



Oluwole Oni

Data Analyst | Business Intelligence | Data Storyteller

- Excel • PowerPoint • Python • Anaconda • PostgreSQL • Jupyter

Table of contents

1. About Me

Brief introduction, background, and skills.

2. Data Analysis Project Overview and Tools

Overview of tools and technologies used in the projects: Python, SQL, Tableau, Excel, and Power BI.

3. Project 1 – GameCo Video Game Sales Analysis.

An analysis of video game sales data to uncover popularity trends and performance by platform and region.

4. Project 2 – Preparing for Influenza Season.

Predictive analysis to forecast flu trends and inform healthcare resource allocation.

5. Project 3 – Rockbuster Stealth Data Analysis.

Exploration of customer and rental data to support business strategy for a global movie rental company.

6. Project 4 – Instacart Grocery Basket Analysis.

Customer segmentation and purchasing behavior analysis to inform targeted marketing strategies.

7. Project 5 – Gun violence in the U.S.A (2013 – 2018)

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

About Me

I hold a Bachelor's degree in Mechanical Engineering from a reputable university in Nigeria and an Automotive Master Certificate from the Community College of Baltimore County, USA. My zeal for excellence and continuous learning has driven me to earn multiple professional certifications, including becoming an Automotive Service Excellence (ASE)-Certified Technician.

Over the years, I've worked with both independent and corporate organizations as an automotive technician, where I relied heavily on diagnostic tools, performance metrics, and sensor data to identify and resolve complex technical issues. These hands-on experiences revealed the immense power of data-driven insights in improving decision-making and system performance.

Gradually, my curiosity expanded beyond mechanical systems to understanding how data patterns, trends, and visualizations can transform raw information into actionable insights. This passion for data led me to pursue a Data Analytics certification with CareerFoundry, complete an Introduction to Data Analysis using Microsoft Excel course on Coursera, and receive private mentorship in R programming.

Today, I'm channeling my analytical mindset, technical background, and problem-solving skills into the field of data analytics, where I aim to uncover insights that drive innovation and efficiency.

Data Analysis Project Overview and Tools

□ Project 1- GameCo video game sales analysis

I examined and refined the GameCo dataset to identify sales trends and regional performance differences. Metrics were summarized and visualized through clear, data-driven charts to communicate insights and support strategic decision-making

Duration: 3 weeks

Tools Used: Excel, Microsoft word and PowerPoint

□ Project 3 - Rockbuster Stealth Data Analysis

I structured and stored data efficiently, created entity diagram and data dictionary, performed advanced SQL operations including filtering, cleaning, joining tables, subqueries and common table expressions and presented the analyzed results to derive actionable insights.

Duration: 4 weeks

Tools Used: Excel, Microsoft word, PostgreSQL, PowerPoint, DBvisualizer, Tableau.

□ Project 2 - Preparing for influenza season

Designed and executed a data research project by sourcing and profiling high-quality data(s), conducting statistical analyses and hypothesis testing.

Creating clear visualizations including temporal, statistical, comparative and spatial charts to forecast trends and support informed decision-making for influenza season preparedness.

Duration: 3 weeks

Tools Used: Excel, Microsoft word, Power Point, Tableau

□ Project 4 – Instacart Grocery Basket Analysis

I performed data wrangling, consistency checks, and variable derivation on all the data used for the analysis, then grouped and aggregated data to uncover more information about sales patterns visualizing key insights with Python for better business understanding.

Duration: 8 weeks

Tools Used: Jupyter notebook, Python, Excel.

□ Project 5 – Gun violence Analysis

I conducted comprehensive exploration of gun violence in the United States using a dataset of over 260,000 recorded incidents from January 2013 to March 2018. The analysis combines descriptive statistics, visualizations, clustering, and trend analysis to understand patterns, identify high-impact incidents, and assess the relationship between firearms involved and victim outcomes.

Duration: 3 weeks

Tools Used: Microsoft word, Python(jupyter notebook), Tableau

Project 1- GameCo.



Problem:

- GameCo aims to leverage data-driven insights to guide the development of new games and forecast their potential performance in the market

Objectives:

- Perform descriptive analysis to identify overall sales trends and patterns
- Examine sales pattern across different region
- Identify top-performing genres and assess competition across markets.
- Analyze changes in game popularity and sales over time.
- Visualize key insights to support GameCo's strategic, data-driven decisions.



Data and Methodology

Data:

- Historical sales of video games Dataset sourced from VGChartz.

About the data set:

- It tracks the total number of units of games sold (not financial figures) from 1980 to 2016.
- The numbers represent units sold in millions. When you see the number “1.2,” for instance, this represents a total of 1.2 million units sold.

Methodology:

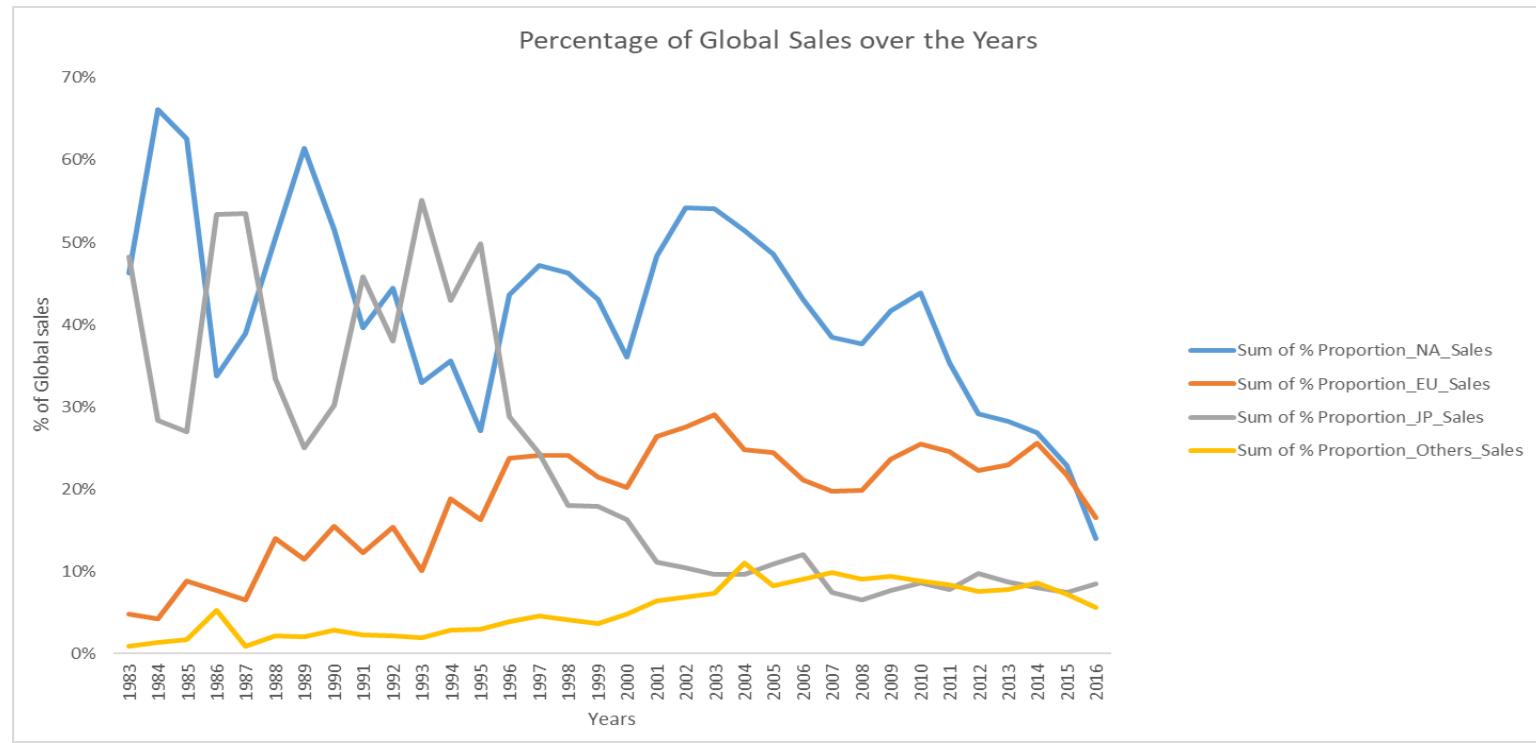
Data Cleaning: The GameCo dataset was thoroughly cleaned to ensure accuracy and consistency before analysis. Missing values were handled through mean imputation for regional sales columns, while the *NA Units* column and an empty final row were removed. Duplicates and blank entries were corrected, special characters in game names were standardized, and all sales figures were rounded to two decimal places. These steps ensured a reliable and well-structured dataset for accurate analysis and visualization.

Data Processing: Descriptive statistical techniques were applied to summarize sales data across genres, publishers, and regions. Data was grouped and aggregated using **Excel** and **Pivot tables** to identify sales trends, regional performance variations, and changes in game popularity over time. Visualizations such as bar charts and line graphs were created to highlight key insights, allowing for clear comparisons and trend identification across multiple markets.

Data Exploration

Expectation

- GameCo current understanding assumes that the sales of video games for the various geographic regions have stayed the same in year.



- North America:** Led sales, peaking over 60% before declining after 2008.
- Europe:** Grew steadily, holding 20–30% of total sales.
- Japan:** Strong early performance, then declined after the mid-1990s.
- Other Markets:** Gradual growth, driven by emerging regions.

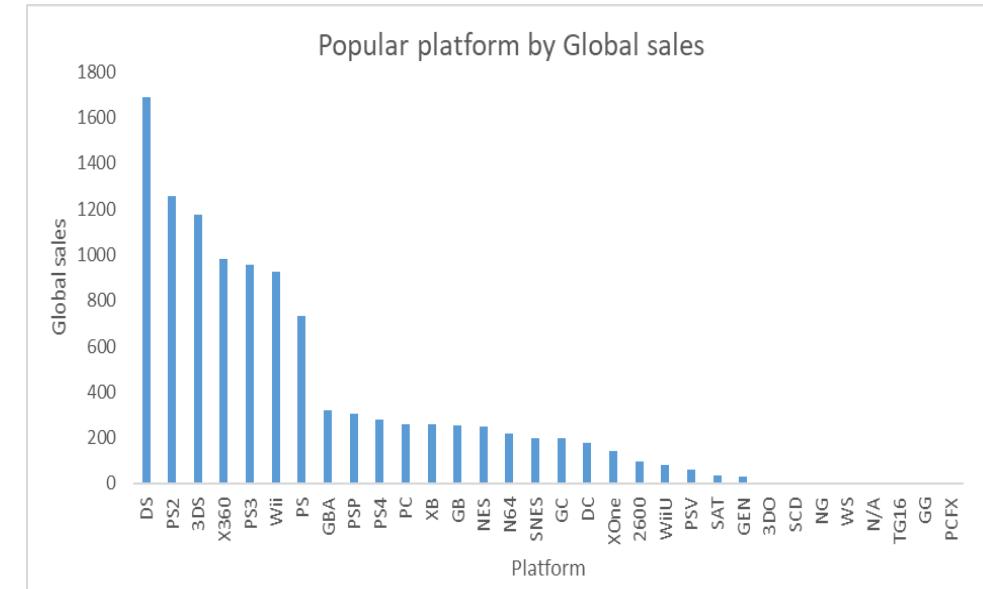
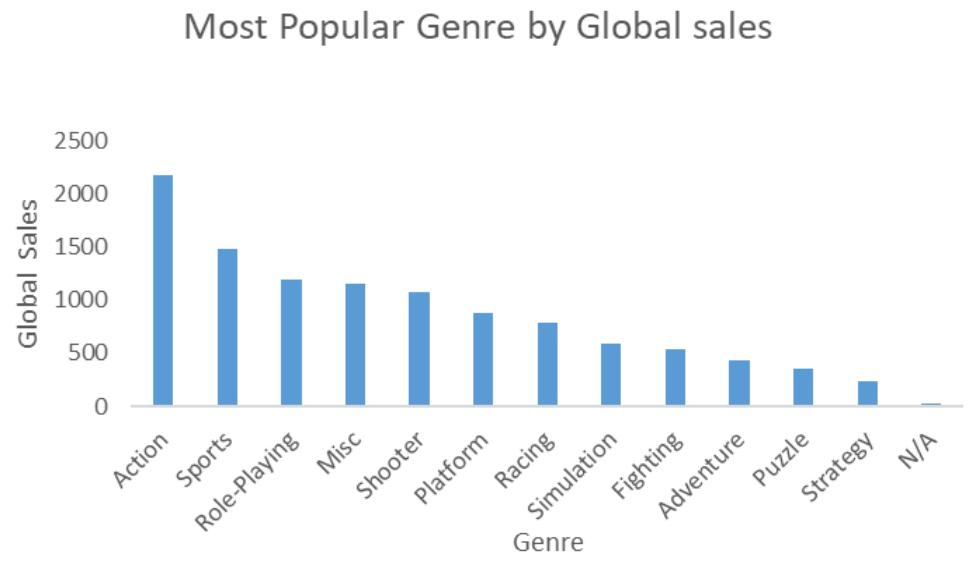
The analysis of global sales from 1983 to 2016 reveals significant shifts in regional market dynamics.

North America consistently dominated game sales, accounting for the largest share of global performance, while Europe showed steady growth and Japan's influence declined after early peaks. Emerging markets gradually expanded, contributing to a more balanced global distribution over time.

These insights challenge GameCo's initial assumption that sales performance across geographic regions has remained consistent, highlighting evolving consumer trends and market opportunities worldwide.

Data Exploration

- Identify top-performing genres and assess competition across markets



Action games consistently lead global sales as, making them the most popular genre, with Sports and Role-Playing genre also demonstrating strong worldwide demand.

On the platform side, the Nintendo DS emerges as the highest-selling system, supported by strong performance from other Nintendo platforms such as the Wii and Game Boy series. This trend highlights Nintendo's long-standing market dominance as one of the top-performing publishers over the years.

Overall, these insights help pinpoint where consumer demand has been strongest and provide valuable direction for strategic planning, product development, and competitive benchmarking.

Data Exploration

The performance of various game genres across key global regions: (North America, Europe, Japan, and Other markets).

Overall, North America and Europe account for the majority of sales across almost all genres, showing that these two regions drive most global demand.

