JORDAN NOVELLI

DATA ANALYST

PORTFOLIO WEBSITE

JORDANMNOVELLI@GMAIL.COM



Tool Images are clickables Links

Image link

TABLE OF CONTENTS

To demonstrate data ethics for security and privacy, I have tried to refrain from using mock personal data. Any personal data that may be showcased within these projects is fabricated.

Image link

GAME CO.

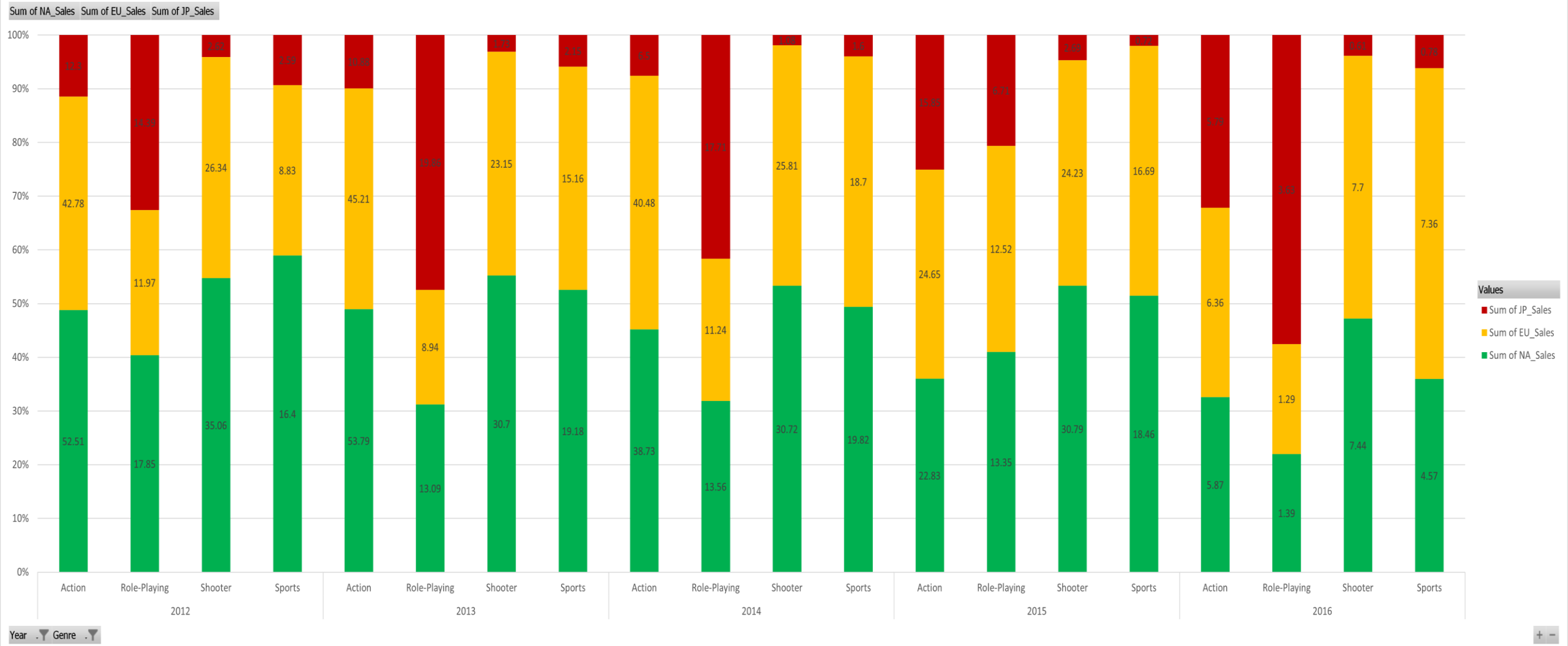
- Conduct descriptive analysis of video game data and build a budget for 2017. Obtain insights into leading genres, comprehend the gaming market, and track the popularity of video games over time.
- Utilize Excel to increase data accuracy and quality, resolve any missing information, create pivot charts, and generate visual representations.

Table of Contents

Table of Contents

GAME CO.

- Explore top game genres by year and region.
- Action and Shooter genres have been top performers among the North American and European regions over a 5-year period (2012-2016)



GAME CO.

- Europe's gaming popularity began peaking in 2016, outperforming North America; helping to curb the budget in Europe's favor.

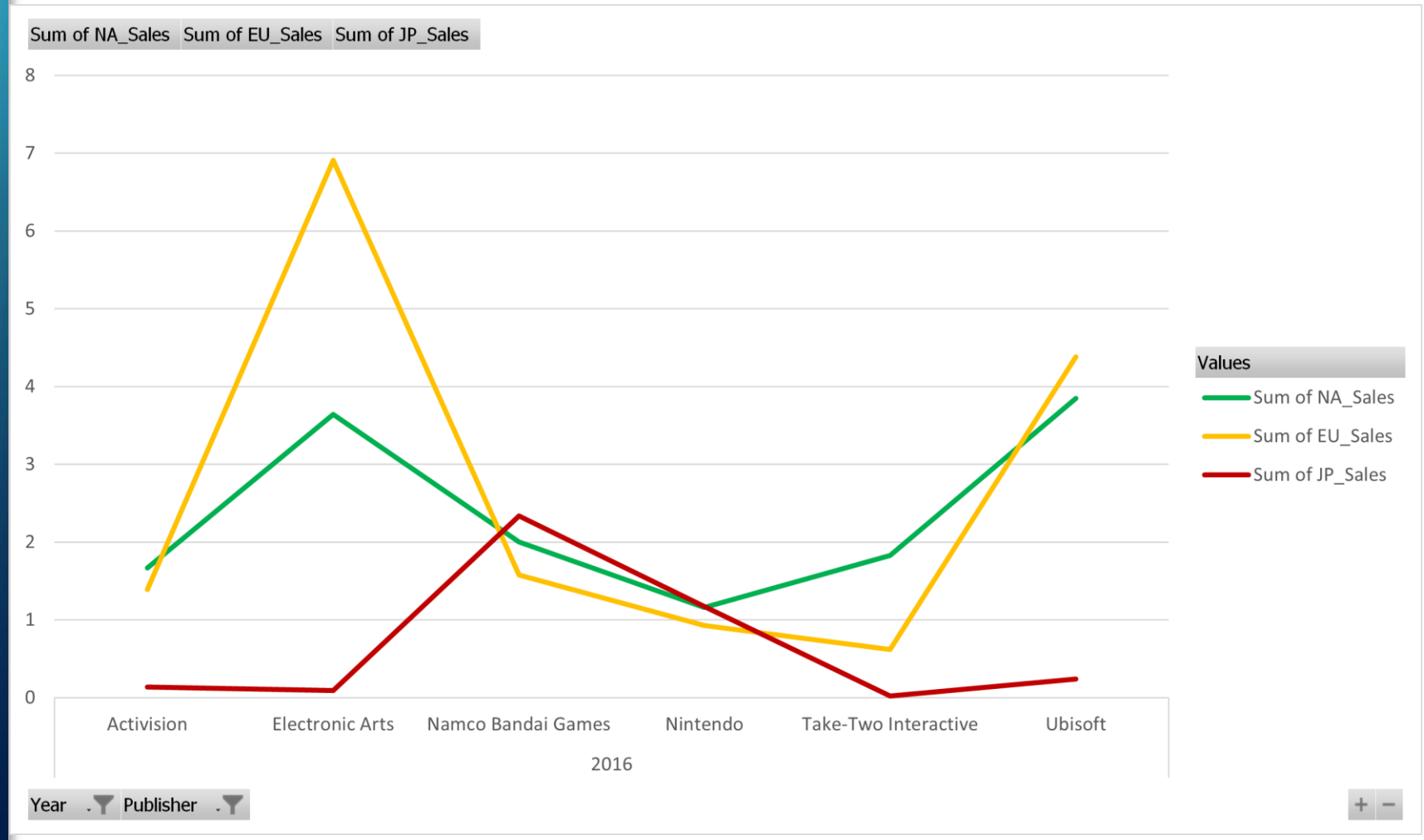


Table of Contents

GAME CO. CONCLUSION

- Video game popularity shifts each year. Publishers and genres have a massive effect especially when comparing what regions, they are popular in.
- Data from this project show that Europe generated more game sales. Recommendations for 2017 is to give Europe the budget to further support their action, shooters and sports genres.
  - Further market development can begin for North America to promote more action and shooter genres.
  - Japan's market can focus on promoting more role-playing and action genres.
- Challenges presented within the data were derived from the years and sales columns.
  - Years: missing values for the years a game was published
  - Sales: sales numbers for specific games or genres were zero.

