Data Analytics Portfolio

Gwyn Reniers

Table of Contents

Project 1

GAMECO ANALYSIS

GAMECO SUMMARY

Excel was used to perform analysis of globally gathered data to help company understand how to better market their product in the world of online gaming.

Data

- Data: VGChartz Website
- From 1980-2016
- Global Units Sold

Skills

- Data cleaning
- Data integration
- Pivot tables
- Visual analysis

Tools

- Excel
- Tableau

Project 1 Analysis

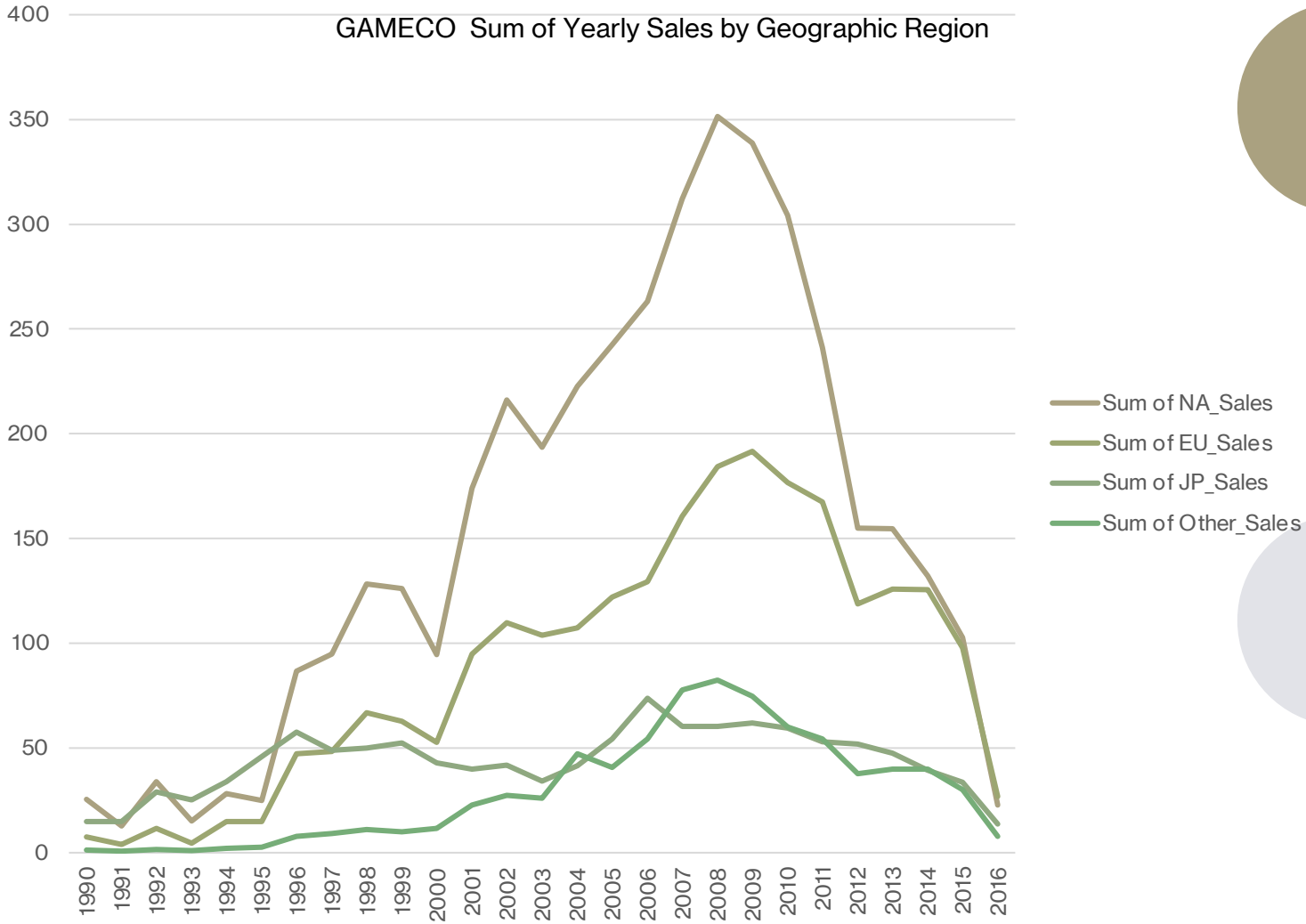
Sales by Year

Procedure

- Analyzed video game sales trends from 1980-2016.
- Developed global business strategy for future

Generated Assessments On:

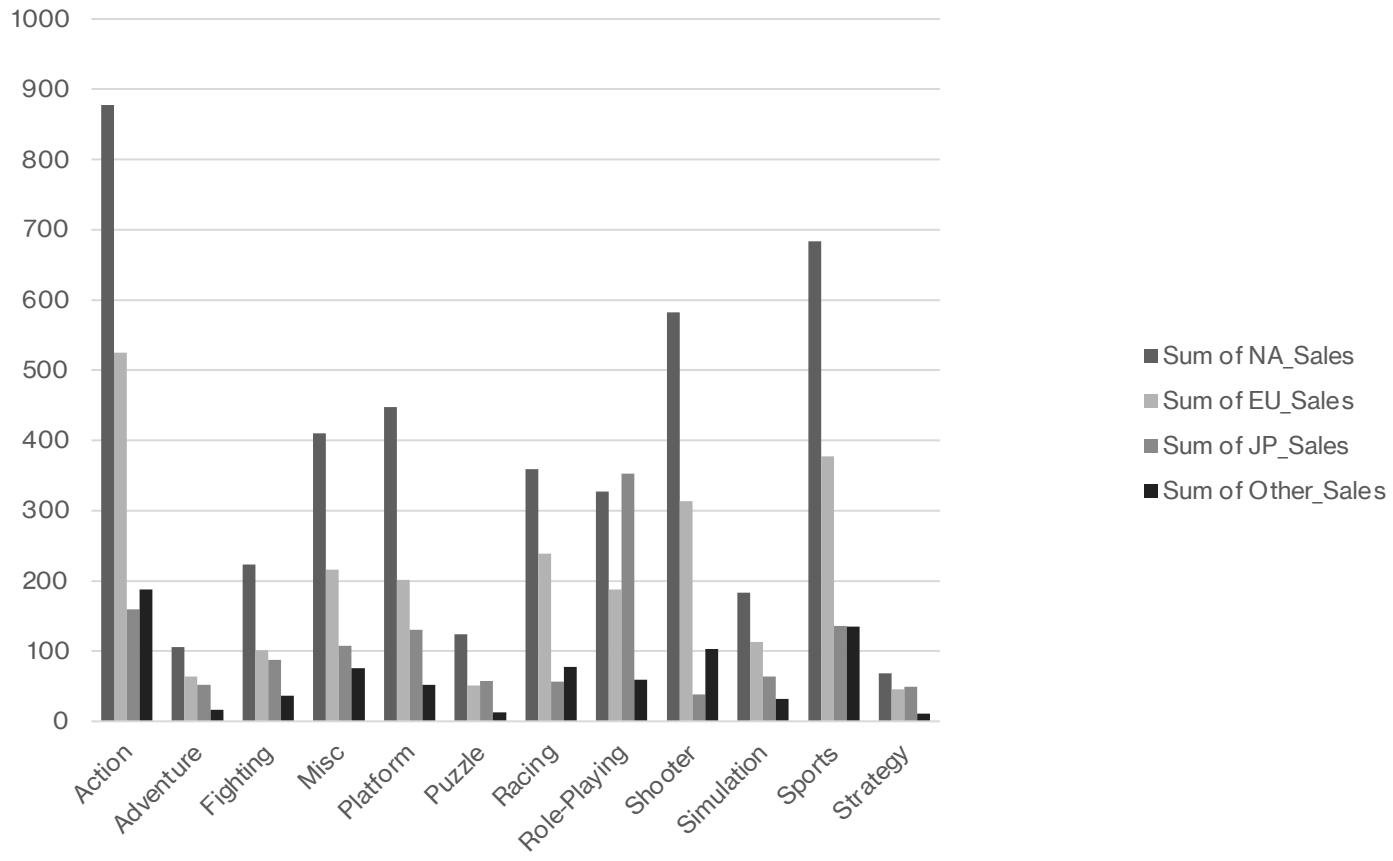
- North America Sales
- Japan Sales
- Europe Sales
- Other Regions Sales



*Sample visual from analysis showing sum of yearly sales per region from 1990-2016.

