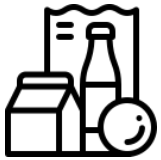
Data Analytics Portfolio

Jesse Holmes

Project List

Analyzing Global Video Game Sales



Marketing Strategy for Online Grocery Store

Flu Season Preparation

Analysis for Global Bank

Analysis for Online Video Rental Business

Gun Violence Data

Gameco Analysis

Analyzing Global Video Game Sales

Key Questions:

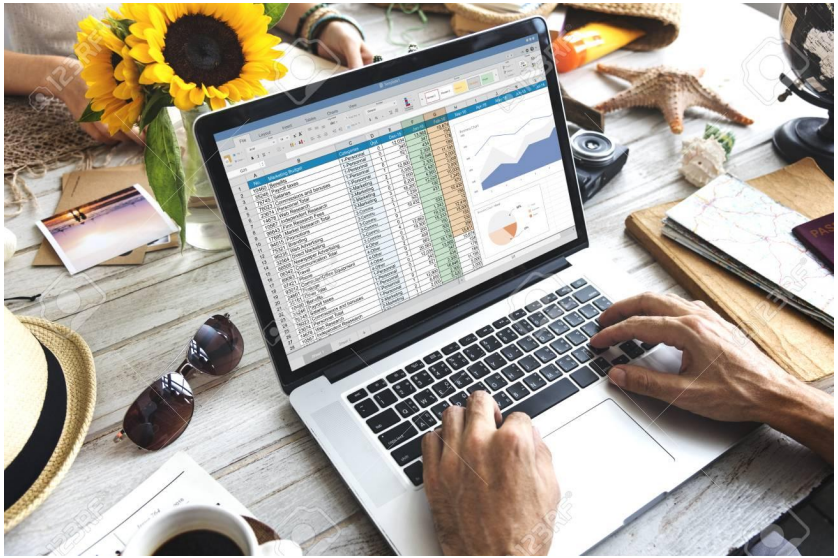
1. Are certain types of games more popular than others?
2. What other publishers will likely be the main competitors in certain markets?
3. Have any games decreased or increased in popularity over time?
4. How have their sales figures varied between geographic regions over time?

Gameco Analysis

Analyzing Global Video Game Sales

Skills & Tools:

- Excel
- Grouping & Summarizing Data
- Descriptive Analysis
- Creating Visuals in Excel
- Presenting Results

