Obinwanne Alisigwe

- **DATA ANALYTICS PORTFOLIO**

Links:

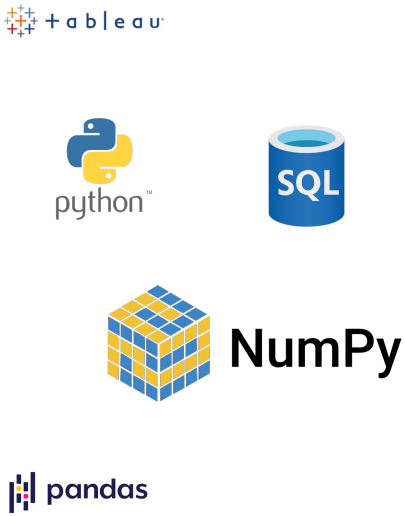
Hello, I’m Obi

Welcome to my current portfolio. I am a data analytics with background in oil & gas, sales, project management, and strong analytical and organizational skills. My technical expertise includes Microsoft Excel, Python, SQL, Tableau, and various web platforms. I have led quality assurance and safety initiatives in automotive, bioplastics, and oil & gas industries. Fluent in English and proficient in German, I possess strong leadership and communication abilities, with a passion for data analytics, teamwork, and innovation. My approach is guided by reliability, flexibility, and a commitment to continuous improvement.

Projects

1. Gun Violence Data
2. GameCo Regional (Sales Analysis)
3. Influenza Season (Epidemic Analysis and forecast)
4. Rockbuster (Inventory Analytics)
5. Instacart Basket (Online Grocery)
6. Pig E. Bank (Finance)

Tools



GUN Violence Data

Objective

The primary objective of the gun violence dataset is to provide comprehensive and detailed information on gun-related incidents in the United States from 2013 to 2018. This data aims to facilitate the analysis and understanding of gun violence trends, demographic impacts, and geographical distributions. By offering insights into the nature and scope of these incidents, the dataset supports the development of targeted interventions, informed policy decisions, and community-based strategies to reduce gun violence and its effects.

Requirement

The chosen data must meet specific criteria so that you can conduct the procedures explored in this Achievement (and develop the necessary skills for a junior analyst). You'll likely need to source multiple data sets while working through the Achievement; however, we advise you to start with one main data set. Keep in mind that data sourcing can be very time-consuming!

Data Set

The gun violence dataset from Kaggle provides detailed information on gun-related incidents in the United States from 2013 to 2018. This dataset supports comprehensive analysis and helps inform strategies to combat gun violence.

