# Data Analytics Portfolio


Geetha Lakshmi

Data Analyst


Tableau repository

GitHub

LinkedIn Profile

# Projects Overview

## Video Game Popularity Data Project

- **Hypothetical Company Name:** "Game Co"
- **Project Goal:** To better understand the gaming industry market.

## Preparing for Influenza Season in the United States

- **Company:** A medical staffing agency
- **Project Goal:** To determine when and where to send medical staff, and how many, to each state.

## Rockbuster Stealth Data Analysis Project

- **Company:** Rockbuster Stealth LCC
- **Project Goal:** To help with the launch strategy for a new online video service.

## Instacart Grocery Basket Analysis Project

- **Company:** "Instacart"
- **Project Goal:** To uncover more information about the sales patterns.

## "Pig E. Bank" Project

- **Company:** "Pig E. Bank"
- **Project Goal:** To identify leading indicators for customers likely to leave the bank.

## "NYCitiBike_Project"

- **Company**: "Citi Bike"
- **Project Goal**: To uncover as many insights as possible about the station's locations and customer base.

# Video Game Popularity Data Project

## Project Goal

To better understand the gaming industry market.

## Tools used

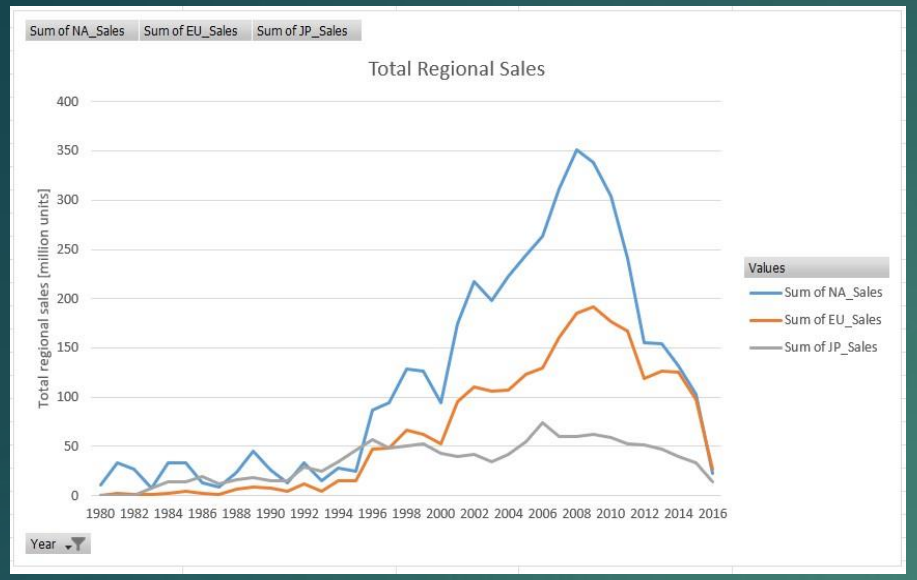
Microsoft Word
Microsoft Excel
Microsoft PowerPoint

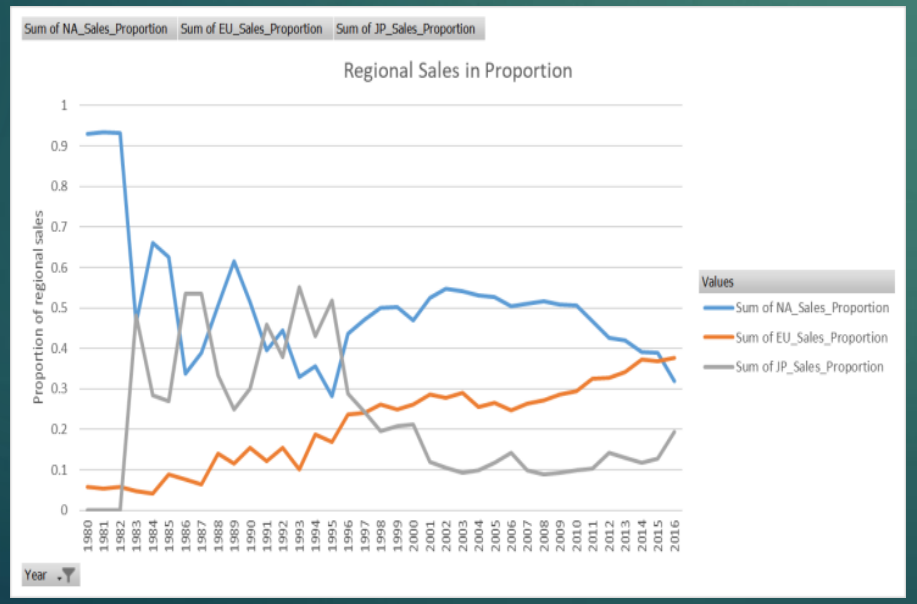
## Data Sets

https://www.vgchartz.com/

## Skills used

Data cleaning
Data grouping and summarizing
Conducting a descriptive analysis
Developing insights
Visualization
Storytelling




Challenge:
The sales were behaved more likely the same in all the regions by comparing the data with years. NA_sales, EU_sales, JP_sales were increased from the year 2001 to 2009. From 2010 to 2016, the sales were started decreasing gradually.

Hypothesis:
Explosion in digital distributed content
Less sales in physical sold games

Solution:
Examine relative sales
Consistent increase in EU sales
> Assumption of accuracy in relative data
>> Conduct analysis based exclusively on relative sales data

- North American Sales are decreasing during the last ten years (2006-2016). On the contrary, the European market has increasing with sustained a steady growth rate during the last 10 years. Europe dominated North America in total sales in 2016.

- We know that sales in NA in larger compare to EU and JP sales. But the sales are also increasing from the year 2000 in EU by observing percentage of regional sales. So I would like to suggest that if Game-co provides good marketing in EU then there are the changes to grow the market in the future.

# Project Links

**Project Brief**

- Project-Brief_Intro_to_Data_Analytics.pdf

**Final Report**

- Project Reflections 1.10 –Geetha Lakshmi.pdf

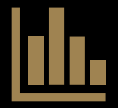
**PowerPoint Presentation**

- Final Project Presentation –Geetha Lakshmi.pdf

# Preparing for Influenza Season in the United States

## Project Goal

To determine when and where to send medical staff, and how many, to each state.