Table of Contents

FLU SEASON IN THE US

- Using CDC surveys and other data we can develop descriptive analysis to determine when the influenza virus is at its peak. We can create a plan that would benefit medical staffing agencies in boosting influenza treatment with adequate staffing.
- Hypothesis: If hospitals and clinics are staffed appropriately in states that are hit the hardest by the flu, then more patients will be treated sooner and recover quicker.


Image link

Table of Contents

FLU SEASON IN THE US

- Using Excel spreadsheets, we produced descriptive data such as variance, standard deviation and averages.
- Bringing the data together in Tableau the top states effected by influenza were targeted.

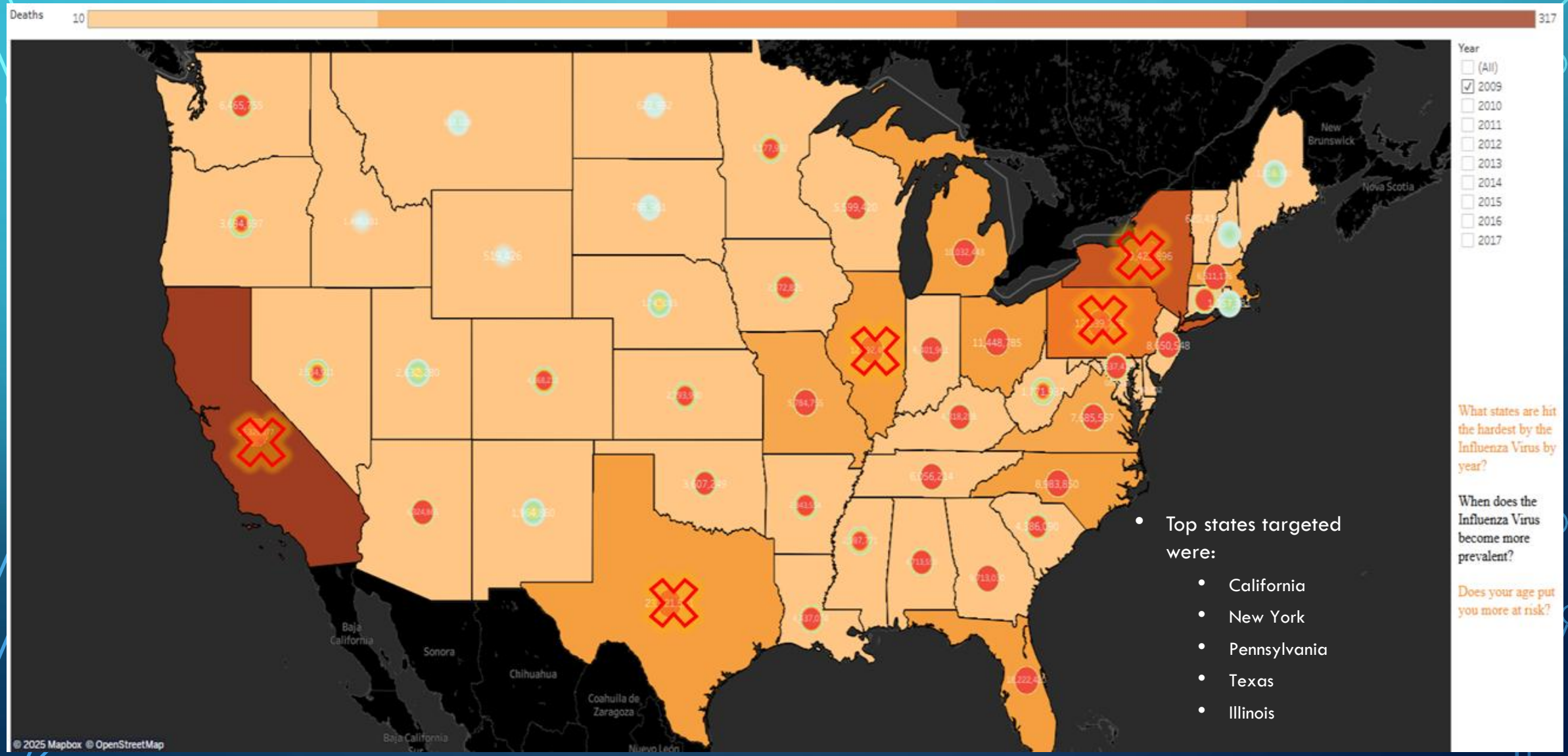
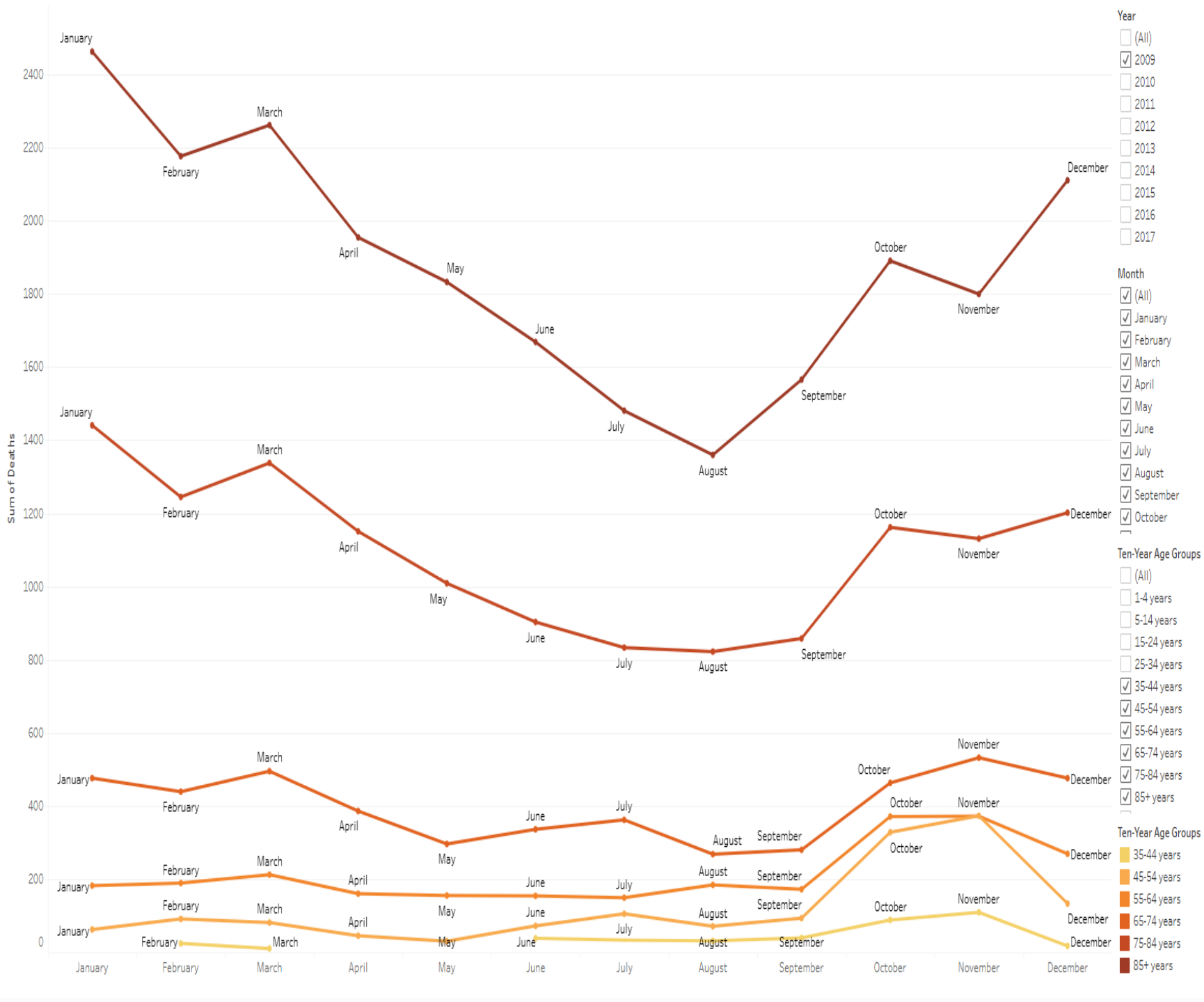