Skills & Tools

Geographical analysis

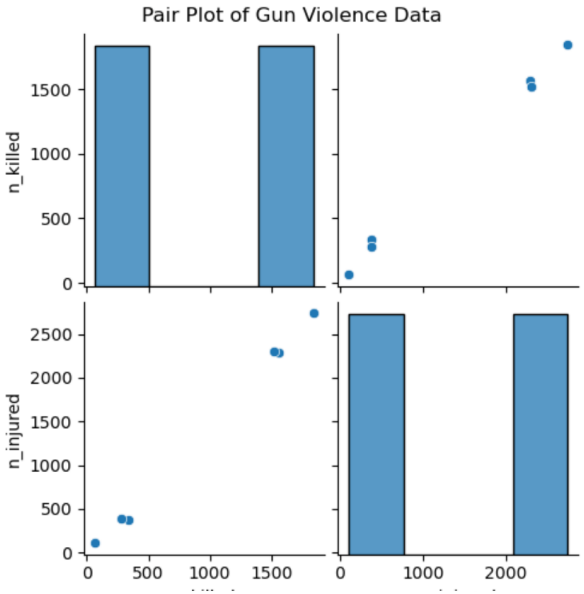
Exploratory analysis

Cluster analysis

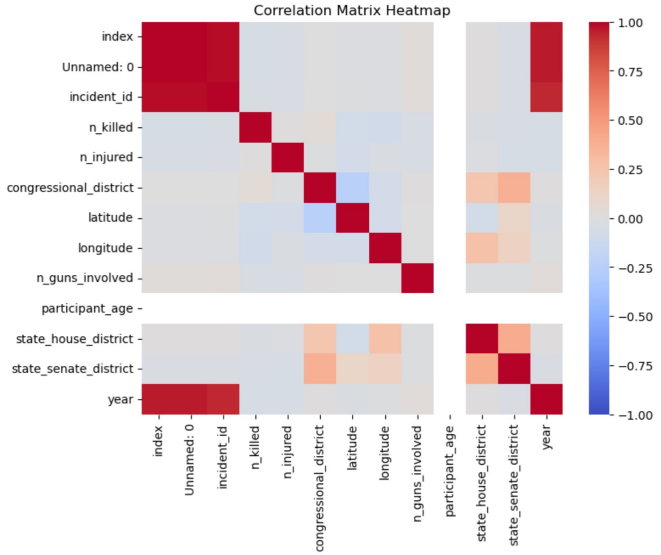
Incident Count

Time-series analysis

GUN Violence Data

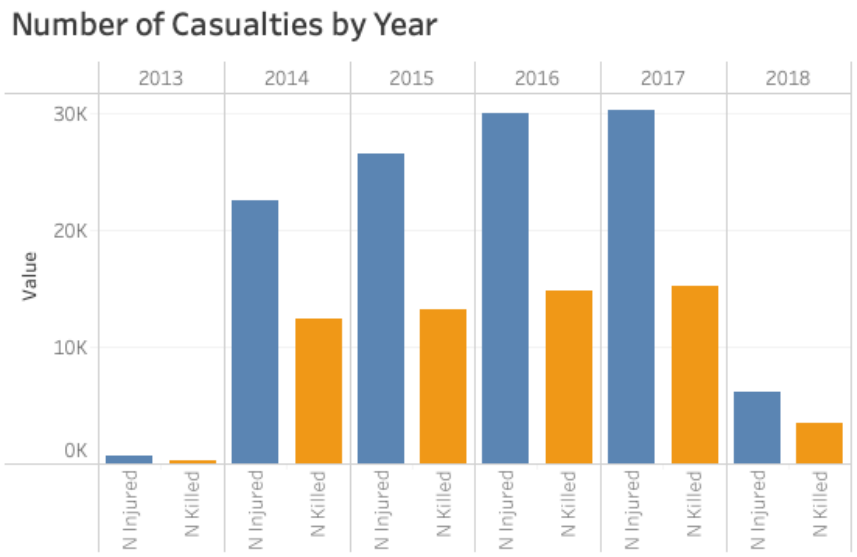


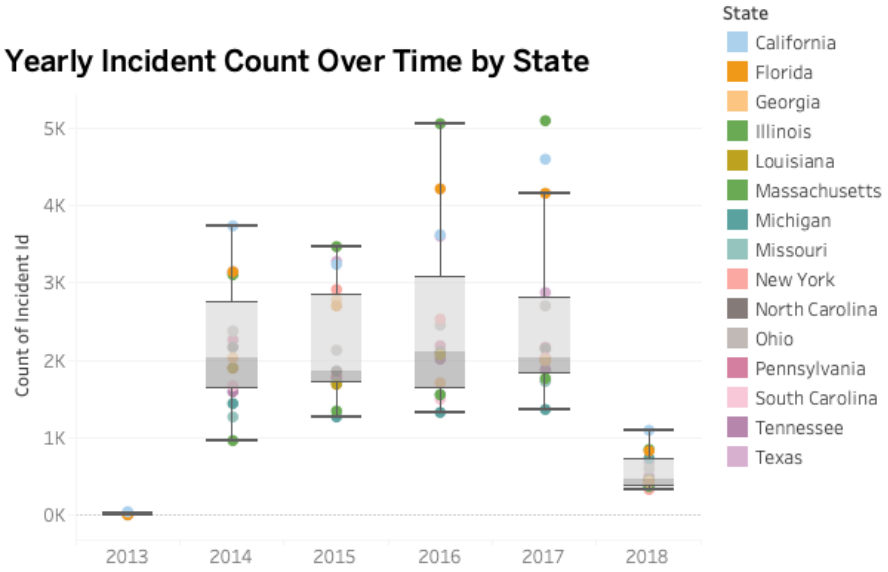
The pair plot provides a visual overview of the distribution of each variable and their relationships. The diagonal shows the distribution of each variable, while the off-diagonal plots show scatterplots of variable pairs. Variables that show strong correlations in scatterplots are worth further exploration, like n_killed and n_injured.



Strong correlations can indicate potential areas for further analysis or potential causative relationships. For example, if 'n_killed' and 'n_injured' are strongly correlated, this may warrant further investigation into factors that simultaneously affect both.
Weak correlations might suggest that other variables or non-linear relationships should be explored.
By visually examining the heatmap, we can quickly identify which pairs of variables have significant relationships. This can guide subsequent analyses, such as regression modeling or hypothesis testing, to better understand the underlying patterns in the data.

Yearly Incident Count





Observation

There is a sharp rise in the incident count from 2013 to 2014. This significant increase suggests a potential change in reporting practices, legislation, or actual increase in gun violence incidents.

The most noticeable trend is the steep decline in incident counts from 2017 to 2018. This sharp drop could be attributed to various factors, including improved law enforcement, successful intervention programs, or changes in data collection methods.

Conclusion:

The yearly incident count provides valuable insights into the trends and changes in gun violence incidents over the years. Understanding these trends can help in formulating better policies and intervention strategies to address gun violence more effectively.

Result & Recommendation

Result

The analysis of the gun violence dataset from 2013 to 2018 reveals critical insights into the prevalence and characteristics of gun-related incidents across the United States. Key findings include:

High Incidence States: States such as California, Illinois, Texas, Florida, and New York report the highest number of gun violence incidents. This concentration is largely driven by their large urban populations and diverse socio-economic factors.

Demographic Impact: The data shows varying impacts across different age groups, genders, and races. Certain demographics are disproportionately affected by gun violence, highlighting the need for targeted prevention efforts.

Incident Types: The dataset categorizes incidents into different types (e.g., homicides, accidental shootings), providing insights into the nature of gun violence and guiding intervention strategies.

These results underscore the need for data-driven approaches to address gun violence, particularly focusing on high-incidence states and at-risk demographics.

Next Steps

Data Enrichment: Integrate additional data sources, such as socio-economic indicators and policy changes, to provide a more comprehensive understanding of gun violence.

Geographic Focus: Conduct in-depth studies on specific high-incidence states and cities to develop localized intervention strategies and policies.

Demographic Insights: Further explore demographic factors and their correlation with gun violence to tailor prevention and support programs more effectively.

Policy Evaluation: Assess the impact of existing gun control policies and interventions to determine their effectiveness and identify areas for improvement.

By addressing these limitations and following through with the proposed next steps, stakeholders can enhance their understanding of gun violence and develop more effective strategies to mitigate its impact.

GameCo
Regional Sales Analysis

Objective

You're an analyst for a new video game company, GameCo, which wants to use data to inform the development of new games. As such, you've been asked to perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.

Data Set

Throughout this course, you'll be using a data set that covers historical sales of video games (for games that sold more than 10,000 copies) spanning different platforms, genres, and publishing studios. This data was drawn from the website VGChartz.