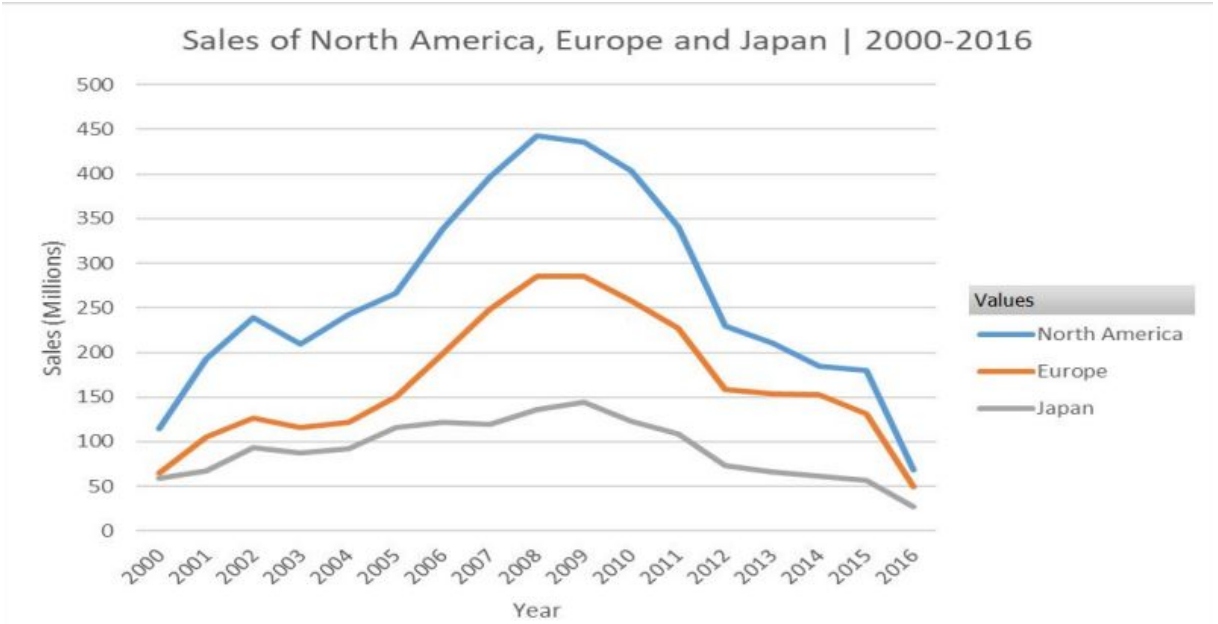


Dataset: VGChartz website - Download

Analysis

The data analysis focused on observing trends and patterns of video game sales across different markets over time, with an additional emphasis on specific genres and games.

The goal was to develop insights into the market's evolution to inform future decisions.



Our top-selling genres remain Action, Sports, Shooter, and Role-Playing games, with strong sales in North America, Europe, and Japan. North American sales declined over the past decade, while Europe experienced steady growth, and Japan remained stable.

Recommendations

1. GameCo's sales strategy should involve diversifying their game portfolio while maintaining a strong focus on their top-selling genres, such as Action, Sports, Shooter, and Role-Playing games.

2. They should prioritize expanding their presence and marketing efforts in Europe due to its steady sales growth, address the decline in North American sales through revamped strategies, and continue to maintain stability in the Japanese market.

3. Additionally, staying agile with market research and community engagement will be crucial for adapting to evolving gaming trends and ensuring long-term success.

Flu Season Preparation

Key Objectives:

1. Providing information to support a staffing plan for the distribution of medical assistance throughout the United States

2. Determine if there are trends or patterns related to influenza throughout the calendar year

3. Categorize each State as low, medium, or high, based on it’s vulnerable population count & needs

Flu Season Preparation

Skills & Tools:

- Excel
- Determining Business Requirements
- Data Cleaning, Integration & Transformation
- Statistical Hypothesis Testing
- Visual Analysis
- Forecasting
- Creating Visuals in Tableau
- Presenting Results

Datasets: Center for Disease Control and Prevention (CDC); Download links 1; 2; 3 and 4
US Census Bureau; Download

Analysis

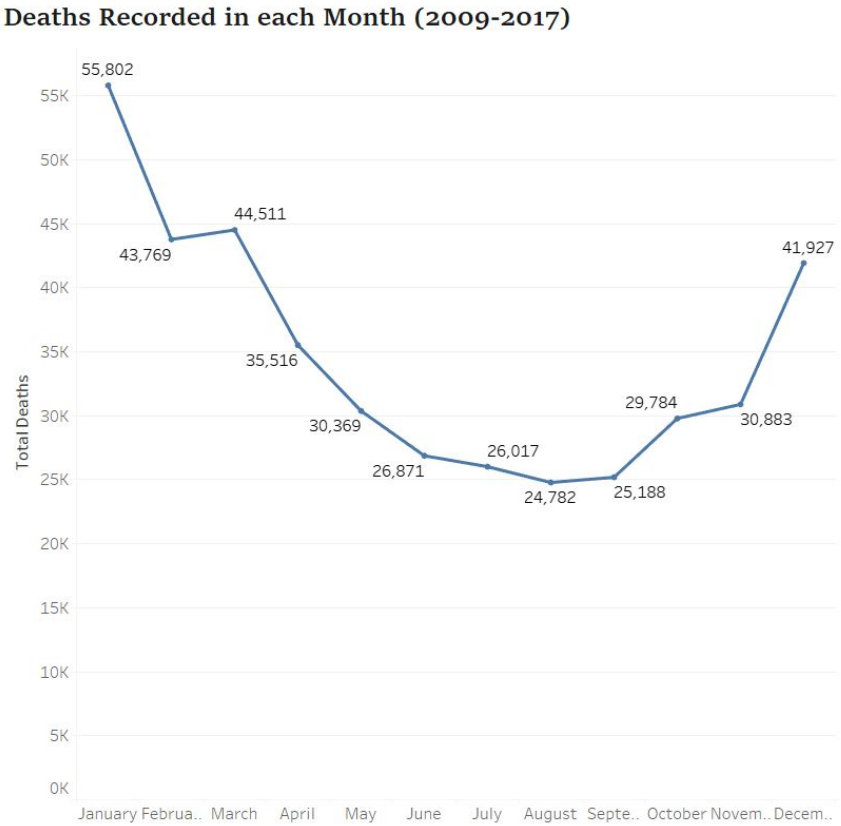
Our motivation is to protect our vulnerable populations, as they are more likely to experience complications related to influenza sickness.

