



Sruthy Sreekanth

Data Analytics Portfolio



About me

- ❖ I am Sruthy Sreekanth, a Data Analyst with over a decade of experience in the banking domain, working both as a banker and as a banking consultant for an IT company. During my tenure as a banker, I managed retail branch banking operations and handled customer issues while also managing account-related data. As a banking consultant, I leveraged my expertise to analyse complex banking issues using tools like SQL & Excel on the data provided by the banks to help them manage their operations efficiently.
- ❖ Building on this experience, I have now retrained in data analytics, equipped with the latest tools and techniques like Tableau & Python to leverage the power of data to drive business growth. With the knowledge gained from my previous experience and data analytics skills, I am well-positioned to help organizations make data-driven decisions and achieve their goals.



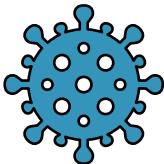
Projects



GameCo

- Perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.

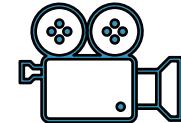
[More Info...](#)



Medical Staffing Agency

- Assist medical staffing agency in planning for influenza season. Examine trends in influenza and how they can be used to plan for staffing needs across the country

[More Info...](#)



Rockbuster Stealth LLC

- Assist business intelligence department with a launch strategy for new online video service by utilizing a relational database management system (RDBMS) and SQL to analyze the data and answer business questions

[More Info...](#)

Projects



Instacart

- Perform an initial data and exploratory analysis of data in order to provide insights into how this strategy may be implemented to ensure that Instacart targets the proper client profiles with the right products.

[More Info...](#)



Bank Customer Segmentation

- Perform exploratory and advanced analysis in order to identify distinct groups of bank customers based on their behavior or other traits that enables efficient marketing strategy.

[More Info...](#)



GameCo

Analyzing global video game sales

Project Brief



Project Goals

- Analyze the video game sales data and understand the prevailing gaming trends and the future outlook for the gaming business to ensure that GameCo maximizes return on investment.



Data

- [Videogame sales dataset](#) was taken from [VGChartz](#) and it tracks the total number of units of games sold from 1980 to 2016.
- The numbers represent units sold in millions
- [Videogame sales cleaned data](#)



Tools used

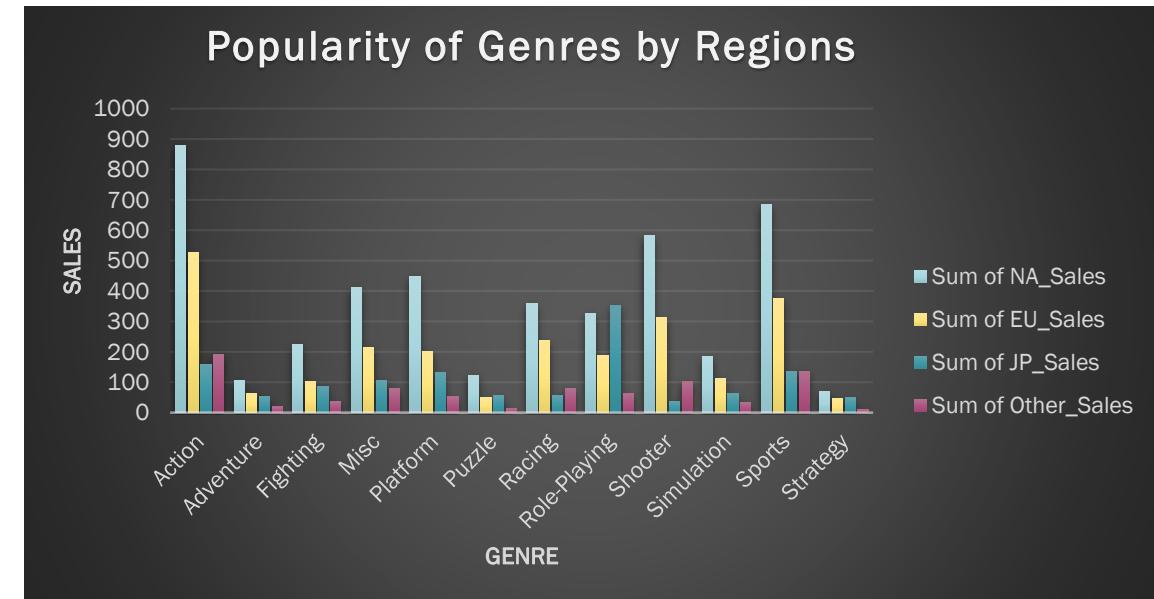
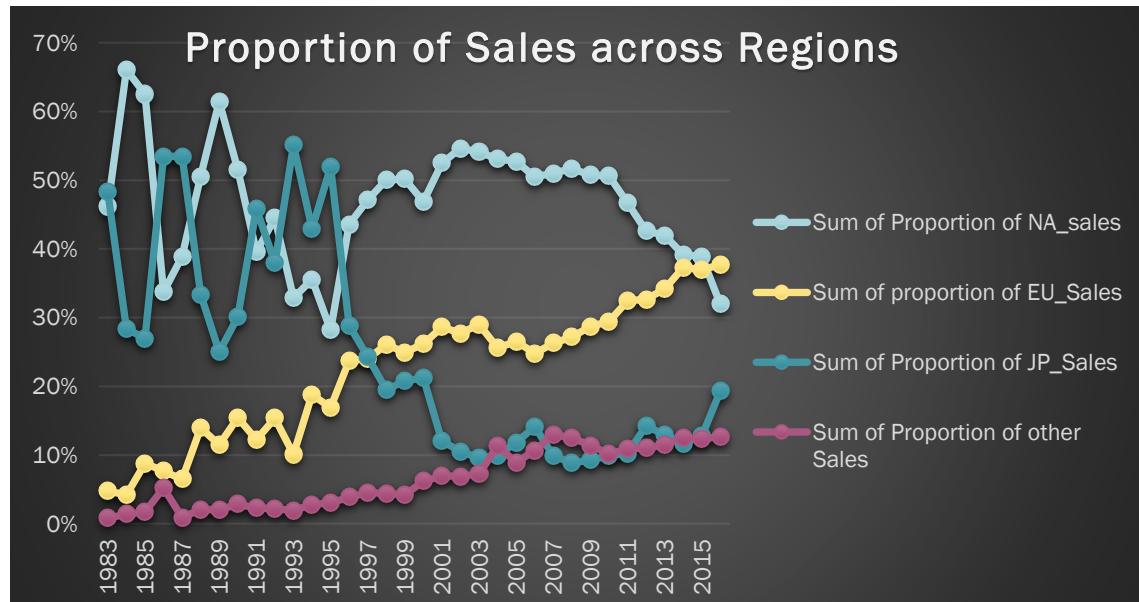
- Excel
- PowerPoint



Skills

- Data Cleaning & Grouping
- Statistical Analysis
- Pivot Tables & Charts
- Data Visualization in Excel & Storytelling

Analysis



- For the majority of the years, North America had the highest sales figures, however, there has been a decline in recent years.
- Since the beginning of the years, European sales have been steadily increasing, and in 2016, the European market overtook the North American market.

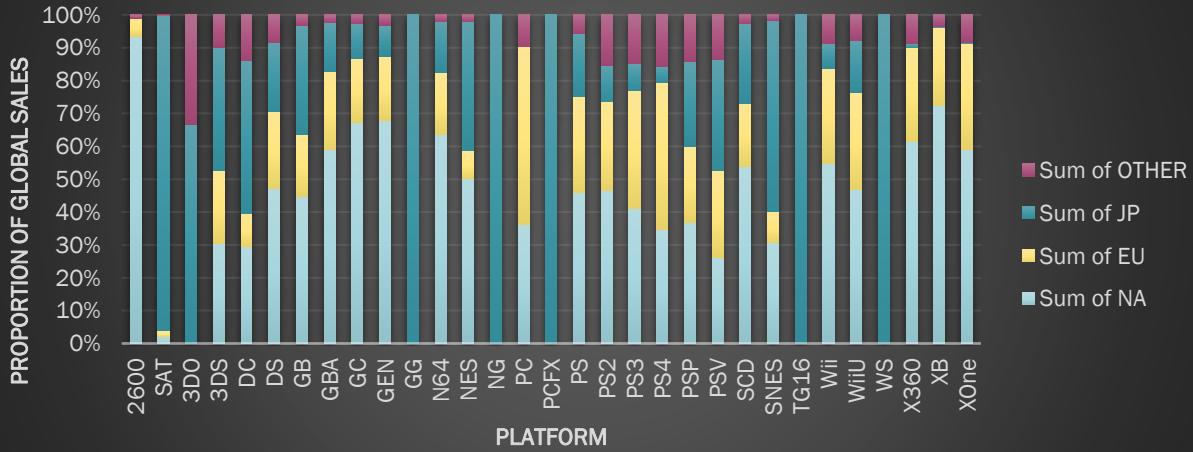
Top 3 popular Genres in North America & Europe

