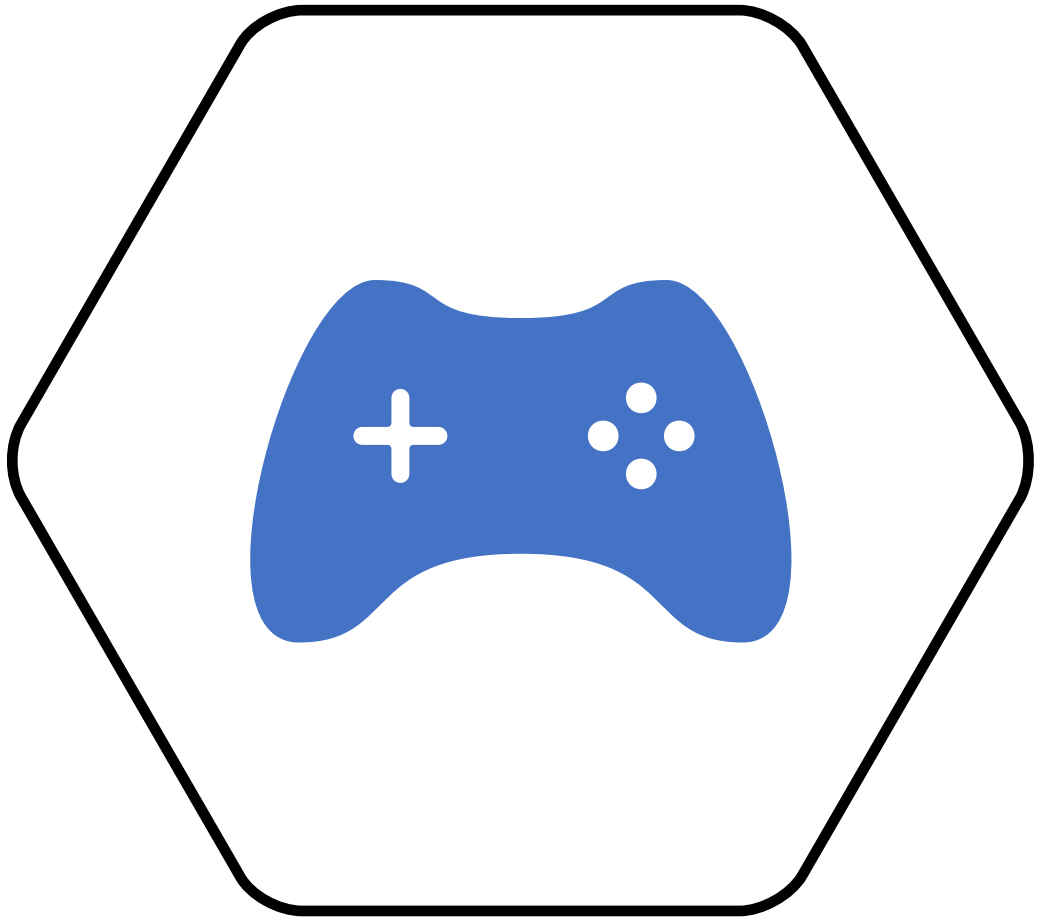
Data Analytics Portfolio

Lionel Eyong

CASE STUDY 1: GAMECO ANALYZING GLOBAL VIDEO GAME SALES



GAMECO: ANALYZING GLOBAL VIDEO GAME SALES

GameCo is a fictional video game seller present globally with its strongest markets in North America, Europe, and Japan. The goal of the analysis was to uncover insights from historical sales data to recommend changes in the marketing budget.

DATA

- Historical sales of video games spanning different platforms, genres, and publishing studios.
- Source: Video Game Sales

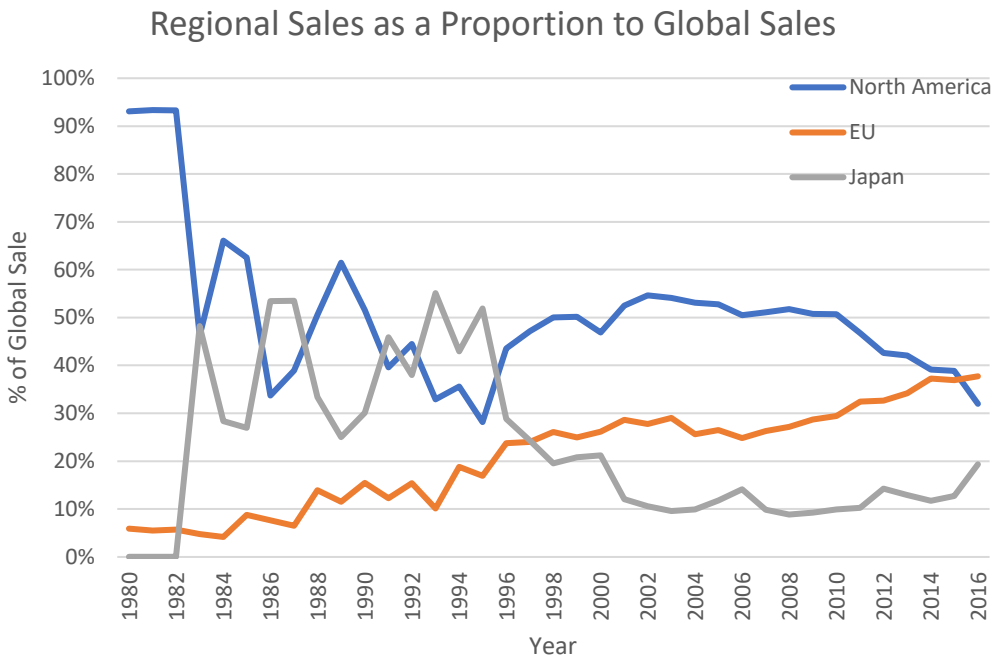
TOOLS:

- Excel
- PowerPoint

SKILLS USED

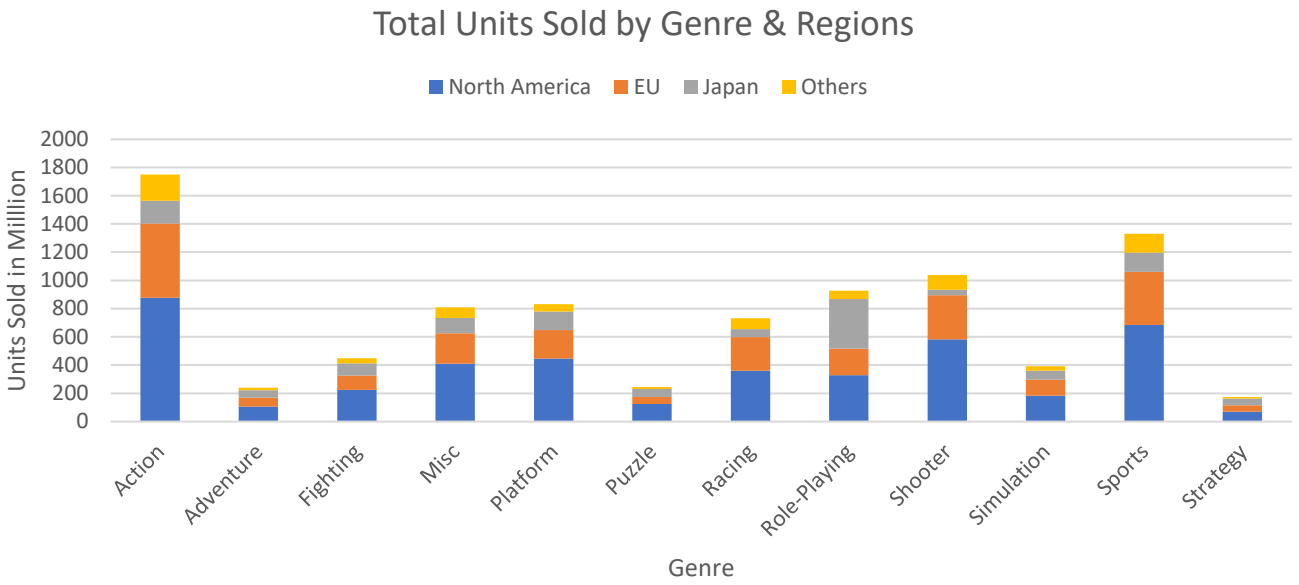
- Data cleaning
- Pivot tables
- Grouping & Summary
- Descriptive analysis
- Data visualization and storytelling

