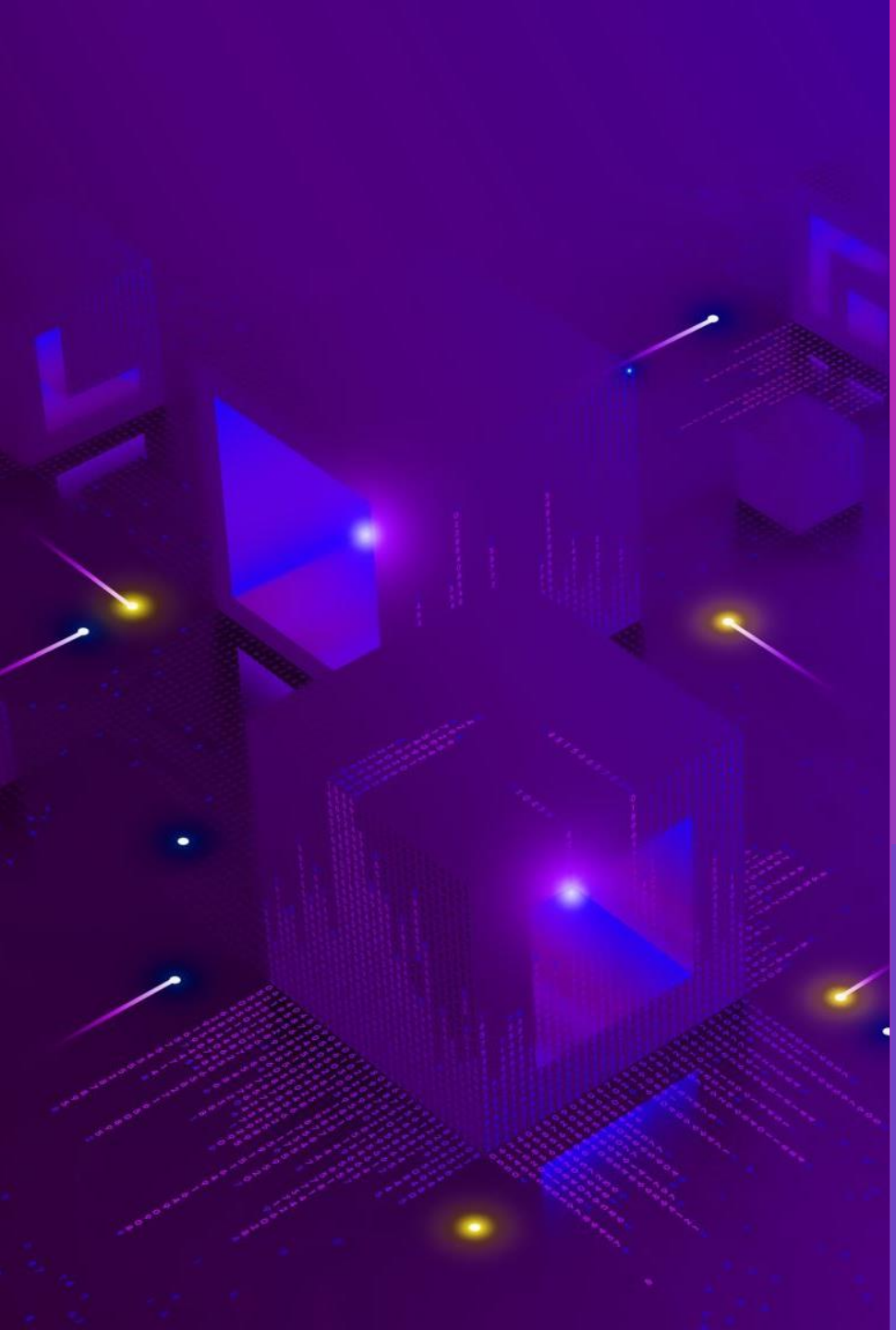


DATA ANALYTICS PORTFOLIO

NUPUR PARIKH

ABOUT ME

Welcome to my portfolio, thank you for stopping by!

I am a Data analyst with two years of experience in the behavioral health industry and seven years of experience in the customer service industry, skilled in observing and pulling associations between different data sets, recognizing trends, and using insights to extract information that delivers resolutions.

With a desire to understand the meaning of every situation I encounter; I seek to uncover critical insights and offer tailored solutions to address the client's needs. I help lead companies to make informed decisions by determining trends and extracting information from datasets.

Analytical Skills

Data Visualization

Data Cleaning

Problem Recognition

Problem-Solving

Querying

Statistical Programming

Tools

Excel

Power Point

Tableau

SQL

Python

Soft Skills

Collaborative
Communication
Analytic
Curious
Adaptability

FIVE MAIN PROJECTS

- GameCo Analysis
- Preparing for Influenza Season
- Rockbuster Stealth
- Instacart Basket Analysis
- Gun Violence Analysis

GAMECO ANALYSIS

PROJECT OVERVIEW

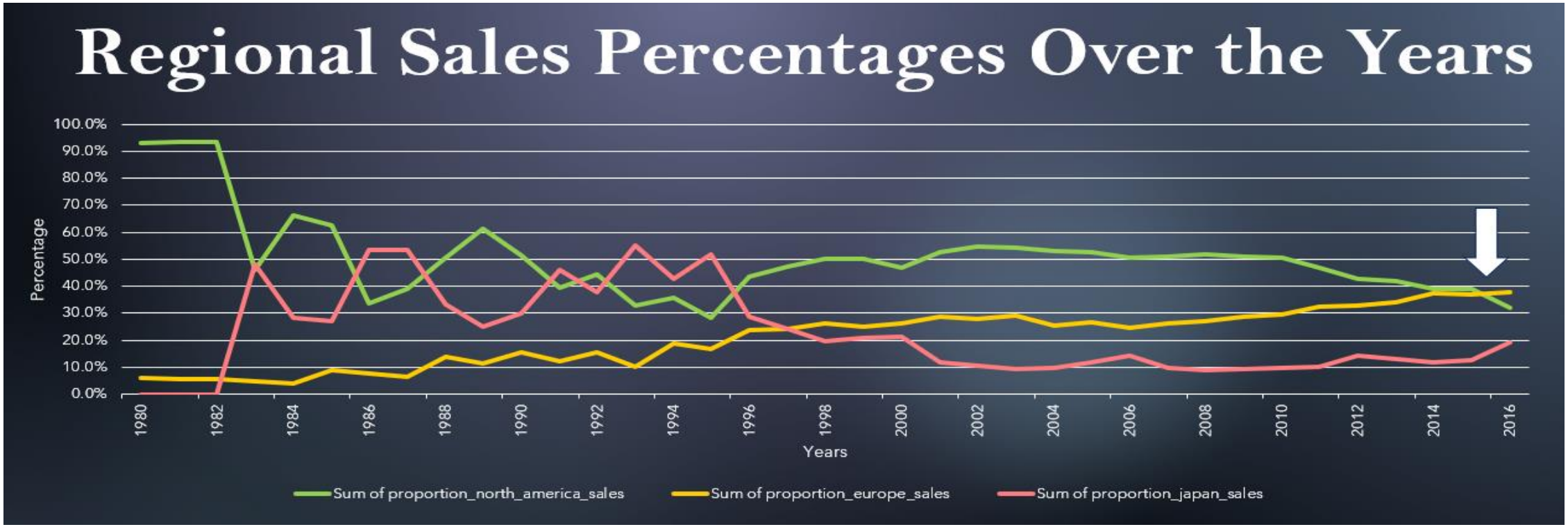
- **Objectives**
  - Perform a descriptive analysis of a video game sales data set to foster a better understanding of how GameCo's (fictional company) new games might fare in the market.
  - Compare assumptions about historical regional sales with the reality of current market conditions.
- **Key Questions**
  - How have their sales figures varied between geographic regions over time?
  - What are the most popular genres?
  - What are the most popular gaming platforms?

Skills
Grouping Data
Descriptive Analysis
Summarizing Data
Visualizations

Tools used


Dataset: VGChartz website; Download link here; methodology here.
Presentation: full powerpoint presentation here.
Reflections here.

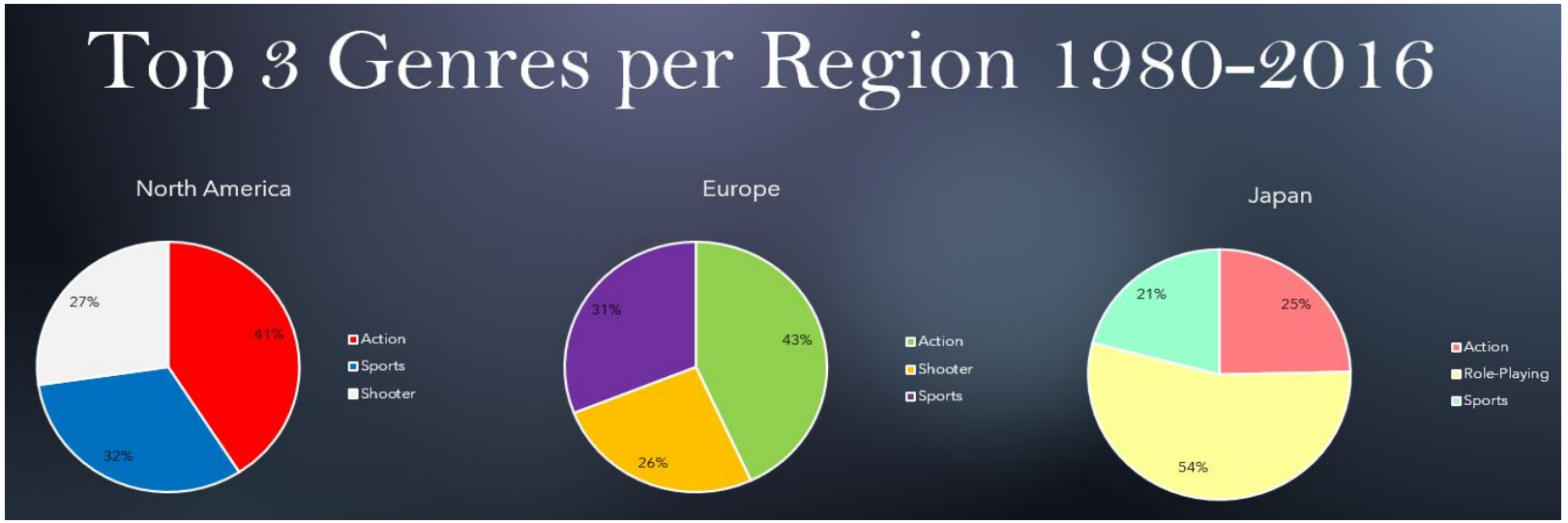
HOW HAVE THEIR SALES FIGURES VARIED BETWEEN GEOGRAPHIC REGIONS OVER TIME?