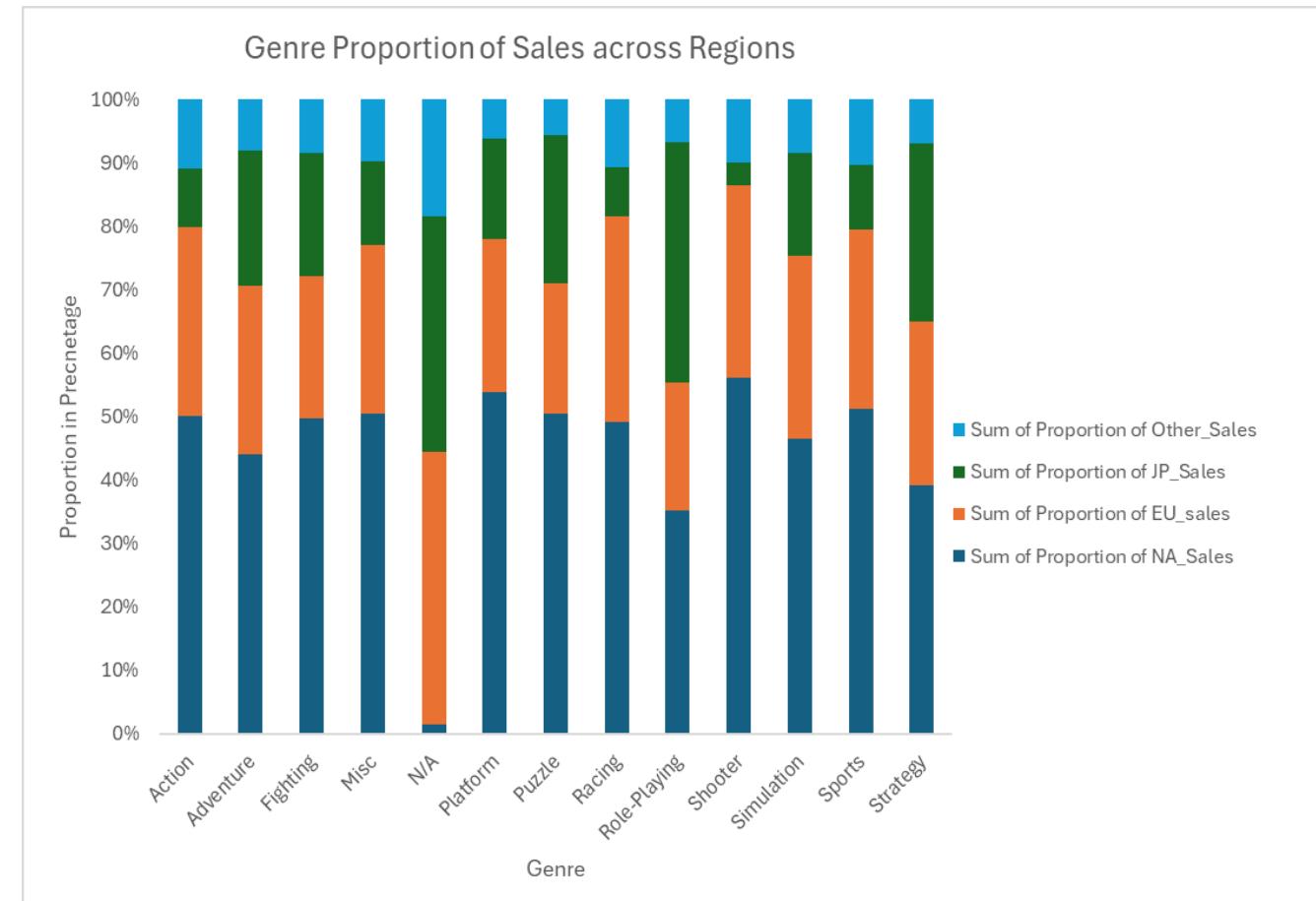
Genres such as **Action, Shooter, Sports, and Racing** show particularly high proportions of sales from these Western markets, with North America alone contributing nearly half of total sales in many cases.

In contrast, Japan shows more selective genre preferences. Genres like **Role-Playing and Fighting** receive a significantly higher share of Japanese sales compared to others, reflecting stronger cultural alignment and long-standing market demand for these styles. For most other genres, Japan contributes a smaller portion relative to North America and Europe.

The Other region contributes modestly across all genres but remains consistent, suggesting emerging or smaller markets with balanced but lower overall impact.

In terms of total sales, North America leads with 4,392.95 million units (40% of global sales), followed by Europe with 2,435.03 million units (22%). Japan accounted for 1,291.30 million units (12%), while Other regions contributed 822.24 million units (8%).

Overall, the data highlights strong Western market dominance, unique Japanese genre preferences, and consistent but smaller contributions from emerging regions. These trends can guide companies like GameCo in focusing on the most profitable genres and tailoring marketing or development strategies to regional demand.



Insights

- **Regional sales assumptions are outdated:** GameCo's assumption of consistent sales distribution across regions no longer holds, as each region shows distinct preferences and market contributions.
- **Action games lead globally:** Action genre consistently top global sales charts, confirming their universal appeal.
- **North America dominates genre sales:** Action games record the highest unit sales in North America, while Shooter games have the largest proportional share within the region, highlighting the market's strong demand for fast-paced genres.
- **Platform performance:** The Nintendo DS emerges as the highest-selling platform, reflecting the brand's enduring popularity. Nintendo, as a publisher, continues to demonstrate long-term market dominance across multiple regions.
- **Regional genre preferences:** Japan favors Role-Playing (RPG) and Fighting games, Europe shows balanced preferences across genres, and North America emphasizes Action, Shooter, Sports, and Racing games.
- **North America drives global sales:** The region accounts for the largest share of worldwide sales, underlining its strategic importance in revenue and market planning.
- **Emerging markets are modest but steady:** The Other region contributes consistently across genres, suggesting potential growth opportunities in smaller or developing markets

Recommendations

Redistribute marketing budget strategically:

- Increase marketing investments in Japan and Europe to boost their contribution to global sales.
- In North America, focus on enhancing the performance of underrepresented genres like Puzzle, Adventure, and Strategy games to maximize overall unit sales.

Prioritize underperforming genres globally:

- Genres such as Adventure, Simulation, Role-Playing, and Strategy currently contribute minimally to global sales. Allocate higher marketing and promotional resources to these genres across all regions to increase visibility and adoption.

Tailor products to regional preferences:

- Focus on Action and Shooter games in North America, Role-Playing and Fighting games in Japan, and maintain a balanced genre mix in Europe.
- Develop localized campaigns that align with each region's gaming culture and audience.

Strengthen presence in emerging markets:

- Invest in the Other regions with targeted campaigns, digital distribution, and affordable pricing strategies to expand market share in smaller but consistent markets.

Maintain competitive advantage with key franchises:

- Continue supporting high-selling platforms, particularly under Nintendo, to sustain long-term market dominance.

Use data-driven sales forecasting:

- Replace assumptions of uniform sales distribution with region-specific analytics to inform inventory, marketing, and development strategies.

Explore cross-genre and cross-platform opportunities:

- Introduce hybrid or crossover games to appeal to a broader audience and leverage global market potential.

Monitor market trends and consumer behavior:

- Regularly track emerging genres, platform adoption trends, and player preferences to adapt strategies and maintain leadership in key regions.

Project 2 - Preparing for influenza Season

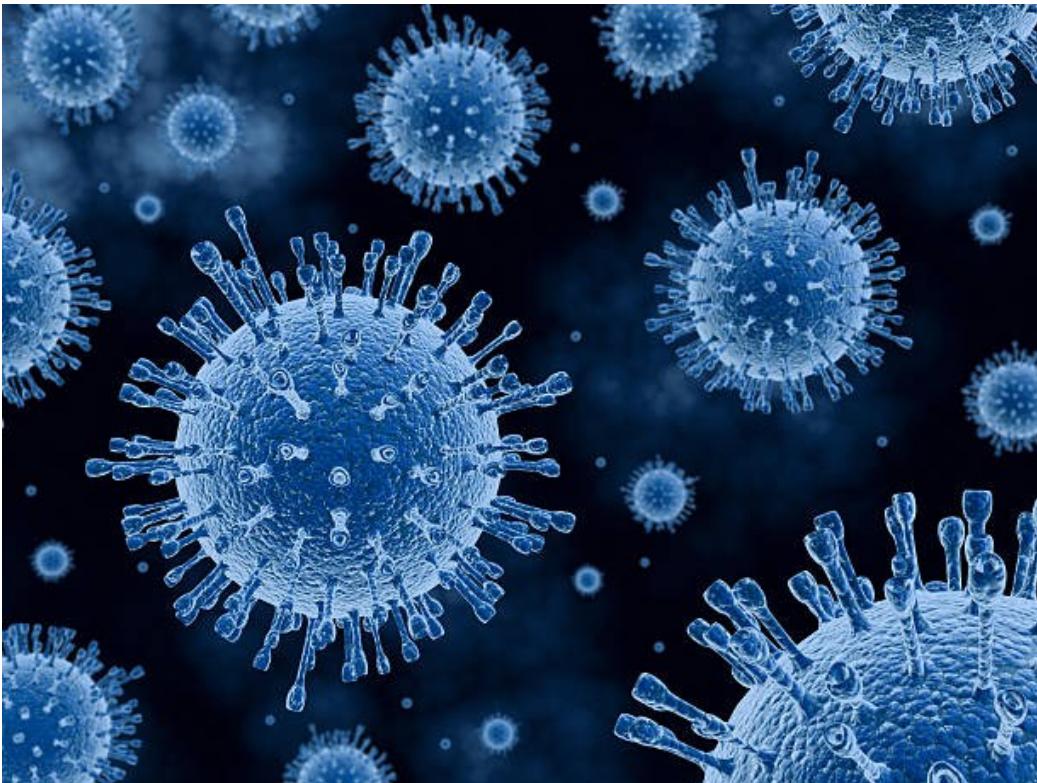


Goal:

To use influenza trends to provide insights to help medical staffing agency that provides temporary workers to clinics and hospitals on an as needed basis.

Objectives:

1. Analyze state-level patterns in flu-related deaths (2009–2017).
2. Focus on deaths among vulnerable populations (65+).
3. Use visualizations to guide staffing and resource allocation.



Data and Methodology

Data:

- Influenza deaths by geography sourced from U.S Centers for Disease Control and Prevention.
- Population data by geography, time, age, and gender Sourced from US Census Bureau.

About the data sets:

- Influenza Mortality data provides a detailed count of influenza-related deaths by state and age group in the United State, spanning the years 2009 to 2017.
- Population data provides a detailed breakdown of the population count across defined age group, as well as by states and counties in the United State, covering the period from 2009 to 2017.

Methodology:

Data Cleaning:

Performed comprehensive data cleaning on both datasets, profiling and integrity checks to ensure consistency and accuracy. This included renaming and formatting variables, validating data types, and addressing missing values. Conducted data quality assessments to confirm completeness and uniqueness, resulting in a reliable and well-structured dataset for analysis.

Data processing:

Integrated flu mortality data with state population data to build a unified, analysis-ready dataset.

Conducted statistical hypothesis testing to assess the relationship between the vulnerable population (age 65+) and flu related deaths across states. Developed a range of Tableau visualizations including temporal trends, forecasts, statistical plots, and spatial maps to uncover key patterns, highlight geographic disparities, and support data-driven insights.

Influenza Season

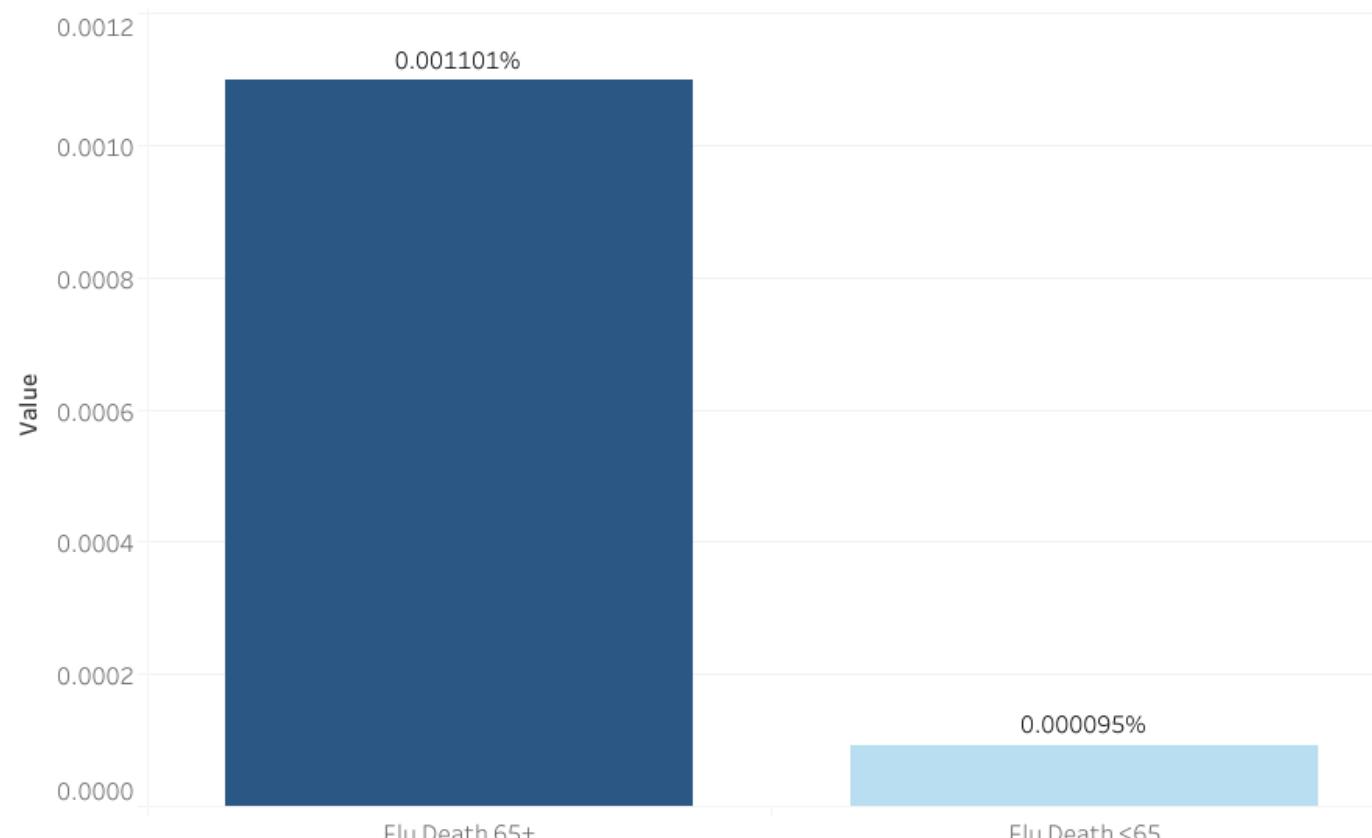


Data Exploration

Older Adults Face the Highest Risk of Flu-Related Deaths

Age plays a major role in this analysis, and the chart clearly highlights the difference in flu deaths between age groups. People aged 65 and older die from the flu at a much higher rate (0.001101%) than those under 65 (0.000095%). This big gap means older adults 65+ are over 10 times more likely to die from influenza compared to younger individuals.