Sales as a Percentage of Global Sales from 1980-2016.



- Sales have decreased since 2008. In terms of market shares
- EU Sales has taken over North America Sales in 2016.

Total Units Sold in Millions by Genre and Region



It is clear that certain genre are more popular than other in specific geographical region
For e.g for

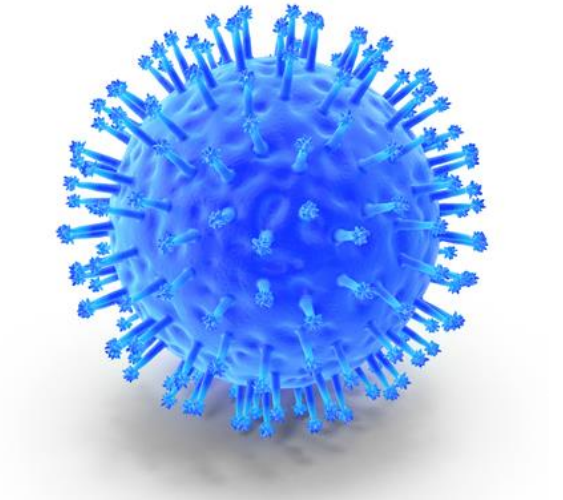
- North America the most popular genre is Shooter games
- In Europe their seems to be a preference for Action games
- Japan has developed a particular taste for role-playing games

CONCLUSIONS & RECOMMENDATIONS

- Decrease in retail sales since 2008 may be due to a shift to online sales
- The EU has overtaken North America as the biggest market in 2016
- Top Genres in North America and the EU are Action, Shooter, Sport and Role-playing
- Top genres in Japan are Role-playing, Action, and Misc
- GameCo should focus on marketing the most popular genres in each region

Click to see full presentation

CASE STUDY 2: PREPARING FOR FLU SEASON IN THE US



PREPARING FOR FLU SEASON IN THE US

The project was motivated by the fact that the United States experiences flu seasons when more people than usual suffer from the illness. The medical staffing agency that provides temporary workers to clinics and hospitals needs to develop a staffing plan for the upcoming flu season.

DATA

- Influenza deaths by geography, time, age, gender from the CDC (Dataset)
- Population data by geography from the US Census Bureau

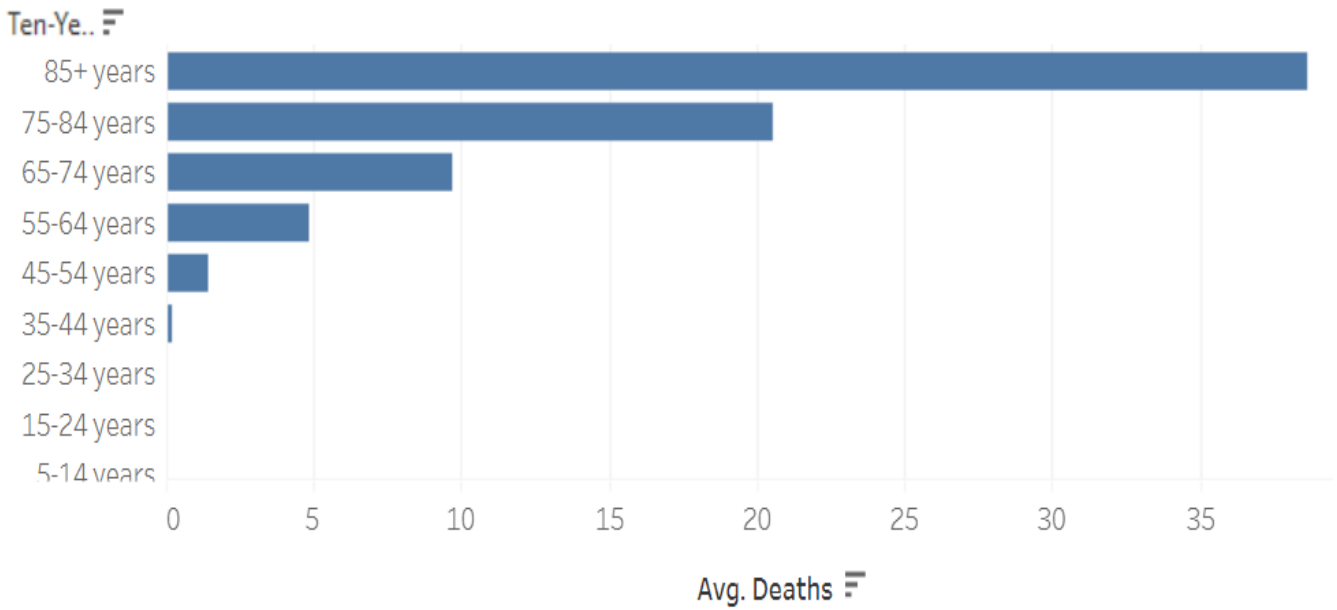
TOOLS

- Excel
- Tableau

SKILLS USED

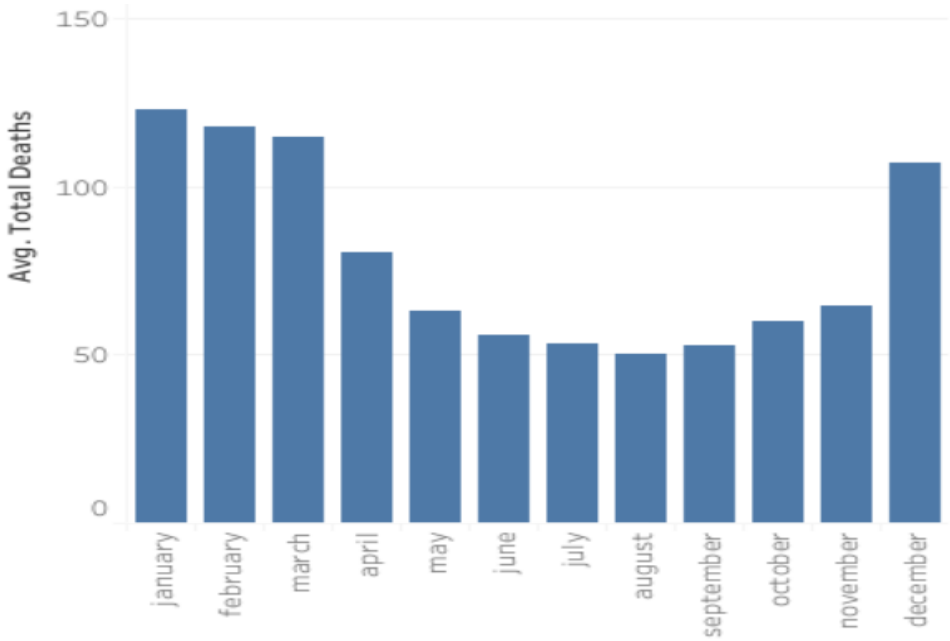
- Data Cleaning
- Data Integration & Transformation
- Statistical Hypotheses Testing
- Visual Analysis & Forecasting
- Story Telling in Tableau