In order to make this a successful initiative, we will explore a support-staffing plan based on the findings presented today.

Our findings included:

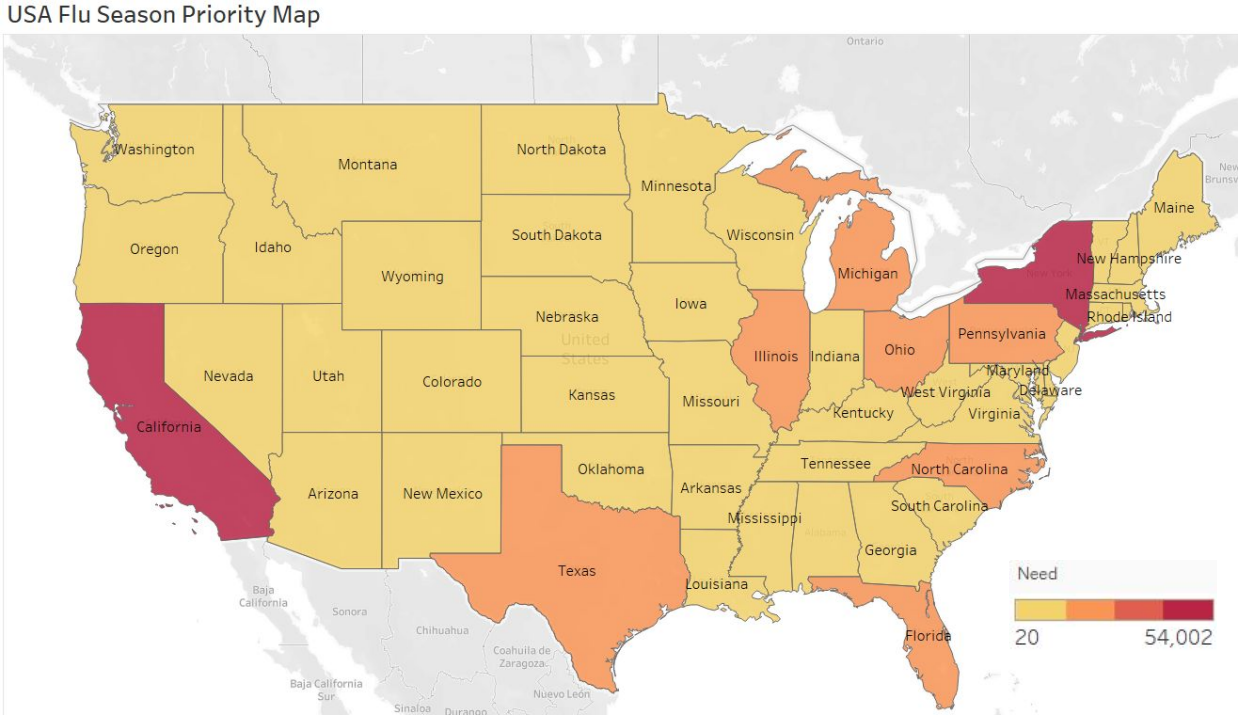
- It can be assumed that there is a clear geographical demand in the Northeastern region of the United States.

