Data Analytics Portfolio

Asawer Maknoon

PROJECTS

1. GameCo

Global market analysis of video game sales

2. Preparing for Influenza Season

Staff deployment planning for influenza season

3. Rockbuster

Launching Rockbuster Stealth online movie service

4. Instacart

Market segmentation analysis to uncover sales

5. Pig E. Bank

Customer churn analysis

6. Divvy's Bikeshare System (2015-2019)

Analyzing Usage Trends

GameCo

Global market analysis of video game sales

Skills applied:

Data Cleaning, Pivot Tables, Descriptive Analysis, Data Grouping and Summarizing, Visualization in Excel and Tableau, Presenting Results.

Objective:

The objective of this presentation is to provide GameCo, a new video game company, with new insights by summarizing findings, trends, and actionable insights for the future market budget.

Data:

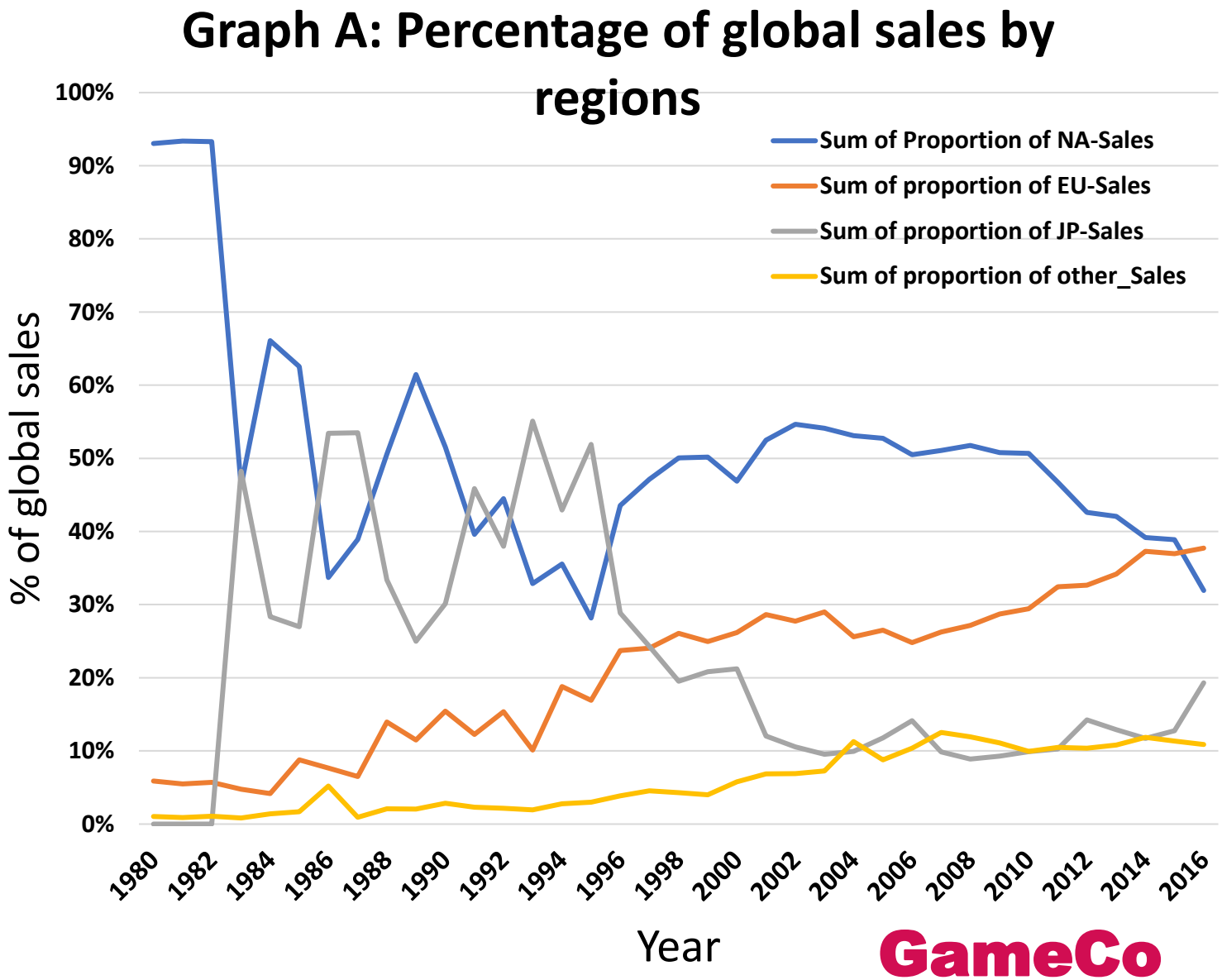
The available data was obtained from VGZChartz (https://www.vgchartz.com/) from 1983-2016.

Generally,
GameCo current understanding is that the sales had the same trend for North America, Europe, Japan and other regions over time.

GameCo

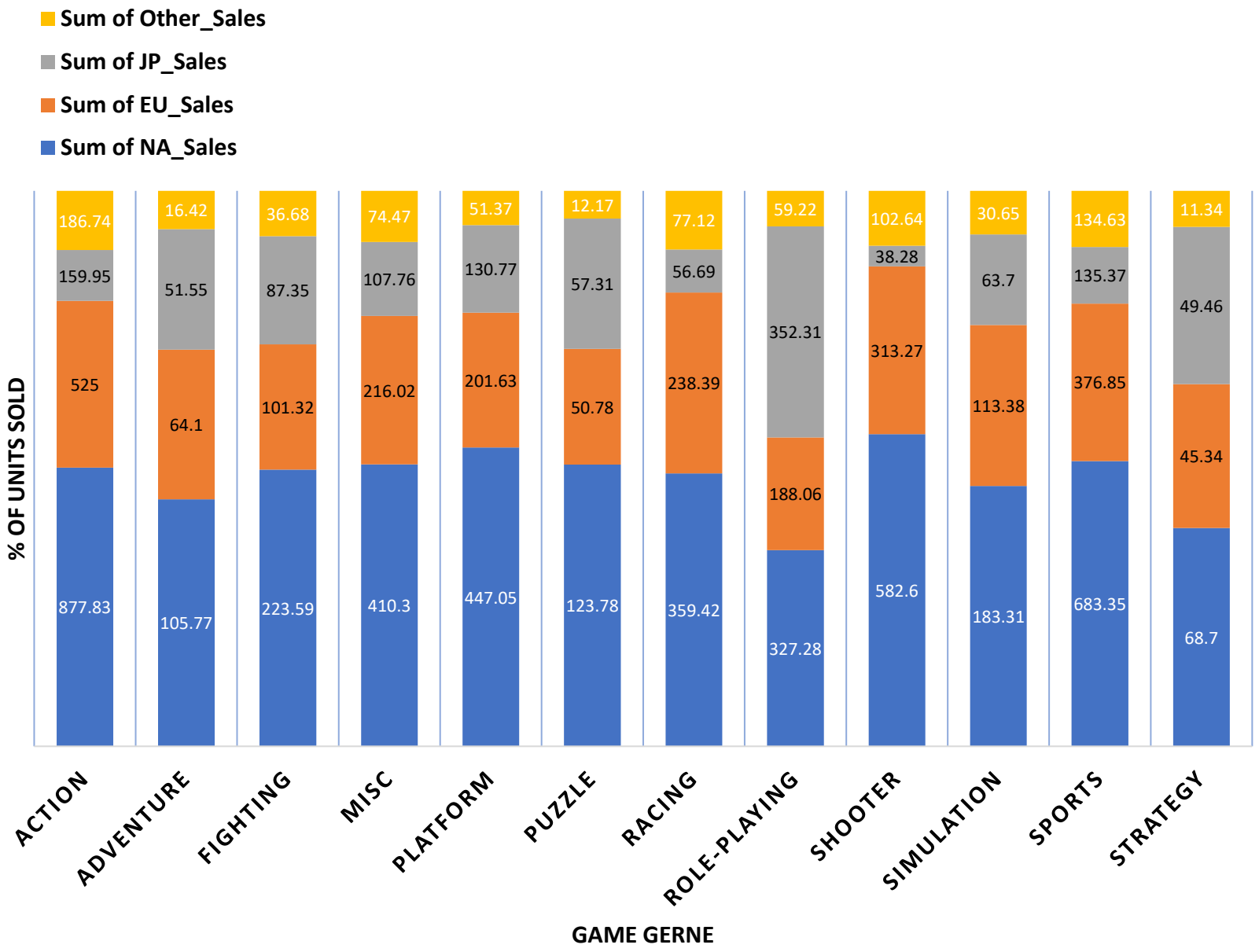
From the chart:

- ✓NA and JP showed fluctuations between 1983-1995.
- ✓NA and JP showed an opposite or a mirror pattern.
- ✓NA-sales dominates the global sales.
- ✓EU and other regions sales increases steadily over years.
- ✓JP showed a dramatically drop in sales in 1996.
- ✓Contrary to expectations, these fluctuations in sale might tell new insights for the future market budget.



GameCo

GRAPH B: GAME GERNES BY REGION



- ✓ The most popular games gernes are action, shooter and sports specially for NA, EU and other regions.
- ✓ In Japan, play-rolling showed the highest sales over years.
- ✓ Almost all regions showed lowest sales for strategy games.

GameCo

Summary

Despite the fluctuations in sales trend for NA between 198 1996, NA-sales dominates game markets compared to th other geographic regions.

The EU showed steadily increasing over time which indicat promising markets for GameCo's future insights.

Other regions sales followed the same pattern of EU ar needs more attention to maintain growing in sales over time.

The most popular game types are action, shooter and sport f NA, EU and other regions.

Japanese most interesting game is play-rolling. Action gam and sport tend also to increase over time.

GameCo

Recommendations

- Localizing game content to include language and cultural references. by doing this the company will ensure the spread of their product globally.
- Developing new genres of games can attract different audiences.
- Investing in campaigns that reach the right audience through channels like social media, influencers.
- Identifying popular gaming platforms in different regions and optimize their games for those platforms.
- releasing regular updates features and new content (seasonal content) can maintain interest and attract more customers.

GameCo

Preparing for Influenza Season

Staff deployment planning for influenza season

Skills applied:

Data Cleaning, Integration and Transformation, Statistical Hypothesis Testing, Visual Analysis Forecasting, Story Telling in Tableau, Presenting Results.

PROJECT OVERVIEW

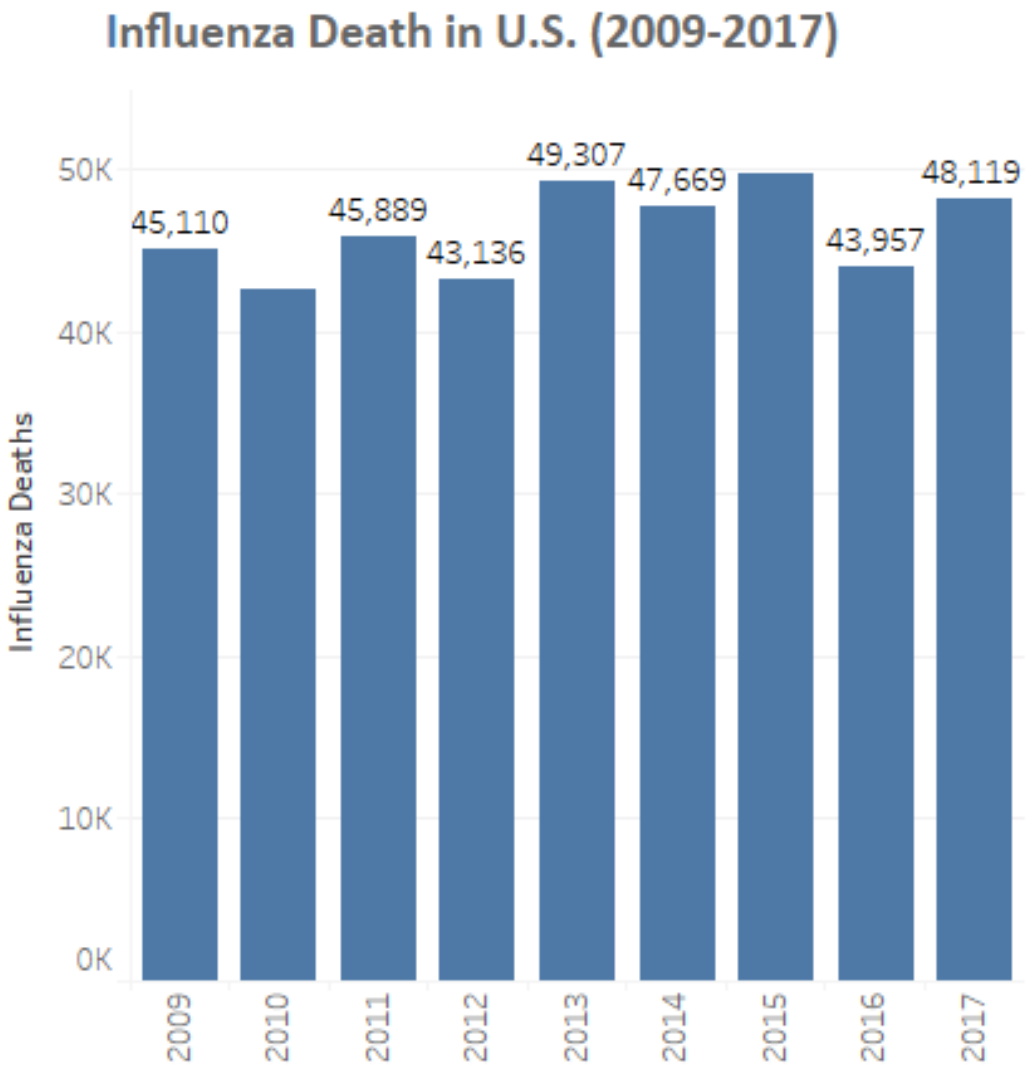
What Is Influenza?

Seasonal influenza is an acute respiratory infection caused by influenza viruses which cause illness that ranges from mild to severe. In the worst cases, influenza can lead to death.

Project Motivation: During influenza season, the increase of demand on additional staff required in most of U.S. hospitals and clinics to deal with the increase of patients, particularly those in vulnerable populations, who suffer from serious complications.

Objective: Determine when to send staff, and how many, to each state.

