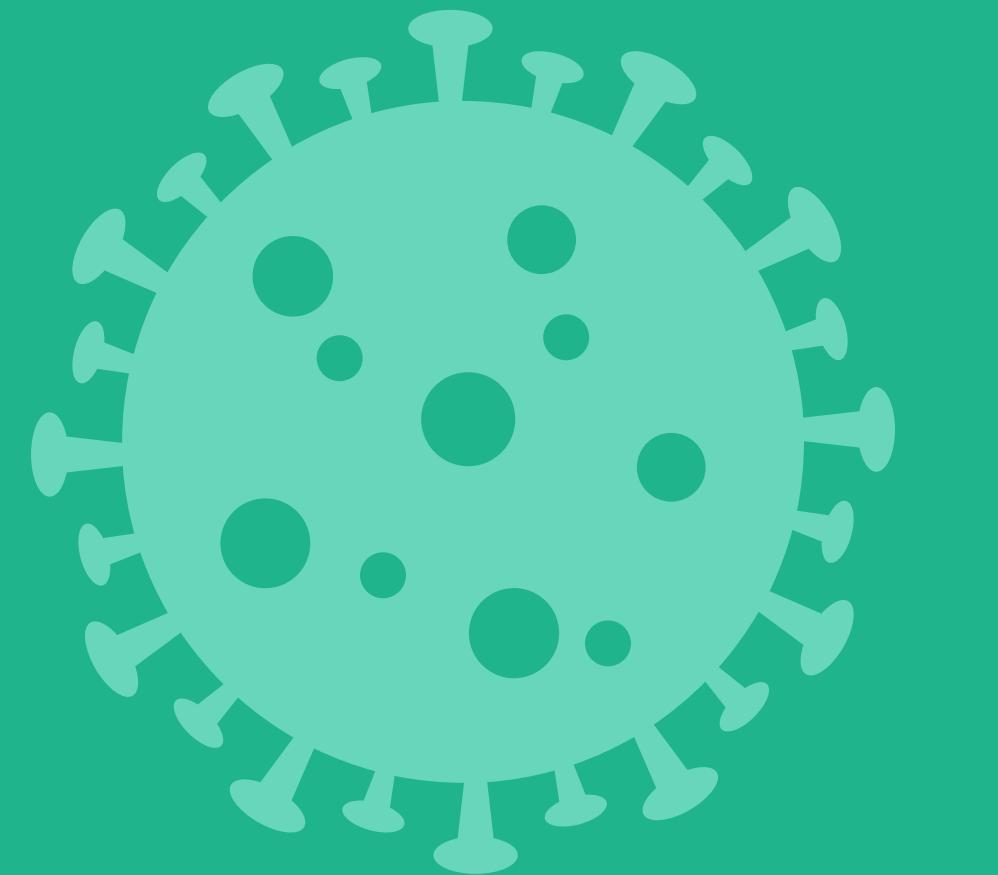
Business Objective & Data

Data Analysis Approach

Key Insights & Visualizations

Flu Season

1



Context and Project Objective

- Help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis.
- Determine when to send staff, and how many, to each state.

Data Sets



- Influenza deaths by geography - (66,097 Rows)
- Population data by geography, time, age, and gender - (28,986 Rows)
- Counts of influenza laboratory test results by state (survey) - (24,952 Rows)
- Survey of flu shot rates in children - (28,466 Rows)

Data Analysis Methods Used



- Data Cleaning
- Data Integration: influenza deaths + population.
- Hypothesis testing.
- Correlation coefficient testing.
- Visualizing with excel and Tableau

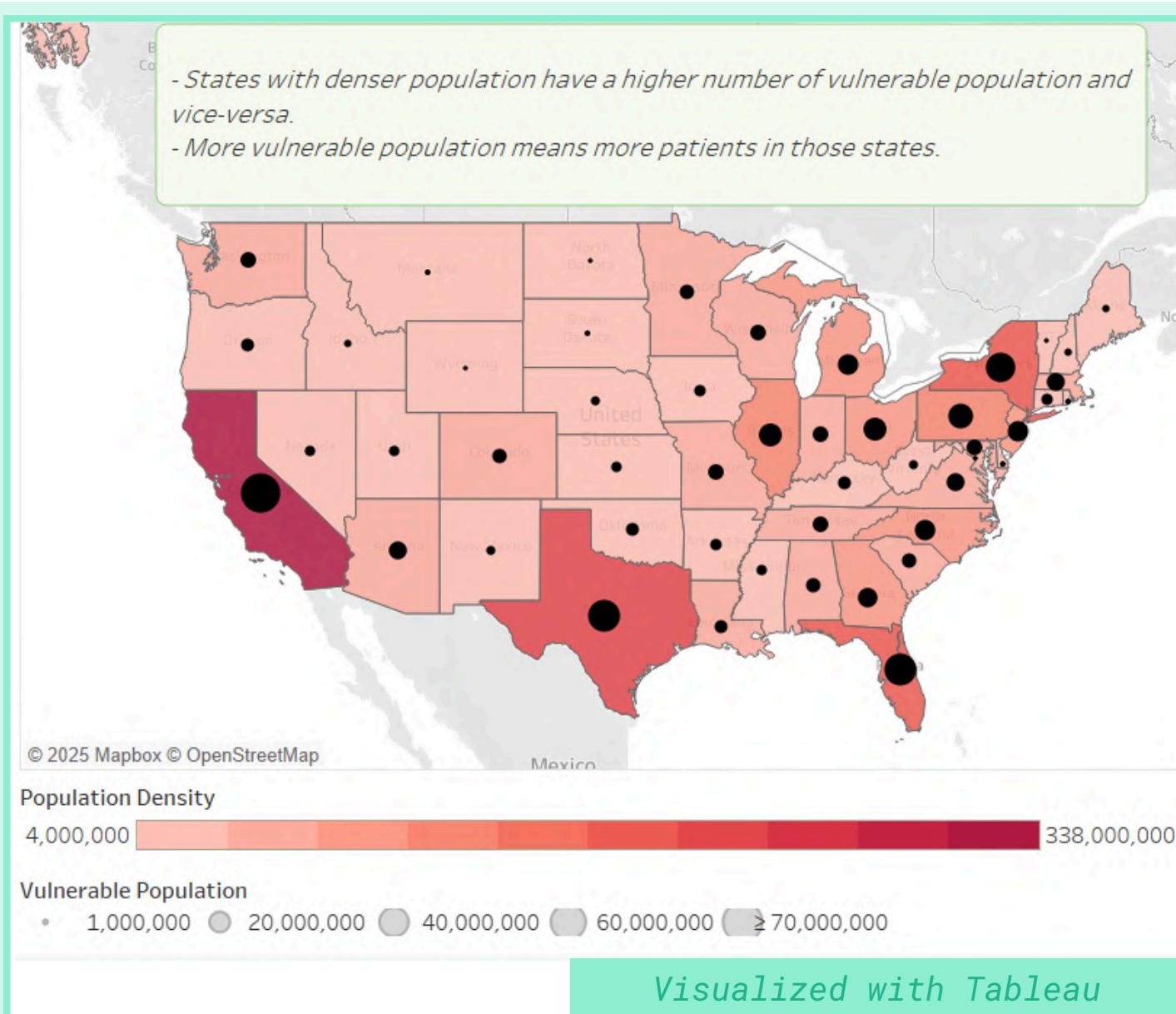
Analysis Insights: Flu Deaths VS. Population



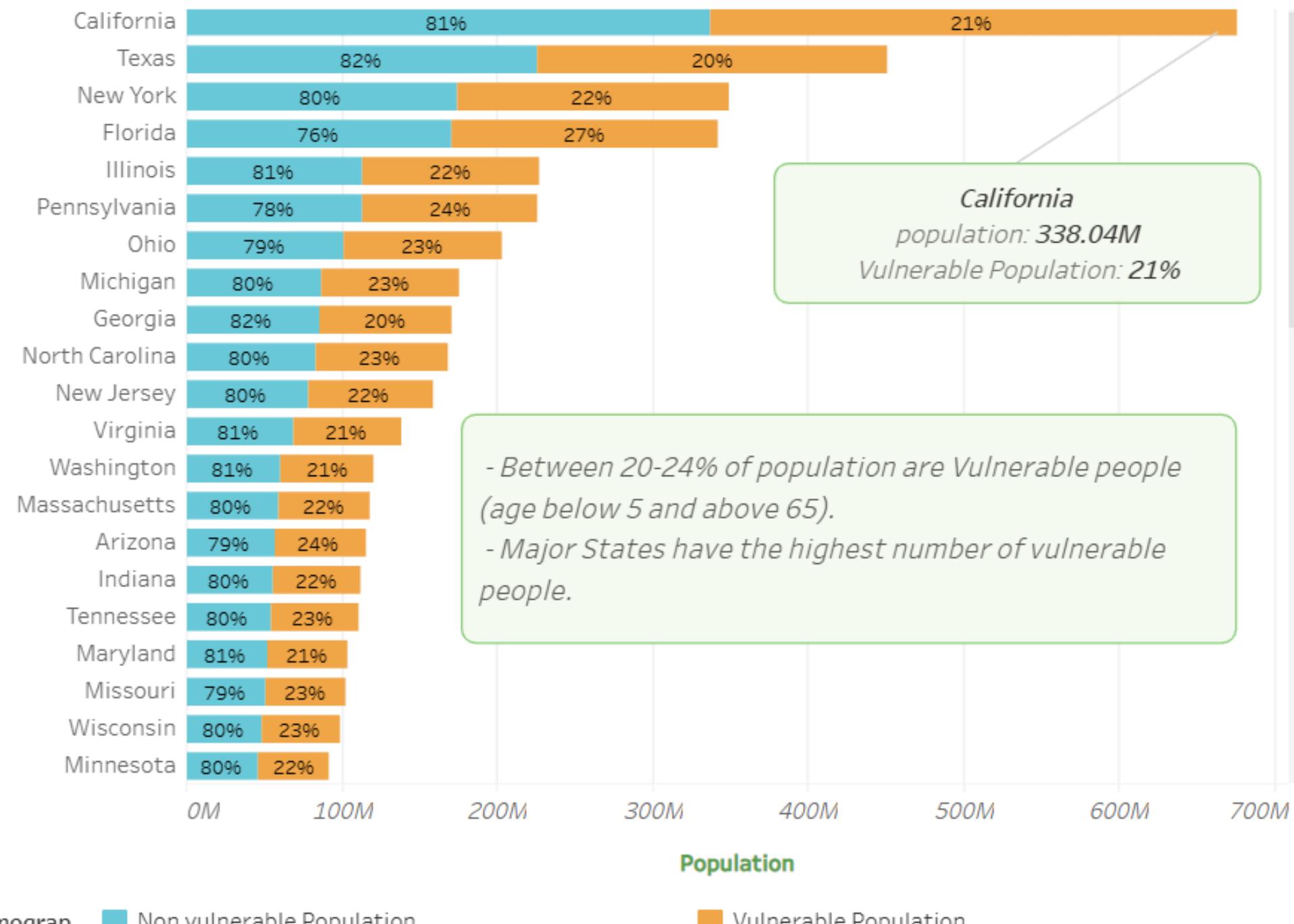
- After cleaning up the data and visualizing the relationship between Flu Deaths and Population, There is a clear positive relationship between them. However, the ratio of deaths are bigger in some states than others.

