DATA ANALYST PORTFOLIO

Denise Olson

dcolson66@gmail.com

Project Case Studies

Game Co.	• Market Analysis global video game sales.
Preparing for Influenza Season in US	• Analysis performed to help healthcare organizations plan for influenza season.
Rockbuster Stealth	• Market Analysis and response to business questions for online video rental company.
Instacart	• Customer and marketing strategy analysis for online grocery store.
Final Project	• Details coming

GameCo.

Project Objective

Perform descriptive analysis of video game data set to gain better insights to how GameCo's new games will perform in the market. Respond to Executive team's key questions regarding data.

Project Documents

GameCo Project Brief

GameCo Presentation

DataSet

- Video Game Sales (1980-2016), Source -VGChartz

Software Tools

- Excel

- PowerPoint

Skills

- Data Set understanding

- Data Cleaning

- Data grouping & summarization (Pivot Tables)

- Analytical Methods

- Descriptive Analysis/Developing Insights

- Visualization (build bar & column charts, whisker & scatter plots)

- Storytelling Data Presentation

GameCo. Analysis

Visualization 1
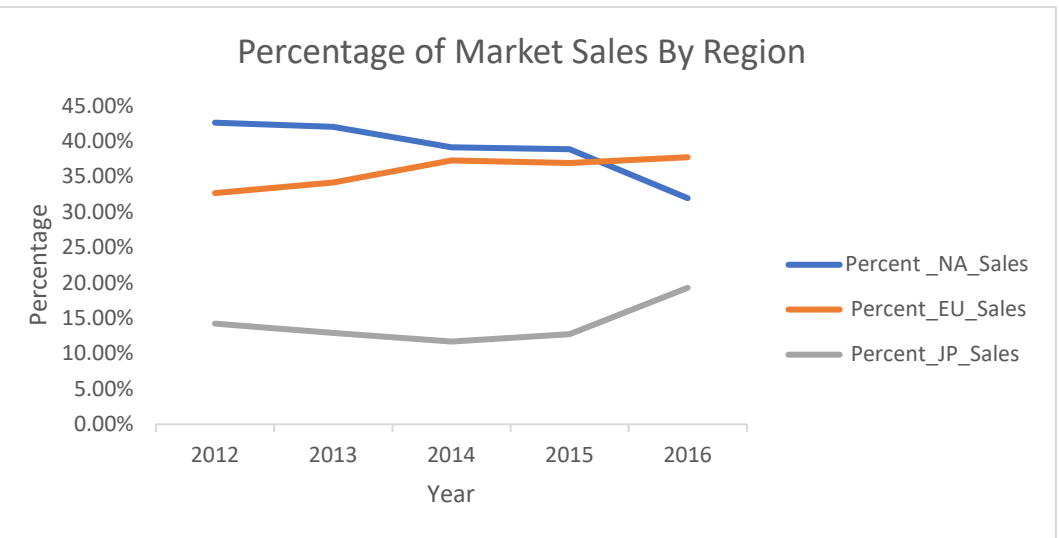Shows five-year historic sales by geographic regions in millions

Key Insight: Sales by geographic region **have not** stayed the same.
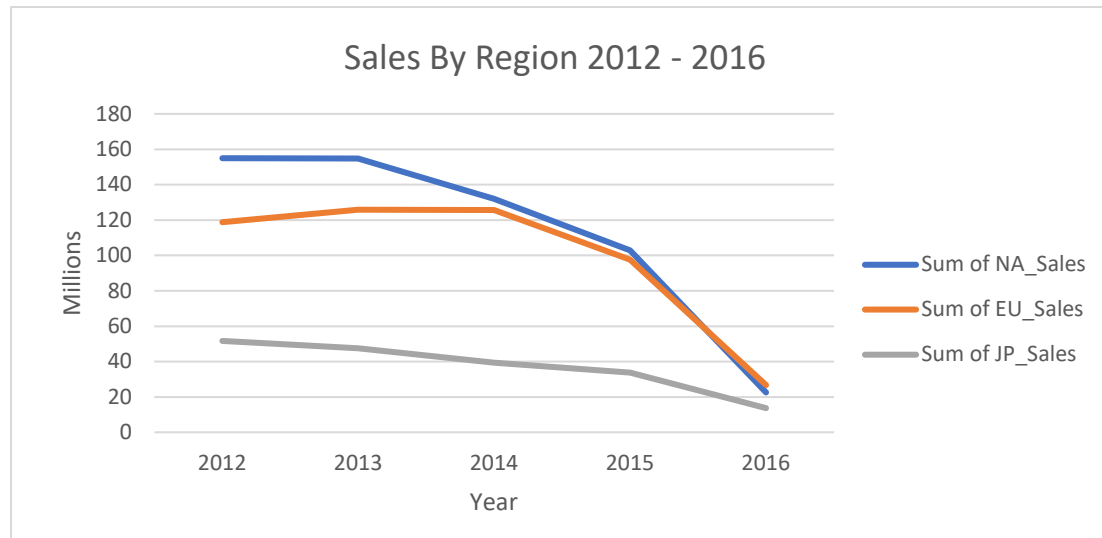
Visualization 2
Shows five-year Percentage of Market Share by Region

Key insight: Europe is surpassing North America sales in 2016, and Japan sales are trending upward in 2016.
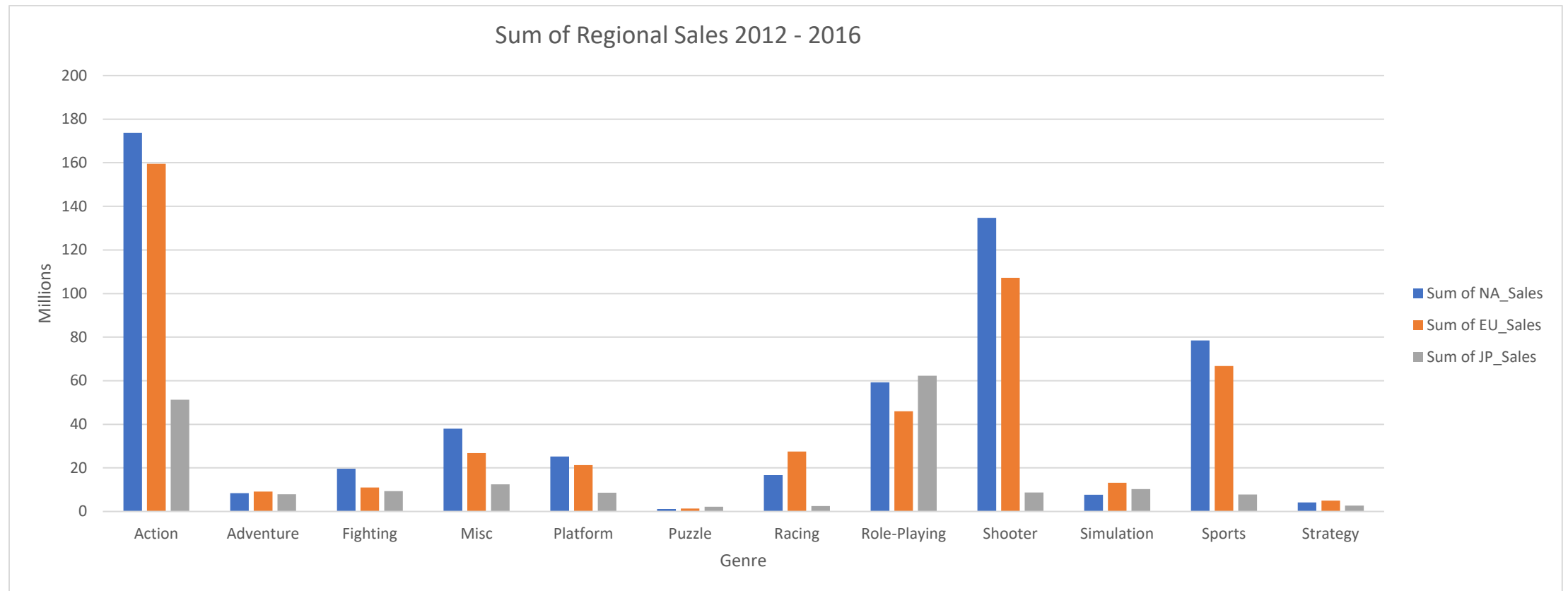
GameCo. Analysis