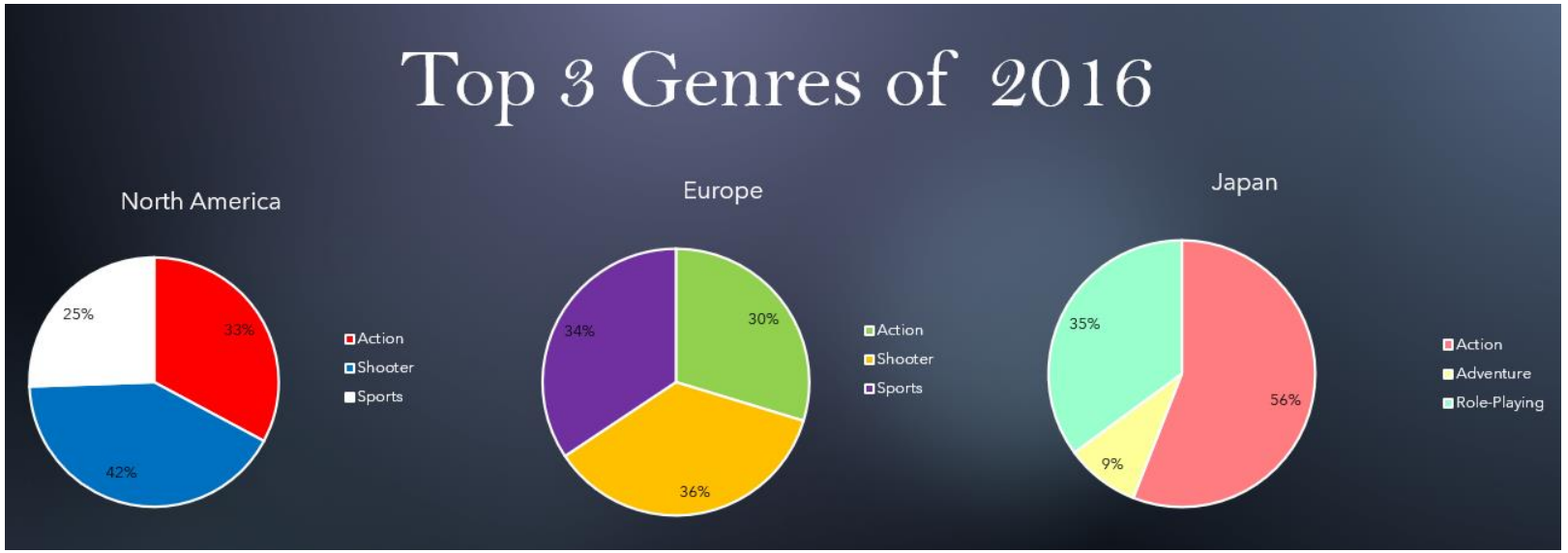
- **1980-1983**: North America held the highest number of sales.
- **1983-1996**: North America and Japan are constantly fluctuating between who has the greatest number of sales.
- **1997**: Europe's sales surpass Japan
- **1996-2015**: North America holds the largest number of global sales until 2015, despite sales decreasing over time. However, as the arrow shows, in 2015 Europe surpassed North America's percentage of sales
- **Key Takeaway**: Japan's sales have remained volatile throughout the years, while North America's top sales decreased by 2015, leaving Europe as the country with the top percentage of sales by 2016.

WHAT ARE THE MOST POPULAR GENRES

- From 1980 to 2016, North America and Europe have the same top 3 genres of Action, Shooter, and Sports.
- Japan has Action and Sports in the top 3. However, Role-Playing is more popular here and accounts for more than half of their video game sales.

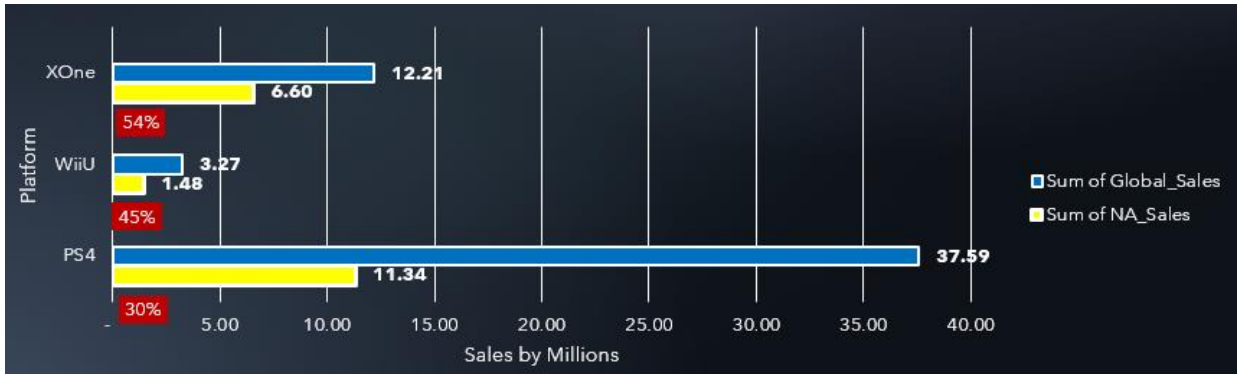


- For 2016, the top 3 genres stayed the same for all regions. However, now Shooter is the most popular genre in North America and Europe.
- For Japan, Action is their most popular game now, accounting for more than 50% of their sales. Role-Playing is still in the top 3, but less sales.



WHAT ARE THE MOST POPULAR GAMING PLATFORMS? (2016)

North America

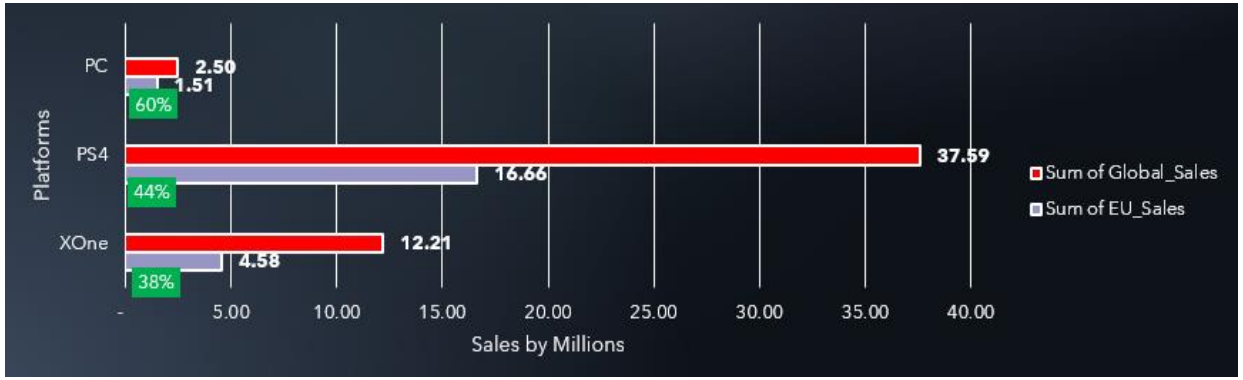


North America's Xbox sales account for more than half of all global sales.

The Wii platform's sales mostly come from North America as well.

Xbox, Wii, and PlayStation have consistently been the best-selling platform in North America for the last decade

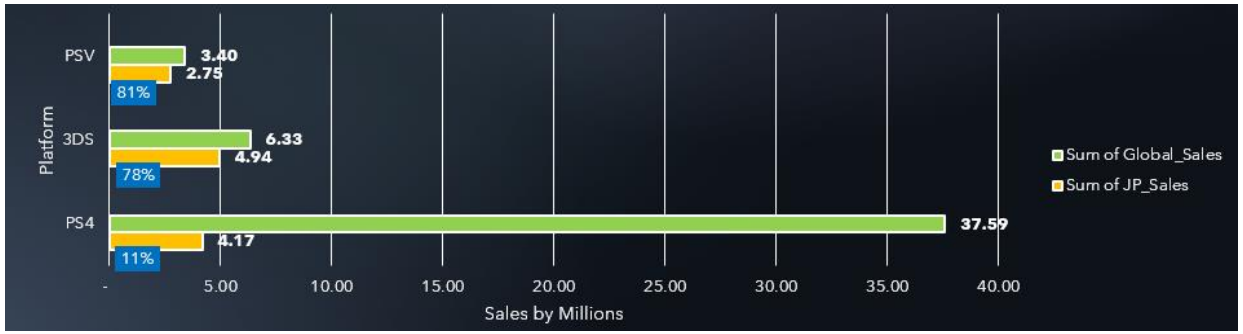
Europe



The PlayStation and Xbox platforms perform well in Europe.

The PC platform is highly popular in Europe, taking up more than half of the global sales. This could be due to the rise of sites such as Discord, released in 2015, and Twitch, released in 2011.

Japan



The DS platforms are very popular in Japan. It was just shy of the top selling platform in 2016. It was slightly beat out by the PlayStation Vita. Both the PSV and 3DS accounted for more than three-fourths of the total global sales separately.

RECOMMENDATIONS

Focus most of the budgeting on the European market, since Europe now holds the most global sales. Focus especially on PC games since Europe's PC sales account for 60% of total sales.

Focus on North America budget as well, since they still hold the greatest number of sales over time. Focus on the Xbox platform here, as their sales account for more than half of the total global Xbox sales.

The budget for Japan should focus solely on PSV and 3DS platform budget since Japan's regional sales for both each account for more than three-fourth of the global sales.

For North America and Europe, focus budgeting on the Shooter, Action, and Sports genres since most of their sales are in these genres.

Focus on Adventure and Role-Playing games for Japan, as Role-Playing is the highest grossing genre there and in 2016, Adventure games held the most sales.

PREPARING FOR INFLUENZA SEASON

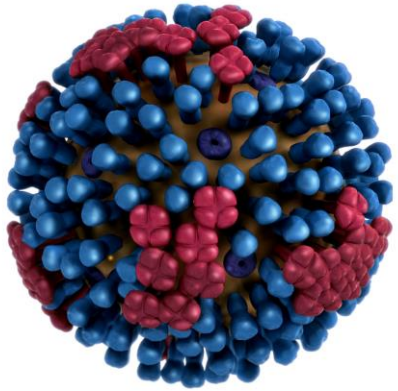
PROJECT OVERVIEW

Objectives:

- Provide a recommendation to a medical staffing agency for the distribution of temporary medical staff for the upcoming influenza season in the United States.
- Analyze the seasonality of the virus and the location of vulnerable populations.