- Action
- Sports
- Shooter

Top 3 popular Genres in Japan

- Role-Playing
- Action
- Sports

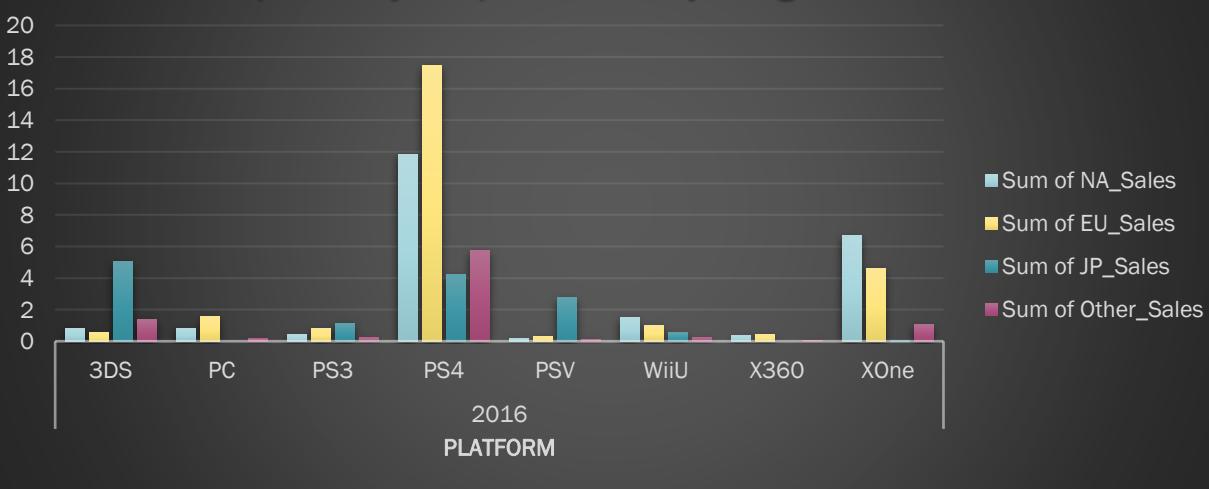
Regional sales proportion by Platforms



Top 3 Platforms in 2016

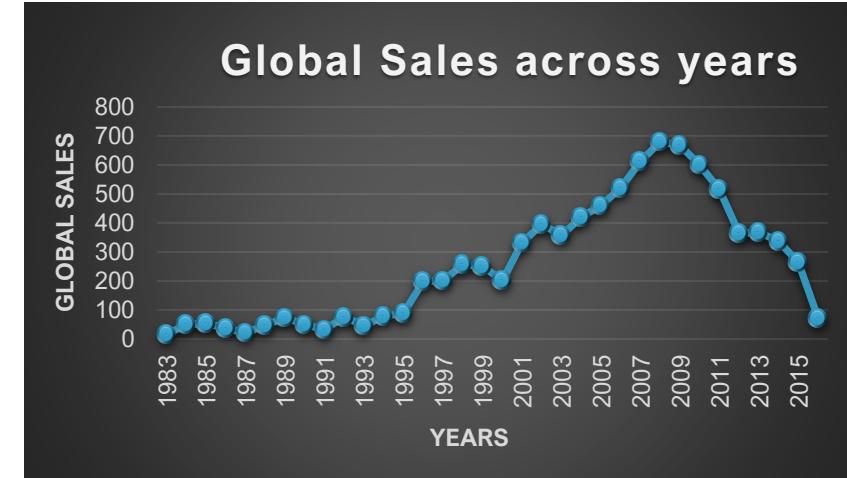
- North America – PS4, Xone, Wiiu
- Europe – PS4, Xone, PC
- Japan – 3DS, PS4, PSV
- Other – PS4, 3DS, XOne

Popularity of platform by Regions-2016



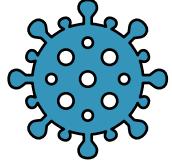
Insights from Data & Recommendations

- The worldwide gaming industry is in a downward trend, and regional video game sales are not consistent over time.
- Although there has been a decline in sales in recent years, North America has had the highest numbers of sales for most of the years and will continue to do so.
- Consider the European market to be the video game market with the most potential for sales and allocate funds appropriately.
- Concentrate on the genres that are most popular locally. Budgeting for action, sports, and shooter games should take precedence over other genres. However, role-playing games receive greater funding in Japan.
- The most popular gaming platforms are PS4 and XOne.



[Click below to view detailed Presentation & Report](#)





Medical Staffing Agency

Preparing for Influenza Season in US

Project Brief



Project Goals

- Assist medical staffing agency in planning for influenza season. Examine trends in influenza and how they can be used to plan for staffing needs across the country
- Determine when to send staff, and how many, to each state in US.



Data

- [Influenza deaths by geography, time, age, and gender](#) obtained from [CDC](#)
Note: Death counts below 10 are suppressed by CDC
- [Population data by geography](#) obtained from US Census Bureau
- [Cleaned & Integrated Dataset](#)



Tools used

- Excel
- Tableau

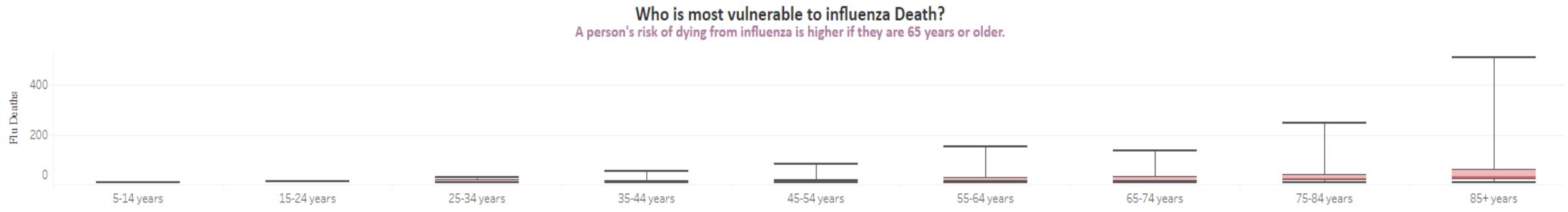


Skills

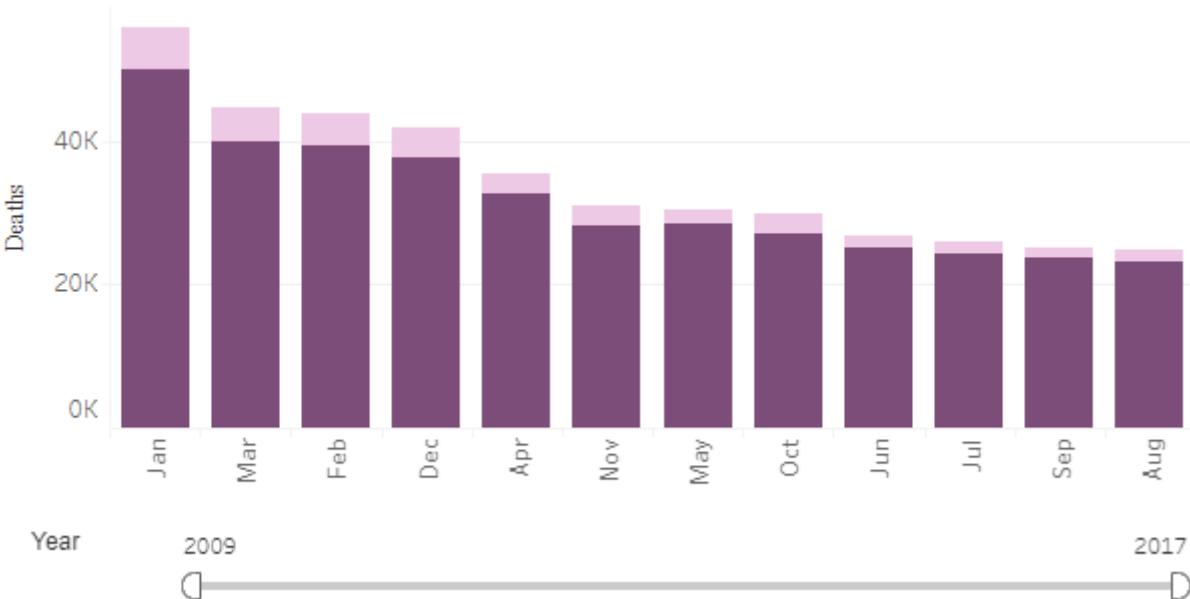
- Translating Business Requirements
- Data Cleaning, Integration, Transformation
- Statistical Hypothesis Testing
- Forecasting
- Data Visualization & Storytelling in Tableau

Analysis

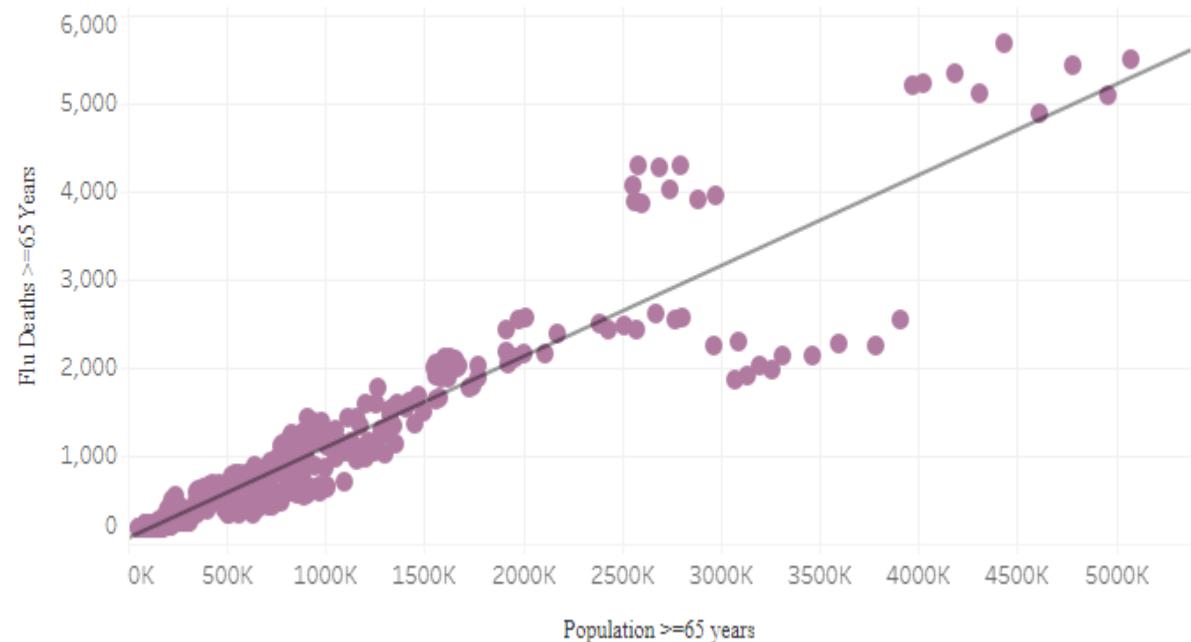
Who is most vulnerable to Influenza?