Total Average Death Per Age Group



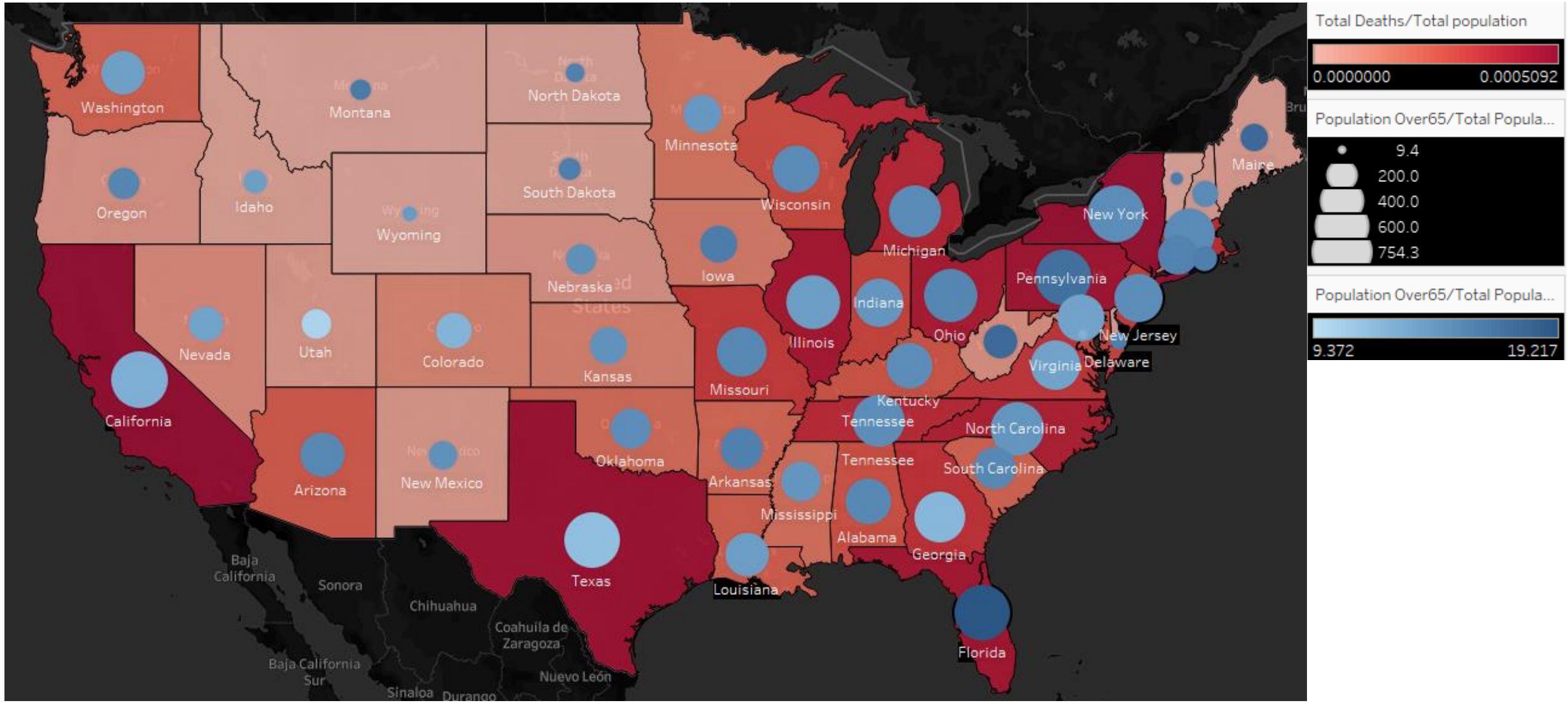
- Age group that mostly affected by influenza are 65 and above, they account for most fatalities
- Young age population (below 34 yrs) almost minimal average death

Total Average Death Per Month



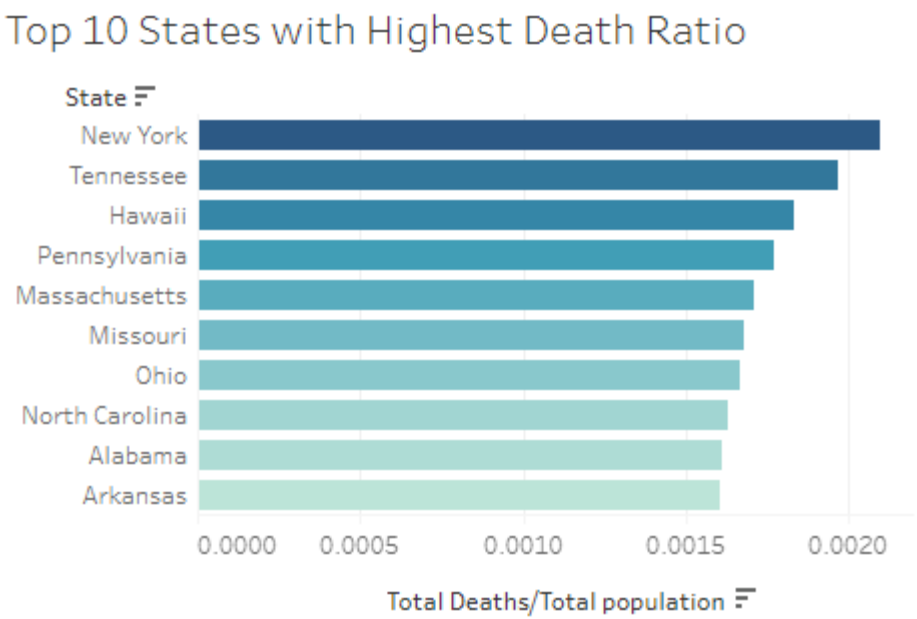
- Influenza is seasonal
- Influenza hit the most from December to March
- It is the worst in January with most casualties