Table of Contents

FLU SEASON IN THE US



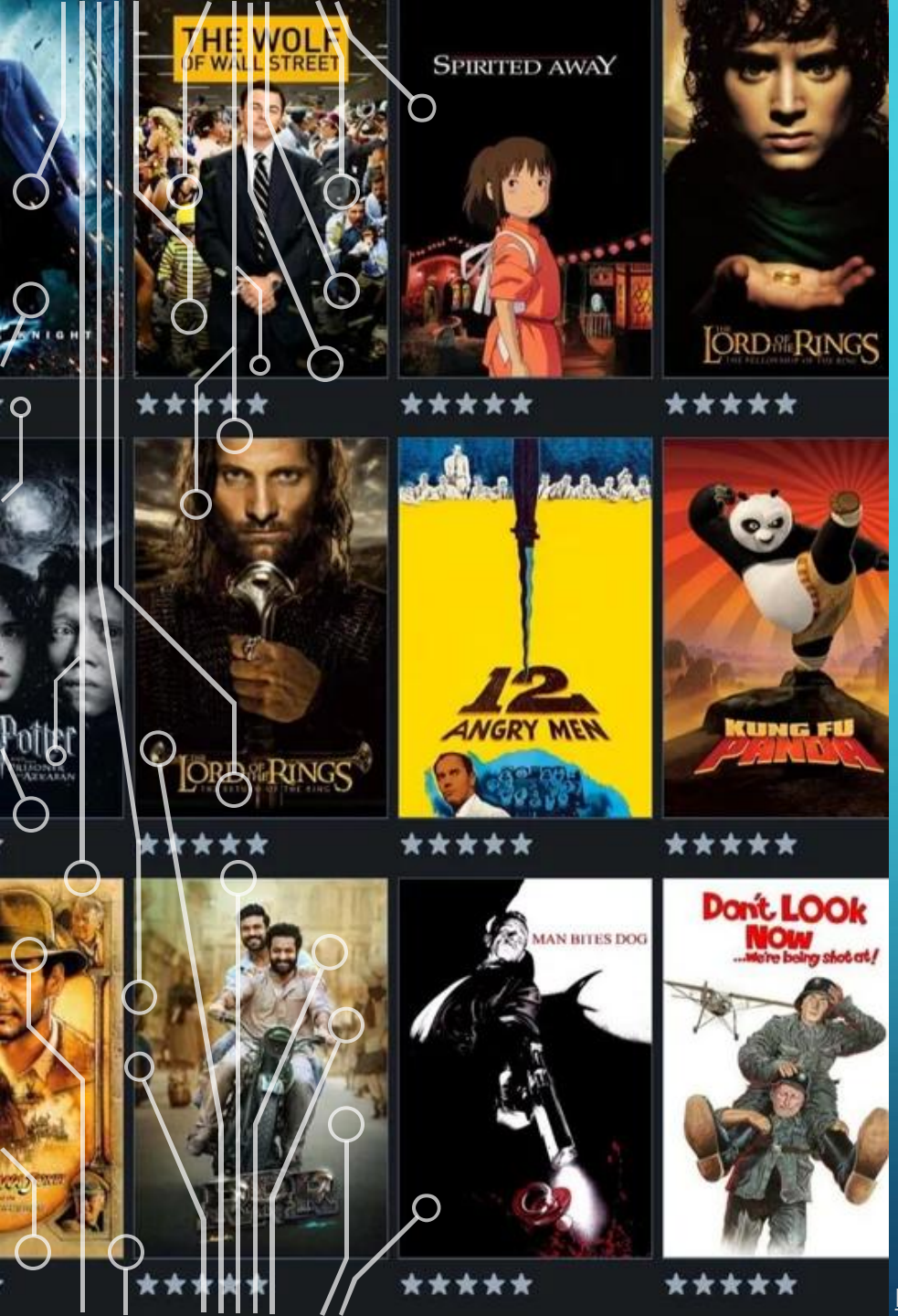
- Peak season for Influenza is January – March and again in October – December.
- Of these peak months, the people more at risk are aged 75-84 years and +85 years old.
- With this combined data we can determine the top 5 states needing increased staff and during which months for specific age groups.

Table of Contents

FLU SEASON IN THE US CONCLUSION

- Limitations and challenges of the data consist of:
  - Total population as it only contains years 2009-2017
  - Deaths documented for ages 9-20 years old are labeled as suppressed data and were not numerical values. This can affect the data set when looking for children that are affected by influenza.
  - Documentation of patients visiting clinics were only collected between 2010-2019 but reported weekly.
  - Additionally, documented deaths by influenza did not include other chronic or acute illnesses that may have contributed to patient death.

Table of Contents



ROCKBUSTER

- Assist Rockbuster in planning an online rental service as they compete with companies such as Netflix and Amazon Prime.
- Explore the data using Structured Query Language (SQL) and cross platform universal database tools such as DbVisualizer.

Image link

Table of Contents

Table of Contents

ROCKBUSTER

- Using SQL, I can produce tables to get precise information.
- Explore database schema's and understand characteristics, keys, and relational tables.

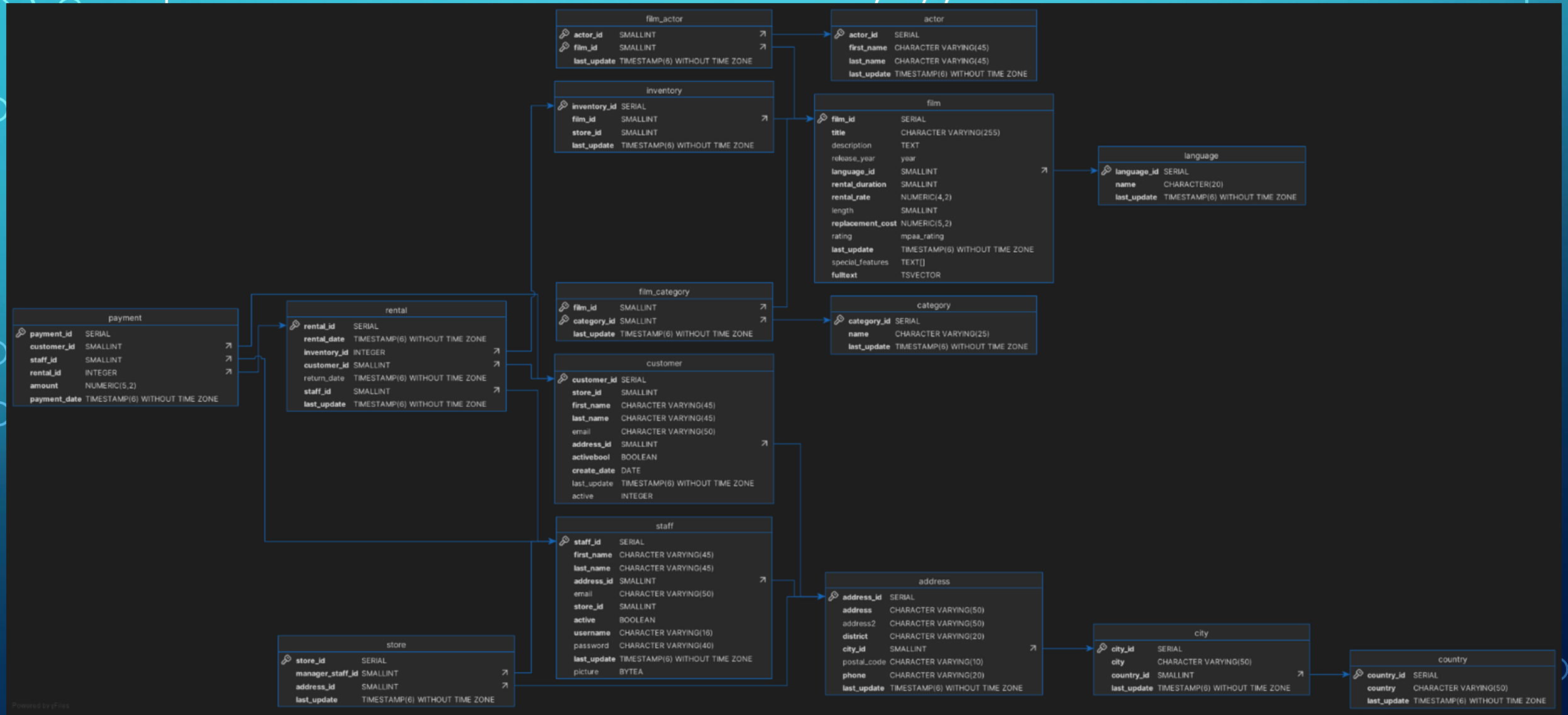
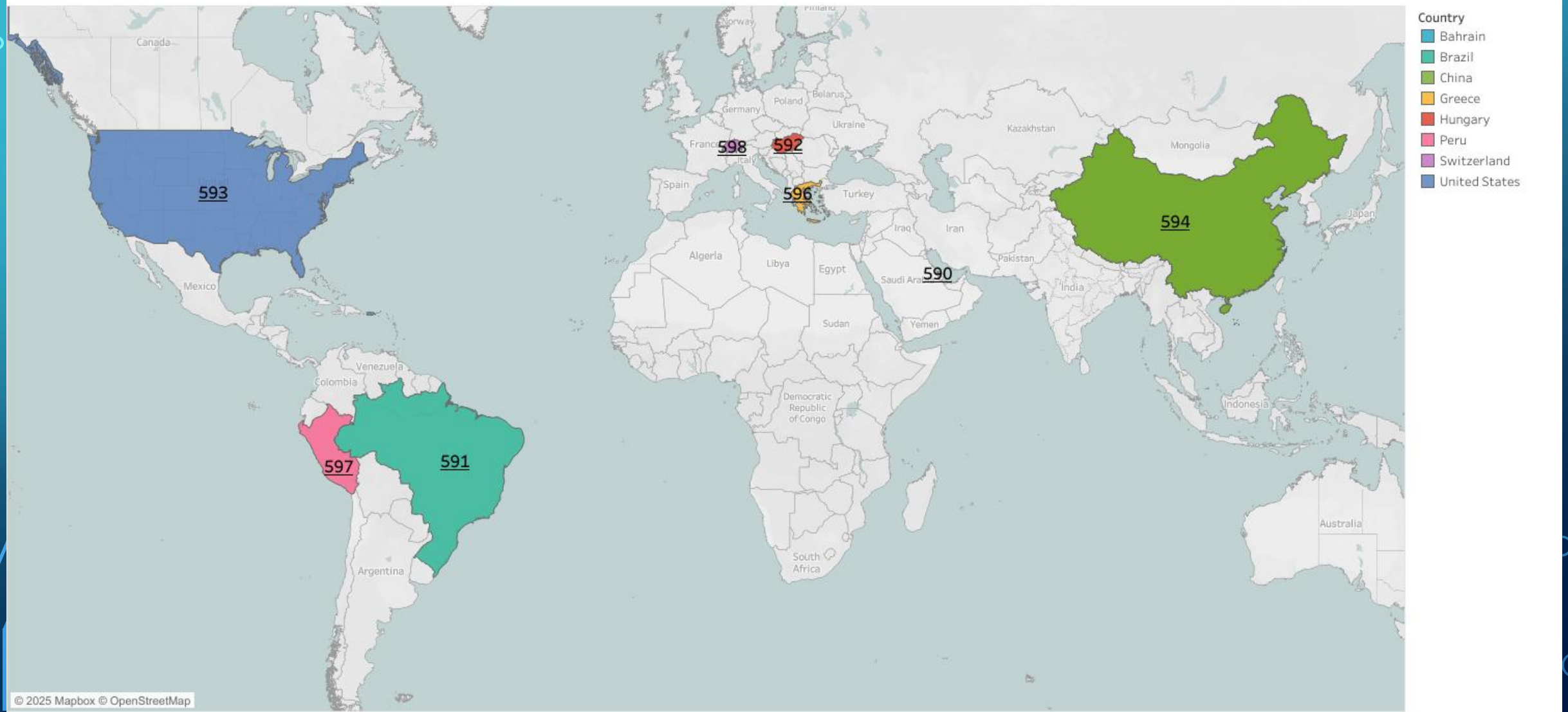