Influenza deaths - Vulnerable Population (65 or older) Vs Non Vulnerable Population (below 65)



Positive Correlation Between Population & Deaths - 65 or older

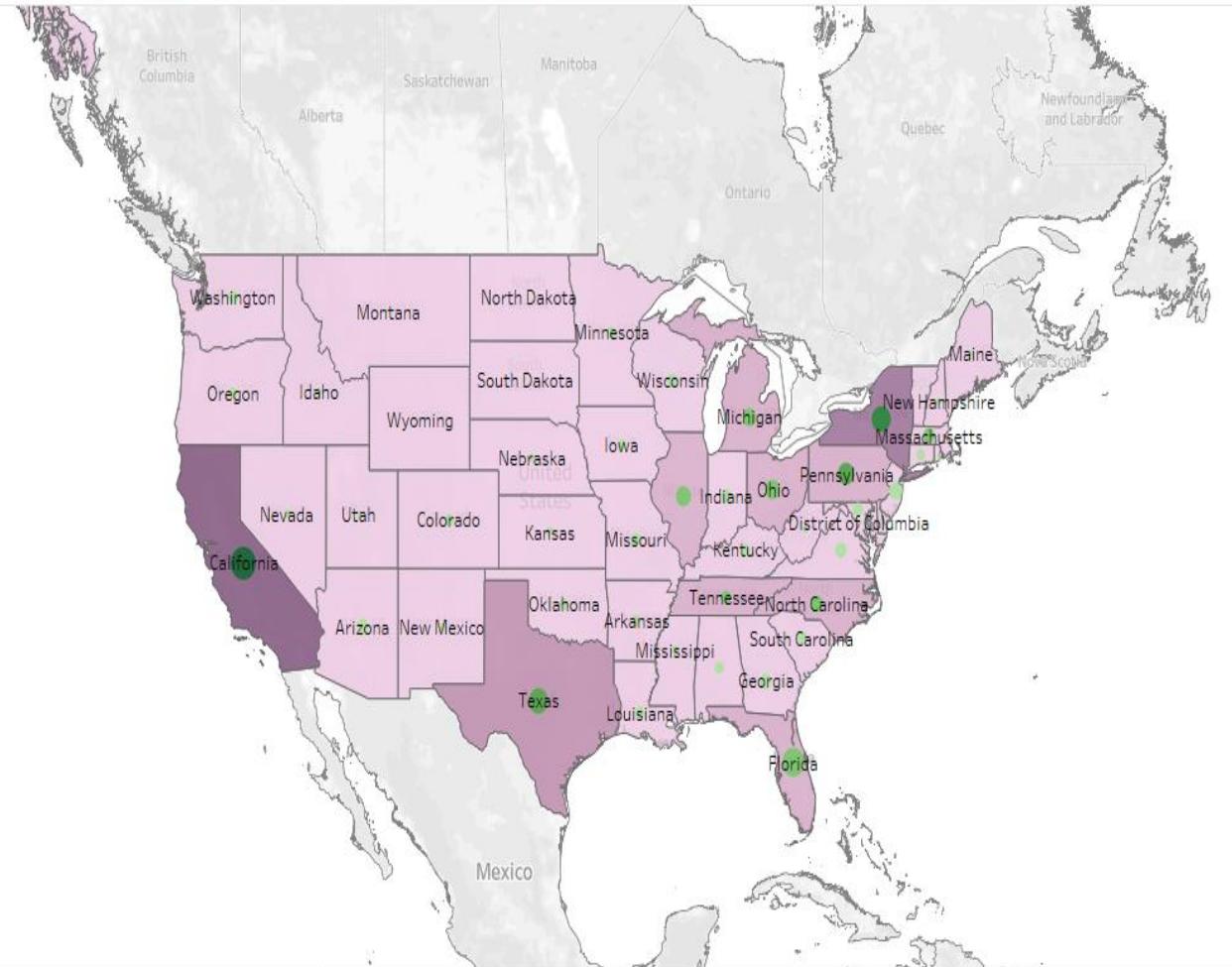


What states need staffs more?

Where are the most staffs needed?

States with highest number of Influenza deaths(Total & 65+) and 65+ population

The darker the state, the higher the Flu deaths. The bigger the circle, higher the 65+ population and darker the circle, higher the 65+ Flu deaths.



© 2023 Mapbox © OpenStreetMap

Total Flu death

Flu Deaths >=65 Years

Population >=65 years

5,940

57,162

1,620

47,483

516,575

10,000,000

20,000,000

30,000,000

40,346,770

States with highest number of Influenza deaths

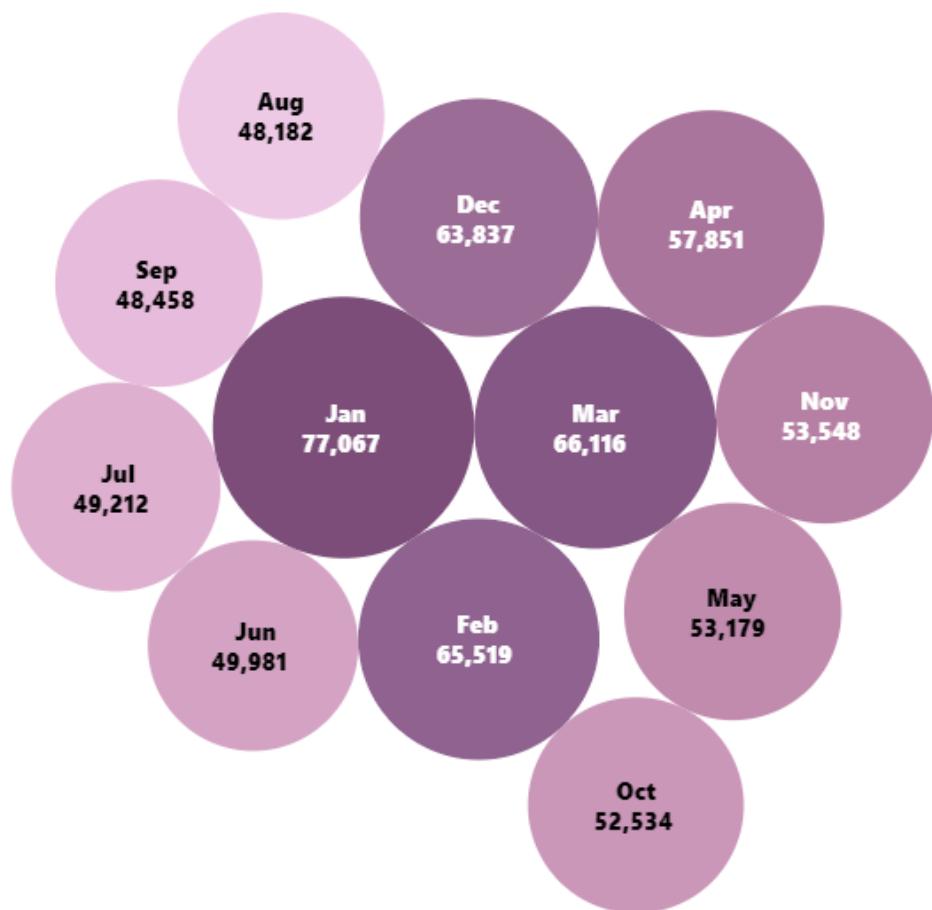
	Flu death Total	Flu Deaths >=65 Years
California	57,162	47,483
New York	44,215	36,576
Texas	30,049	22,140
Pennsylvania	26,109	20,596
Florida	25,624	19,148
Illinois	23,689	18,019
Ohio	22,702	16,866
North Carolina	19,392	13,941
Michigan	18,145	12,962
Tennessee	16,317	11,086
Massachusetts	15,658	11,187
Georgia	15,585	10,262
Virginia	14,755	10,074
New Jersey	14,238	9,646
Missouri	14,225	9,626
Indiana	12,381	7,836
Wisconsin	12,061	7,652
Maryland	12,057	7,587
Alabama	11,920	7,178
Kentucky	11,381	6,842
Louisiana	10,259	5,702
Arizona	10,259	5,730
Washington	10,157	5,692
South Carolina	9,755	5,289
Arkansas	9,499	5,129
Oklahoma	9,464	4,963
Connecticut	9,397	5,072
Minnesota	9,362	5,026
Mississippi	9,317	4,830
Iowa	9,285	4,947
Kansas	9,107	4,774
Colorado	8,602	4,194
Nevada	7,959	3,531
Hawaii	7,800	3,480
Oregon	7,736	3,382
West Virginia	7,629	3,291
Nebraska	7,084	2,764
Utah	6,944	2,624
New Mexico	6,678	2,358
Maine	6,571	2,251
New Hampshire	6,415	2,095
Rhode Island	6,354	2,034
Idaho	6,311	1,991
South Dakota	6,275	1,955
Montana	6,181	1,861
North Dakota	6,049	1,729
Delaware	6,014	1,694
Wyoming	5,981	1,661
Vermont	5,955	1,635
District of Colu..	5,940	1,620
Alaska	5,940	1,620

When are Staffs needed the most?