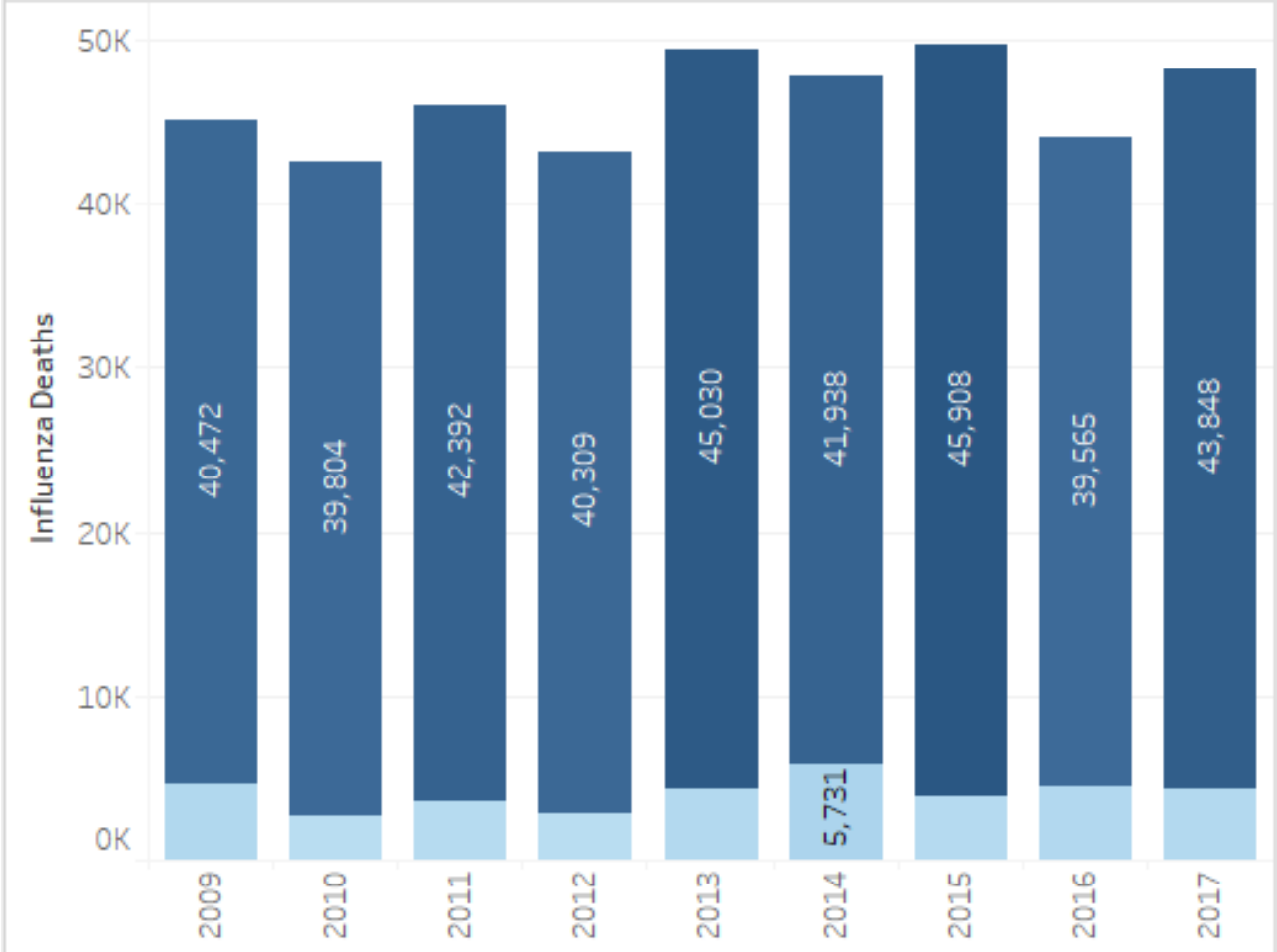
Scope: The project will plan for the upcoming influenza season covering all hospitals in U.S



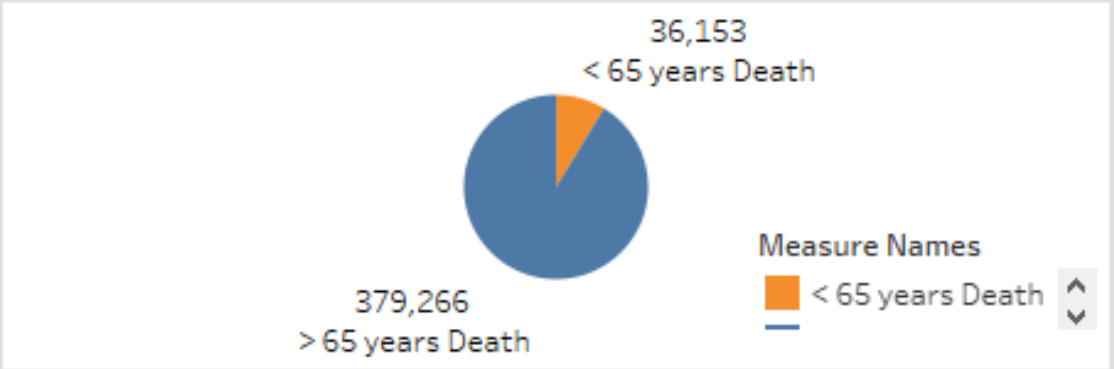
Project Hypothesis

States with higher population of 65+ years age will have higher Influenza deaths compared to states with lower population of 65+ years age.

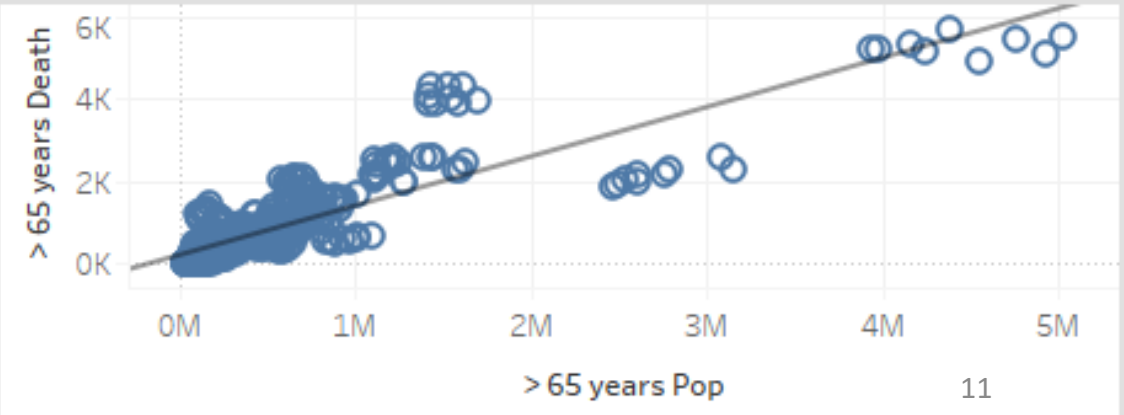
Influenza Death in U.S. (2009-2017) by Age Group



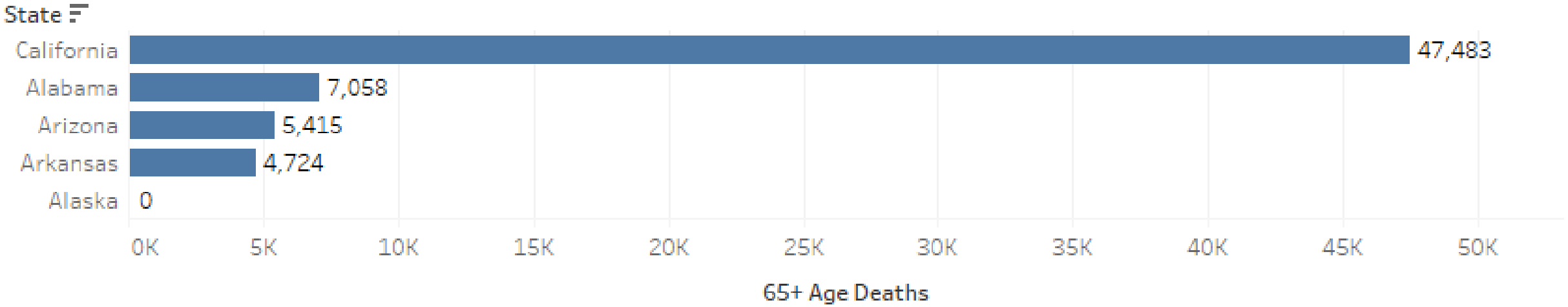
Influenza Death in U.S. by Age Group (2009-2017)



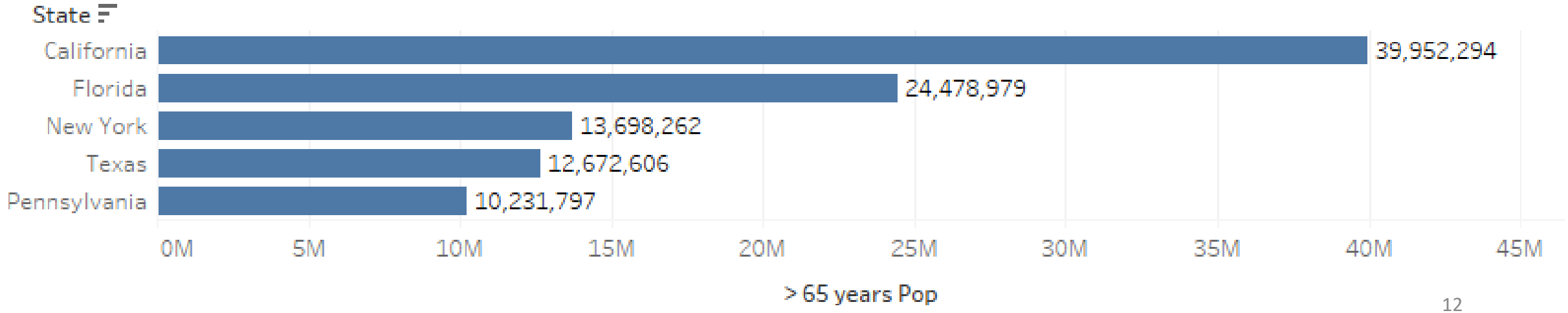
The Relationship between Influenza Deaths of 65+ and Population of the Same Group Age



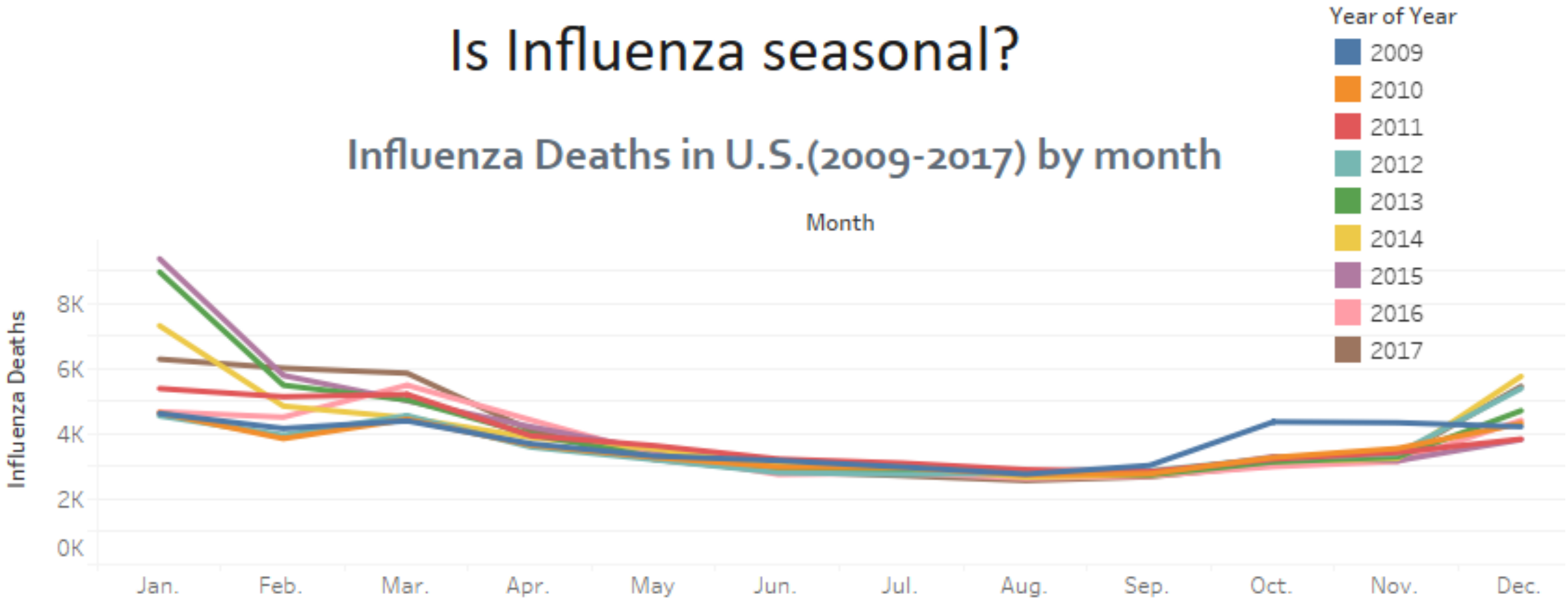
Top 5 Influenza Deaths for 65+ Age in U.S. (2009-2017) by state



Top 5 Population of 65+ in U.S. (2009-2017) by State



Is Influenza seasonal?



Influenza Seasonality Trend

This line chart shows that Influenza has seasonality trend. we can see that Influenza increases from November to March. After March it starts to decrease until August. It can be concluded that Influenza reachs its Peak in December and January.

Conclusion & Recommendations

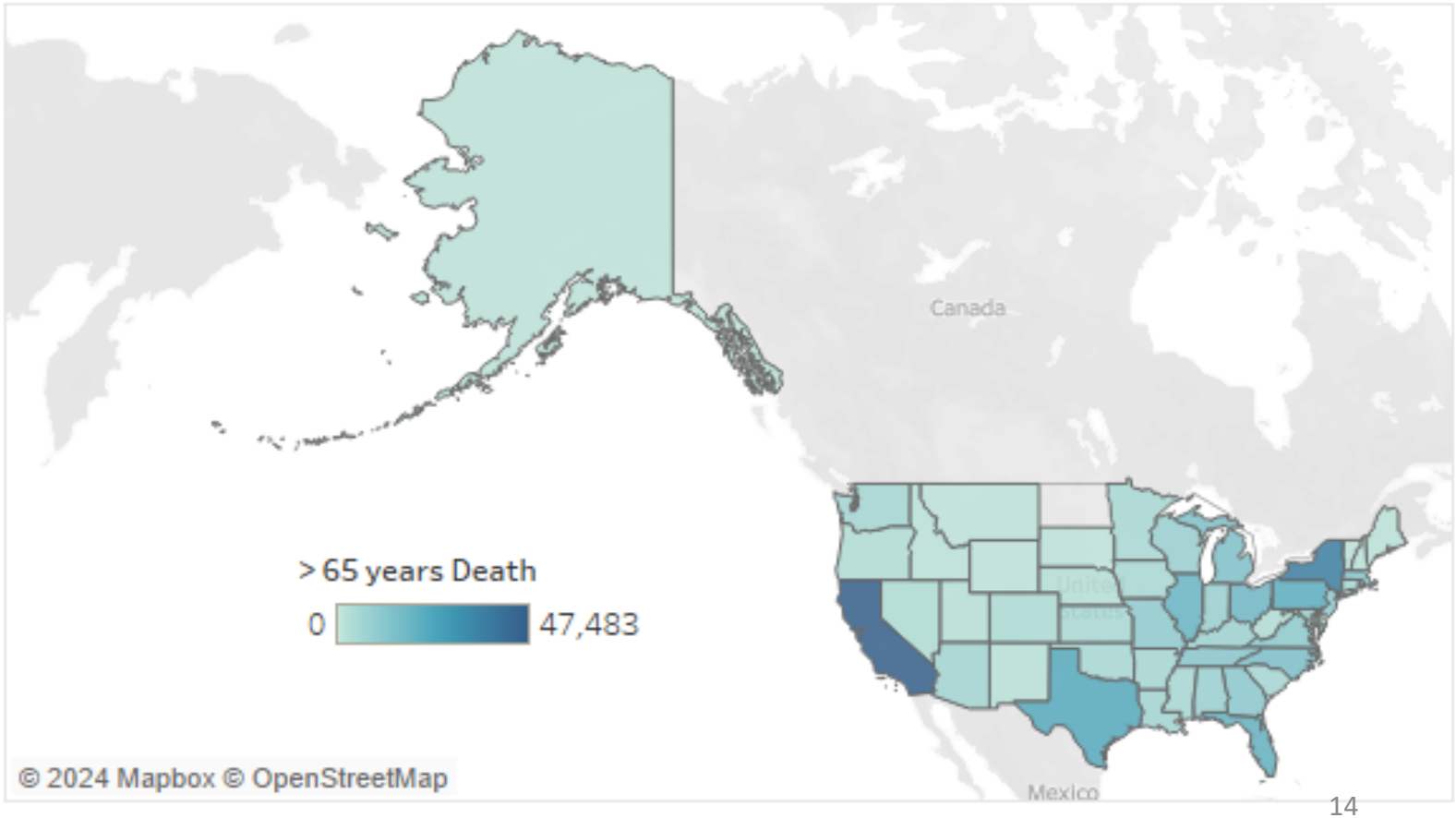
It can be concluded according to projct findings that 65+ age group populations are more vulnerable to Influenza complications and deaths than yunger age group.