## Tools used

Microsoft Word
Microsoft Excel
Microsoft PowerPoint
Tableau

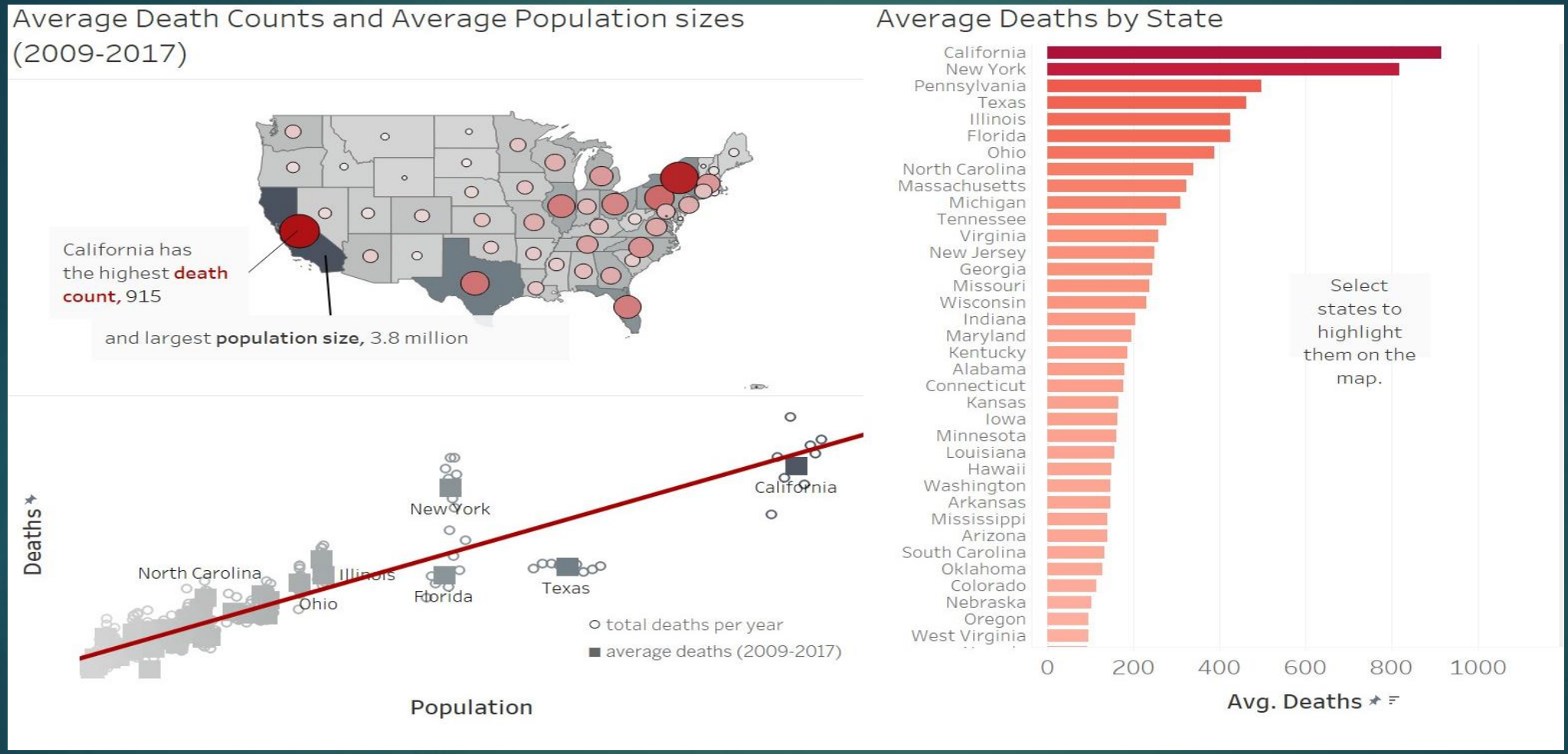
## Data Sets

CDC
CDC Flu View
Surveys of Flu Shots

## Skills used

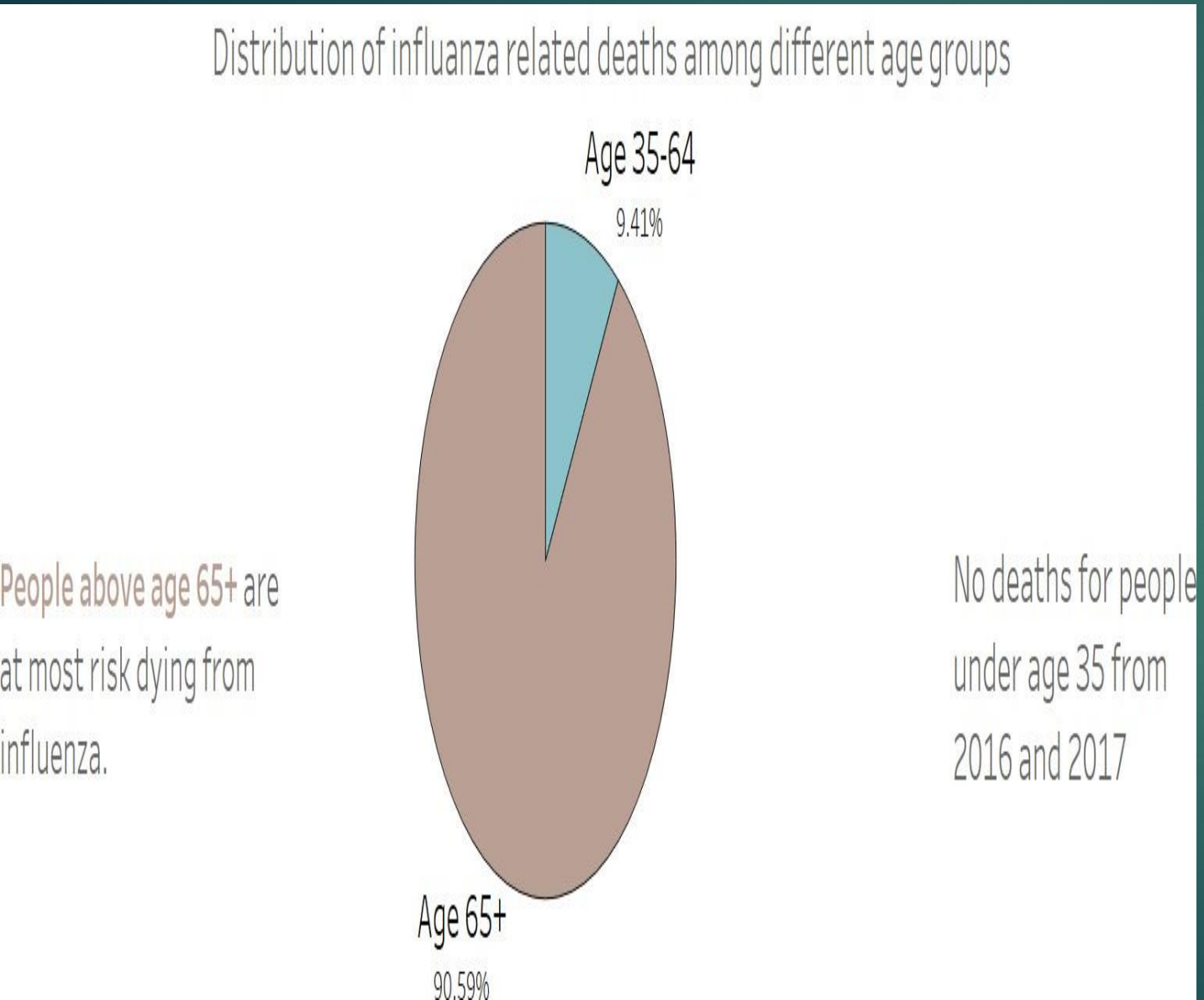
Data cleaning
Spatial analysis
Textual analysis
Visualizations and Forecasting
Storytelling with Tableau
Presenting findings to stakeholders