Key Questions:

1. When does Influenza season occur?
2. Who is part of the vulnerable population?
3. Where does staff need to be sent?
4. What are the next steps?

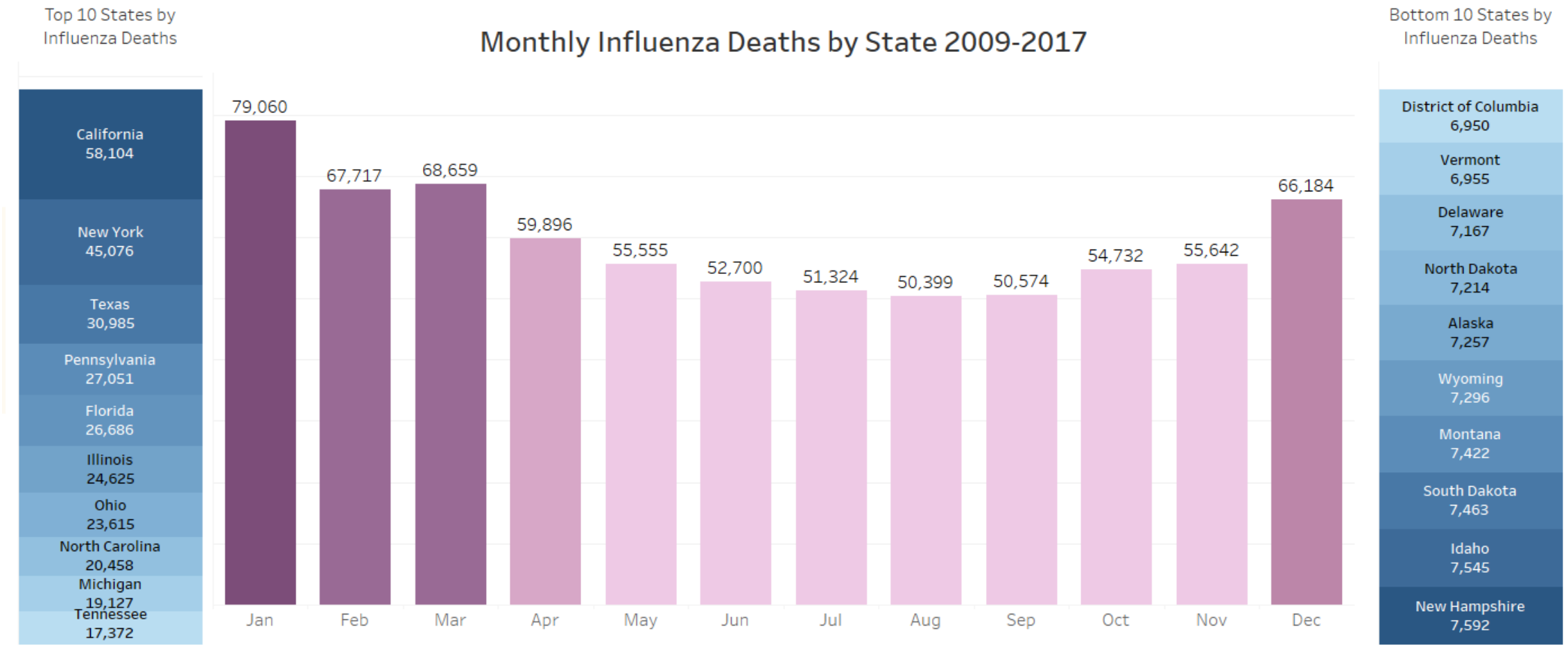


Skills
Data Cleaning
Data Transformation
Data Integration
Statistical Hypothesis Testing
Tableau Visualizations
Presenting Results

Tools Used

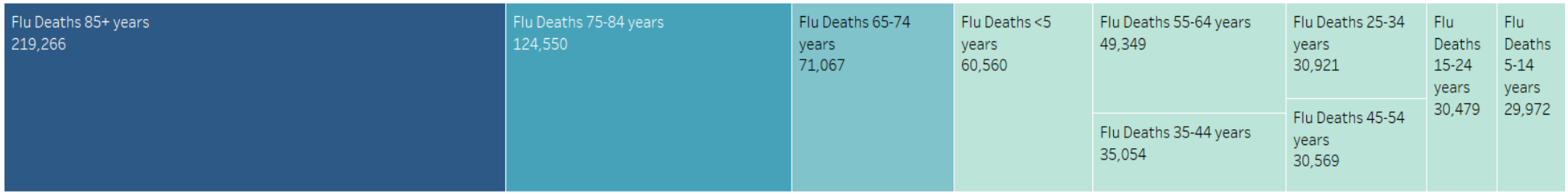
Data set: CDC Flu Deaths & US Census Data
Tableau: full storyboard here.

WHEN DOES INFLUENZA SEASON OCCUR



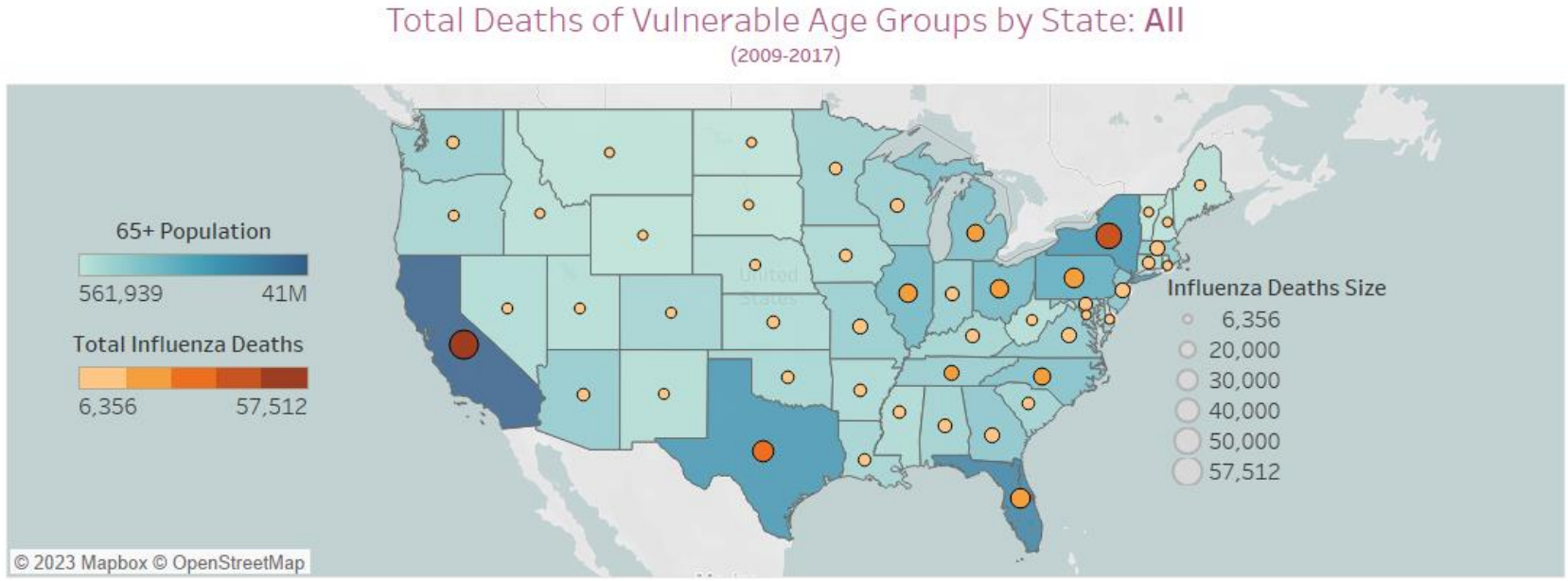
- The peak of influenza season runs from December to March, with the most deaths occuring in January.
- The states with the highest risk follow this pattern while states with the lowest risk have the deaths more dispersed.

WHO IS THE MOST VULNERABLE?



- When comparing age groups, people that are 65+ have more deaths than those <65.
- The 85+ population is mostly vulnerable, accounting for twice, thrice, quadruple, and more, deaths than any other population!

WHERE TO SEND STAFF?



- States with higher populations face higher influenza deaths, especially those with greater 65+ populations.
- The 85+ population faces the greatest number of deaths for most states.

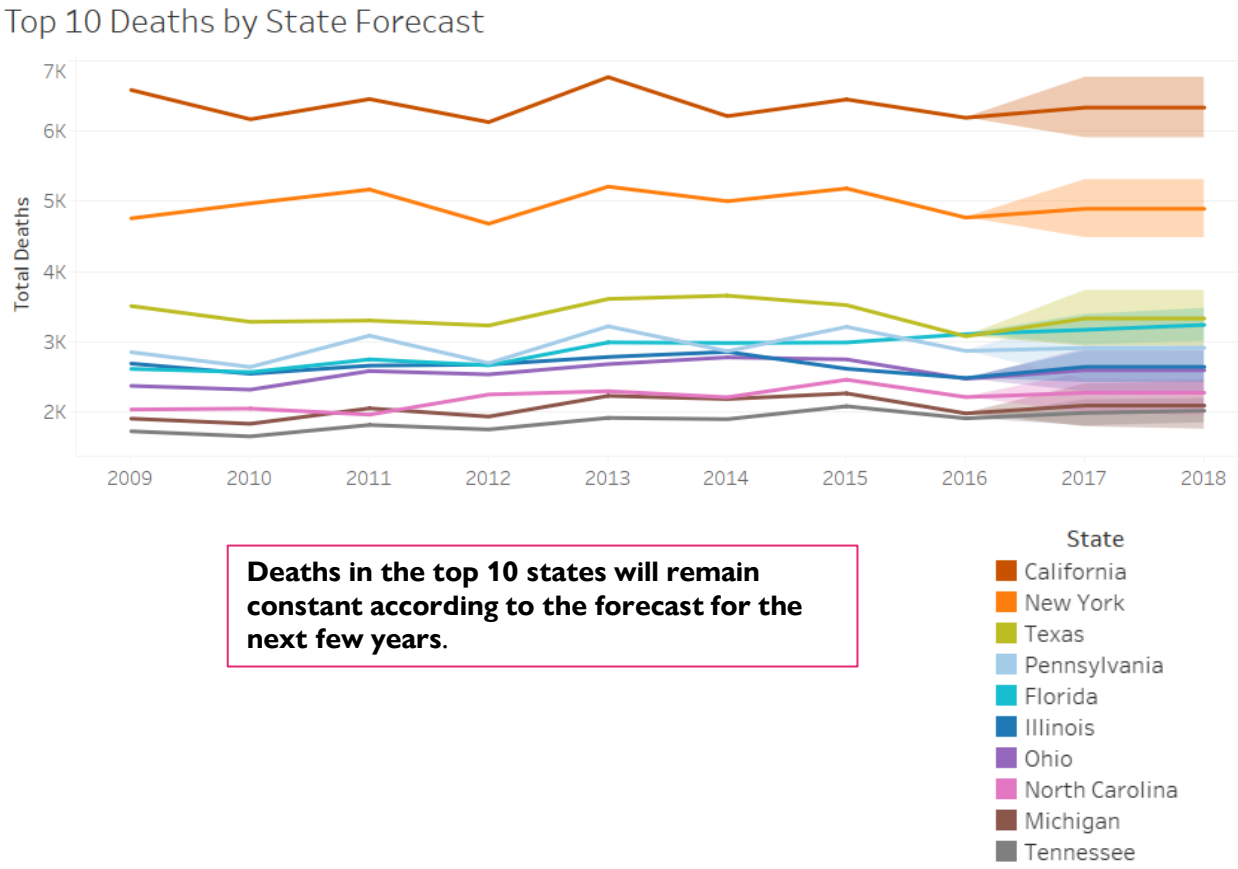
NEXT STEPS?