Visualization 3
Shows Sum of Genre Regional Sales

Key Insight: Top Genre sales across all regions are Action, Shooter, Sports and Role-Playing

GameCo. Insights

- North America and Europe regions top selling Genres (Action, Shooter, Sports)
  - Continue to allocate marketing dollars to top selling genres
  - Allocate additional marketing dollars to promote Role-Playing genre in these markets to try to generate additional interest /sales

- Japan Region
  - Increase marketing dollars in top selling Genres of Action, Shooter, Sports to try to boost regional interest and sales
  - Continue to allocate marketing dollars to Role – Playing genre

Preparing for Influenza Season in US

Project Objective
Analysis of CDC and Census Historical Immunization Data to determine vulnerable populations, and where to send additional medical staffing resources for upcoming flu season.

Project Documents

Influenza Project Brief

Influenza Project Management Plan

Influenza Project Interim Report

Tableau Presentation

Vimeo Video Presentation

Dataset(s)

- Influenza deaths by geography, time, age, and gender(2009-2017) – Source CDC

- Population data by geography(2009-2017) – Source US Census Bureau

Software Tools
- Excel
- Tableau
- Vimeo – video

Skills
- Data Preparation & Analysis (Requirements. Project Plan)
- Data Profiling & Integrity
- Data Quality Measures, Transformation & Integrations
- Statistical Analysis & Hypothesis Testing
- Data Visualization & Storytelling (Tableau)
- Composition & Comparison Charts
- Statistical Visualization (Histograms, Box Plots, Scatter Plots, & Bubble Charts)
- Spatial & Textual Analysis
- Storytelling Data Presentation

Influenza Season Analysis

Visualization pie chart displays total number of vulnerable and non-vulnerable population totals from 2009-2017

Planning for Influenza Season
- If states have a high number of vulnerable populations and deaths due to flu, then hospitalizations will be greater and medical staffing needs will be higher in demand.
- Vulnerable Population: Identified by CDC as children under 5 years and adults over 65 years.

FluDeath NonVP
12,835,121

FluDeath VP
25,693,351
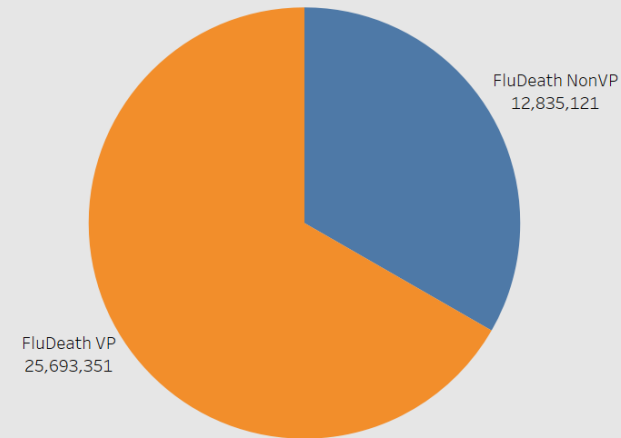
Influenza Season Analysis