Analysis Insights: Vulnerable Population Distribution

- This chart visualized using Tableau shows the number and the percentage of vulnerable population in each state.
- Although some States with higher population have higher vulnerable population, the ratio remains around the same.

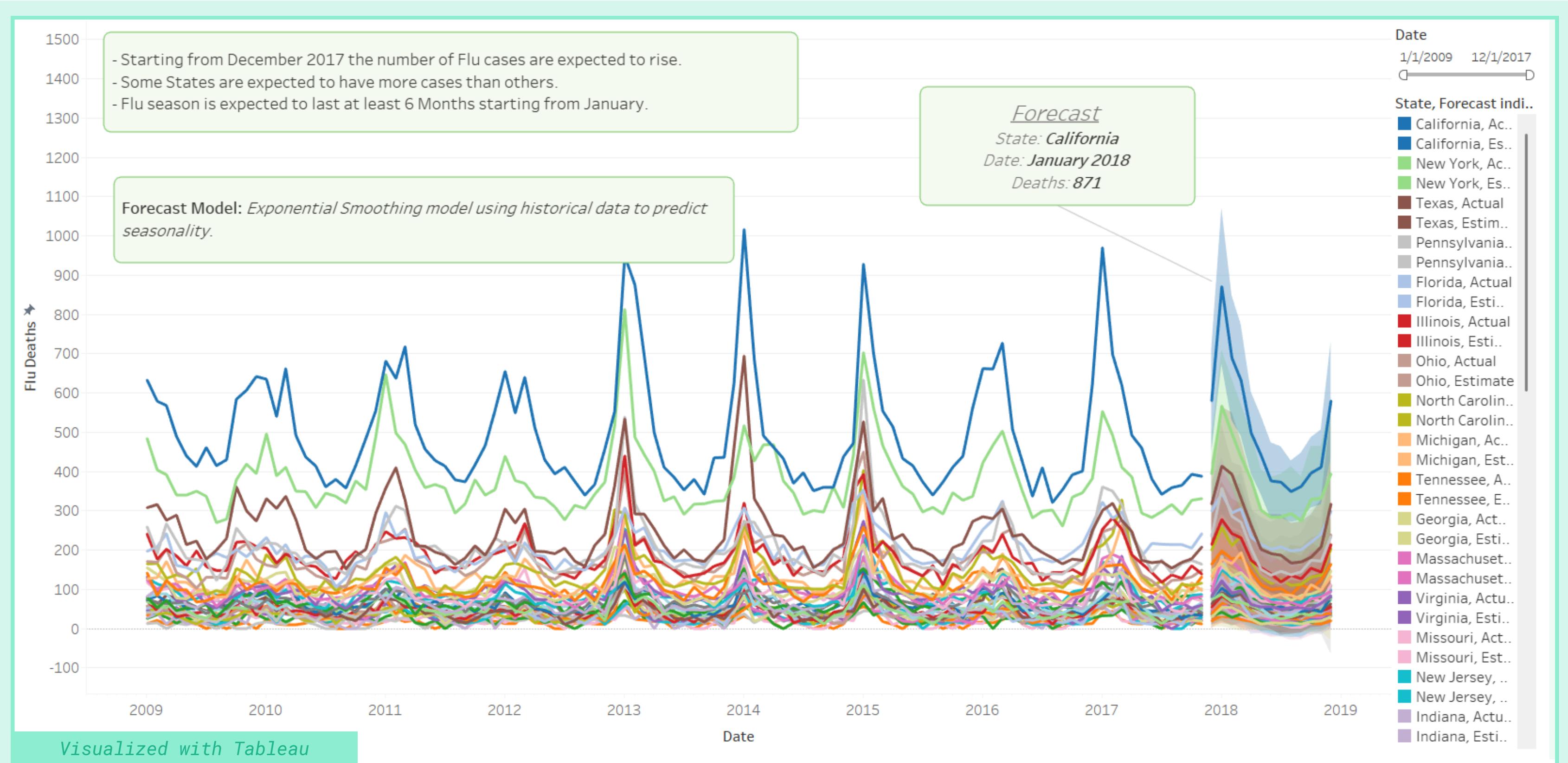


Vulnerable and Non-Vulnerable Population in Each State



- The heatmap shows the density of population in each state along with Vulnerable vulnerable population.

Analysis Insights: Flu Season Forecast



- The chart above shows the Flu Season prediction for 2018 using **Exponential Smoothing Model** in Tableau.
- According to the forecast, Flu Season is to start at the Start of first quarter of 2018 based on historical data.

Insights / Recommendations / Next Steps

Recommendations

- Staffing for different hospitals and clinics should start before January 2018.
- Staffing needs should be according to the number of population taking into account the fatality rate in each State and the vulnerable population numbers.

Next Steps

- Find out the number of available staff that are ready to be sent to the hospitals for the flu season and decide how many to send where.
- During the flu season, survey patients who are treated in the hospital to see the effectiveness of the project. furthermore, add complaint boxes on the hospitals to gather indirect feedback.
- After the flu season has ended, prepare the final analysis of the surveys to conclude the effectiveness of the project.

LINKS:



[Tableau Dashboard](#)



[Influenza Deaths by Geography](#)



[Population Data by Geography](#)



[Counts of Flu lab tests](#)



[Survey of Flu shot rates in Children](#)



[Video Presentation](#)



Source: [CDC](#)



Source: US Census Bureau



Source: [CDC](#)



Source: [CDC](#)

2

Rockbuster Stealth

Context and Project Objective

- **Context:** Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.
- **Objective:** Help with the launch strategy for the new online video service.



Data Sets

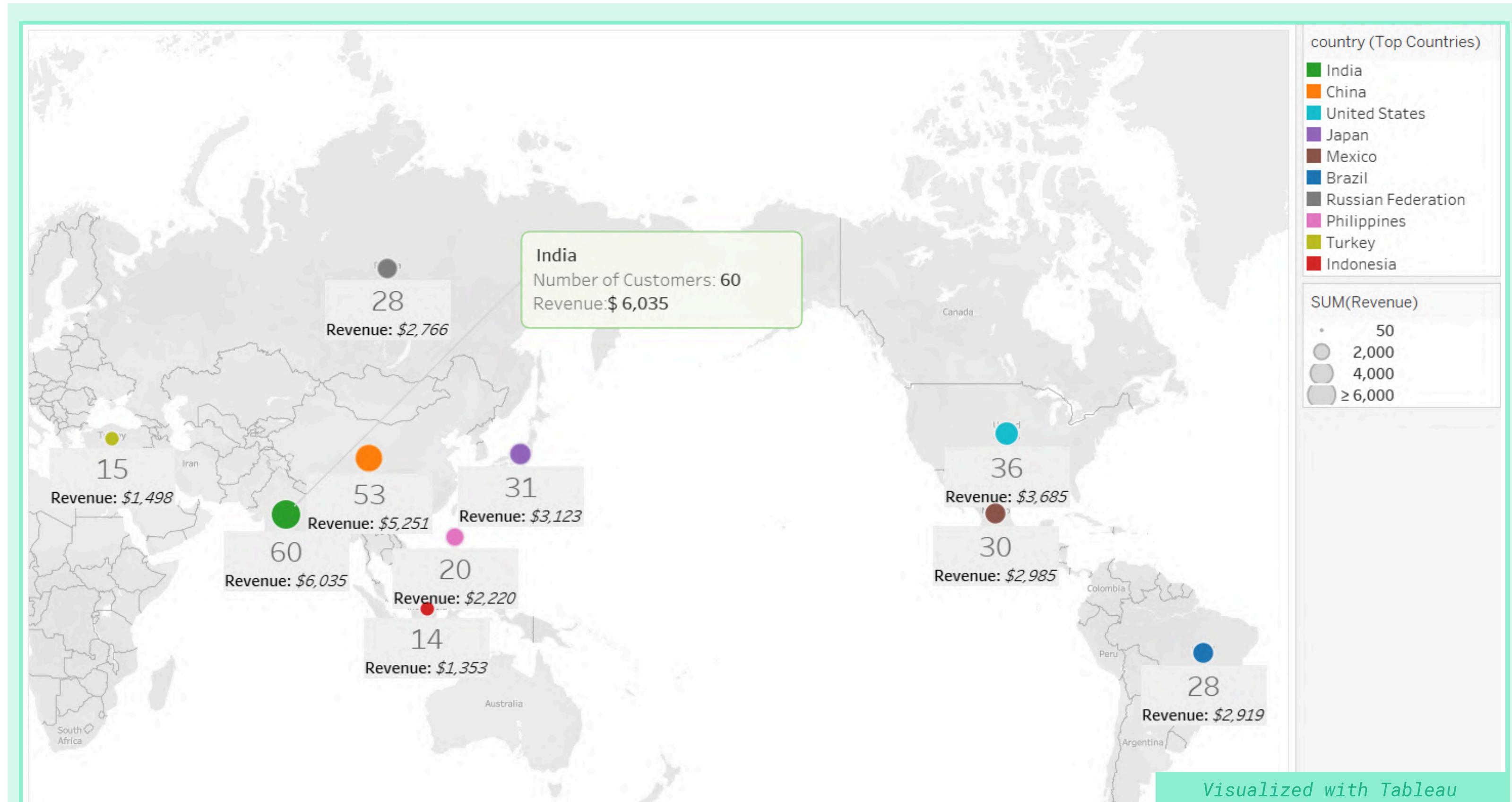
- Data Set includes information about Rockbuster's film inventory, customers, and payments, among other things.



Data Analysis Methods Used

- Data Cleaning
- Writing moderately complex SQL queries to answer business questions:
 - Creating CTEs, table joins, and subqueries.
- Descriptive Analysis using SQL.
- ERD Creation
- Data Dictionary Creation.
- Creating visualizations in Excel and Tableau.