Table of Contents

ROCKBUSTER

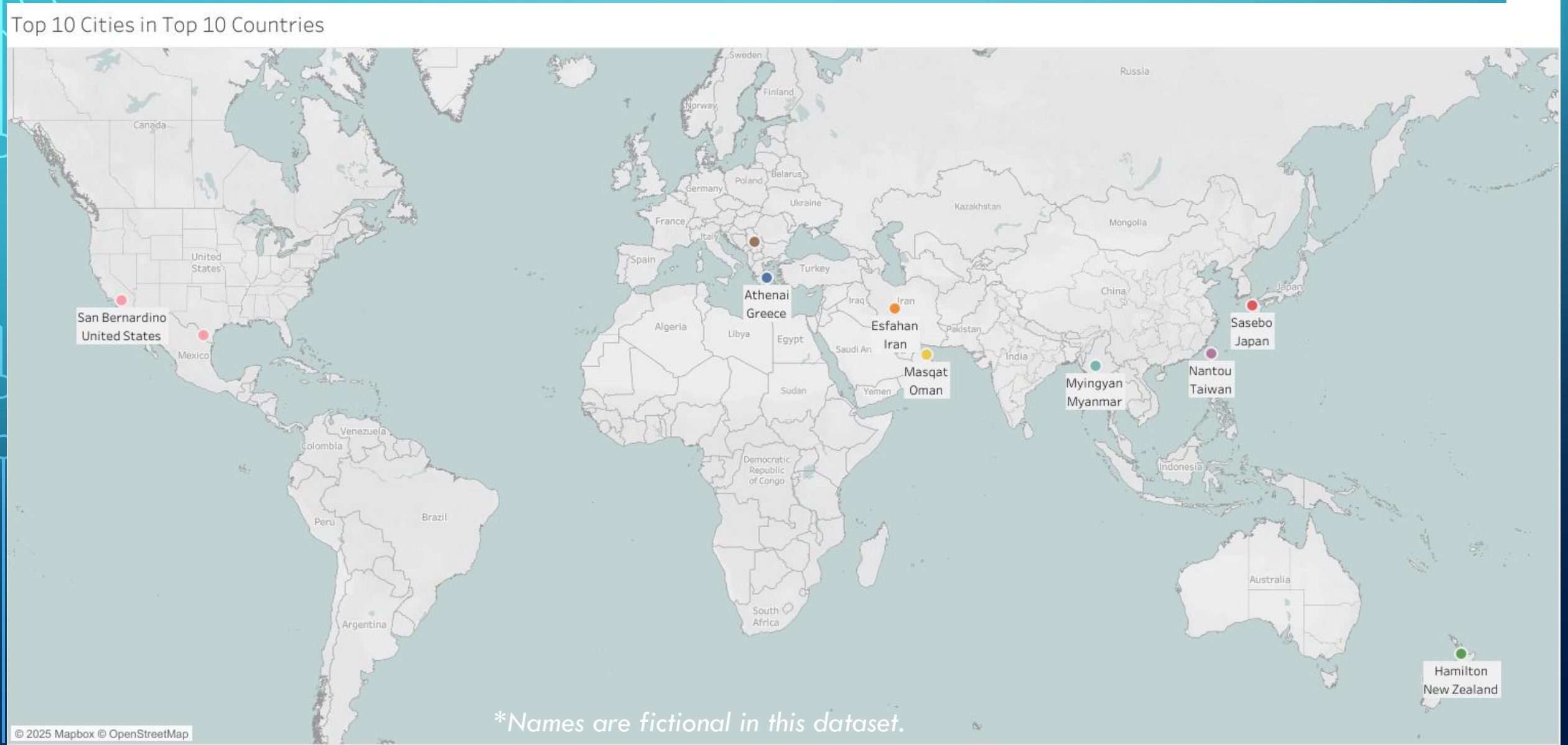
- Using features like Tableau, I can visualize the top 10 countries that Rockbuster's clientele live.



ROCKBUSTER

Table of Contents

- Within the top 10 countries the top 10 cities were pin-pointed. Rockbuster's highest sales were further identified



**Names are fictional in this dataset.*

ROCKBUSTER CONCLUSION

- Challenges presented within this project:
  - Inputting additional movie genre's i.e.: Thriller, Crime, Mystery, Romance, War for improved categorizing.

Table of Contents

INSTACART

- Develop promotional strategies from sales data to produce relevant ad campaigns for top products.
- Discover Instacart's busiest times of day and busiest days of the week.
- Explore popular product sales.