States that have high numbers of vulnerable populations are impacted the most with influenza deaths (2009-2017)
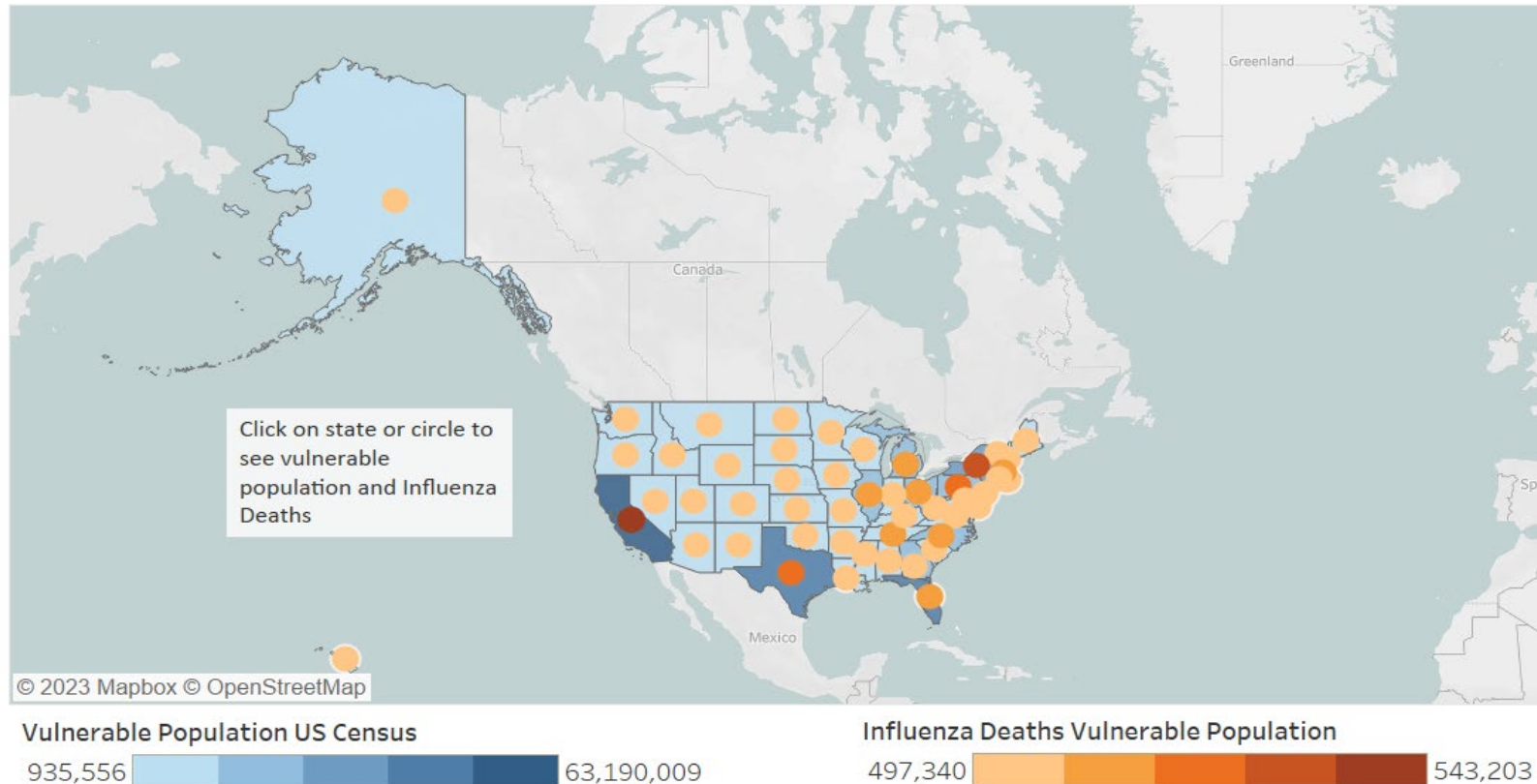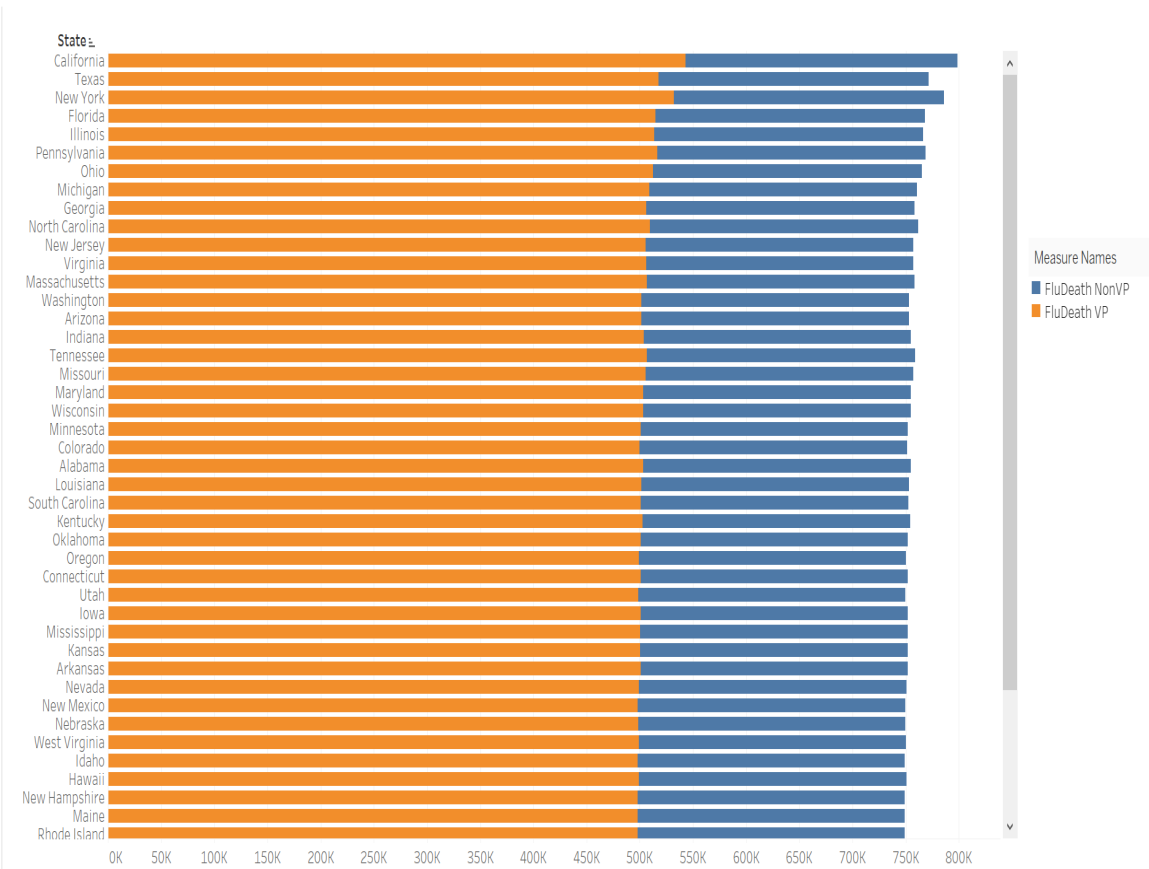
Influenza Season Analysis

This Scatter Plot map shows, as the vulnerable population number increases, the number of influenza deaths of vulnerable populations are expected to rise.
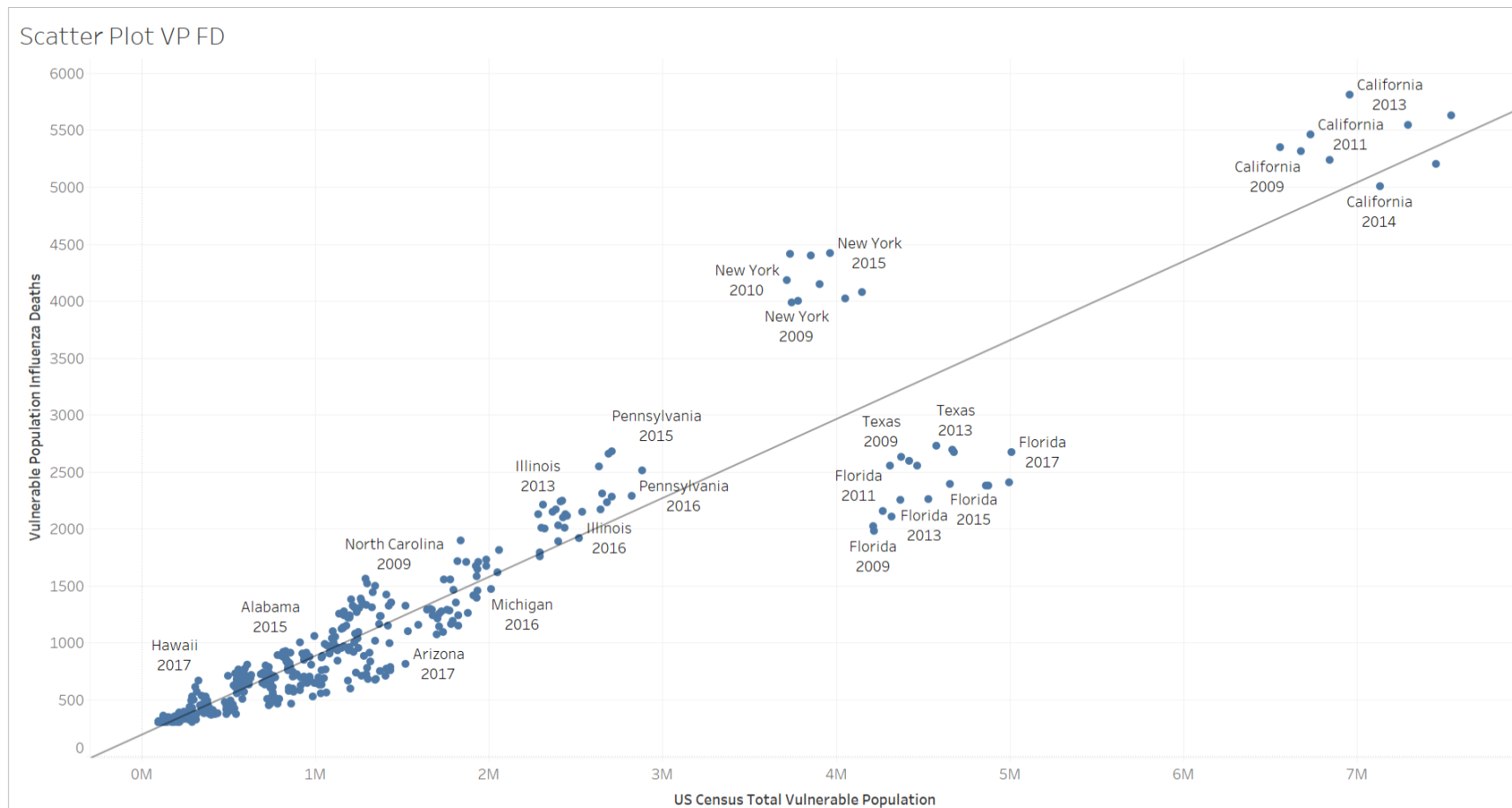
Influenza Season Analysis