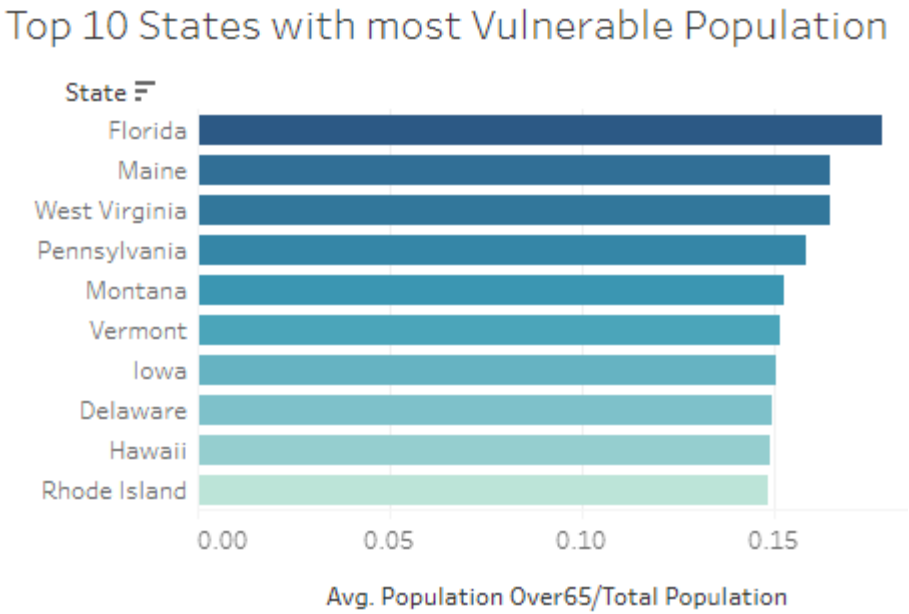
Vulnerable Population and Average Total Death Per State



State with higher number of population over 65 are more likely to have higher mortality

CONCLUSIONS & RECOMMENDATIONS



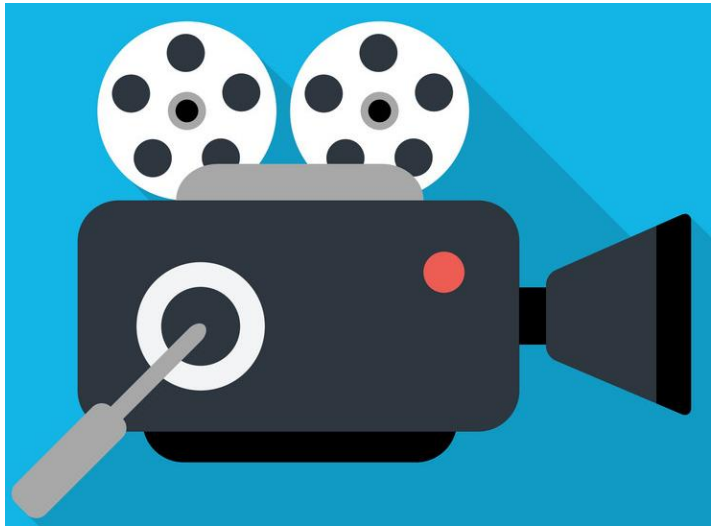


- Those states need to be prioritized as they have the highest death ratio and highest number of vulnerable population.
- Seasonality should be taken into consideration to allocate staff when most needed

Click to see Tableau Storyboard

Click to see video Presentation

CASE STUDY 3: ROCKBUSTER STEALTH DATA ANALYSIS



ROCKBUSTER STEALTH DATA ANALYSIS

This project focused on using SQL to analyze the data of a fictional movie rental company (Rockbuster Stealth) in order to help them successfully launch their own streaming service.

DATA

- The dataset is a relational database containing information about Rockbuster's film inventory, customers, and payments