Because of this, it is important to focus more on protecting the 65+ population through vaccinations, early treatment, and making sure they have the right healthcare support during flu season.



[Click link to view Tableau Dashboard](#)

Influenza Season



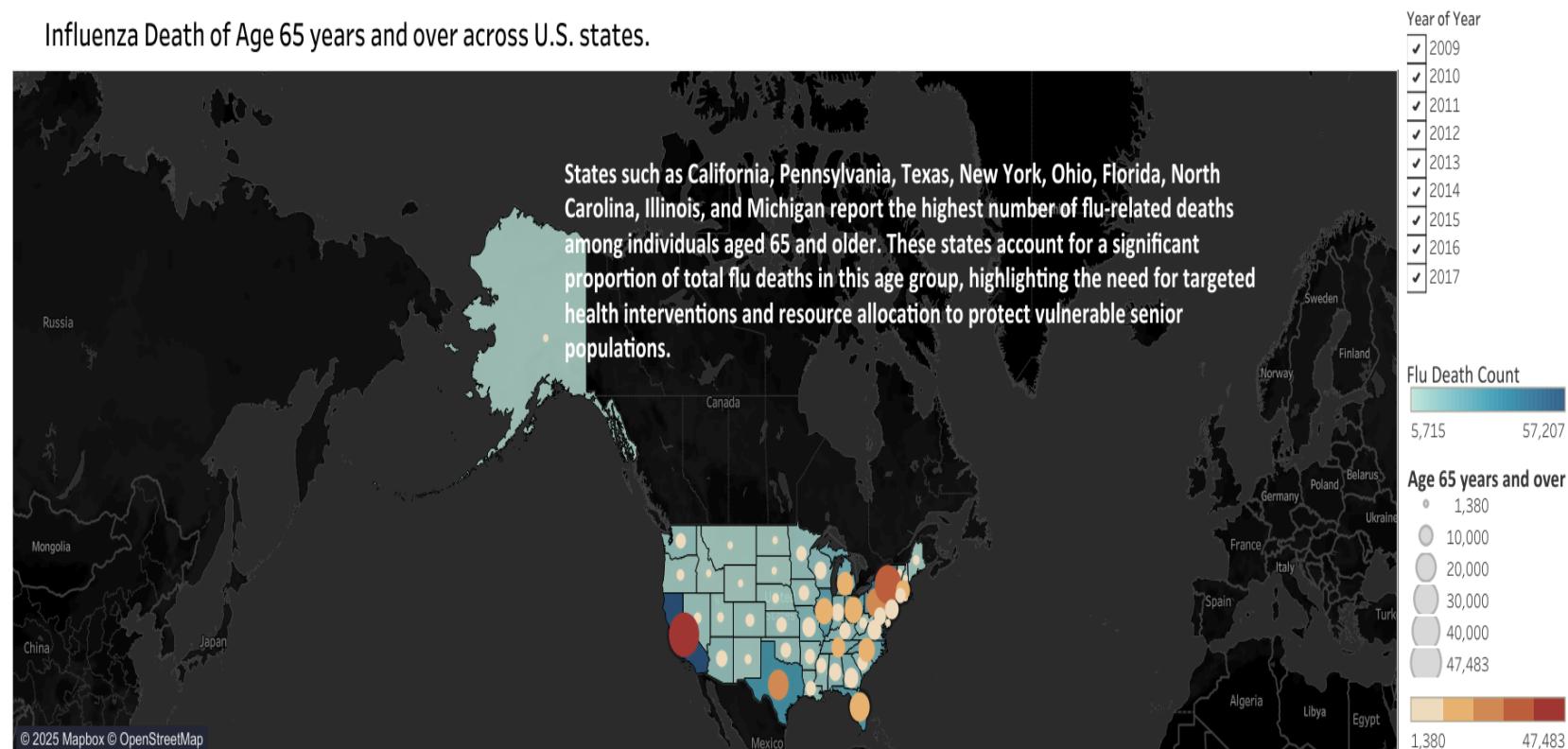
Data Exploration

Identifying High-Risk States for Senior Flu Deaths

The map highlights clear geographic patterns in flu-related deaths among individuals aged 65 and older across the United States. States such as **California, Pennsylvania, Texas, New York, Ohio, Florida, North Carolina, Illinois, and Michigan** report the highest numbers of flu deaths within this age group. These states also represent some of the largest senior populations, contributing significantly to the overall flu mortality count.

The combination of high flu death totals and large 65+ populations in these regions underscores the need for targeted public health strategies, including enhanced vaccination programs, early treatment initiatives, and strategic resource allocation. This visualization emphasizes the importance of focusing prevention efforts in states with both large senior communities and elevated flu mortality levels to better protect vulnerable older adults.

Influenza Death of Age 65 years and over across U.S. states.



Click link to view Tableau Dashboard

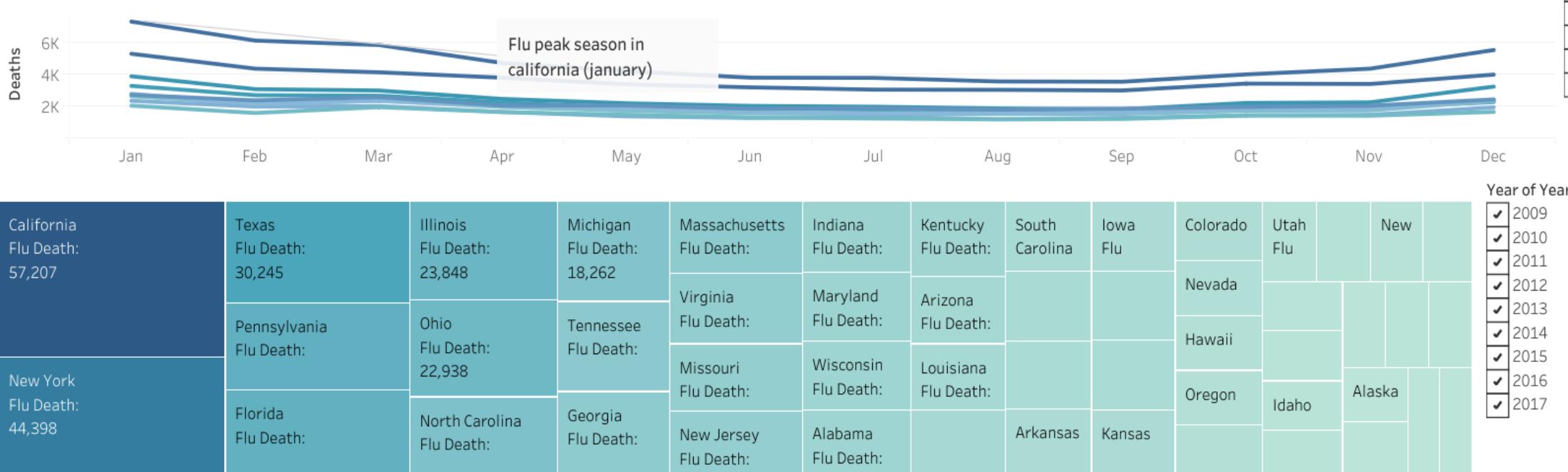
Influenza Season



Data Exploration

Influenza at its peak

Flu outbreaks impact different regions with varying intensity, with some states experiencing severe spikes that strain hospitals and clinics. The line chart and tree map linked together provide valuable insights into both the timing and severity of these outbreaks, helping to predict when flu cases are most likely to peak. As shown in the visualization, flu season typically reaches its highest levels during the first three months (January to March) and last two months (November to December) of the year. By analyzing these seasonal trends, healthcare systems can better anticipate demand and ensure clinics and hospitals are adequately staffed and resourced during peak periods. (line graph is connected to tree map to highlight and view flu trends)



Click link to view Tableau Dashboard

Influenza Season



Insights and Recommendations

Key Findings and Trends

- Flu-related deaths vary across states, but individuals aged 65 and older consistently face the highest risk nationwide.
- States such as California, New York, Pennsylvania, Texas, and Florida report notably higher mortality in this age group.
- Flu activity peaks during the winter and fall months, especially from January to March and October to December

Recommendations

- Deploy Additional Medical Staff to High-Risk
- Prioritize Elderly Patients (Aged 65+) in All States
- Enhance Public Awareness Campaigns
- Implement Seasonal Preparedness Protocols
- Monitor and Report Flu Trends in Real-Time.

Next Steps

- Conduct a Resource Assessment
- Develop a Deployment Plan
- Implement Monitoring and Evaluation Mechanisms
- Schedule Regular Progress Reviews



 [Click link to view Tableau Dashboard](#)

 [YouTube Video Presentation](#)

Project 3 - Rockbuster Stealth Analysis



Goal:

Rockbuster Stealth LLC, aims to transition into the online video rental market to remain competitive against streaming giants like Netflix and Amazon Prime. To support this shift, the management team seeks data-driven insights from its existing customer and rental data to guide strategic decisions for their business plan and ensure a successful digital launch.

Objectives:

- Use PostgreSQL for advanced database querying and extraction
- Create and use a data dictionary and entity-relationship diagram
- Analyze customer behavior, sales performance and rental behavior across the globe.



Rockbuster Stealth Analysis



Data and Methodology

Data:

Data set that contains information about Film Inventory, Customers Information, and Payment Transactions sourced from Rockbuster's internal Record.

Methodology:

Data Cleaning:

To prepare the Rockbuster Stealth dataset for analysis, all available data was first loaded into a relational database management system (RDBMS) to ensure proper storage, structure, and integration across tables. SQL was then used extensively to explore and clean the data. This included querying the database to identify inconsistencies, filtering out irrelevant or duplicate records, and handling missing or incorrect values.

Additional SQL operations were performed to summarize key variables, standardize formats, and ensure the dataset was accurate, well-organized, and ready for deeper analysis.

Data Processing:

During the Rockbuster Stealth analysis, several SQL-based data processing techniques were applied to transform and prepare the dataset for deeper insights.

DbVisualizer was used to create an entity relationship diagram (ERD), helping to clearly understand table connections and database structure. Data from multiple tables was combined using SQL joins to build complete and meaningful views of customer, film, rental, and payment information. Subqueries were employed to extract targeted metrics and perform more complex filtering. Common Table Expressions (CTEs) were also used to simplify multi-step transformations, improve query readability, and manage intermediate results efficiently. Together, these techniques ensured a smooth and structured data workflow throughout the analysis.

Rockbuster Stealth Analysis

Data Exploration

Descriptive Statistics

MINIMUM

Rental duration = 3days

Rental Rate = \$0.99

Film Length = 46mins

Replacement cost = \$9.99

AVERAGE

Rental duration = 5days

Rental Rate = \$2.98

Film Length = 115.27mins

Replacement cost = \$19.98

MAXIMUM

Rental duration = 7days

Rental Rate = \$4.99 Film

Length = 185mins

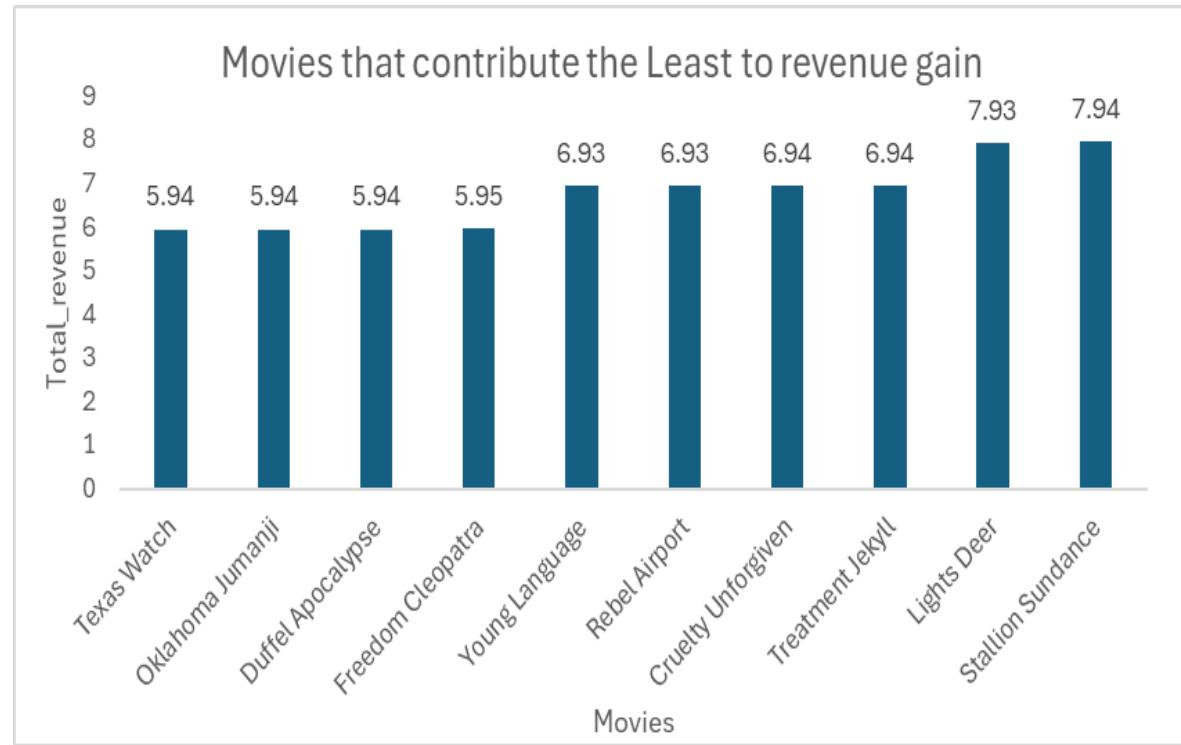
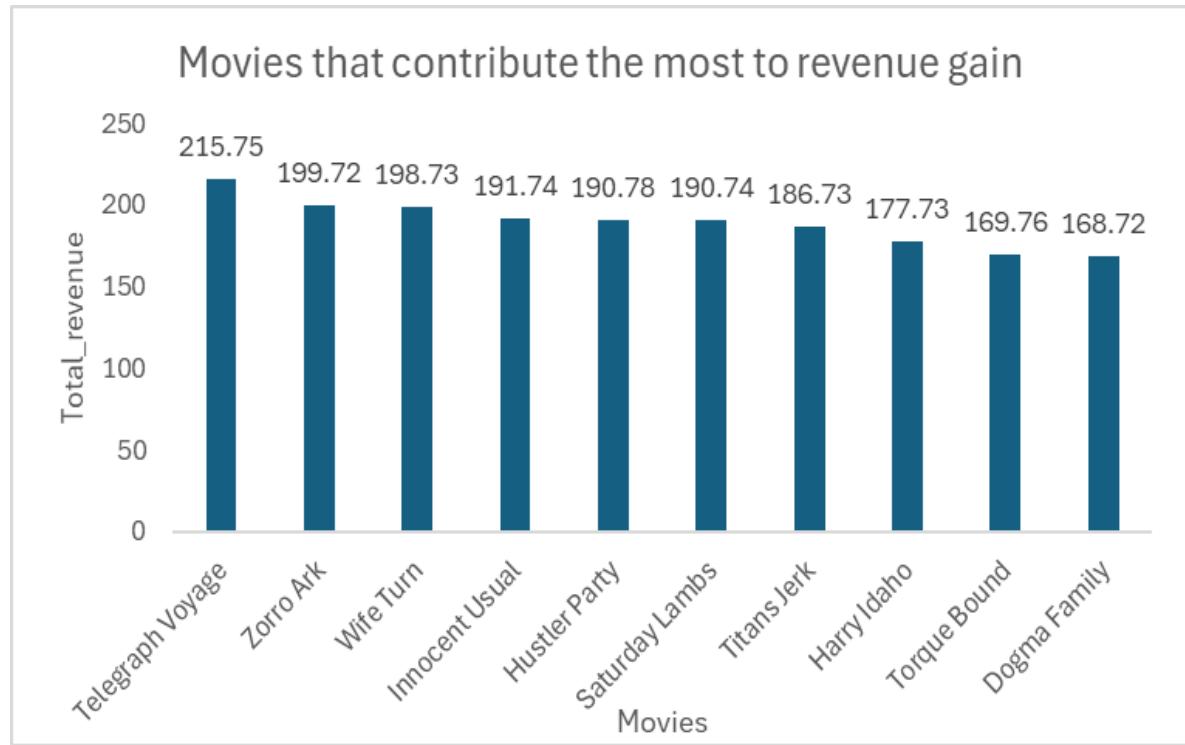
Replacement cost = \$29.99