Requirement

a.Are certain types of games more popular than others?

b.What other publishers will likely be the main competitors in certain markets?

c.Have any games decreased or increased in popularity over time?

d.How have their sales figures varied between geographic regions over time?

Skills & Tools

Data cleaning

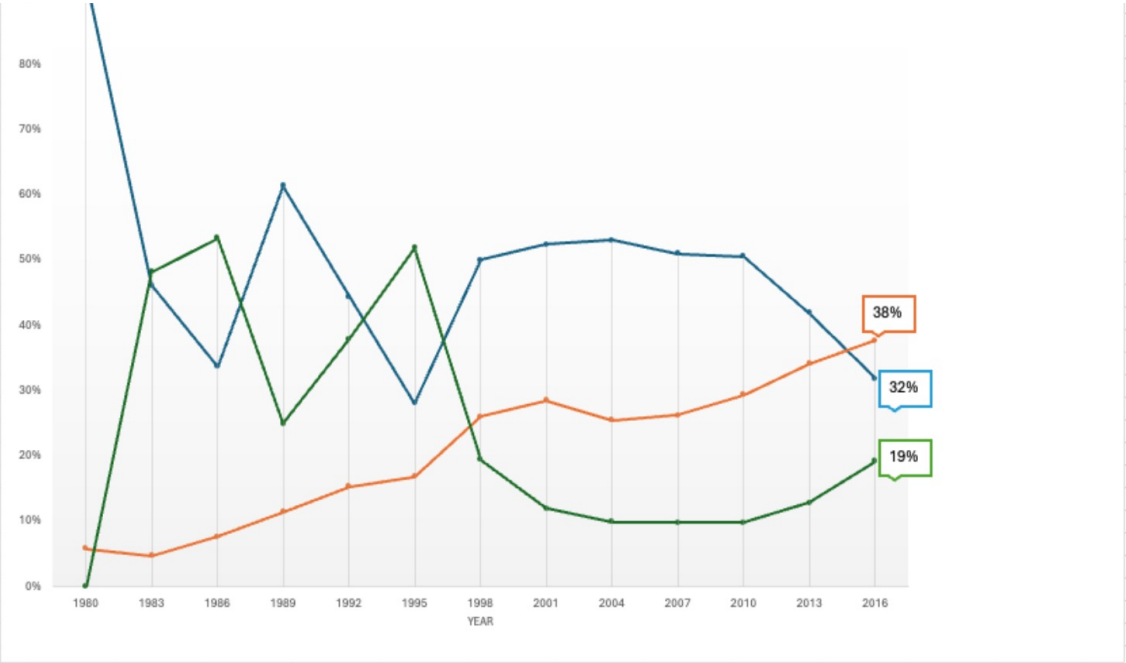
Grouping & summarizing data

Descriptive analysis

Utilizing Pivot Tables

Visualization charts in MS Excel/Power Point

Data Reveal New Insights

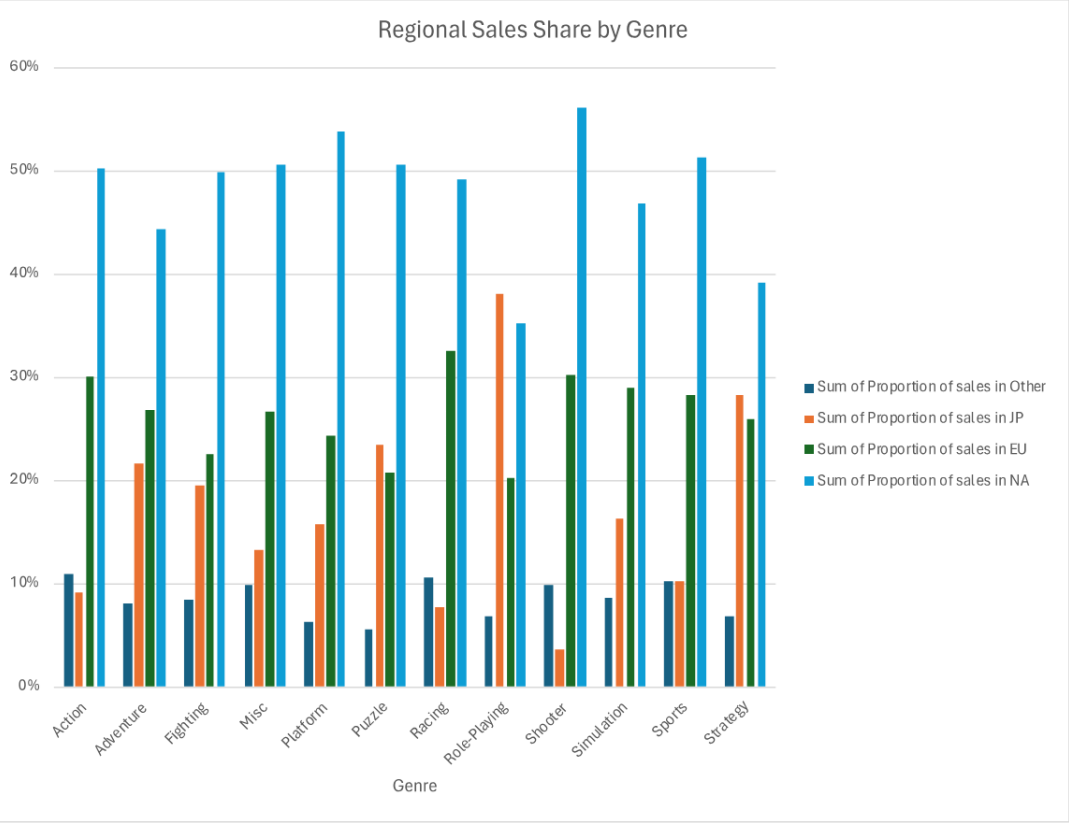


In the year 2016, the Europe sales went up even more than the leading North America for global market share of video games sales