Image link

Table of Contents

INSTACART

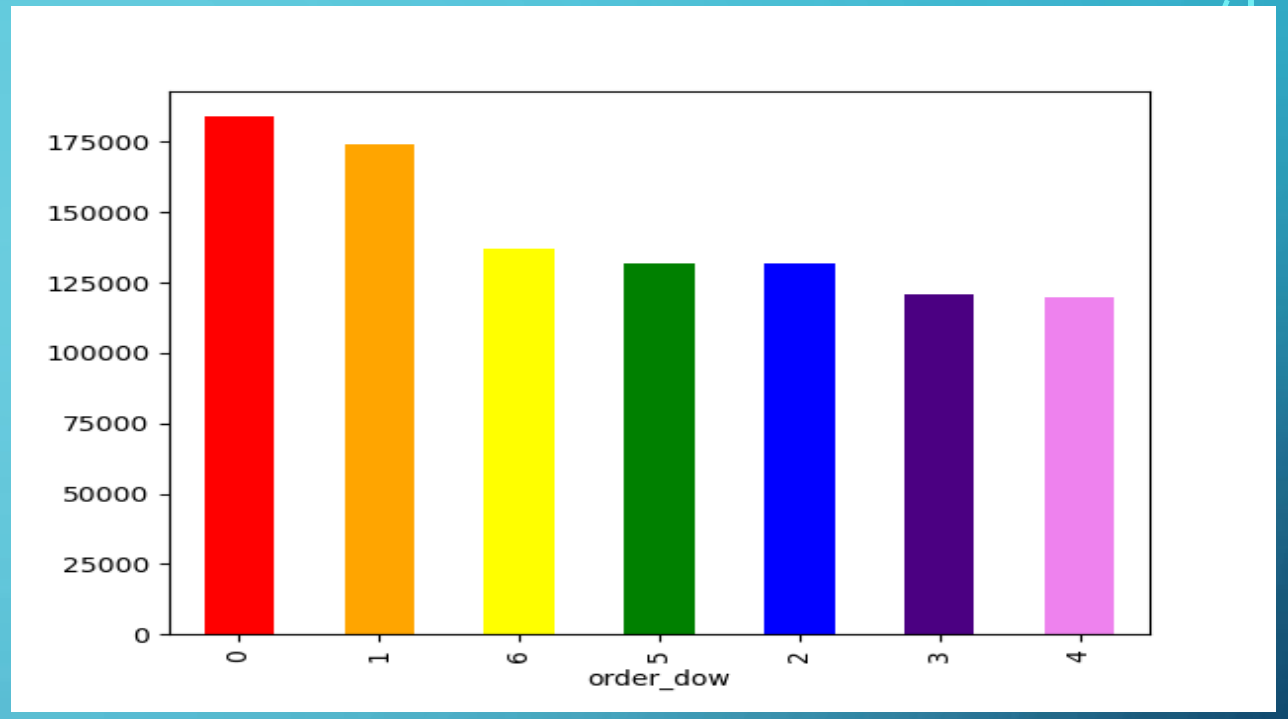
- Answering stakeholder sales questions lead to the discovery that Instacart's busiest days of the week are Saturday and Sunday.
- The histogram shows peak hours are 7am – 6pm. (hours are in a 24-hour period)

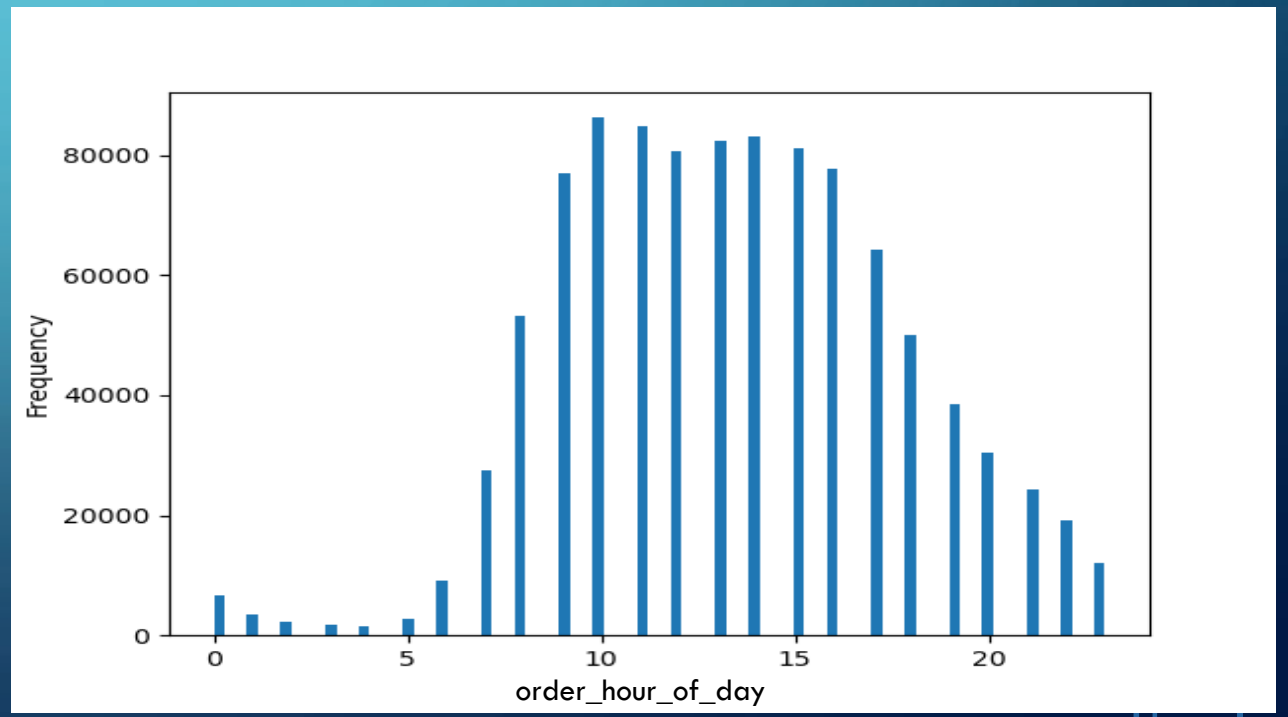
Note on Instacart "orders_dow" Variable

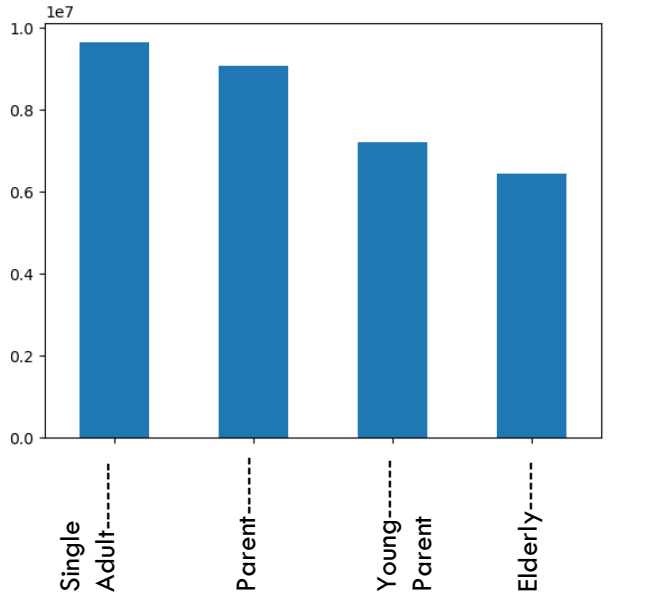
One of the variables in the data is "orders_dow", with "dow" meaning "days of the week". Each day corresponds to a number, as follows:

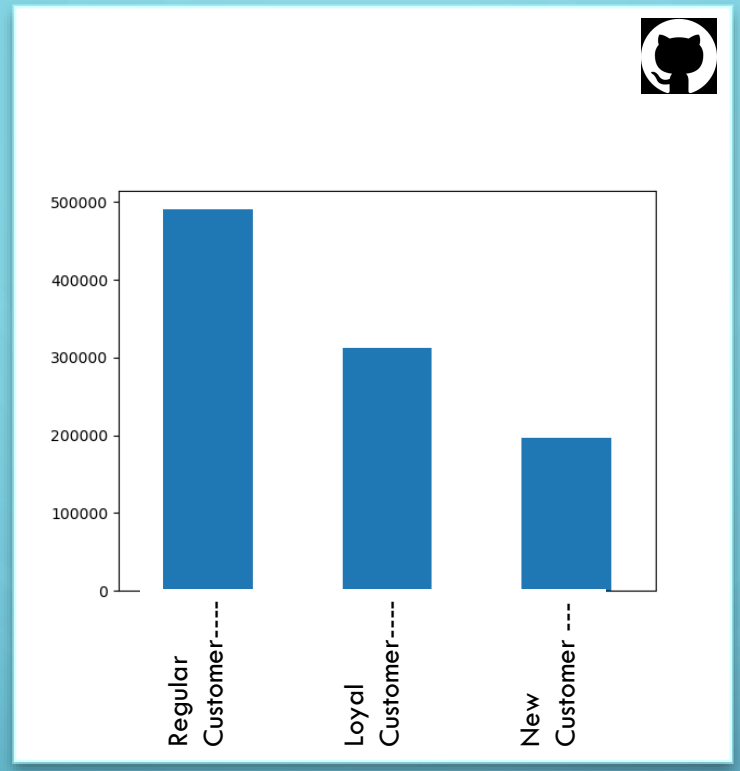
- 0 = Saturday
- 1 = Sunday
- 2 = Monday
- 3 = Tuesday
- 4 = Wednesday
- 5 = Thursday
- 6 = Friday

