Anjan Pakhrin

Data Analyst Portfolio



About Me

- Mechanical Engineer with 5+ years' experience in CAD design and PLM consulting graduated **Master's degree in Mechanical Engineering** from **Anhalt University of Applied Sciences**, Sachsen Anhalt, Germany

- Skilled in tools such as **CATIA, Creo, Autodesk Fusion**, and experienced in implementing **PLM systems** to improve product development processes

- Currently transitioning into **Data Analytics**, combining strong analytical thinking and problem-solving skills from engineering with new expertise in **Python, SQL, data visualization, and statistical analysis**

- **Data Analytics program at CareerFoundry**, **Germany**

Portfolio Overview

CASE STUDIES	TOOLS
01. GameCo: Video Games Popularity Analysis	Excel, PowerPoint
02. Staffing Agency: Preparing for Influenza	Excel, Tableau
03. Rockbuster Stealth: Movie Rental Data Analysis	Excel, SQL, Tableau, PowerPoint
04. Instacart: Sales Pattern Analysis	Excel, Python
05. Pig E. Bank Churn Prediction	Excel
06. US Traffic Accidents Analysis	Excel, Python, Tableau

Project 1: Video Game Popularity Analysis

Background

GameCo assumed regional sales of video games remained consistent and sought to confirm this with historical data before finalizing the 2017 marketing budget.

Objectives

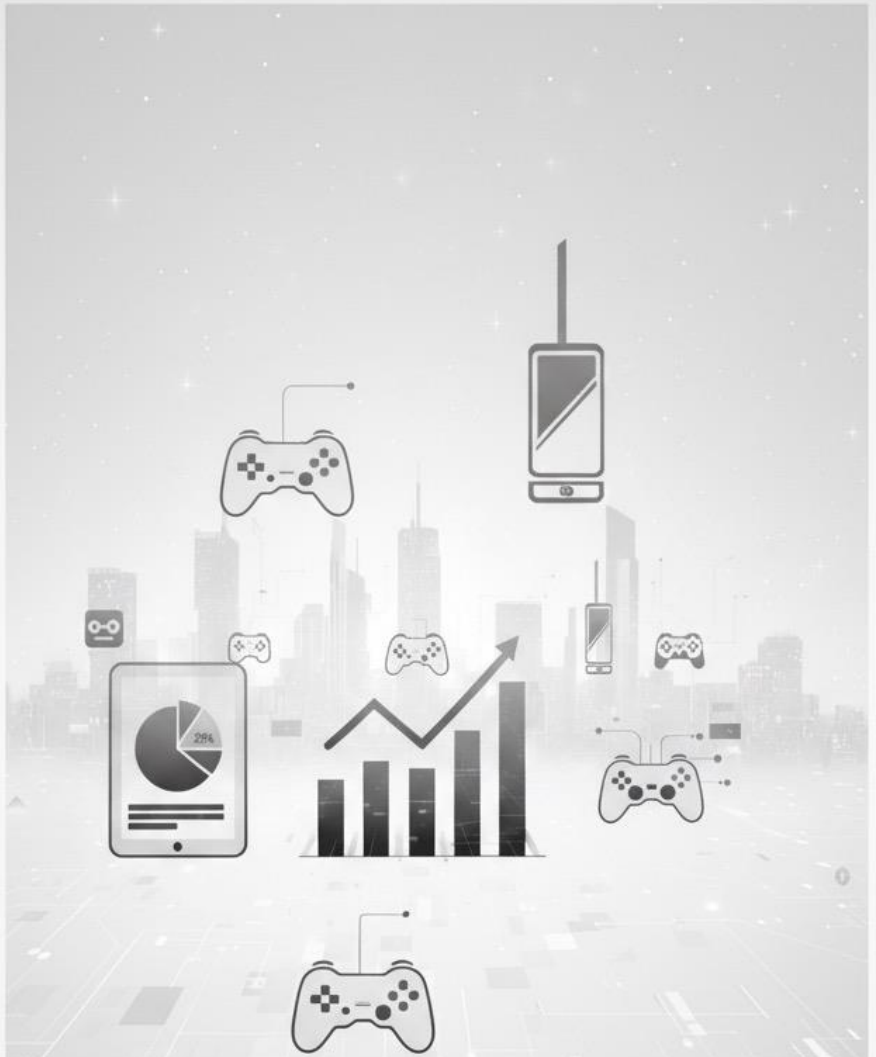
- Identify the most popular game genres and platform
- Analyze key competitors and their market presence across regions
- Track trends in game popularity over time
- Compare sales performance across different geographic regions

Methods

- Organize, clean and prepare data in Excel
- Descriptive statistics and data summarization
- Visualizations of findings with charts and graphs