States with have high numbers of Vulnerable Populations(VP) are impacted the most with influenza deaths (2009-2017)

Preparing for Influenza Insights

➢ It can be concluded from the historic data and visualizations evaluated from the years of 2007-2019, which states consistently have high numbers of vulnerable populations and influenza deaths.

➢ These states clinics and hospitals should further be consulted with to determine if during influenza season they will have additional medical staffing resource needs for the upcoming influenza season.

Rockbuster Stealth

Project Objective
Provide insights and answer several key business questions to leverage existing movie license to launch online video rental service.

Project Documents

Rockbuster Stealth Project Brief

Rockbuster Stealth Presentation

Rockbuster Stealth Data Dictionary

Dataset(s)
- Dataset loaded into PostgreSQL
    - Contains information about Rockbuster's film inventory, customers, and payments.

Software Tools
- Excel
- PowerPoint
- Tableau
- PostgreSQL

Skills
- Relational Databases
- Structured Query Language (SQL)
- Summarizing and cleaning Data in SQL
- Table Joins, Performing Subqueries
- Common Table Expressions(CTE's)
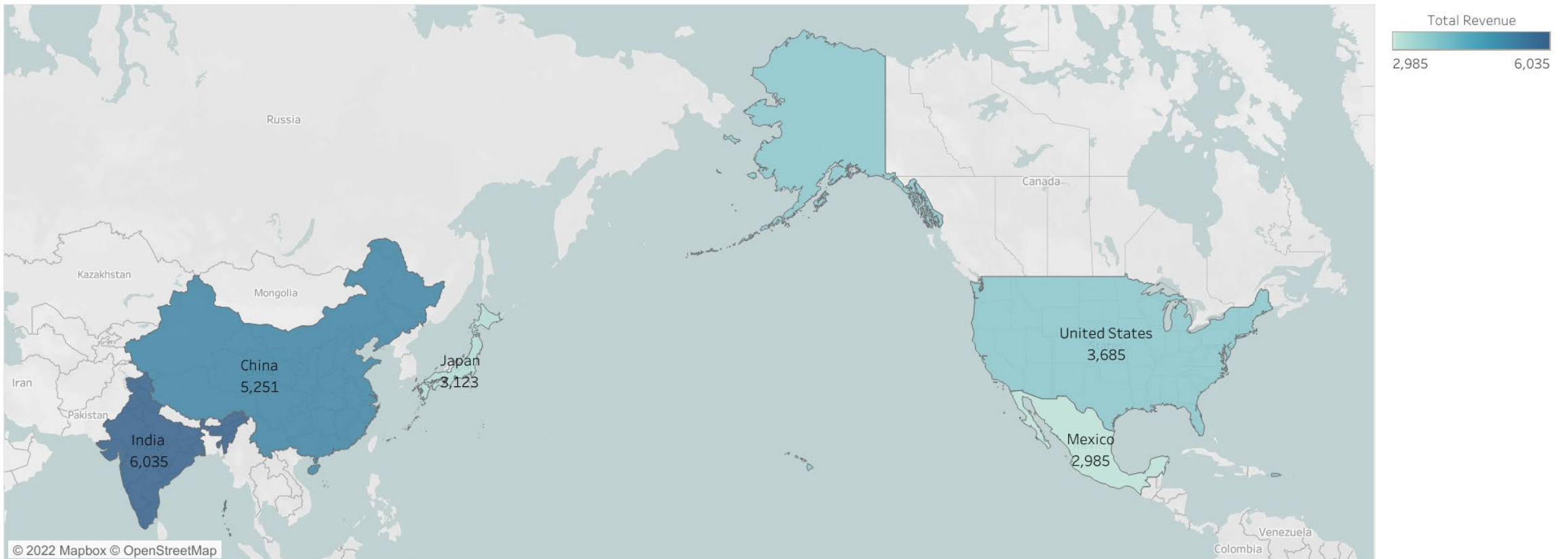- Visualizations of SQL results and Presentation