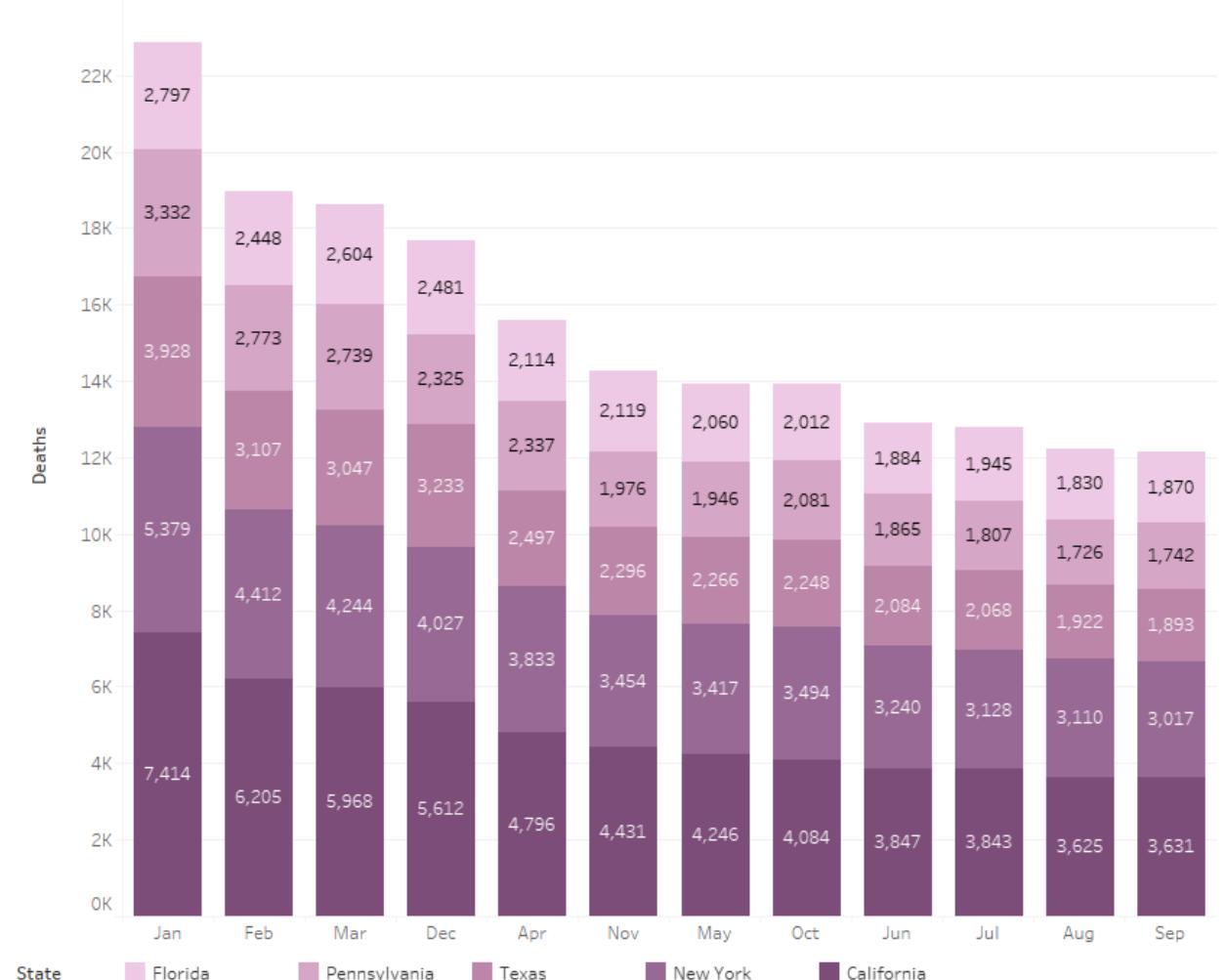
When is Flu season?

The most deaths from influenza occur during the colder winter months.

December, January, February, March



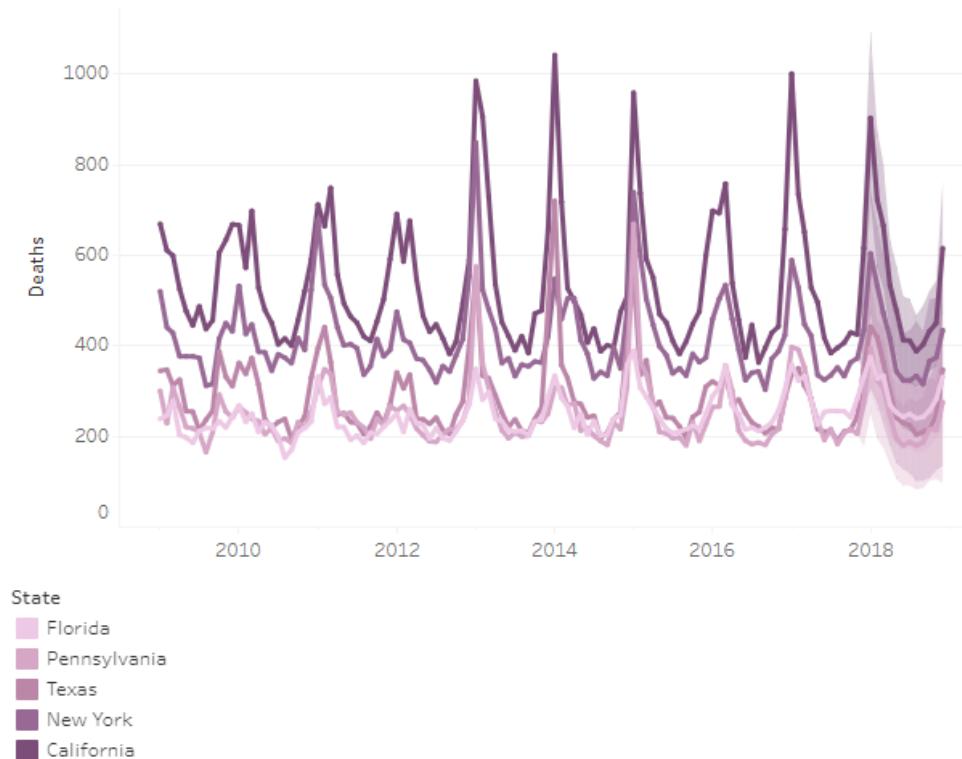
Top 5 states - Influenza deaths by month

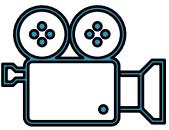


Insights from Data & Recommendations

- The risk of flu-related death is higher in the population aged 65 and above.
- The states with the highest number of flu deaths and greatest number of vulnerable people are California, New York, Texas, Pennsylvania & Florida.
- January, February, March & December have seen the highest number of flu- related deaths between the years of 2009 and 2017.

Influenza death 2018 Forecast
The prediction below indicates that the 2018 flu season should follow the trends from the previous nine years.





Rockbuster Stealth LLC

Answering business questions for an
online video rental company

Project Brief



Project Goals

- Assist business intelligence department with a launch strategy for new online video service by utilizing a relational database management system and SQL to analyze the data and answer business questions posed by the management board.



Data

- [Rockbuster Dataset](#) is provided by CareerFoundry.
- [Data Dictionary](#)



Tools used

- SQL (PostgreSQL)
- Excel
- PowerPoint
- Tableau



Skills

- Extract Entity Relationship Diagram & Create Data Dictionary
- Database Querying
- Filtering ,aggregating & Cleaning Data
- Joining Tables ,Subqueries & Common Table Expression
- Data Visualization & Storytelling

Analysis

Which Countries generate the most Revenue?

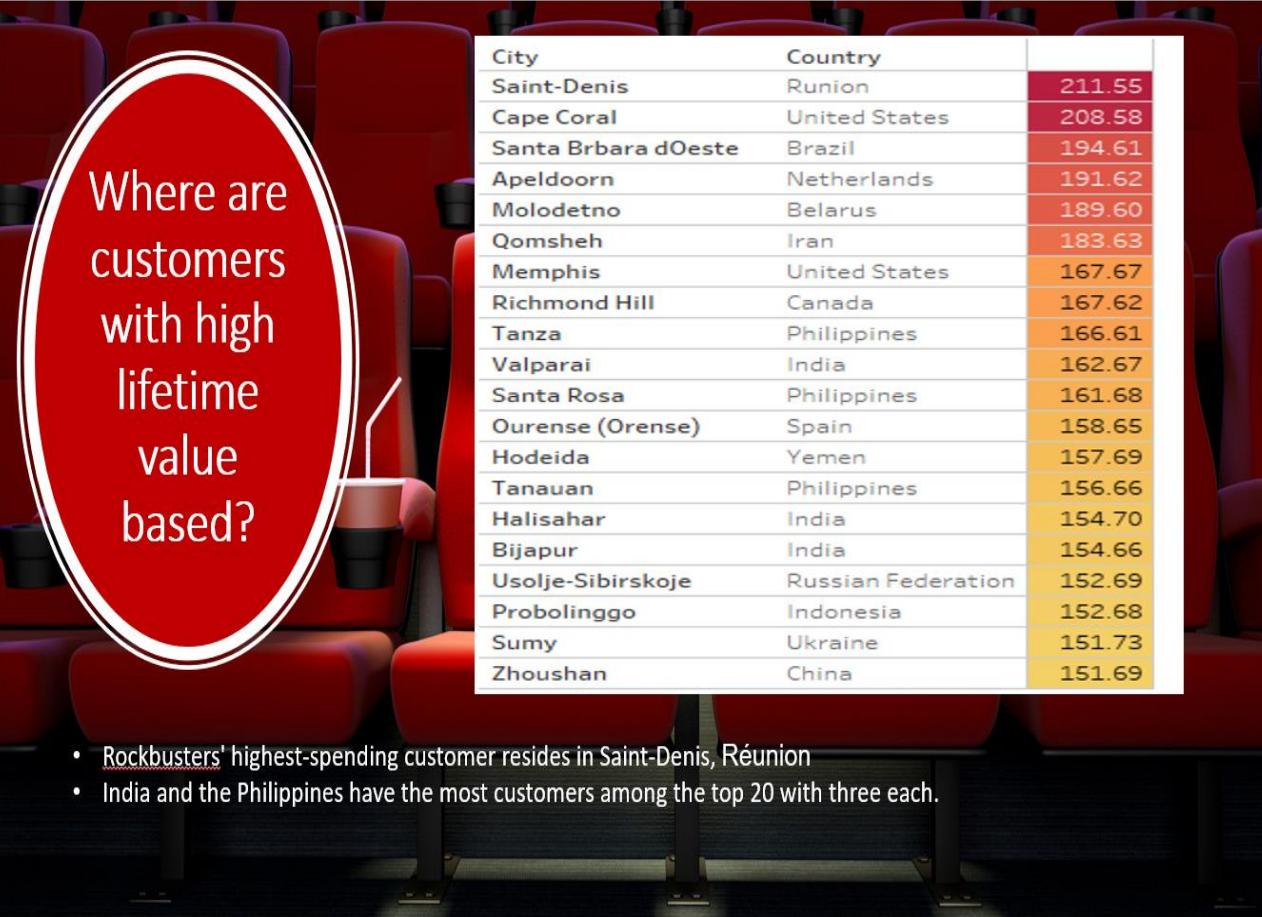


India and China are the two countries that generate the most revenue.

What are the top countries with the most customers?