# Where is the Highest demand?



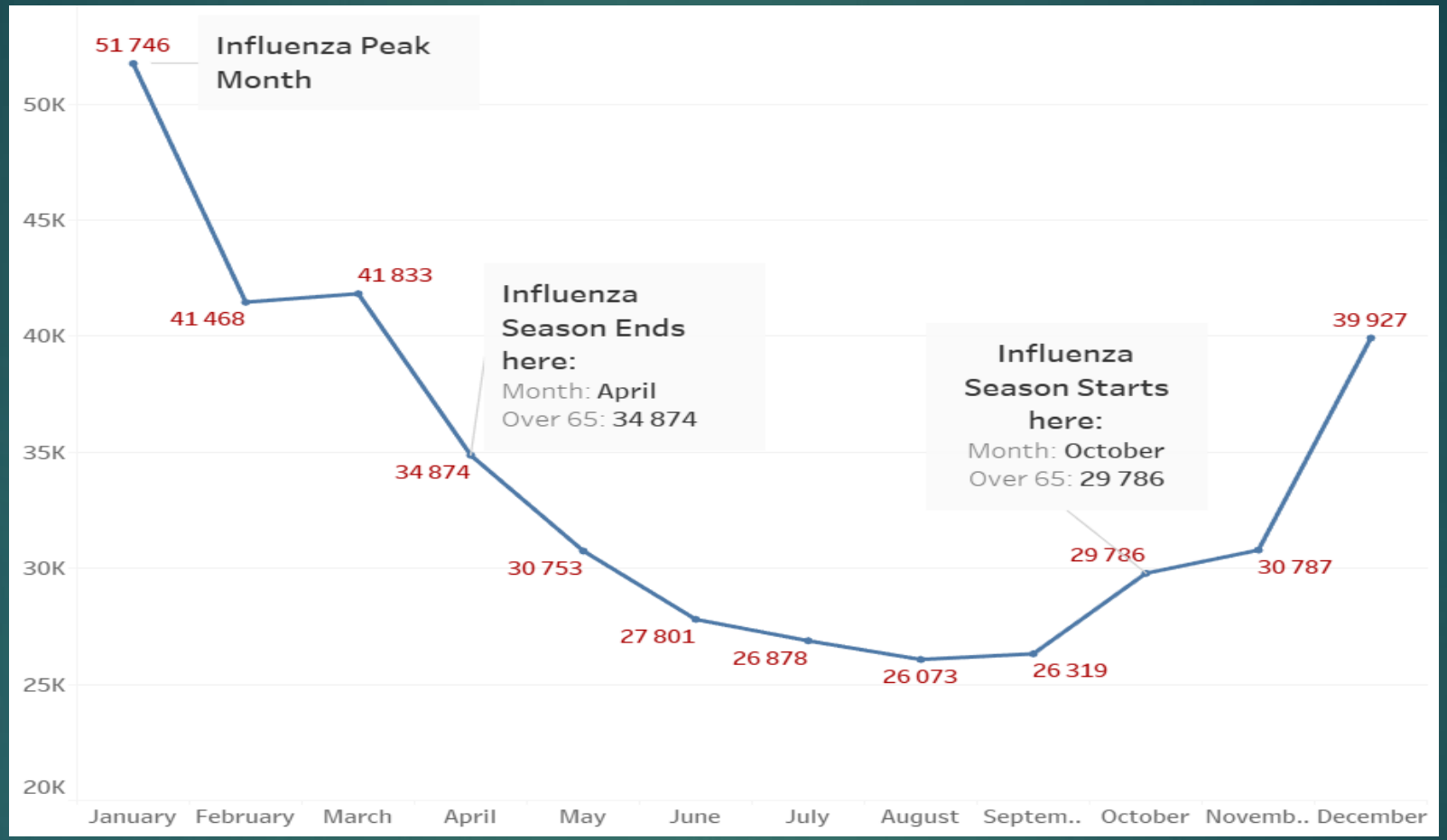

States with larger populations like California, New York and Texas suffer greater losses and therefore require more support.

# Who is at more risk?




- People above 65+ are at most risk from Influenza virus.
- There are very less number of deaths in-between the ages of 35 – 64
- There are no deaths for the people under the age of 35.

# Preparing for Influenza Season in the United States




- We see here that Influenza season occurs during the months of October to April(during the cold months)

# Project Links

**Project Brief**

- Project brief pdf

**Final Report**

- Tableau

**PowerPoint Presentation**

- YouTube

# Rockbuster Stealth Data Analysis Project

## Project Goal

To help with the launch strategy for a new online video service.