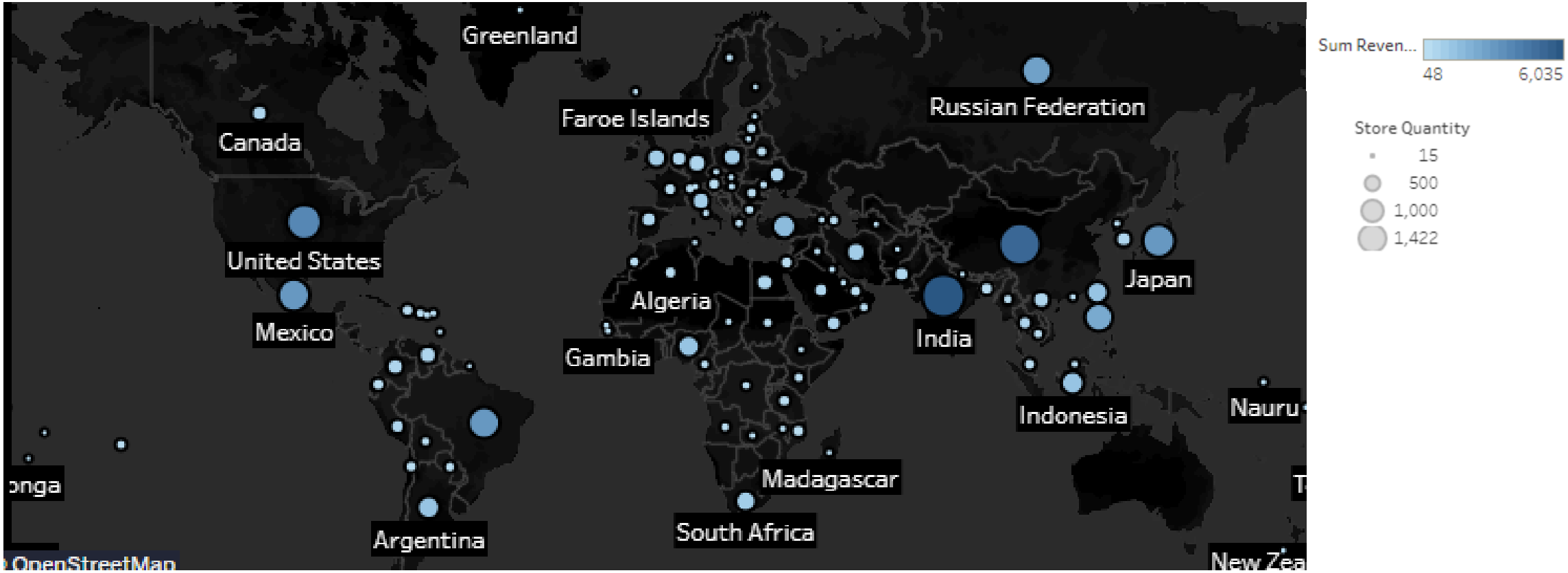
TOOLS

- PostgreSQL
- DB Visualizer
- Tableau

SKILLS USED

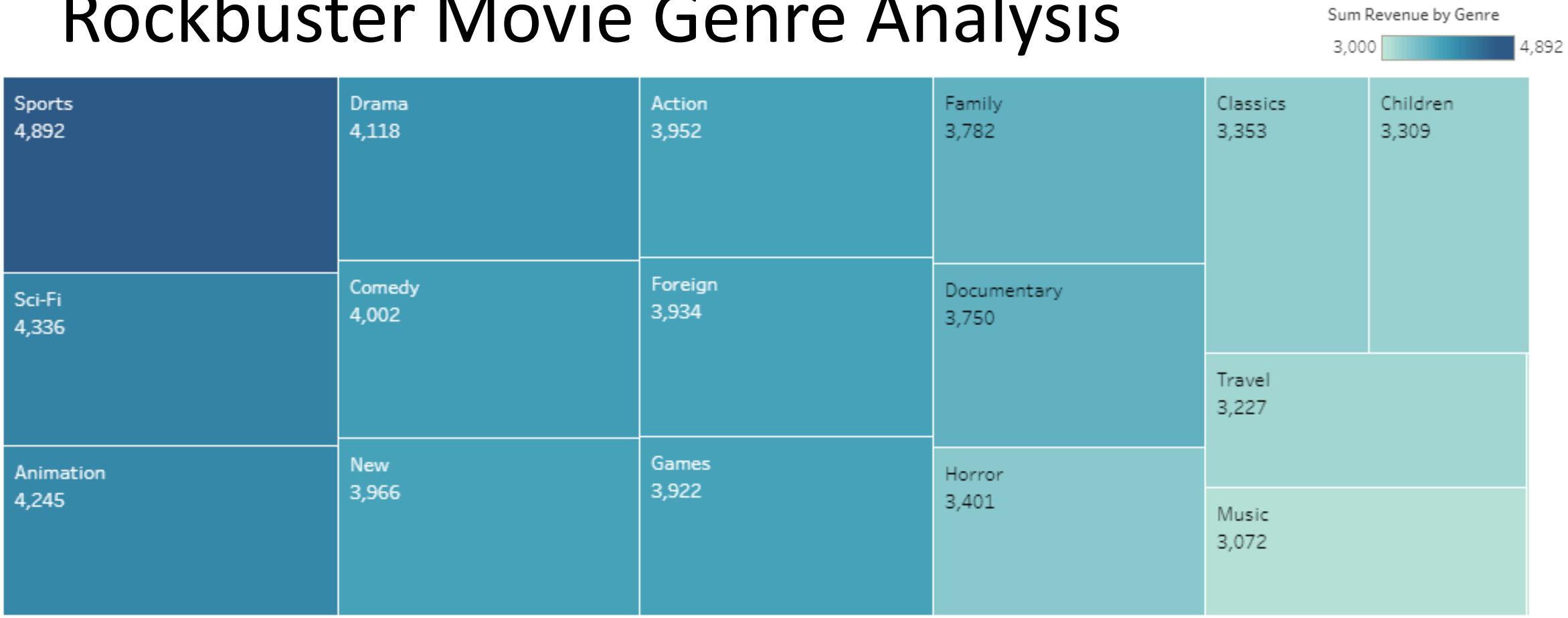
- Cleaning data
- Filtering data
- Joining tables
- Subqueries
- Common Table Expression
- Creating visualizations

Rockbuster Store Presence around the world



- Rockbuster Stealth has store presence in 161 countries around the world
- India has the highest number of stores (885) and revenue (3810 M)
- Switzerland has the lowest number of stores (13)
- American Samoa has the lowest number of revenue (47.85 M)

Rockbuster Movie Genre Analysis



Movies Sum Revenue by Genre

- Sports make the highest contribution to revenue, followed by Sci-Fi and Animation
- Worst performing Genre is Music, followed by Travel and Children

CONCLUSIONS & RECOMMENDATIONS

Rockbuster should prioritise:

- Top 5 Countries with highest revenues: India, China, US, Japan and Mexico
- Top 5 Genres with highest revenue: Sports, Sci-Fi, Animation, Drama and Comedy.
- Top 5 Countries with highest customer base: India, China, US, Japan and Mexico

Should also consider a loyalty scheme, to reward regular customers and retain existing customers thus maintaining a sustainable growth

CASE STUDY 4: INSTACART GROCERY BASKET ANALYSIS

INSTACART GROCERY BASKET ANALYSIS WITH PYTHON

Instacart is an online grocery store that operates through a mobile application. The stakeholders are most interested in the variety of customers in their database along with their purchasing behaviors to implement a targeted marketing strategy.

DATA

- Open-sourced from Instacart (Customer Data Set)
- Some of the datasets contain over 32M observations