- ❖ Descriptive statistics reveal that Rockbuster videos are rented for an average of 5 days at a typical rate of \$2.98. Films run about 115 minutes on average, and the standard replacement cost is approximately \$19.98.

Rockbuster Stealth Analysis



Revenue Contributions

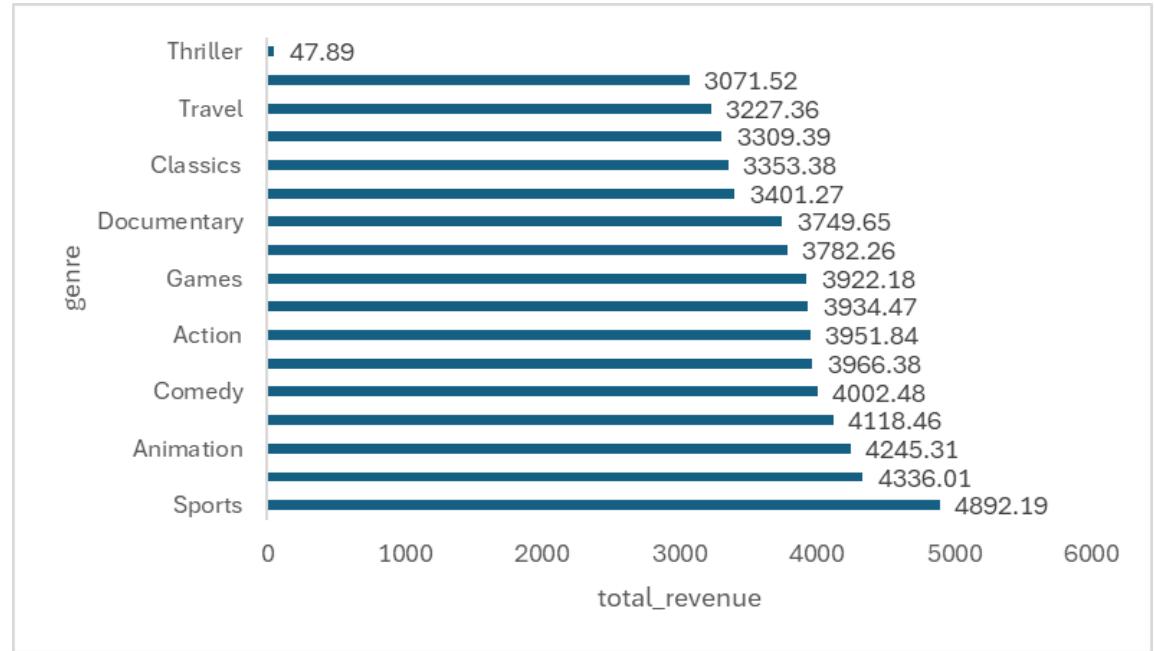
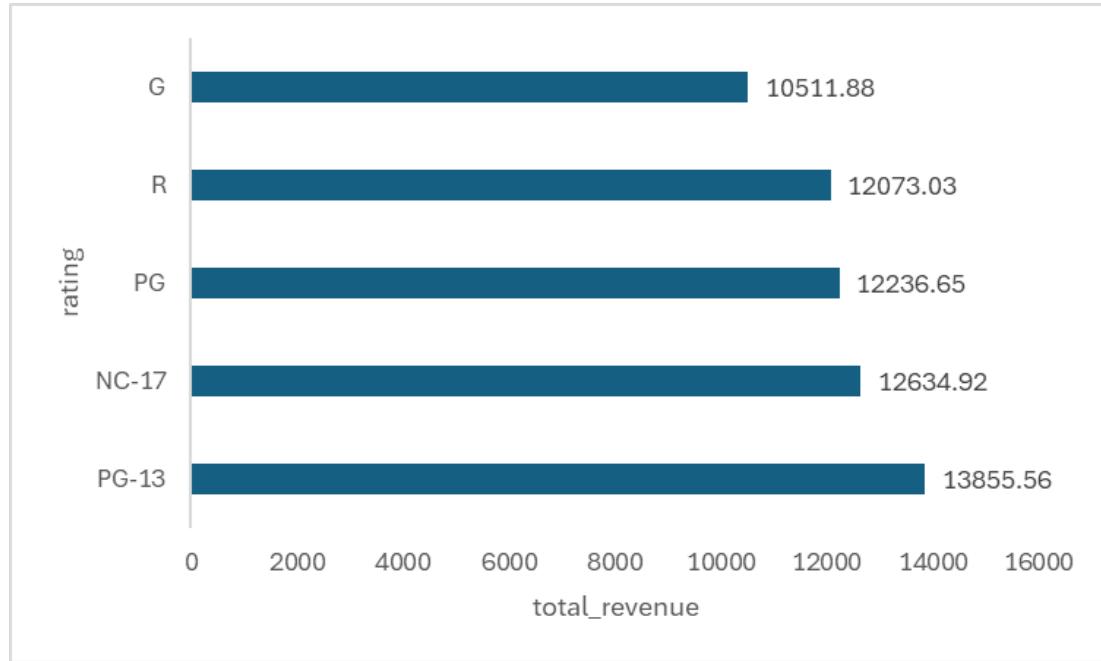


The analysis highlights a strong difference in revenue performance across Rockbuster's movie catalog. Top-performing films such as *Telegraph Voyage*, *Zorro Ark, and *Wife Turn* generate the highest total revenue, each earning between \$168 and \$216, indicating strong customer demand and consistent rental activity. In contrast, the lowest-earning titles including *Texas Watch*, *Oklahoma Jumanji*, and *Duffel Apocalypse* bring in significantly less, with revenues ranging from \$5.94 to \$7.94. This suggests limited rental interest and minimal financial impact on overall earnings. Overall, the charts show that a small group of high-performing titles contributes a substantial portion of Rockbuster's revenue, while many lower-performing films have minimal influence on total earnings. This insight can help guide future inventory decisions, marketing focus, and promotional strategies.*

Rockbuster Stealth Analysis



Revenue by Rating and Genre



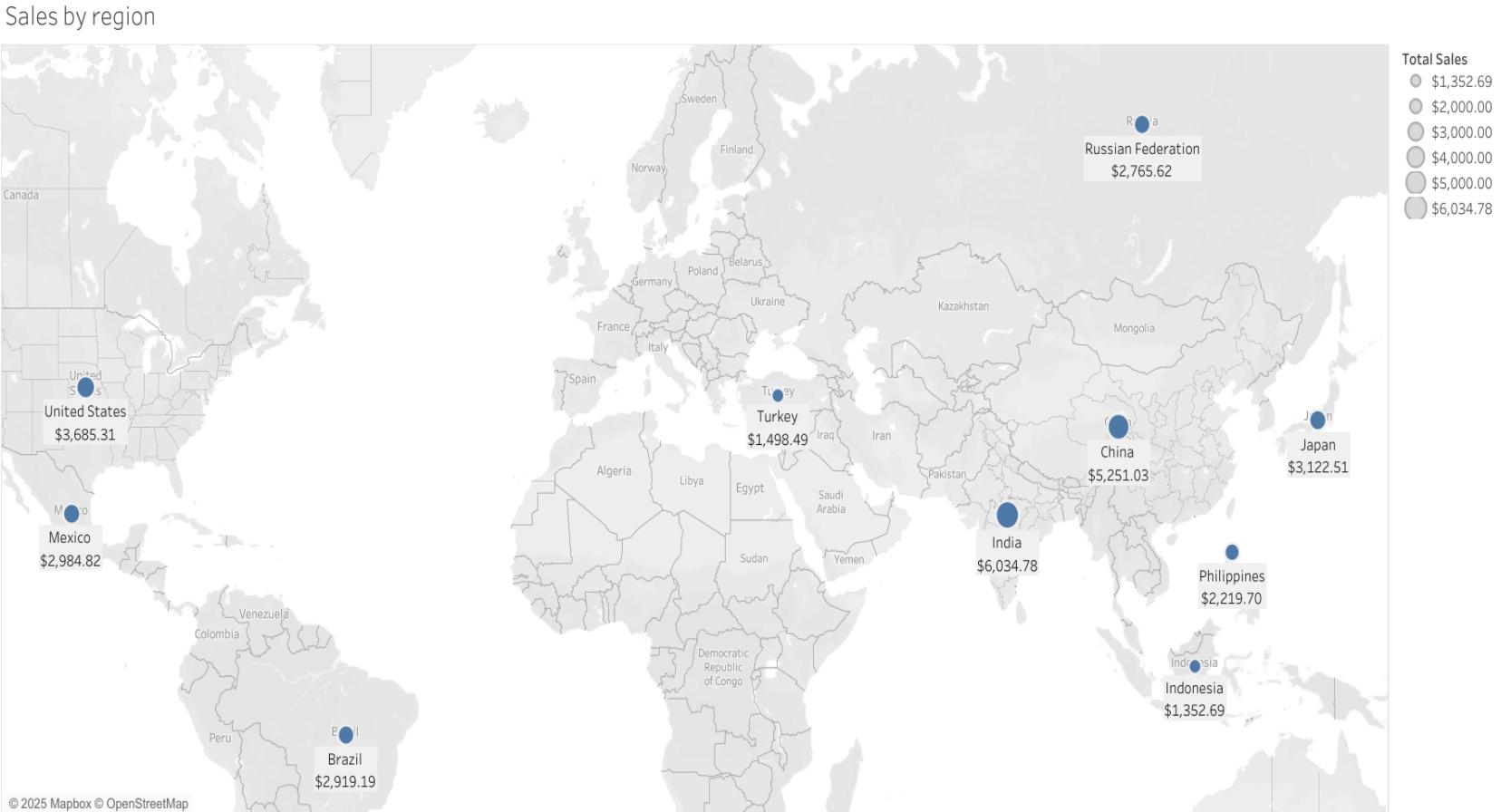
- ❖ The bar chart shows that PG-13 films and Sports games generated the highest revenues in their categories, while G-rated films and Thriller titles recorded the lowest

Rockbuster Stealth Analysis



Regional Differences in Sales Performance

- ❖ Sales vary widely by region. Asia leads with India, China, and Japan, while the U.S. tops the Americas and Brazil follows. Europe and nearby regions, including Russia and Turkey, show lower sales.