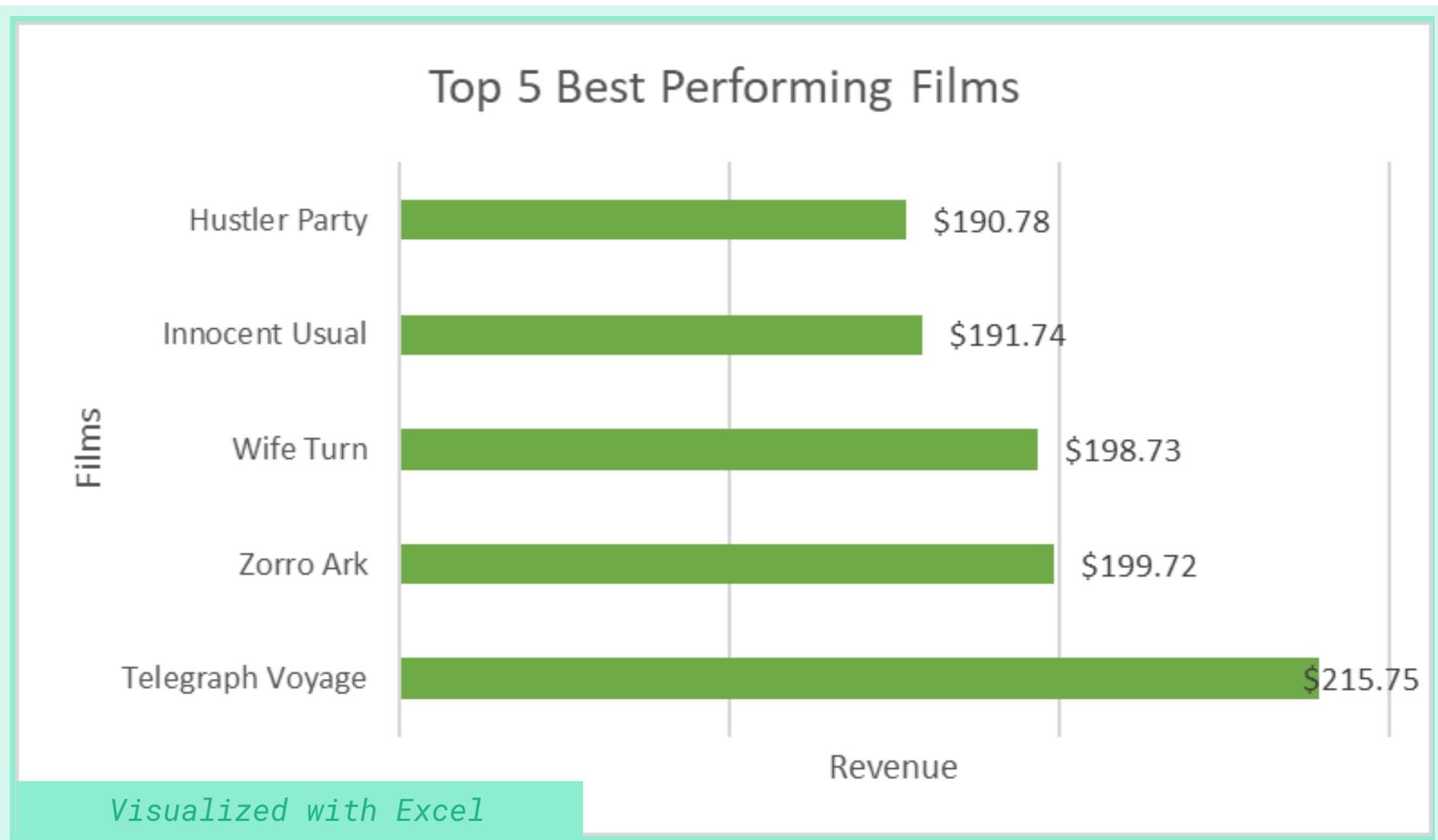
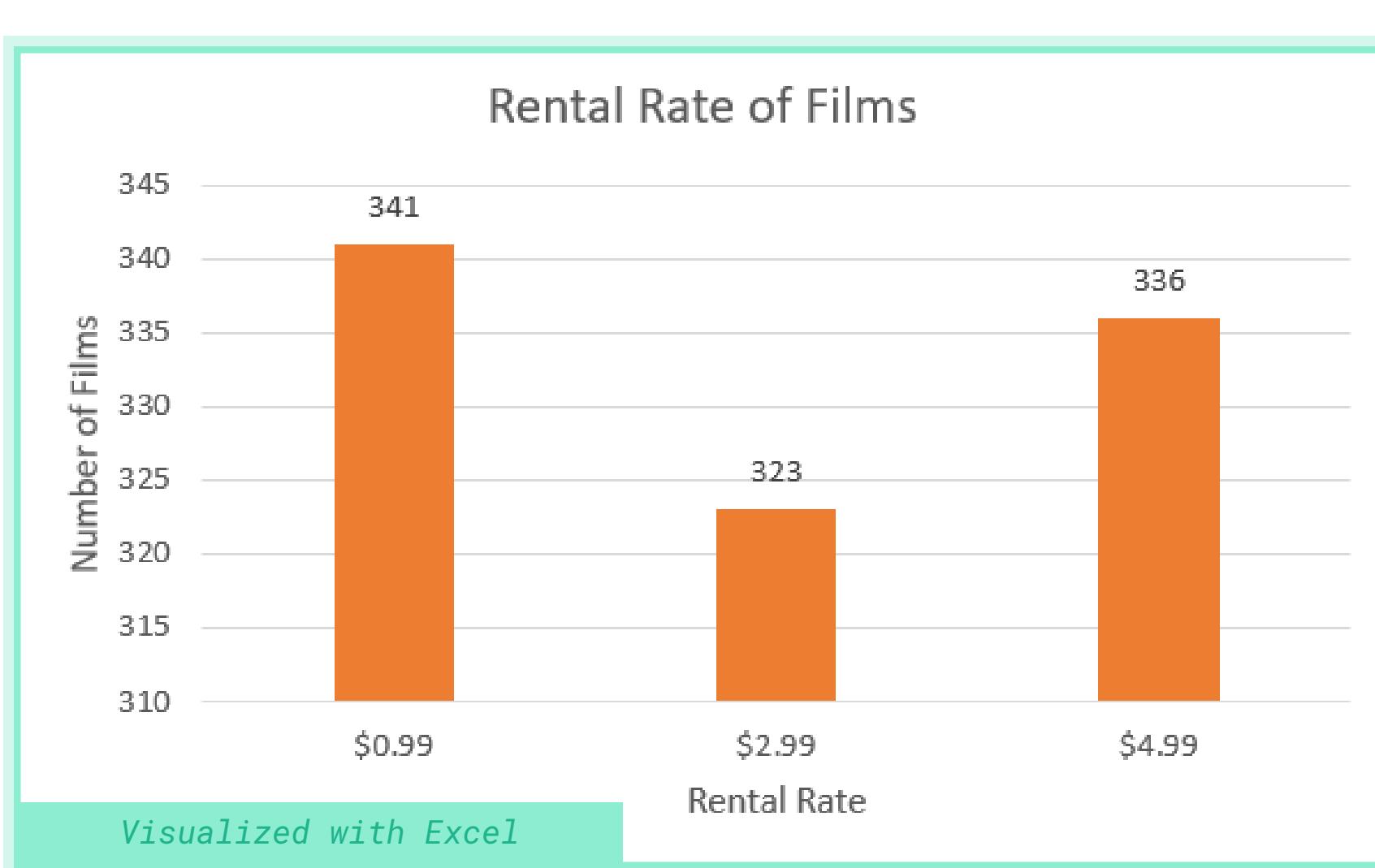
Analysis Insights: Where are Top customers based?



- This visualization shows the number of Customers and revenues brought by Top 10 countries.

Analysis Insights: Best Movies & Rental Rates

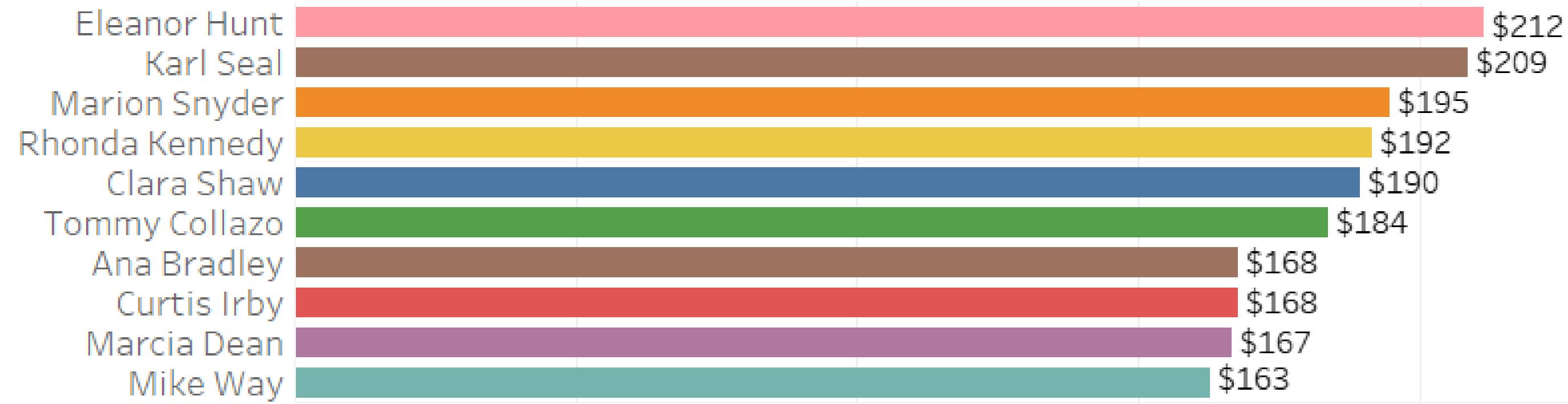
- The Top 5 films revenue range from \$190 to \$215 with "Telegraph Voyage" on the lead



- Majority of the films Rockbuster provides are affordable.
- The number of films that are expensive are the second highest.

Visualized with Tableau

Top 10 Customers



Country

United States	Brazil	India	Netherlands	Runion
Belarus	Canada	Iran	Philippines	

- Customers that have paid the most come from different countries.
- TOP paying customers also come from countries with less amount of customers.

Insights / Recommendations / Next Steps

Recommendations

- Focus marketing effort on Top 5-10 countries as they bring the most revenue Majority of the Top countries are rich countries.
- To get more customers, decrease the renting price and replacement based on the country. eg. discord's Nitro subscription.
- Make the replacement cost cheaper and in turn increase the rental price to increase customer loyalty.

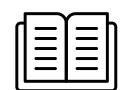
Next Steps

- Develop targeted marketing campaigns. for Top customers and Top regions with the marketing team.
- Explore partnerships and licensing deals in highest traffic locations.

LINKS:



[Tableau Dashboard](#)



[Data Dictionary](#)



[Github](#)



[Rockbuster Data Set](#)

3



Instacart Grocery Basket



Project Objective

- Uncover insights about sales patterns and suggest strategies for better Segmentation.



Data Sets

- Open-Source dataset by Instacart which includes data on:
 - Customers (206,209 Rows)
 - Orders (3,421,083 Rows)
 - Orders Products Prio (32,434,489)
 - Products (49,693)



Data Analysis Methods Used

- Data cleaning using Python (*Jupyter notebooks*).
- Data Analysis using **Python** and relevant libraries (*pandas, NumPy, os, matplotlib, scipy, and seaborn*).
- All of the data sets merged into **A single data set**.
- Conducting descriptive analysis after importing of datasets.
- Deriving new columns relevant to the analysis. (*age group, income class, loyalty flag*)
- Visualizing important data using python.