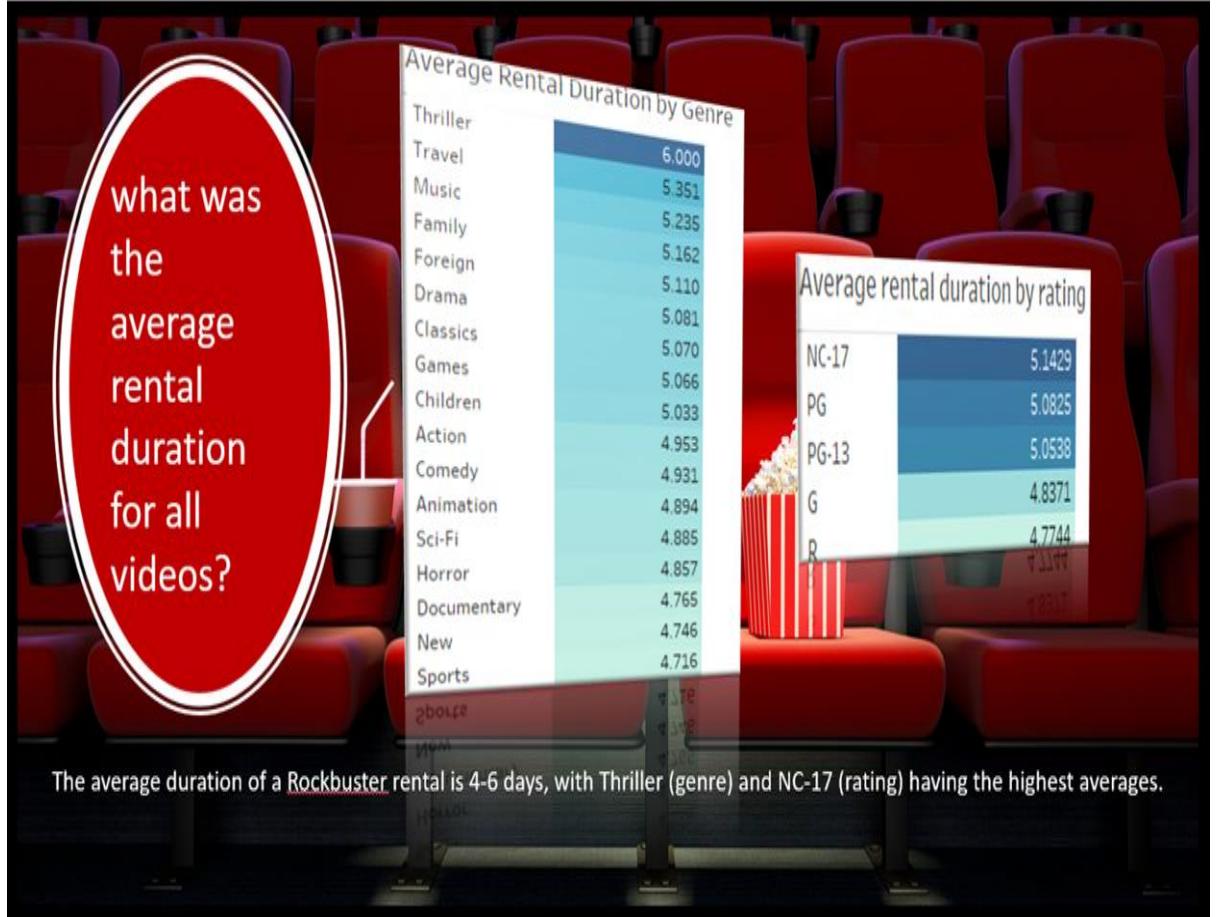
- Rockbuster has locations in 109 countries.
- This map lists the top 10 countries where Rockbuster should focus on promoting its online renting services.
- India should receive the most marketing since it has the largest customer base.



Where are customers with high lifetime value based?

City	Country	
Saint-Denis	Runion	211.55
Cape Coral	United States	208.58
Santa Barbara d'Oeste	Brazil	194.61
Apeldoorn	Netherlands	191.62
Molodetno	Belarus	189.60
Qomshah	Iran	183.63
Memphis	United States	167.67
Richmond Hill	Canada	167.62
Tanza	Philippines	166.61
Valparai	India	162.67
Santa Rosa	Philippines	161.68
Ourense (Orense)	Spain	158.65
Hodeida	Yemen	157.69
Tanauan	Philippines	156.66
Halisahar	India	154.70
Bijapur	India	154.66
Usolje-Sibirskoje	Russian Federation	152.69
Probolinggo	Indonesia	152.68
Sumy	Ukraine	151.73
Zhoushan	China	151.69

- Rockbusters' highest-spending customer resides in Saint-Denis, Réunion
- India and the Philippines have the most customers among the top 20 with three each.



what was the average rental duration for all videos?

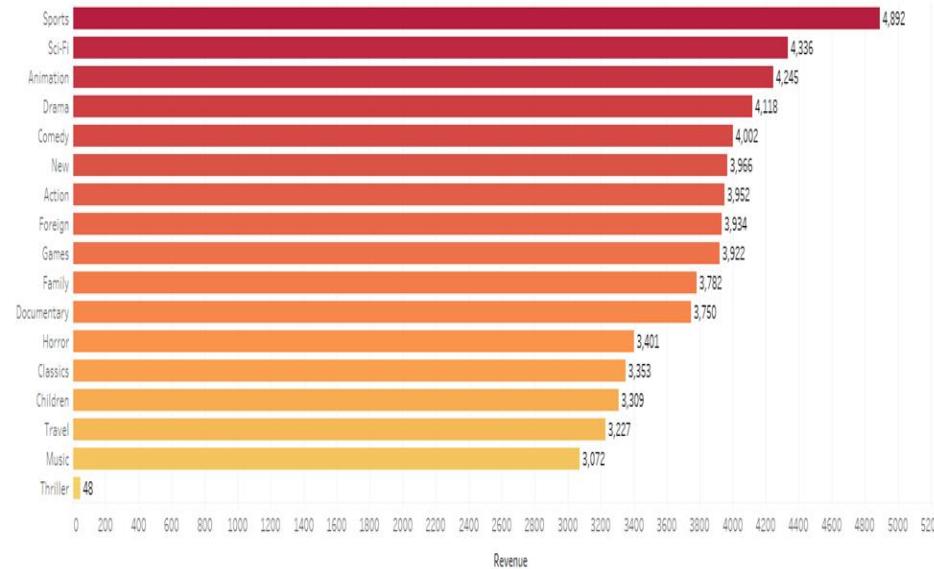
Average Rental Duration by Genre	
Thriller	6.000
Travel	5.351
Music	5.235
Family	5.162
Foreign	5.110
Drama	5.081
Classics	5.070
Games	5.066
Children	5.033
Action	4.953
Comedy	4.931
Animation	4.894
Sci-Fi	4.885
Horror	4.857
Documentary	4.765
New	4.746
Sports	4.716

Average rental duration by rating	
NC-17	5.1429
PG	5.0825
PG-13	5.0538
G	4.8371
R	4.7744

The average duration of a Rockbuster rental is 4-6 days, with Thriller (genre) and NC-17 (rating) having the highest averages.

Which of Rockbuster's Genres are most profitable?

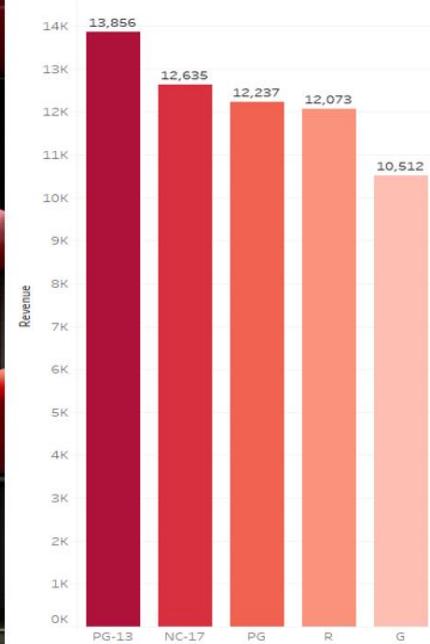
Revenue By Genre



- Sports is the category with the highest revenue at \$4,892.
- The least popular category, thriller, brought in \$48.

Which of Rockbuster's Ratings are most profitable?

Revenue by Rating



- The most profitable classification for Rockbuster is PG-13, which also has the highest revenue of all the ratings

- The least profitable rating is G. This might be an indication that kids' movies aren't as popular as others, based on other ratings.

Insights from Data & Recommendations

- The majority of Rockbuster's customers are from Asia, with India and China having the greatest numbers of customers. Therefore, these regions should be the focus of marketing activities.
- Additionally, Rockbuster needs to focus its customized marketing and promotions in regions with high customer lifetime value.
- Movies about sports, sci-fi, and animation have the largest revenue shares, followed by dramas and comedies. This data can be used to guide Rockbuster's decision regarding which movies to provide on their new online video service.

Click below to view detailed Presentation ,Entity Relationship Diagram,SQL Queries & Visualizations





Instacart

Marketing strategy for an online grocery store

Project Brief



Project Goals

- The task is to perform an initial data and exploratory analysis of some of their data in order to provide insights into how this strategy may be implemented to ensure that Instacart targets the proper client profiles with the right products.