Recommondations:

Additional Staff need to be sent to States with higher vulnerable populations: California, Florida, New york and Texas.

These additional staff should be sent in the beggining of November.

A map of U.S states with highest 65+ vulnerable populations



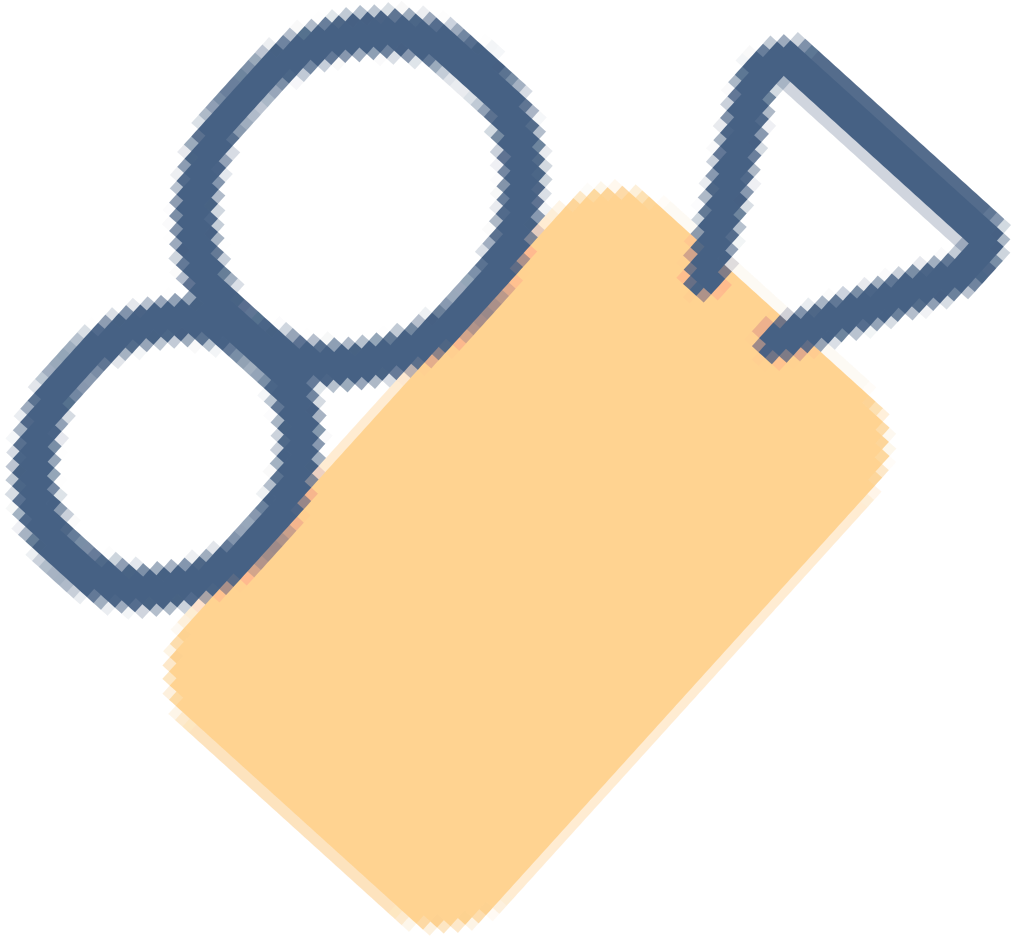
ROCKBUSTER STEALTH LLC

Launching Rockbuster
Stealth online movie service

Skills applied:

Relational Data Basis, SQL, Creating Data Dictionary, Database Querying, Data Filtering, Data Cleaning and Summarizing, Joining Tables, Subqueries , GitHub.

PROJECT OVERVIEW



What is Rockbuster Stealth LLC is a fictional global movie rental company facing competition from streaming giants like Netflix and Amazon Prime. To stay competitive, they plan to launch an online video rental service using their extensive movie license portfolio.

Objective: this project will use data analytics to address key business questions, informing Rockbuster's 2020 strategy as they shift to an online service model. These insights will guide decision-making and strategic planning.

KEY QUESTIONS

1. Which movies contributed the most/least to revenue gain?
2. What was the average rental duration for all videos?
3. Which countries are Rockbuster customers based in?
4. Where are customers with a high lifetime value based?
5. Do sales figure vary between geographic regions?

DATA OVERVIEW

Films: 1000

Film Genres: 17

Film Runtime: Average 115 min

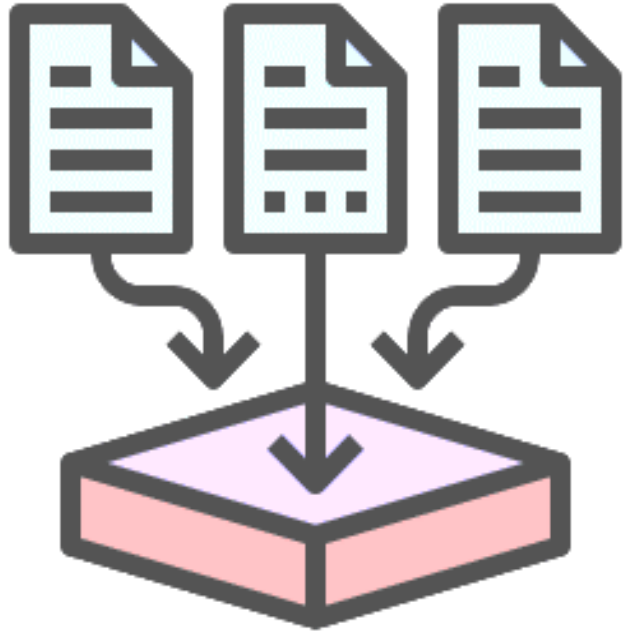
Average Rating: PG-13

Replacement Cost Average: $20

Average Rental Duration: 5 Days

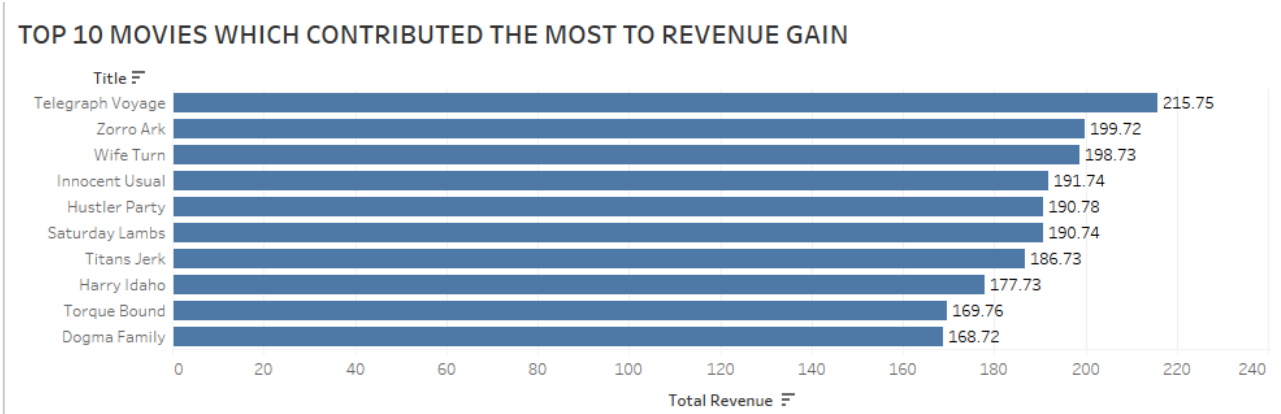
Average Rental Rate: $2.98

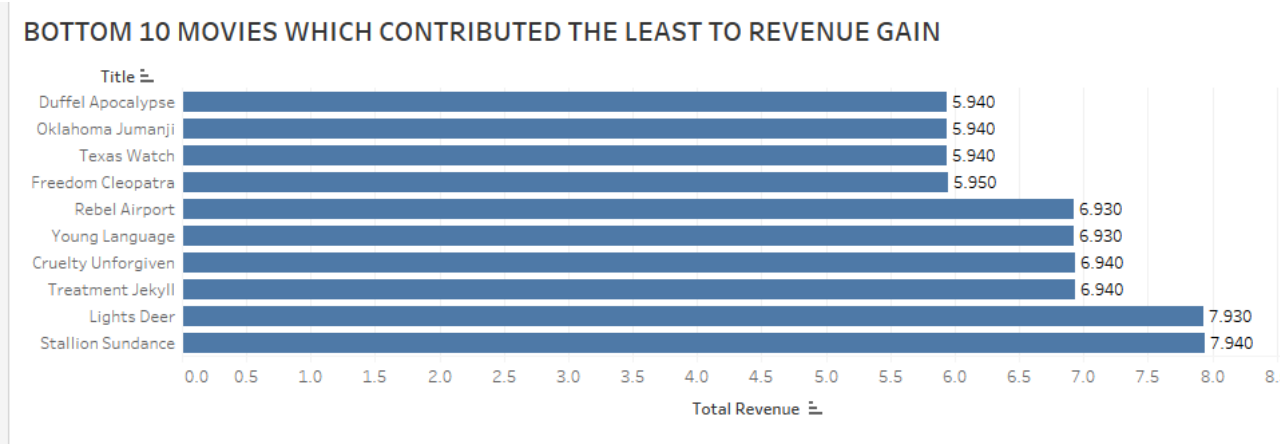
Customers: 599



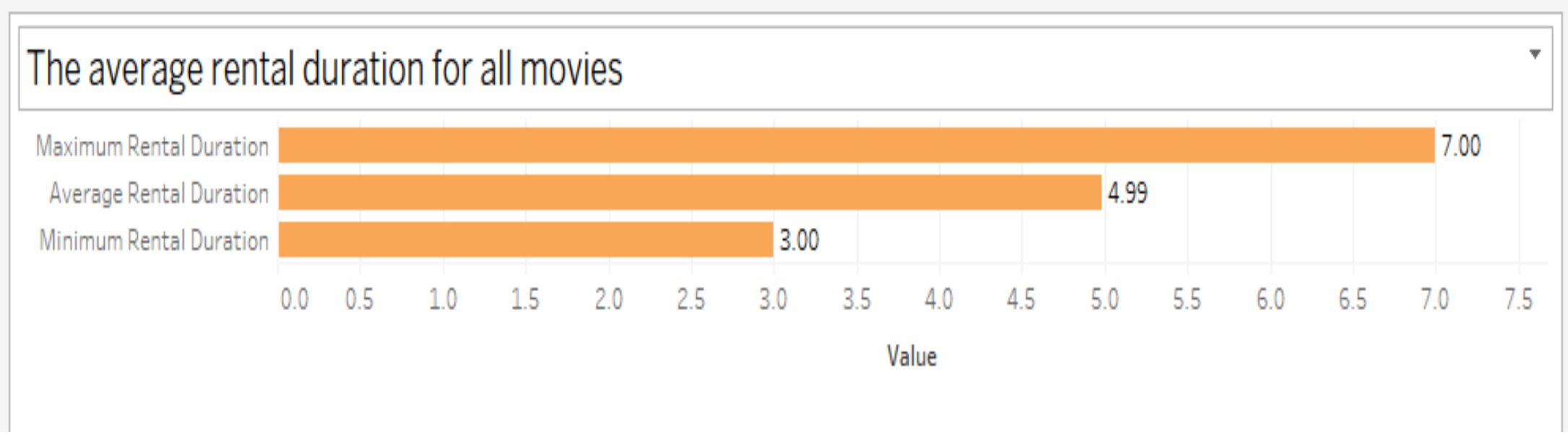
1 WHICH MOVIES CONTRIBUTED THE MOST/LEAST TO REVENUE GAIN ?

Title	Name	Total Revenue
Telegraph Voyage	Music	215.75
Zorro Ark	Comedy	199.72
Wife Turn	Documentary	198.73
Innocent Usual	Foreign	191.74
Hustler Party	Comedy	190.78
Saturday Lambs	Sports	190.74
Titans Jerk	Sci-Fi	186.73
Harry Idaho	Drama	177.73
Torque Bound	Drama	169.76
Dogma Family	Animation	168.72





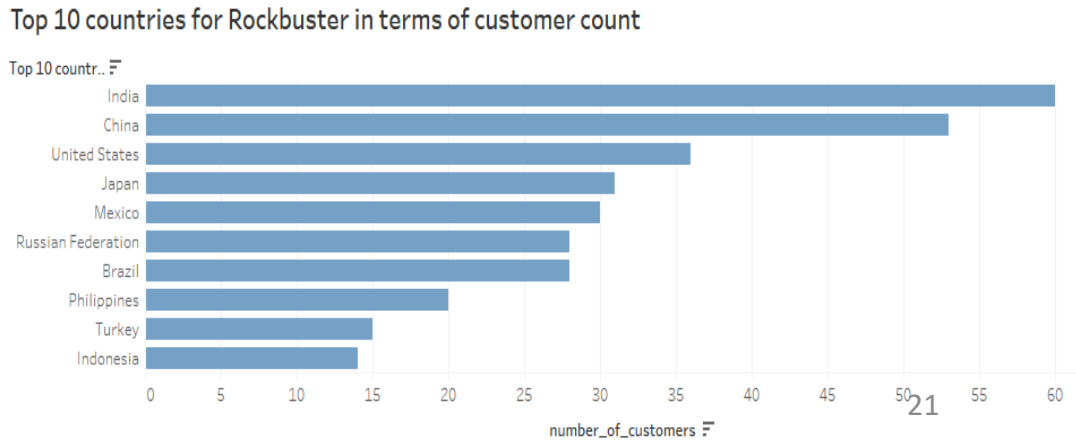
2 WHAT WAS THE AVERAGE RENTAL DURATION FOR ALL VIDEOS?