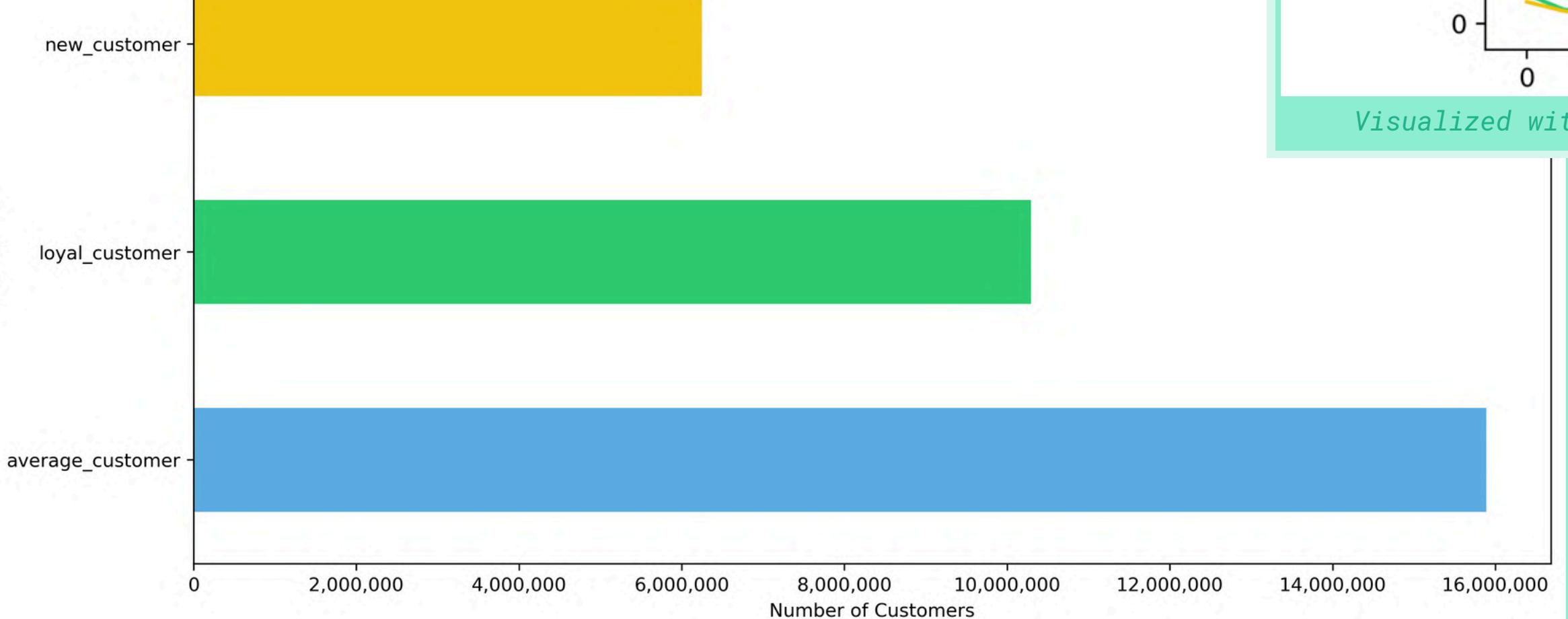
Analysis Insights: Customer Segments & Behavior

- The line chart shows the behavior of different loyalty segments at different hours of the day.
- Loyal and average customers seem to be more active during the early hours of the day.
- It is also evident that the busiest hours of the day remain between 9 AM to 2 PM.

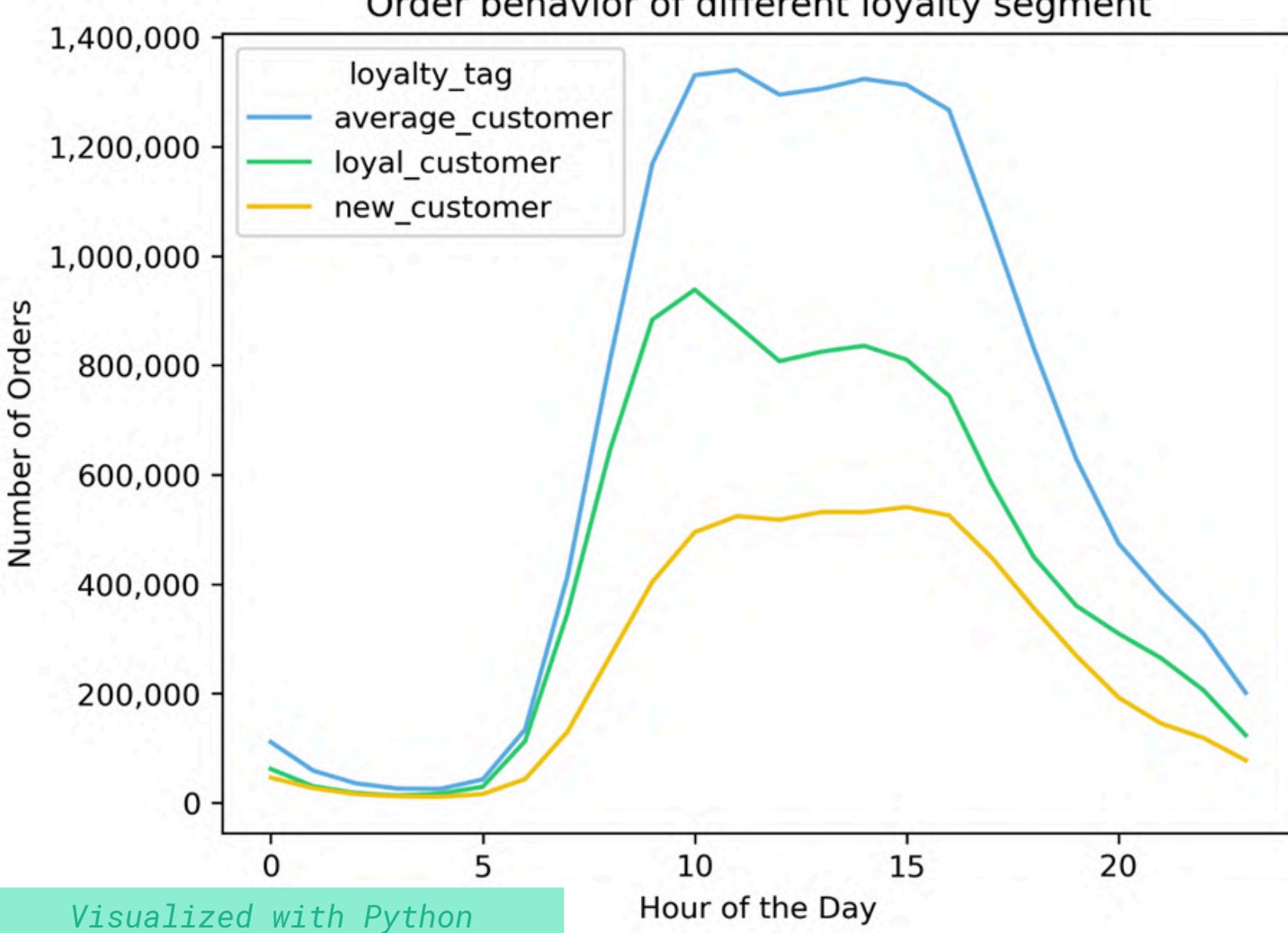
Visualized with Python

Customer Loyalty chart

Loyalty Tag



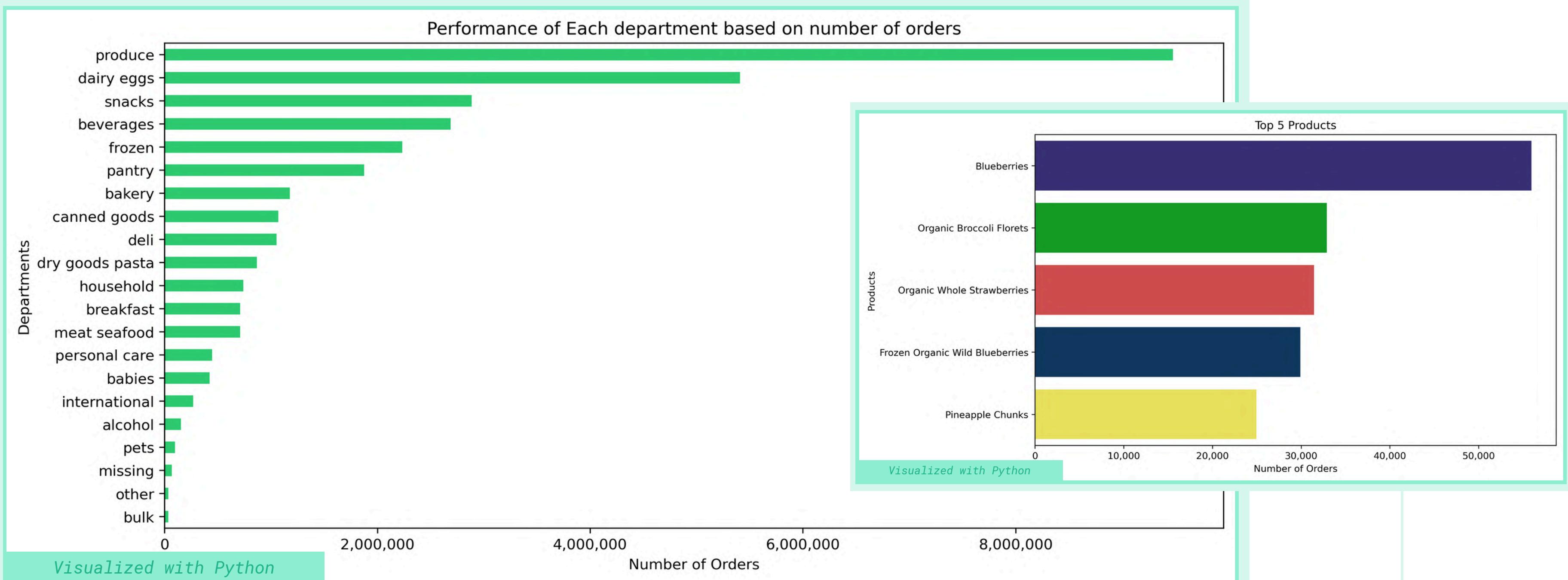
Order behavior of different loyalty segment



Visualized with Python

- The bar chart shows the number of existing customers based on their loyalty.
- The number of Average customers are considerably higher than loyal customers.