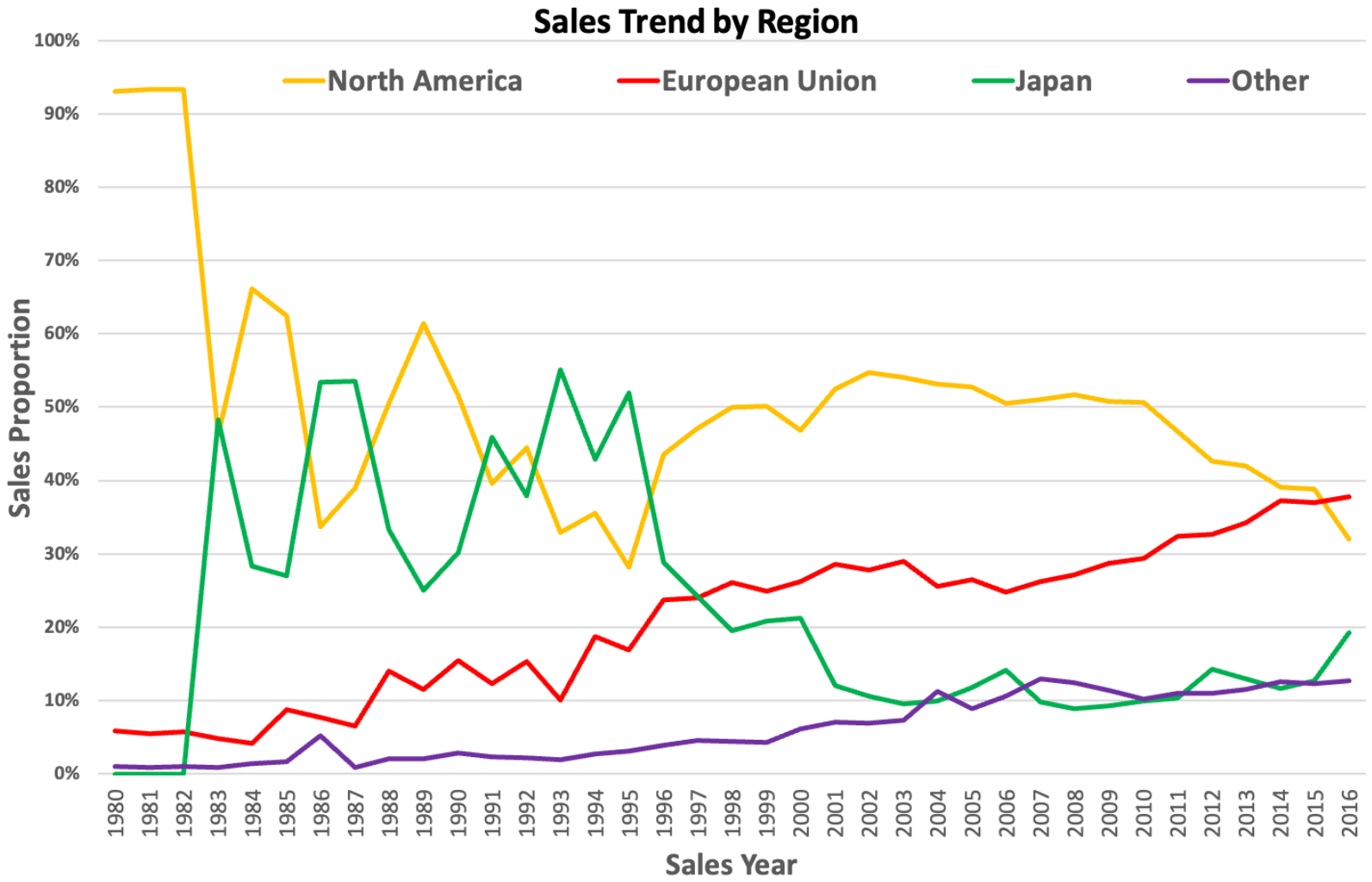
Data

- Video game sales (1980 – 2017) from VGChartz



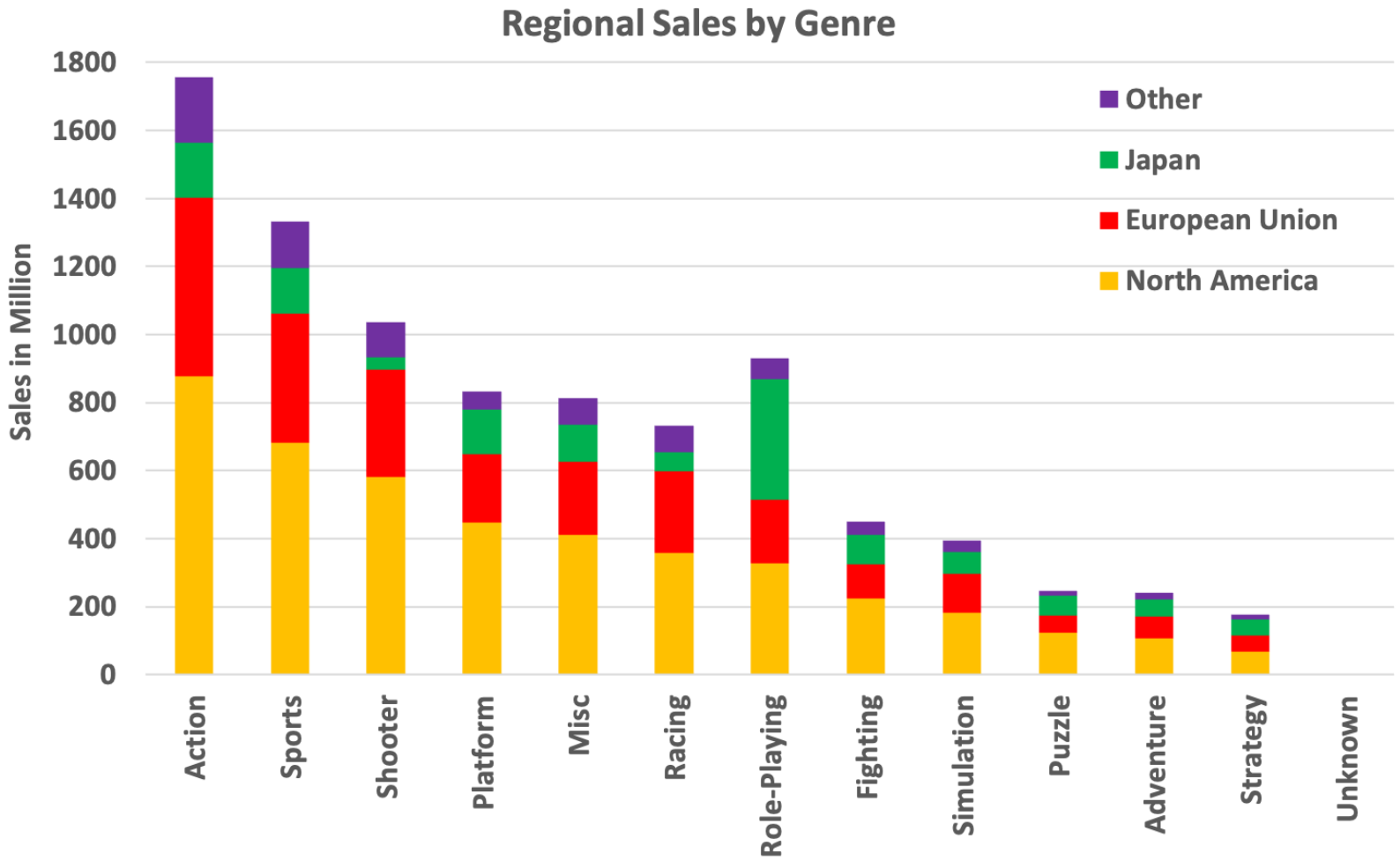
Analysis Tools

Regional Sales Trend Changes over Time



- Sales in NA declined significantly from 93% in 1980 to 31% by 2016
- EU overtaking global dominance with 37% sales over NA by 2016
- Japan's share dropped after mid-1965 losing it's market share from ca. 50% to 19% by 2016
- Other regions show slight but continuous growth