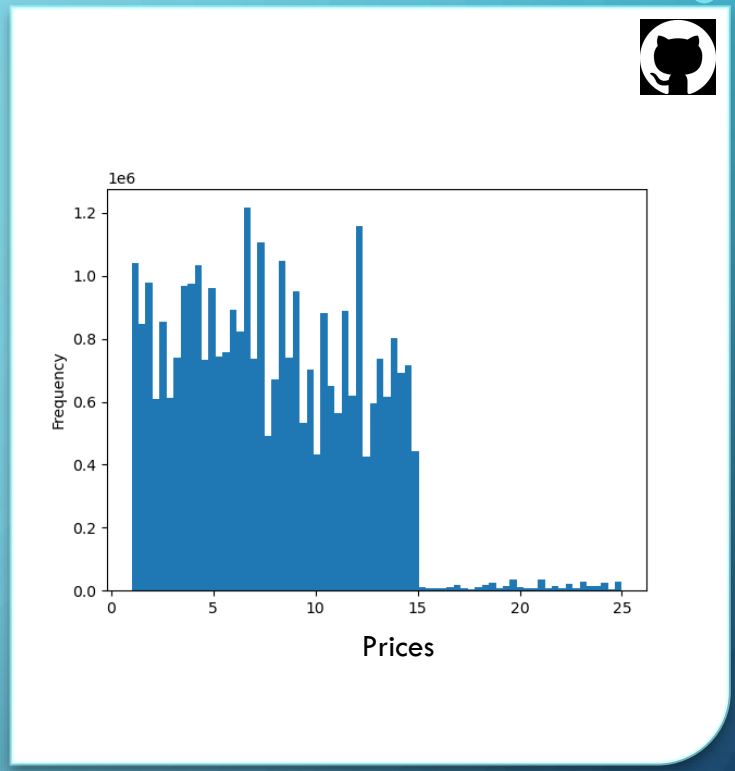
Table of Contents











INSTACART

- (LEFT) Instacart was interested in their clientele based on their spendings.
- (MIDDLE) This bar graph represents the type of clientele that Instacart serves based on the customer loyalty flags built from the number of orders placed.
- (RIGHT) The histograms shows price points. Item frequency shows the number of items within that price range.

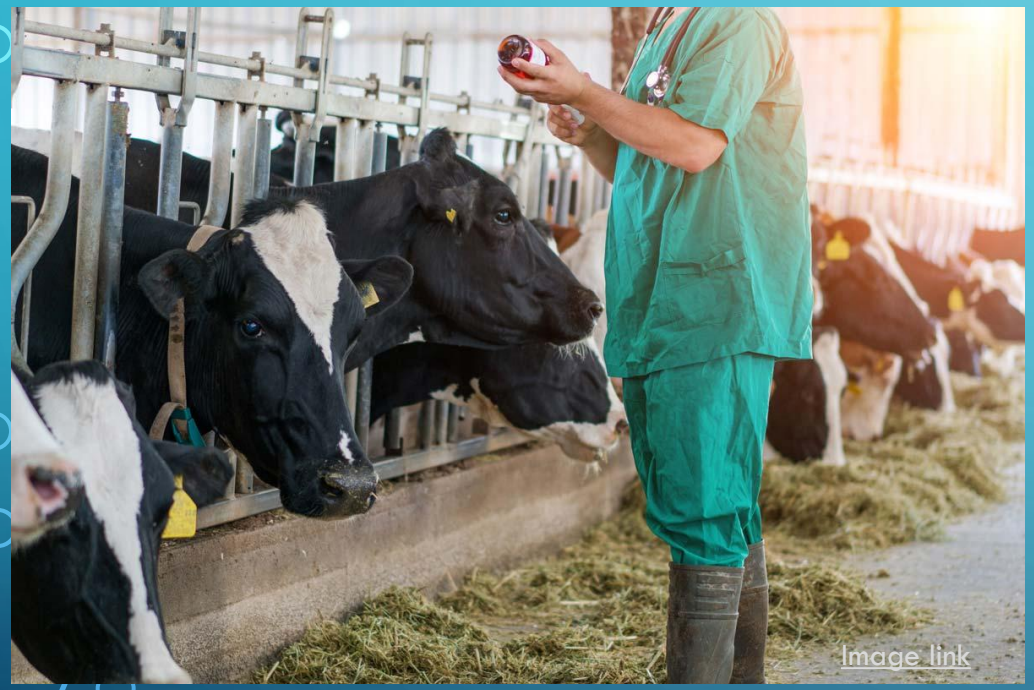
Table of Contents

INSTACART CONCLUSION

- We can infer that during the week more advertisements can be conducted Monday-Friday outside of peak sales hours.
- A few challenges within this data came from having to combine two data sets to achieve the insight.
- Transposing the products data set as it was wide instead of long, this allowed it to be read more easily.
- Changing column titles for easier understanding, resolving mixed data types, removing missing data that couldn't be imputed and removing duplicate information.
- Writing If-Then Statements and Loc() codes to gain further insight into customer loyalty and price points of products.