Analysis Insights: Best performing departments & Products

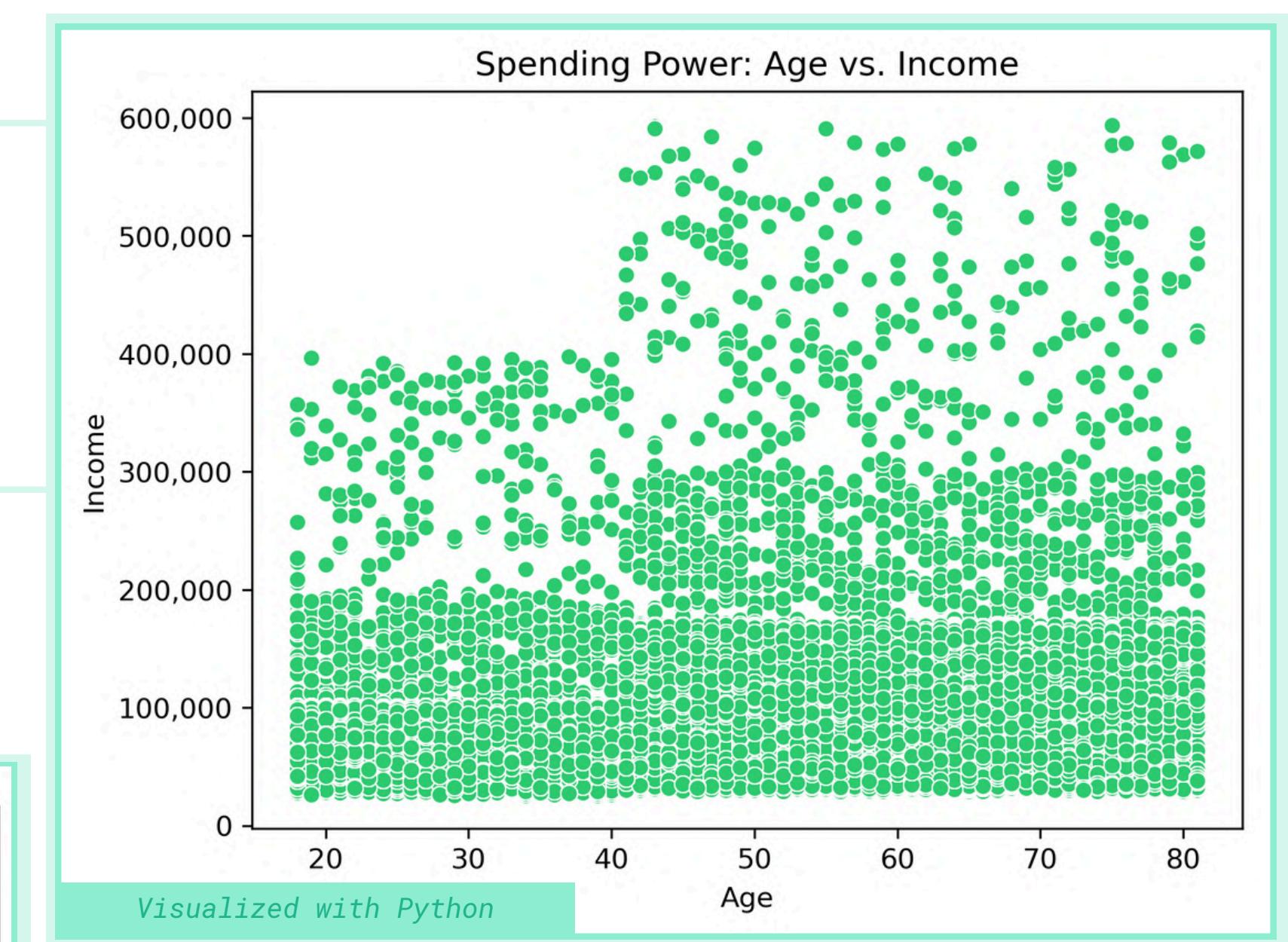
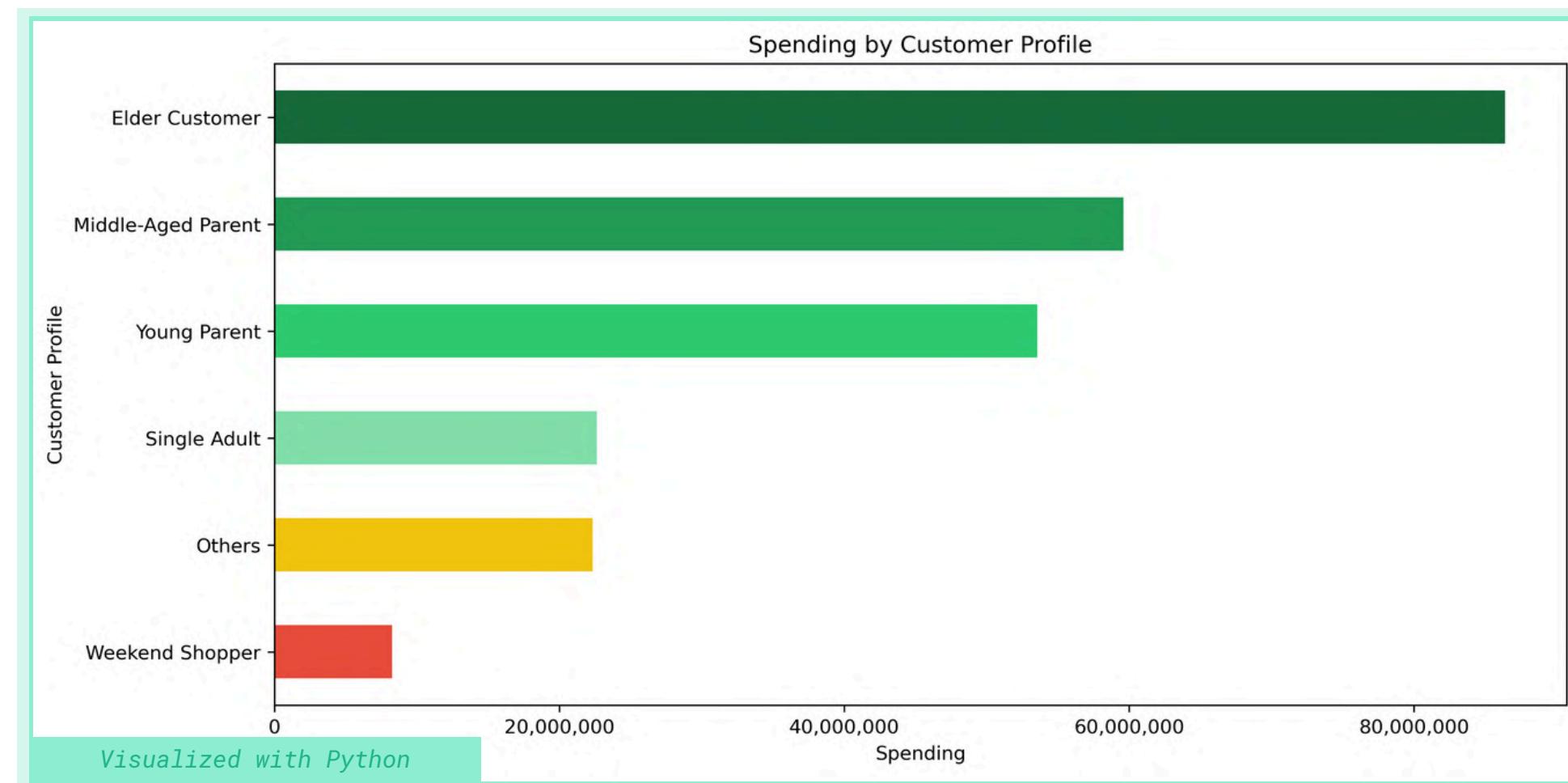


- The chart shows the best performing departments based on the number of orders.
- Produce and dairy eggs are among the top 2 departments dominating the charts.

- The Top 5 products chart shows the most popular products between customers.

Analysis Insights: Spending Power - Age & Marital status

- The scatterplot shows that there is a slightly positive relation between the age and spending power.
- The spending power of a proportion of the customers increase after the age of 40.



- The spending behavior across different demographics vary greatly. Elder customers over 60 take up a decent portion of the spending.
- Couples spend way more than Single Adults.

Recommendations / Next Steps

Recommendations

- focusing ads between 9 AM to 2 PM for optimal results.
- Produce, Dairy Eggs and snacks are among the top product departments with the majority of the sales. Have sufficient stock of products in those departments.
- A decent percentage of customer are new customers so more attention should be given to building customer loyalty.

Next Steps

- Work with the marketing department to tailor marketing campaigns toward different segments of customers.
- Collect data on what products are ordered together.
- Get feedback from stakeholders.

LINKS:



[Data Dictionary](#)



[Github](#)



[Customer Data Set](#)



[Orders Products Prior](#)



[Orders and Products Data Set](#)



[Department Data Set](#)

Source: ["The Instacart Online Grocery Shopping Dataset 2017"](#) Accessed via Kaggle.

4

Pig E. Bank



Project Objective

- Identify risk factors that have contributed to customers leaving the bank.
- Determine the probability of customers leaving the bank.



Data Sets

- Pig E. Bank's client data set (992 Rows)