There is an increase in the Japan sales up 19% it's about 8 points higher when you compared it with the sales in the last 10-15yrs

There is a consistent decline in the North American sales in the last 6yrs compared to the way it was 10yrs prior

Recommendation



An increase in advertising in the European market should be a continuous process and the Other region needs to be considered

Europe has the most population, an establishment of GameCo office in Europe should be considered and Japan as well

An increase in 2017 budget should be revised and reflect an increase in the JP and Other region

A reduction in the marketing budget for the North America should be considered and the budget should be channeled to the EU, JP and Other since the NA region still maintain a huge number of sales regardless of the recent decline

For more info read here

Influenza Season
Epidemic Analysis and forecast

Goal

To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

Requirements

Provide information to support a staffing plan, detailing what data can help inform the timing and spatial distribution of medical personnel throughout the United States.

Determine whether influenza occurs seasonally or throughout the entire year. If seasonal, does it start and end at the same time (month) in every state?

Prioritize states with large vulnerable populations. Consider categorizing each state as low-, medium-, or high-need based on its vulnerable population count.

Assess data limitations that may prevent you from conducting your desired analyses.

Data Set & Tools

Influenza deaths by geography (Source: CDC)

Population data by geography, time, age, gender (Source: US Census Bureau)

Skills & Tools

Data cleaning, integration, & transformation

Statistical hypothesis testing

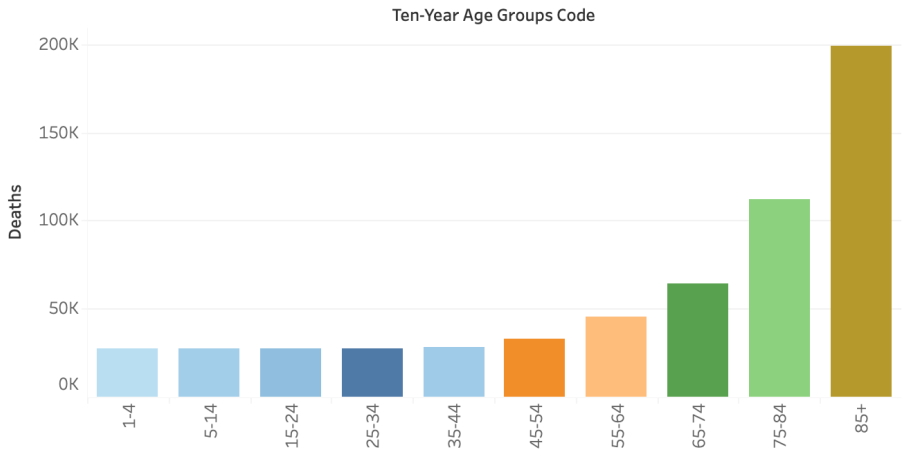
Visual analysis

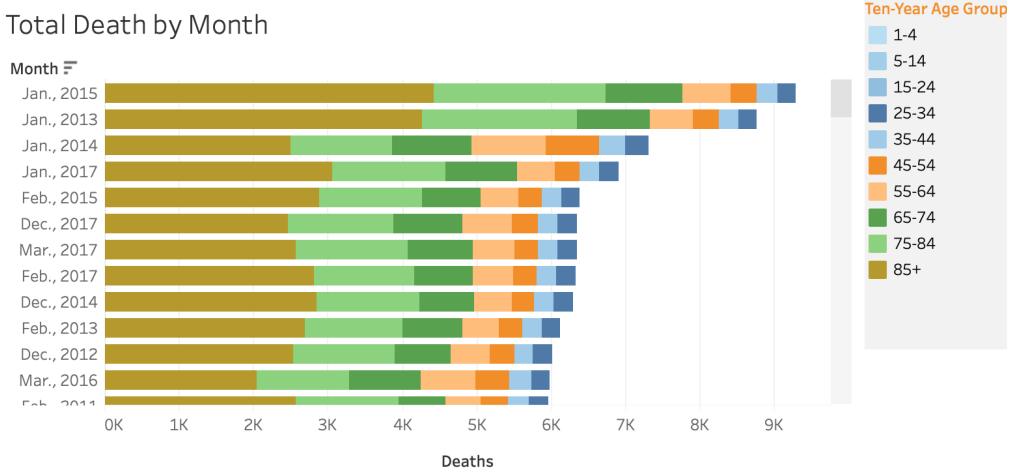
Forecasting

Storytelling in Tableau

Presenting results

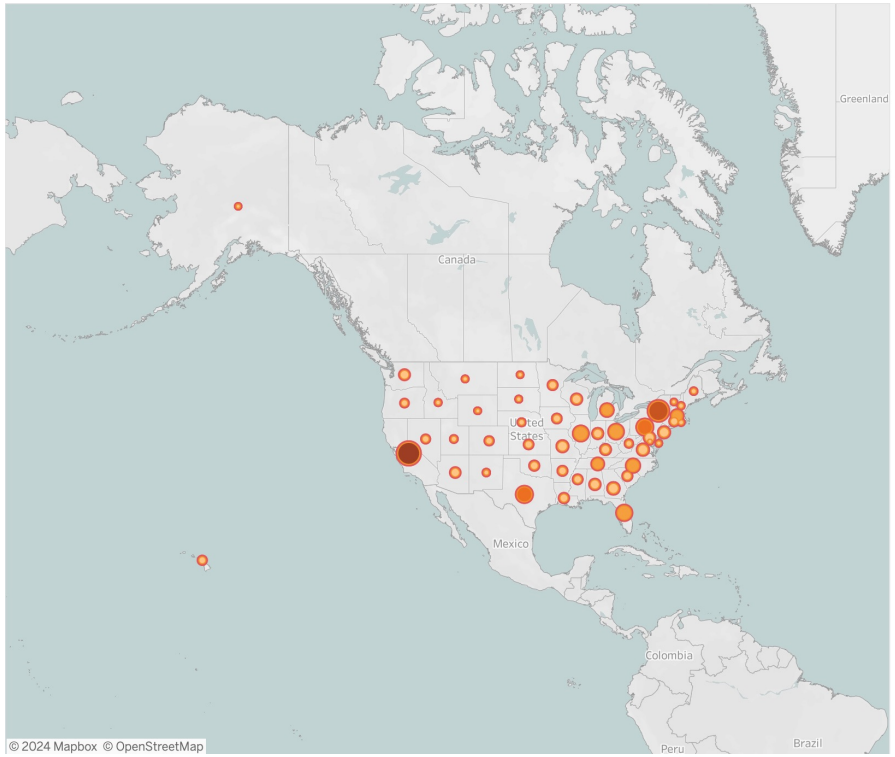
Influenza Season
Epidemic Analysis and forecast