Project 1 Analysis

Sales by Genre



Genre Observations

- 75% of market share belongs to the Shooter, Action, and Sports genres
- Shooter, Sports, Role-playing and Action genres have consistently been the most lucrative, and therefore should receive marketing priority

Project 1 Result

PROJECT BRIEF

FINAL PRESENTATION

PROJECT REFLECTIONS

- Sales have not remained constant over the past 10 years
- For past 10 years North America has been leading market but began to decline in 2010
- Europeans Sales have continued to rise
- Japanese sales are lowest, but have risen since 2015
- Top selling genres globally: Action, Sport, Shooter, Role Playing
- Lowest producing genres: Strategy, Puzzle and Adventure

*Data Limitations: Data was not gathered in its entirety after 2016. Some areas may not have reported data due to lack of access underrepresenting some areas globally.

Project 2

Influenza Analysis

Influenza Summary

Analysis and visual presentation of cleaned and combined data from the CDC and U.S. Census to determine past trends in Influenza spread to help avoid over and under staffing

Data

- CDC Influenza death rates
- U.S. Population Census Bureau
- CDC Influenza lab test results
- CDC Flu shot rates in children

Skills

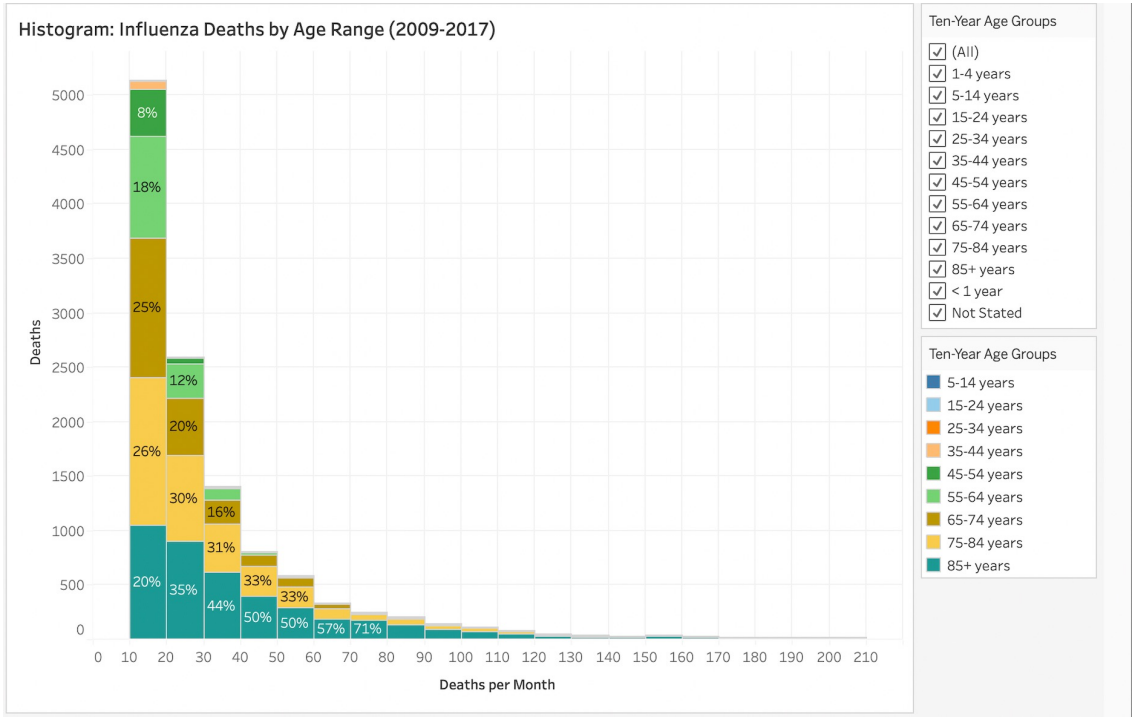
- Data cleaning
- Data integration
- Story telling in Tableau
- Visual analysis
- Statistical hypothesis testing