## Tools used

Microsoft Word, Microsoft Excel, Microsoft PowerPoint, PostgreSQL, Tableau, DbVisualizer

## Data Set

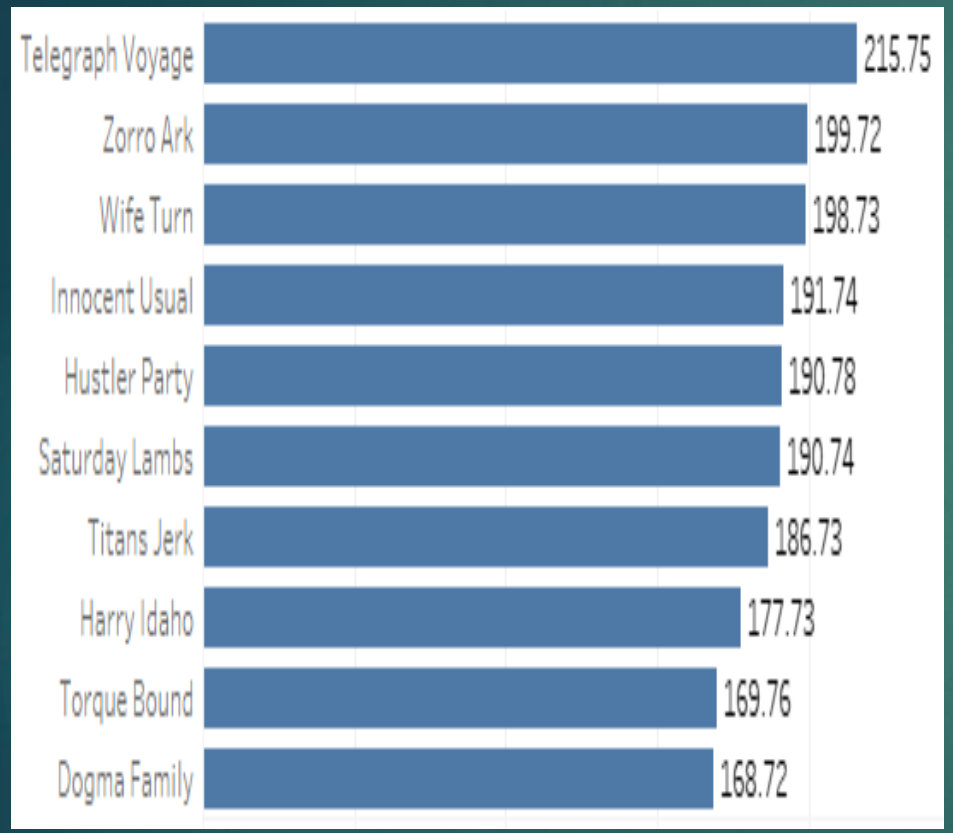
dvdrental.zip

## Skills used

Data cleaning, Relational Database, Summarizing and cleaning data in SQL, Filtering, Joining tables of data, Performing subqueries, Presenting SQL results

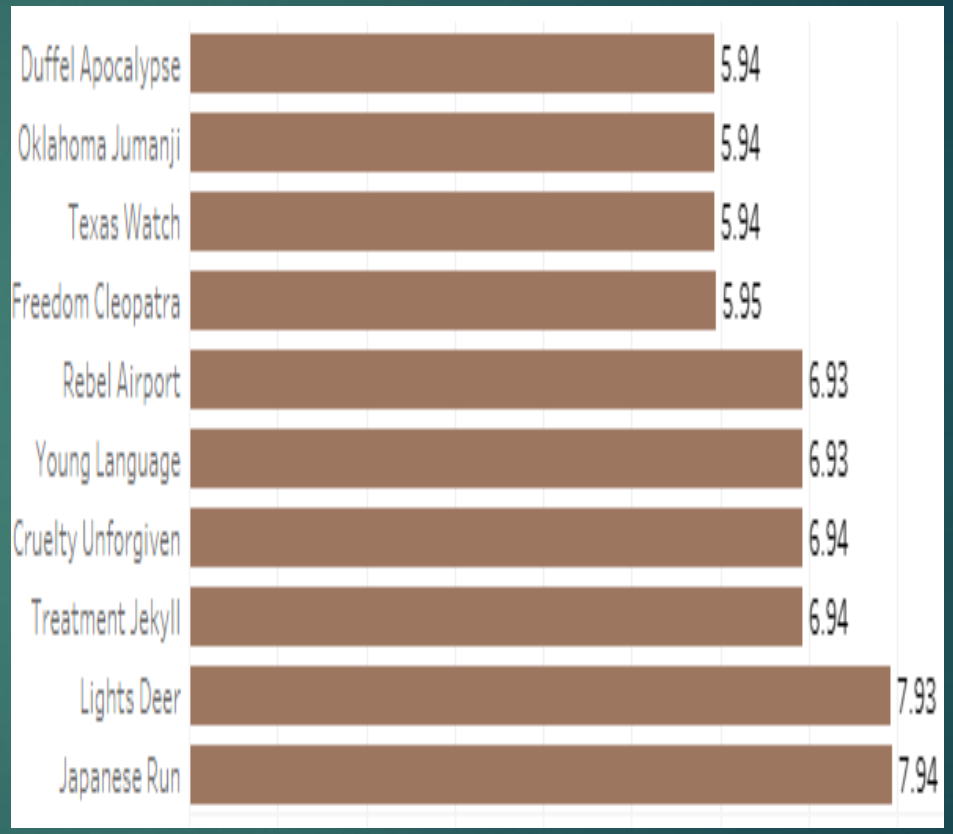
# Which movies contributed the most/least to revenue gain?