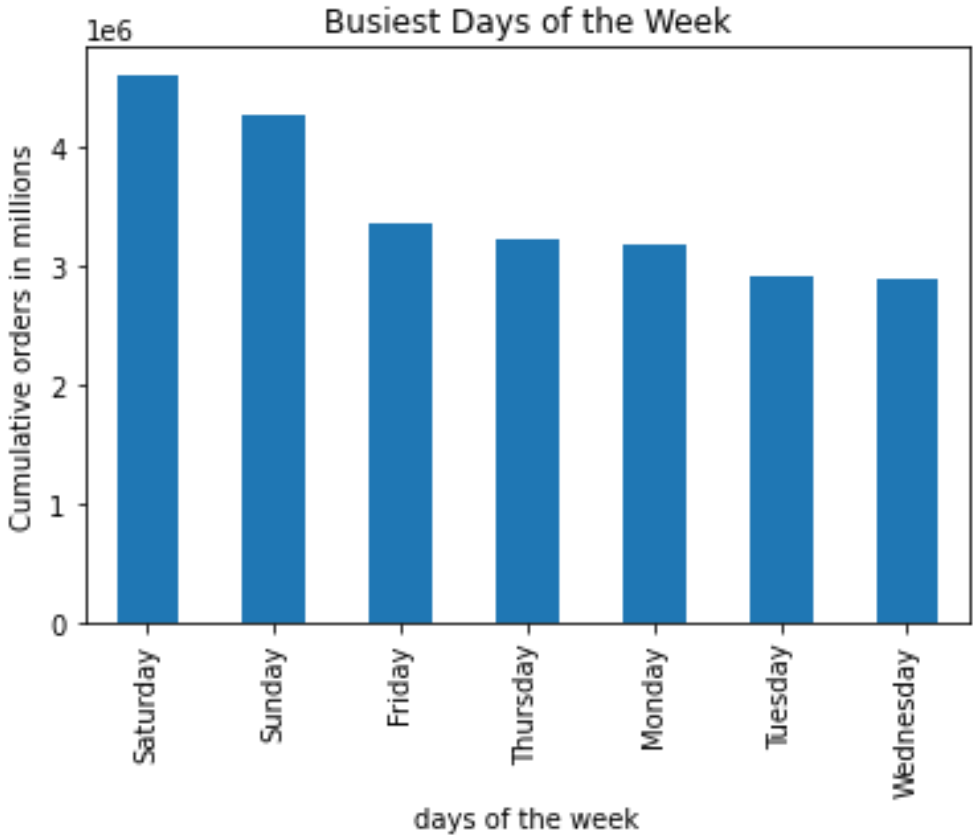
TOOLS

- Python and Jupyter
- Libraries: Pandas, Numpy, Matplotlib, Seaborn, Scipy

SKILLS USED

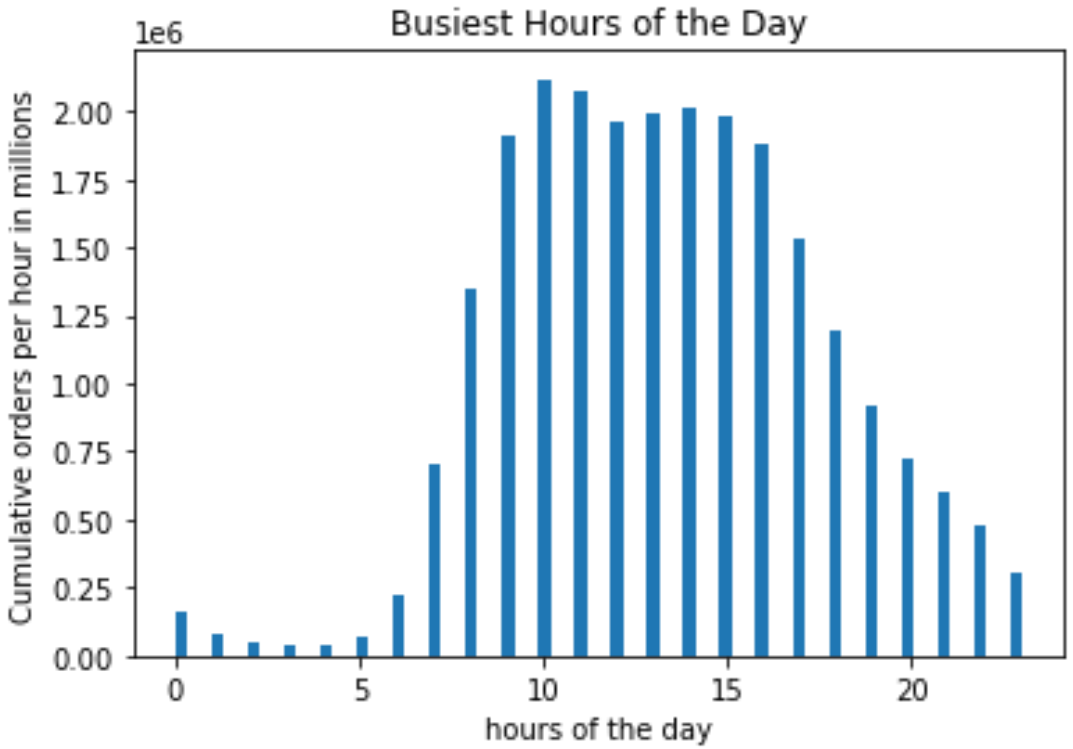
- Cleaning data
- Merging datasets
- Deriving new variables
- Grouping & aggregating data
- Creating visualizations

Busiest Days of the Week and Busiest Hours of the Day



Frequency of Orders by week

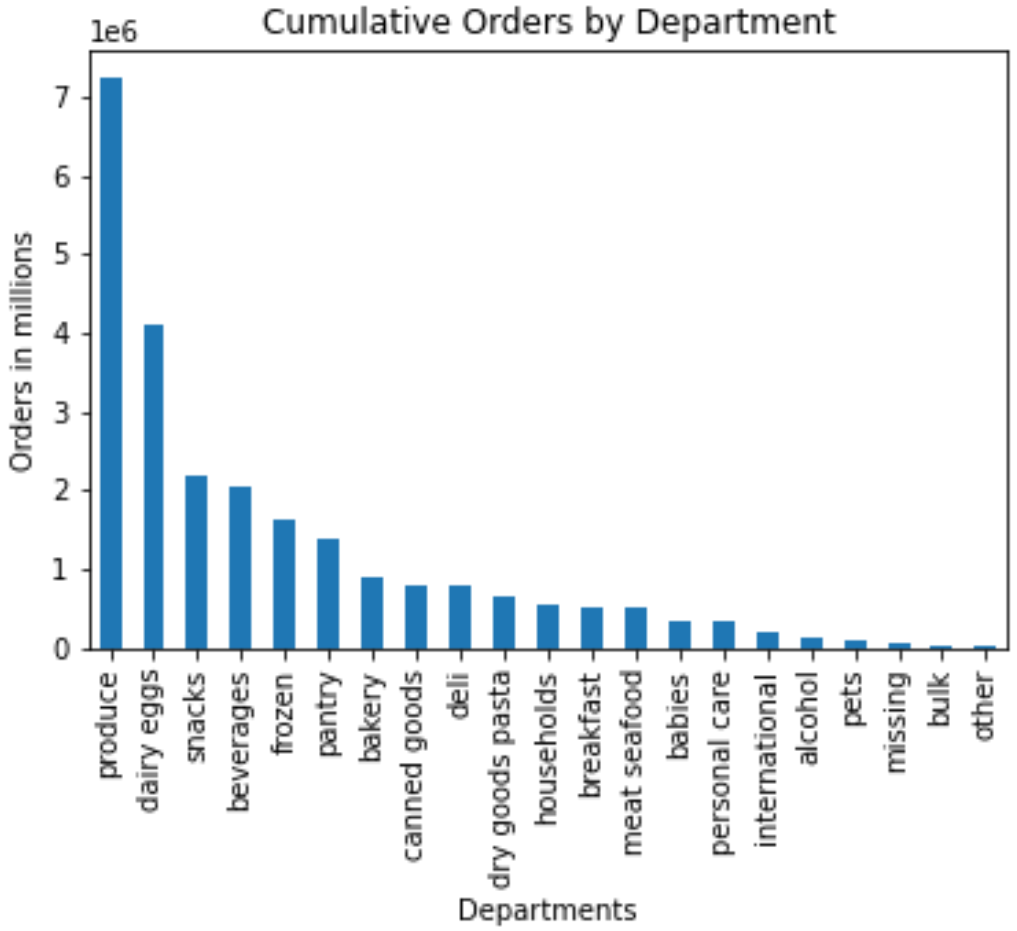
The busiest days of the week are the weekends (Saturday and Sunday) followed by Friday.

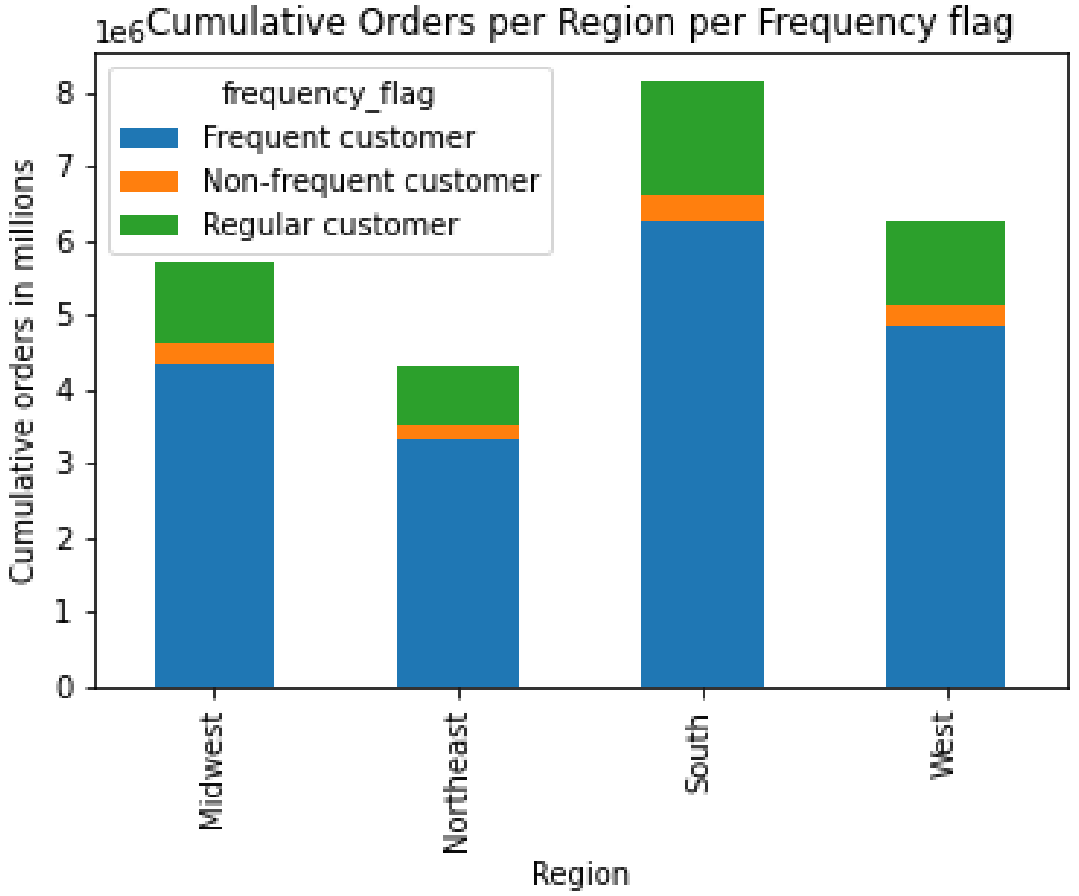


Frequency of Orders by the hour

The busiest hours of the day are between 10 am and 16.