Data

- The dataset comprises of information on the following parameters: Orders, Products & Departments
- The data used is an open-source data by Instacart from 2017 and is available [here](#) and also used a [customer data set](#) created and included for the purpose of this project



Tools used

- Python, Anaconda, Jupyter Notebook
- Libraries (Pandas, Numpy, Seaborn, Matplotlib, Scipy)
- Excel

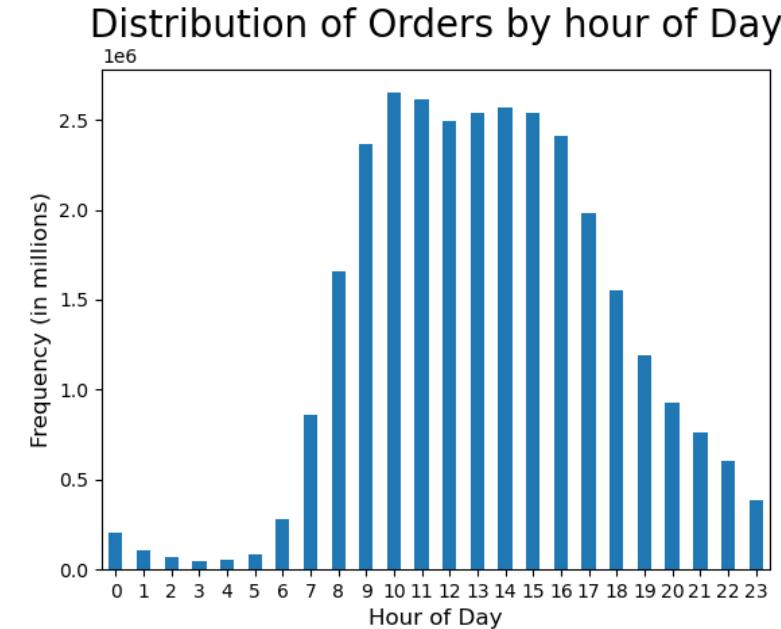
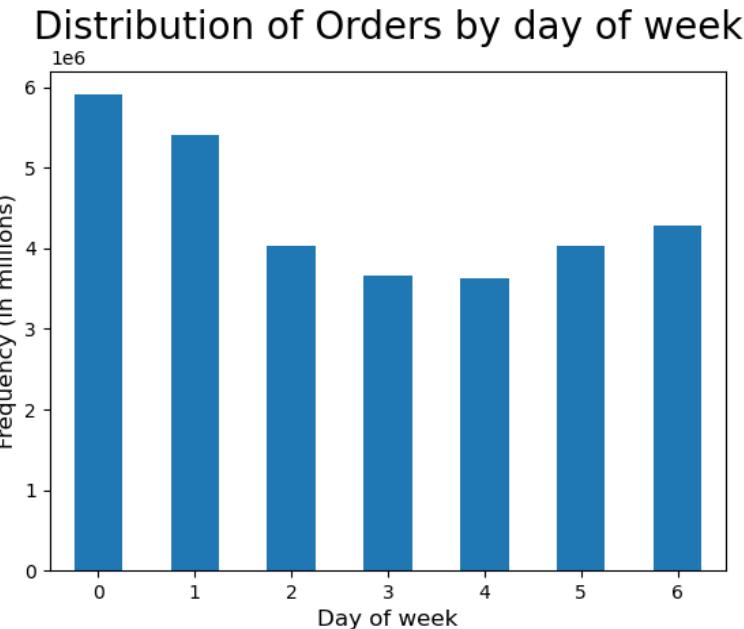


Skills

- Data Wrangling & Merging
- Deriving Variables
- Grouping & Aggregating Data
- Reporting in Excel & Population Flows
- Visualization & Storytelling with Matplotlib & seaborn

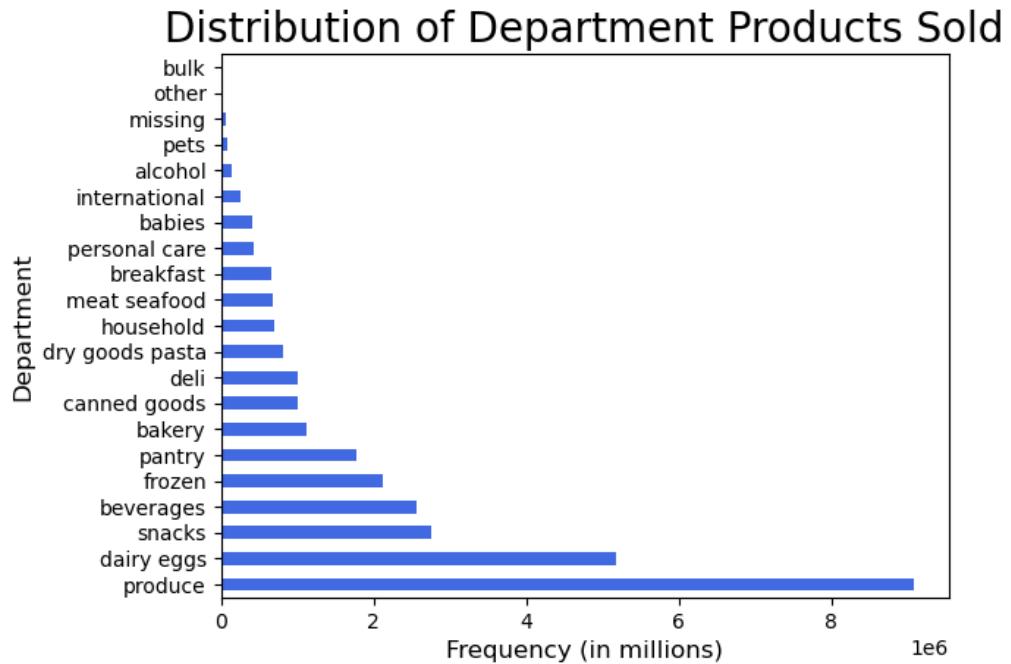
Analysis

What the busiest days of the week and hours of the day?



- The busiest days of the week are Saturday and Sunday.
- The hours between 9am and 4pm had the highest number of orders, with 10am being the peak hour for orders

Products that are more popular than others?

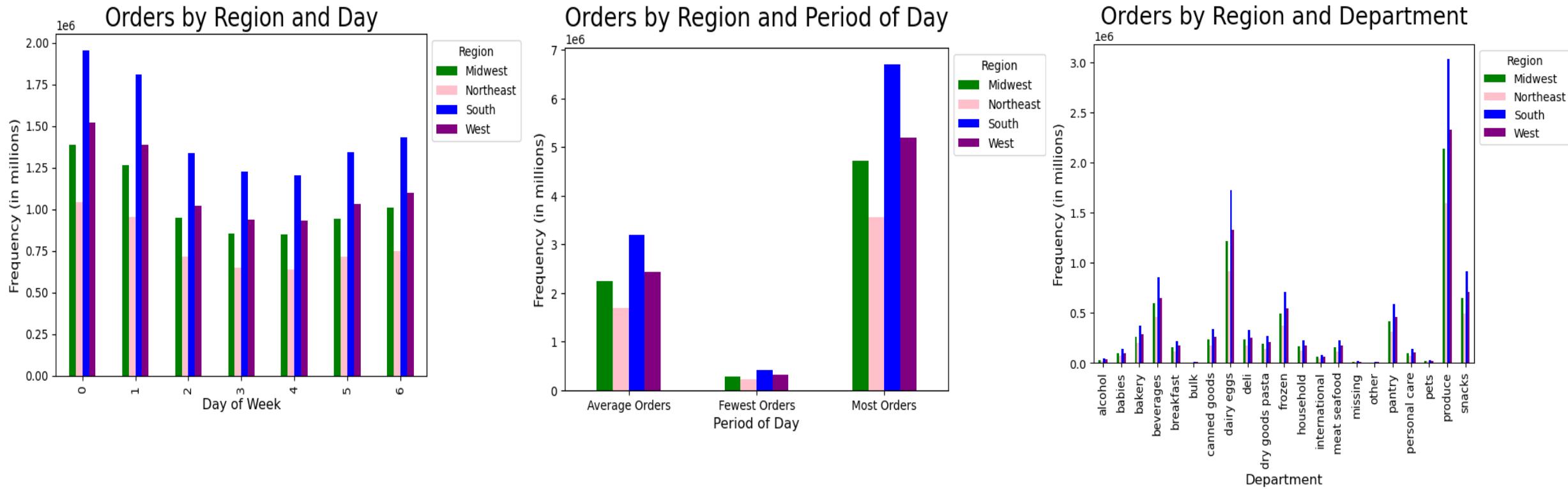


produce	9079273
dairy eggs	5177182
snacks	2766406
beverages	2571901
frozen	2121731
pantry	1782705
bakery	1120828
canned goods	1012074
deli	1003834
dry goods pasta	822136
household	699857
meat seafood	674781
breakfast	670850
personal care	424306
babies	410392
international	255991
alcohol	144627
pets	93060
missing	64768
other	34411
bulk	33451

The following are the top 5 departments in terms of the frequency of product orders:

- **Produce**
- **Dairy eggs**
- **Snacks**
- **Beverages**
- **Frozen**

Orders by Region



- The **South region** places the most orders, followed by the West region. Northeast region places the least orders.
- The orders placed across the day, the hours, and the department have not significantly changed buyer's purchasing habits.

Insights from Data & Recommendations

- Busiest days of the week are Saturday and Sunday
- Busiest hours of the day are between 9am to 5pm
- Top performing departments are Produce and Dairy Eggs
- Southern region has the highest amount of purchasing orders
- To encourage people to keep using Instacart's services, a loyalty/rewards program needs to be established. Based on the factors mentioned, it is also recommended to improve advertising and promotions.
- To ensure Instacart's sustained success, a further marketing campaign for certain criteria will result in an increase in purchases.

Click below to view detailed Report, Scripts & Visualizations





Bank Customer Segmentation

Analysis of Customer Demographics & Transaction data

Project Brief



Project Goals

- Identify distinct groups of customers based on their behavior or other traits. It enables efficient marketing resource allocation, the maximization of point-based approach to each customer group, as well as sales opportunities. It helps in enhancing customer service and retaining customers which are crucial for the banking industry.



Data

- The data was collected as a part of research project in collaboration with a bank. The dataset includes Customer demographics and transactions data from an Indian Bank.
- This data is an open-source data and downloaded from [Kaggle.com](https://www.kaggle.com).



Tools used

- Python, Anaconda, Jupyter Notebook
- Libraries (Pandas, Numpy, Seaborn, Matplotlib, Scipy, datetime, calendar, folium, sklearn, pylab, statsmodels.api)
- Tableau



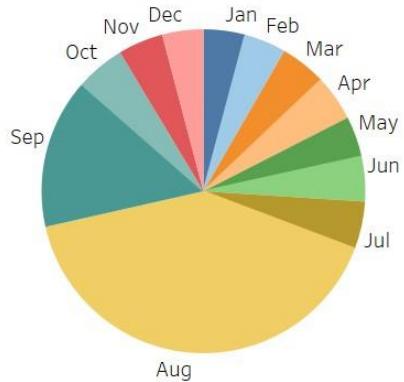
Skills

- Data Sourcing, Wrangling & Deriving Variables
- Exploratory Visual Analysis
- Geospatial Analysis & Time-Series Analysis
- Machine Learning –Linear regression & Kmeans Clustering
- Visualization & Storytelling

Exploratory Analysis

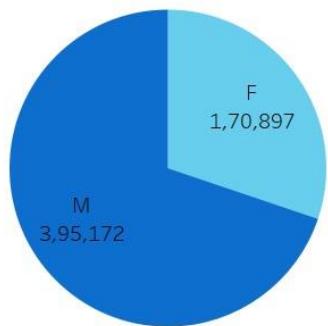
Month with the most and expensive transactions

The most transactions, as well as the most expensive transactions, occur in August.