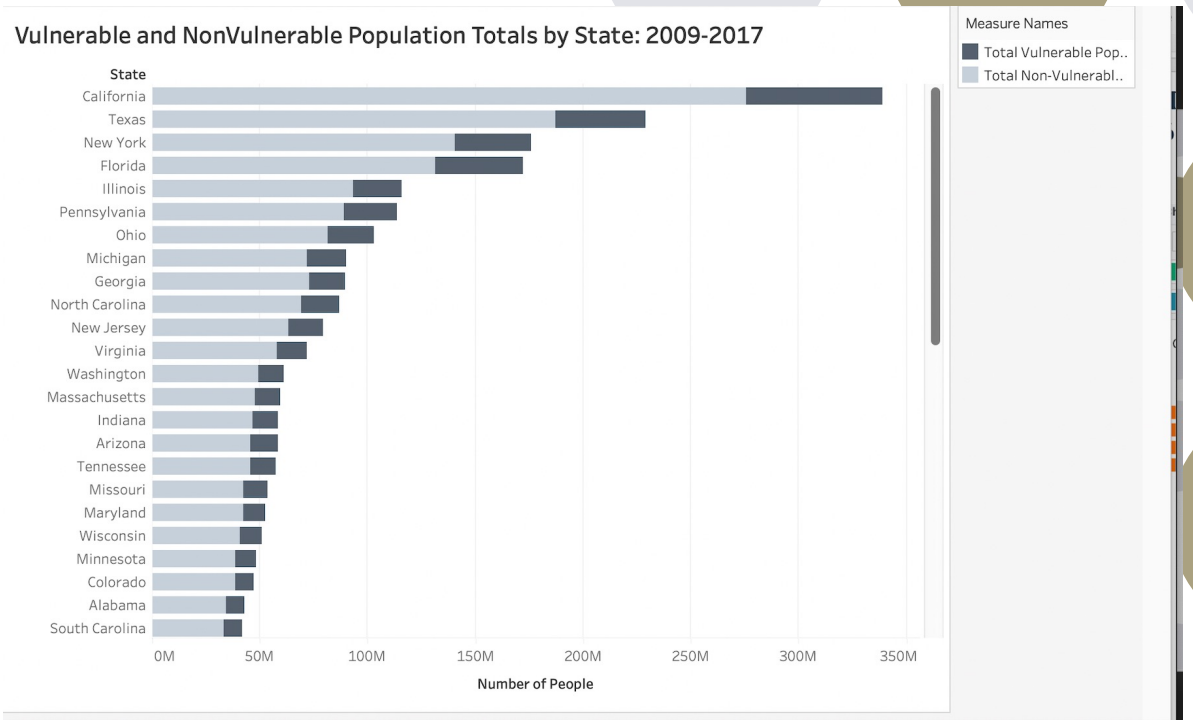
Tools

- Excel
- Tableau

Project 2 Analysis

Influenza and Vulnerable Populations





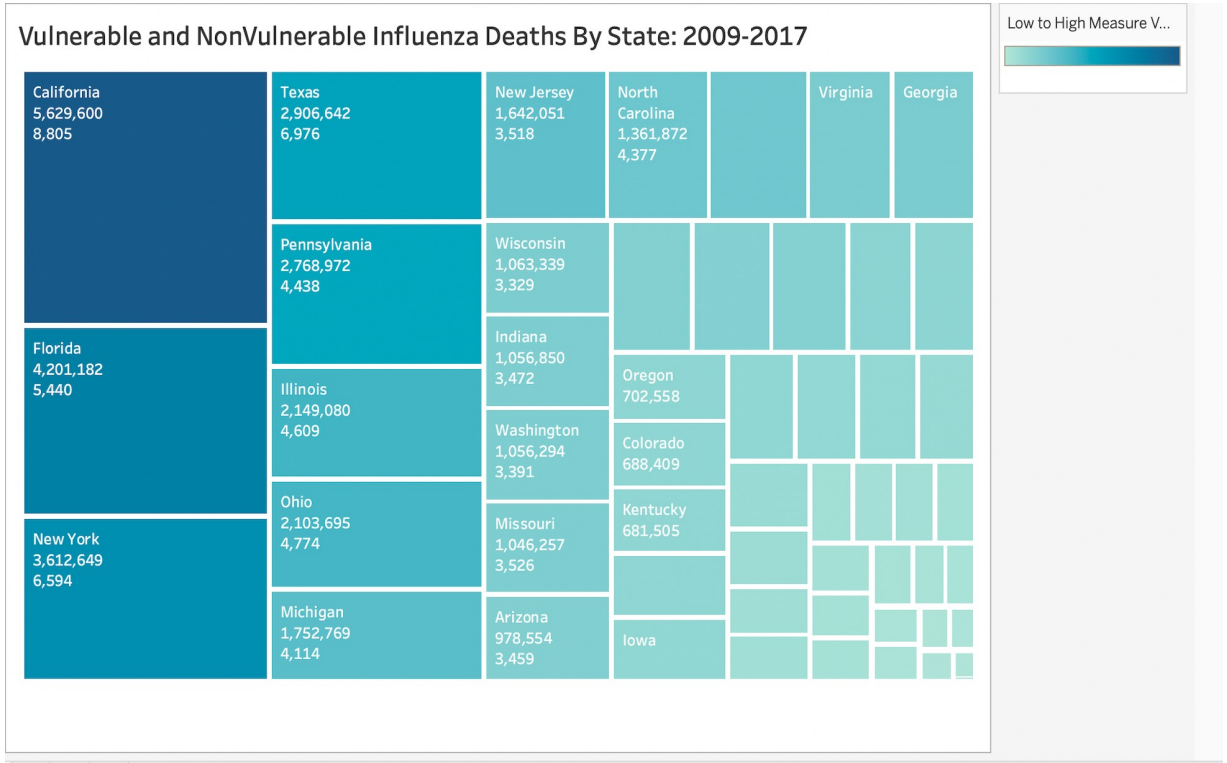
- Positive correlation between population and influenza. States with large vulnerable populations have the highest influenza rates.
- Total population of the state is the driving factor for influenza, with deaths occurring within the vulnerable population increasing as this population rises.

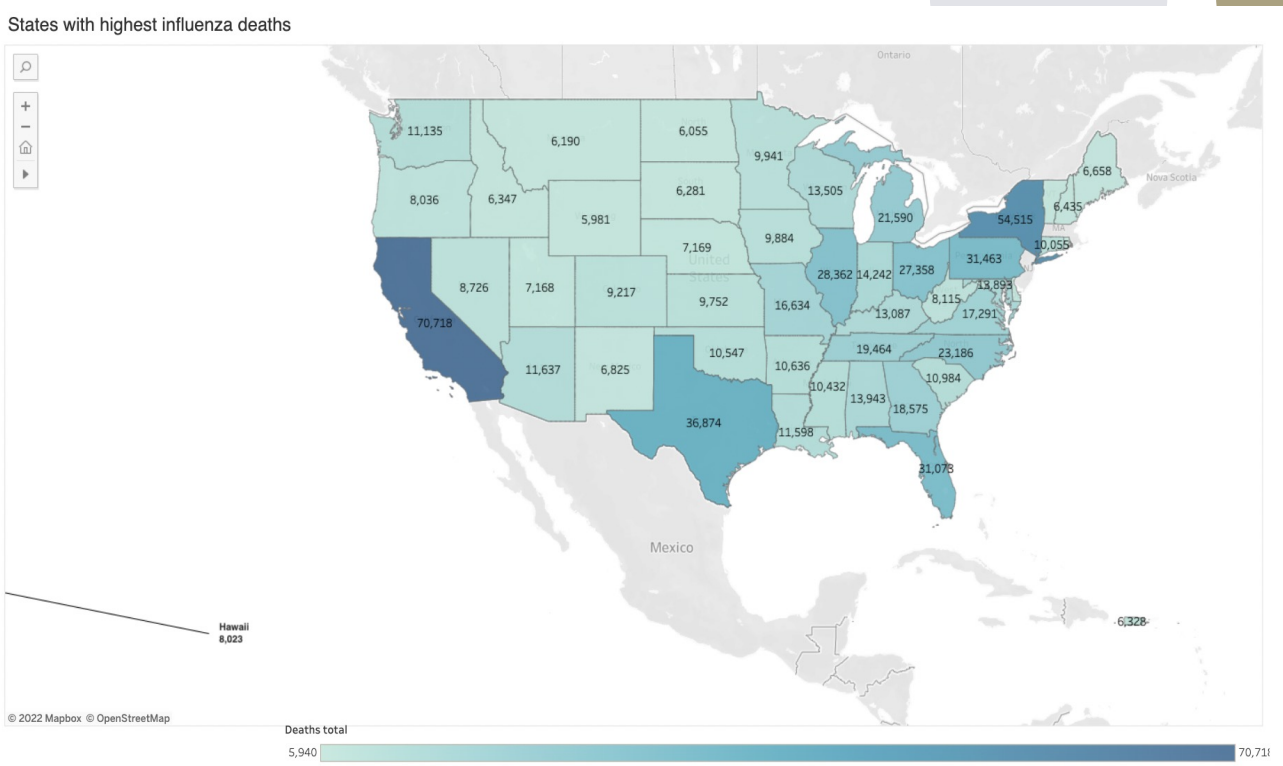
Project 2 Analysis

Influenza and Geography

Hypothesis: States with vulnerable populations (<5 years and 65+) would have the largest amount of influenza deaths and cases, requiring more staff and support

Conclusion: 10 states with the highest populations also had the highest vulnerable populations and the highest influenza rates. More staff should be sent to
CA, NY, TX, PA, FL, IL, OH, NC, TN, MA, MI, GA, VA, NJ and MS





Project 2 Result

RECOMMENDATIONS

- Staffing should be centered around the most populous states and those with the highest vulnerable populations.
- Staffing in these areas need to consider the level of care that will be needed
- Consider impact of Holiday season (End of Nov- Beginning of January)
- Availability of traveling nurses or health care staff for extra coverage

PROJECT BRIEF

FINAL PRESENTATION

DATASET

• DATA/ANALYSIS LIMITATIONS

- The us census population data is 10 years old, and we know that the US population is aging so this will increase the number of individuals that fall into the vulnerable group.

- As stated at the beginning, the vulnerable population in this analysis only takes age into consideration, and therefore other vulnerable groups will be affected by Influenza.

- The data gathered from the CDC was limited in its scope and therefore this analysis is only as accurate as the data provided. With over 80% of data returned as suppressed, and this data representing numbers less than 10, there was a large degree of estimation and possibly error.