Recommendations

• The total deaths will continue to remain consistent for the high-risk states, so actions need to be taken.

• Monitoring of of staff-to-patient ratio, as well as mortality and vaccination rates, will be critical to swerve influenza death rates.

• Allocating more staff to high-risk areas to assist vulnerable patients will alleviate the growing demand, as higher population require more help. Sending staff during late October or early November. right before the higher numbers begin, will allow for help earlier on.



Deaths in the top 10 states will remain constant according to the forecast for the next few years.

ROCKBUSTER STEALTH

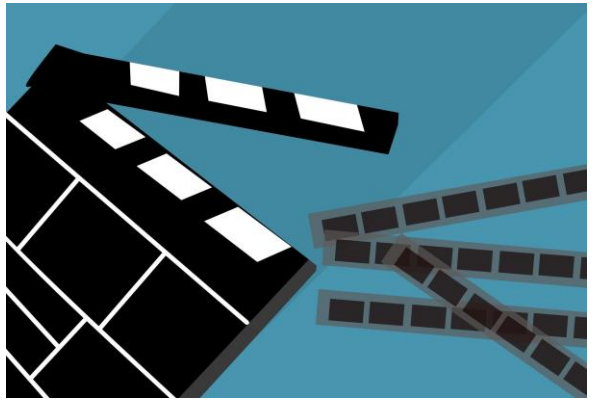
PROJECT OVERVIEW

Objectives

- Perform an analysis of past movie rentals and customer trends.
- Make strategic recommendations for the (fictitious) company to transition to online movie streaming.

Key Questions

- ➢ Which movies contributed the most/least to revenue gain?
- ➢ What was the average rental duration for all videos?
- ➢ Which countries are Rockbuster customers based in?
- ➢ Where are customers with a high lifetime value based?
- ➢ Do sales figures vary between geographic regions?



Skills
Database Querying
Joining Tables
Common Table Expressions
Subqueries
Creating a Data Dictionary
CRUD Operations

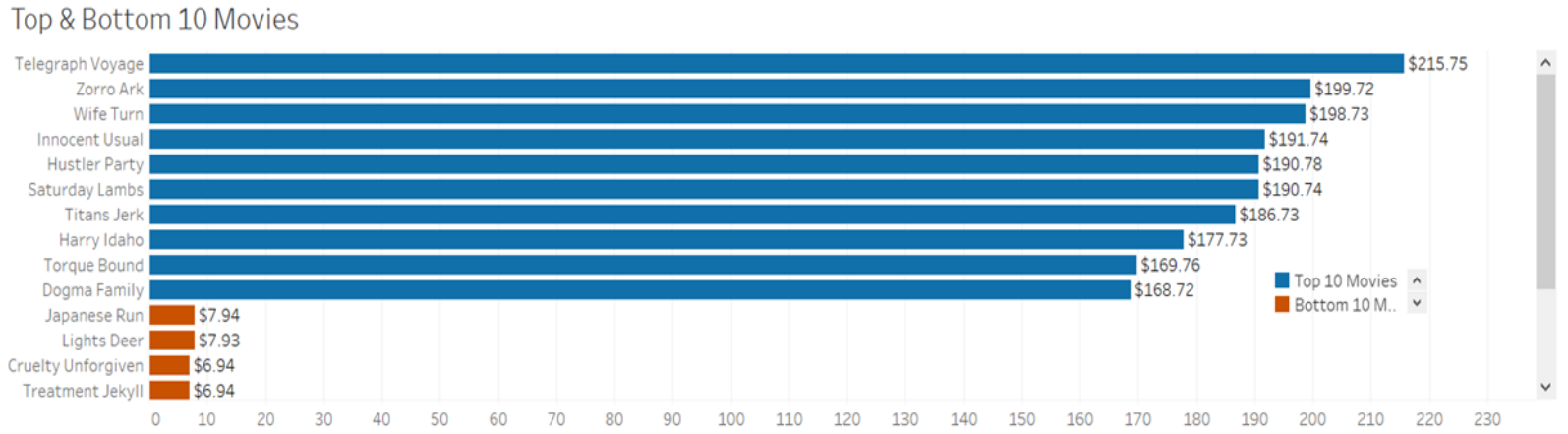
Tools Used

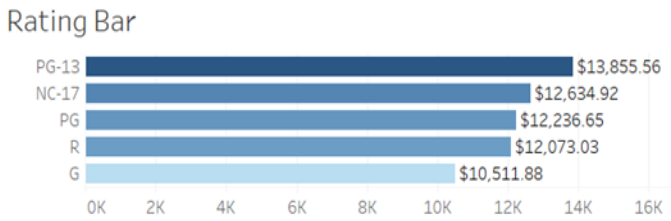
Tableau: full storyboard here.
Powerpoint: full presentation here.
Github: full repository here.

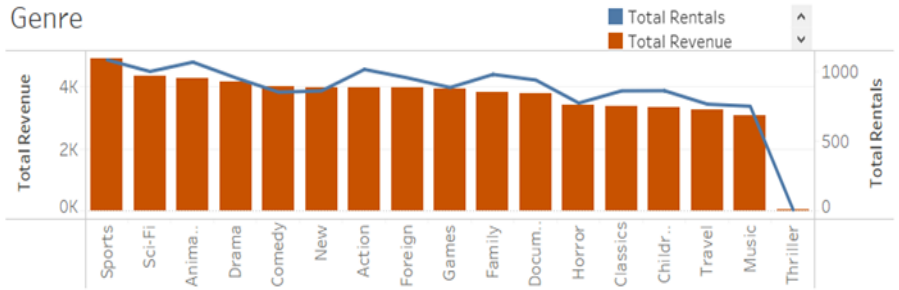
REVENUE ANALYSIS

Process

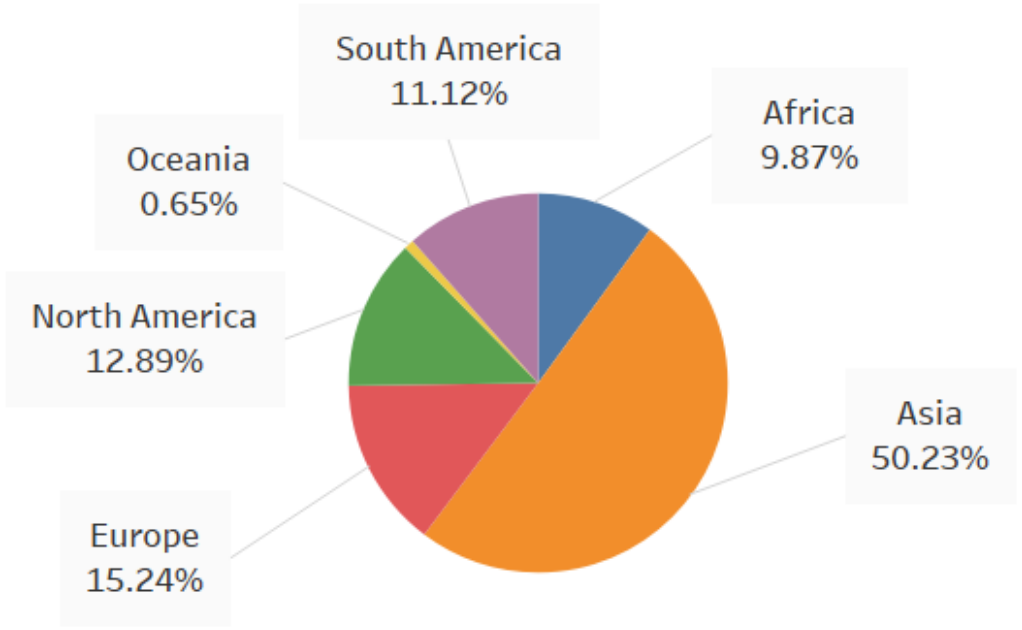
- Utilized CRUD commands and proper syntax to summarize and clean dirty data.
- Filtered, grouped, and summarized the data to generate insights on revenue streams.