- Is Influenza Seasonal?
- Following the trend of influenza death for each year, it seems that influenzas death start to increase from Nov. and decrease after March.
- With the peak month for influenza death occurring in January.
- States with the highest count of vulnerable death
- California has the highest number of vulnerability, and Massachusetts is the list

Influenza Season
Epidemic Analysis and forecast



State	
California	53,382
Florida	23,339
Georgia	15,070
Michigan	16,645
New York	42,155
North Carol..	17,912
Ohio	20,952
Pennsylvan..	24,744
Texas	27,544

Top 10 Populated States

State	
California	338,021,212
Texas	225,790,532
New York	174,835,891
Florida	171,082,777
Illinois	113,534,600
Pennsylvan..	113,148,158
Ohio	101,550,516
Michigan	87,857,710
Georgia	85,669,964
North Carol..	84,319,476

tableau

Conclusion
Through this research project, it was identified that adults ages 65yrs and above are most vulnerable population when affected when affected by influenza and influenza season starts from November and ends after March

Recommendations
States like California, New York, and Texas should be priortised for allocation of additional staffing support when constucting the staffing plan.

These additional medical staff should be sent across the mostly affected states at the beginning of November when the influenza season starts.

Deliverables

- View complete storyboard here on Tableau Public
- View presentation to stakeholders here

Rockbuster
Inventory Analytics

Objective

Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive.

Requirement

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?

Data Set

Rockbuster's relational database

Skills & Tools

Relational databases

Database querying with SQL

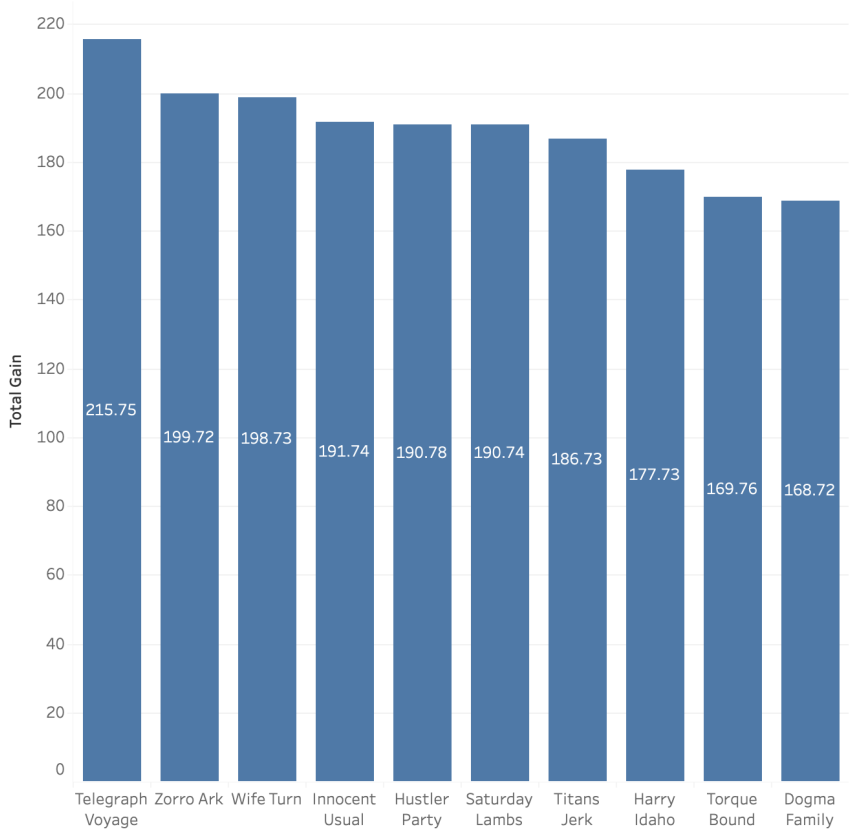
Filtering, cleaning and summarizing data with SQL