**TOP 10 MOVIES**




In the top 10 movies, the most revenue contributed movies are Telegraph Voyage and Zorro Ark

**Bottom 10 movies**

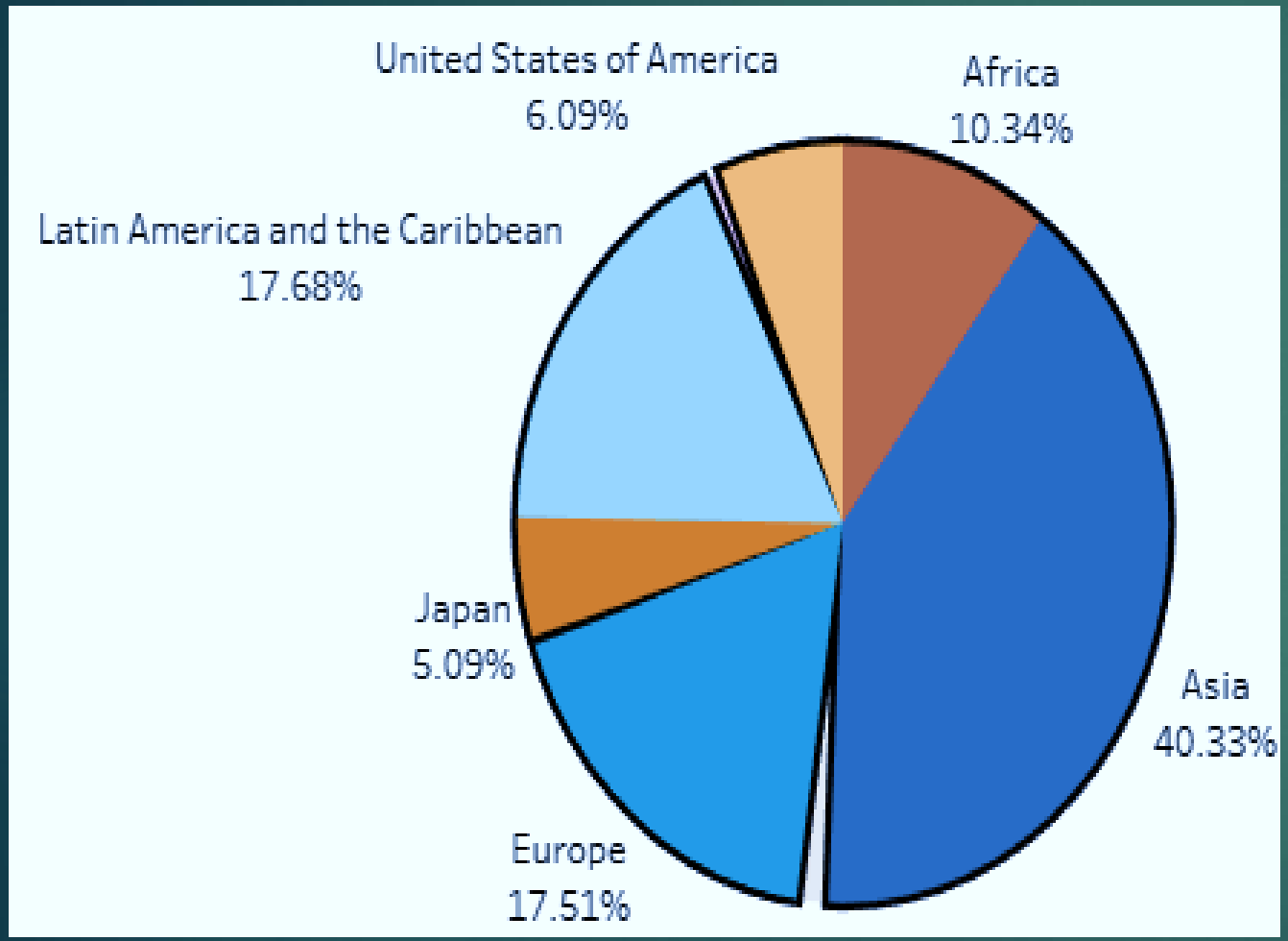



In the bottom 10 movies, the least revenue contributed movies are Duffel Apocalypse, Oklahoma Jumanji and Texas Watch.

# Which countries are Rockbuster customers based in?


The Rockbuster serves 109 countries within 6 regions with 2 stores located in Canada and Australia.

# Revenue by Region




**Total Revenue = $61,312**

Asia, Europe and Latin America has the highest Revenue followed by Africa, US and Japan.

# Project Links

**Project Brief**

- Project Brief PDF

**Data Dictionary**

- Dictionary

**PowerPoint Presentation**

- Presentation

**SQL codes**

- SQL codes and outcomes

# Instacart Grocery Basket Analysis Project

## Project Goal

To uncover more information about the sales patterns.

## Tools used

Microsoft Word, Microsoft Excel, Microsoft PowerPoint, Python, Jupyter, Pandas, Numpy, Matplotlib, Seaborn, Scipy.