GLOBAL REVENUE



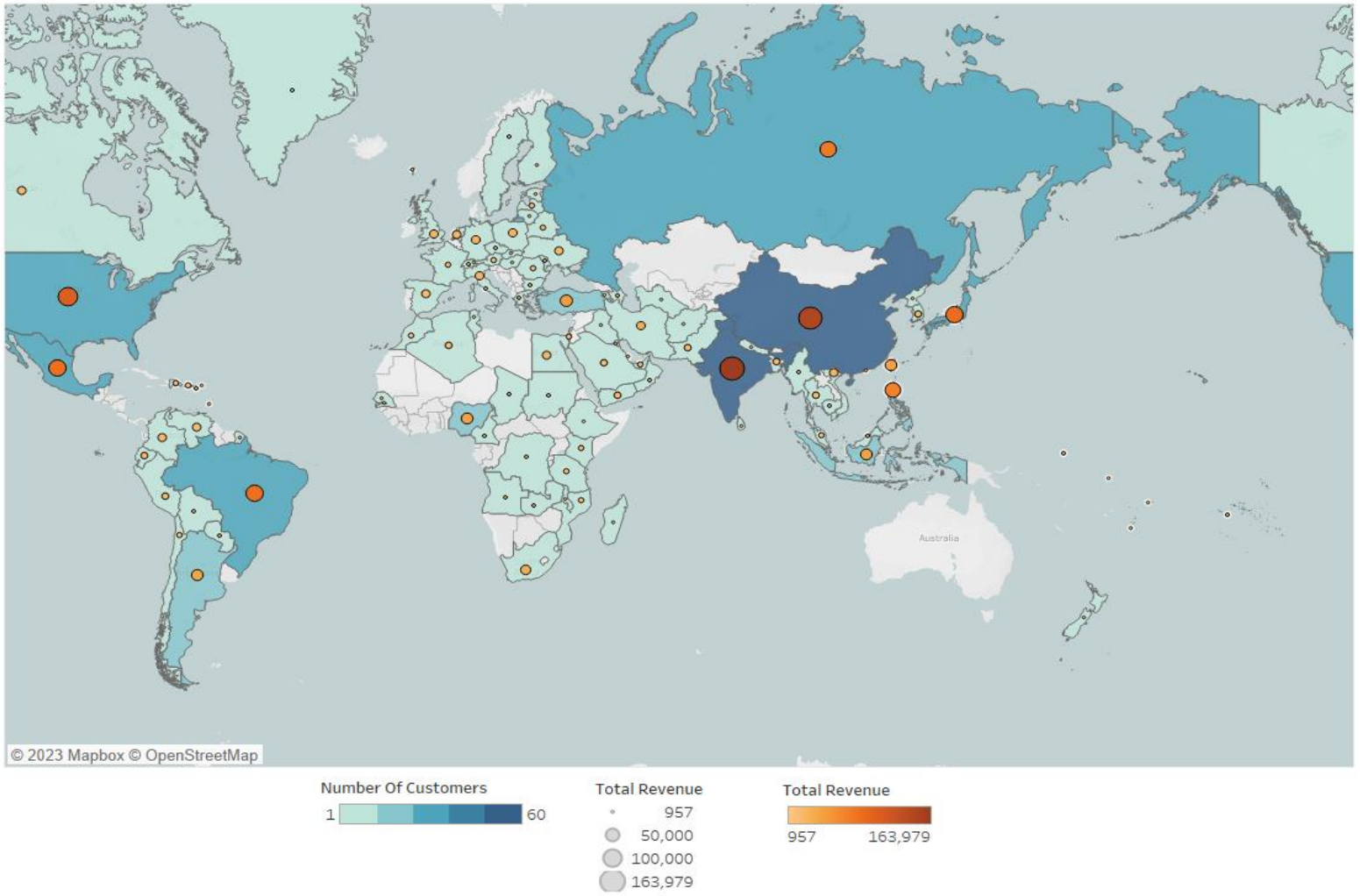
Regions	Number of Customers	Total Revenue
Asia	301	$856,637
Europe	87	$259,875
North America	77	$219,889
South America	69	$189,709
Africa	60	$168,361
Oceania	5	$11,076

- Asia accounts for half of the global sales.
- Asia also accounts for 50% of the overall customers.

CUSTOMER ANALYSIS

- Analyzed Rockbuster database and joined tables to highlight the relationship between sales by country and the number of customers. In the example above, I summarized the data by customer count and revenue.

- I also performed subqueries and CTEs to answer more complex business questions such as who were the top 10 customers within the top 10 countries by sales



RECOMMENDATIONS

- Acquire more titles under the Sports, Sci-Fi, and Animation genres, as they generate the most revenue and rentals
- Promote Sports, Sci-Fi, and Animation movies.
- Invest in movies in different languages to attract customers that do not speak English.
- Promote and acquire movies in the PG-13 movies, as they tend to have the most rentals and garner the most revenue.
- Focus marketing efforts in Asian and African countries as they have the most customers.
- Open a rewards program for loyal customers.

INSTACART BASKET ANALYSIS

PROJECT OVERVIEW

Objectives

- Uncover more information about grocery sales patterns.
- Derive insights and suggest strategies for better segmentation based on the provided criteria.

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 types of products that are more popular than others?
5. What different types of customers can be identified and how do their ordering behaviors differ?

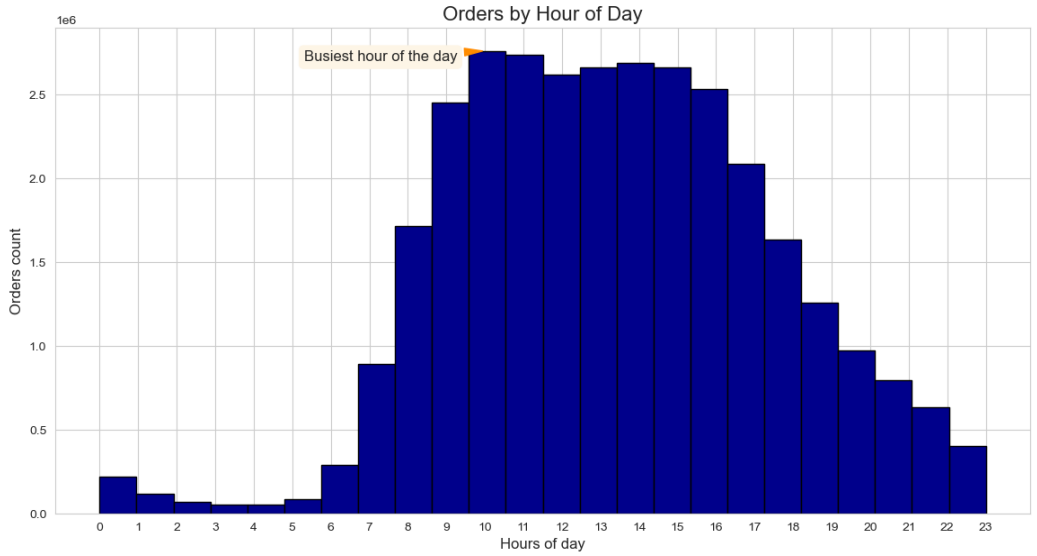
instacart

Skills
Data Wrangling
Data Consistency Checks
Grouping & Aggregating Data
Combining Data
Data Visualization with Python

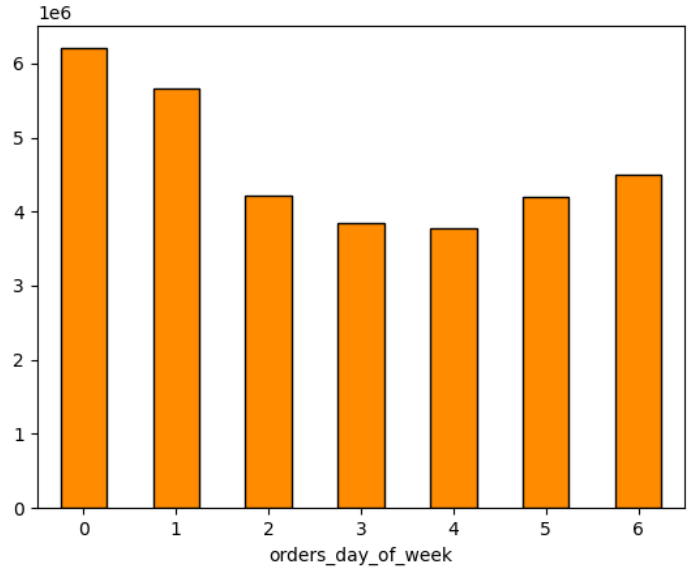
Tools Used

Github: full repository here.
Excel: full report here.

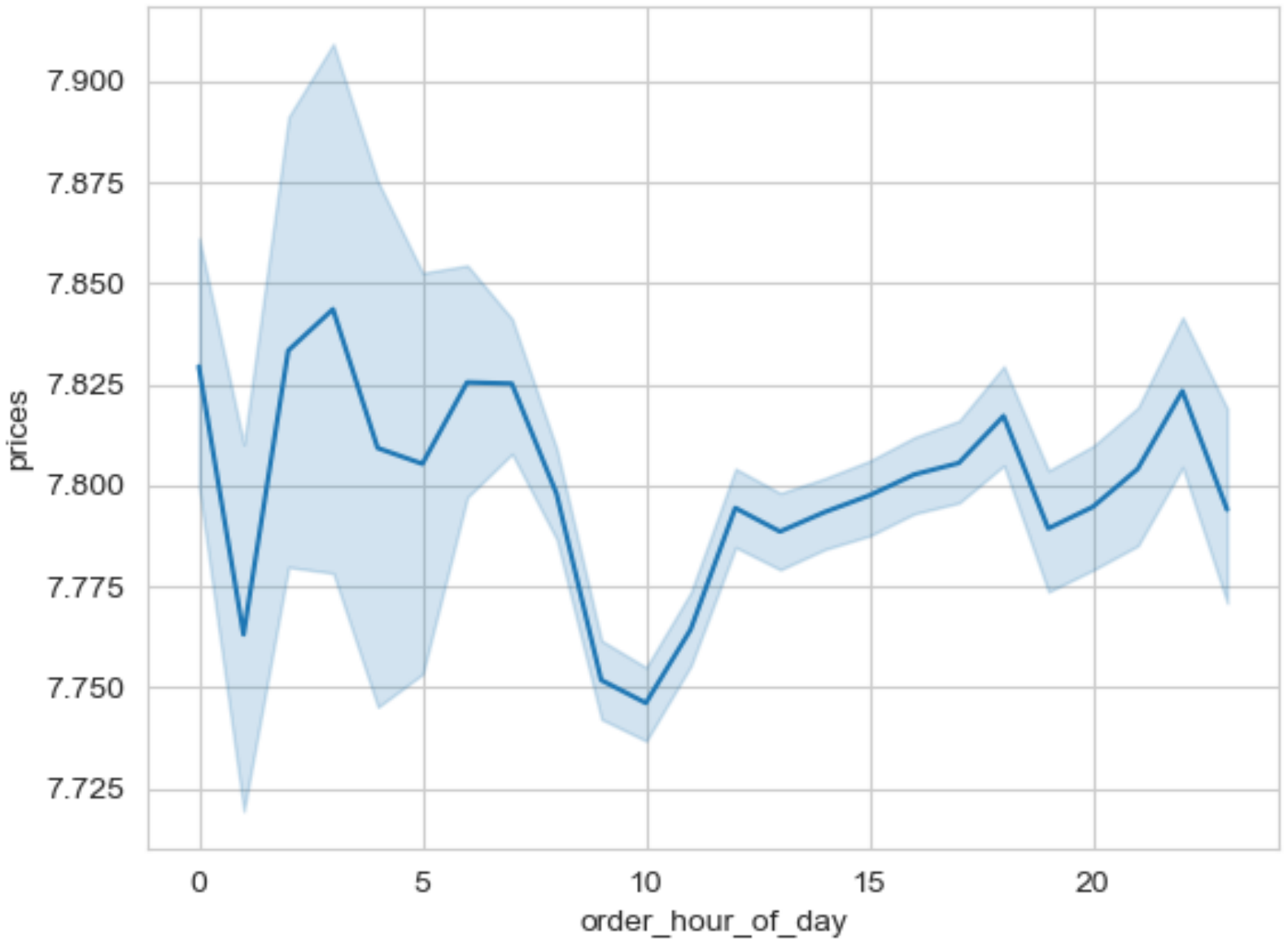
BUSIEST HOURS AND DAYS



This graph depicts the busiest hours of the day. It shows that the busiest time of day is between 10 am and 4 pm, with the peak hour being 10 am.