- Expanding the designated "Flu Season" to include March is advisable



Recommendations

1. Priority be given to our Most Populated States
2. Priority be given to our Population aged 65 years of age and higher
3. Geographical priority should got to our States in the North Eastern region of the country



Rockbuster Stealth Analysis

Analysis for Online Video Rental Business

Key Questions:

1. Which countries are Rockbuster customers based in?
2. Where are customer with a high lifetime value based?
3. Which movies contributed to the highest & lowest revenue?
4. What was the average rental duration?
5. Does revenue vary between geographical regions?

Rockbuster Stealth Analysis

Analysis for Online Video Rental Business

Skills & Tools:

- Relational Databases
- SQL
- Database Querying
- Filtering, Cleaning and Summarizing
- Joining Tables
- Subqueries
- CTE

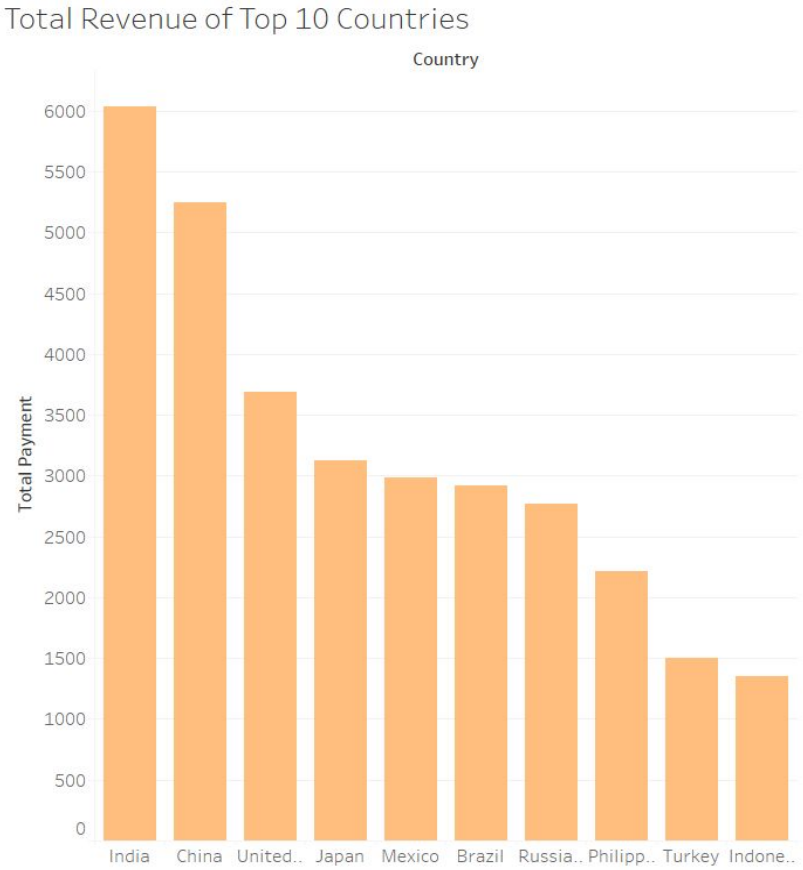
Dataset: PostgreSQLTutorial - Download

Analysis

In order to compete with streaming services like Netflix, and Amazon Prime, Rockbuster Stealth is in the process of launching an online video rental service in order to stay competitive.

Our findings included:

- While sizable segments of our customer base hail from North and South America, our efforts to establish a presence in the European market are progressing more slowly in comparison to other regions.
- Our primary markets revolve around India and China, serving as the key regions for our business. Almost 50% of our customer base comes from the Asia-Pacific region.