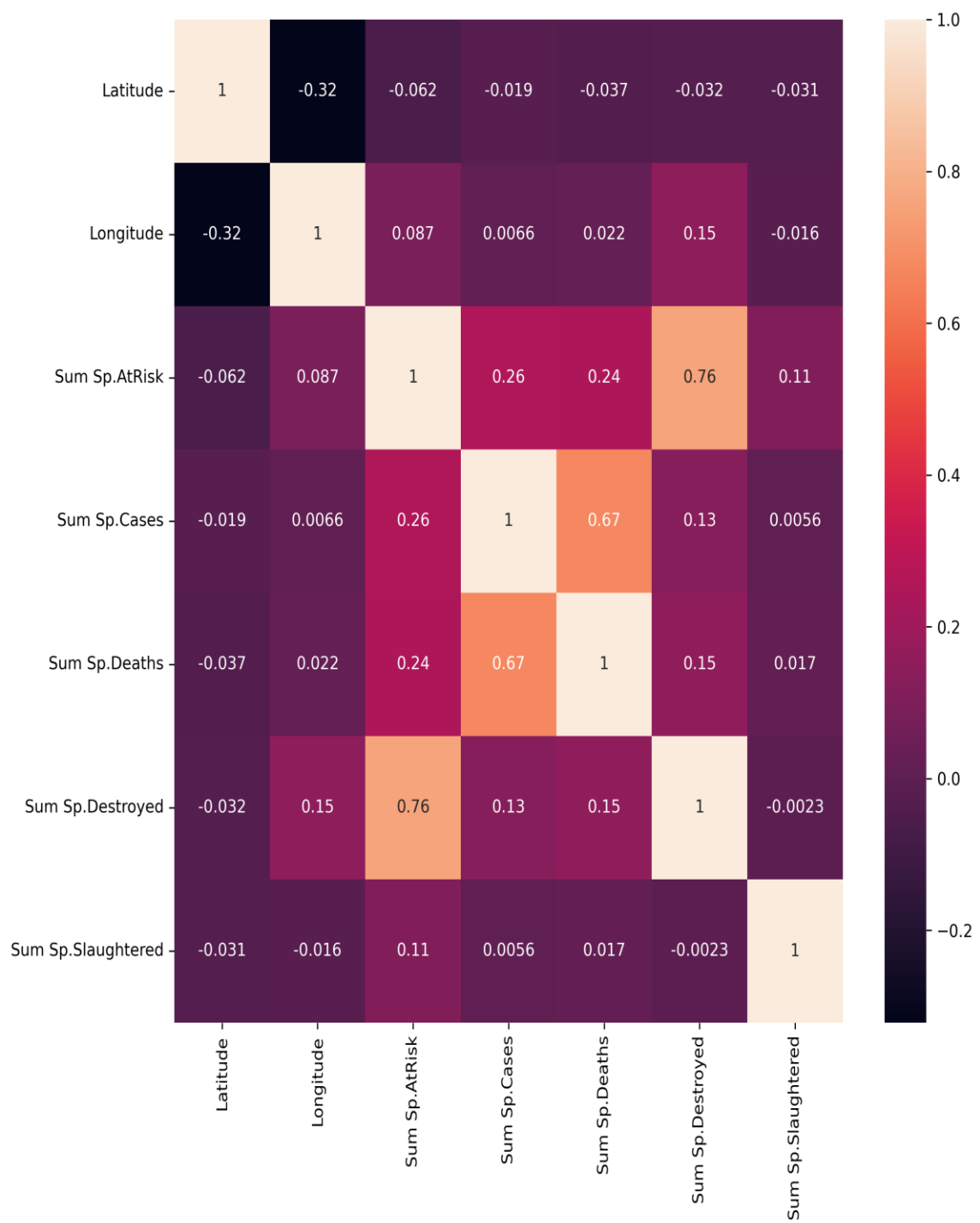
Table of Contents

GLOBAL ANIMAL DISEASE SURVEILLANCE

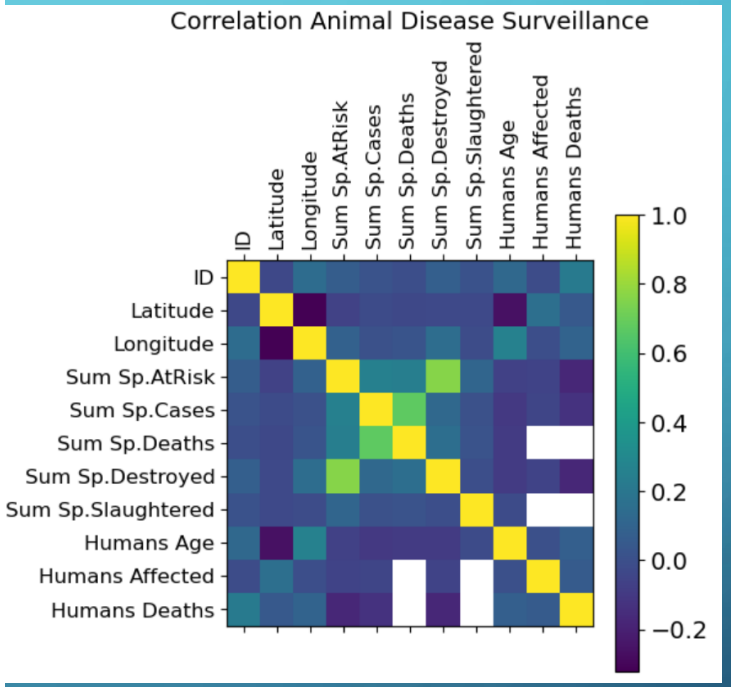


- Build an interactive dashboard visually showcasing well-curated results of advanced exploratory analysis conducted in Python.
- Source data from open-source websites such as Kaggle and generate questions about the data used.
- Using Python explore relationships, create geo-graphic visuals, explore regression, clustering and analyze time series data (time-series data is not related to the main data source).
- Using Tableau, create visuals to answer questions about the diseases, their prevalence in each region and in the animal species.

Table of Contents



GLOBAL ANIMAL DISEASE SURVEILLANCE



- Exploring relationships with correlation heatmaps allowed for better visualization of positive relationships.
- In this data visual values closer to 1.0 have a strong relationship that should be explored.

Table of Contents

GLOBAL ANIMAL DISEASE SURVEILLANCE

- A choropleth map was created using a JSON file. The JSON file was created using geojson-maps.kyd.au and was incorporated into the original data.
- To interact with the map, you can check it out here.

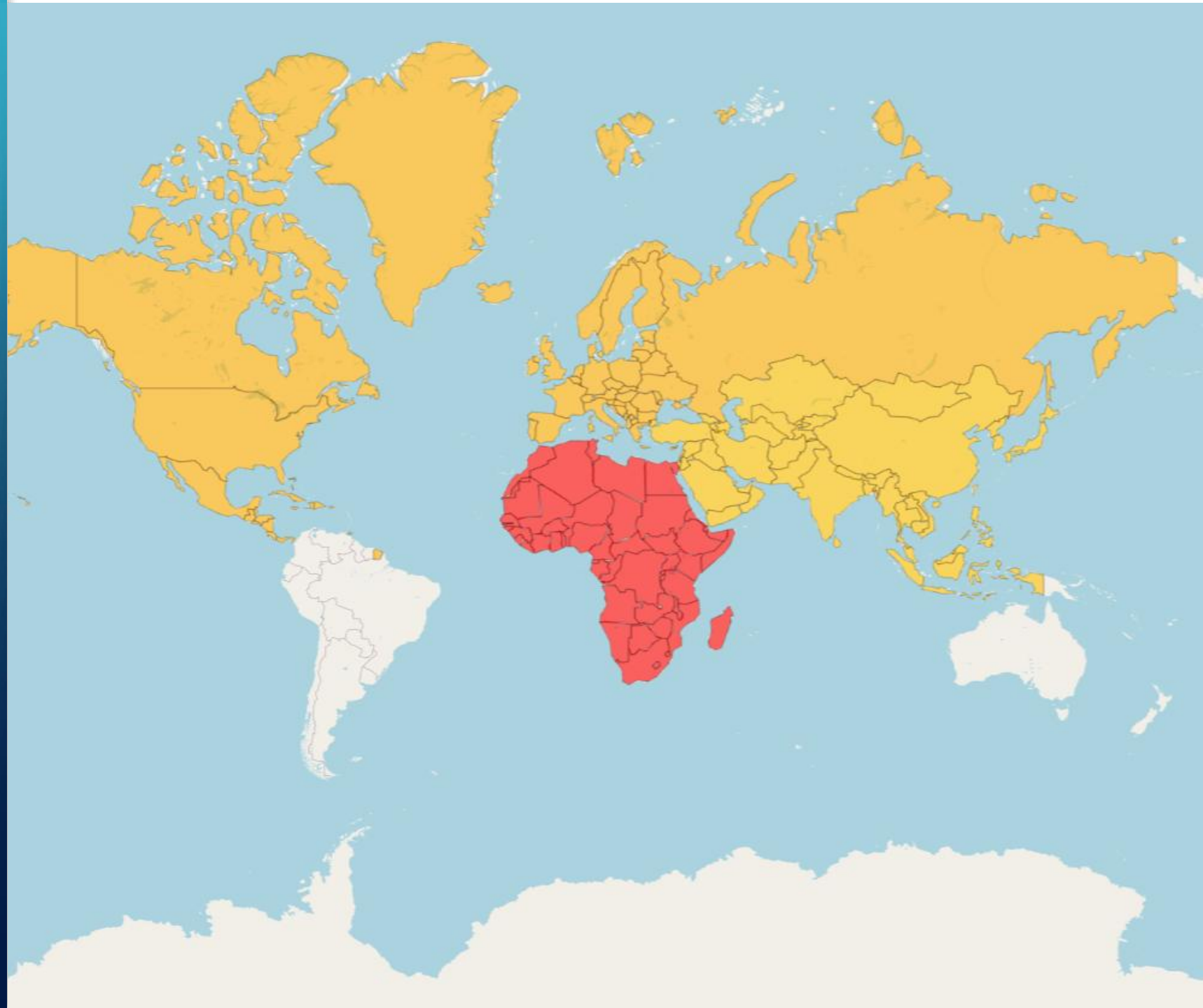
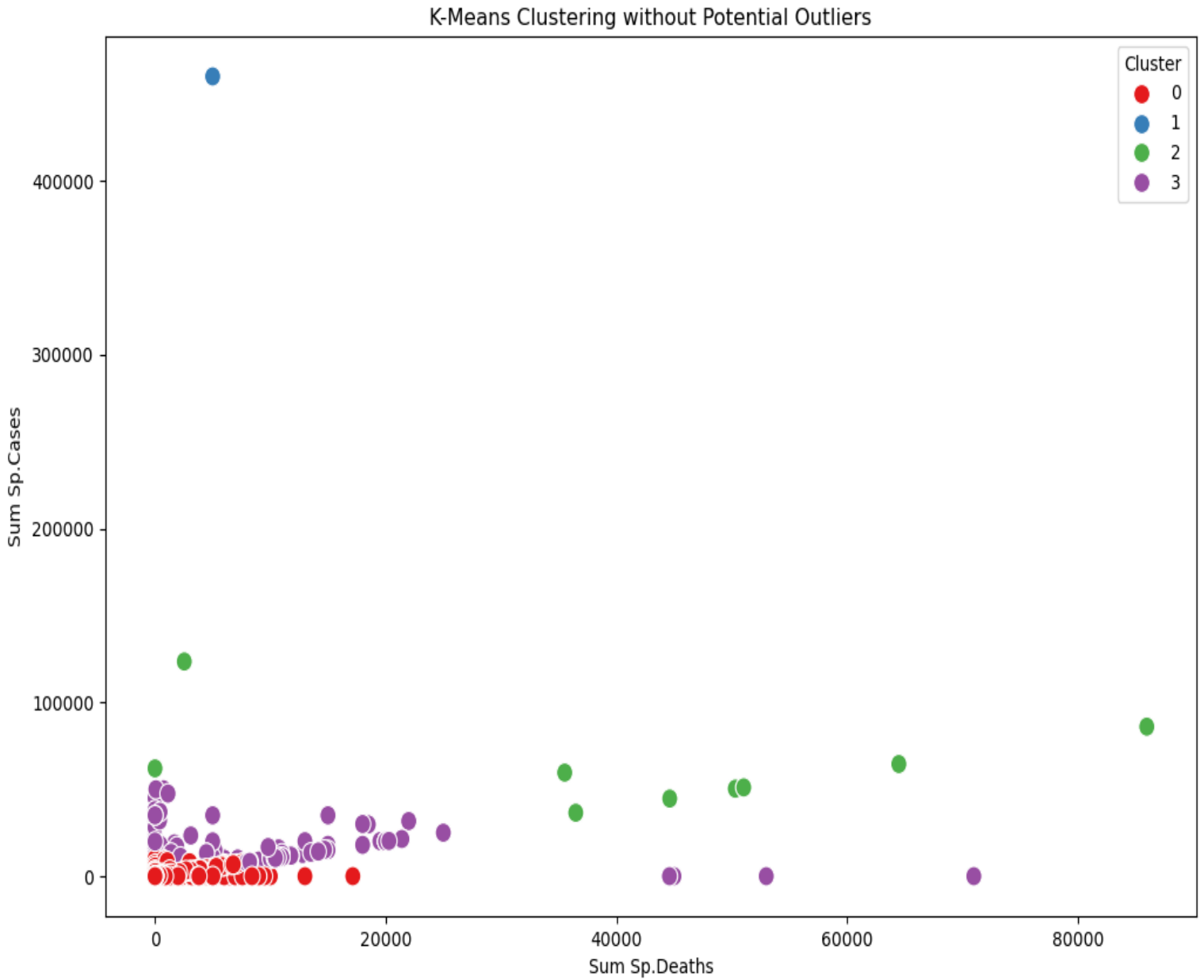