Joining tables

Performing subqueries

Common table expressions

Geo Distribution & Top 10 Movies by Revenue

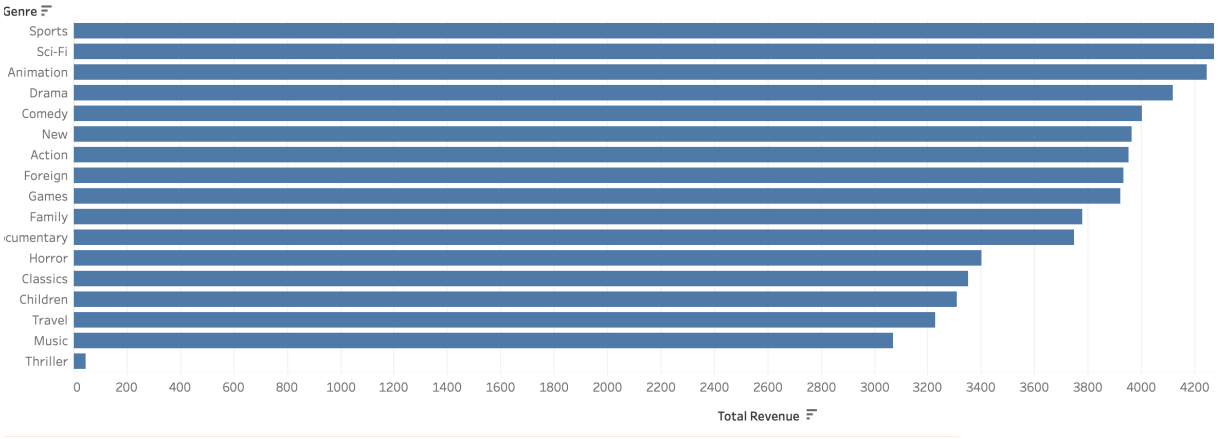


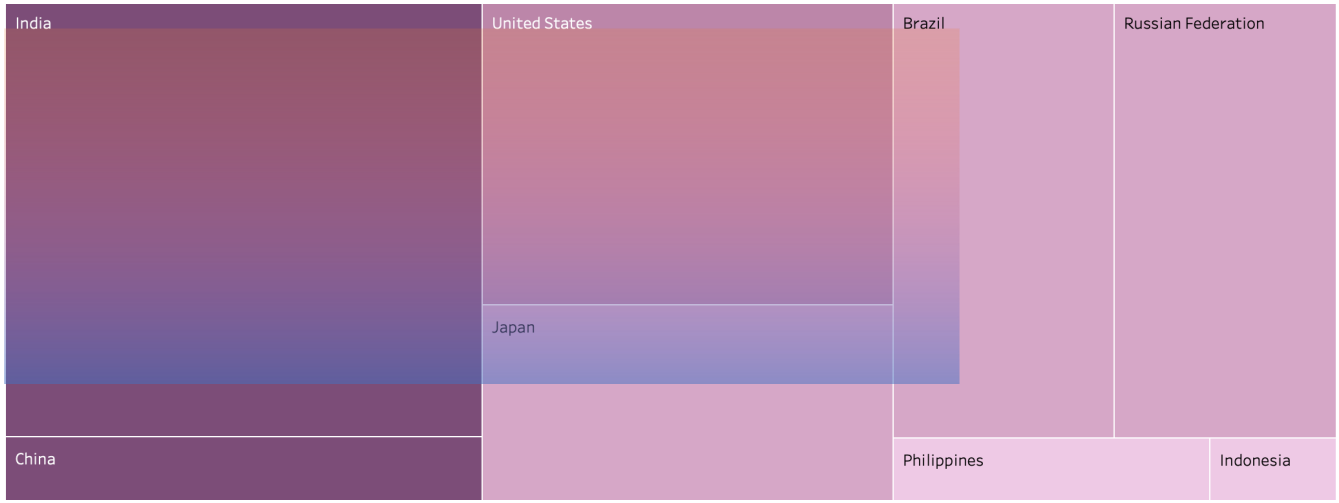
- **Key Insights**

The top 10 countries with the most customers are: India, China, United States, Japan, Mexico, Brazil, Russia, Philippines, Turkey, and Indonesia

The revenue of the top 10 ranges from $215,75 to 168,73, indicating a narrow range and consistent performance among the top performers.

Movie Genre Performance & Top 10 Countries by Customer Count





- Key Insights
- Action and Thriller genres should receive increased marketing efforts to take advantage of their growing popularity. Additionally, continue investing in Sports, Sci-Fi, and Animation genres, as they demonstrate high rental counts and revenue.
- India top the customer by counts followed by China and United States while Indonesia is the list.

Three Actionable Insights

1. Content Powerhouse:

• Dominate Popular Genres: Focus your library on the genres that resonate most with your audience.

• Promote Top Performers: Shine a spotlight on your most successful movies to keep them in front of viewers.

2. Smart Pricing & Rentals:

• Dynamic Pricing: Implement data-driven pricing models that adjust based on demand and rental duration.

• Smarter Acquisitions: Use data insights to guide content acquisition, focusing on replicating what works.

3. Customer Champions:

• Personalized Marketing: Develop personalized marketing strategies and loyalty programs to retain high-value customers.

• Targeted Engagement: Motivate mid-tier customers with targeted promotions and incentives, driving deeper engagement.

Conclusion

1. Leverage High-Performing Content and Target Key Regions: Prioritize popular genres and promote top performing movies. Tailor marketing campaigns to regions with concentrated customer bases.
2. Optimize Pricing and Rental Strategies: Implement dynamic pricing models based on demand and rental duration. Utilize data-driven insights to guide content acquisition and replicate successful factors.
3. Enhance Customer Loyalty and Engagement Programs: Develop personalized marketing strategies and loyalty programs for high-paying customers. Engage mid-tier customers with targeted promotions and incentives.

Links:

Data Dictionary: Comprehensive documentation of data elements. Here

Excel Workbook: Detailed data analysis for technical colleagues. Here

Full Presentation: In-depth analysis and findings. Here

Visualizations: Created in Tableau. Here

Instacart Basket Online Grocery

Objective

Instacart uncover more information about their sales patterns. Your task is to perform an initial data and exploratory analysis of some of their data to derive insights and suggest strategies for better segmentation based on the provided criteria.

Data Set

Open-source datasets from Instacart (Kaggle) Instacart is a real company with online data, but this project's contents are fabricated for the assignment.