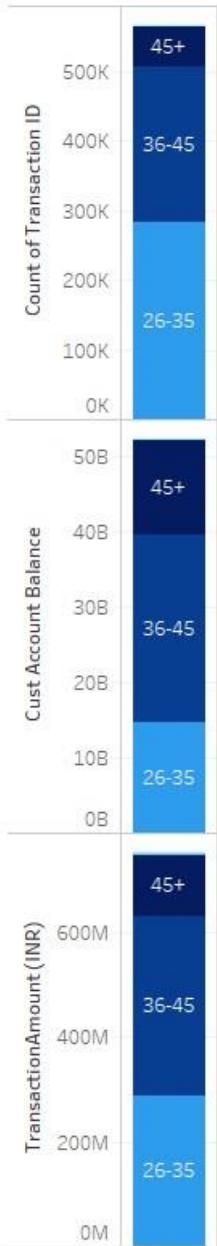
Gender Categorization

Comparing male and female customers, male customers are slightly more likely to spend money, have greater account balances, and make more transactions.



Age Categorization

Age groups **26 to 35** have more transactions than other age groups, while **36 to 45** have more expensive transactions and higher account balances than other age groups.

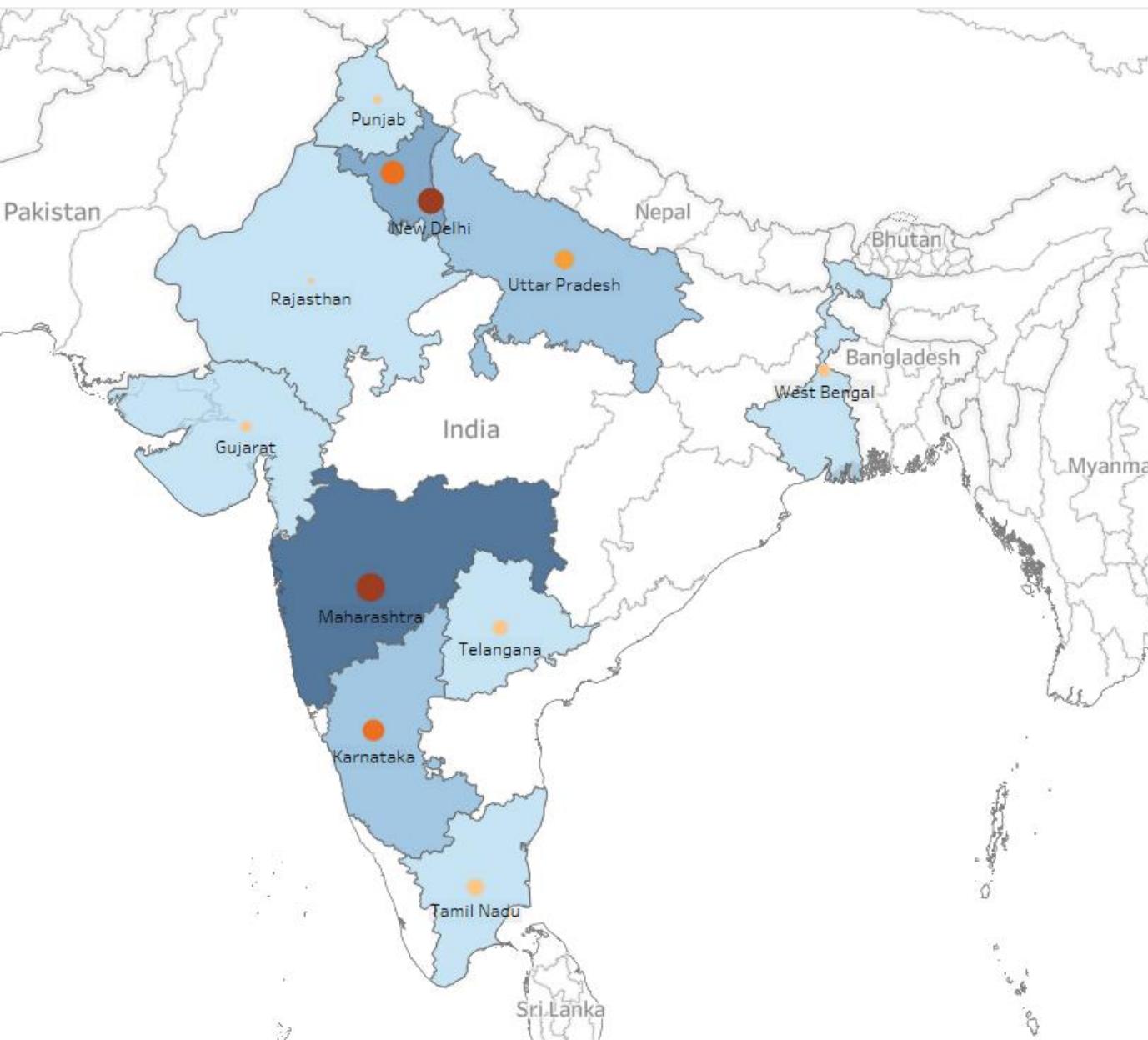


GeoSpatial Analysis

States with high customer balances, high customer spending, and high transaction volumes

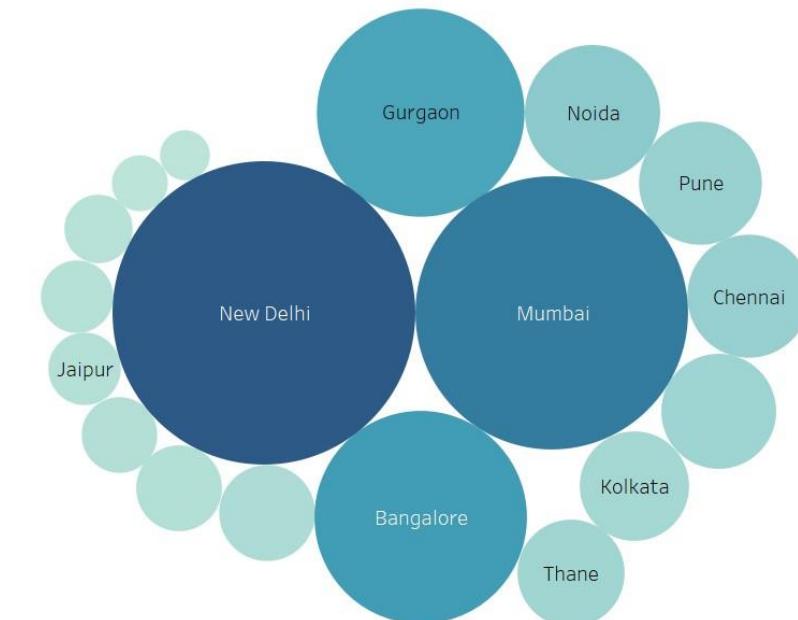
The darker the state, the higher the customer account balances.

The bigger the circle, higher the customer spending and darker the circle, higher the transaction volume.



Cities with high spending customers and high transaction volume

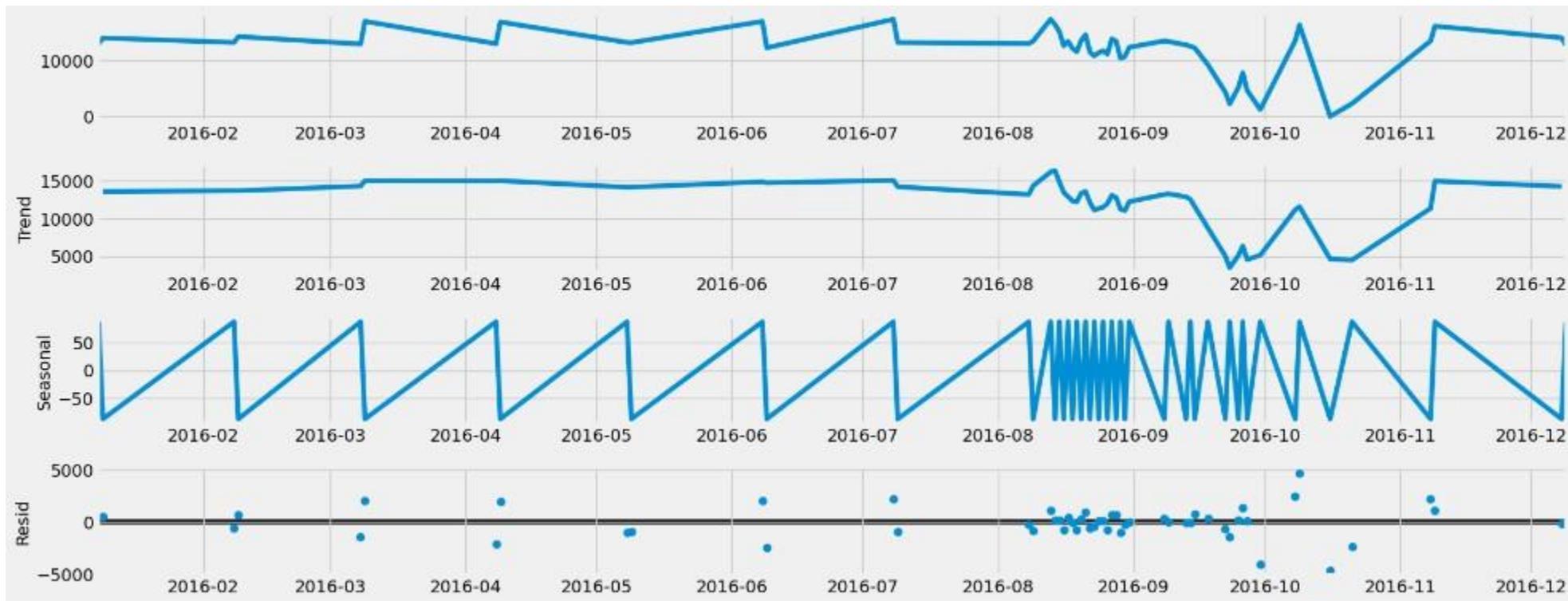
The darker the state, the higher the transaction volume. The bigger the circle, higher the transaction value