Project 3

Rockbuster Analysis

Project 3 Summary

Objective: Analyze data from global online movie rental company to determine launch strategies for new online video service

Data

Fictional Film and rental data loaded into RDMS

DATA DICTIONARY

Skills

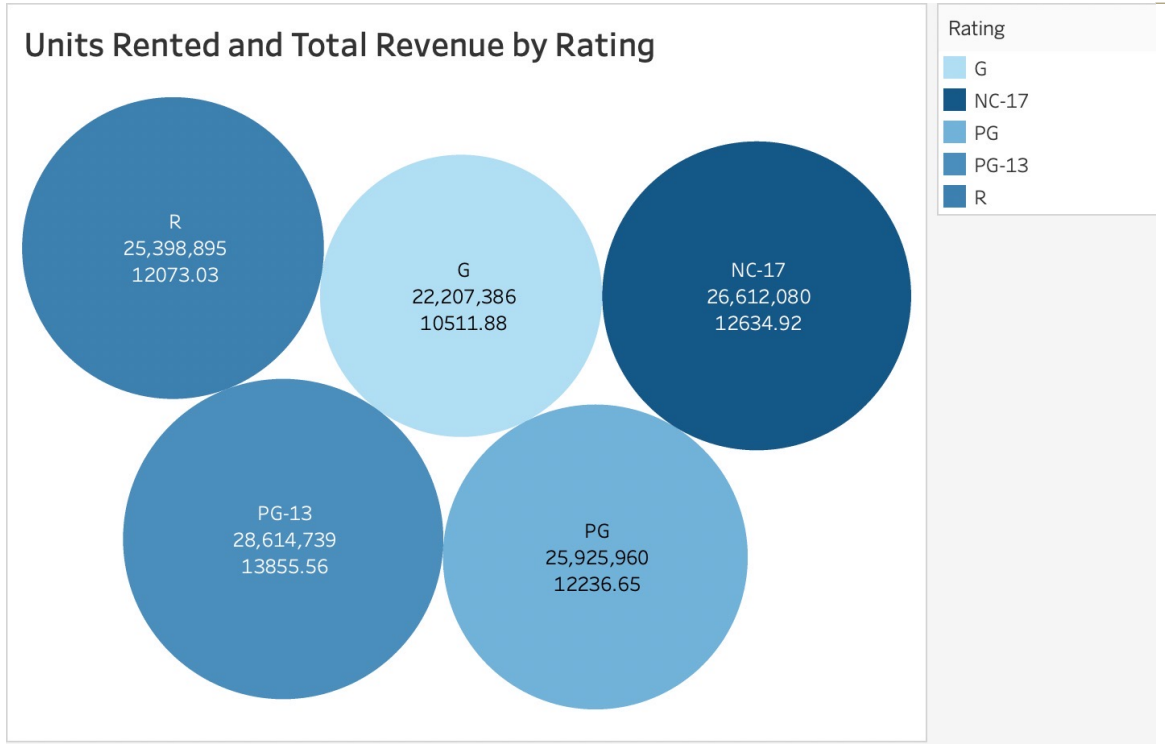
- Relational databases
- Database querying
- Filtering
- Cleaning/Summarizing
- Joining tables
- Subqueries
- Common table expressions

Tools

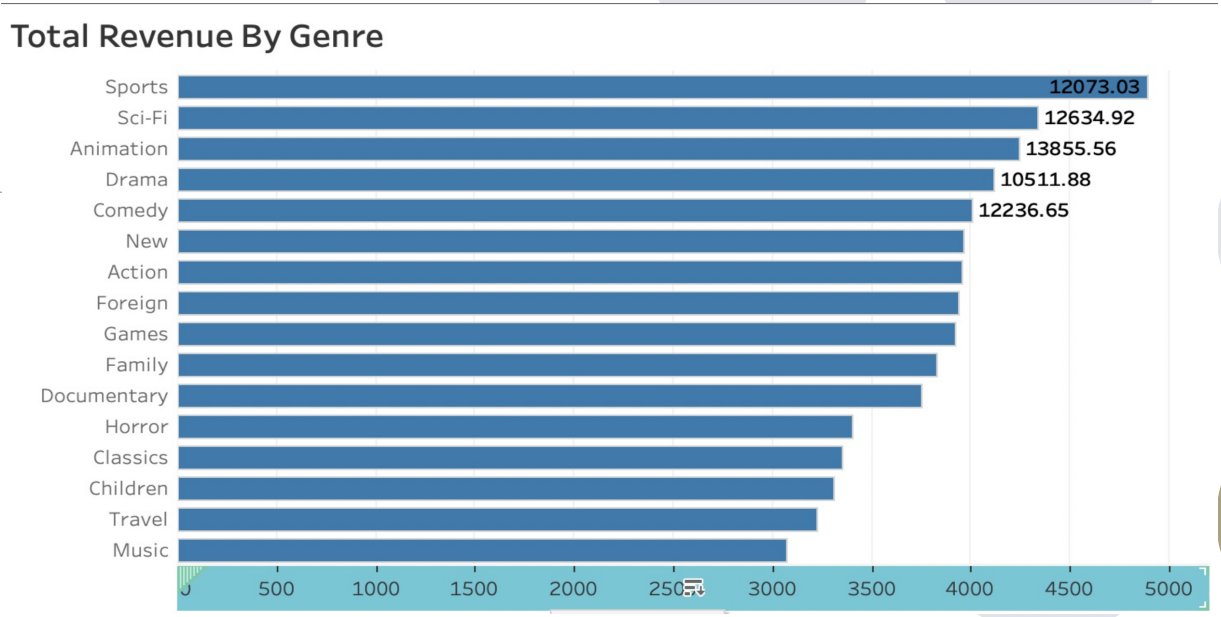
- PostgreSQL
- Tableau

Project 3 Analysis

Revenue by Rating and Genre



Increase Revenue by promotion and availability of most popular genres and ratings.



Project 3 Analysis
Rental Geography

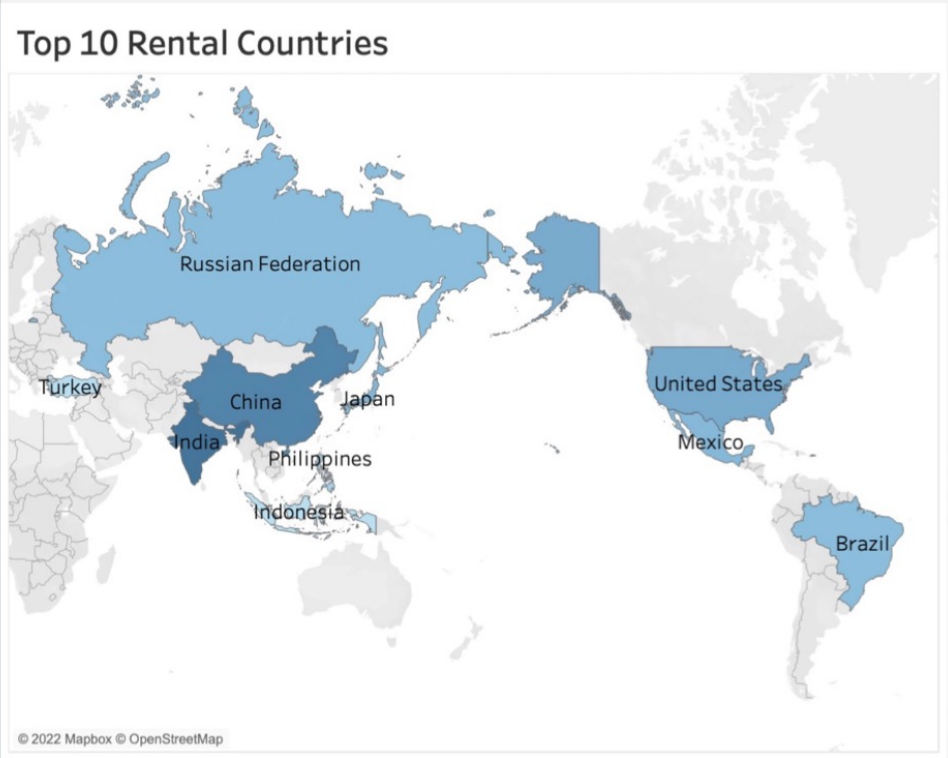
- **REVENUE BY RATING**

- Target audience ratings through movie promotions of specific ratings.

- Be country specific- identify cultural boundaries and use respect for this as campaign.