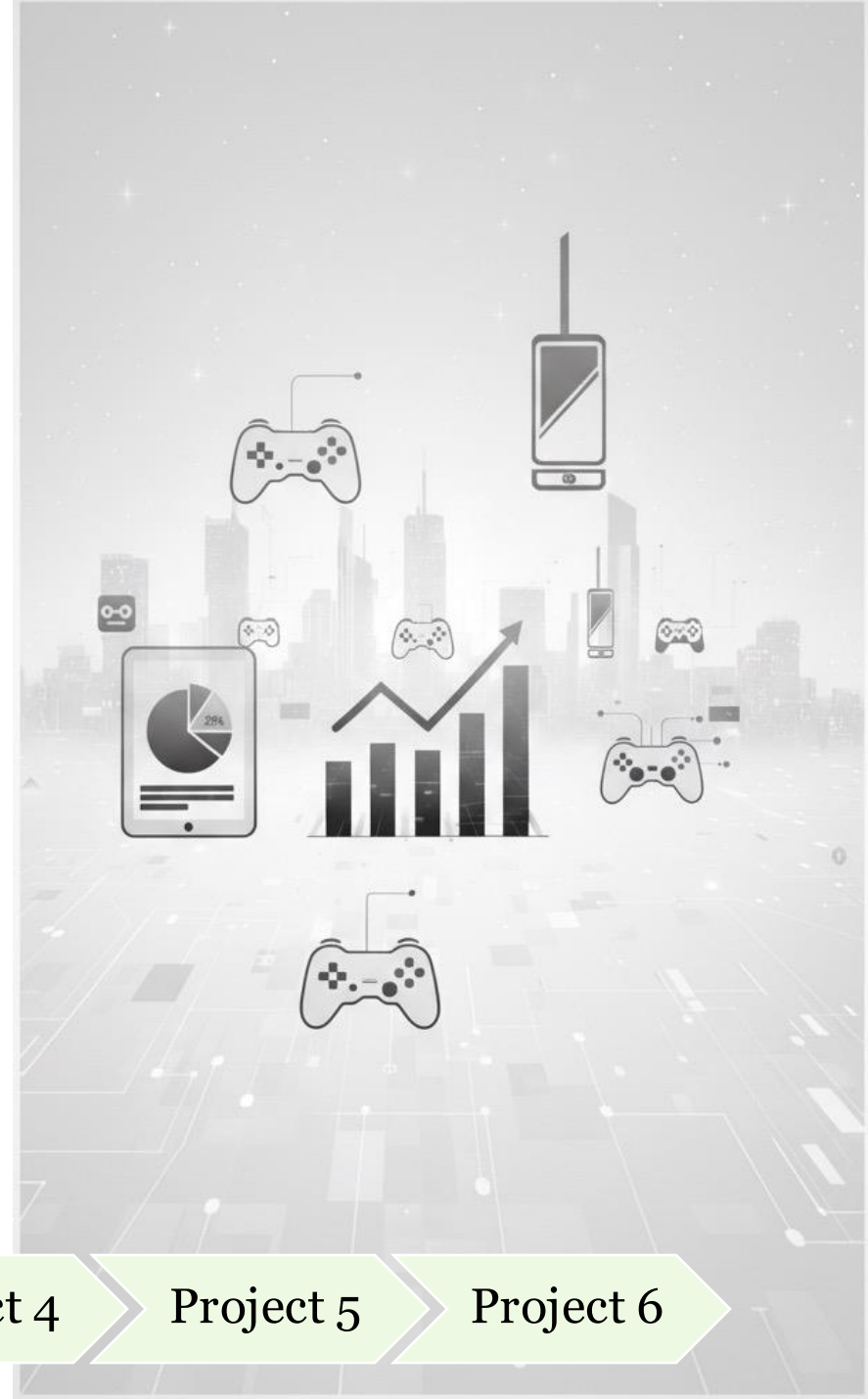
Regional Sales Revenue by Genre



- Action, Sports, and Shooter genre bring the most revenue in North America, European Union, an other regions. It also demonstrates their popularity in these regions accordingly
- In the other hand, Role-Playing genre brings the largest revenue in Japan

Insights & Recommendations

- Market for GameCo is not static as it's assumption → it shifts over time as the sales proportion in the **North American** market is declining
    - Focus on high-performing genre like "Action", "Sports", and "Shooter"
    - Explore mobile or digital platforms to revive growth
- **Europe** may now be a more important market than previously thought → relocate marketing budget for 2017 accordingly
- GameCo is losing its market in **Japan** → requires re-evaluation of investment in Japan
- Explore emerging opportunities in Other Regions (China, India) – steady growth suggests long-term potential , worth's considering small test campaign



Project 2: Preparing for Influenza Season

Background

A U.S. medical staffing agency needs proper plan to balance the resource-distribution for upcoming flu-season across US-states

Objectives

- Identify vulnerable and non-vulnerable groups to influenza
- Track the trends to prioritize vulnerable groups and states
- Forecast optimized staffing nationwide