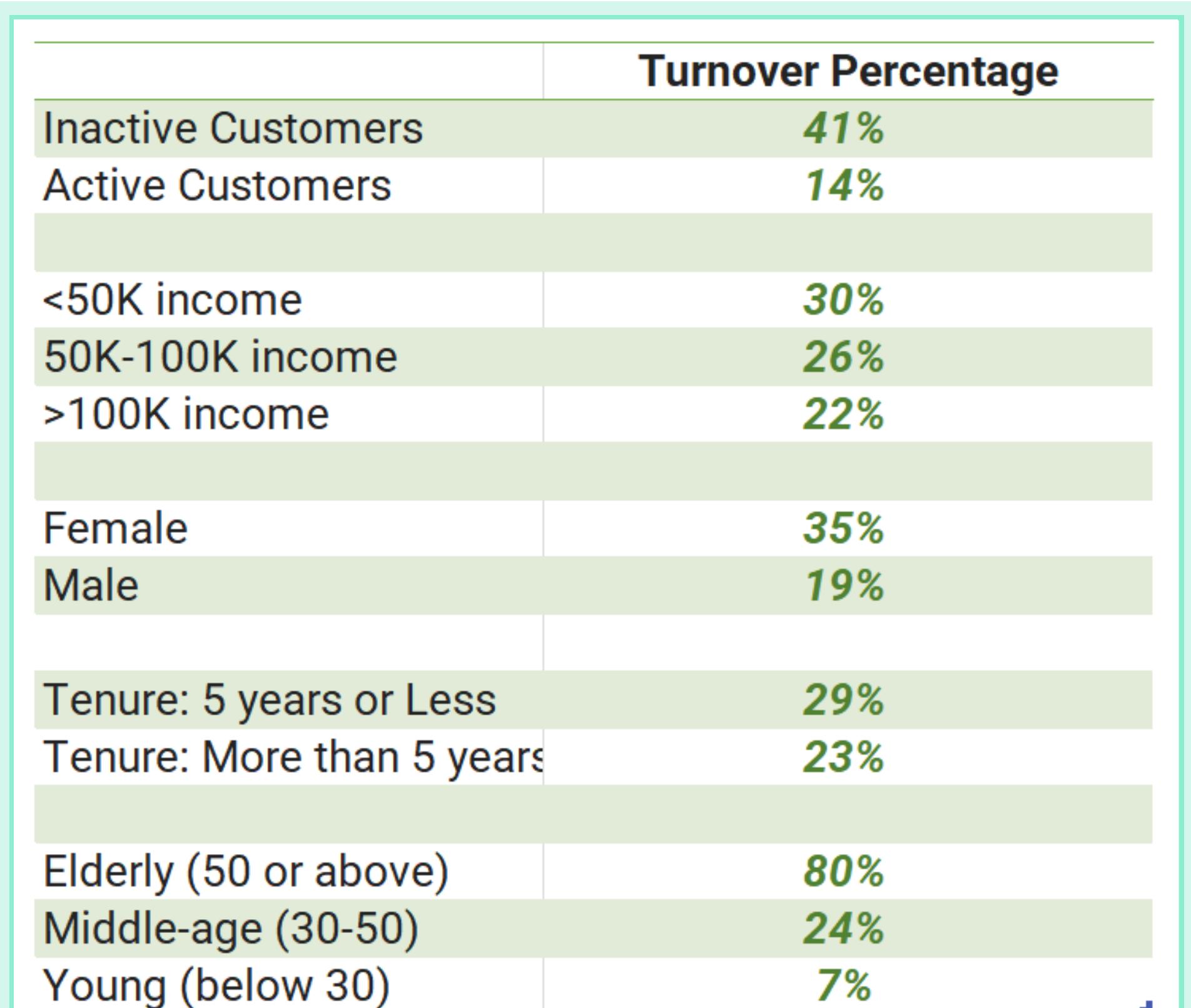


Data Analysis Methods Used

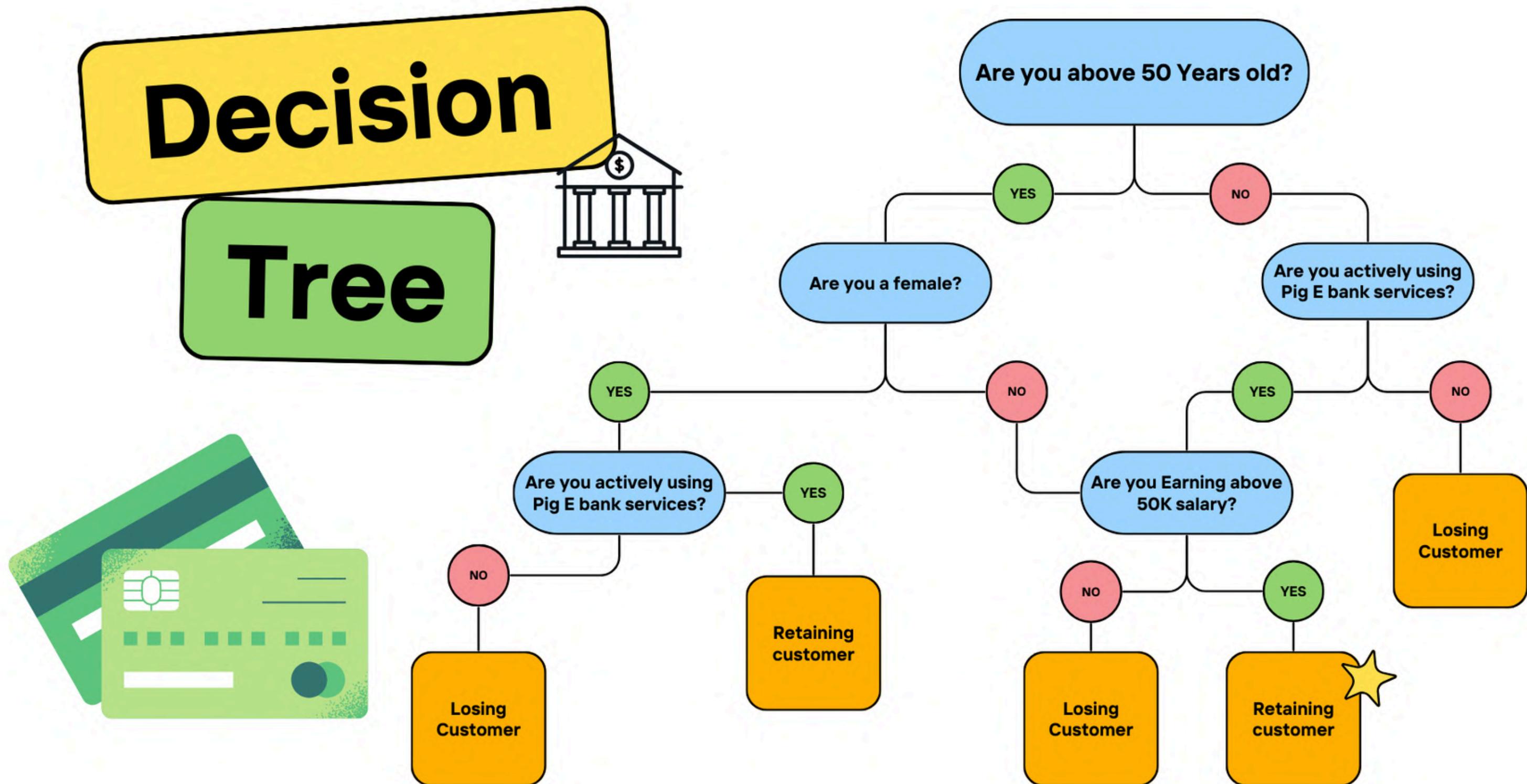
- Data cleaning using Excel
- Conducting Descriptive analysis in Excel.
- Creating Decision-Tree based on the insights uncovered.

Analysis Insights: Descriptive Analysis

- The Description Analysis shows the turnover percentage of different classes of people who are likely to leave the bank.
- Inactive customers have 41% chance of leaving the bank right after people above 50 years old with 80% chance of leaving.



Analyzed with Excel



Recommendations / Next Steps

Recommendations

- Conduct a survey from old people to see why they are leaving.
- Make banking simpler for older people to be able to use it easily.
- Create incentivized engagement for inactive customer to keep them engaged and active. Make the experience fun.
- Conduct a research on why female customers leave more than male customer to find the reason behind it.

Next Steps

- Collaborate with the marketing team to create campaigns for customer retention.
- Gather more info to find the reasons for customers leaving Pig E. Bank.
- Present the results to stakeholders

LINKS:



[Pig E. Bank Customer Data Set](#)



5

Looker E-Commerce



Project Objective

- Uncover sales trends across regions and analyze customer behavior

Data Sets



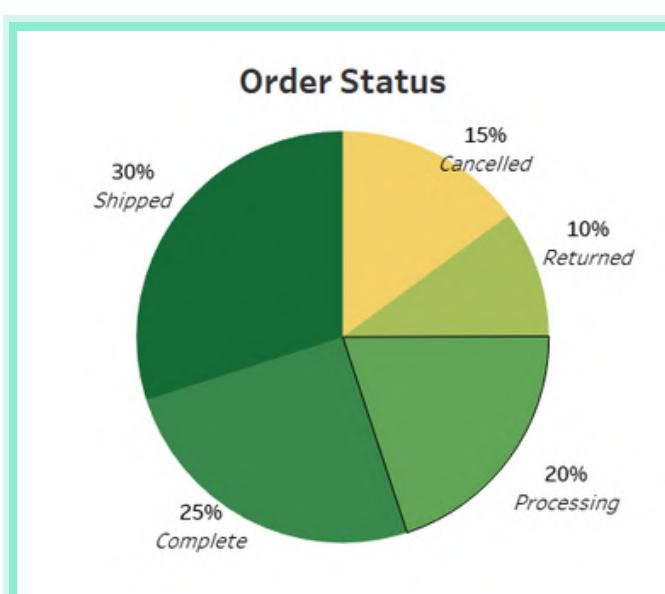
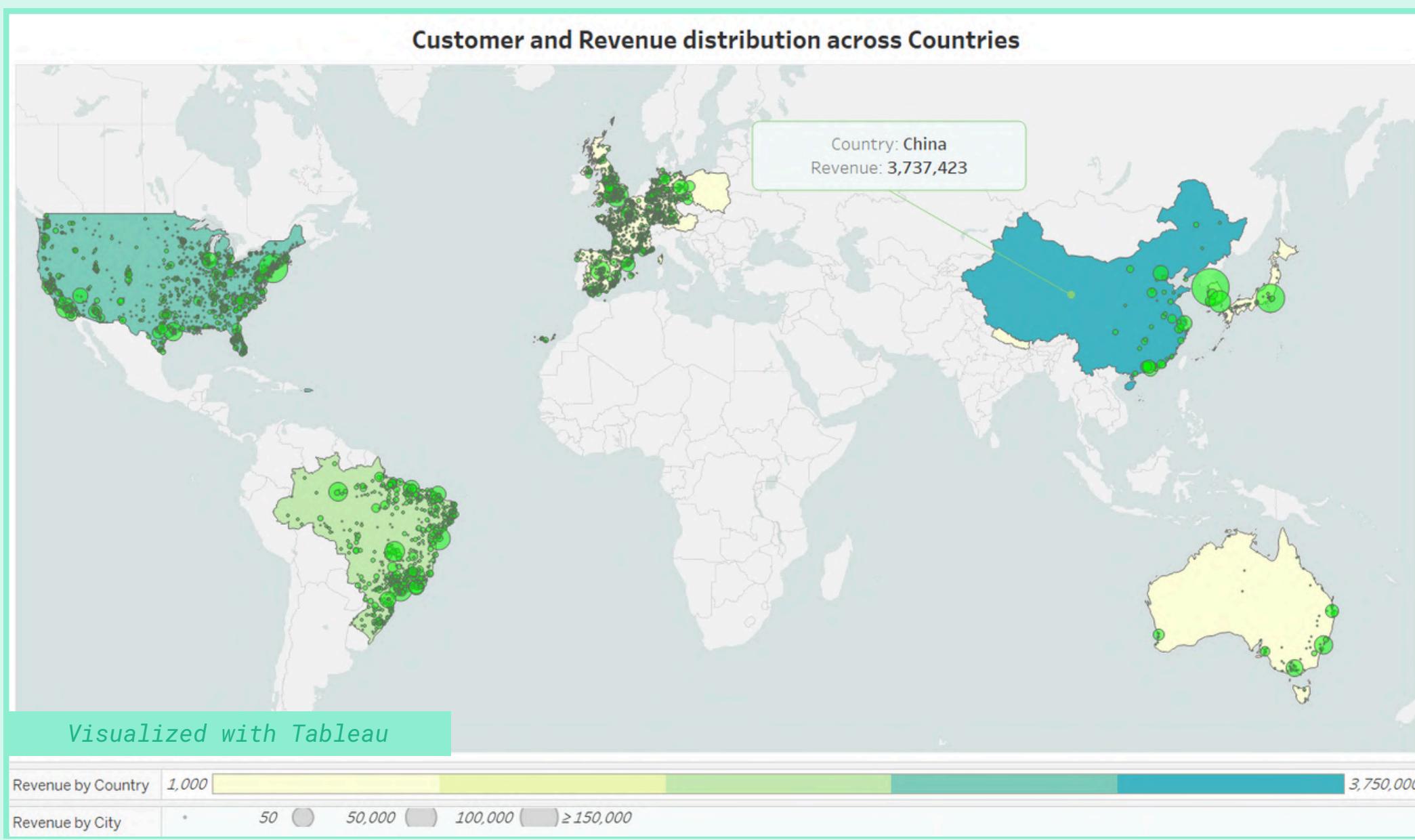
- Open-Source Looker E-Commerce dataset from [Kaggle](#) which includes data on:
 - Orders: 125,226 rows
 - Products: 29,120 rows
 - Users: 100,000 rows
 - Inventory items: 490,705 rows
 - Order items: 181,759 rows
 - Events: 2,431,963 rows
 - Distribution Centers: 10 rows