TOP 10 RENTAL COUNTRIES

INDIA	$6033
CHINA	$5247
UNITED STATES	$3694
JAPAN.	$3122
MEXICO	$2985
BRAZIL	$2919
RUSSIAN FEDERATION	$2766
PHILIPPINES	$2220
TURKEY	$1498



Project 3 Results

Recommendations

PROJECT BRIEF

FINAL PRESENTATION

DATA

STATISTICS

Total Customers.....599
Total Films.....1000
Average Rental Rate.....$2.98
Average Replacement Cost.....$19.98
Most Common Language.....English
Most Common Rating.....PG-13

RECOMMENDATIONS

- Explore subscription style model
- Market Country specific ratings (consider culture)
- Increase number of movies in top selling genres and ratings

*There were no data limitations to this dataset as it was all fictional

Project 4

Instacart Analysis

Project 4 Summary

Objective: Gather information on consumers and their usage of Instacart by using exploratory analysis to generate marketing insights and increase engagement and revenue

Data

Instacart open source
Fictional customer data

Skills

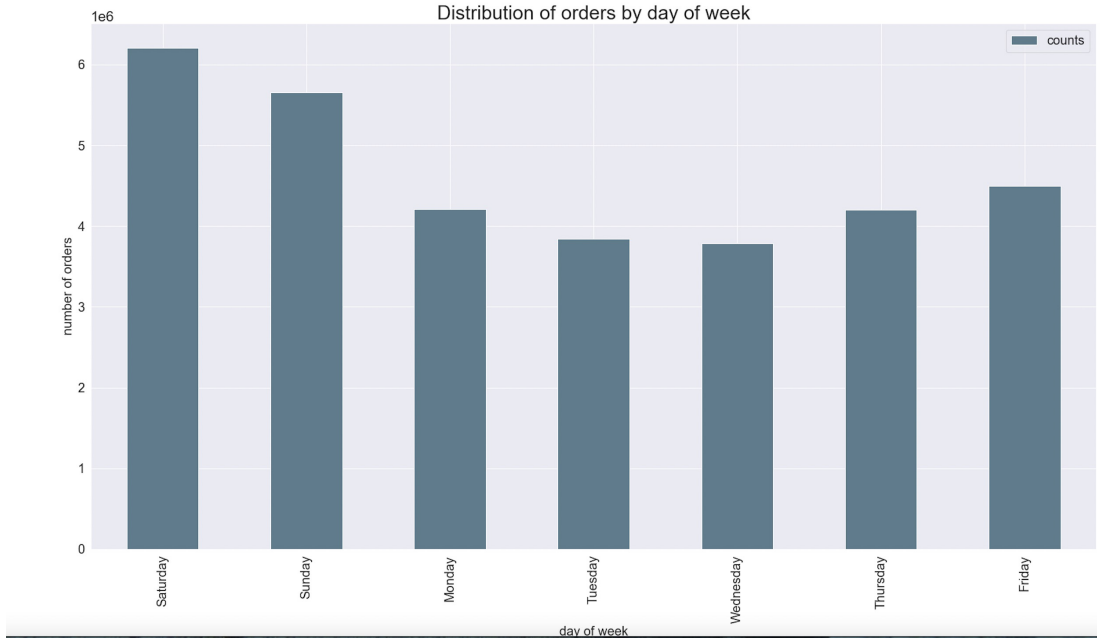
- Python3
- Jupyter Notebook
- Pandas, Seaborn, Numpy & Matplotlib libraries

Tools

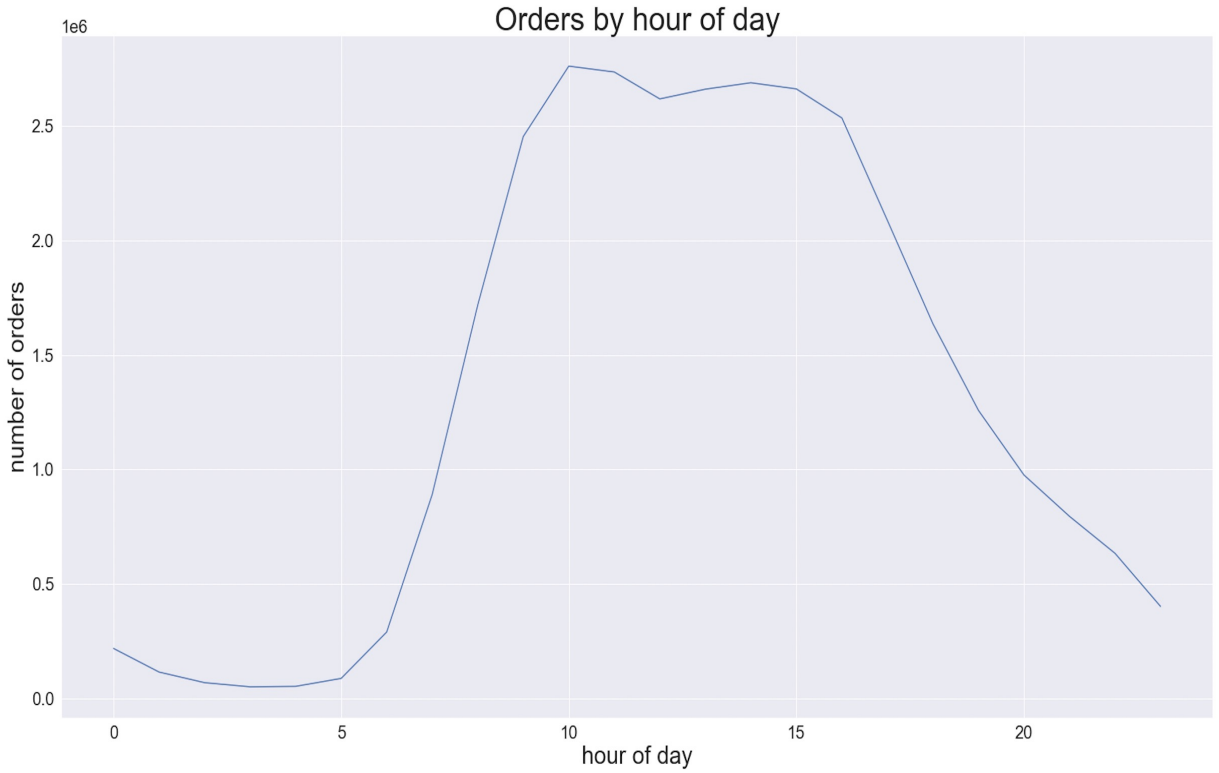
- Data cleaning and wrangling
- Data merging
- Deriving variables
- Grouping and summarizing
- Aggregating data
- Population flows

Project 4 Analysis

Order Day/time



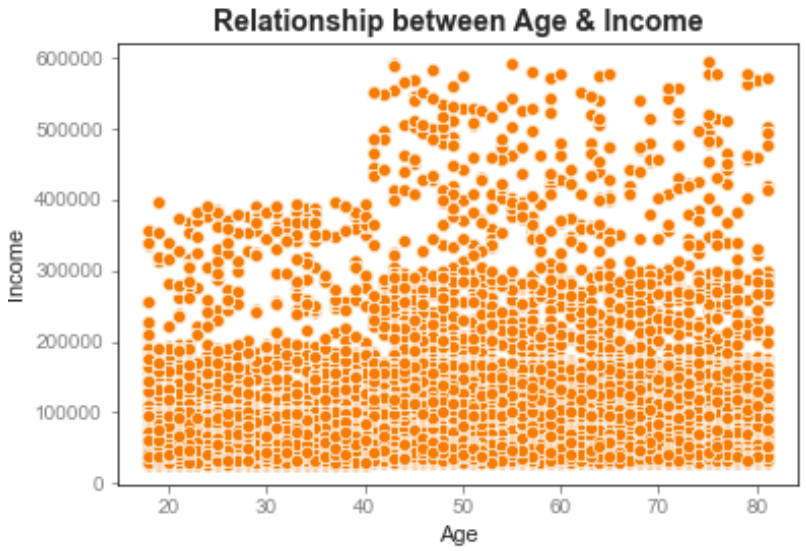
To better target consumers, ads should be placed between 7pm and 11pm, and between midnight and 7am



Analysis of the data showed the busiest days of the week were Saturday and Sunday, and the busiest hours of the day were between 8am and 6pm.