Methods

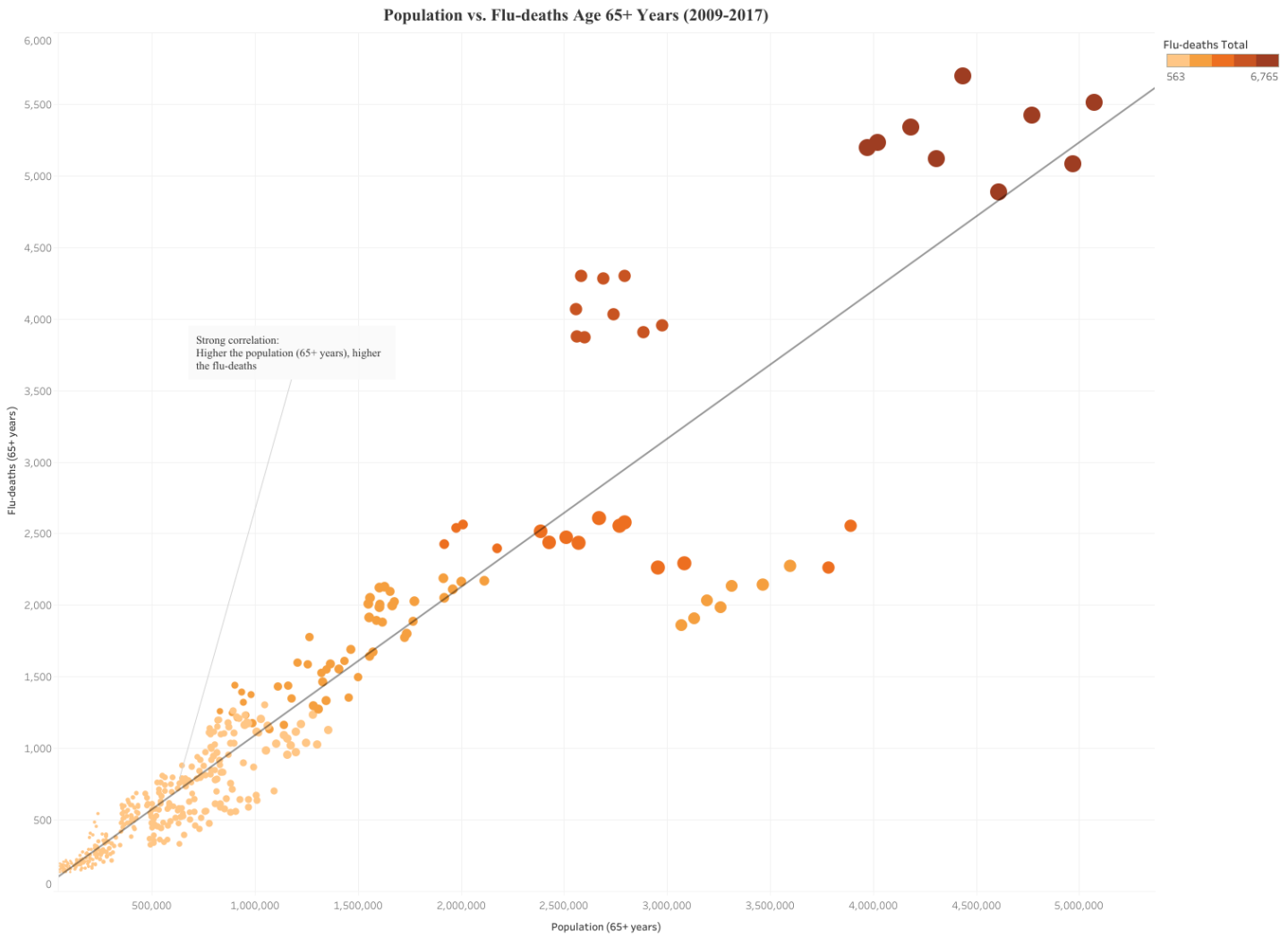
- Data cleaning, integration and transformation
- Statistical hypothesis testing
- Correlation analysis between different variable (flu-deaths, population, seasons)

Data

- Influenza deaths by geography (CDC), Population data (US Census Bureau)
- Influenza lab test result (CDC), Survey flu shot rates (CDC)