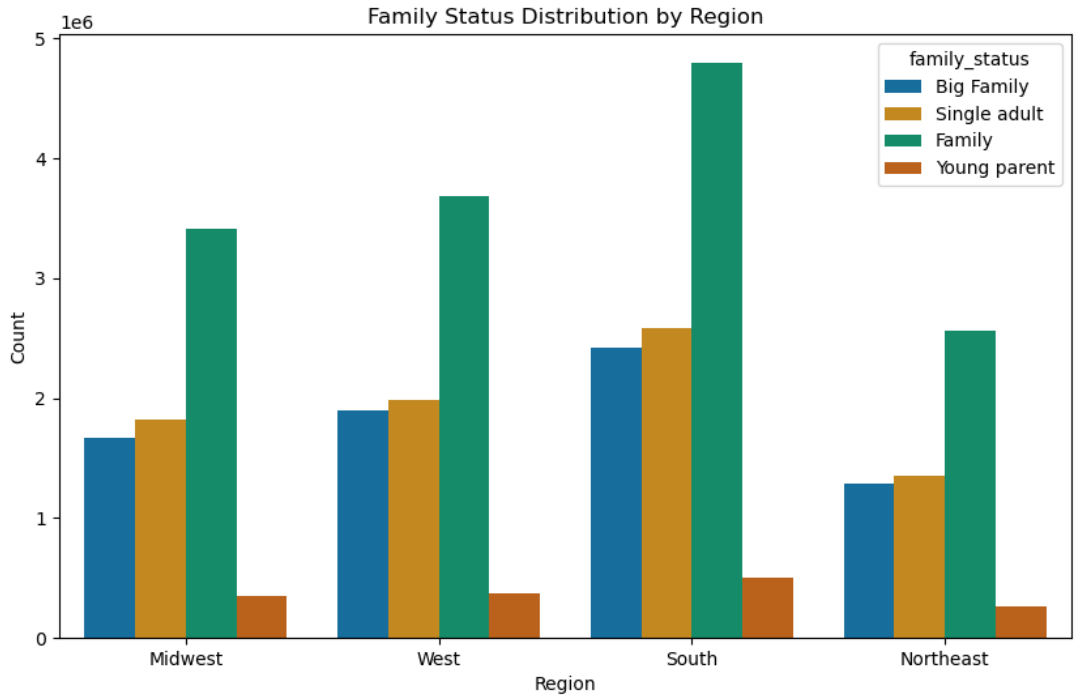
Histogram showing the busiest, and least busy, days of the week. Saturday and Sunday are the two busiest days, which makes sense considering the weekend is when most people have time to shop or need groceries.



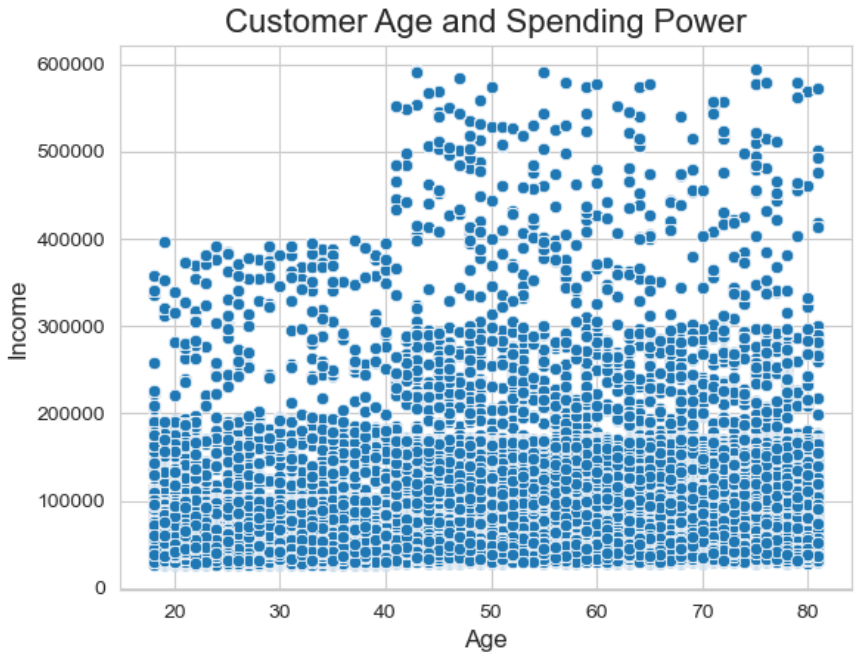
TIMES OF DAY WHEN MONEY IS MOST SPENT

The most money is spent between 2 am and 10 am, despite that being a time when there are less orders. The overall expenditure seems to remain the same throughout the day.

CUSTOMER PROFILE ORDERING HABITS



The distribution of orders based on family status per region is showed. The Southern region has the most orders, with families of married parents and 2 or less dependents having the most orders for all regions.



While the scatterplot shows that most people of all ages are situated below the 200000 mark for income, most of the customer base's spending power comes from people 40+. This makes sense since 40 and over is usually families are established and settled.

RECOMMENDATIONS

Conducting ads and promotions targeting slower times of day, such as discounts and speedier delivery times. Additionally, to optimize the user experience, consider offering personalized recommendations and promotions that align with the identified shopping patterns, such as promoting weekend deals and targeting marketing efforts between 9 a.m. and 5 p.m.

The sales team should promote sales on higher-priced items before 7 a.m. and during late afternoons when most money is spent. Additionally, having a reward incentive for ordering a high amount for each order, promoting more significant additional dollar purchases with incentives like free or express delivery.

Create incentives for newer customers. A promotion could be run that says that once they pass the ten orders mark, they get discount incentives if they make the order within 20 days. For regular customers, once they pass 40 orders, they earn more incentives for becoming loyal customers.

Running promotions and ads in the South since they bring in the most money. Additionally, promoting ads in the Northeast, such as discounted prices, will help to boost sales there. Promotions geared toward young parents, such as discounts or sales on baby products or snacks, could help boosting their sales as well.

Since most of the users are older and have families, pushing ads and promotions towards that population would be beneficial. However, pushing deals towards people with middle or low income, especially young parents and adults, would provide them a bit of relief if they are not buying due to financial issues.

GUN VIOLENCE ANALYSIS

PROJECT OVERVIEW

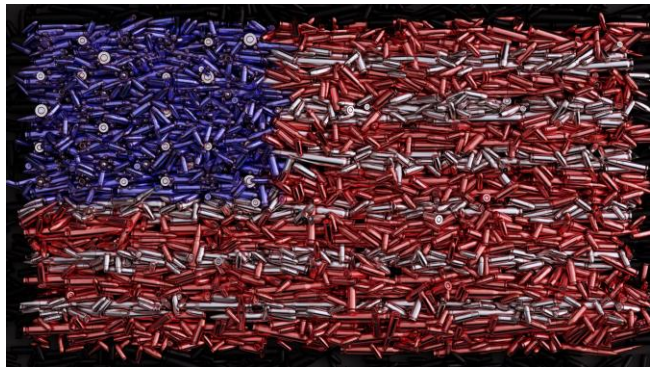
Objectives

- Find key variables to figure out any key areas where gun violence occurs the most.
- Use these variables to study relationships using linear analysis

Key Questions

1. Does the frequency of gun violence incidents vary significantly across different states and cities?
2. Are there temporal patterns in gun violence incidents, such as monthly or yearly trends?
3. Are males more disproportionately affected by gun violence than females?

Dataset: from Kaggle, full download link here.
Github: full repository here.
Tableau: full storyboard here.



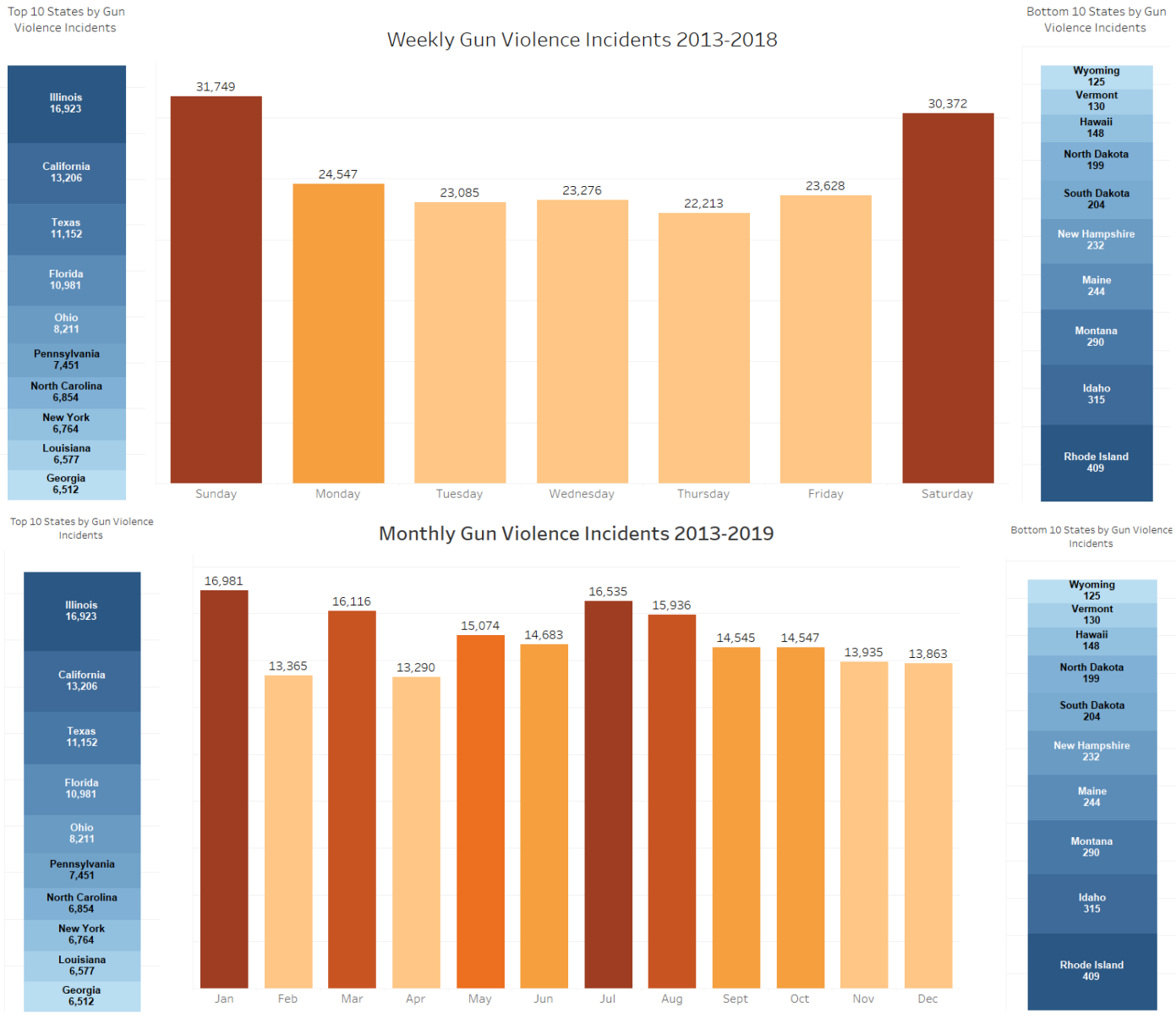
Skilled Used
Sourcing open data
Data cleaning, wrangling, and subsetting
Performing exploratory visual analysis
Creating geographical visualizations
Supervised machine learning with linear regression
Unsupervised machine learning with k-means clustering
Sourcing and analyzing time-series data
Creating a data dashboard