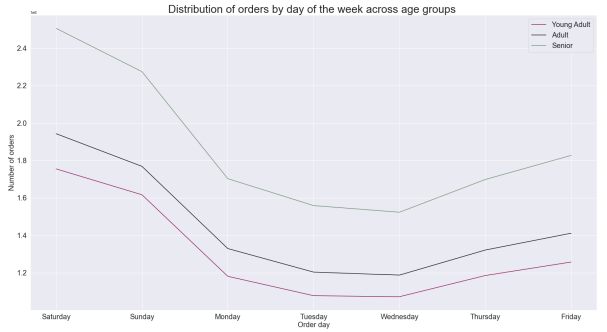
Project 4 Analysis

Age Profile

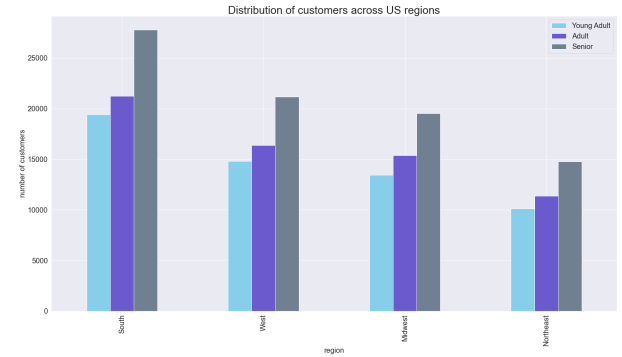


Correlation between age Income and Instacart use

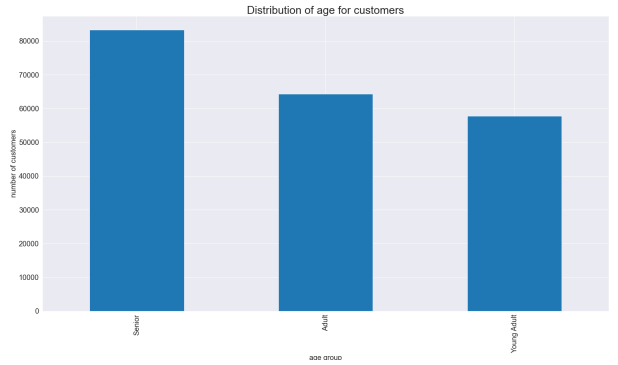
The data shows that customers have similar ordering patterns Regardless of age with orders being higher at the beginning of the week and dropping during weekdays



More orders were placed by seniors, followed by middle aged adults and lastly younger adults. Ads Should be targeted too young Adults to increase market share

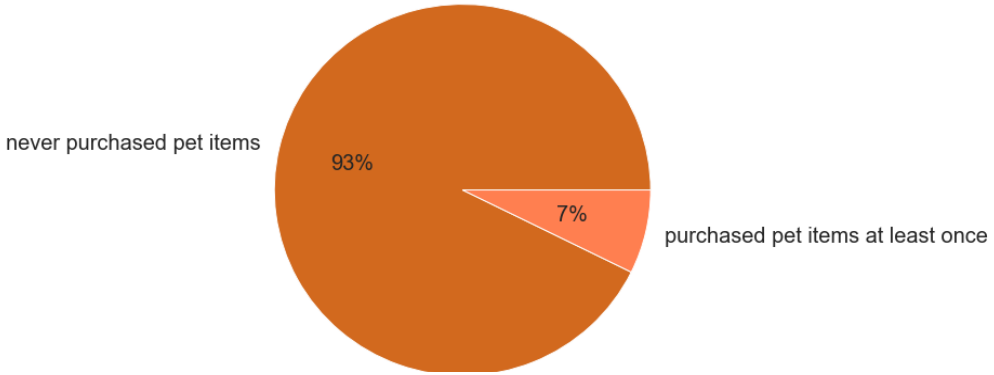


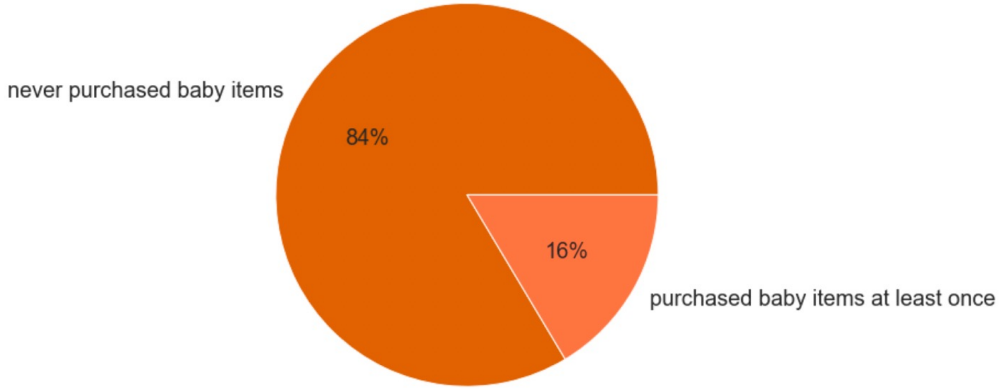
Regionally, orders were placed similarly, with seniors and older adults placing the most orders. The South places the most orders followed by the West, but population must be considered.

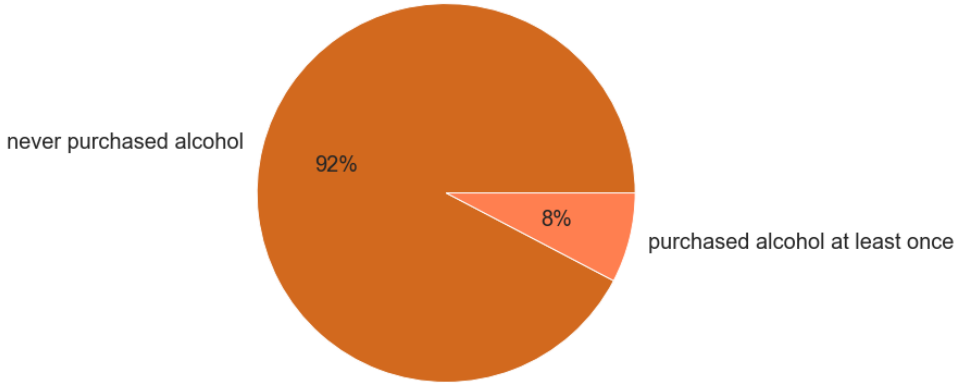


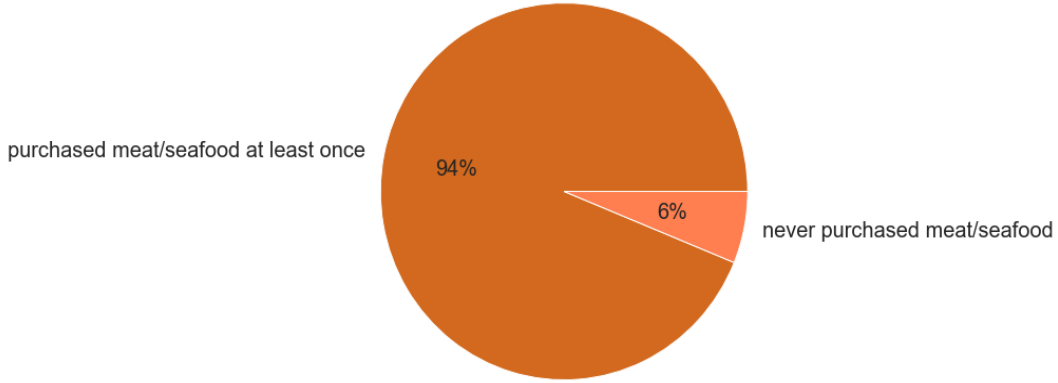
Project 4 Analysis

Customer Profiles









Profile flags created based on a customer's buying habits (alcohol drinkers, vegans, people with babies and pet owners) show areas of growth for marketing specific ads.

Project 4 Results

PROJECT BRIEF

FINAL REPORT

GITHUB

FINDINGS

- Busiest days of the week: Saturday and Sunday
- Busiest hours between 10AM and 3PM
- The top five departments in order of frequency:
  Produce, Dairy/eggs, Snacks, Beverages and frozen foods
- Seniors place the most orders followed by middle aged adults
- Customers are most likely to be mid to high income
- There is very little regional variation on spending by income and age
- Most expensive items are purchased between 12AM and 7am

Dataset limitations/Issues: **PRIVACY**

"We sometimes share information with third parties to process information on our behalf or to otherwise provide certain services (such as delivery services, advertising services, or information to better tailor our services to you)" June 15, 2022

Project 5

Pig E. Bank

Project 5 Summary

Explored big data, data mining, predictive models, regression, and forecasting to provide analytical support bank to reduce money laundering.
Explored ethics and responsibility in data analysis.