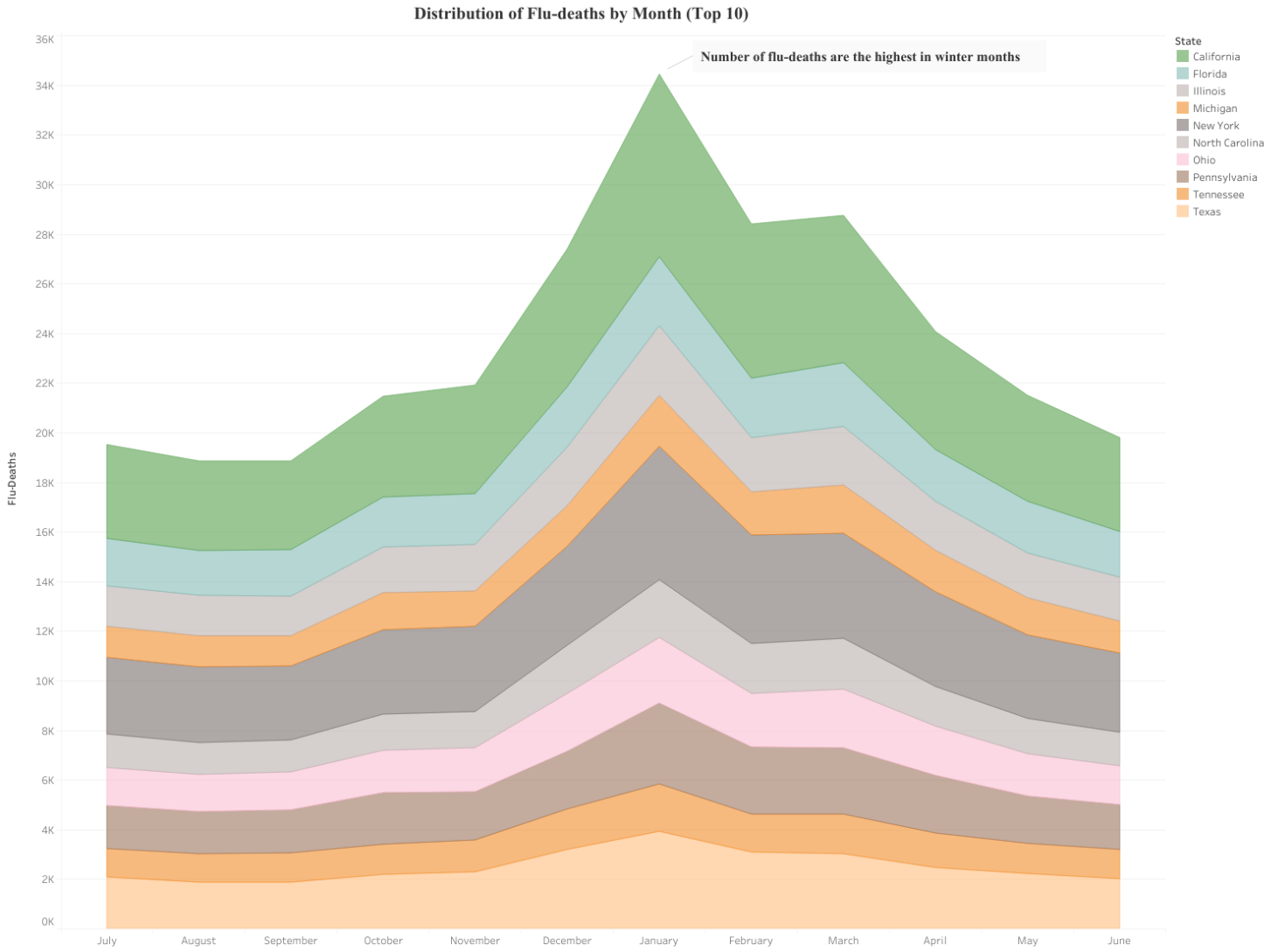
Analysis Tools

High Risk Population



- There is strong positive correlation between 65+ population and flu-deaths.
- This means, as the population aged 65 and above increases, flu-related deaths also tend to rise.

Influenza Death Peak Season



- Flu-related deaths consistently rise between November and March each year. This trend aligns with winter months, when colder weather leads to increased indoor gatherings and higher transmissions rates

Insights & Recommendations

- Vulnerable Population (65+) has strong positive correlation with flu-deaths → States with higher vulnerable population are significantly impacted by higher influenza related deaths
- Deaths vary form year to year but seasonal trends are highly predictable → Peak in winter months (November – March)

- Begin staff deployment in late October, especially in high-risk states like Wyoming, Vermont, Florida considering high flu-deaths and high 65+ population
- Maintain elevated staffing levels through March, aligned with peak flu season

Tableau Storyboard

YouTube Presentation

Project 3: Rockbuster Stealth LLC

Background

Rockbuster Stealth LLC, once a global leader in movie rental, is facing stiff competition from streaming services such as Netflix and Amazon Prime and planning to digitalize its service.

Objectives

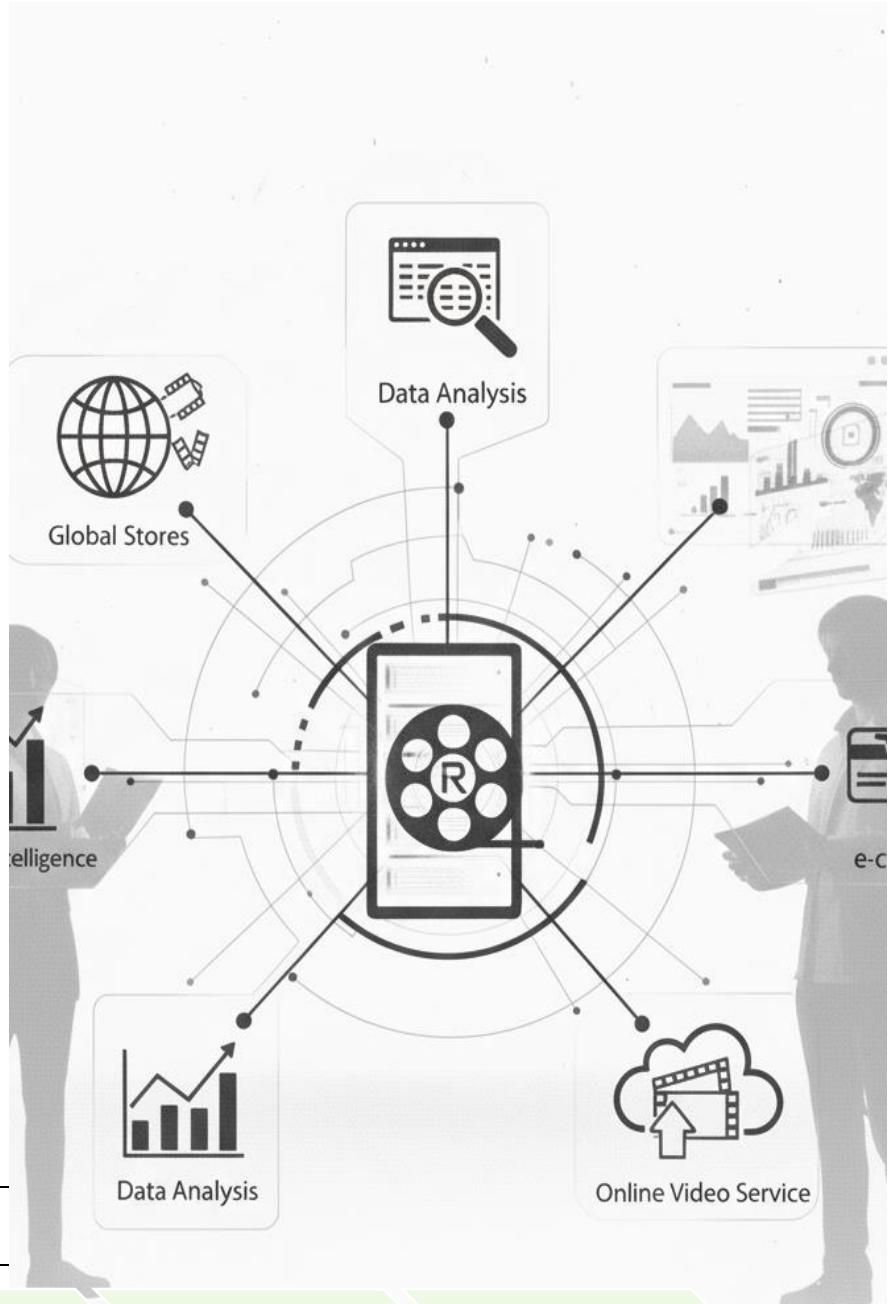
- Analyze movie revenue, rental duration, and customer distribution
- Identify high-value customers and compare sales across regions

Methods

- Load Rockbuster's dataset into PostgreSQL RDBMS
- Filter, clean, and summarize data with SQL
- Use SQL to query and extract insights

Data

- Rockbuster data set (Film inventory, customer & payment details, country & regional data)