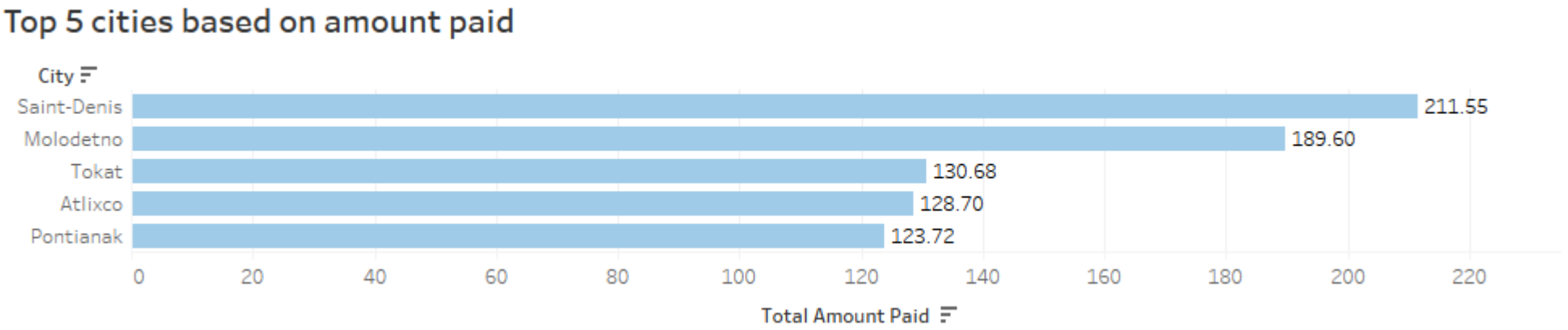
- The rental durations of the movies are between 3 – 7 days.
- The average duration of rental is 5 days.
- The most frequent duration of rental was 6 days.

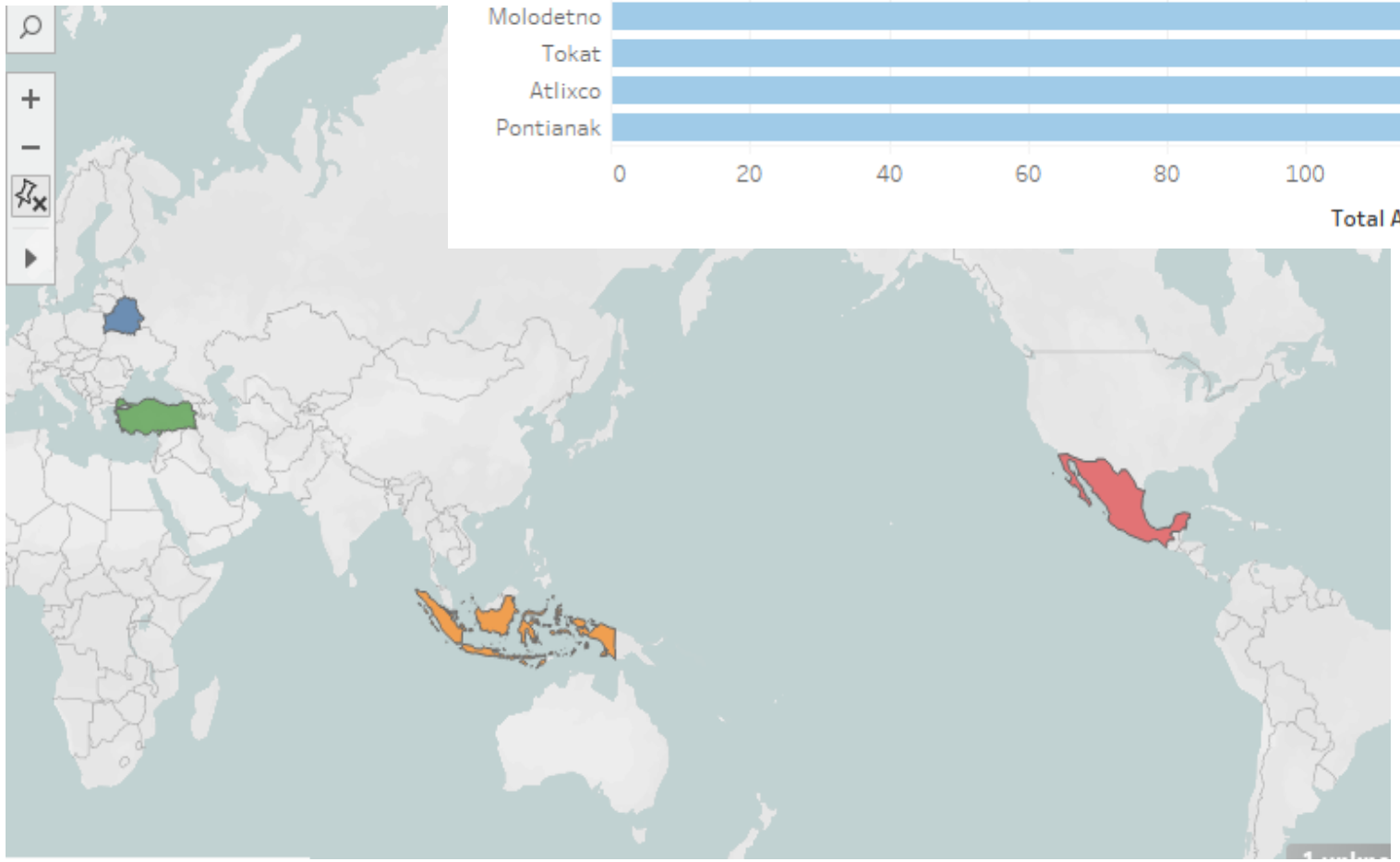
3 WHICH COUNTRIES ARE ROCKBUSTER CUSTOMERS BASED IN?

Top 5 countries:
India
China
United States
Japan
Mexico



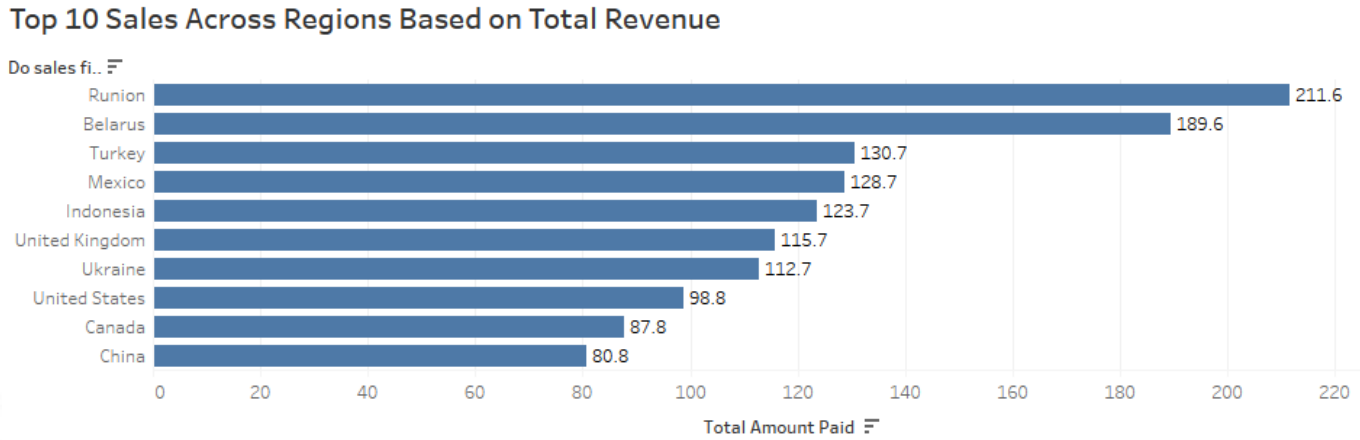
4 WHERE ARE CUSTOMERS WITH A HIGH LIFTIME VALUE BASED?

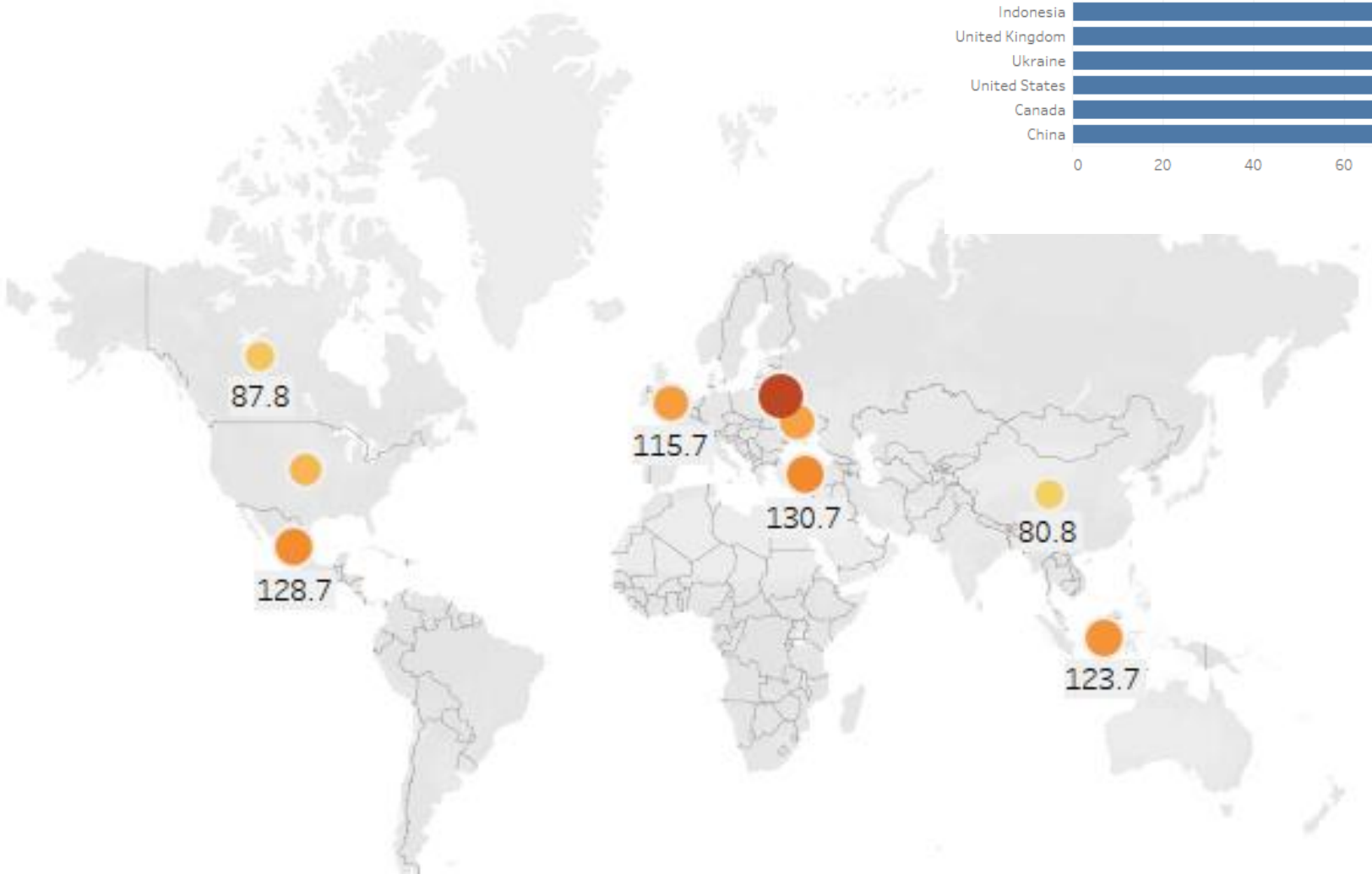




City	Total_amount_paid
Saint-Denis	211.55 $
Molodetno	189.6 $
Tokat	130.68 $
Atlixco	128.7 $
Pontianak	123.72 $

5 Do sales figures vary between geographic regions?





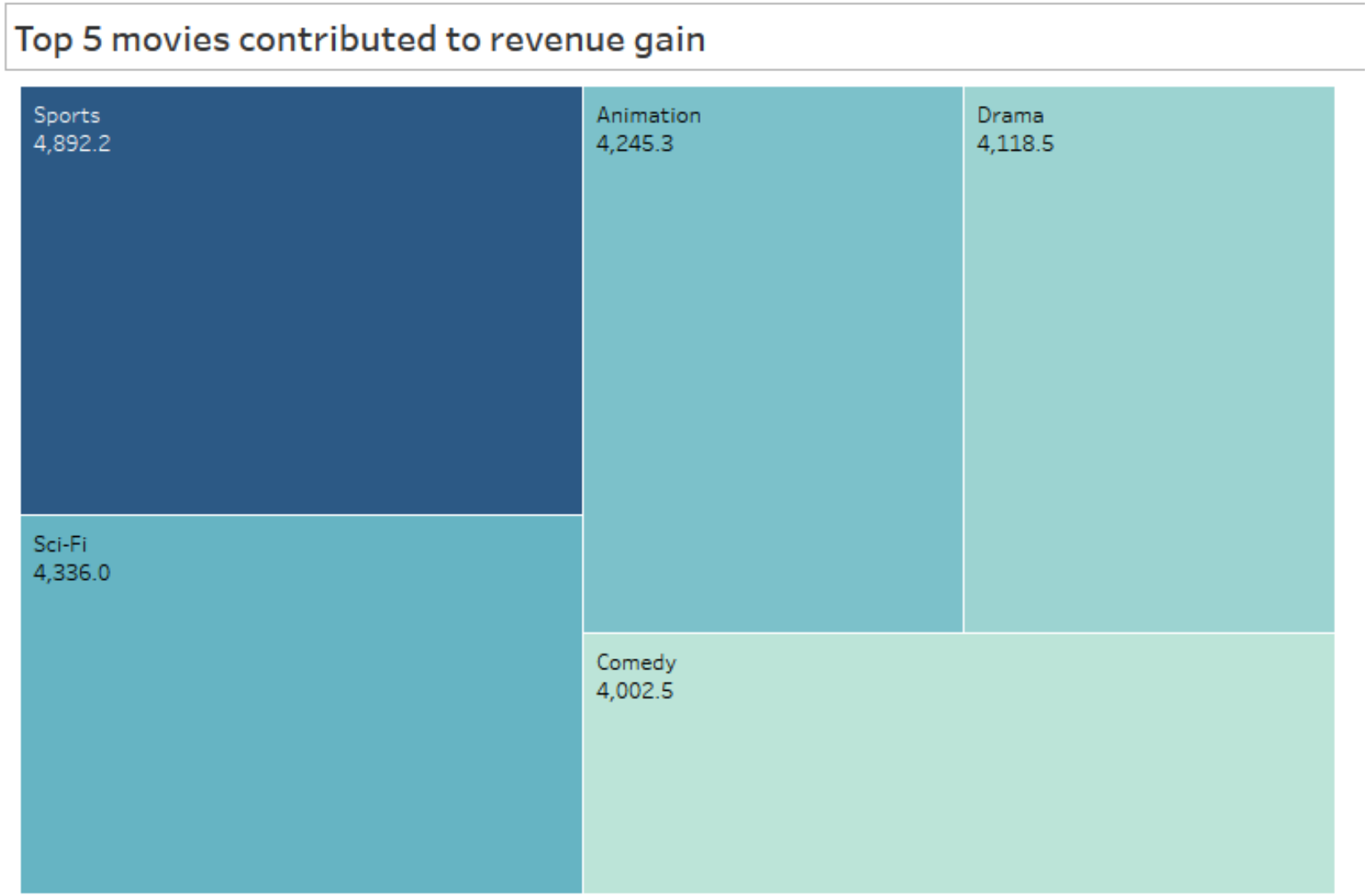
Top 3 countries with high sales based on revenue gain:

Runion

Belarus

Turkey

MOST POPULAR GENRES



CONCLUSION

- The top 5 genres contributing to revenue gain are sports, sci-fi, animation, drama and comedy.
- Top 5 movies contributing to revenue gain are Telegraph Voyage, Zorro Ark, Wife Turn, Innocent Usual, and Hustler Party.
- Top 10 countries based on number of customers are India, China, United States, Japan, and Mexico.
- Our top 5 cities with highest sales are Saint-Denis, Molodetno, Tokat, Atlixco and Pontianak.
- The majority of the high lifetime value customers are from the highest performing countries.