Recommendations

1. Intensify marketing initiatives in the top 10 countries/territories with substantial customer counts and high lifetime values.

2. Implement a standardized movie rental duration of 5 days to enhance the efficiency of the online rental business.

3. Expand the movie database as the business grows, prioritizing the inclusion of additional films in top-earning genres and those with high movie ratings.

Additional:

- Presentation Link - Here
- Tableau Link - Here
- Github repository - Here

Instacart Analysis

Marketing Strategy for Online Grocery Store

Key Questions:

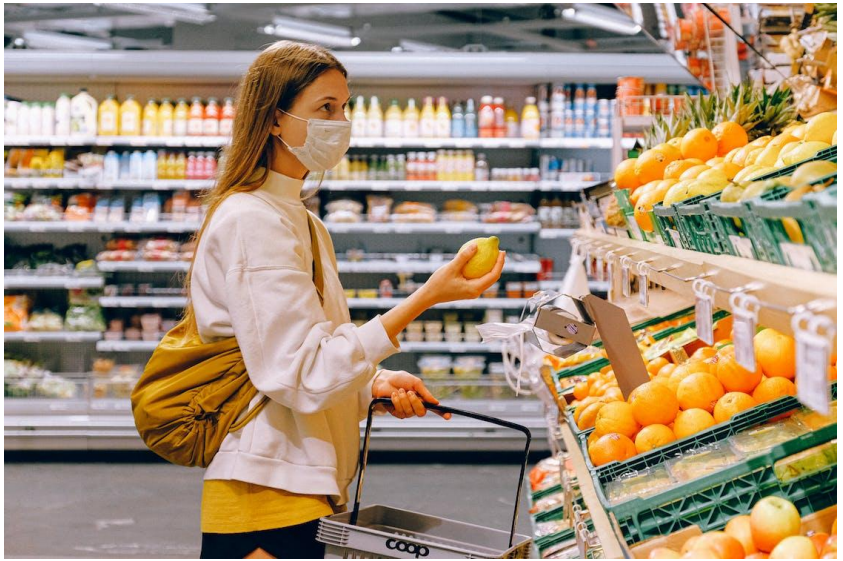
1. What are the busiest days, and hours, of the week?
2. Are there particular times of the day when people spend the most money?
3. How can price range groupings be simplified?
4. Are there certain products that are more popular than others?
5. What types of customers can be identified, and what are their ordering traits?

Instacart Analysis

Marketing Strategy for Online Grocery Store

Skills & Tools:

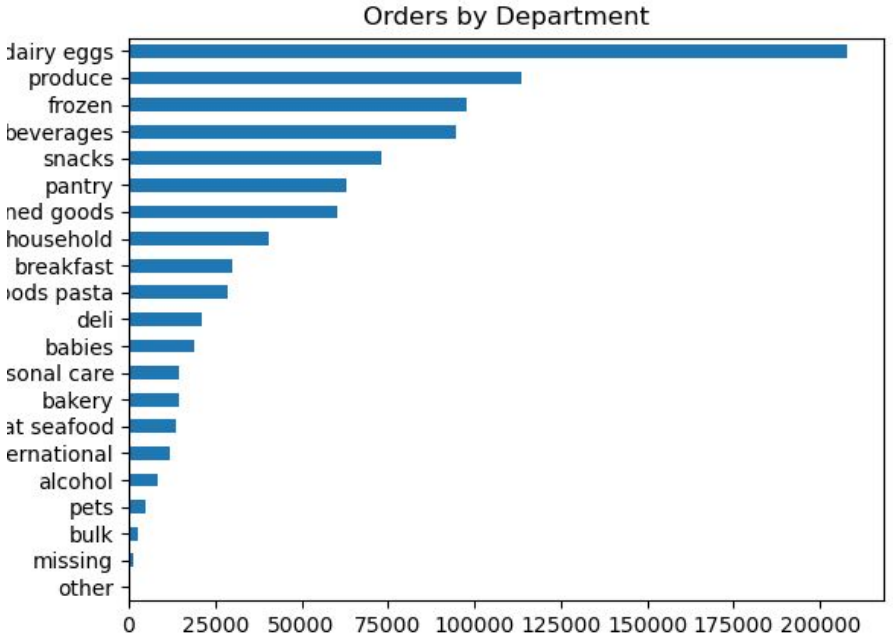
- Python
- Data Wrangling & Merging
- Deriving Variables
- Grouping & Aggregating Data
- Excel Reporting