Analysis Tools

Key Statistics

- Number of stores: 2
- Number of active customers: 599
- Customers based countries: 109

Maximum

- Rental rate: $4,99
- Rental duration: 7 days
- Replacement cost: $29.99
- Film length: 185 min

Minimum

- Rental rate: $0.99
- Rental duration: 3 days
- Replacement cost: $9.99
- Film length: 46 min

Average

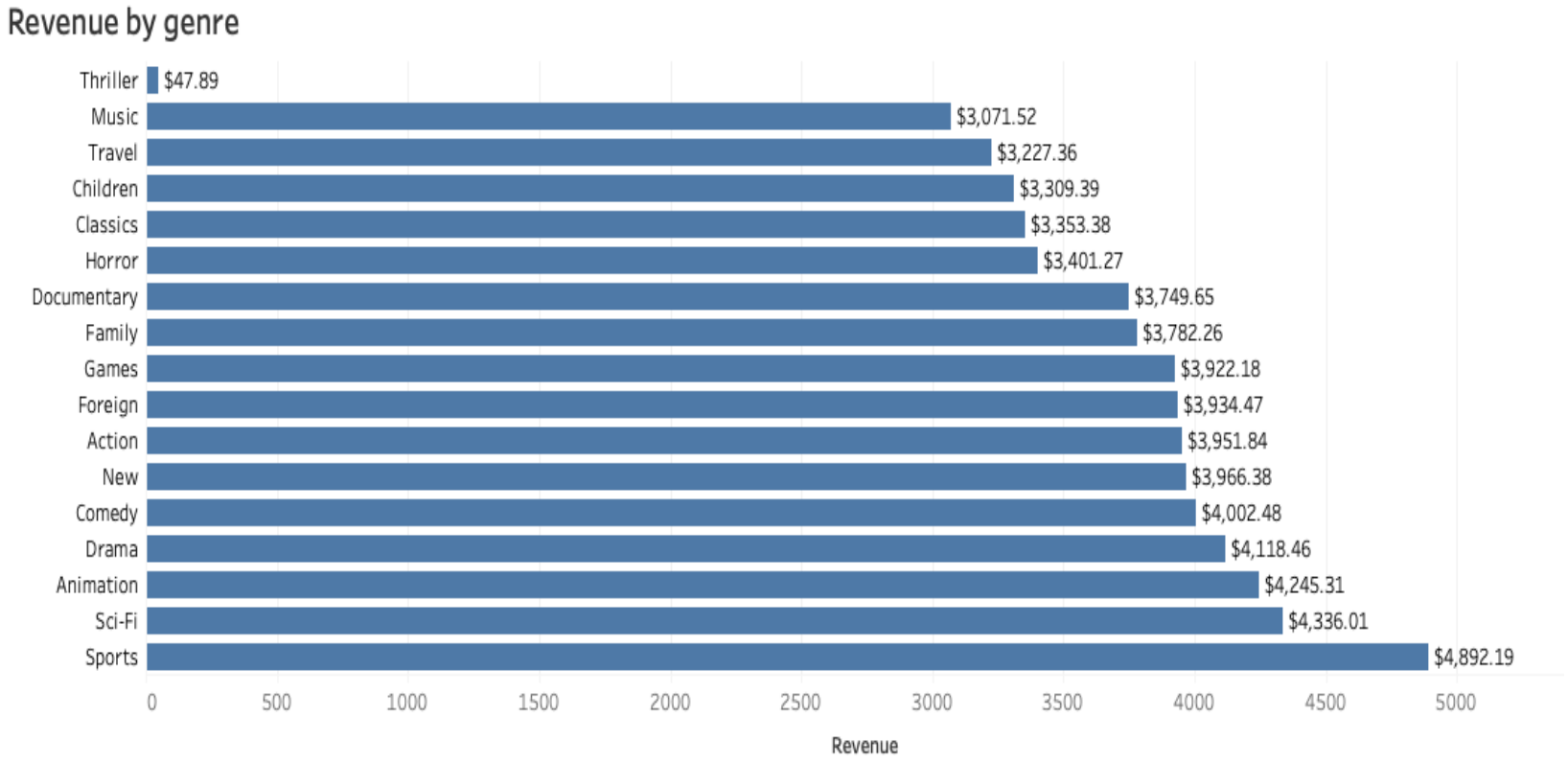
- Rental rate: $2.99
- Rental duration: 5 days
- Replacement cost: $19.99
- Film length: 115.272 min

Top 10 Countries by Revenue

- **India, China, the United States, and Japan** generated the highest revenues, while Mexico, Brazil, Russia, and the Philippines showed strong market presence.
- The map also highlights Turkey and Indonesia as promising markets for increased future investment