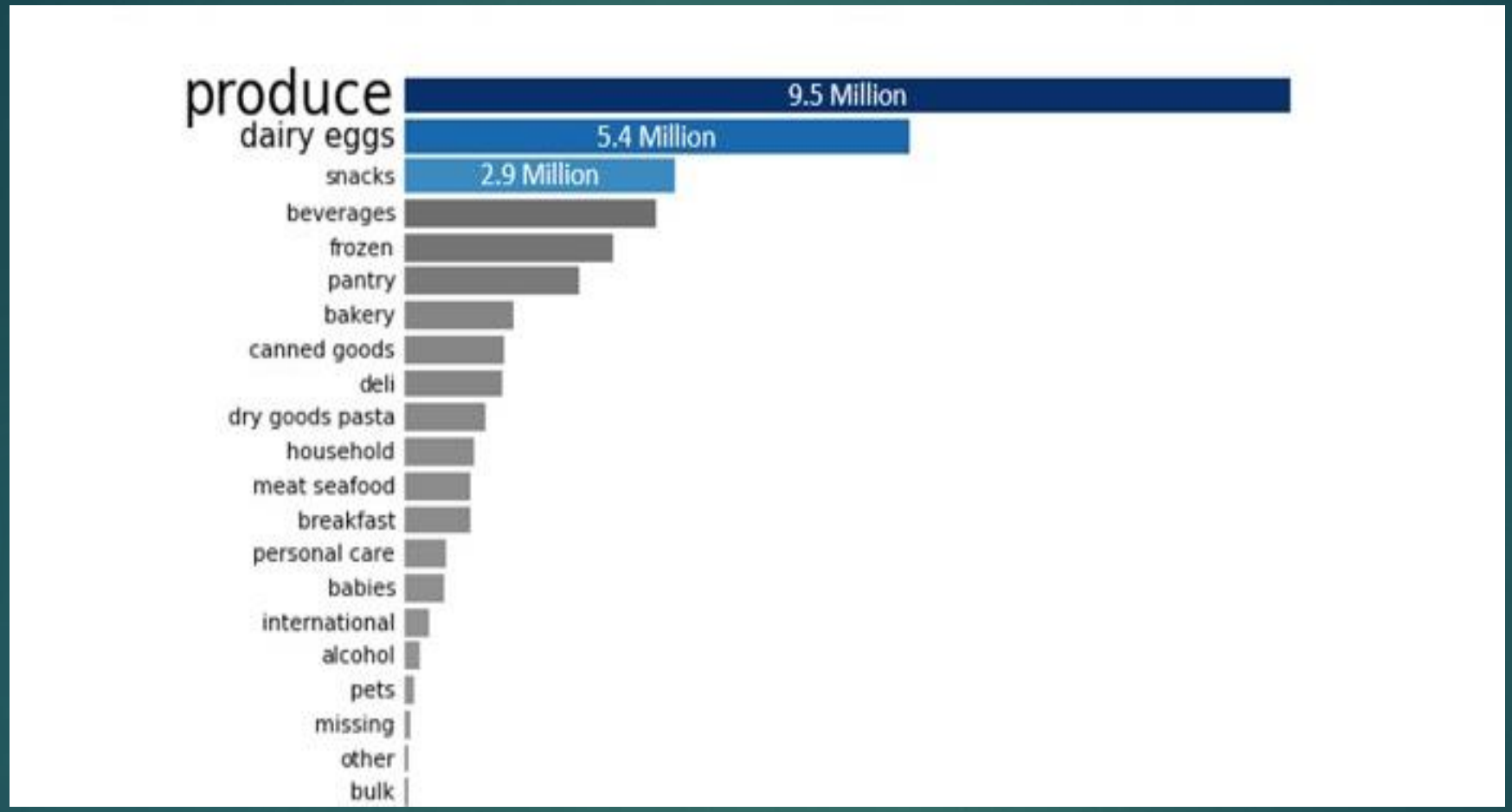
## Data Set

"The Instacart Online Grocery Shopping Dataset 2017", Accessed from via Kaggle

## Skills used

Data cleaning, wrangling & subsetting. Data consistency checks. Combining and Exporting. Deriving new variables. Grouping and aggregating variables. Data visualization with Python. Coding etiquette and Excel reporting.

# Departments with Highest Item sales




Dairy eggs and Snacks are the highest sales compared to all other products.

Please find my GitHub repository for the Python task.
GitHub repository:
https://github.com/GeethaLakshmi13/instacart_analysis_python

# Project Links

**Project Brief**

- Project Brief PDF

**Final Report**

- GitHub

# “Pig E. Bank” Project

## Project Goal

To identify leading indicators for customers likely to leave the bank.

## Tools used

Microsoft Word, Microsoft Excel, Microsoft PowerPoint, GitHub.

## Data Set

Pig E. Bank

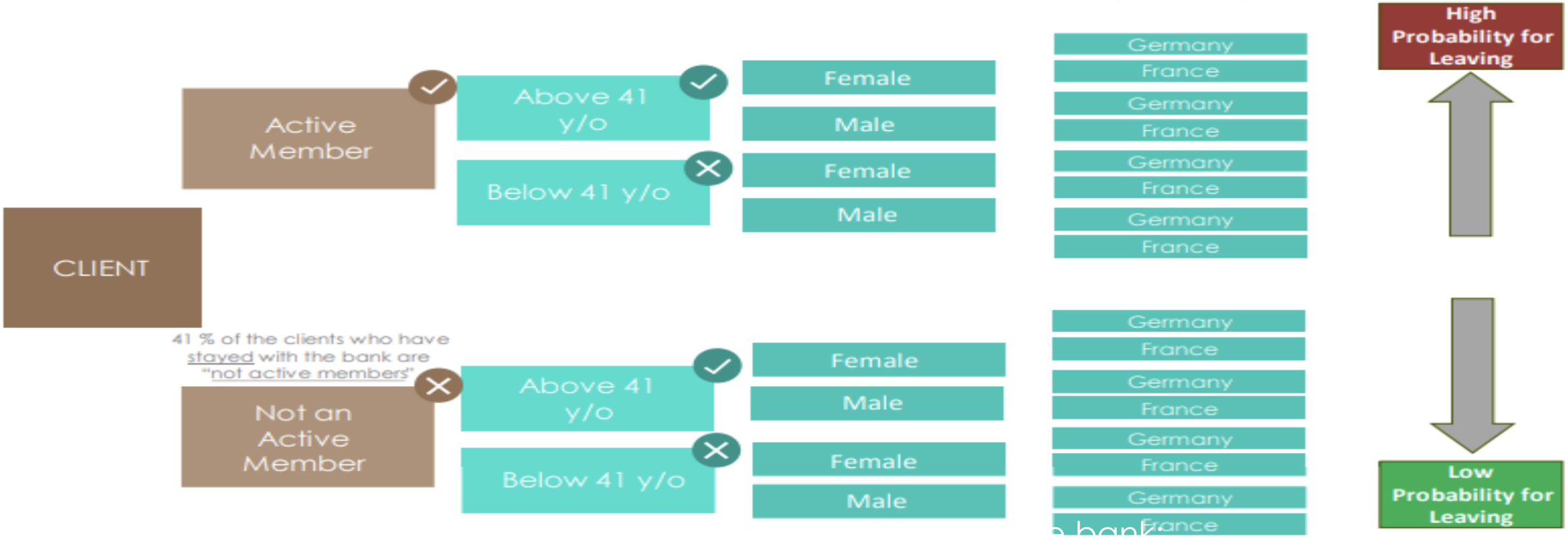
## Skills used

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

# "Pig E. Bank" Project

## Decision Tree

This decision tree map shows the probability of customers **leaving** the "Pig E. Bank".




- Active member;
- Above 41 years of age;
- Female;
- Living in Germany

# Project Links

**Project Brief**

- Project Brief PDF

**Final Report**

- Excel

# “NYCitiBike_ Project”

## Project Goal

To uncover as many insights as possible about the station’s locations and customer base.