32,404,859 rows/entries

Requirement

Determine peak days and hours for orders to optimize ad scheduling.

Identify peak spending times to tailor product advertisements.

Simplify product pricing into actionable price range groups.

Analyze popular product types across departments to guide marketing efforts.

Explore customer segmentation and ordering behaviours to enhance targeting strategies.

Skills Tools

Python

Data wrangling , merging, and aggregation

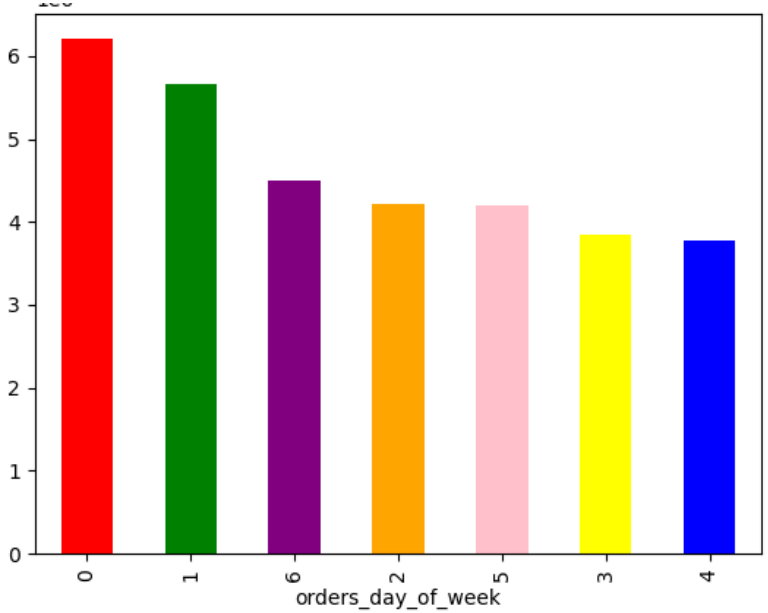
Deriving variables

Grouping and aggregating dataframes

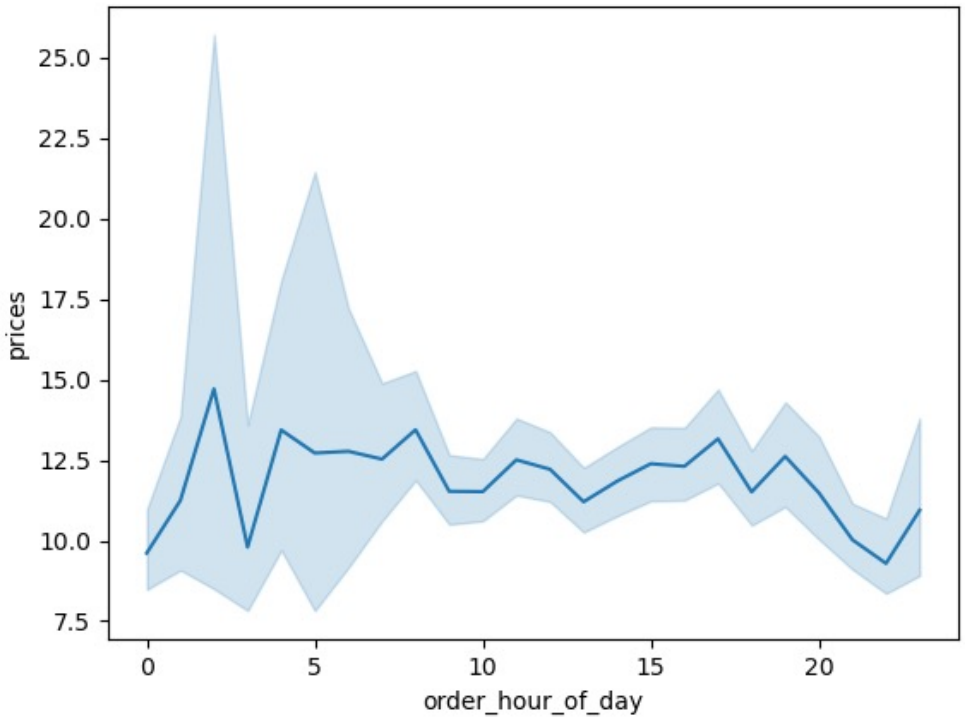
Reporting in Excel

Populations flows

Time of order

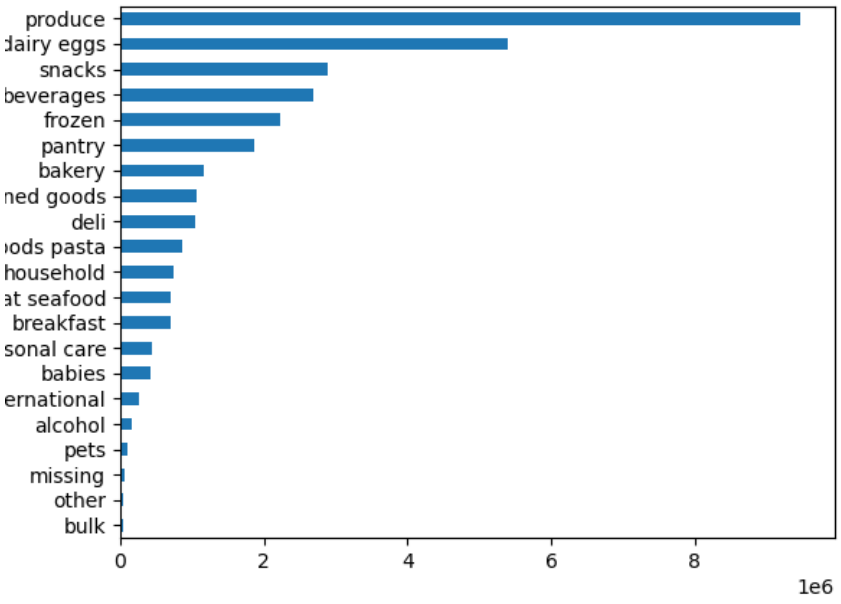


Saturday and Sunday are the days with most orders.
(Saturday=Red, Sunday=Green)