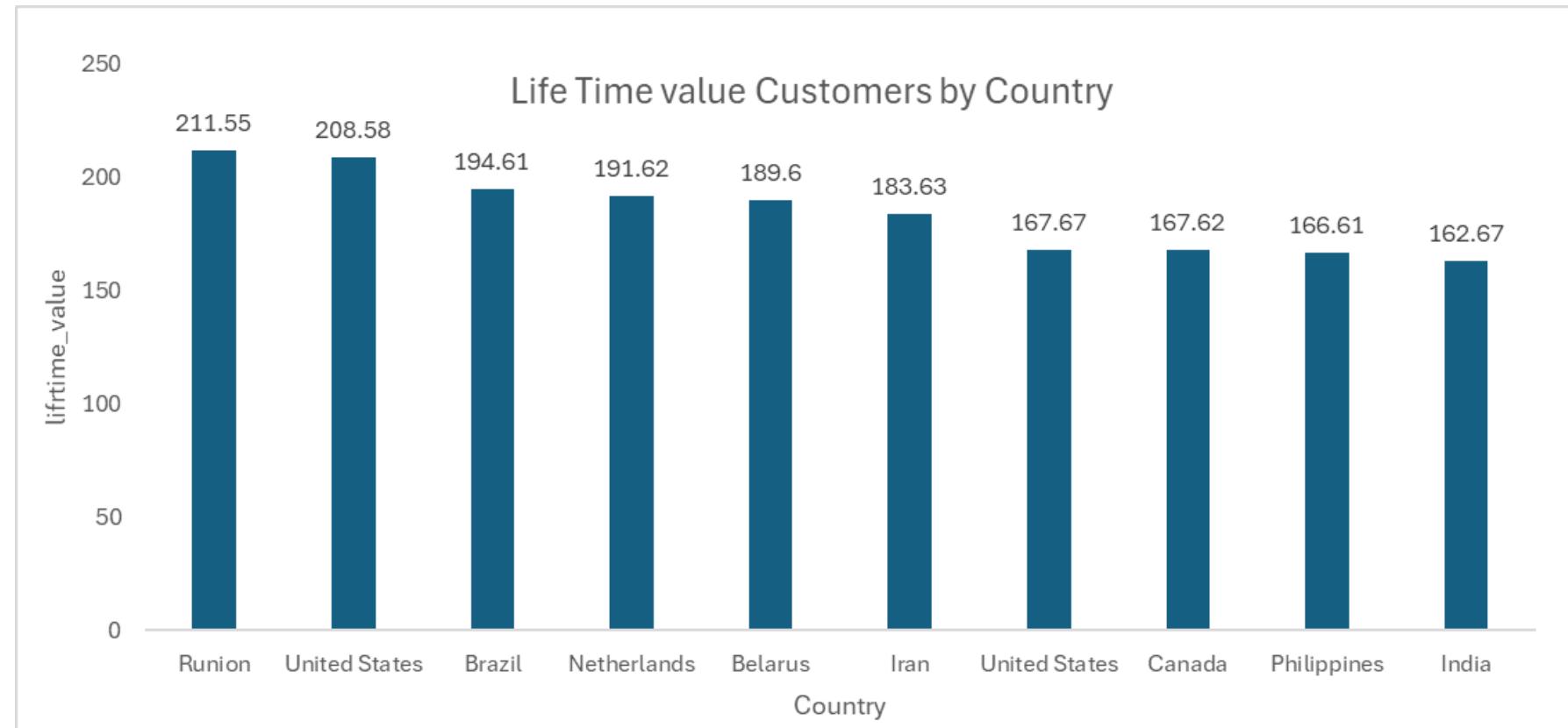
[View Tableau Dashboard](#)

Rockbuster Stealth Analysis



Geographic Distribution of High-Lifetime-Value Customers

- ❖ The lifetime value analysis reveals which customers have generated the highest total revenue over time, helping to identify Rockbuster's most valuable customers and their geographic distribution.



Rockbuster Stealth Analysis



Insights and Recommendations

Key insights

- **Revenue & Content:** PG-13 films and Sports titles, especially Telegraph Voyage, drive the most revenue, while some older or less popular movies earn very little.
- **Customer Base:** Rockbuster's strongest markets are in Asia and Latin America, with India and China leading.
- **Regional Sales:** Asia leads overall revenue, followed by North America; Europe lags behind.
- **High-Value Customers:** Regions like Réunion, the U.S., Brazil, and the Netherlands have highly engaged, high-spending customers.

Recommendations

- **Content Strategy:** Prioritize PG-13 and Sports content; phase out or repackaging underperforming titles.
- **Regional Marketing:** Focus on high-revenue markets (India, China, U.S., Japan) with aggressive digital campaigns; target emerging markets (Philippines, Turkey, Indonesia) with low-cost models and localized content.
- **Customer Retention:** Engage high-value customers via loyalty programs, premium tiers, and personalized recommendations.
- **Budget Transition:** Shift funds from low-traffic physical stores to digital infrastructure.

Project 4 – Instacart Grocery Basket Analysis



Goal:

Instacart, an app-based online grocery platform, already enjoys strong sales but aims to analyze its sales data more deeply to uncover patterns and provide insights for improved customer segmentation and strategic decision-making.



Objectives:

- Leverage Python and key libraries (Pandas, Polars, NumPy, Matplotlib, Seaborn, SciPy) to analyze large-scale datasets.
- Conduct exploratory data analysis (EDA) on Instacart's sales and customer data.
- Identify trends and patterns in product departments and individual product performance.
- Generate actionable insights to support sales optimization and targeted marketing strategies



Data and Methodology

Data:

- Instacart Online Grocery Shopping Dataset (2017) sourced from Kaggle.
- Customer Dataset provided by (Career Foundry) specifically for analysis conducted in this project.

Methodology:

Data cleaning:

All Instacart datasets were imported into Python and prepared for analysis through careful data cleaning. This included wrangling and sub-setting to focus on relevant data, as well as performing consistency checks to correct duplicates, handle missing values, and standardize formats. These steps ensured the datasets were accurate, reliable, and ready for insightful analysis.

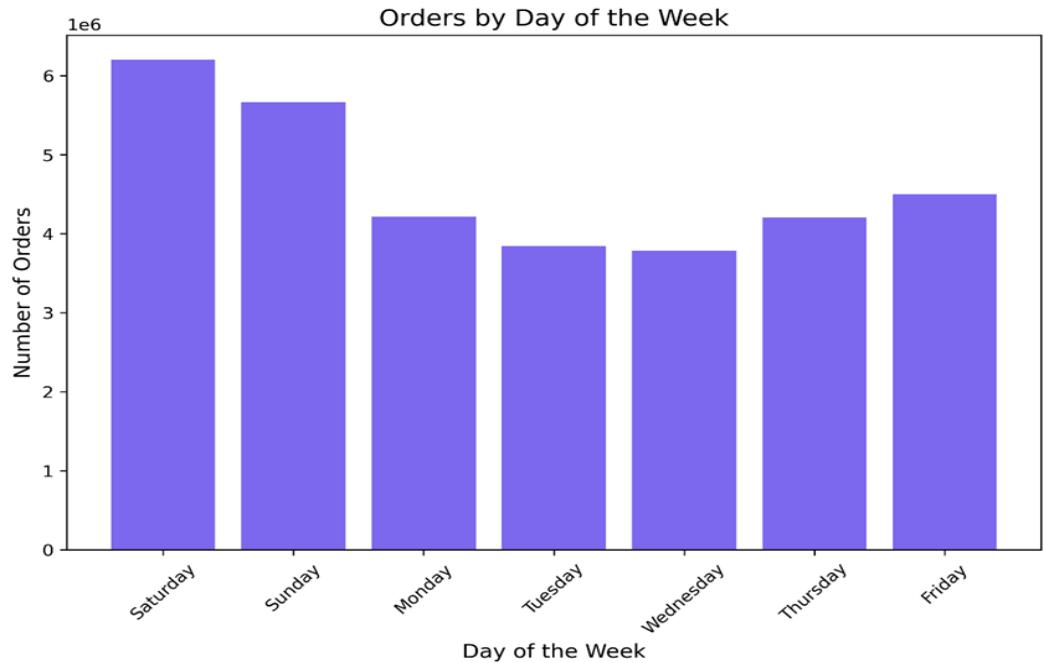
Data processing:

All Instacart datasets were processed to generate meaningful insights. This included combining datasets for a comprehensive view, deriving new variables such as `loyalty_flag` and `busiest_period_of_day`, grouping and aggregating data to summarize trends across products, customers, and departments.

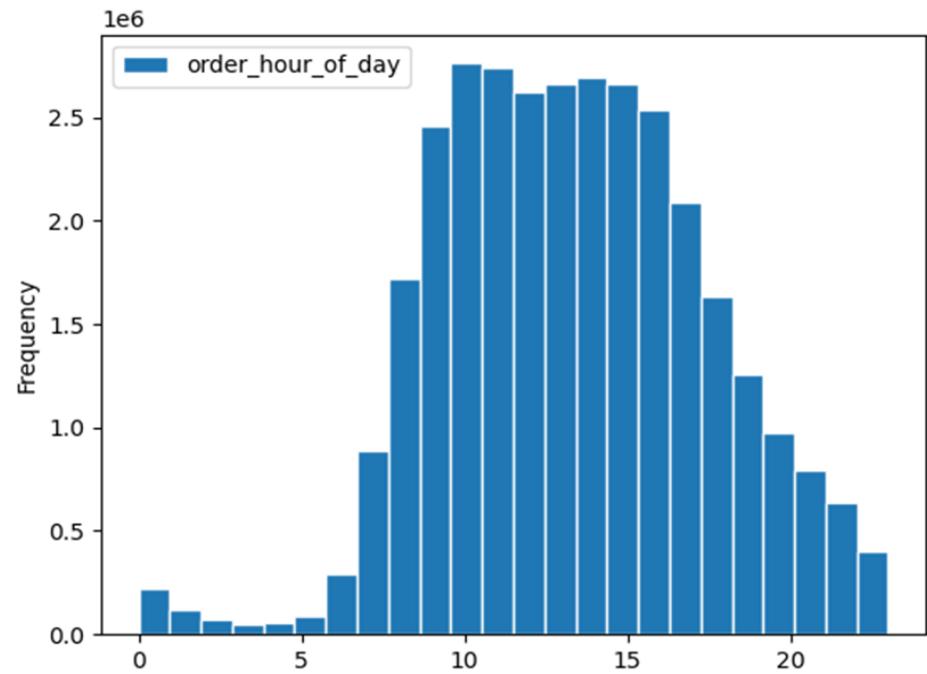
Data Exploration

Key Question

- Which days of the week and hours of the day are the busiest?



- ❖ Saturday, Sunday and Friday are the busiest days of the week, with Saturday showing the highest order volume. while Tuesday and Wednesday has the least order activities.
- This indicates that customers are weekends shoppers and are most active in the late morning, likely placing orders for groceries to be delivered or picked up for the day or week ahead.

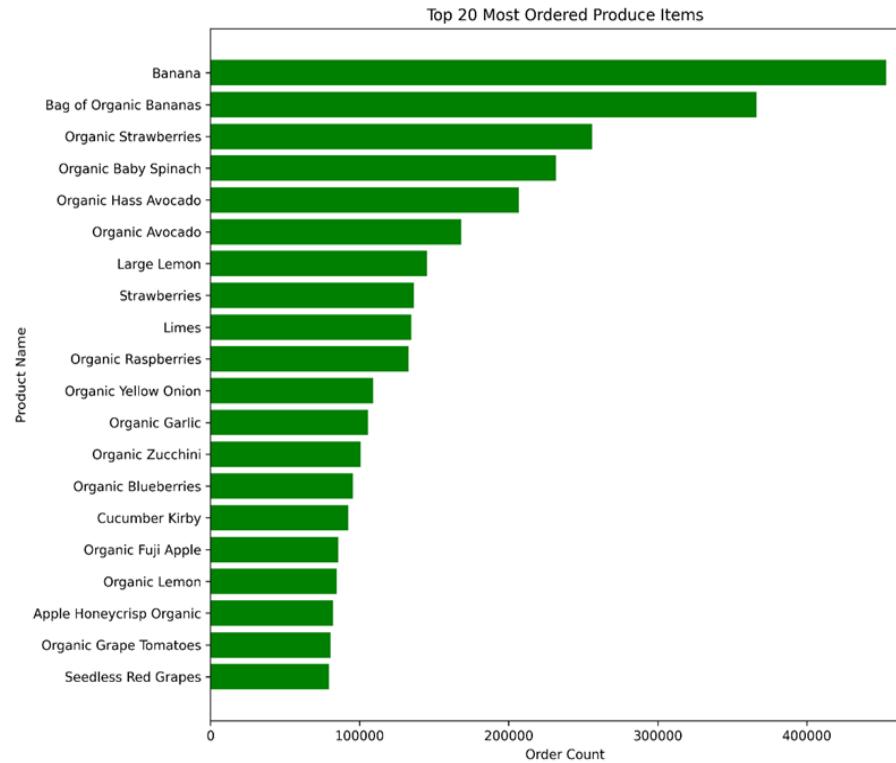
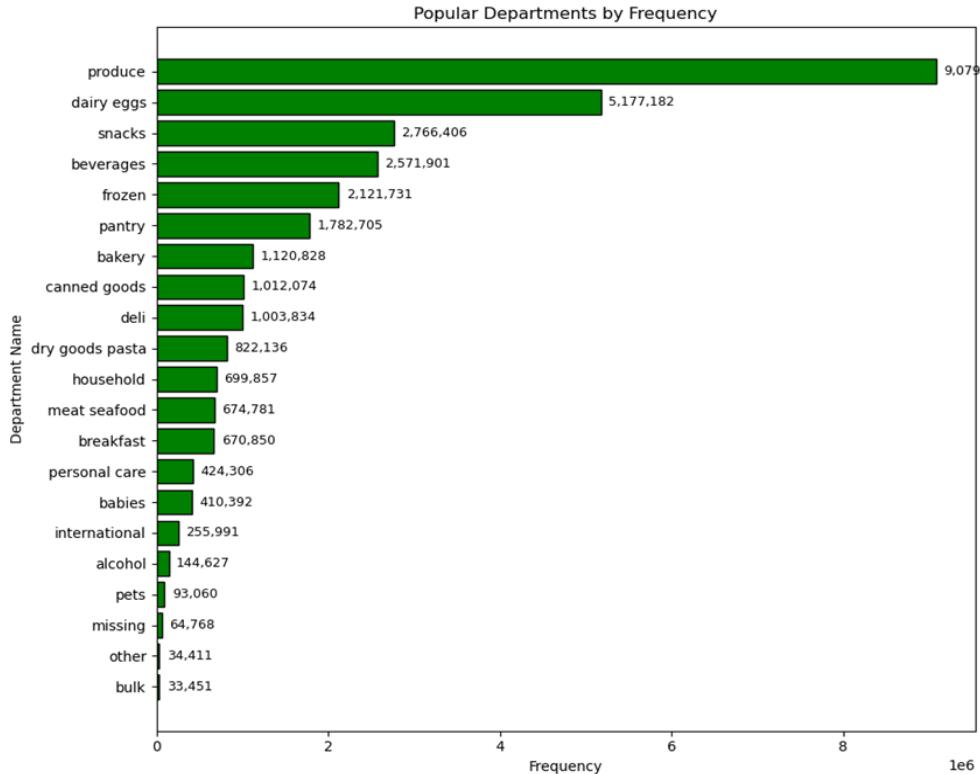


- ❖ The histogram shows that order activity is minimal from midnight until about 5 a.m. It begins to rise sharply after 6 a.m., increasing strongly between 8 and 10 a.m. Order counts remain consistently high from roughly 10 a.m. to 3 p.m., then decline noticeably after 5–6 p.m. Overall, the data reveal that customers place most of their orders in the late morning and early afternoon, with very little activity during the night.

Data Exploration

Key Question

- Are there certain types of products that are more popular than others



❖ Produce department has the highest frequency of product ordered, followed by dairy eggs , snacks, beverages and frozen department.