Link to Tableau

Revenue by Genre

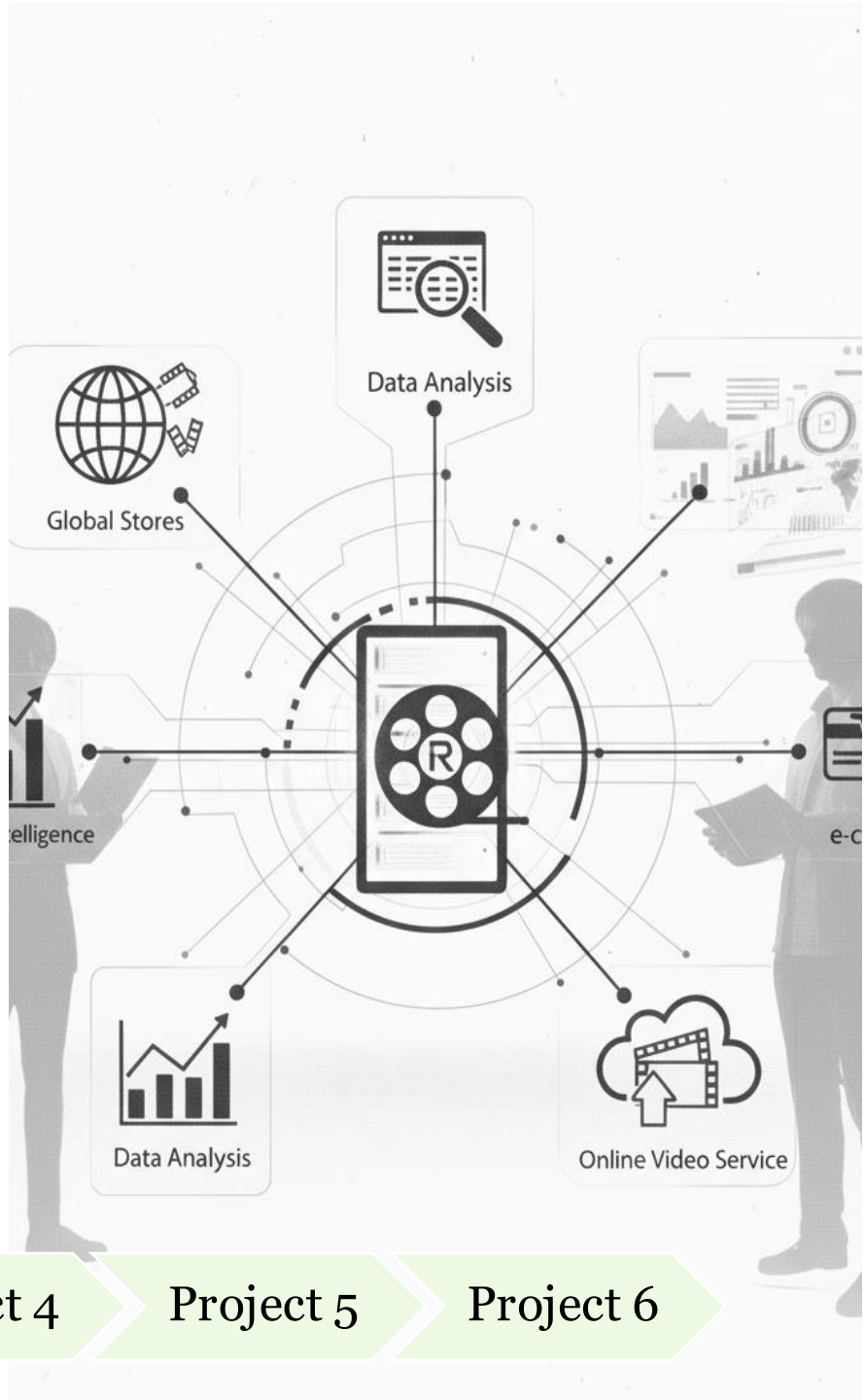


- **Sports, Sci-Fi, Animation** make highest revenue followed by Drama, Comedy and Action.
- Surprisingly, Thriller is the genre contributing to generate least revenue

Link to Tableau

Insights & Recommendations

- The **USA, China, and India** are top revenue generating-markets while **Japan, Indonesia, Mexico, Brazil, and Turkey** show significant growth potential
  - Focus marketing and contents on core markets to maintain revenue leadership
  - Invest in growth regions to capture emerging opportunities
- **Sports, Sci-Fi, Animation and Drama-Comedy** are most popular genres among the customers
  - Promote popular genres to attract and retain a wider audience
- Customers with high **Lifetime Value (LTV)** are primarily based in the **USA, Brazil, and the Netherlands**
  - Engage them by building loyalty and premium offerings
- **Refine Content Strategy:** Use insights from low-performing movies to reduce risk and optimize Rockbuster's portfolio



GitHub Repository

Project 4: Instacart Grocery Basket Analysis

Background

Instacart, an online grocery store, wants to uncover more information about the sales pattern to derive insights and suggestions for better segmentation.

Objectives

- Identify sales pattern such as peak shopping hours, days, spending behavior of customers
- Recommend strategies for targeted marketing to increase customer engagement

Methods

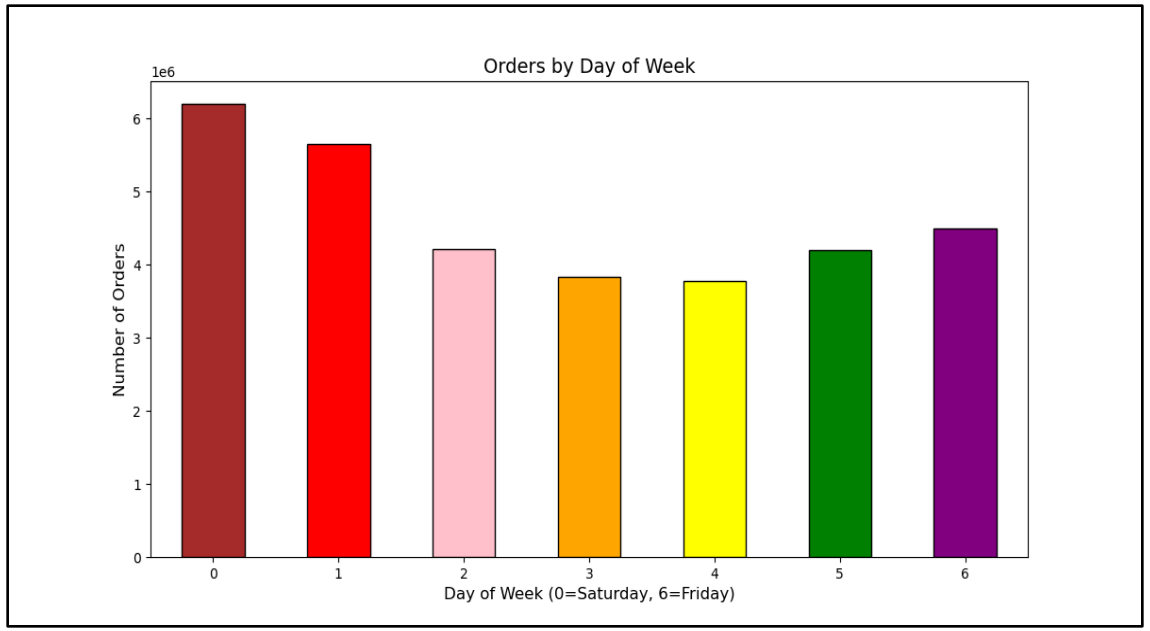
- Analyze sales data with Python (pandas, numpy, matplotlib, seaborn) and visualize
- Conduct data consistency, wrangling and derive new variable with Python
- Exploratory Data Analysis (EDA) to uncover trends and patterns