## Tools used

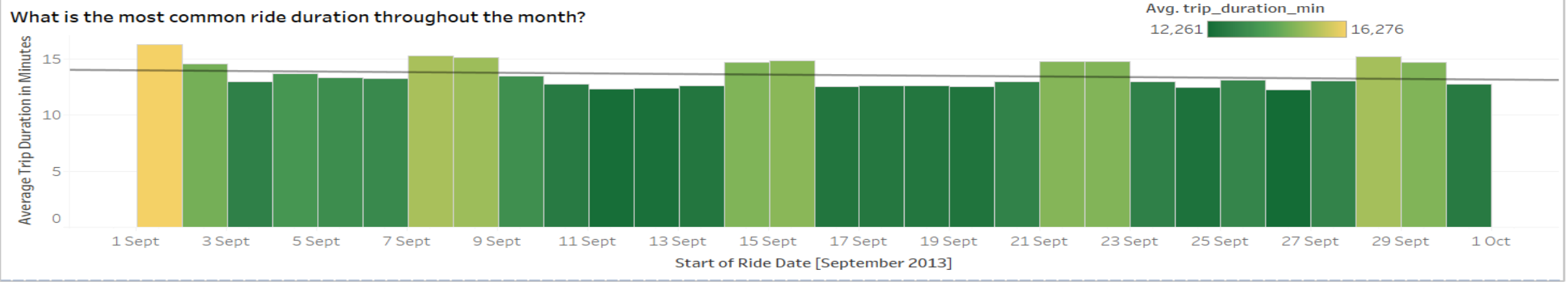
Microsoft Word, Microsoft Excel, Microsoft PowerPoint, GitHub, Jupyter Notebook, Python, Tableau.

## Data Set

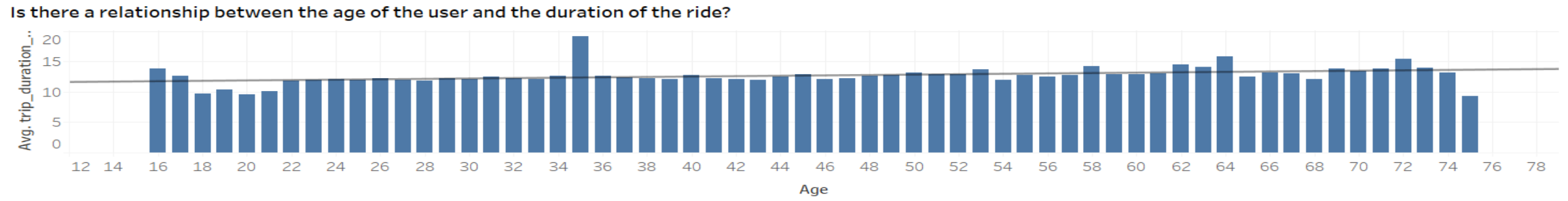
New York Citi Bike

## Skills used

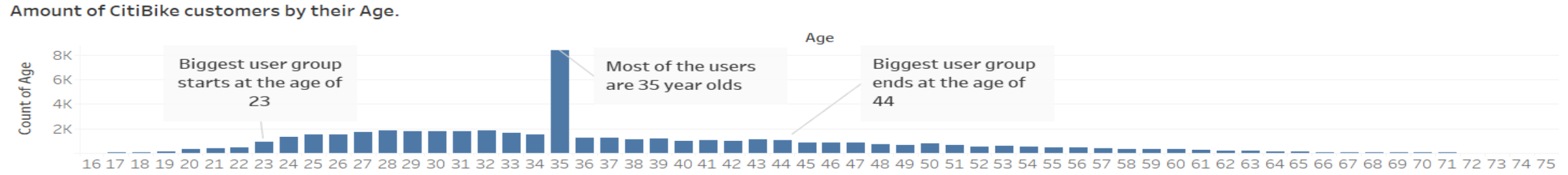
Big Data, Data Ethics, Data Mining, Predictive Analysis, Time Series and Forecasting, Designing and building a Dashboard, Relationships and patterns spotting.




Most common ride duration throughout the month is 14 minutes.



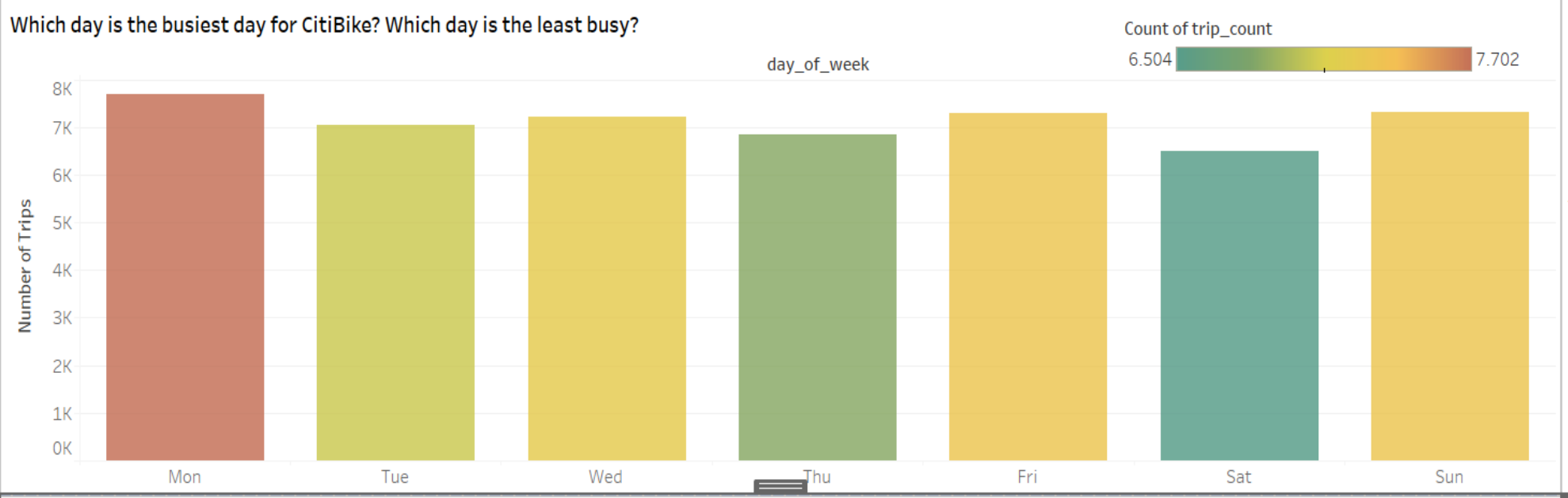

Statistics show that younger people ride approximately 12 minutes on average, while older people drive 13 minutes. But that's not a big of a difference.