Data

Data set for project 5

Skills

- Data ethics
- Big data
- Data mining
- Time series analysis/forecasting
- Predictive analysis
- Decision tree

Tools

- Excel
- Github

Project 5 Analysis

Questions

1.What does the CRISP-DM methodology for data mining entail? What does the data analyst do in each step?

2.Explain what needs to be done before you can forecast time series. In other words, what does time series analysis involve?

3.Discuss the main principles of data ethics, as well as what companies or organizations could do to uphold these principles.

4.Imagine you're an analyst at a financial services company and you have some concerns about bias in the way a predictive model was measured. How would you raise your concerns with your manager? Make some suggestions for how to avoid measurement bias in the future.

5.Explain the difference between regression and classification models in predictive analysis. When would you use one of these models over the other?

Project 5 Result

GITHUB

INTERVIEW QUESTIONS

Project 6

World Happiness Report

Project 6 Summary

Explored World Happiness Data set to determine variables of happiness in South America
Compared how variables affected happiness score in each County
Determined greatest and weakest factors of happiness on the Continent

Data

- World Happiness Report

Skills

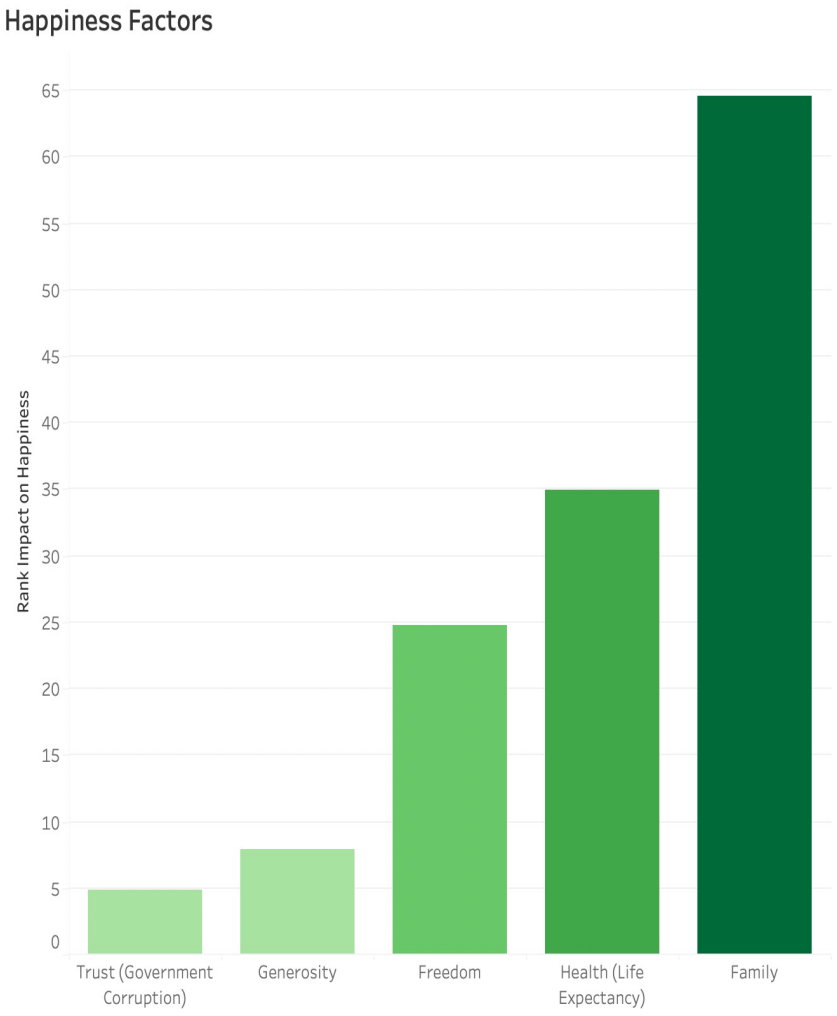
- Data Sourcing
- Data cleaning
- Data Wrangling
- Data Merging
- Regression Analysis
- Cluster Analysis
- Visualization/Dashboards

Tools

- Python
- Tableau
- Github

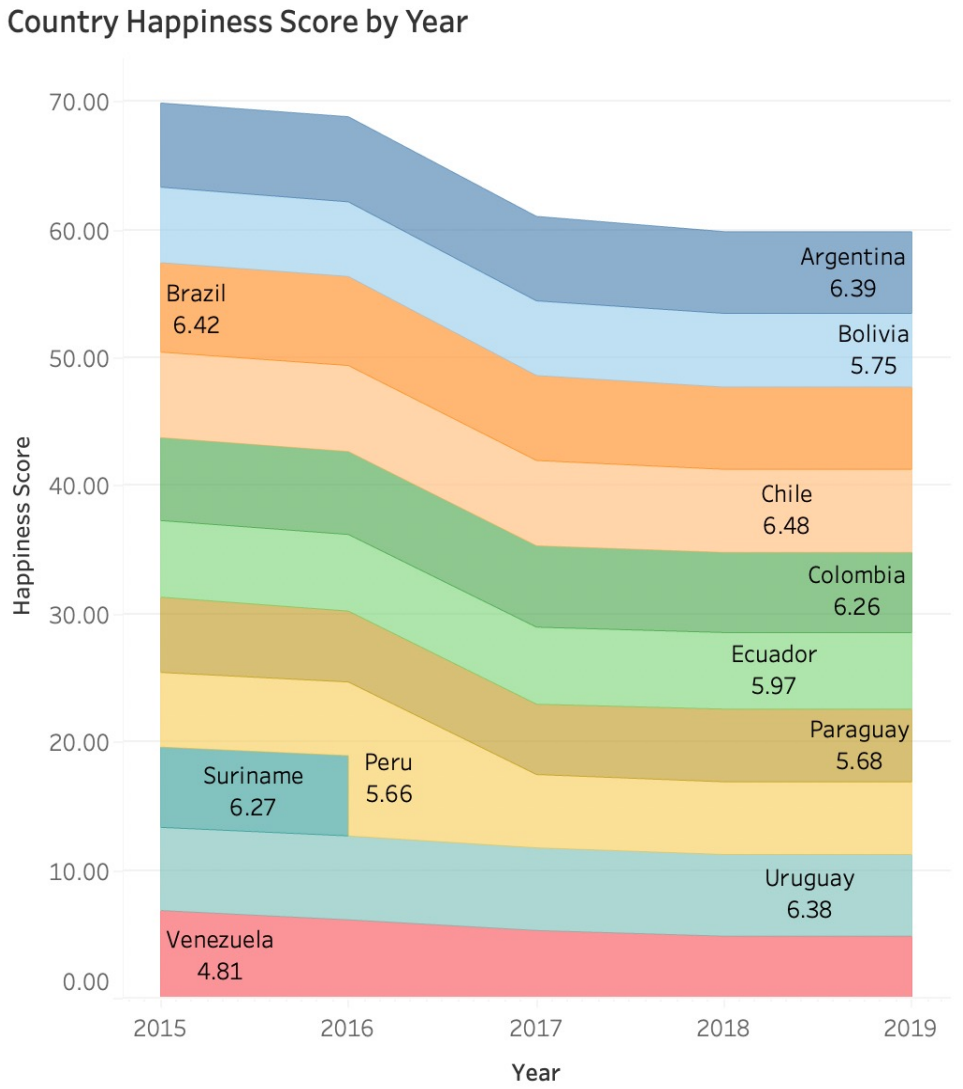
Project 6 Analysis

Happiness Variables/Score



Happiness is affected by variables with Family being the greatest influence of happiness in South America, and Perception of Corruption (Trust in Government) having the least effect.