Data Analysis Methods Used

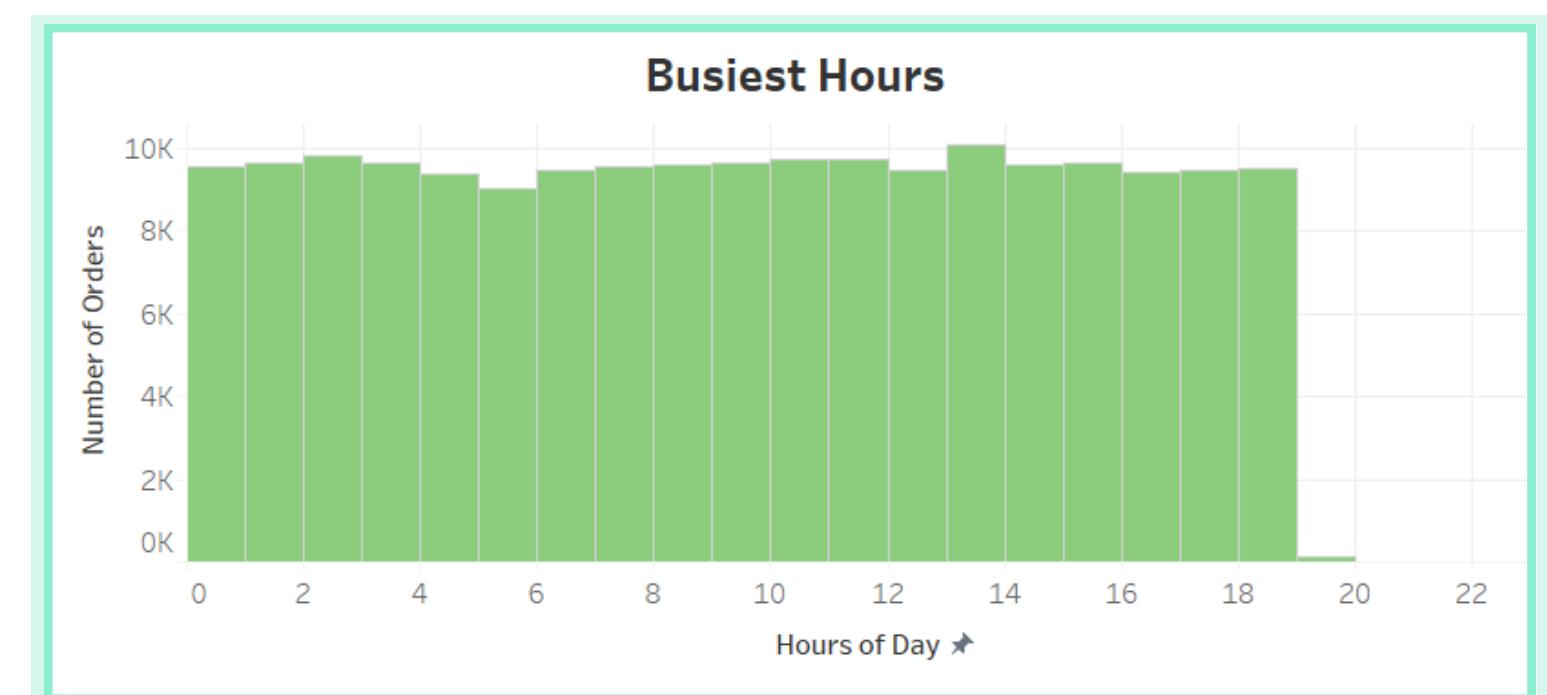
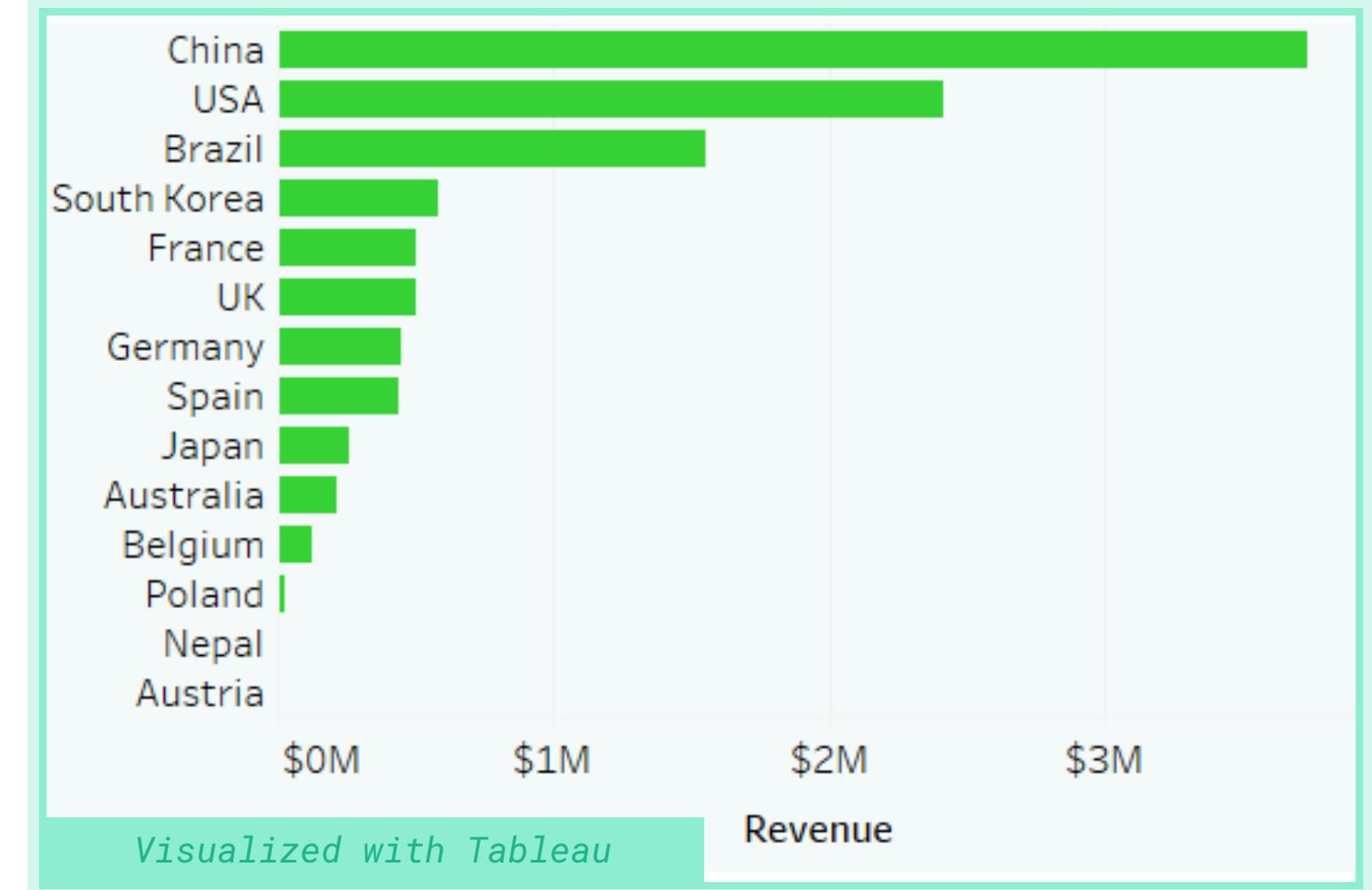


- Data cleaning using Python (*Jupyter notebooks*).
- Relevant data sets merged into **A single data set**.
- Linear Regression in **Python** and relevant libraries (*scikit-learn*).
- Clustering in Python Using k-Means
- Time-Series analysis and Stationarity testing (Dickey-Fuller test)
- Visualizing important data using python and Tableau.