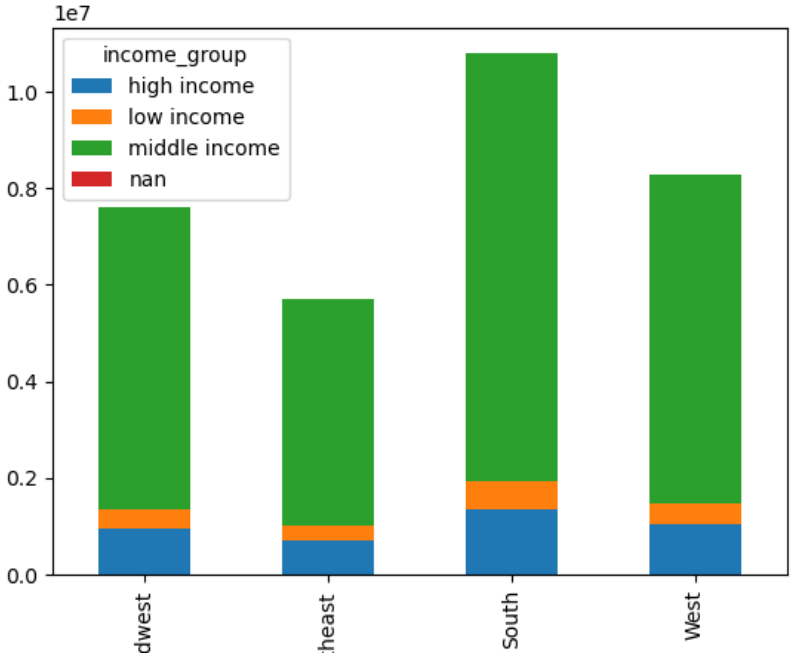
- The orders with the highest average price of items occur between midnight and 6am
- Specifically, between 3am and 4am

Analysis & Insights



Understanding the products

To analyse popular product types, I used Python to create a bar chart of the top 15 items by sales count. This visualization highlights key products driving sales, aiding inventory and marketing decisions.



Assessing Regional Differences

To identify regional differences, I aggregated total order numbers and prices across all regions. The South showed the highest spending, but no distinct patterns emerged; spending remained consistent across regions.

RESULTS & RECOMMENDATIONS

Ad Scheduling: Schedule ads on Tuesday and Wednesday evenings after 6 PM for maximum exposure during low-order periods, as most customers shop on weekdays.

Price Trends: Implement flash sales or bundle offers during peak purchase times at 3 AM, 6 AM, 5 PM, and 8 PM to capitalize on the demand for higher-priced items.

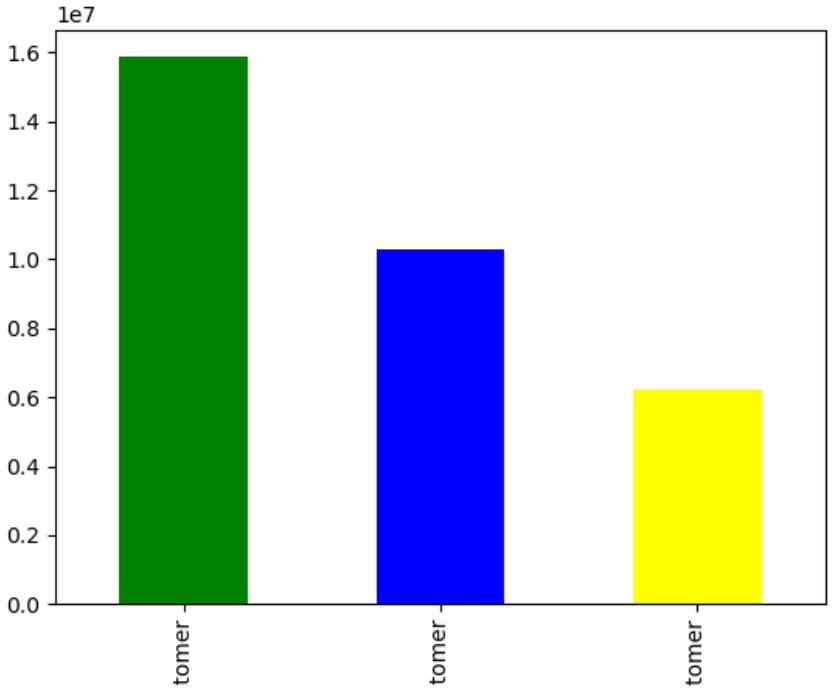
Price Range Strategy: Focus volume-driven promotions on low-range products, broad marketing campaigns on mid-range products, and targeted ads for high-range products to attract higher-income consumers.

Category Promotions: Prioritize ads and promotions for the Produce and Dairy & Eggs departments, with additional focus on Snacks, Beverages, and Frozen items, leveraging their convenience appeal.

Customer Loyalty: Develop loyalty programs to encourage regular and new customers to increase their orders while retaining current loyal customers, as over half are regular customers and one-third are loyal.

Regional Optimization: Tailor resource allocation, inventory, staffing, and advertising strategies to match regional demands, with the South having the most orders and the Northeast the least.

Demographic Focus: Middle-aged single adults are the most frequent shoppers, and parents form 75% of the customer base. Develop family-focused loyalty programs emphasizing fresh produce, animal-sourced foods, and bulk offerings to drive sales and loyalty.



Thank you