Order by Department & Customer Profile





- Department with the most orders are Produce, followed by dairy eggs and Snacks

- Most popular Region is the South
- Most customers are Frequent Customers

CONCLUSIONS & RECOMMENDATIONS

The majority of Instacart customers are :

- between the age of ‘20s and ‘70s
- married and have at least one dependent
- regular and loyal customers
- firstly from South region, followed by West, Midwest and finally Northeast

Instacart should consider target marketing and introduce a loyalty scheme

GitHub Click to see full presentation

CASE STUDY 5: RENTAL APARTMENT IN GERMANY PRICES ANALYSIS

RENTAL APARTMENT IN GERMANY PRICES ANALYSIS

In this project, I will use scraped data from immobilienscount24, the biggest real estate platform in Germany, available on Kaggle to analyze what is it that informs the price of rental apartments in Germany. I conducted EDA to understand the dataset and I used regression ML technique to see if we can predict the total rent amount. Then, Kmeans unsupervised ML techniques is applied to identify clusters within the dataset. Furthermore, geospatial analysis for the German rental market will be conducted by creating a choropleth map.

DATA

- Scrapped from immobilienscount24.de and available on Kaggle
- The datasets contain 49 variables &
  over 260,000 observations

TOOLS

- Python and Jupyter
- Libraries: Pandas, Numpy, Matplotlib, Seaborn, Scipy

SKILLS USED

- Data cleaning and wrangling
- Exploratory Analysis
- Machine Learning techniques: Regression Analysis and Clustering
- Geospatial Analysis
- Analysing Time Series Data

RENTAL APARTMENT IN GERMANY PRICES ANALYSIS

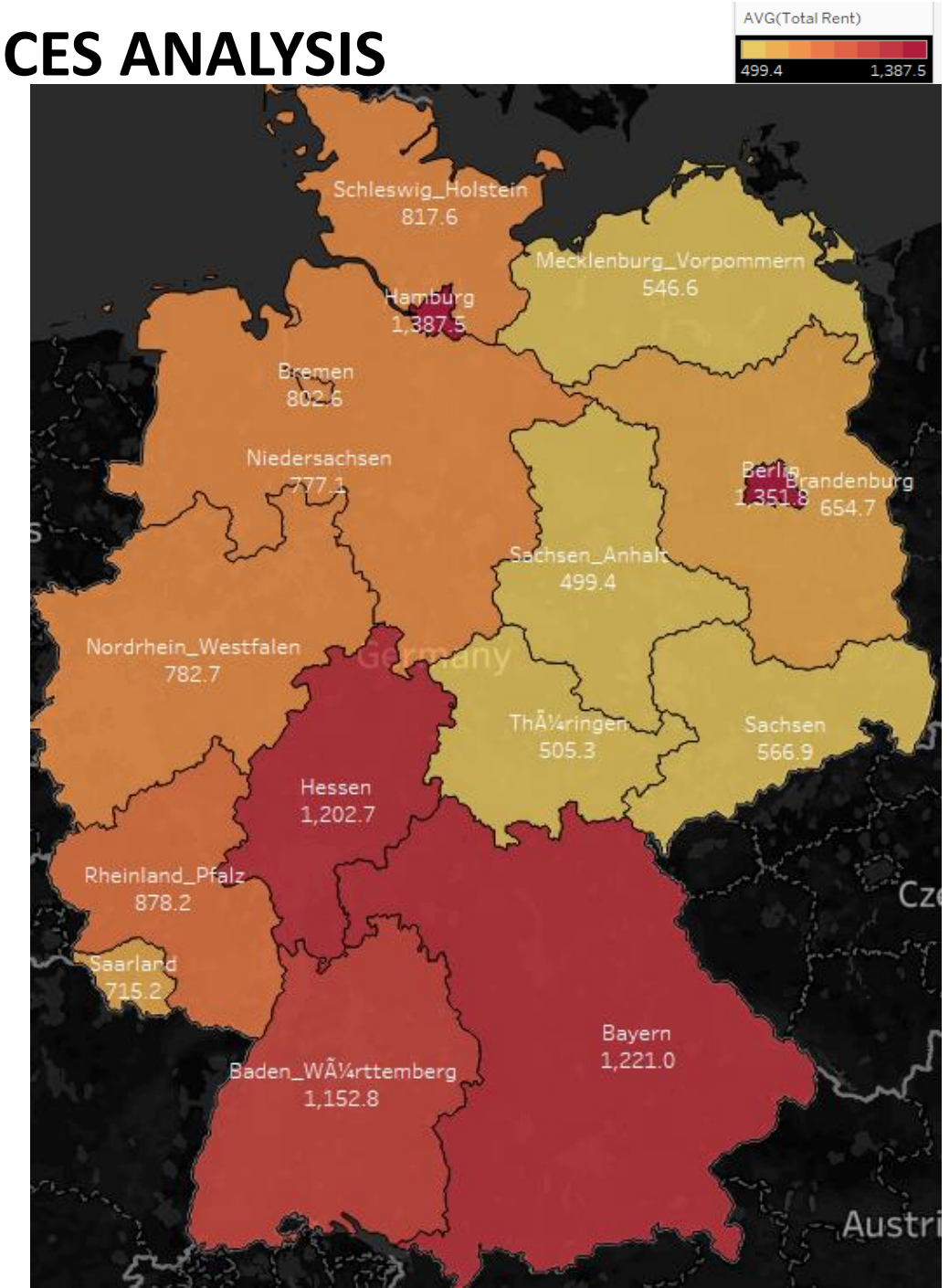
The map on the right shows the States in Germany and the Average monthly Total Rent per State

The 3 Most Expensive States to rent an apartment are:

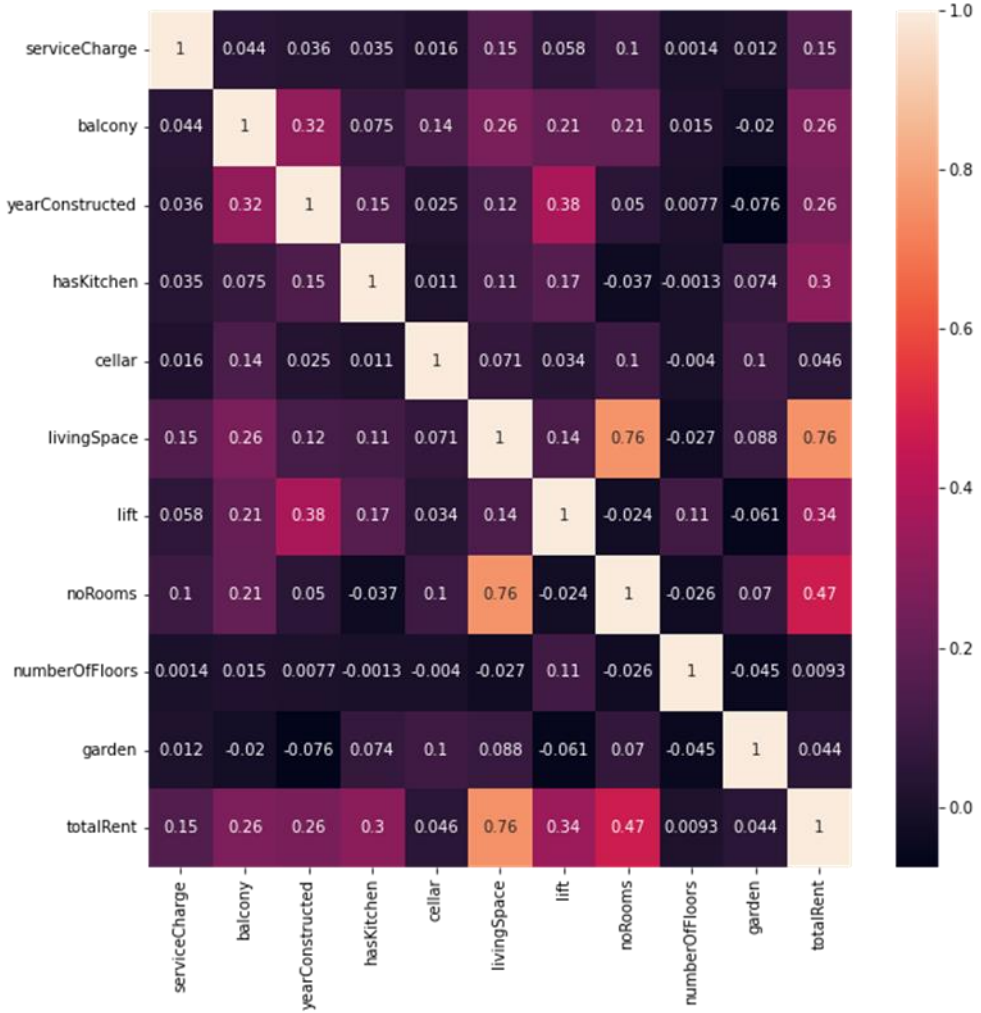
1. Hamburg - Total Rent - €1,387
2. Berlin - Total Rent - € 1,351
3. Bayern - Total Rent - € 1221

The 3 cheapest States to rent an apartment are:

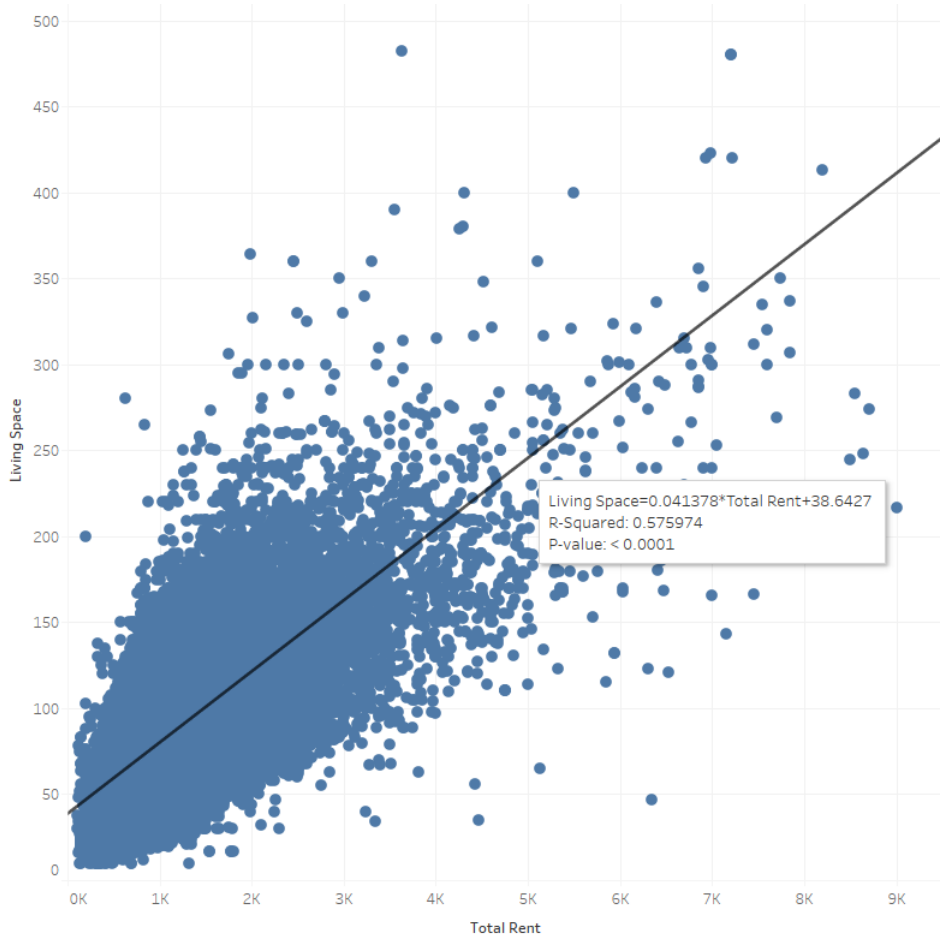
1. Sachen-Anhalt - Total Rent - €499
2. Thüringen - Total Rent - €505
3. Mecklenburg-Vorpommern - Total Rent - €546



Correlation Heat Map



Linear Regression TotalRent vs LivingSpace



- From the heat map above, TotalRent has a strong positive correlation of 0.76 with LivingSpace

LivingSpace is able to contribute to around 57% in explaining the TotalRent amount

CLUSTER ANALYSIS

The most expensive of the clusters is, cluster 2, purple. With:

- Average Total Rent of € 2,357 per Month
- Average LivingSpace of 142 square meters
- Average Number of Rooms is 4

The cheapest cluster is, cluster 3, blue. With :

- Average Total Rent of € 553 per Month
- Average LivingSpace of 54.29 square meters
- Average Number of Rooms is 2

CONCLUSIONS & RECOMMENDATIONS

Conclusions

- Apartments with bigger living space tend to be more expensive
- Apartments with more rooms number also tend to be more expensive

But there are exceptions as this is not always the case

Furthermore, it should be noted that all 4 clusters can be found in all the states

Limitations

- Inconsistency in the data when scrapping from website
- Whether Scrapping data ethical or not

Next Steps

Investigate the linear relationship between several variables (Multiple Linear Regression). I believe more variables will help to better predict the rent prices.

GET IN TOUCH

https://lionel1801.github.io
https://www.linkedin.com/in/lionel1801

write me an email:

lioneleyong@gmail.com