➤ This suggest that customers likes to shop fresh produce items.

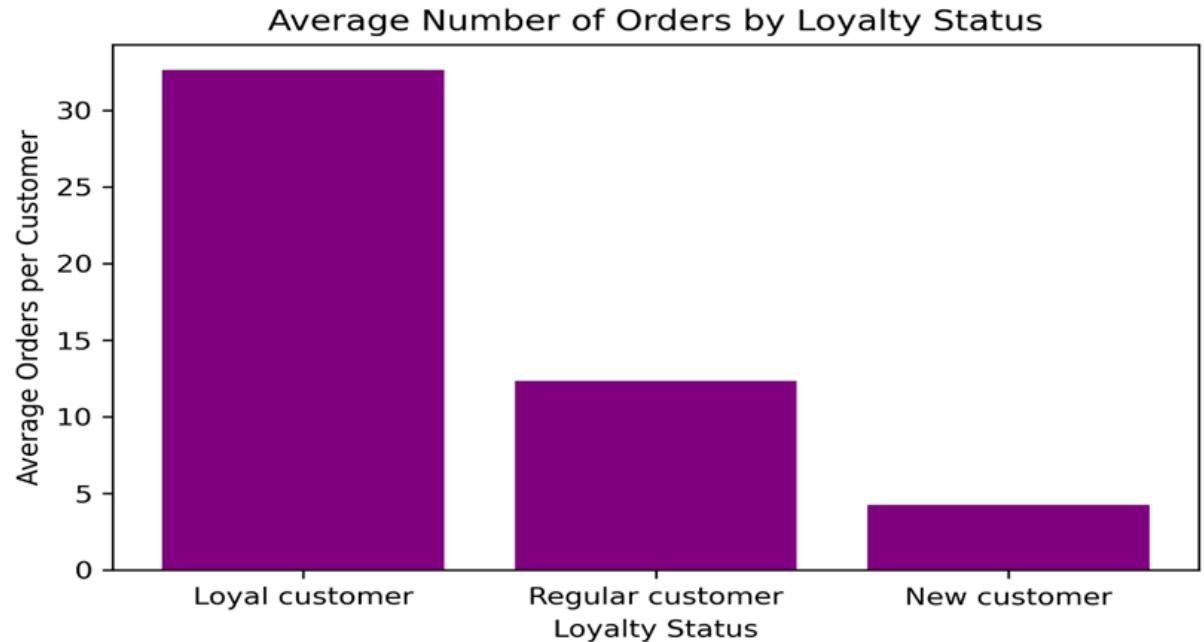
❖ Bananas (regular and organic), organic strawberries, and organic baby spinach are among the top produce items.

Data Exploration

Key Question

- Are there differences in ordering habits based on a customer's loyalty status?

loyalty_flag	avg_orders_per_customer
"Loyal customer"	32.641196
"Regular customer"	12.347452
"New customer"	4.268295

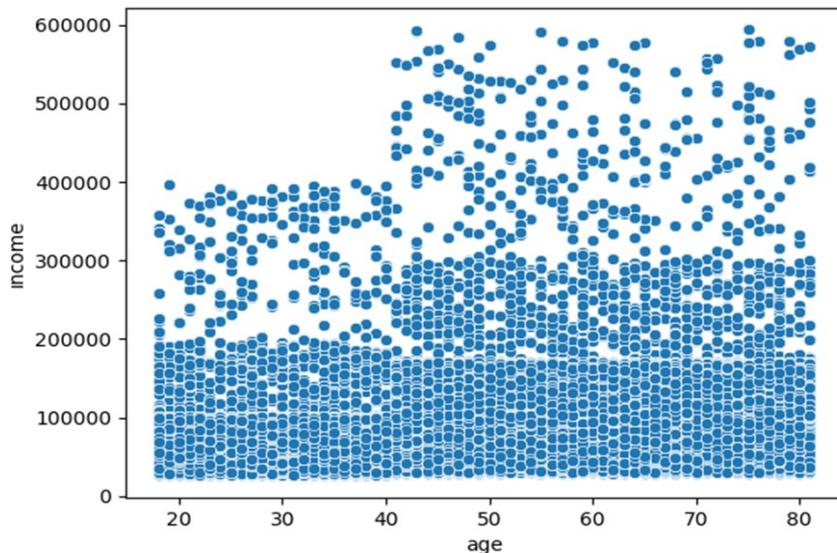


- ❖ **Loyal customers** places an average of about 33 orders, showing a high level of engagement and consistent purchasing behavior.
- ❖ **Regular customers** group shows moderate engagement with orders average of 12, likely representing periodic or seasonal shoppers.
- ❖ **New customers** have the lowest order average With roughly 4 orders, representing group in the early stage of their customer lifecycle

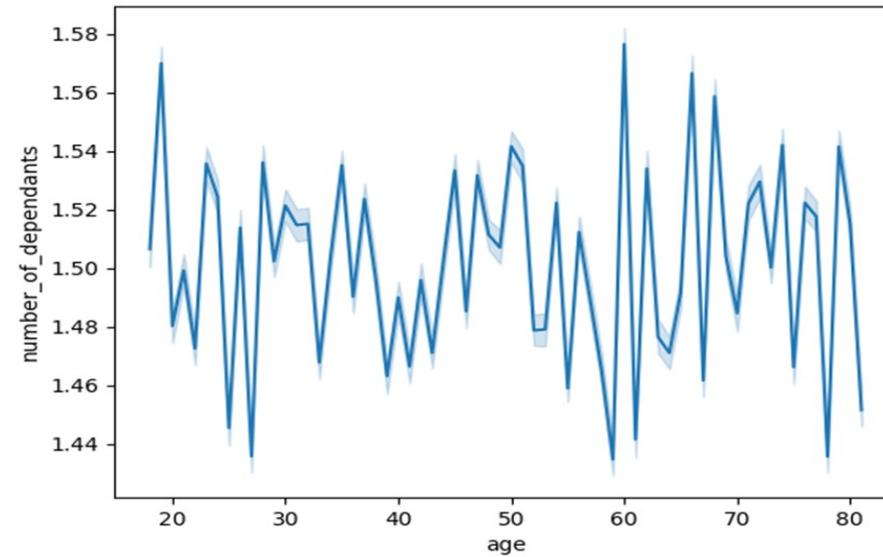
Data Exploration

Key Question

- Is there a connection between income and age?
- Is there a connection between age and family status?



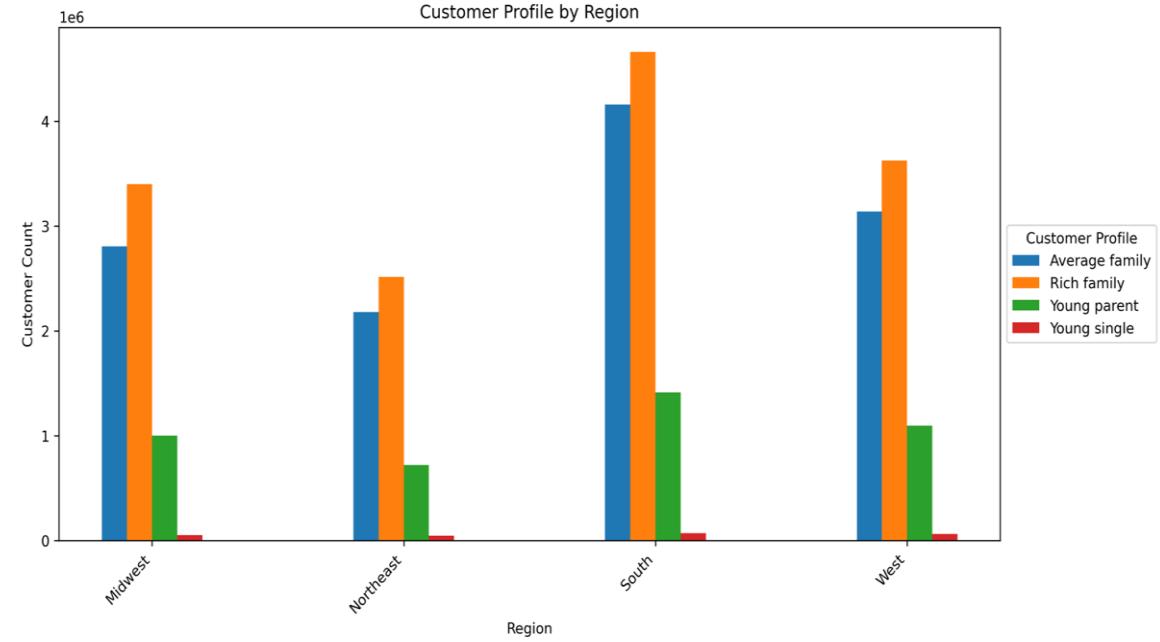
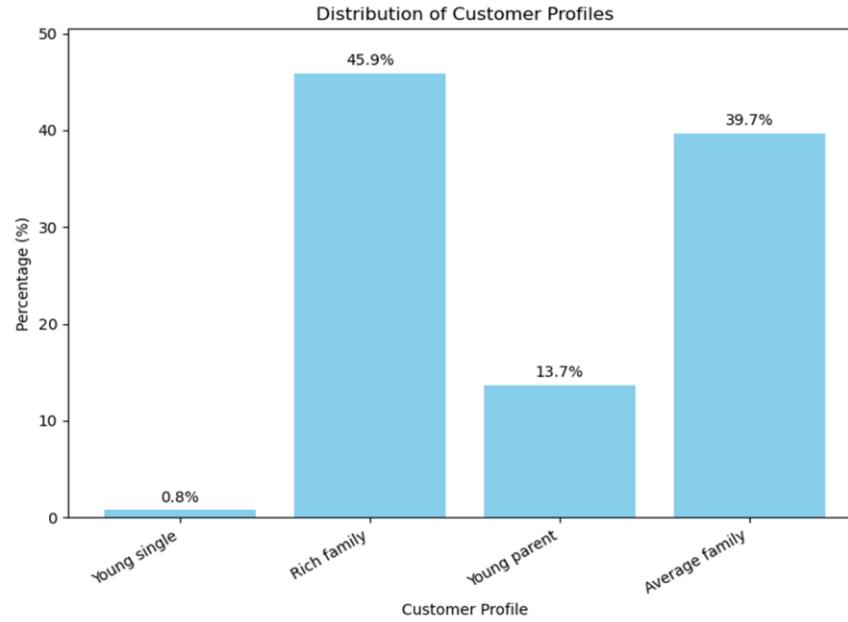
❖ The scatterplot displays age on the x-axis and income on the y-axis. Points are widely scattered with no clear upward or downward trend, suggesting that income is not strongly related to age in this dataset. In conclusion, the plot suggests income varies greatly at every age and age alone doesn't predict higher or lower earnings in this sample.



❖ The chart shows age along the x-axis and number of dependants along the y-axis. Across all age groups, from the late teens to the 80s, the values stay close to about 1.5 dependants, indicating that people of different ages generally have one to two dependants. The pattern remains fairly consistent with no obvious upward or downward trend, suggesting that age doesn't have a strong relationship with the number of dependants in this dataset.

Data Exploration

Comparing Distribution of Customer Profile to regions



- ❖ The customer base is dominated by two major segments: Rich Families (46%) and Average Families (40%). Young Parents represent about 14%, while Young Singles make up less than 1%.
- **This indicates that Instacart's market is largely family-oriented, making these households the most influential in shaping sales trends and product preferences**
- ❖ The South has the highest number of customers, followed by the Midwest and the West. Across all regions, Rich Families and Average Families consistently remain the dominant segments, while Young Singles account for a very small share.
- **This uniform pattern indicates a stable national trend, showing that Instacart's customer base is largely family-driven rather than region-specific.**

Recommendations

- The sales team should schedule digital ads and targeted promotions during lower-traffic periods, particularly early mornings (6–9 AM), evenings (7–10 PM), and mid-weekdays such as Tuesday and Wednesday. Running campaigns outside peak ordering hours increases visibility, reduces competition for customer attention, and improves the likelihood that shoppers will engage with promotional content before placing their orders.
- Prioritize produce in marketing and homepage placement, ensure consistent availability of top-selling fresh items, and bundle popular produce such as bananas, organic strawberries, and baby spinach with complementary products from other departments to boost cross-sales and enhance customer satisfaction.
- Allocate more marketing resources to the South region, which shows higher engagement and purchasing activity, especially among rich and average families.
- Strengthen supply chain for produce and dairy categories, these are consistently top performers across all demographics. Consider regional warehouses in the South to handle the higher customer load efficiently
- Strengthen customer retention by converting new shoppers through personalized offers and first-time incentives, while maintaining loyal customers with exclusive deals and appreciation campaigns. Use insights on frequently purchased products to tailor targeted marketing for each customer group.

Gun violence in the U.S.A (2013-2018)



 [Click to view tableau dashboard](#)
[click to view on GitHub](#)

Goals:

Gun violence is a critical public health issue in the United States, causing tens of thousands of deaths and injuries each year. This recurring problem has always been close to my heart. With the U.S. experiencing a much higher rate of gun-related deaths compared to other high-income countries, I am deeply concerned as a citizen advocating for a safer, non-violent society. As a father who wants a secure environment for his children and as an analyst passionate about using data to address real-world problems, I was motivated to investigate this topic to uncover actionable insights and support the development of potential solutions.

Objectives:

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



Gun violence in the U.S.A (2013-2018)



Data and Methodology

 [Click to view tableau dashboard](#)

[click to view on GitHub](#)

Data:

Gun violence dataset contains over 260,000 recorded gun violence incidents spanning January 2013 to March 2018. Each entry includes up to 29 variables such as incident date, city and state, geographic coordinates, number of fatalities and injuries, weapon details, participant demographics, and links to original source reports.

Methodology:

Data cleaning:

The dataset initially contained 239,677 rows and 29 variables. Data cleaning began with validating and correcting variable data types, followed by a systematic assessment of missing values. Variables with excessive missing data (>70%) or low analytical relevance (e.g., participant details and URL fields) were removed, along with highly granular legislative district fields that were difficult to standardize accurately. Variables that were analytically useful but incomplete (e.g., address, gun type, and gun stolen status) were retained, with missing categorical values filled as “**Unknown**” and numerical values imputed using the **median**. A final duplicate check confirmed data integrity, resulting in a clean and analysis ready dataset.