Dataset: Instacart Online Grocery Shopping Dataset 2017 - Download
Careerfoundry - Download

Analysis

- Monday and Tuesday emerged as the busiest days for order placements
- We see a significant spike in spending from midnight to 2am
- The top performing departments are predominantly in the consumable goods category, including food and beverages
- Instacart leans significantly on its ‘Regular’ and ‘Loyal customer base
- It appears there are no discernible differences in ordering habits based on a customer’s loyalty status
- The demographic information suggests that the driving force behind sales on Instacart is the older, more established, and higher-earning customer segment.



Recommendations

1. Instacart should concentrate its promotions on Saturday and Sunday, while finding ways to boost demand on Wednesdays and Tuesdays.

2. In the popular departments of produce and dairy/eggs, the focus of advertising should be on promoting existing items or expanding the variety of items within the department.

3. Instacart should make efforts to attract new customers, such as creating special offers for their first orders.

4. Young adults with a high income, especially those with children, appear to be underrepresented in the customer base, despite generating the largest relative number of orders and revenue per customer. For this reason, Instacart should focus on retaining and attracting this demographic

Pig E. Bank Analysis

Analysis for Global Bank

Key Objectives:

1. Identifying potential data bias and ethical concerns on how the data is used, shared, collected and protected

2. Using common data mining algorithms to identify leading factor to client loss

3. Understanding predictive analysis and models and how to apply them

4. Acquiring knowledge on time series analysis and forecasting and their potential role in identifying money laundering fraud in the bank

Pig E. Bank Analysis

Analysis for Global Bank

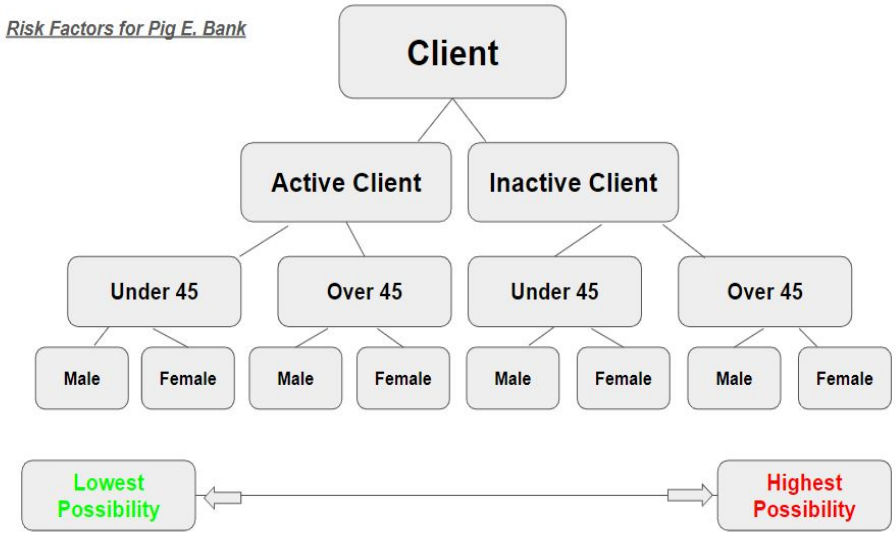
Skills & Tools:

- Using Big Data
- Data Ethics & Mining
- Predictive Analysis
- Time Series Analysis and Forecasting

Dataset: CareerFoundry - Download

Analysis

- Identify and address potential data bias in Pig E. Bank's money laundering flagging model, including biases held by analysts, and suggest ways to communicate ethical concerns and establish a clear data security protocol.