Rockbuster Stealth Analysis

Top performing Genre's by number of films and total revenue. Top 5 Genre's (Sports, Animation, Action, Sci-Fi, and Family) make up 35% of total rental revenue.
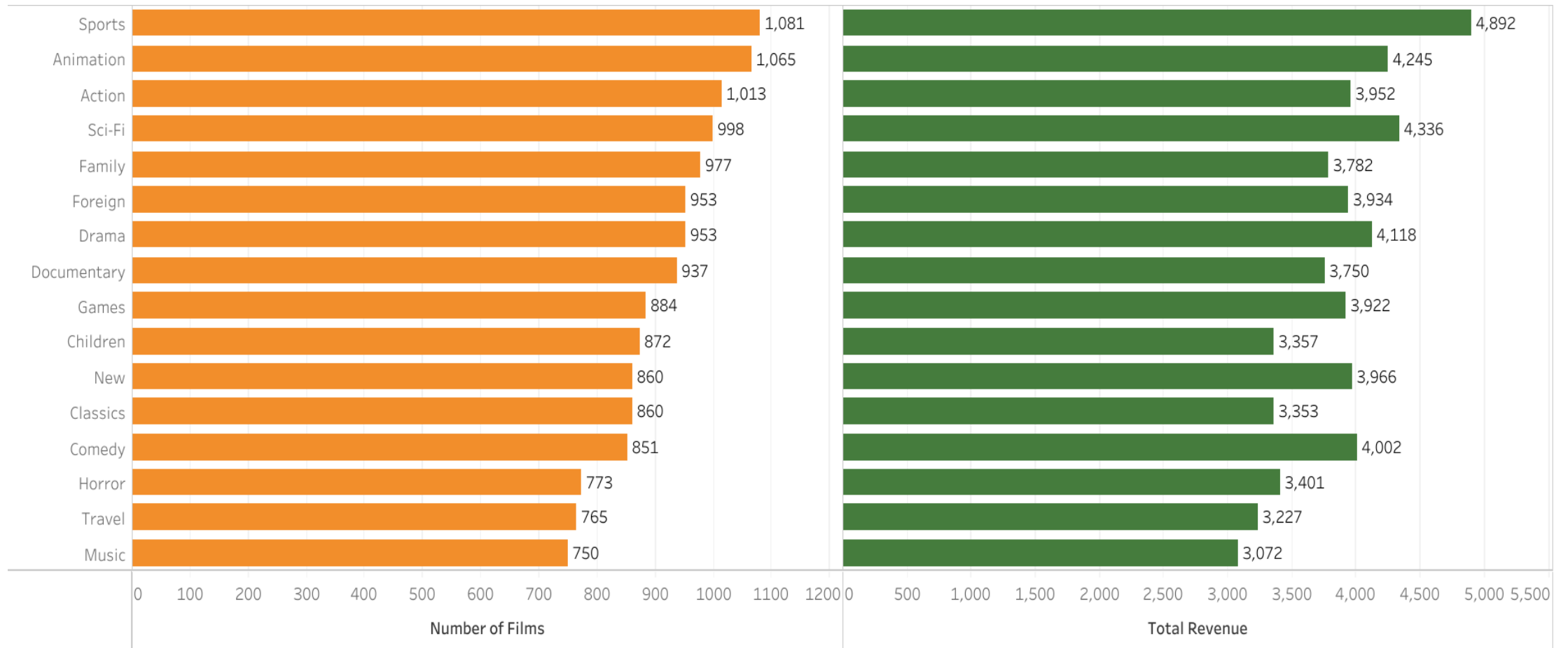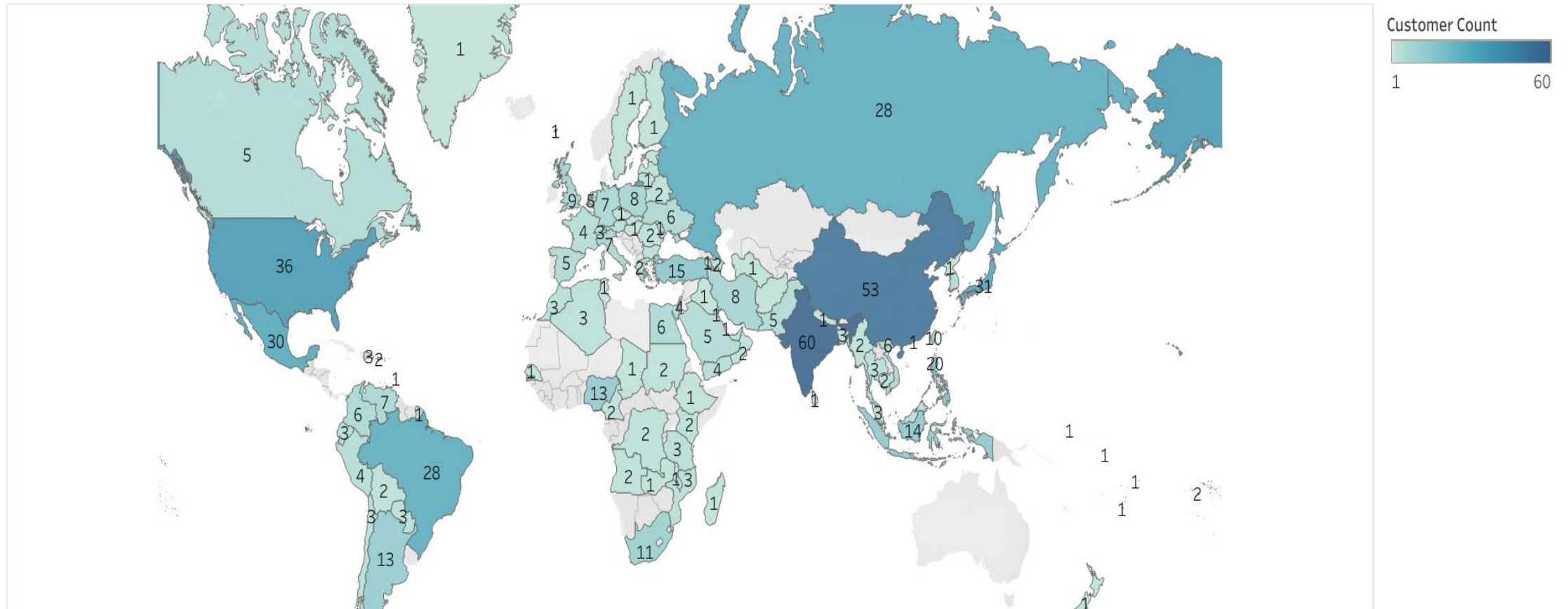
Rockbuster Stealth Analysis

Top Countries in Revenue are India $6035, China $5251, United States $3685, Japan $3123 and Mexico $2985

Rockbuster Stealth Analysis

Top Countries Customer count are India 60, China 53, United States 36, Japan 31 and Mexico 30

Rockbuster Stealth Insights

➢ Rockbuster Stealth Management team should focus making new streaming services available to existing top performing countries.

➢ Create marketing campaign to target locations with high revenue count.

➢ Focus on promoting Sports, Animation, Action, Sci-Fi, and titles for they are the most popular.

Instacart Grocery Store

Project Objective
Perform data and exploratory analysis to derive insights and suggest strategies for better segmentation based on customers online ordering behavior.

Project Documents

Instacart Project Brief

Instacart Final Report

GitHub Repository (for both Rockbuster Stealth & Instacart Grocery)

Dataset(s)
Open-source data sets from Instacart's online ordering, Includes, "customers", "orders" and "products"
- Instacart

Software Tools
- Excel
- Anaconda & Jupyter
- Python, Python datatypes & libraries
- Pandas
- Matplotlib
- Seaborn

Skills
- Data Exploring
- Data Wrangling & Subsetting
- Data Consistency checks, Cleaning, & Merging
- Combining & Exporting Data
- Deriving New Variables
- Data grouping &Data Aggregating
- Data Visualization with Python and presentation
- Coding Etiquette & Excel Reporting

Instacart Analysis

Peak days / hours. Bar chart shows the busiest days are Sat., Sun. and Friday, while the histogram shows that the most orders are placed between 8am and 7pm.