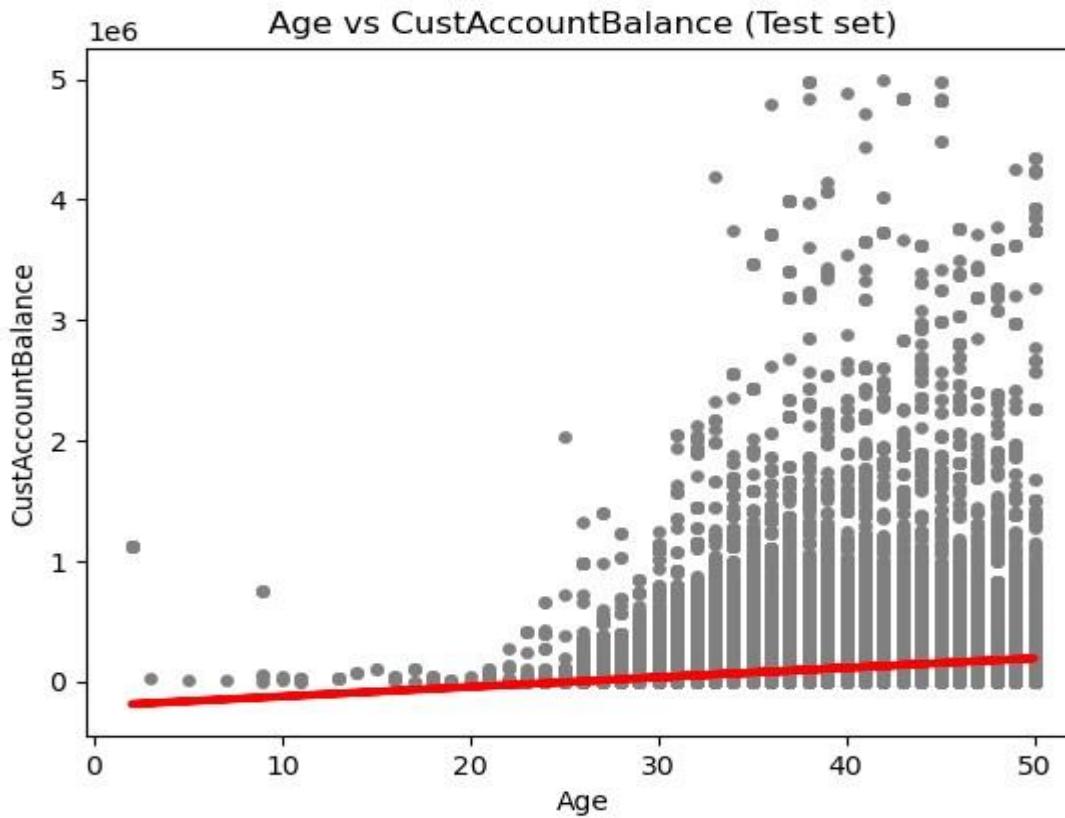
Time Series Analysis

2016 Monthly Transaction Count

The busiest month in 2016 was August, followed by September in terms of transactions.



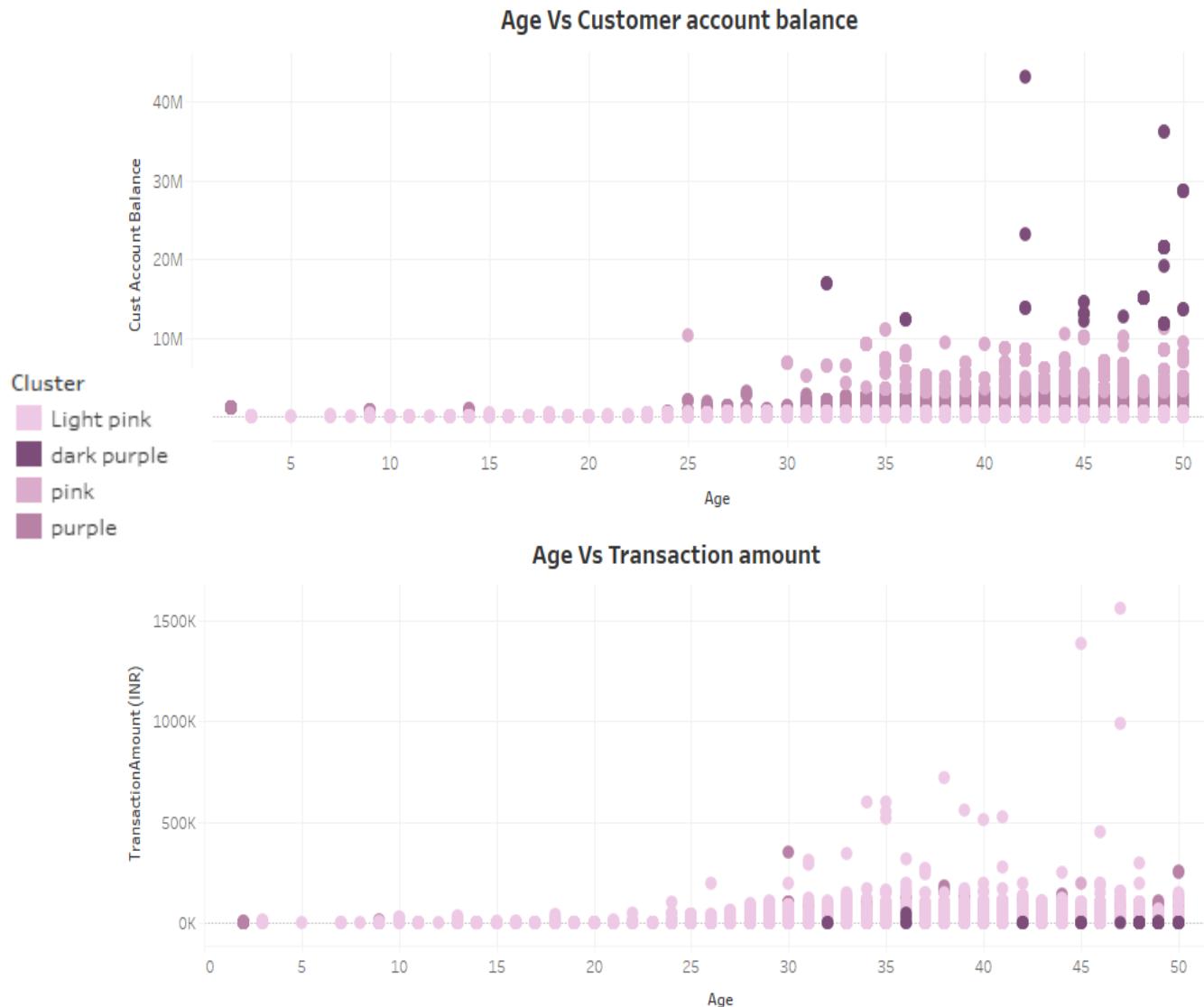
Regression Analysis- Linear Regression



- There is no linear relationship between the variables because most of the data points lie outside of the trend line.
- The R-squared value is always in the range between 0 and 1, with values closer to 1 denoting a regression model that fits the data more closely. Therefore, the extremely low R-squared value of almost zero and the huge Mean Squared Error demonstrate that the Linear regression is not the best model for the dataset.
- Even though the positive slope value indicates a solid association between the variables in this case, the MSE & R2 score value prevents us from utilizing this model to generate predictions.

Clustering Analysis – Kmeans Clustering

Since Linear regression model didn't seem to fit well to the data, the analysis moved on to machine learning using k-means clustering. The optimal number of clusters has been identified to be four using the Elbow technique. After that, a model was created using kmeans clustering and the resulting clusters produced insightful results.



Cluster Analysis Results

- In terms of statistics across all categories, the dark purple cluster outperforms pink, purple, and light pink.
- In comparison to other groups, the dark purple cluster's customers are older, have larger account balances, and do more transactions.
- As a result, it is evident that a customer's account balance and transactional or spending amounts increase with age.

cluster	Age		CustAccountBalance		TransactionAmount (INR)	
	mean	median	mean	median	mean	median
Light pink	36.197638	35.0	4.937166e+04	16385.54	1273.485022	435.56
dark purple	46.563380	48.0	1.794515e+07	15158474.07	2997.812676	2000.00
pink	43.042343	44.0	5.184852e+06	4831162.46	4444.372308	1334.00
purple	40.223630	40.0	1.096728e+06	906665.85	3082.458258	1000.00

Insights from Data

The statistics supported my hypothesis, "Customers' account balances and transactional or spending quantities increase as they get older."



Insights

- Male customers are more likely than female customers to spend money, have higher account balances, and make more transactions.
- August is the month with the most transactions and the most expensive transactions, followed by September.
- The age groups 26 to 35 have the most transactions, whereas the age groups 36 to 45 have the most expensive transactions and the highest account balances.
- The state of Maharashtra has the highest customer balances, customer spending, and transaction volumes, followed by New Delhi and Haryana.
- The cities with the highest spending and transaction volume are New Delhi, Mumbai, and Bangalore.



Limitations

- Since the data was provided by a bank as part of a research project in 2016, the analysis won't take into account recent trends.



Next steps

- The most recent bank transaction data will be updated, and further research will be conducted by adding other variables including the type of transaction.

[Click below to view detailed Report, Scripts & Visualizations](#)



Let's connect and explore how I can help you achieve your business objectives



sruthysreekanth@gmail.com



<https://www.linkedin.com/in/sruthy-sreekanth/>



<https://github.com/sruthysreekanth>



<https://public.tableau.com/app/profile/sruthy.sreekanth>