RECOMMENDATION

✓ **Beta Launch in Strategic Countries:**

- Prioritize a beta launch in countries with high customer engagement, such as India, China, the United States, Mexico, and Russia.

- This approach allows loyal customers to provide valuable feedback and suggestions, improving the service before a full-scale launch.

✓ **Genre-Specific Catalog:**

- Leverage the popularity of specific film genres (e.g., Sports, Sci-fi, Animation, Drama).

- Ensure that the majority of the new online video service's catalog includes titles from these genres.

Instacart Grocery Basket Analysis

Market segmentation analysis to uncover sales

Skills applied:

Python, Data Wrangling, Data Merging, Deriving Variables, Grouping Data, Aggregating Data, Reporting in Excel, Population Flows, GitHub.

ABOUT PROJECT

Objective

As an analyst for Instacart, I aim to uncover deeper insights into sales patterns. Despite our strong sales, my task is to perform an initial data analysis to suggest strategies for better segmentation based on the provided criteria.

Context

Instacart stakeholders are keen to understand
The diversity of their customers and purchasing behaviors. They recognize the need for a targeted marketing strategy to effectively reach different customer segments. My analysis will guide the development of tailored marketing campaigns to boost product sales by targeting the right customer profiles with appropriate products.

ABOUT PROJECT

Project Scale

1 month

Primary Stakeholder

Careefoundry Data Analysis Course.

Skills

- Data cleaning
- Data wrangling
- Data merging
- Deriving variables
- Grouping & aggregating data
- Data Exporting
- Excel reporting

Tools

Python, Jupyter Notebook, pandas, Numpy, Seaborn & Matplotlib libraries

Data

- **Instacart** open source dataset
- **Customer dataset** created by Careefoundry

THE PROCESS

Step1: Data Preparation

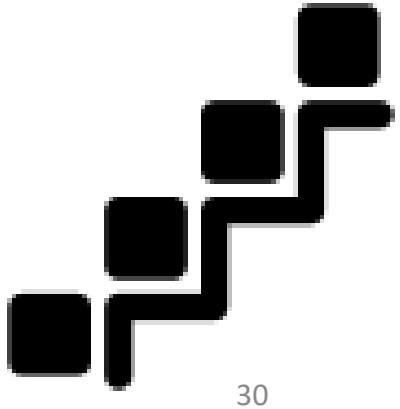
To prepare the Instacart dataset,
I started by loading the data and cleaning it by removing duplicates, handling missing values, and ensuring correct data types. I then dropped unnecessary columns, renamed them for clarity, and merged tables. I checked for consistency and validated ranges.
I performed exploratory data analysis (EDA) with statistics and visualizations. I created new features through grouping and aggregation. Finally, I exported the cleaned dataset for further use.

Step2: Data Analysis

To answer key questions and provide clear insights, I used this approach to analyze the data: I started by finding the busiest days and times for orders and used charts to show these patterns. Next, I looked at when customers spent the most money and visualized these trends. I identified the most popular products and departments and displayed this data with simple charts. Finally, I grouped customers based on their buying habits to understand different types of shoppers.

Step3: Results

Despite the solid foundation of the analysis, data limitations were apparent in all sections of the results. Stakeholders were informed of the findings, with a clear note highlighting these data constraints.



ORDER DAYS/HOURS ANALYSIS

Key Questions

1. Determine the busiest days of the week and hours of the day with fewer orders .
2. Determine the times of the day when people spend the most money

CONSTRAINTS

Analyzing data from a specific time period may miss long-term trends. Seasonal variations also impact patterns. These Temporal constraints make it challenging to gain clear insights.

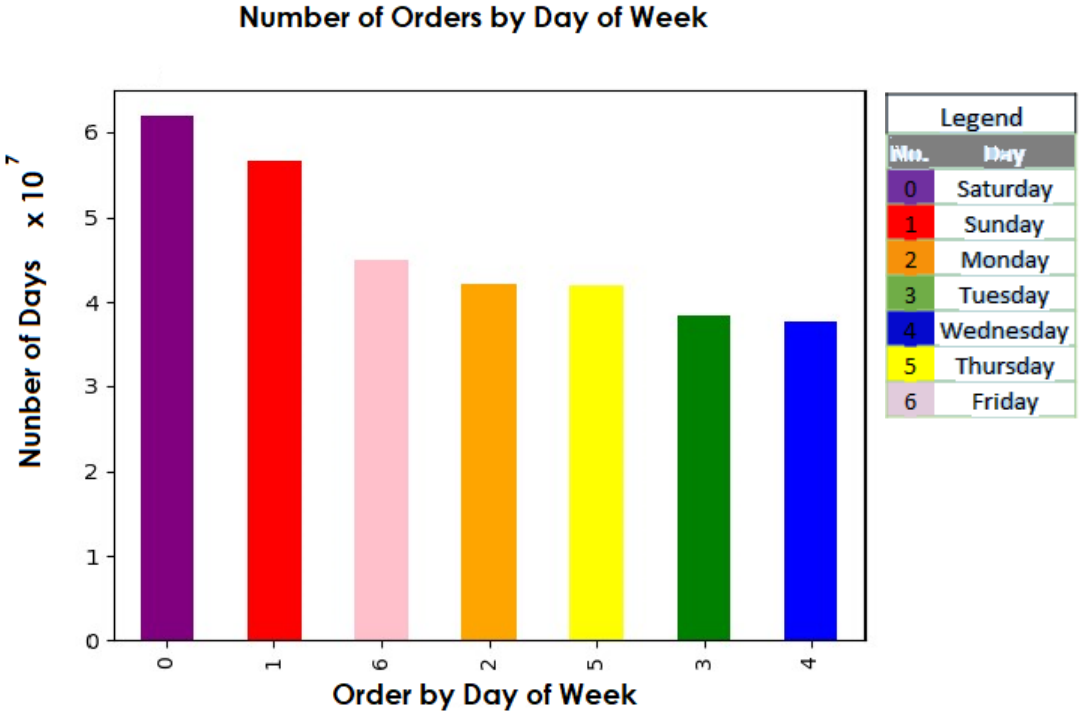
SOLUTION

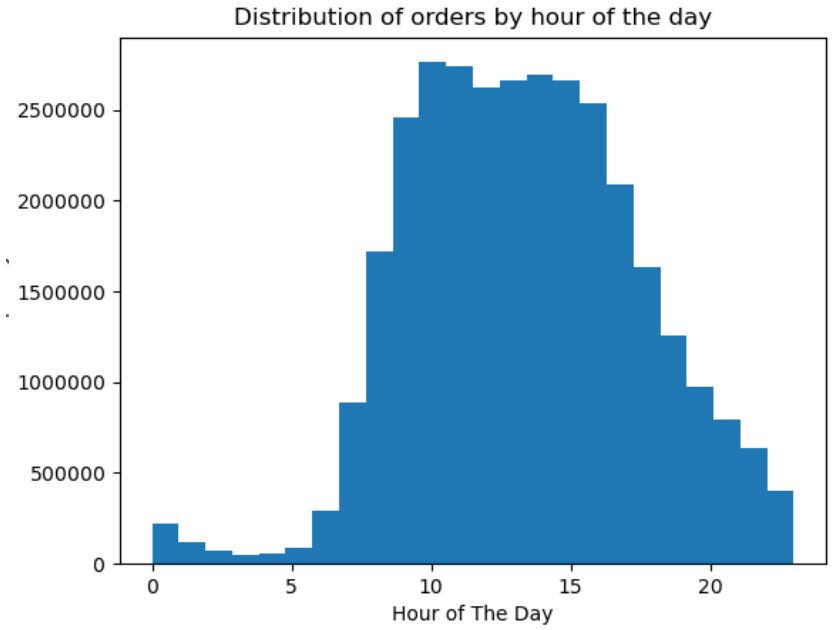
To address temporal constraints in data sets, consider widening the time window for analysis.

ORDER DAYS/HOURS ANALYSIS

1

This Chart shows the Number of Orders by Day of Week. Saturday and Sunday are the busiest days of the week. Wednesday is the least busy day.



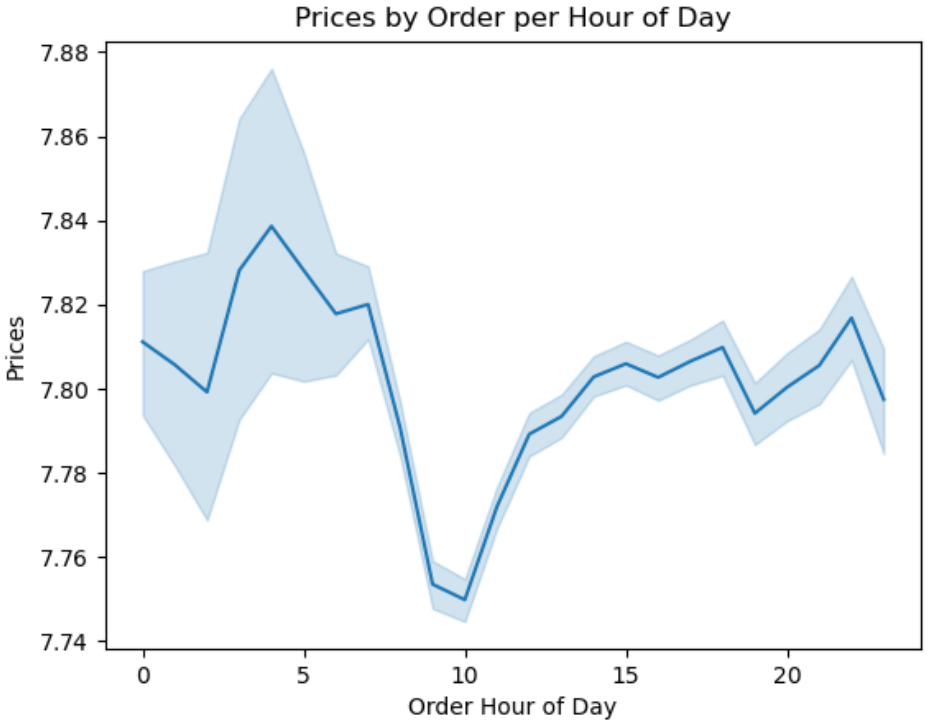


The histogram shows the distribution of orders by hour of the day. The number of orders started increasing between 8:00 am and 4:00 pm. It reached its peak around 10:00 am. After 4:00 pm, orders decreased gradually.

ORDER DAYS/HOURS ANALYSIS

2

The Lline chart shows Prices by Order per Hour of Day. we can see clearly that people spend most of their money between 4:00 am- 6:00 am. After 6:00 am, the line dramatically declined and hit rock bottom between 9:00 and 10:00 am. However, it started to increase again after 10:00 am.



PRODUCTS & DEPARTMENTS ANALYSIS

Key Questions

1. For simplicity ,create a price flag with price ranges (low, medium, high).
2. Determine which departments have the highest frequency of product orders

CONSTRAINTS

Lack of some data like quantities of the ordered items can lead to unclear insights.

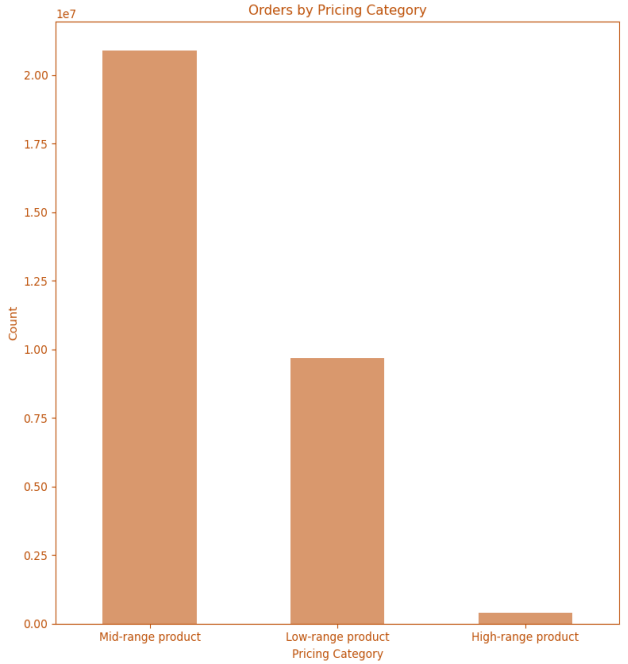
SOLUTION

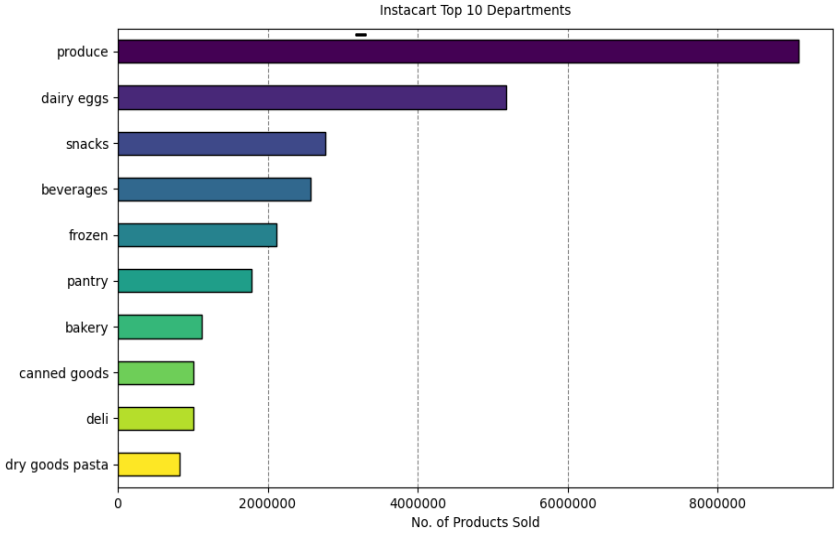
Analysis can be conducted based on some other available features or variables.

PRODUCTS & DEPARTMENTS ANALYSIS

1

This chart shows orders by pricing category.
we can see clearly that the majority of customers prefer the mid-range products as well as the low-range products as a second priority.