- Use exploratory analysis to identify main variables influencing client loss, such as active membership, gender, age, and country, and develop a decision tree to predict customer churn

- Analyze the outputs of linear and logistic regression models, as well as time series and forecasting models, to understand their characteristics, applications, and requirements

Recommendations

1. Use the CRISP-DM methodology to guide the data mining process and prioritize data ethics principles, such as privacy, transparency, accountability, and equality.

2. Select forecasting models based on data characteristics, validate them for accuracy and reliability, and communicate ethical concerns from a problem-solving and preventive perspective.

3. Take steps to avoid measurement biases in predictive models, including targeted training, understanding of variables and potential biases, and constant monitoring to proactively identify emerging biases

Gun Violence Data

Key Objectives:

1. To build an interactive dashboard visually showcasing well-curated results of an advanced exploratory analysis conducted in Python

2. Identifying connections between variables worth further exploration

3. Use various advanced analytical approaches to help test the data

4. Present results using Tableau Dashboard

Gun Violence Data

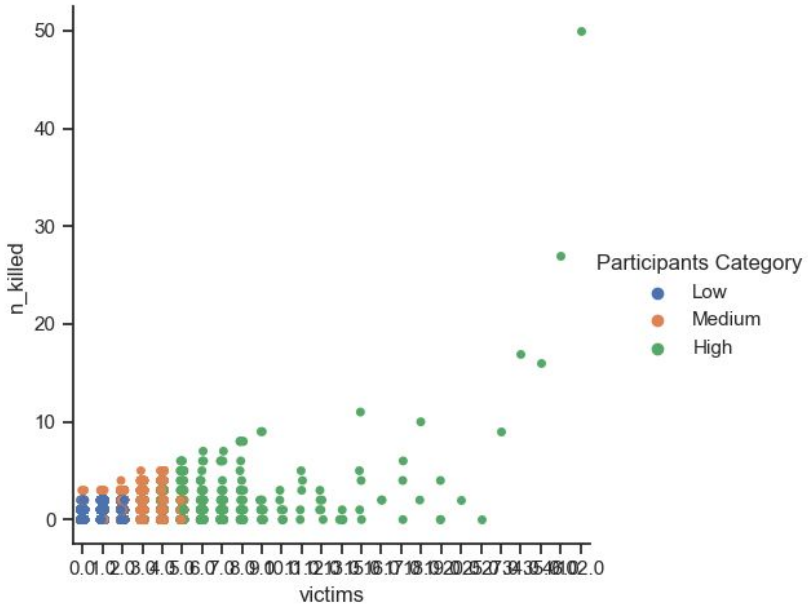
Skills & Tools:

- Data sourcing & transformation
- Exploratory analysis
- Geospatial analysis with Python
- Time-Series and Machine Learning Analysis

Dataset: The Gun Violence Archive (GVA) - Download

Gun Violence Data

- Sourcing, cleaning, and analysis of the data for this project was extensive. We had many potential ethics issues we needed to remove from the dataset, as well as cleaning up empty rows, and mislabeled columns.



- We then looked at how our variables correlated, using a Correlation Matrix

- After we had our hypotheses, we continued by using linear regression, Cluster Analysis, and attempted to use Time-Series Analysis

Recommendations & Limitations

1. The dataset provided daily accounts of each incident that was logged. This is fantastic to have, but we were unable to use it in our Time-Series Analysis in a meaningful way, as the required codes, and processes, needed totals in a year-over-year rather than daily.

2. Once outliers were removed, the Machine Learning Scatter Plot because difficult to produce. We had several thousand (around 18 thousand to be exact) instances, but most fell within a 3-digit range. Perhaps a more expansive plot could have provided a greater visual.

3. Further analysis, and having a grasp of the required information for our 3 main forms of analysis (Correlation Matrix, Machine Learning, and Time-Series Analysis) prior to the investment of time and cleaning of data would have been beneficial.