Happiness scores in South America have decreased since 2015. The largest drop in happiness occurred in 2016. Ecuador ranks 7th out of the 13 countries In terms of happiness



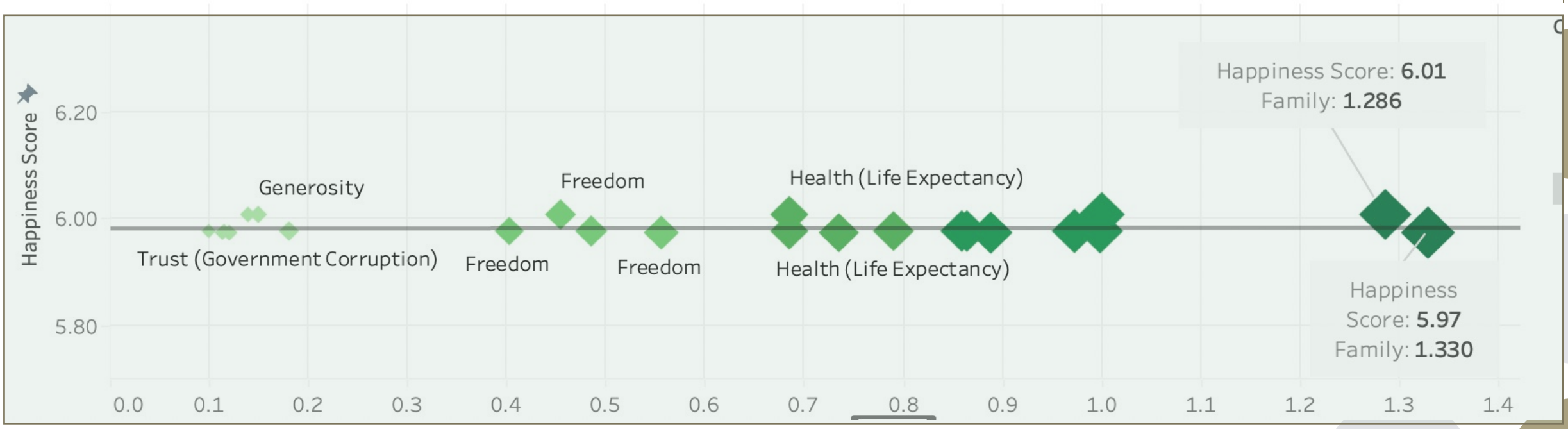
Project 6 Analysis

Ecuador Happiness

FINDINGS

HYPOTHESIS: Ecuador was one of the happier countries in South America. INCORRECT it is 7th out of the 13 countries surveyed. It ranks .6 points down from the highest country in happiness, Argentina, and only .4 points above the most unhappy country, Venezuela

HYPOTHESIS: FAMILY is a driving factor of happiness in South America. CORRECT : There is a positive relationship with family and happiness. GDP contributes 2nd to happiness. Generosity and Perception of corruption (trust in Government) have the least impact.



PROJECT 6 ANALYSIS

Cluster Analysis

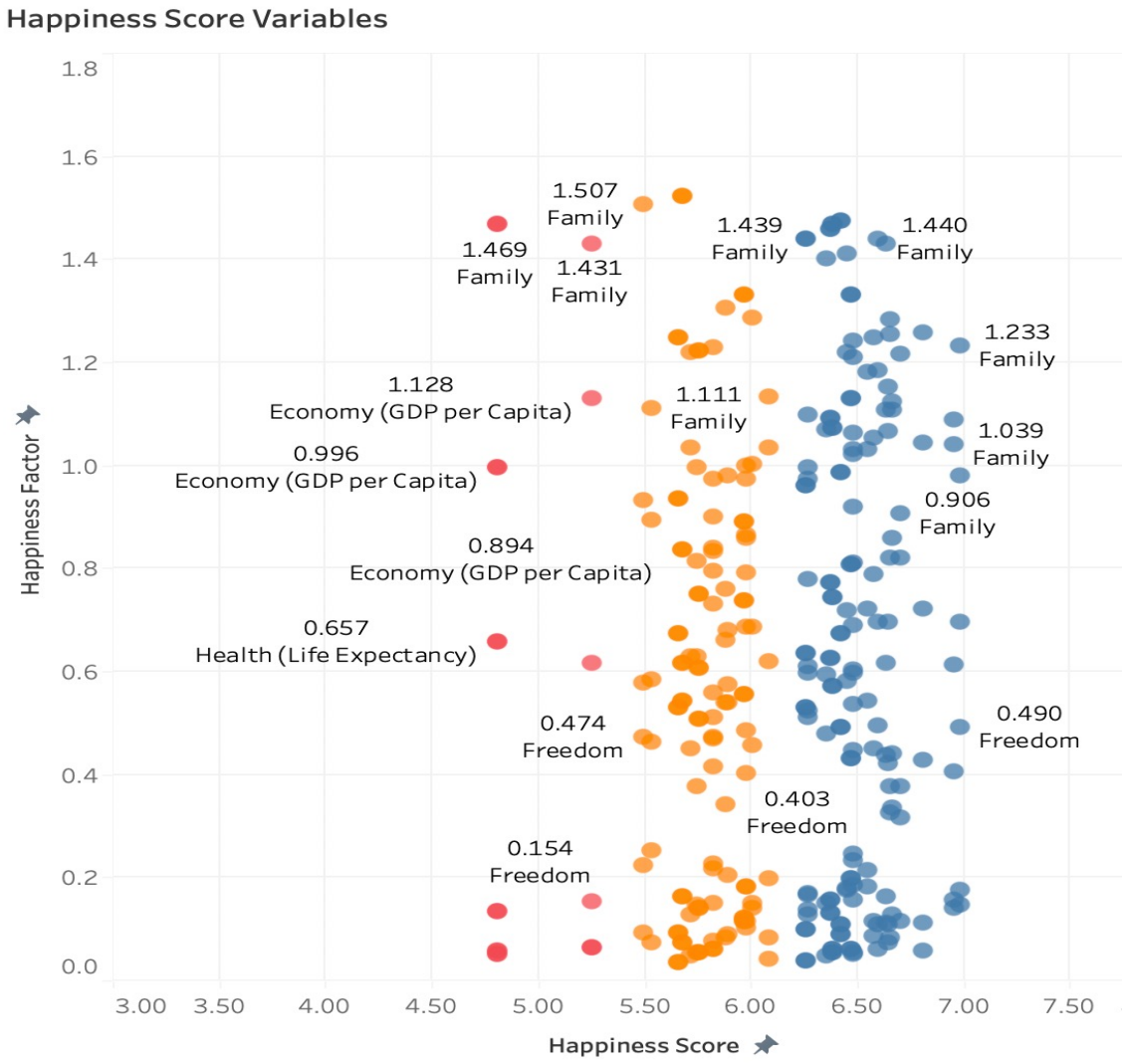
Cluster Analysis was performed to better group the happiness variables of each country in South America and determine the overall driving factors of happiness on the Continent.

Three clusters were used:

RED (Low): Happiness factors with a median = 4.95. These factors have low impact on happiness.

ORANGE (Medium): Happiness factors with a median = 5.80. These factors have a medium effect on happiness.

BLUE (High): Happiness factors with a median = 6.51. These are the most impactful factors on happiness in South America.



Cluster 1

Cluster 2

Cluster 3

StreetMap

Project 3 Results

Summary

PROJECT BRIEF

TABLEAU

GITHUB

South American Happiness Statistics

Leading Factor of Happiness.....Family
Lowest Happiness Factor.....Trust in Government
Happiest Country.....Brazil
Least Happy Country.....Venezuela
Ecuador (My Country).....7th out of 13th
Average Happiness Score of SA.....6.14

Limitations

- Cultural happiness is defined differently and therefore difficult to apply across the globe
- Subjective well-being is difficult to be measured
- Happiness is determined by beliefs and people have vastly varying beliefs
- The sample size of 1000-2000 people is small

CONTACT ME

Github

Tableau

Email: gwynreniers@gmail.com