Data processing:

The gun violence dataset was systematically processed and analyzed to generate meaningful insights through multiple analytical approaches. This included the creation of geographic visualizations to examine spatial distribution across states, time-series analysis to uncover temporal trends and structural shifts, and both supervised and unsupervised machine learning techniques to identify patterns and classify incident types. These methods enabled a deeper understanding of trends, variability, and relationships within the data beyond simple descriptive statistics.

Gun violence in the U.S.A (2013-2018)



Data Exploration

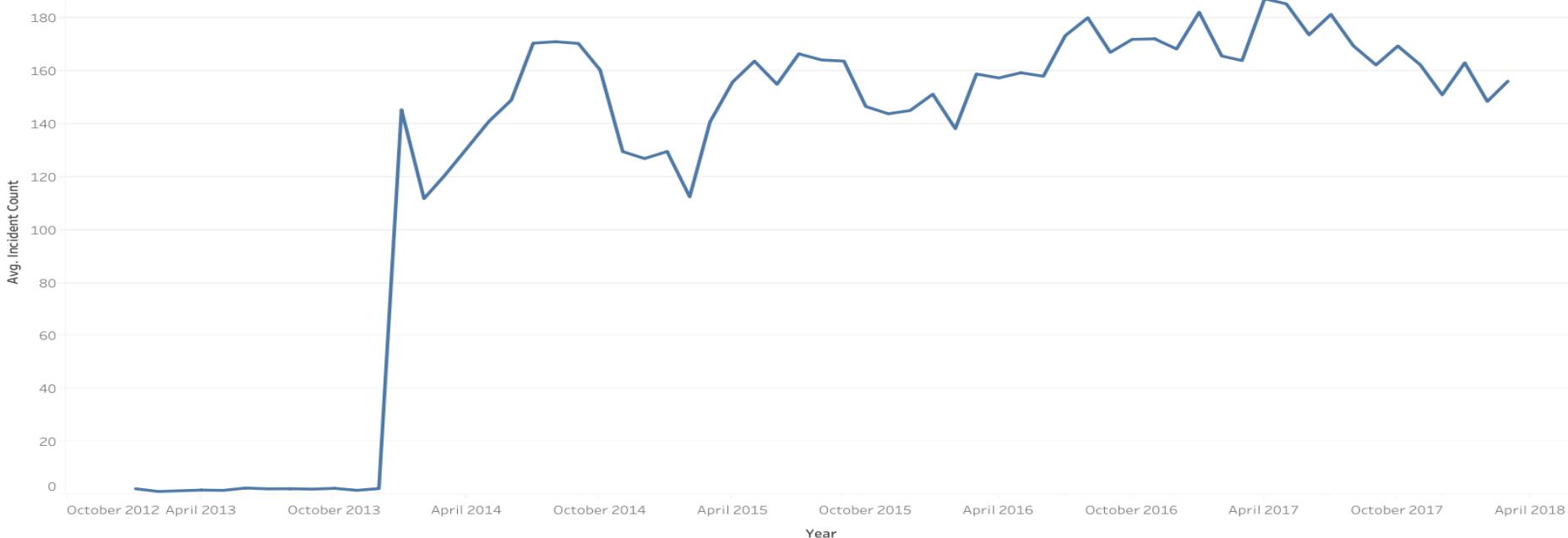
Key Question:

- How have gun violence incidents evolved over time?

[Click to view tableau dashboard](#)

[click to view on GitHub](#)

Gun Evolution Trend from 2013 - 2018



The time-series analysis reveals a clear structural shift in gun violence beginning around 2013–2014, marked by a sharp increase following a previously flat period. Rather than a gradual rise, gun violence moved to a higher baseline level that persisted over subsequent years. From 2014 onward, incident counts became more frequent and volatile, reaching their highest levels between 2016 and 2017, which suggests the influence of sustained and systemic drivers rather than short-term fluctuations. Although there is a slight decline by 2018, incident levels remain well above early-2013 values, indicating that gun violence has become a persistent and elevated public safety challenge rather than a temporary anomaly.

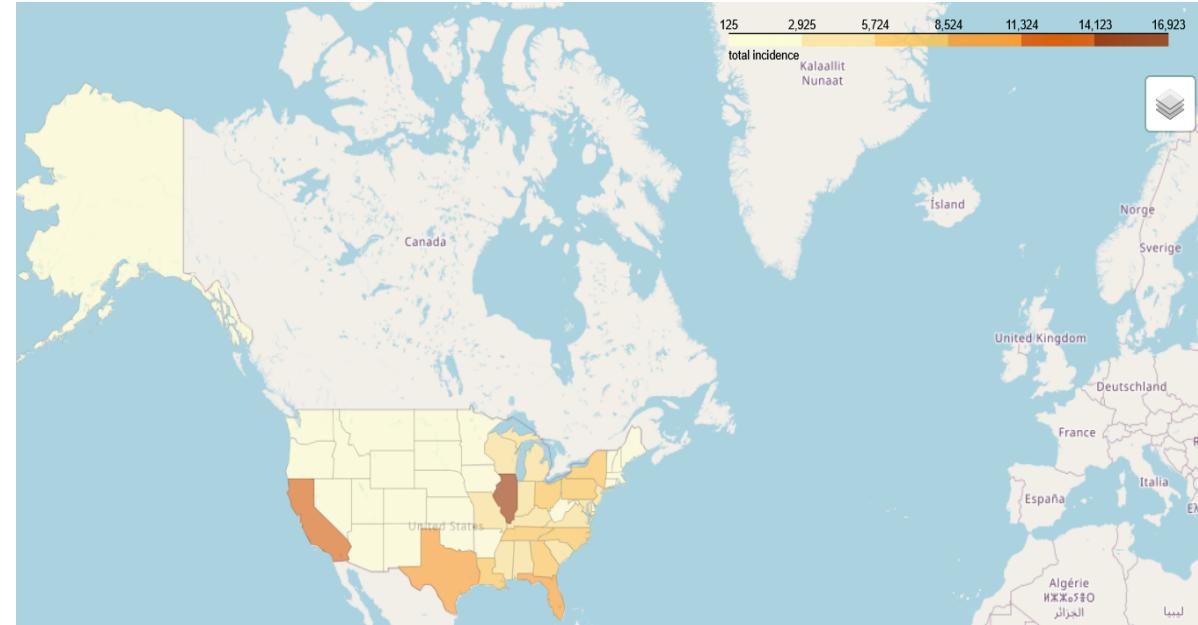
Gun violence in the U.S.A (2013-2018)



Data Exploration

Key Question:

- Which states have the highest number of incidents?



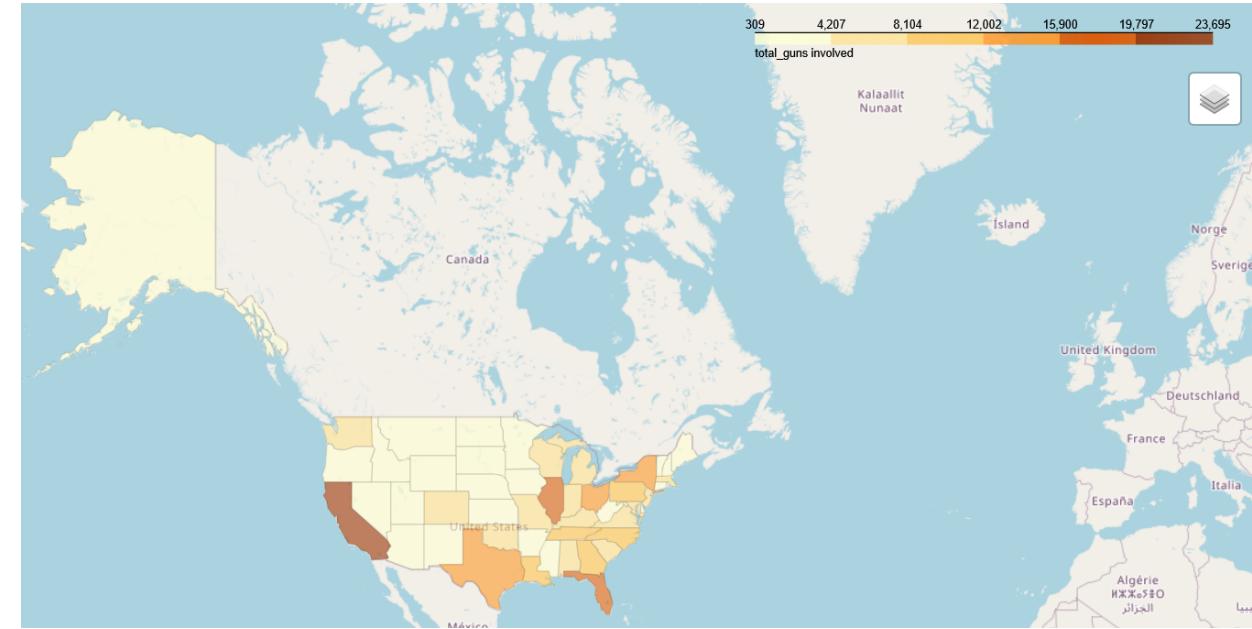
The choropleth map shows that gun violence incidents are heavily concentrated in a few states.

California, Illinois, Texas, and Florida stand out with the highest total number of victims (killed + injured).

Midwestern and Southern states also show moderately high incident levels, while many **Western and Northeastern states** display lower totals.

- Overall, the map highlights clear geographic disparities, with a small group of states accounting for a large share of national gun-related harm.

[Click to view tableau dashboard](#)
[click to view on GitHub](#)



The map is a choropleth visualization showing the total number of guns involved in gun-violence incidents across U.S. states.

Each state is shaded based on how many guns were recorded in these incidents.

Illinois shows the highest number of guns involved, appearing in the darkest shade.

Florida and California also report very high totals, following behind Illinois.

Several states including **Georgia, Ohio, North Carolina, Missouri, and Pennsylvania** fall into the moderately high range.

Most other states show lighter shading, indicating lower numbers of firearms involved in incidents.

- Overall, the map highlights significant regional variation, with a few states accounting for a disproportionately large share of guns used in violent events.

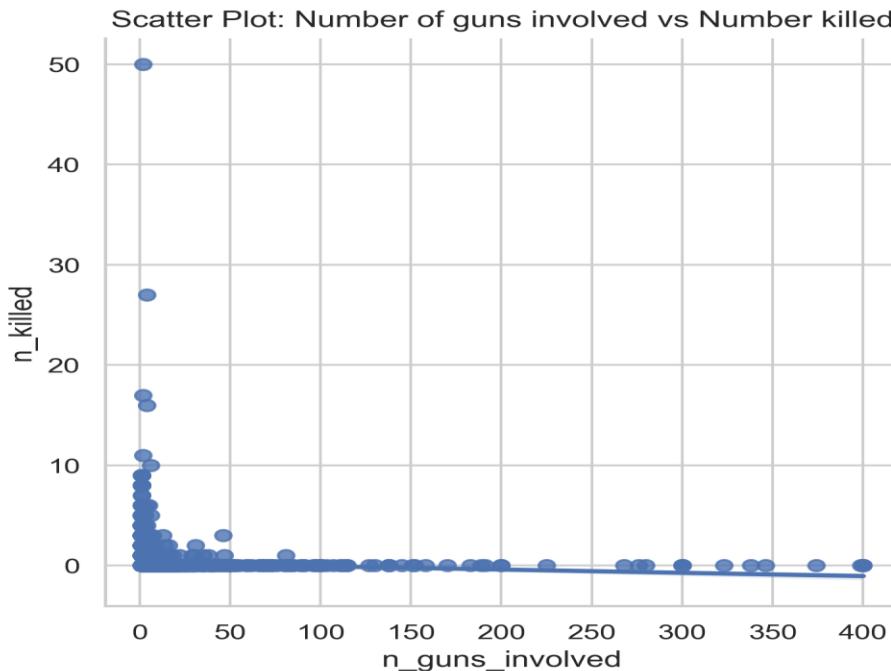
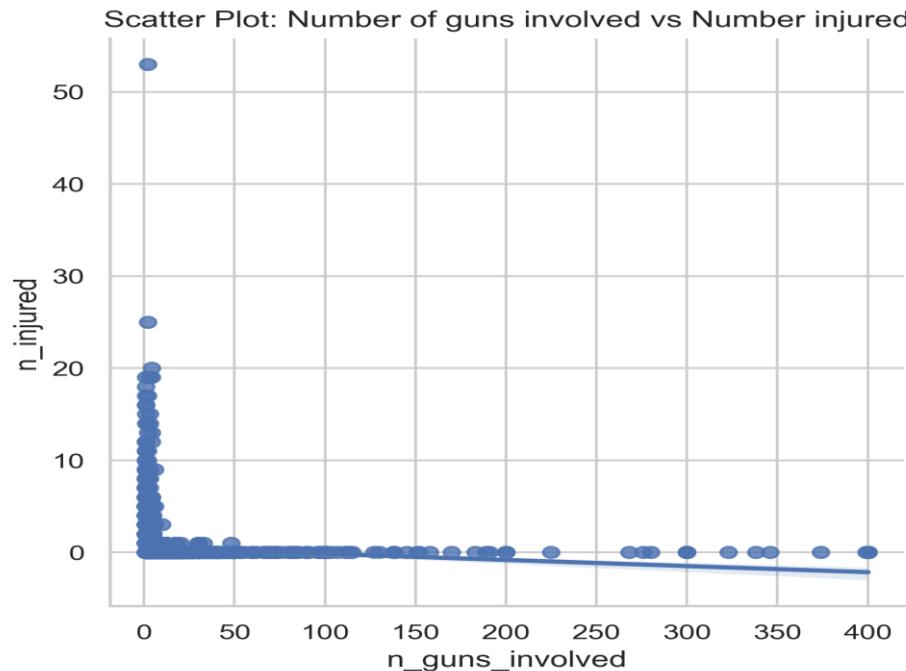
Gun violence in the U.S.A (2013-2018)

Data Exploration

Expectation

- The initial assumption was that incidents involving a higher number of firearms would result in greater numbers of deaths and injuries.

 [Click to view tableau dashboard](#)
[click to view on GitHub](#)



Scatterplot analysis shows that most gun violence incidents involve few firearms (0 - 3) and result in minimal deaths or injuries. Rare outliers with very high gun counts often have few or no victims, likely reflecting non-shooting events such as seizures or recoveries. Regression trends confirm no meaningful relationship between the number of guns involved and victim counts, indicating that fatalities and injuries are driven by factors other than firearm quantity.

- These findings directly challenge the initial assumption that incidents involving more firearms would result in higher numbers of deaths and injuries.