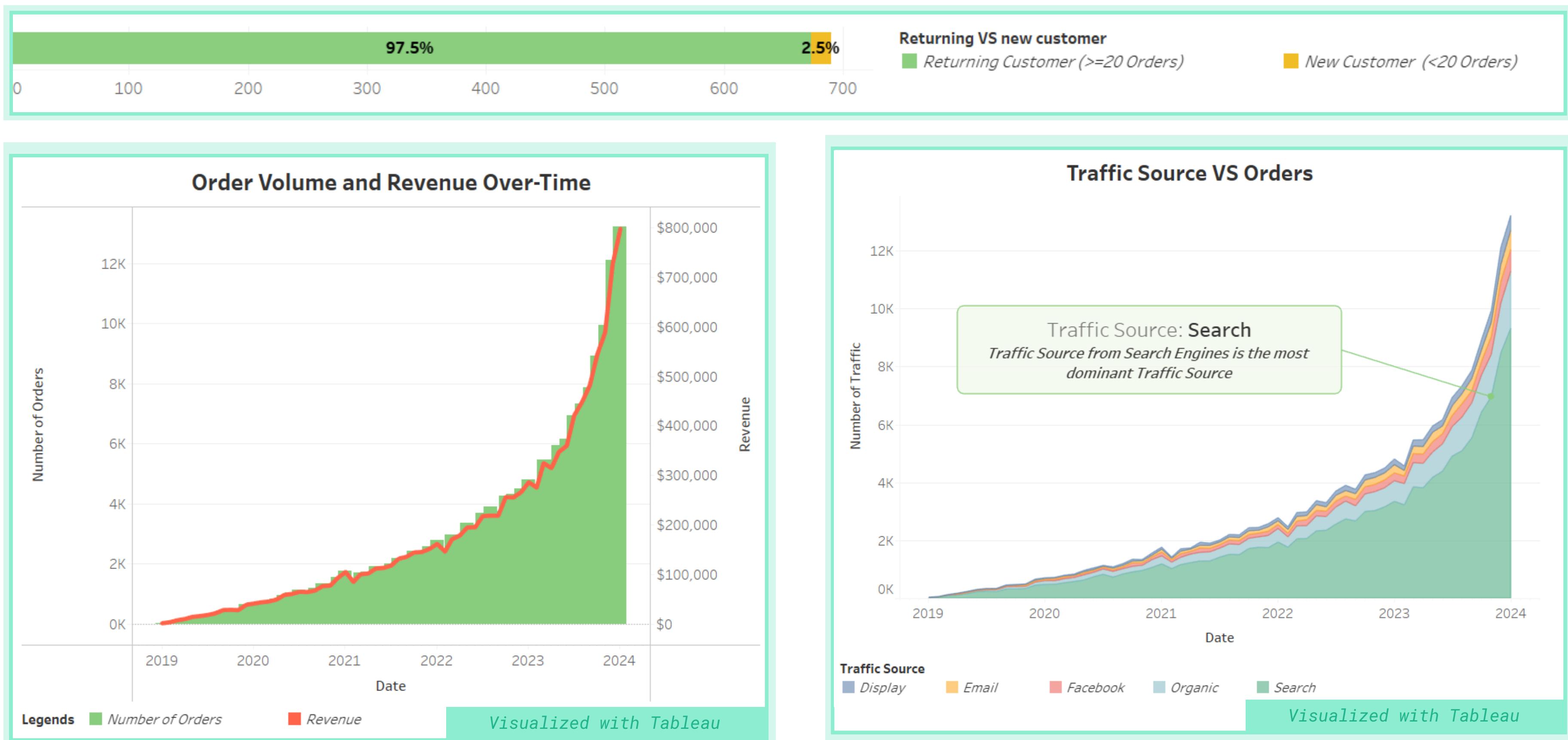
Analysis Insights: Customer Demographics & Revenue



- Top 2 countries with most revenues are China and the USA.
- The busiest hours are between 1AM to 6PM.
- The return rate of products are around 10%.



Analysis Insights: Revenue/Traffic Source/Customers

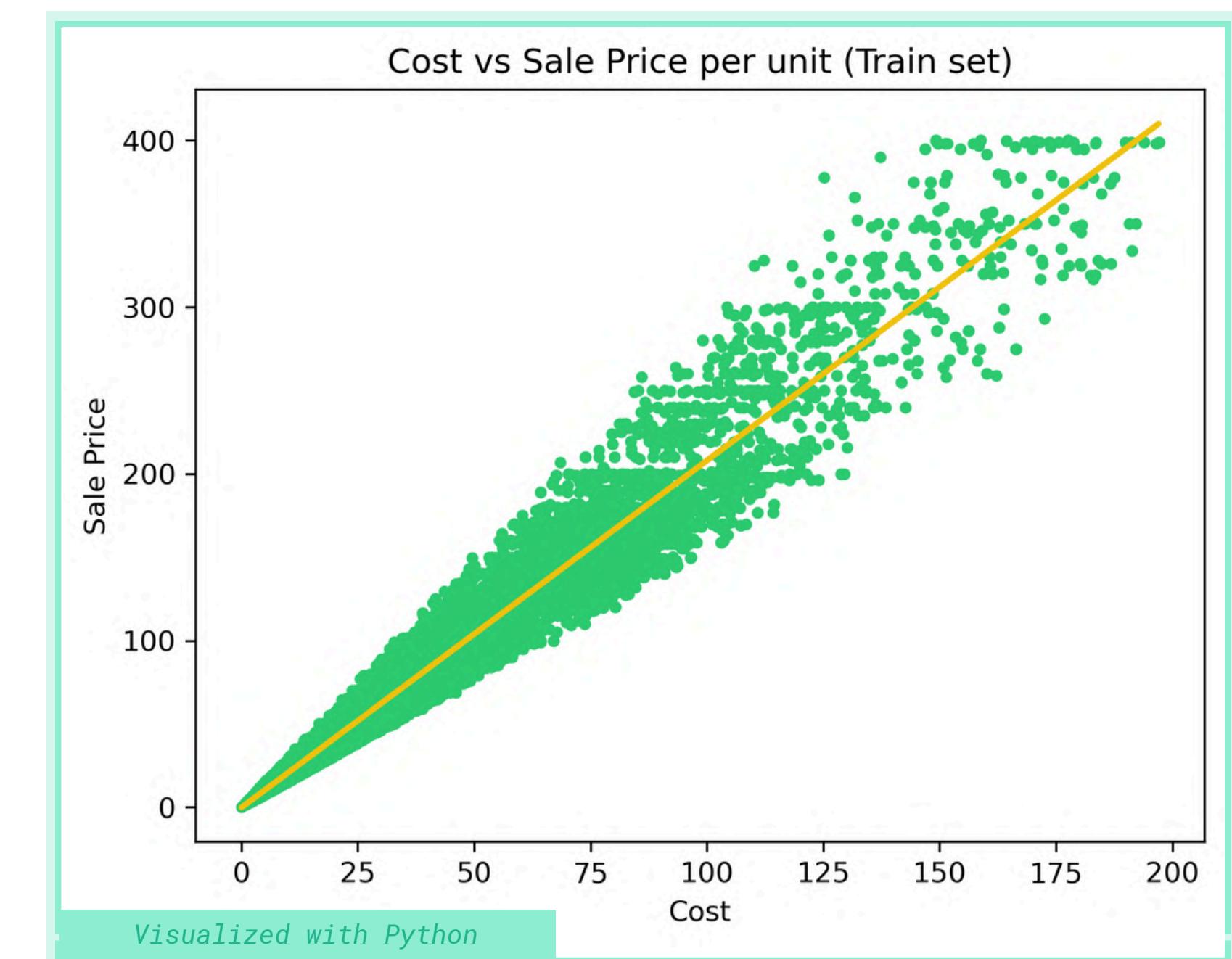
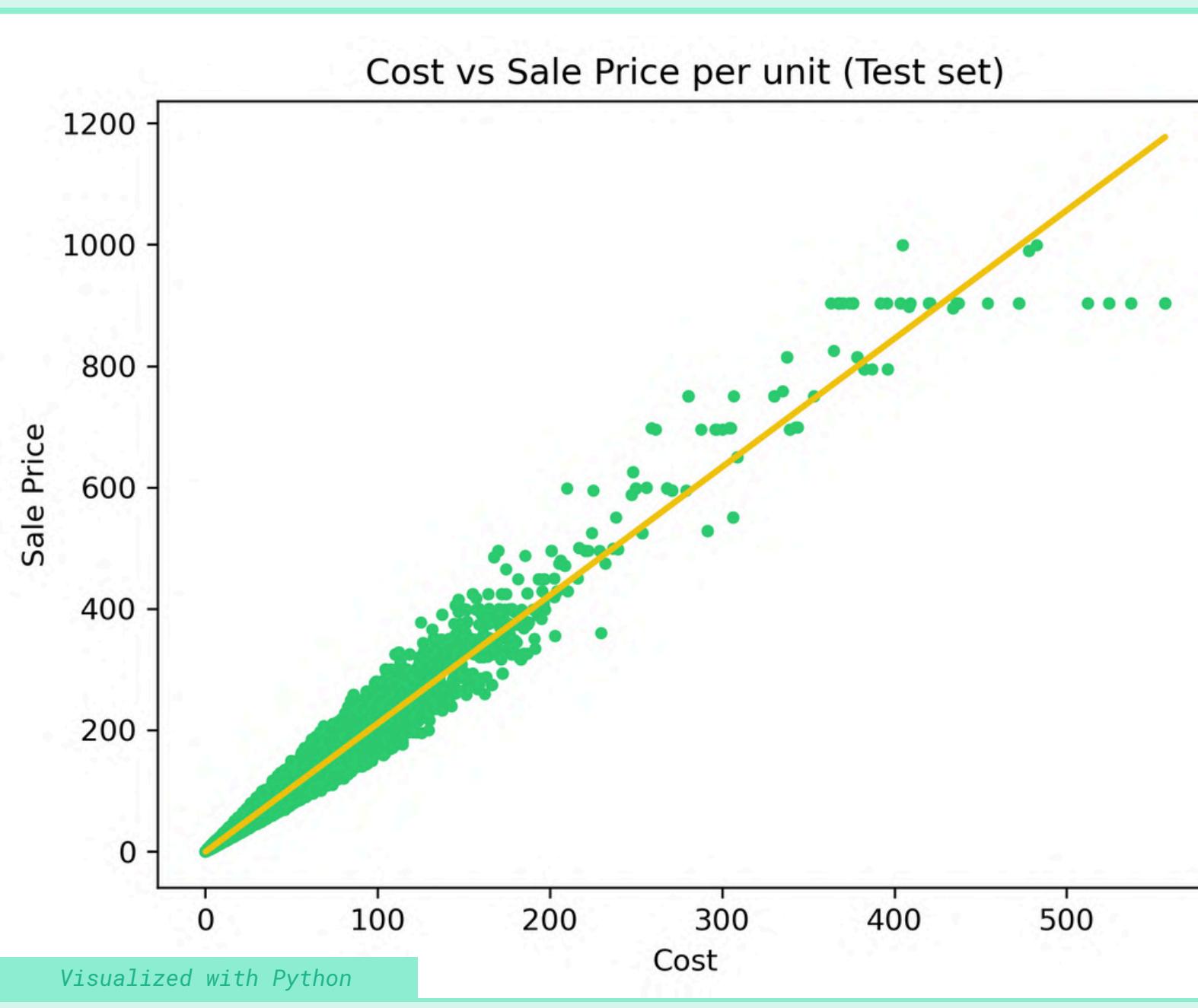


- The order Volume and the Revenue has been on a steady rise ever since launch.

- Majority of the traffic comes from Search Engines.

Analysis Insights: Linear Regression (Python)

- Linear regression on test-set: Clear positive relation between Price and Cost.
- R2: 96.6% Accuracy



- Linear Regression on Training-Set (excluding some outliers): Clear positive relation between Price and Cost.
- R2: 95.7% Accuracy

Recommendations / Next Steps

Recommendations

- Focus marketing campaigns on returning customers, especially those with higher order values.
- Create loyalty programs for existing customers. eg. loyalty credit that accrues after every purchase and can be used in the next order.
- Create personalized product recommendation based on previous purchase history.
- Increase campaigns in top performing countries like China and USA.
- Allocate more budget to SEO and email promotions.

Next Steps

- Build prediction model for product performance.
- Promote best performing products to more customers.
- Use insights to personalize website content and personal recommendations.
- Collect customer reviews for better understanding of customer needs.

LINKS:



[Tableau Dashboard](#)



[Github](#)

Source: [Looker E-Commerce dataset](#) Accessed via Kaggle.

Thank you!