LEGEND
0 = Saturday
1 = Sunday
2 = Monday
3 = Tuesday
4 = Wednesday
5 = Thursday
6 = Friday
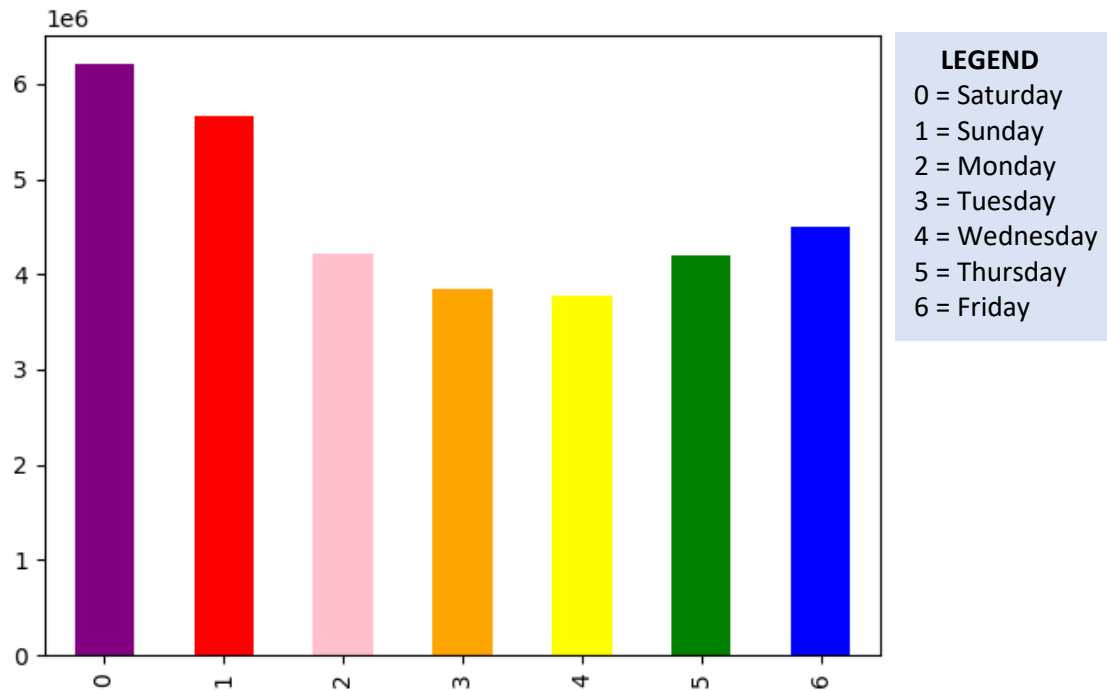
This bar chart represents the days of the week when orders are placed. The busiest days are Saturday., Sunday and Friday.
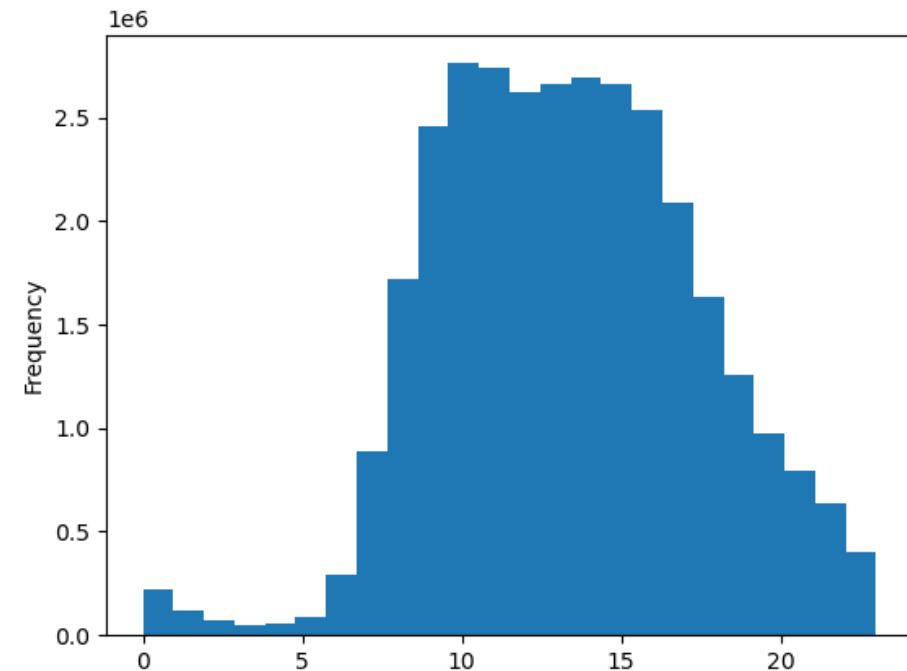
This histogram represents what hours orders are being placed hourly. The busiest hours of the day are between 8am and 7pm.
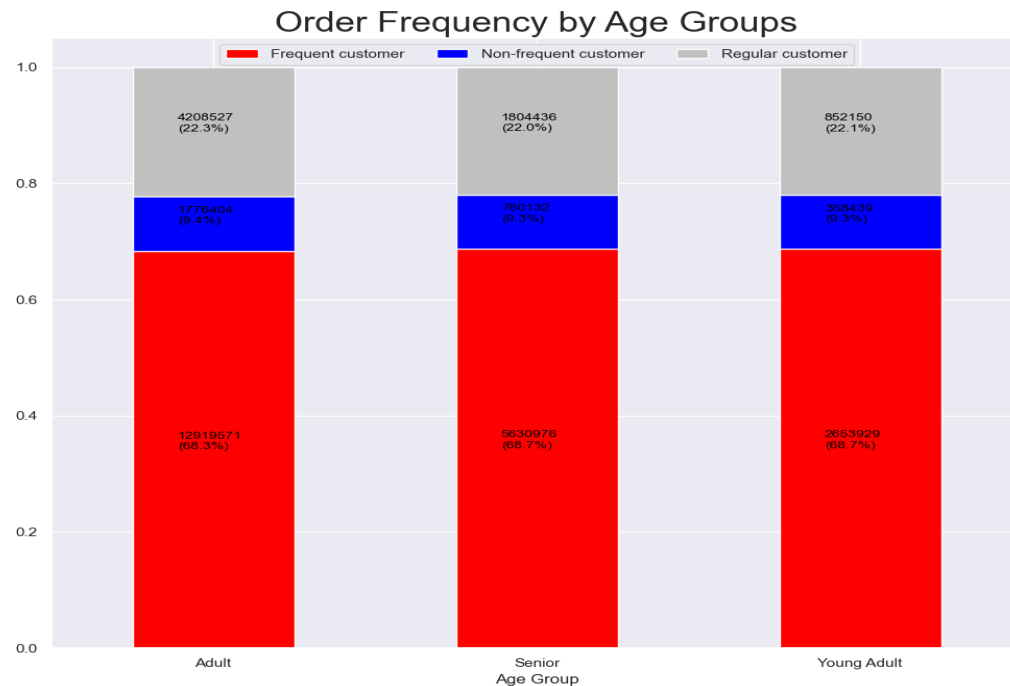
Instacart Analysis