Gun violence in the U.S.A (2013-2018)



Data Exploration

 [Click to view tableau dashboard](#)

[click to view on GitHub](#)

Cluster Analysis

Dark purple cluster (Multi-gun, low-victim incidents)

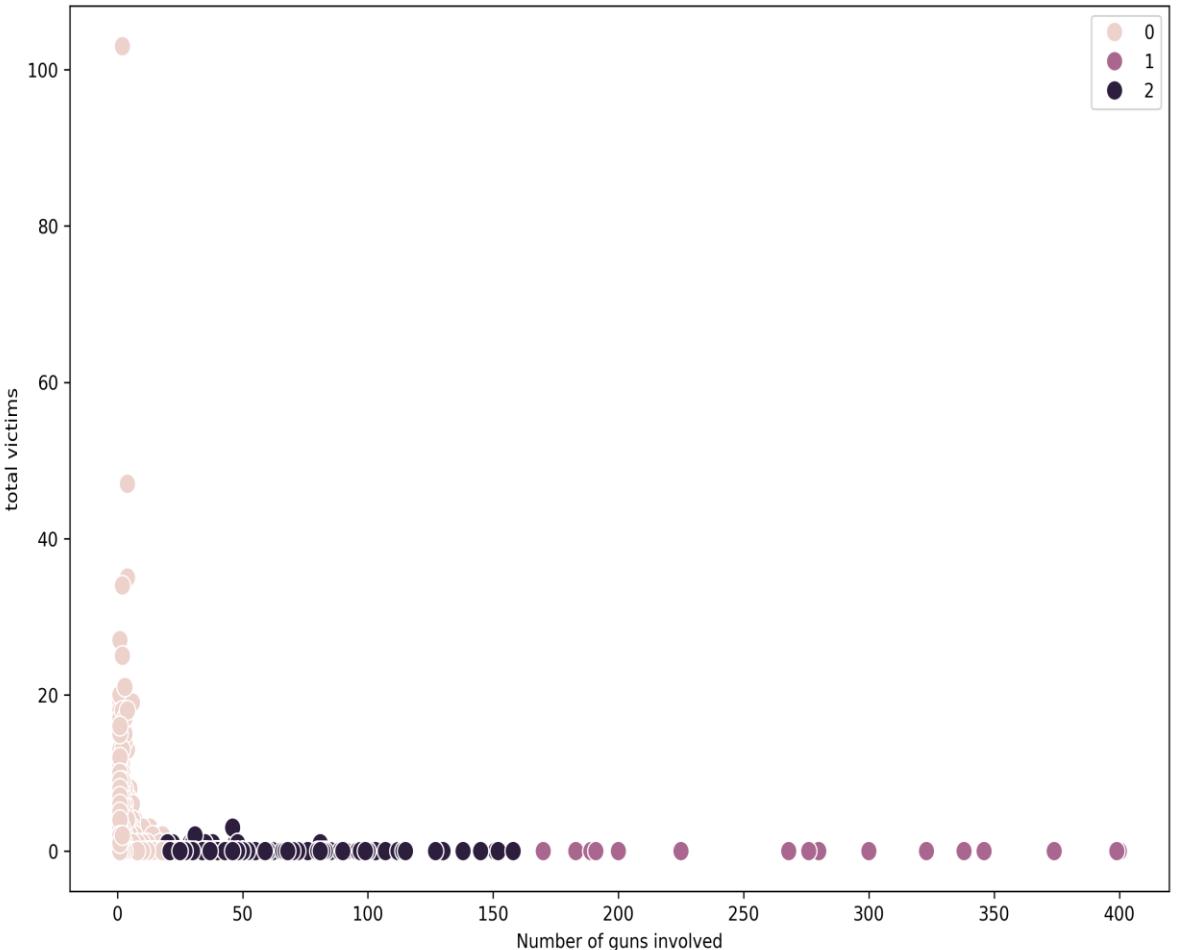
This cluster includes incidents involving large number of firearms, with an average of about 36 guns per event, yet almost no associated victims. Killings and injuries are minimal, with median values of zero across all victim measures. These patterns suggest that the cluster primarily represents non-shooting events, such as gun recoveries, seizures, or possession cases, where multiple firearms are present without violence.

Pink cluster (Typical low-severity incidents)

This cluster represents the most common type of incident in the dataset. Victim counts are generally very low, with an average of fewer than one victim per incident and median values of zero or one for killed, injured, and total victims. The number of guns involved is also minimal, usually one firearm per incident. Overall, these events reflect everyday gun incidents that result in few or no casualties.

Purple cluster (Extreme gun-count outliers)

This small cluster consists of rare incidents with extremely high numbers of firearms involved, averaging nearly 300 guns per event. Despite the large gun counts, these incidents result in no victims, with both mean and median values of zero for all victim metrics. These outlier events heavily influence the clustering results but do not represent typical patterns of gun violence.



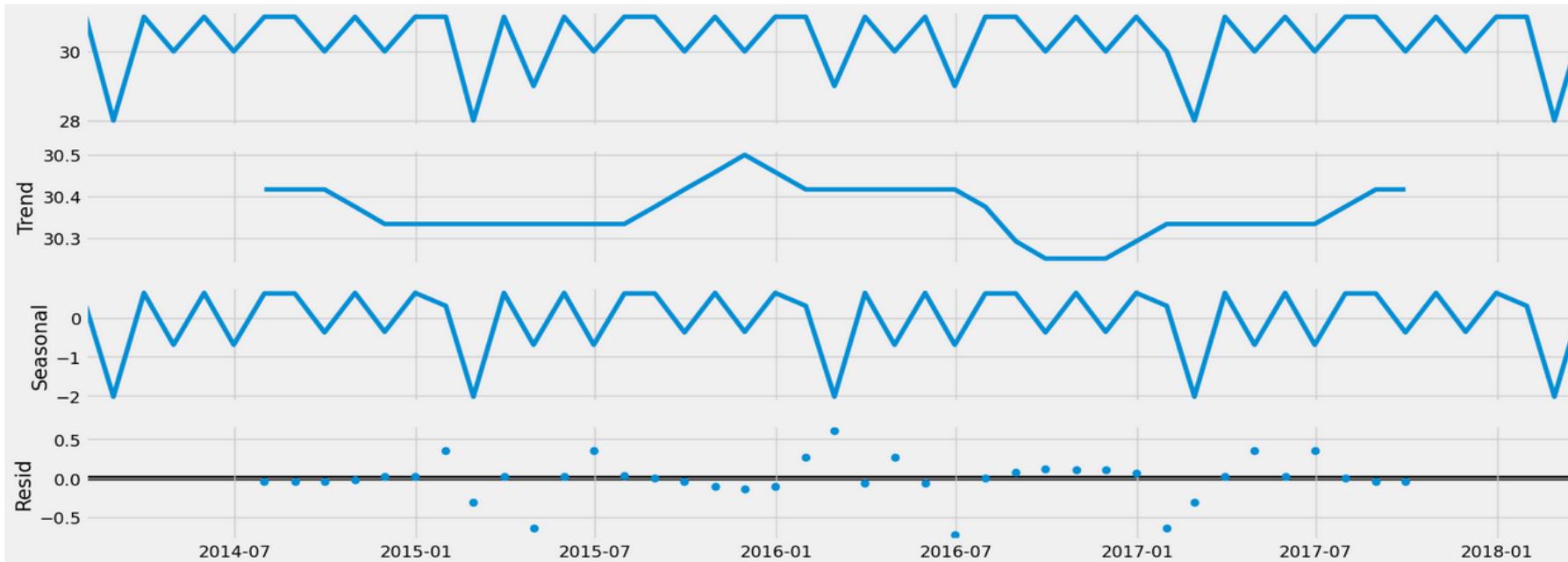
Gun violence in the U.S.A (2013-2018)



Data Exploration

[click to view on GitHub](#)

Time-Series Decomposition of Monthly Gun Violence Incidents (2014–2018)



NOTE:

The early 2013 to 2014 period shows high volatility and a structural break likely due to incomplete reporting, so the analysis focuses on 2014 - 2018 for a stable and reliable time series suitable for modeling and forecasting.

- **Level:** Incidents fluctuate within a narrow range (~30 - 31 per month), showing stable and predictable behavior with no extreme spikes or drops. This consistency validates the focus on the 2014 - 2018 period.
- **Trend:** The series exhibits a shallow, smooth trend with minor variations, indicating no long-term growth or decline. The absence of structural breaks or abrupt changes confirms stability for trend analysis.
- **Seasonality:** Monthly patterns are consistent across years, with small, repeatable deviations, reflecting stationary seasonal effects suitable for seasonal modeling.
- **Residuals:** Residuals are small, centered around zero, and show no systematic patterns, suggesting that trend and seasonality effectively explain the variation.

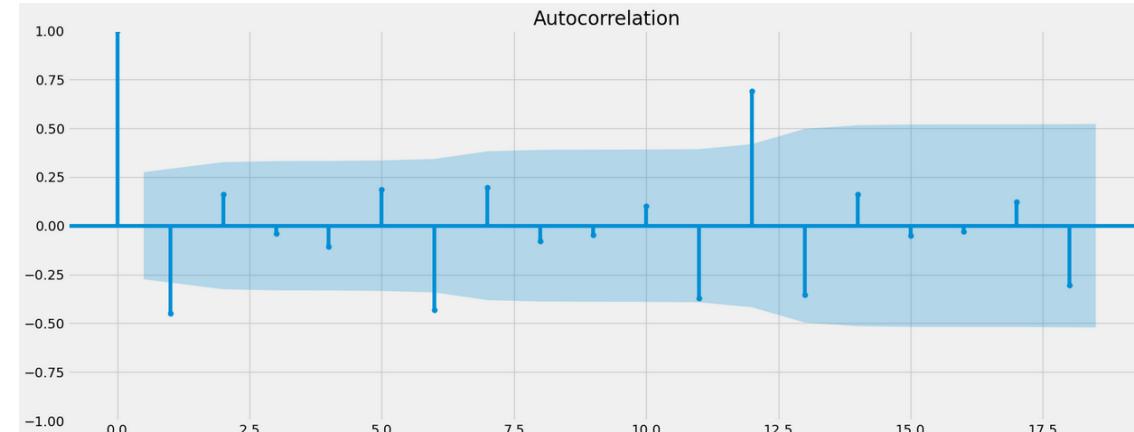
Gun violence in the U.S.A (2013-2018)



[click to view on GitHub](#)

Data Exploration

Testing for stationarity



A few of the vertical lines are beyond the confidence interval, which clearly indicates that the data is non-stationary

Dickey-fuller stationary test:

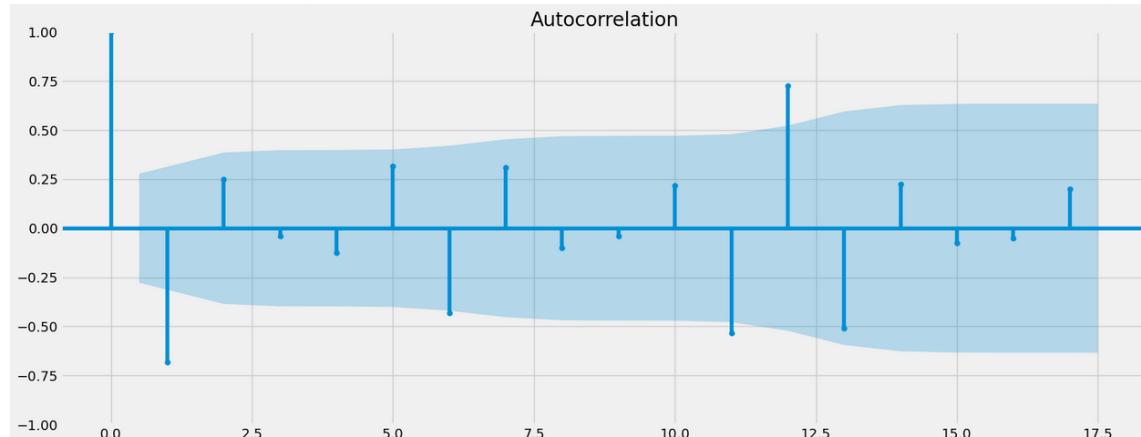
Interpretation

Test Statistic (-2.099147) is greater than all critical values (-3.610400, -2.939109, -2.608063)
p-value (0.244858) is greater than 0.05.

meaning

Test statistic larger than critical value = can't reject null hypothesis.
There is a unit root in the data, and data is non-stationary.

Stationarizing time series data | Method: Differencing



First-order differencing makes the series stationary, with most autocorrelations within confidence bounds; remaining spikes at lag 12 reflect seasonal patterns rather than non-stationarity.

Dickey-fuller stationary test:

Interpretation

Test Statistic (-7.025412e+00) is smaller than all critical values (-3.610400e+00, -2.939109e+00, -2.608063e+00)
p-value (6.383451e-10) is smaller than 0.05.

meaning

Test statistic smaller than critical value = can reject null hypothesis
There is no unit root in the data, and data is stationary.

- The Dickey-Fuller test and autocorrelation analysis confirm that the series is stationary, with no long-term persistence or unit root. The stationarity validation of the gun violence series ensures the data is stable and suitable for building accurate forecasting models.

Gun violence in the U.S.A (2013-2018)



Insights, Limitations and Next Step

Key Insights:

1. Gun violence incidents show stable trends over 2014 - 2018, with predictable seasonal patterns and clear clustering of incident types.
2. Geographic patterns reveal states with consistently higher incident rates, providing targets for policy interventions and community safety programs
3. High-impact shootings often involve few firearms but result in many victims, while some incidents with many guns show minimal harm, highlighting that gun count alone does not drive severity.
4. Cluster analysis reveals distinct incident types: high-impact shootings with few guns, non-violent high-gun incidents, and moderate-gun low-victim events, highlighting that gun count alone does not determine harm.

Limitations:

1. Early data (2013 to early 2014) is incomplete and volatile, limiting the historical scope of the analysis.
2. The dataset may underreport certain types of incidents or lack contextual variables (e.g., socioeconomic factors, law enforcement activity) that influence gun violence.
3. Causal relationships cannot be inferred; the analysis identifies patterns and correlations but not direct causes

Next Step:

1. Apply predictive modeling and time-series forecasting to anticipate future incident trends and seasonal peaks.
2. Integrate additional datasets (demographics, policy changes, crime reports) to enrich analysis and explore causal factors.

Contacts

Email: oluwole.akorede@yahoo.com

<https://github.com/Oluwole-Oni>

[LinkedIn Account](#)

Phone Number: 443-522-6556