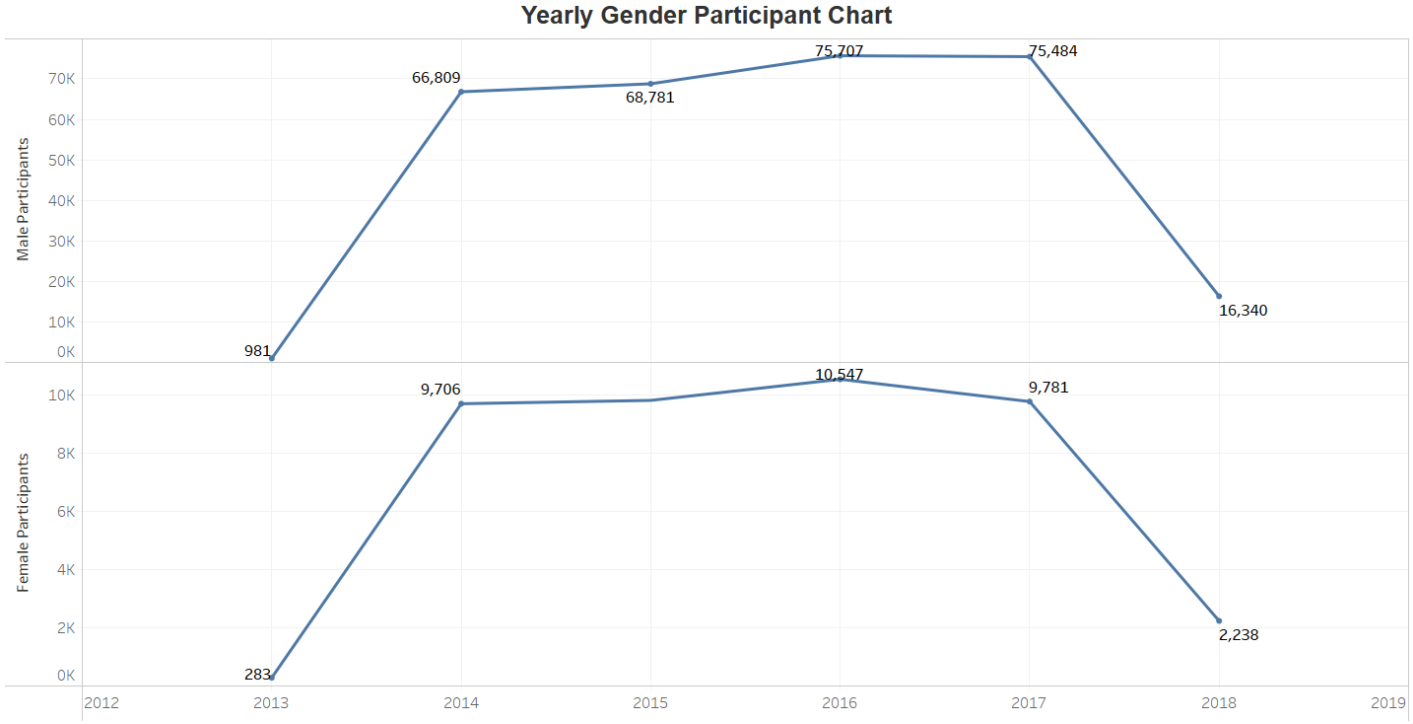
Tools Used

INCIDENTS ANALYSIS

- **Sunday** and **Saturday** are the days with the highest numbers of gun violence incidents.
- **Thursday** is the day with the least amount of gun violence incidents.
- **Illinois**, **California**, and **Texas** are the states with the most gun violence incidents.
- **Wyoming**, **Vermont**, and **Hawaii** have the least gun violence incidents

- **January**, **March**, and **July** are the months with the **highest numbers** of gun violence incidents.
- The season of **summer** has the most incidents (Jun-Aug)
- The season of **fall** has the least number of incidents (Sep-Nov)





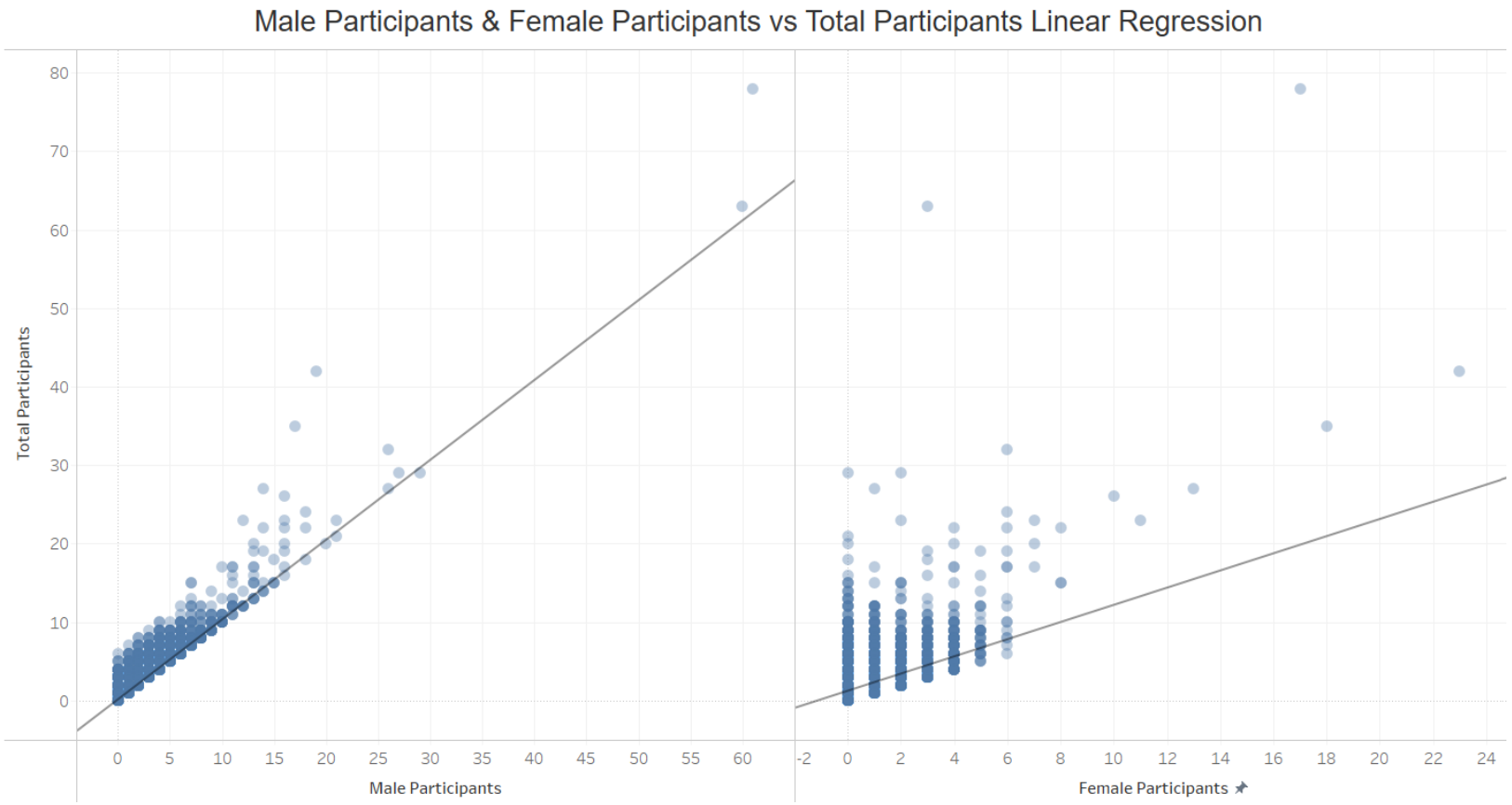
PARTICIPANTS ANALYSIS

• A linear analysis of the number of incidents for male and female participants was conducted. The participants are the number of people involved. That could mean suspects or victims per incident.

• The line chart showcases a similar line pattern for both genders, with the peak for both being in 2016.

• Males make up the vast majority of gun violence incidents, having six to eight times more participants per year then that of women.

LINEAR REGRESSION



Male Participants

- Male participants make up of most of the participants per incident, making up to 66%-100% of the incidents where males are involved.