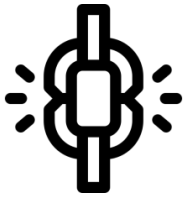
2

The chart shows the top 10 departments of Instacart.
it shows that the most ordered products are produce" followed by airy eggs and then snacks. The less ordered products are dry goods pasta and deli.

Customer Analysis

Key Questions

1. Determine the ordering behavior of the customers.
2. Are there any differences in ordering habits based on demographical and geographical characteristics.



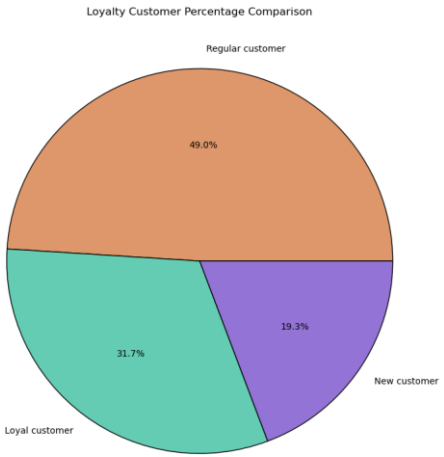
CONSTRAINTS

availability bias in demographic and geographical data make it challenging to make future decision and uncover insights.

SOLUTION

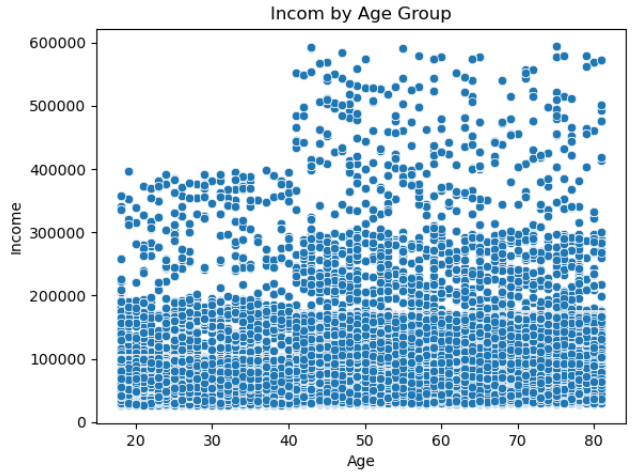
No solution for this kind of data availability issues. The analysis was conducted based on the data available.

Customer Analysis



1

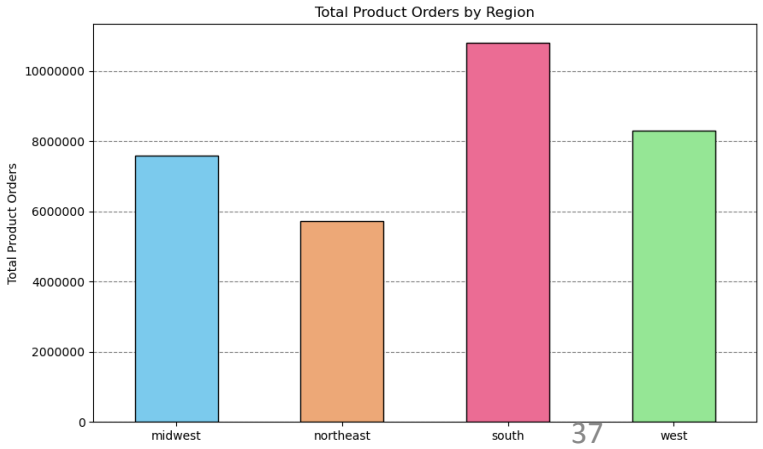
The Pie chart shows that the majority of Instacart customers (51%) are regular customers. Loyalty customers make up 33 % and the new ones make up 15.5 %.



From Scatter plot, we can conclude that customers of age 40 and onward contributed the most in the Instacart ordering habits since the income is higher compared to other age groups.

2

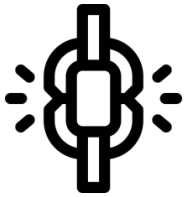
The southern states contribute the most in Instacart ordering habits compared to the other three regions. These findings were expected because the southern states are more densely populated followed by the western region..



Customer Profiling

Key Questions

1. Identify specific customer profiles.
2. Are there any differences in the ordering habits of different customer profiles.



CONSTRAINTS

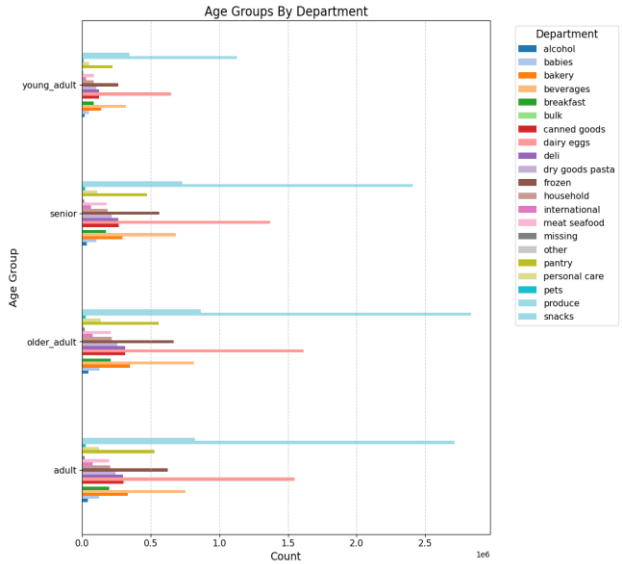
availability bias in demographic and geographical data make it challenging to make future decision and uncover insights.

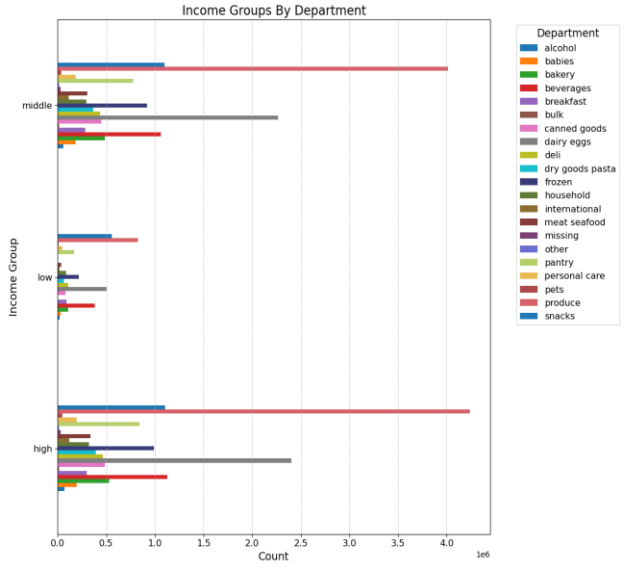
SOLUTION

No solution for this kind of data availability issues. The analysis was conducted based on the data available.

Customer Profiling

- Examining department by age groups, we observe that produce consistently achieves the highest orders across all age demographics, followed closely by dairy and eggs.
- When analyzing department by income groups, we find that produce remains the top-performing department. Interestingly, among low-income customers, alcohol emerges as the second-highest selling department, while dairy and eggs hold that position for other income groups.





Challenges

- When working with the Instacart dataset, I encountered performance issues that led to a slowdown in Jupyter Notebook. The large dataset posed challenges, affecting execution speed and responsiveness.

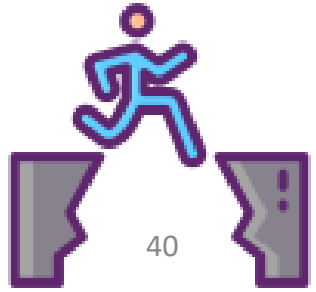
  Solution: I reinstalled all tools I needed to proceed further with my analysis.

- Some codes did not execute in Jupyter Notebook.

  Solution: I realized that having many Jupyter notebooks open simultaneously can consume a significant amount of your computer's memory (RAM) and processing power.

- Managing and processing this vast amount of data was challenging to me due to its sheer scale and complexity.

  Solution: to address this challenge, I incorporated a Table of Contents (ToC) feature. This allowed for seamless navigation and facilitated transferring throughout the notebook.



Recommendations

- Advertise more on Tuesdays and Wednesdays specially between 12 AM and 5 AM.
- Highlight specific products during peak times and run additional ads for other categories.
- Reward loyal customers and encourage new customers with incentives.
- Offer affordable prices on popular youth products.
- Focus on the West and Midwest.

Pig E. Bank

Customer Churn Analysis

Skills applied:

Big Data, Data Ethics, Data Mining, Predictive Analysis, Time Series Analysis, Forecasting.

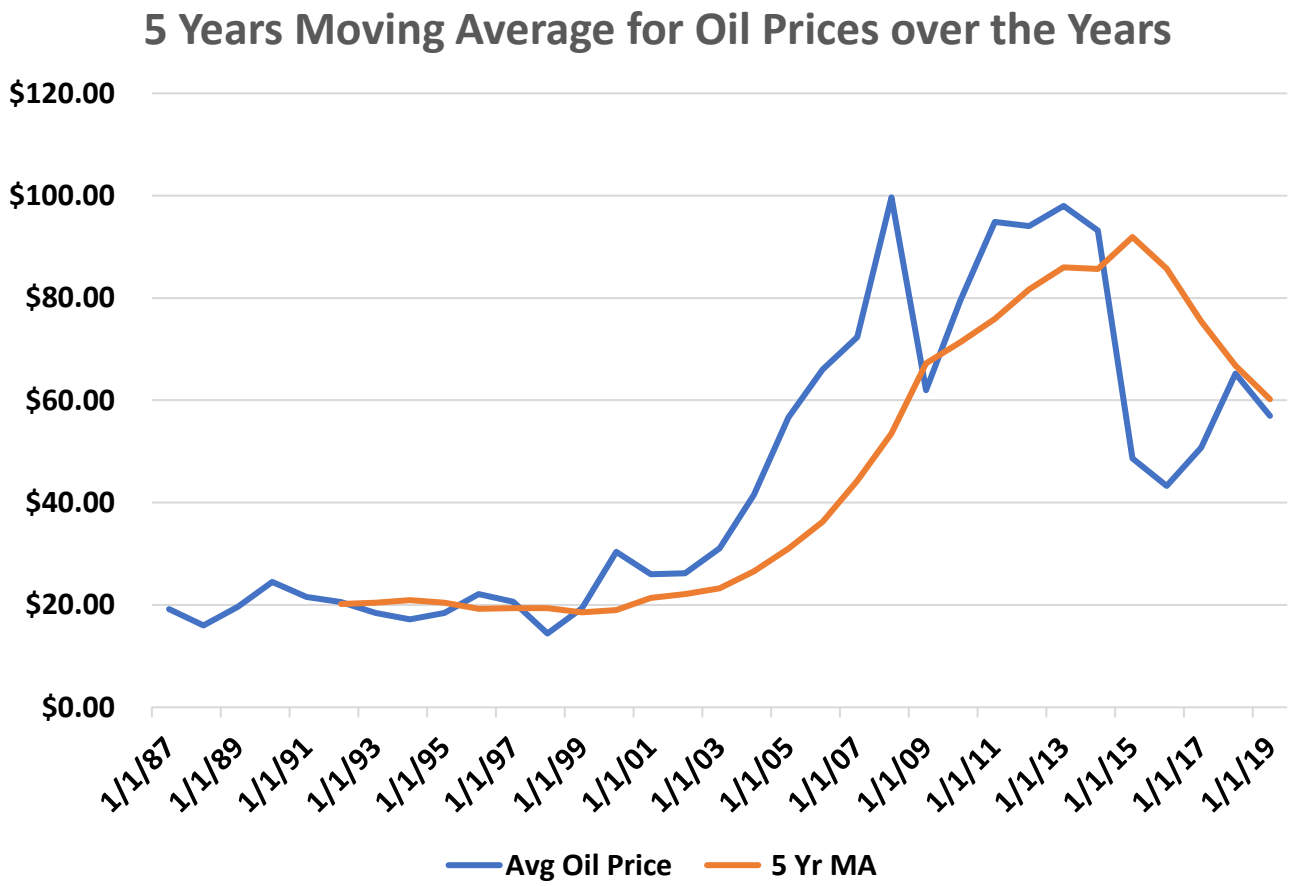
Pig E. Bank Analysis

Objective:

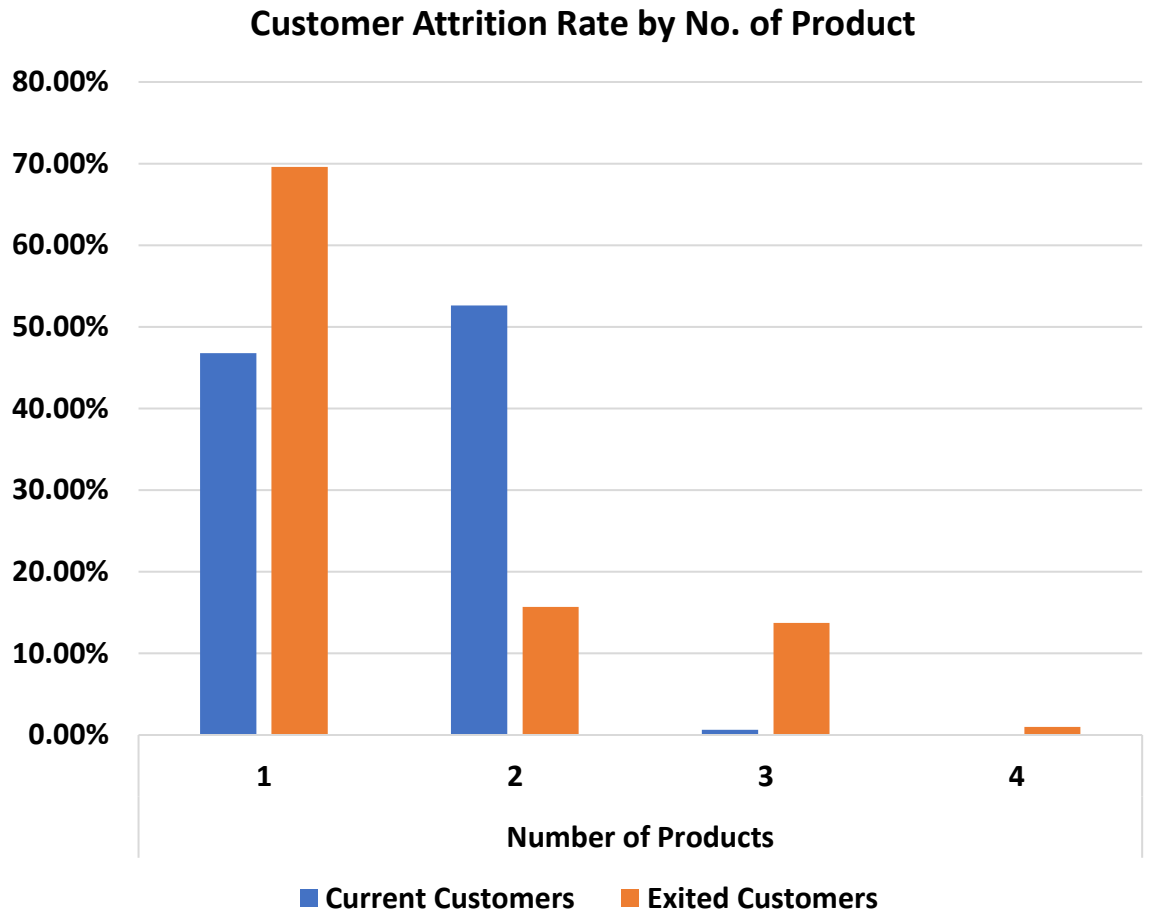
The goal of this project was to thoroughly analyze customer satisfaction data at Pig E. Bank. This analysis aimed to identify the factors contributing to customer attrition, with the ultimate objective of devising effective strategies to enhance customer retention.