Table of Contents



GLOBAL ANIMAL DISEASE SURVEILLANCE

- KMeans clusters created in python applies an algorithm to the dataset.
- The data is grouped into distinct clusters by using a centroid and stabilizing the clusters.
- For example, the 0 cluster represents lower cases reported and lower deaths associated with those cases.

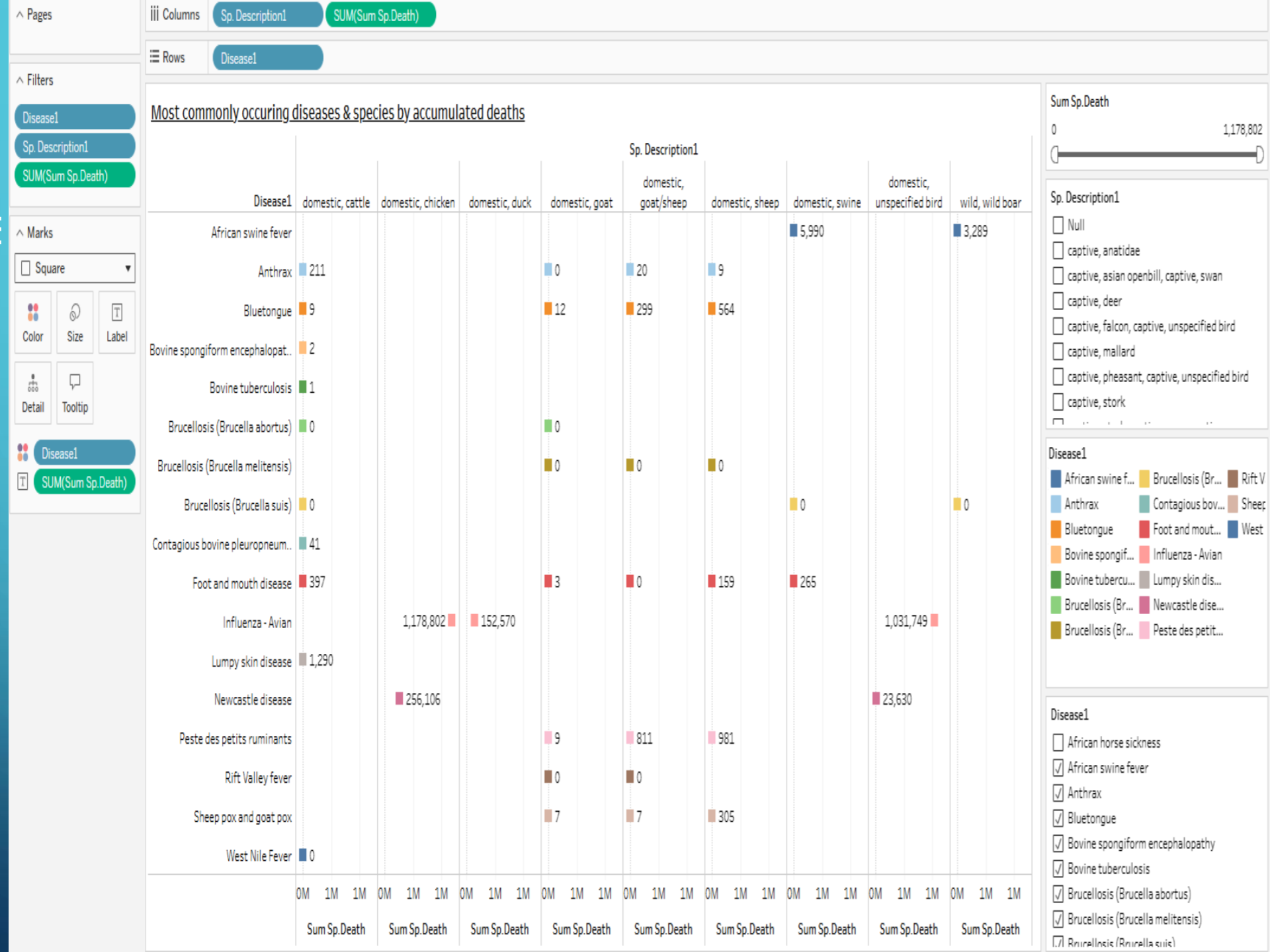
Table of Contents

GLOBAL ANIMAL DISEASE SURVEILLANCE

- Tableau visuals show various interpretations of disease by region, prevalence and species.
- Most common diseases are identified as well as species that are more at risk.

If you'd like to check out the Tableau storyboard, click the icon

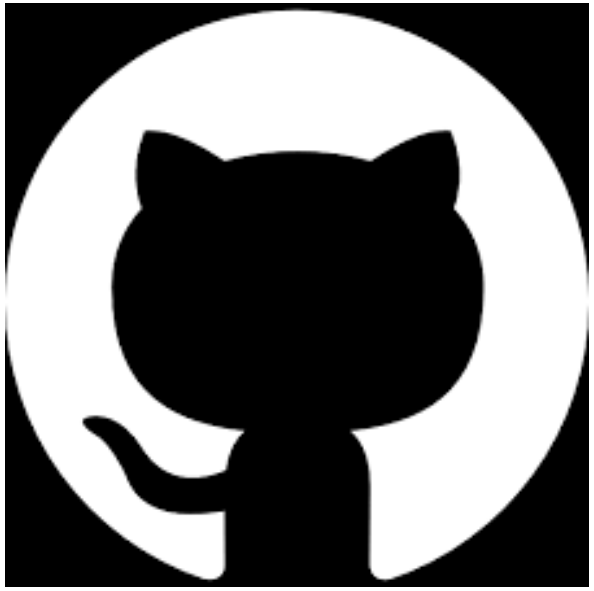
Table of Contents



GLOBAL ANIMAL DISEASE SURVEILLANCE CONCLUSION

- Challenges within the data come from the data itself and the JSON file.
- The data has loose interpretation of the species columns and there are small amounts of human data present. This could have been excluded but it was kept to get a glimpse of what ailments affect humans.
- The species description (Sp. Description) column in the data set had multiple combinations of species that presented diseases. This made it difficult to pin-point diseases by each individual species.
- Additionally, an interesting challenge was finding the correct key in the JSON file as there were about 170 columns and several keys accompanying them. To sync the JSON file with my data frame, I had to rename the Americas to North America under the Region columns. Throughout the project North America also includes South America.
- Conclusions were easily drawn from the Tableau visuals to understand the disease prevalence in each region, by disease and by species.

Table of Contents



Project and examples can be found within GitHub and Tableau.

CONTACT INFO

JORDAN NOVELLI – JORDANMNOVELLI@GMAIL.COM

PORTFOLIO WEBSITE

Table of Contents