The bar chart shows the 5 product departments with highest frequency of product orders are produce, dairy eggs, snacks, beverages, and frozen. Instacart has a lot of products with different price tags, the below pie chart shows price ranges from low, mid and high to help understand customer ordering.
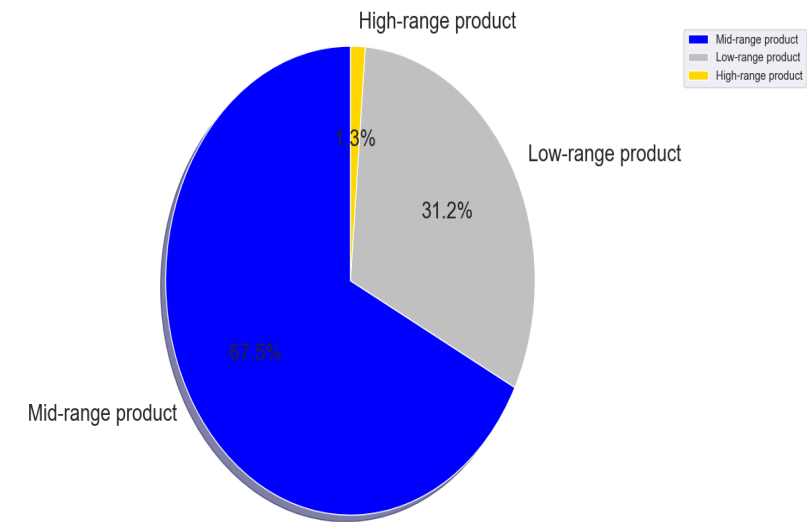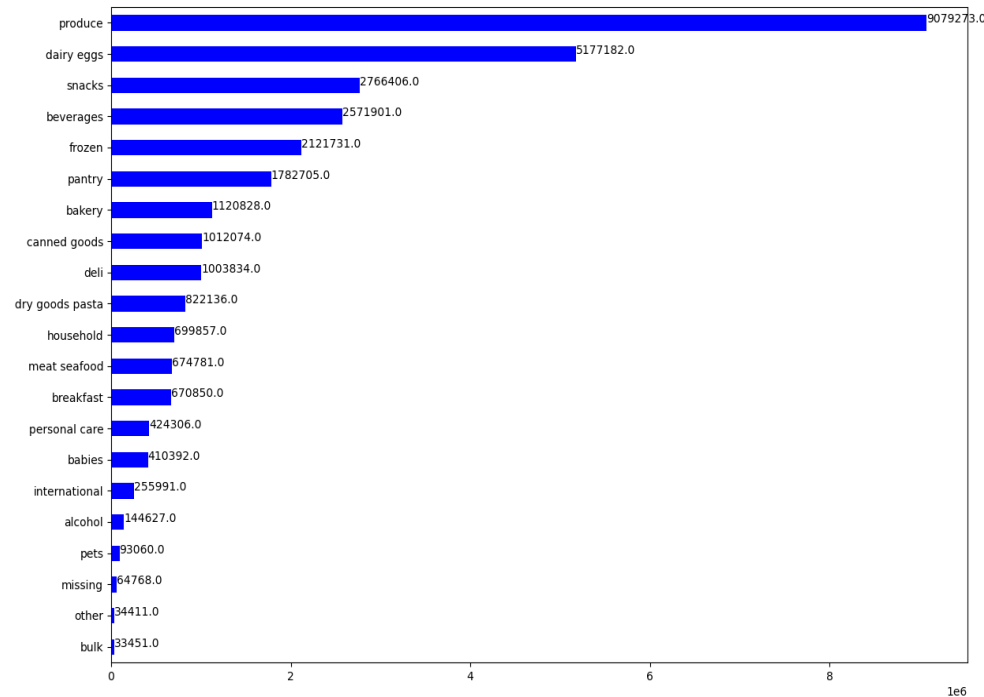

Price ranges were sorted by low-range, mid-range and high-range product categories. Pie Chart displays the proportions of each category.
If product is $5 it is considered low range product
If product price is > $5 but <= $15 mid range product
If product price is > $15 it is considered high range product

Instacart Analysis

Instacart customer base segmented by age group and order frequency and customer loyalty

The above stacked bar chart shows order frequency by age groups. The distribution among each group is nearly the same. Age Group is defined by:

Age <= 25 = Young Adult

Age > 25 and <65 = Adult

Age >= 65 = Senior
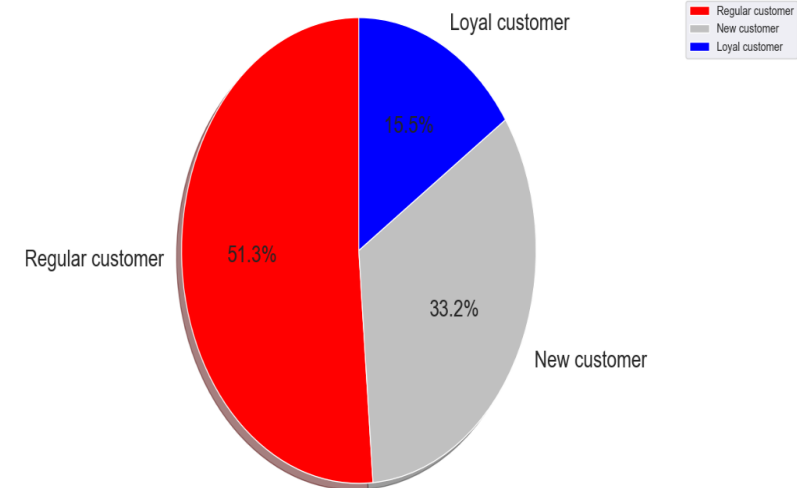
Customers are ranked as New, Regular and Loyal based on:

Max order <= 10 = New Customer

Max order >10 but <= 40 = Regular Customer

Max order > 40 = Loyal Customer

Instacart Insights

➢ For peak shopping days and times, Instacart marketing team should promote Ads and incentives to encourage ordering outside of busy days and hours.

➢ Top 5 product departments with highest frequency of product orders are produce, dairy eggs, snacks, beverages and frozen. Marketing can focus on the top 5 product departments but also can try to use ads to increase sales of some lower departments like alcohol and personal care.

➢ To increase ordering of different products, could have target ads marketing the low and higher range products to try to increase sales across each age and marital status group. Marketing should focus advertisements to regular and new customers to increase max orders for them to become loyal customers.

Boat View Analysis

Project Objective
Conduct exploratory analysis of boat data to examine which variables, specifically country, price in EURO's, length and width may impact the number of views a boat receives.

Project Documents

Boat Views Analysis Tableau

GitHub Repository

Dataset(s)
- Dataset sourced and provided by Kaggle (boat_data.cvs)
 - Kaggle boat_data

Software Tools
- Excel
- PowerPoint
- Tableau
- PostgreSQL
- Python, Python datatypes & libraries
  - Quandl
  - Pandas
  - Numpy
  - Seaborn
  - Matplotlib, Matplotlib.pyplot
  - Statsmodel.api
  - Folium
  - Plotly
  - Quandl
- Anaconda & Jupyter
- GitHub

Skills
- Sourcing Open Data
- Exploring Relationships (scatterplots, correlation heatmaps, pair plots& categorical plots)
- Geographical Visualizations
- Supervised Machine Learning: Regression
- Unsupervised Machine Learning: Clustering
- Sourcing & Analyzing Time Series Data
- Creating Data Dashboard