Key Question

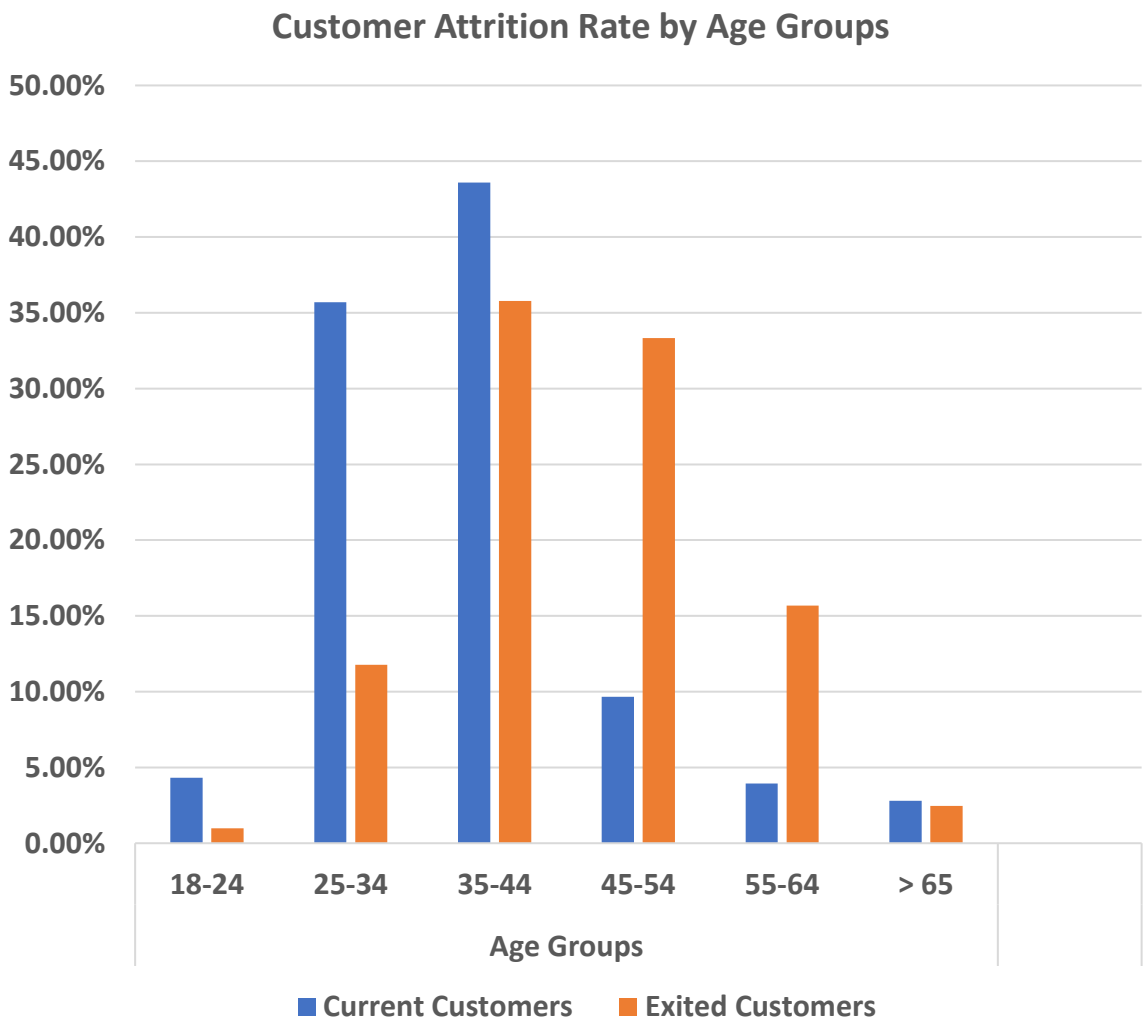
What are the key risk-factors in identifying customers who are most likely to churn?

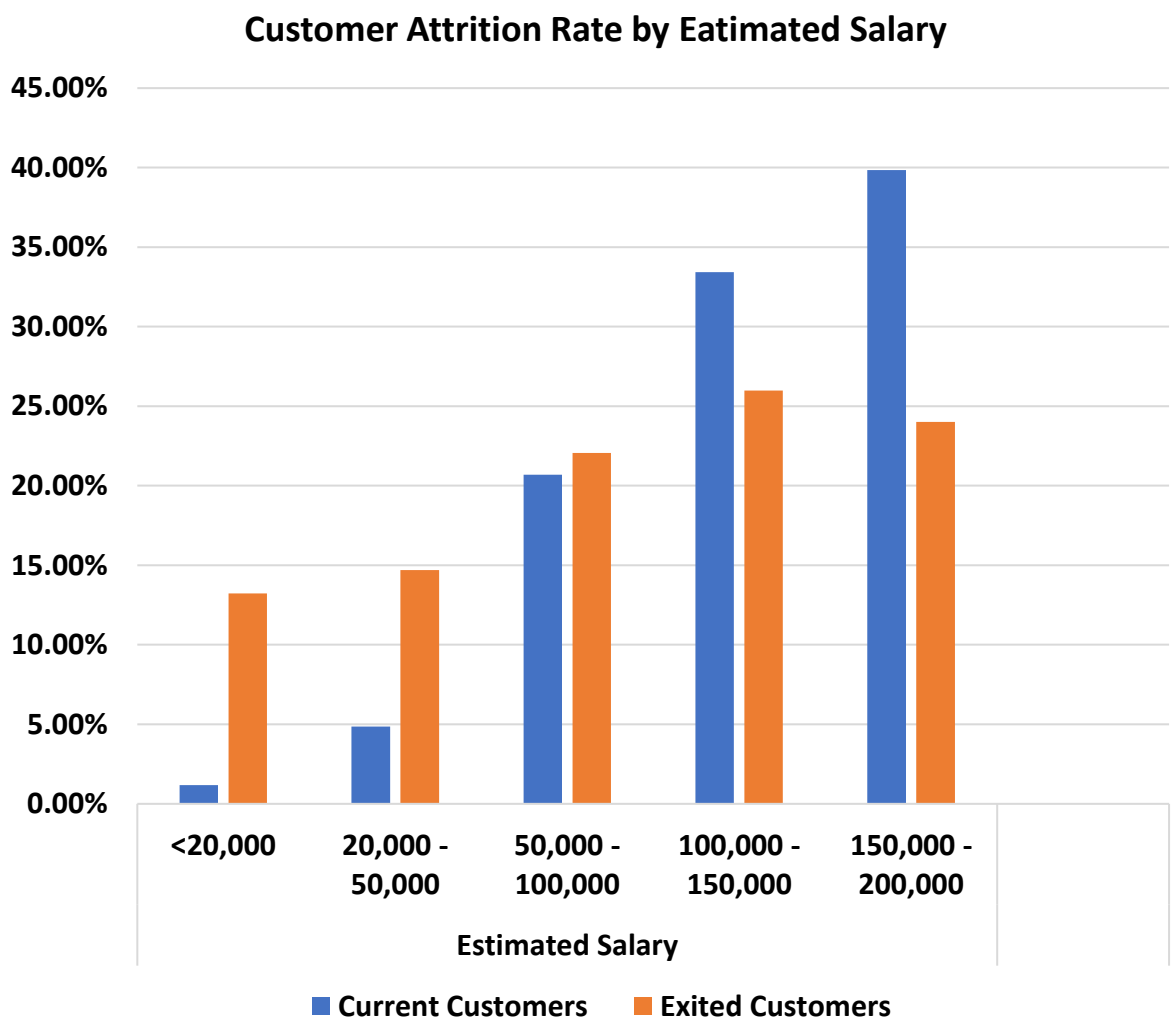


Pig E. Bank – Analysis



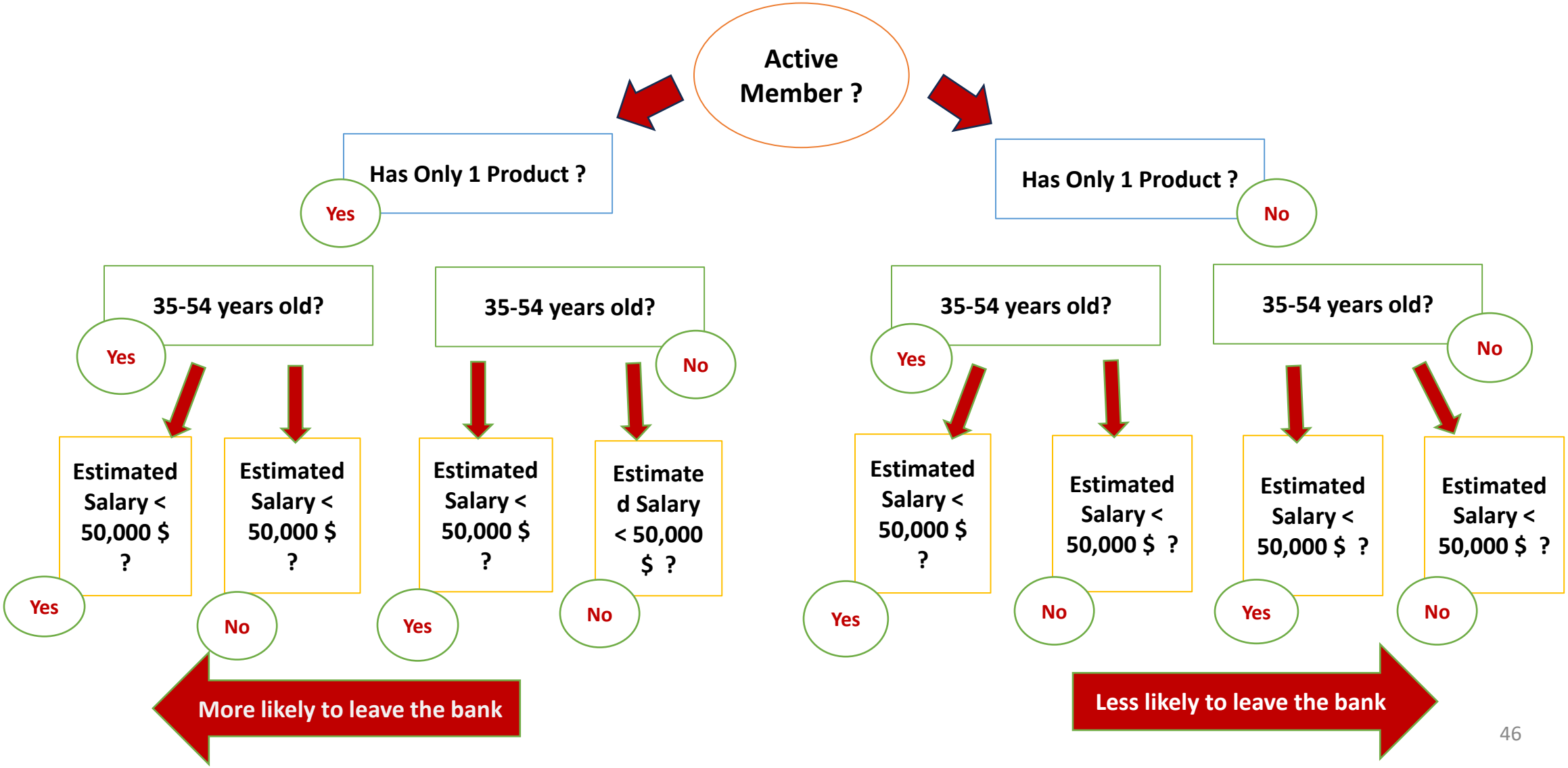
Pig E. Bank – Analysis





Analysis – Decision Tree

How Likely is a client to leave the bank?



Pig E. Bank – Key Insights

The major contributing factors in leaving the bank

1. **Customer Membership:** is a crucial contributing factor in leaving the bank.

2. **Number of Products:** 69 % of customers who have only one product left the bank, while customers with more than one did not leave.

3. **Age Group:** 33% of customers of age range 45-54 left the bank compared to the current customers of the same age can be an indicator that age is a crucial factor influences customers leaving.

4. **Estimated salary :** customers with estimated salary < 50,000 can be considered as a crucial factor which can lead to customers leaving.

Pig E. Bank – Recommendations

1. Customer Membership: Enhance membership programs with personalized services and loyalty rewards to increase engagement and satisfaction.

2. Number of Products: Promote cross-selling and educate customers on the benefits of using multiple products to encourage them to adopt more services.

3. Age Group: Develop targeted communication and financial planning services for the 45-54 age group to address their specific needs and concerns.

4. Estimated Salary: Introduce affordable products and financial assistance programs for customers with lower estimated salaries to support their financial stability.

Divvy's Bikeshare System (2015-2019)

Analyzing Usage Trends

Skills applied:

Python, Data Wrangling, Predictive Analysis, Time Series Analysis, Forecasting.

Divvy: Chicago's Bikeshare System

Project Hypothesis

Subscribers use the Divvy bikeshare service significantly more frequently than Casual users, particularly during weekday rush hours.

Divvy is Chicago's bikeshare system, launched in 2013 to promote sustainable transportation and healthy living. With over 16,500 bikes and 800 docking stations, it covers 190 square miles, including Chicago and Evanston.

Looking ahead, Divvy plans to expand its fleet, introduce advanced e-bikes, and adopt greener practices. This evolution aims to foster a strong biking culture and ensure safety for all users, contributing to a more sustainable and connected Chicago.