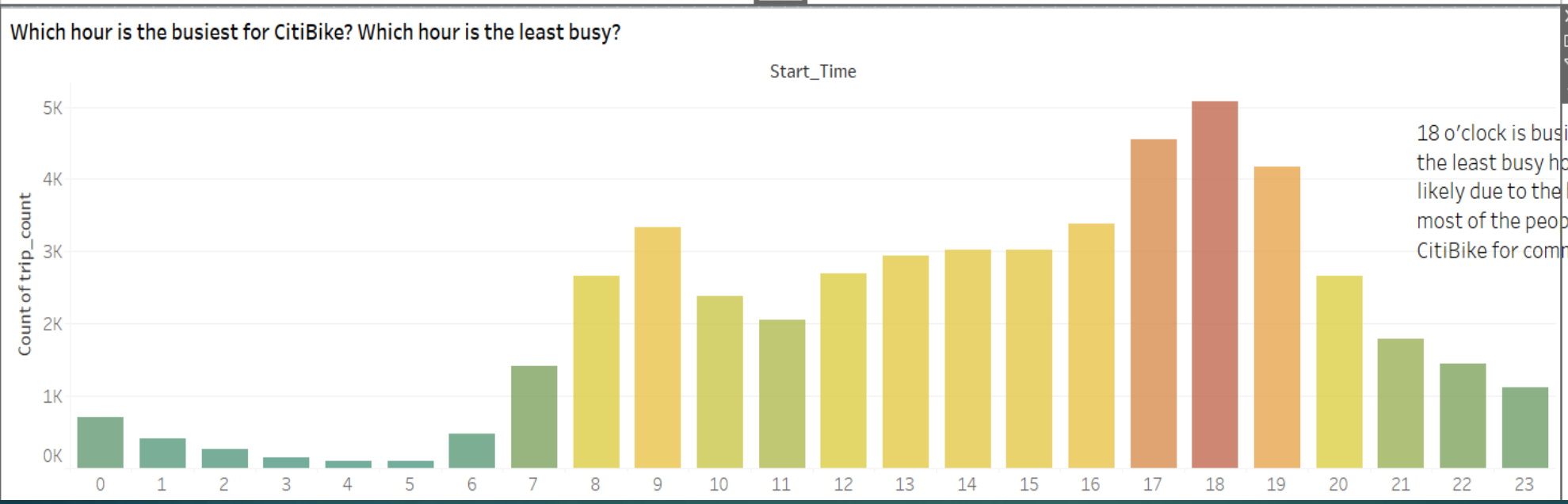

Most of the users are young and middle-aged people around (23 and 44 years old).

- During my analysis I found that most common rides are around 14 minutes. That older people ride on average 1 minute longer than younger people. And that the prevailing customer age is between 23 and 44 years old.

# "NYCitiBike_Project"



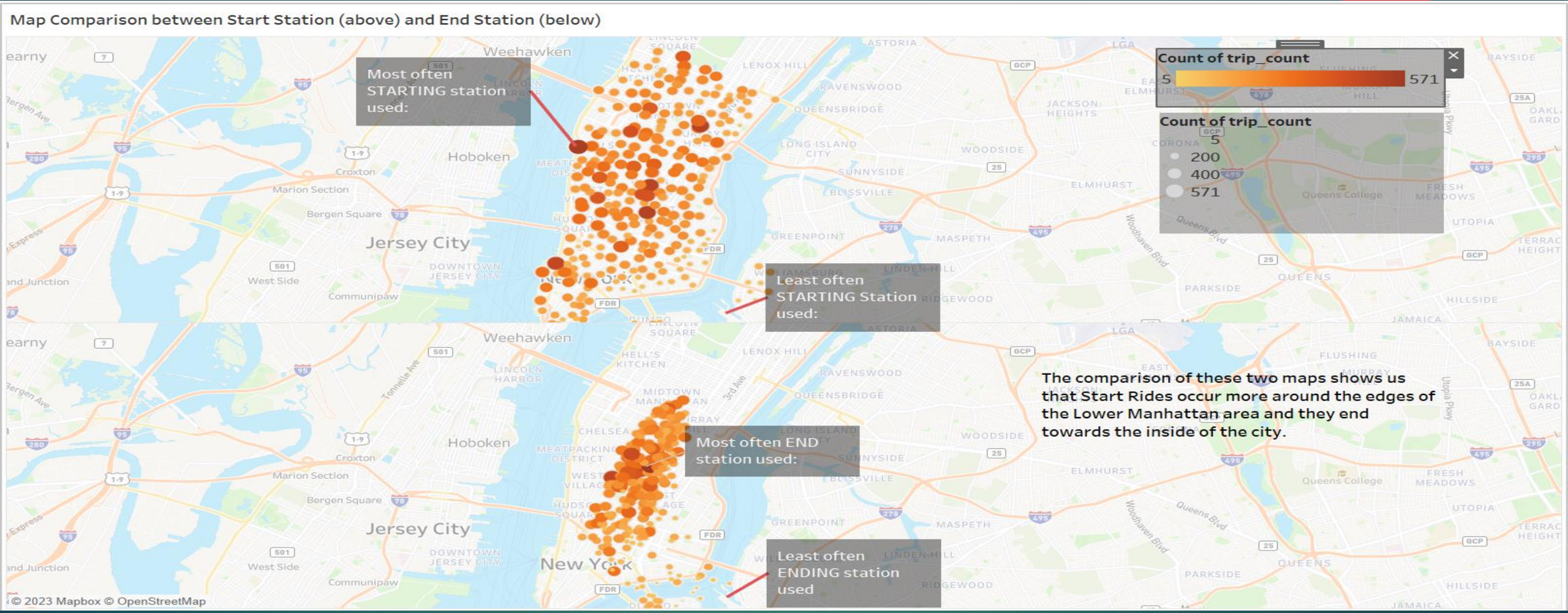

Busiest day in terms of trip counts is Friday/ Least busy is Saturday.




18 o'clock is busiest and 5 AM is the least busy hour. This is most likely due to the hypothesis that most of the people are using CitiBike for commuting purpouses.

Further into the analysis I found out that the busiest day in terms of trip counts is Monday and especially around 18 o'clock. Which led me to assume that most of the bike usage is for a commuting purposes.

# "NYCitiBike_Project"




- With the help of this interactive map, I uncovered that most of the rides occur in an inwards direction (from the outside towards the inside of the city).

- One more thing that I saw was the which starting and ending stations are mostly used and least used.

# "NYCitiBike_Project"

# Project Links

**Project Brief**

- Project Brief PDF

**Final Report**

- Tableau

# Thank you

Geetha Lakshmi
Data Analyst