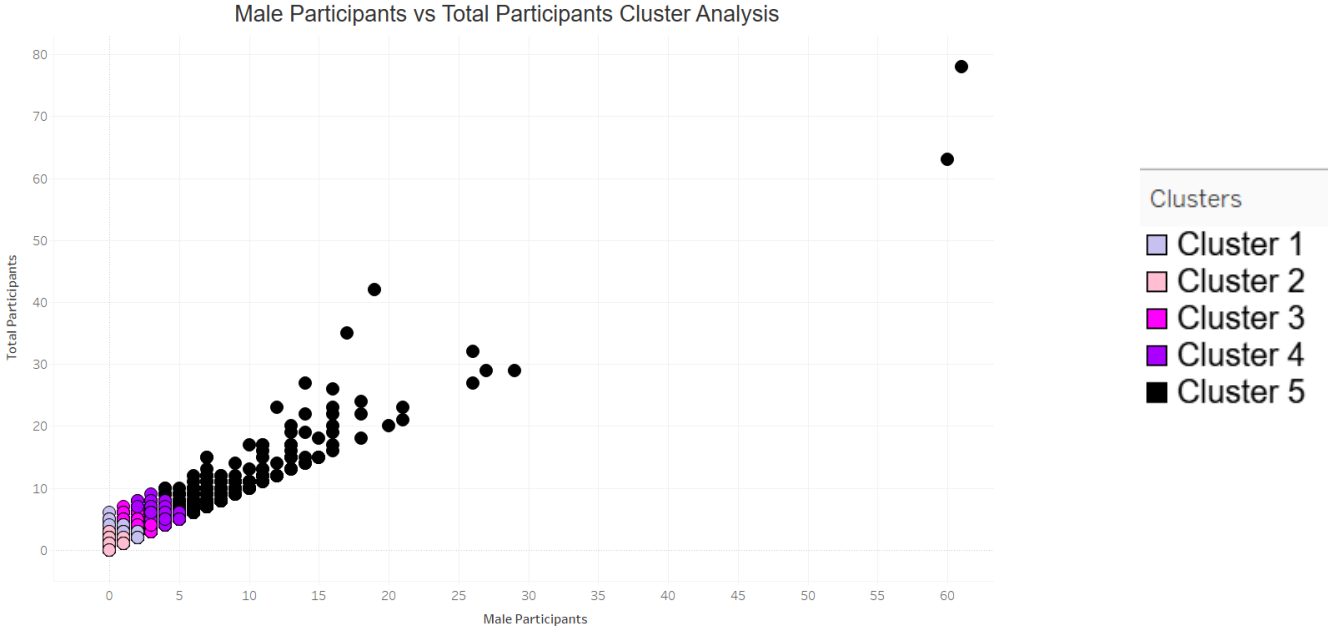
- The R-squared value for the line of the regression model is 0.8516, meaning that there is a strong correlation between the two variables.

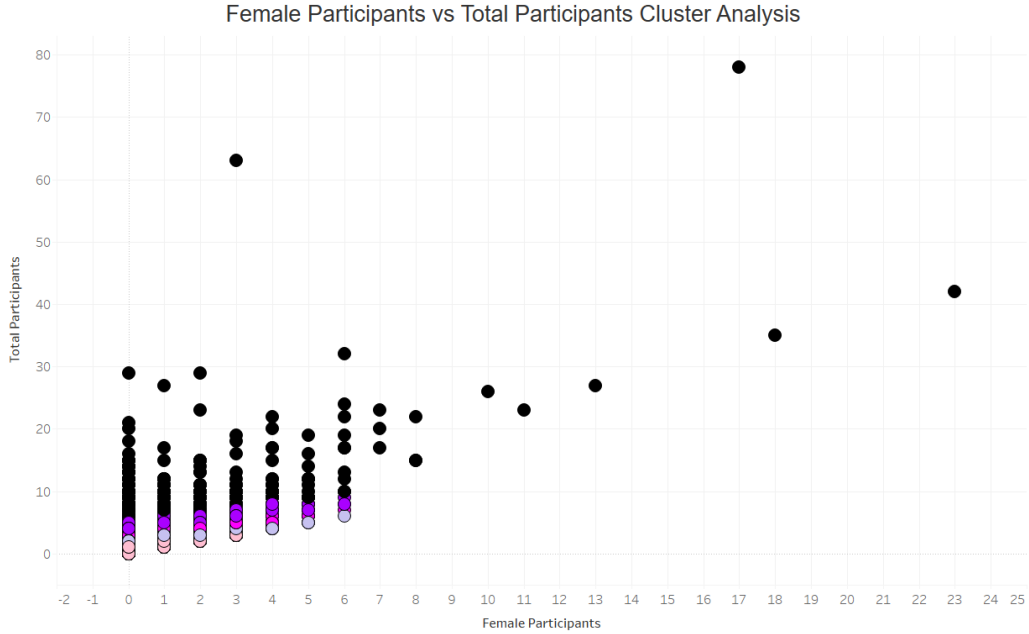
Female Participants

- Female participants make up between 0%-65% of the participants per incident, whether if a suspect or participant.

- The R-squared value for the line of regression is 0.1779, meaning there is a weak correlation between the two variables.

K-MEANS CLUSTERING





The k-means clustering was conducted to take a further look at the data of male participants vs. total participants and female participants vs. male participants.

This cluster analysis identified five different clusters, based upon the severity of violence each group faces. This factor is based upon the total incidents and total participant averages, as well as the proportion of male to female participation.

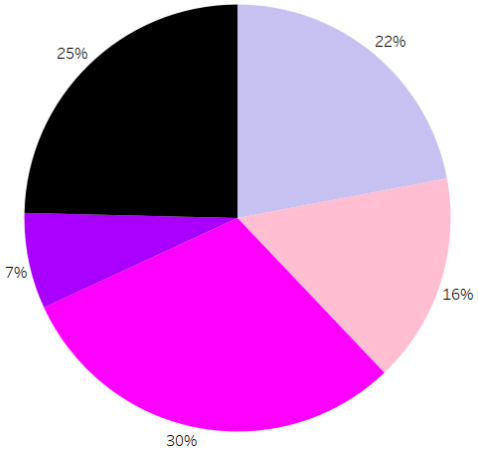
Many of the clusters were stacked on top of each other due to there being many overlapping values within the clusters.

Averages per cluster 1-5 (respectively) for male participants: 0.97, 1.35, 3.29, 1.17, 1.04

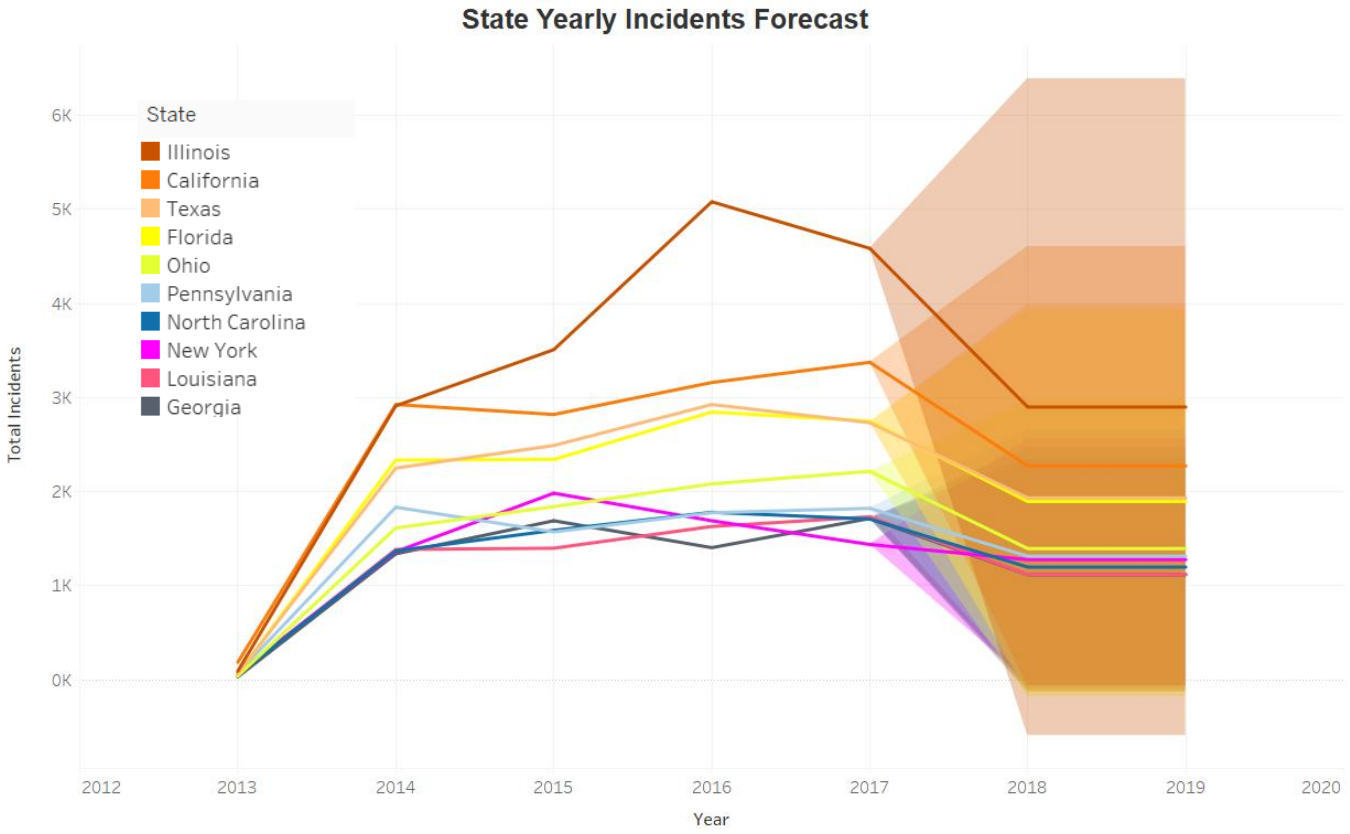
Averages per cluster 1-5 (respectively) for female participants: 0.13, 0.20, 0.48, 0.18, 0.12

Male participants have much higher values!

The magenta cluster (cluster # 3) has the highest average of male participants and consists of about 30% of all incidents. This suggests that there are a higher number of male individuals involved in incidents within this cluster. Conversely, the purple cluster (cluster #4) has a lower average and consists of about 7% of the cluster data. This indicates fewer male participants on average per incident in this cluster. Cluster #3 has the largest number of unique incidents, suggesting that most incidents fall into this cluster area and that there are about 3 males per incident involved 30% of the time.



RECOMMENDATIONS



A forecast was created to predict what the rest of 2018 and what 2019's data would look like for the top 10 states with the most gun violence incidents. The numbers for Illinois stay relatively high, comparatively to the others, but the numbers do not go back to their peak.

Recommendations

Steps for the top clusters (magenta, black, and lilac)

- Increase community outreach and engagement efforts in areas associated with the top 3 clusters. Collaborate with local leaders and residents to identify community-specific challenges and implement solutions.
- Provide social support services, especially within the top 3 clusters, to address underlying issues that may contribute to higher male participation, which might involve partnering with social service organizations, mental health professionals, and community support networks.
- Establish a feedback mechanism to gather insights directly from residents within the top 3 clusters. This participatory approach ensures that interventions align with community needs and preferences.

General next steps

- Launch public awareness campaigns highlighting the risks and consequences of gun violence, particularly during peak months and among specific demographics (top 10 states and males). These campaigns can include social media, community events, and educational materials.
- Provide specialized training for law enforcement to handle situations involving male individuals, emphasizing de-escalation techniques and community-oriented policing.

NUPUR PARIKH

USE THE LINKS TO SEE MY WORK OR CONTACT ME!

Contacts & Links