The dataset: 'Divvy Data' from 2017 – 2019

Data Overview

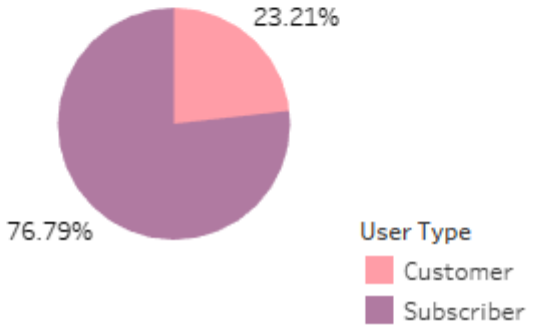
Total Trips

Total stations

Total Bikes

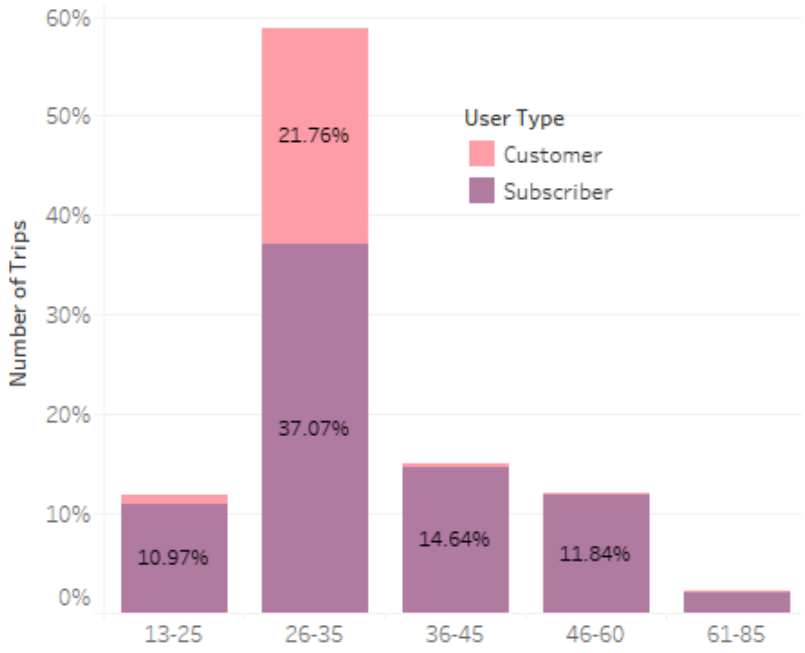
Average trip duration

Subscribers vs. Customers

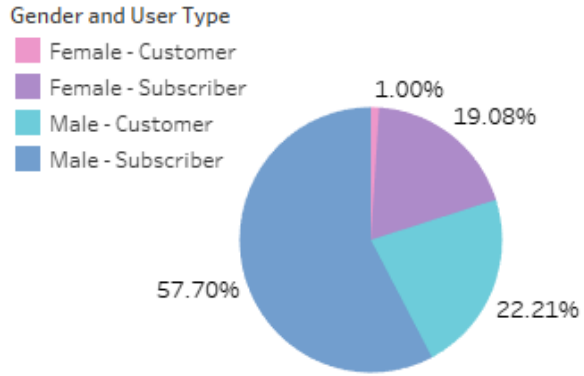


User Demographics and Behavior

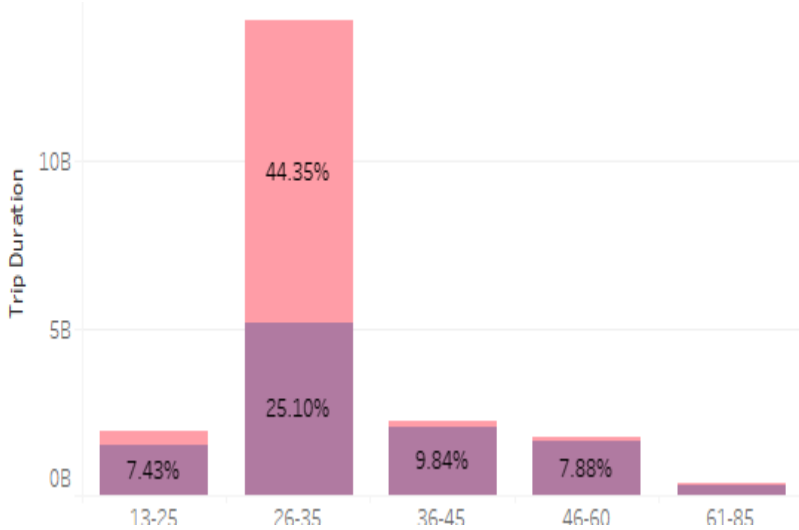
Distribution of Trips by Age Groupe and User Type



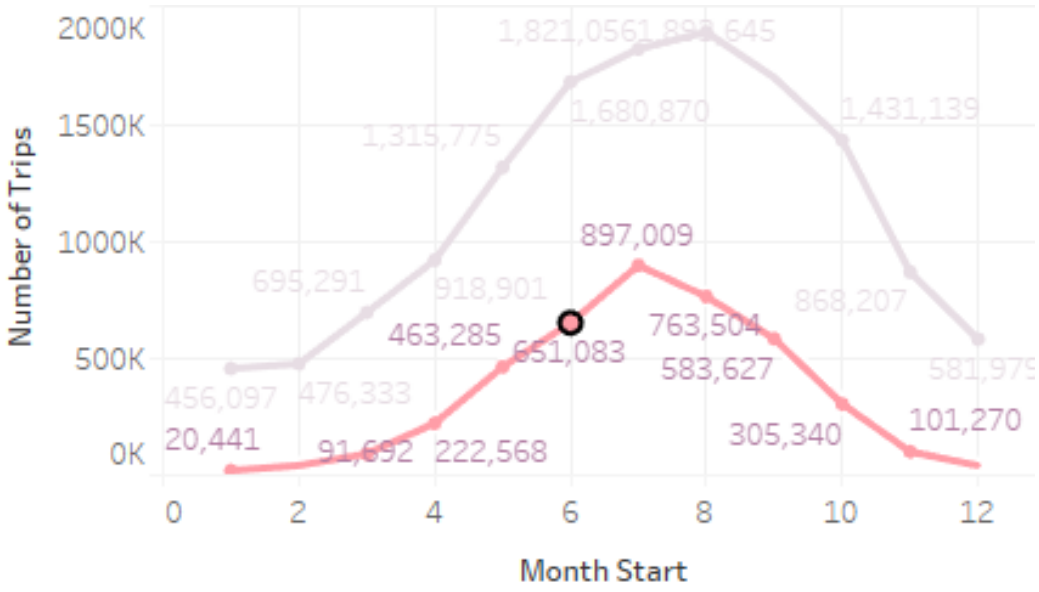
Distribution of Trips by Gender and User type

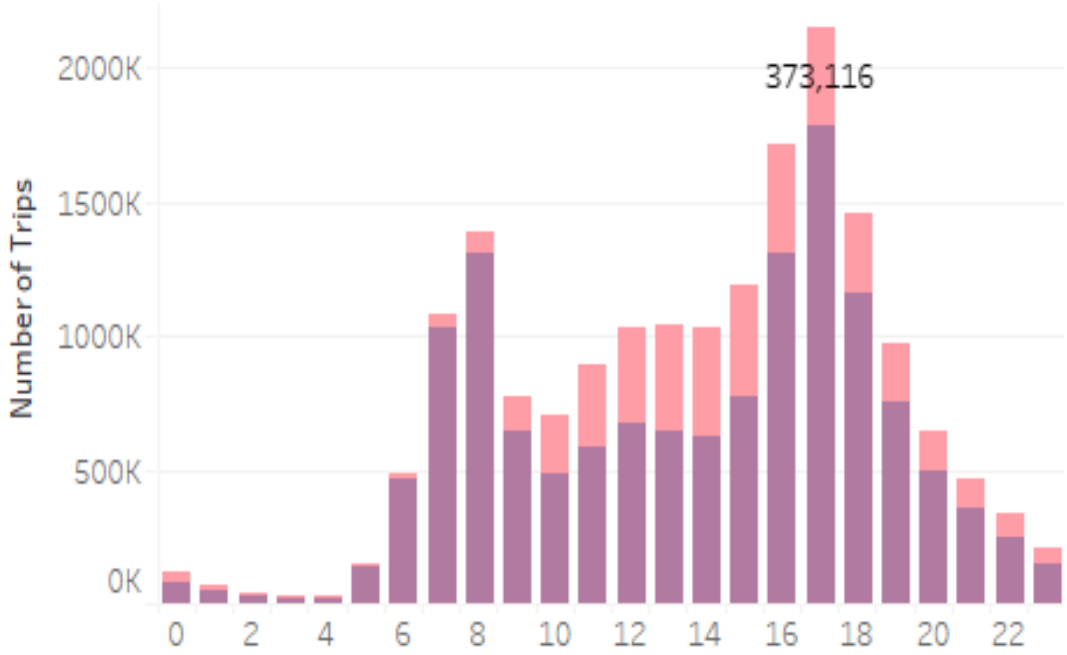


Comparative Trip Durations Among Different Age Groups

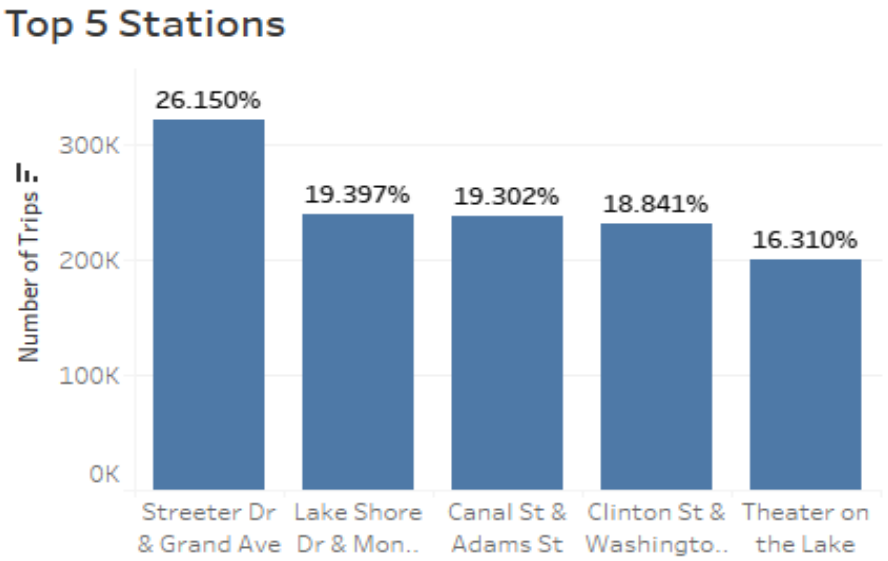


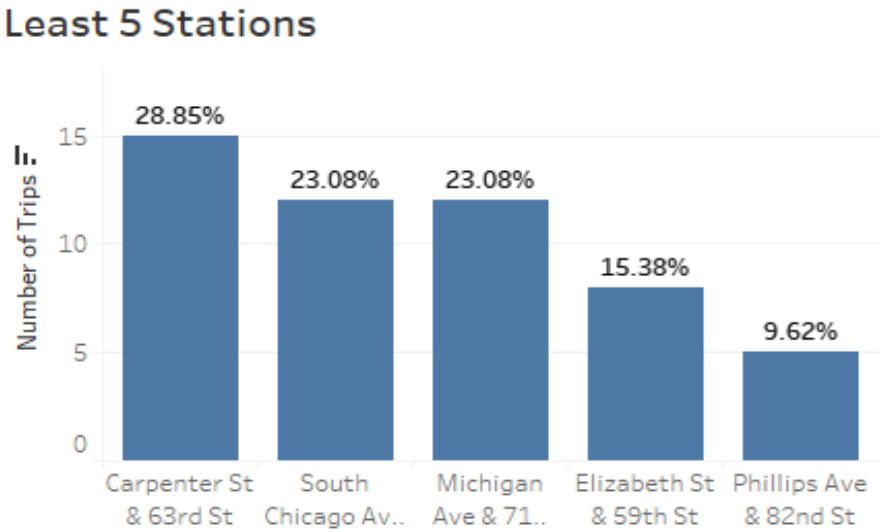
Seasonal and Temporal Trends





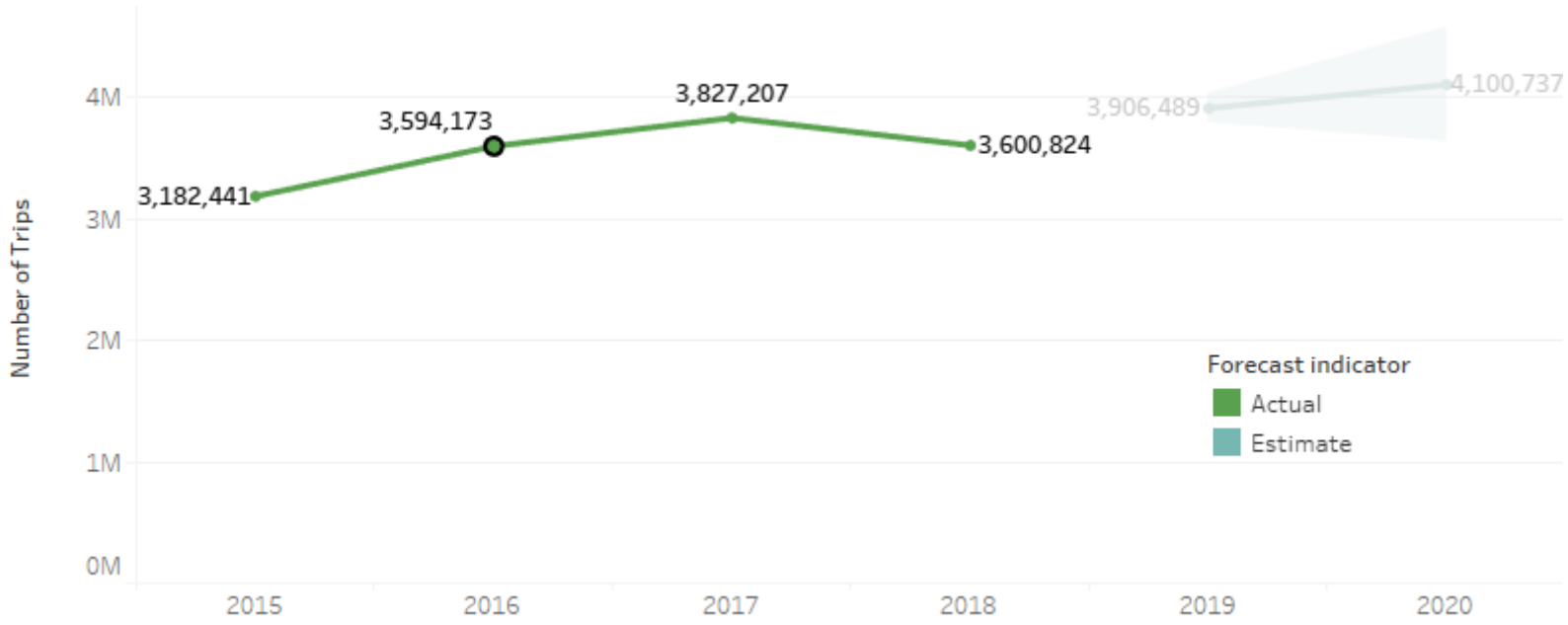
Popular Stations





Predictive Insights

Forecasting Trip Trends: Analyzing Trip Data Over Time (2015-2019)



Conclusion & Insights

Conclusion: The analysis of Divvy's trip data provides key insights into user behavior by examining trips by User Type (Subscribers vs. Customers) and Gender. Subscribers use the service more frequently, especially during weekday rush hours and warmer months, with a notable presence of female riders in the Subscriber segment. Additionally, the majority of Subscribers belong to the 31-45 age group, indicating regular commuting patterns.

Limitations: The analysis is limited by the absence of unique user identifiers, hindering accurate quantification of unique riders and potentially distorting user behavior insights. Additionally, underrepresentation of certain stations or times in the dataset may introduce bias and restrict nuanced conclusions about individual user patterns.

Insights: The analysis suggests that Divvy should enhance services during peak commuting hours for Subscribers and target younger Casual users with promotions. The presence of female riders in the Casual segment presents an opportunity for targeted marketing campaigns, while seasonal trends can help optimize fleet availability during high-demand months.

Thank You

Click to connect