Exploratory Data Analysis

- Exploring the dataset using Python
- Finding connections between variables
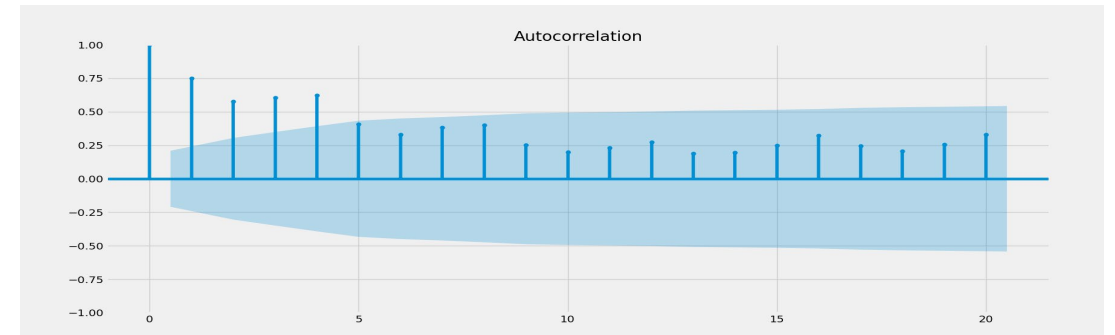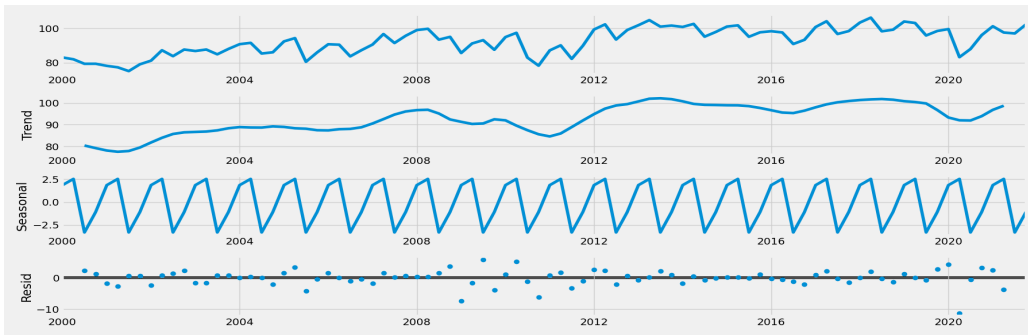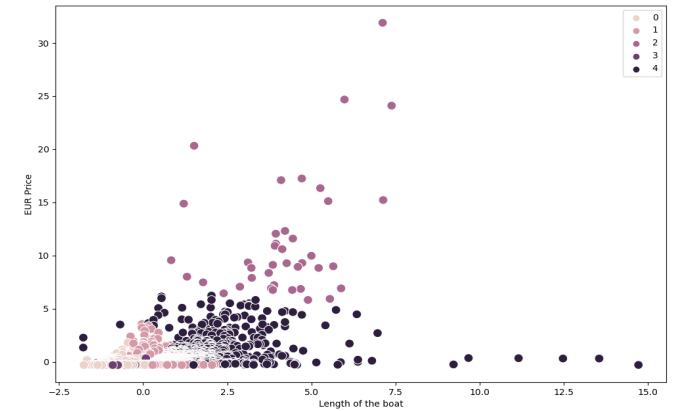- Formulating and testing hypothesis with analytical approaches
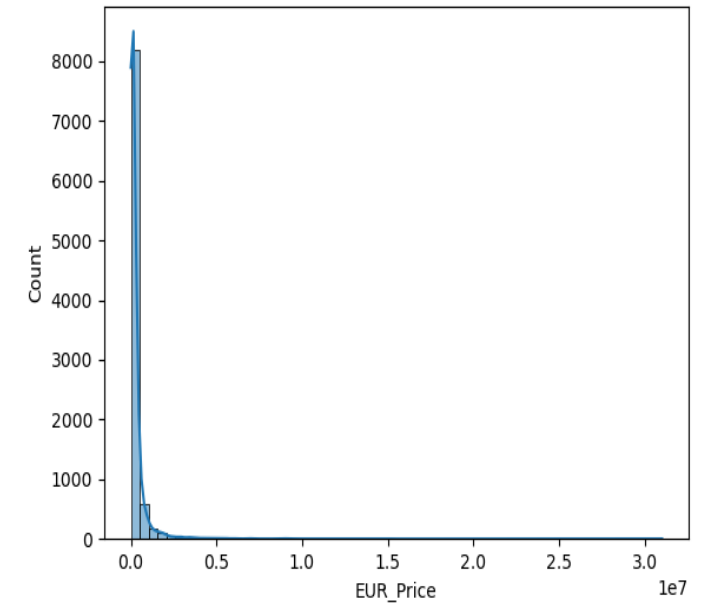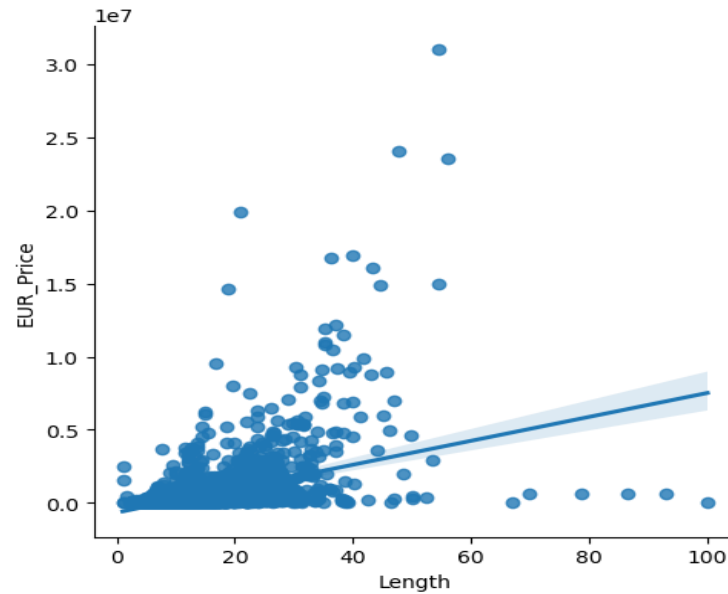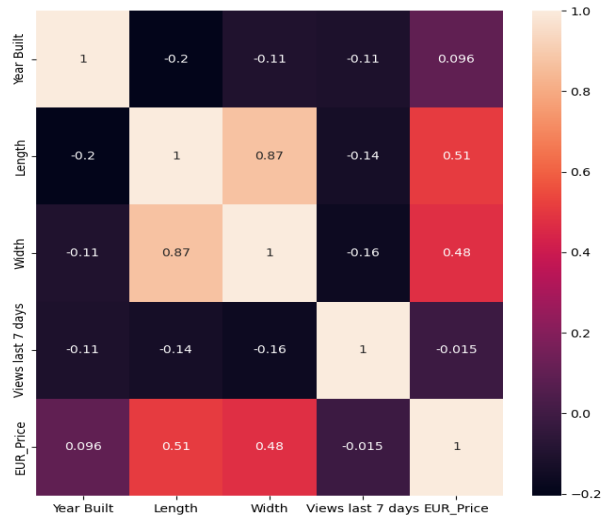
Machine Learning &Time Series Analysis

Boat Length vs EUR Price (Test set)
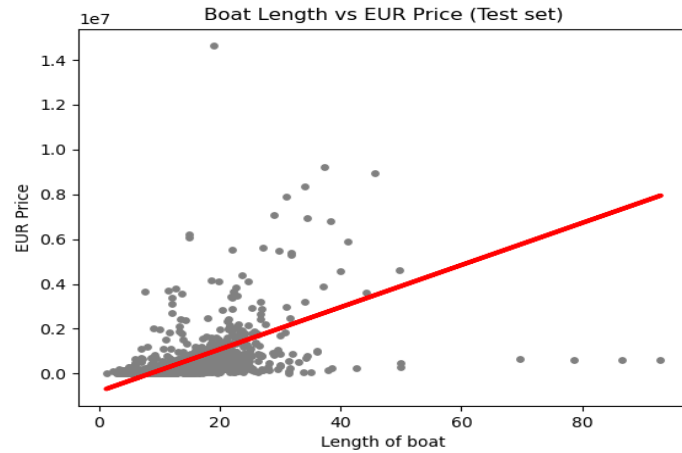
Using Machine learning Analysis to examine:
- Linear Regression
- K-means Clustering
- Time Series

Autocorrelation

Boat View Insights

➢ In analysis of the boat data and variables of Euro Price, Length, and Width there is no apparent relationship with the number of views a boat receives.

➢ In determining which factors impact views and price of boats other variables need to be considered which may include age, material, boat model, location etc., of the boat.

➢ Recommendation is for additional analysis performed to help gain better insight to what other variables may impact the number of boat views.