Data

- Customer dataset by CareerFoundry
- Instacart Data Dictionary

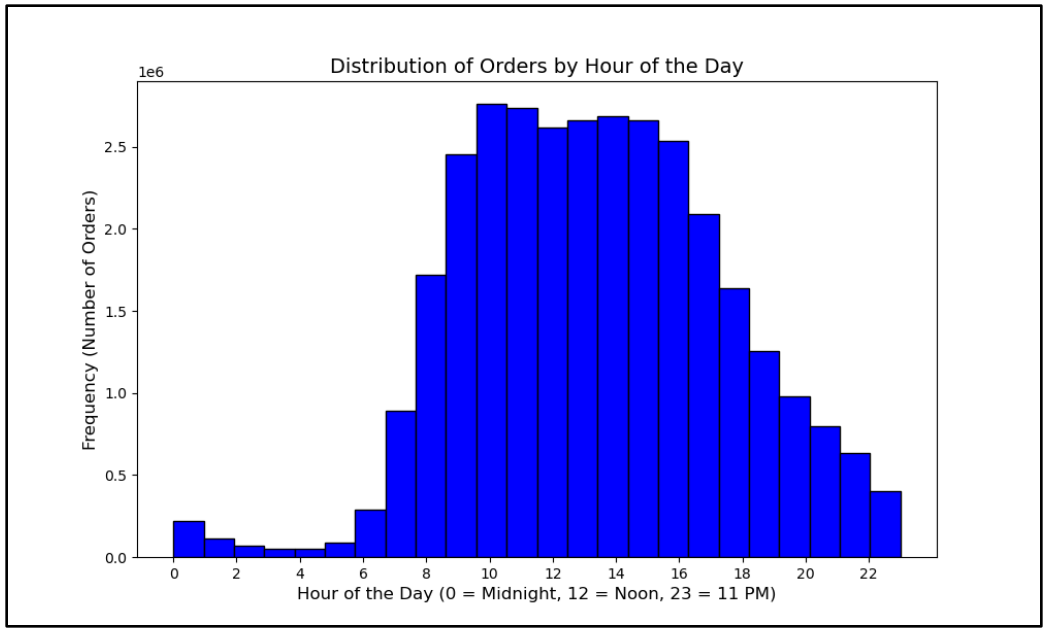
Analysis Tools

Busiest days of the week



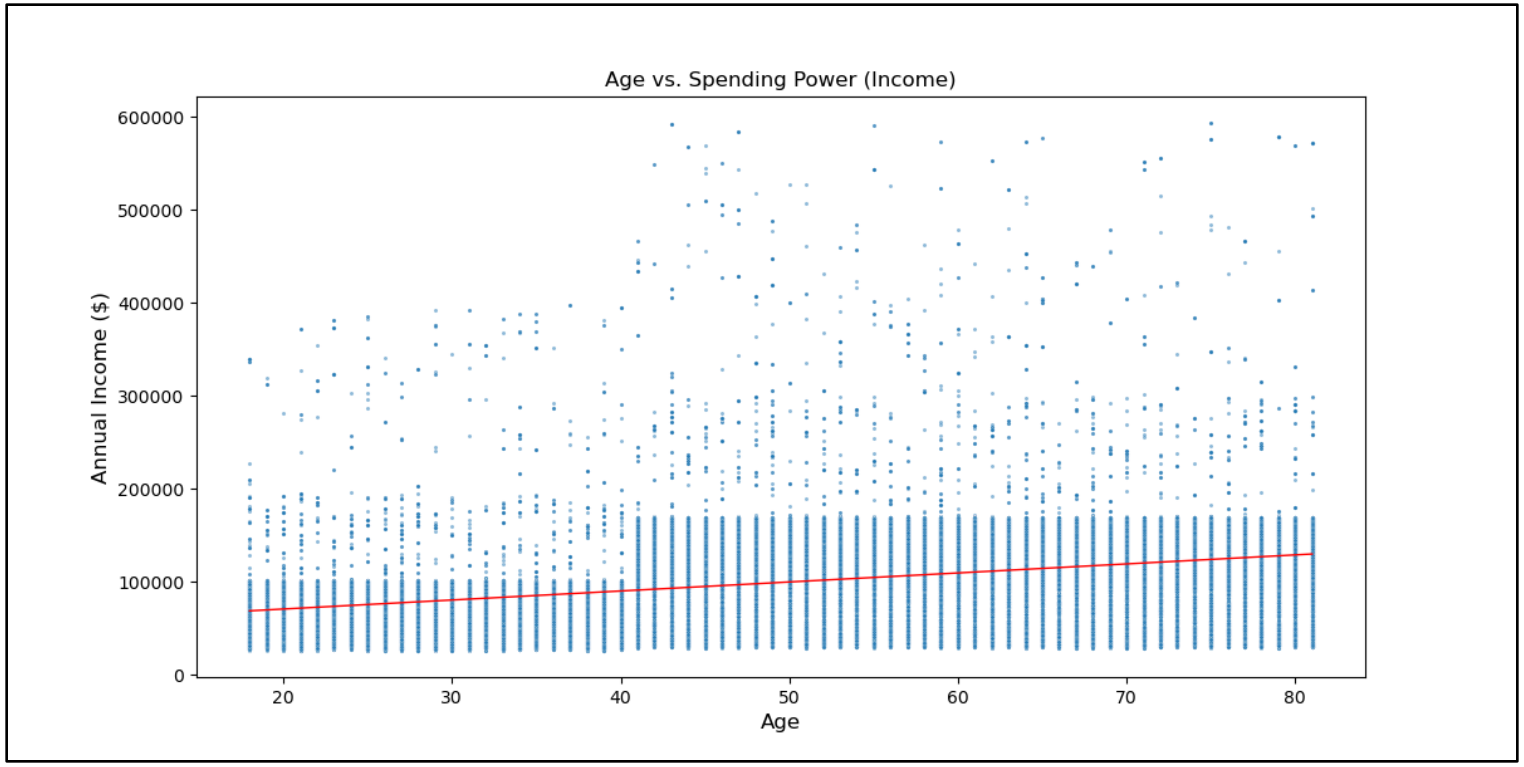
- Customers are more active on weekend (**Friday, Saturday and Sunday**) and place more orders, which is normal as expected in comparison on weekdays.

Busiest hours of the day



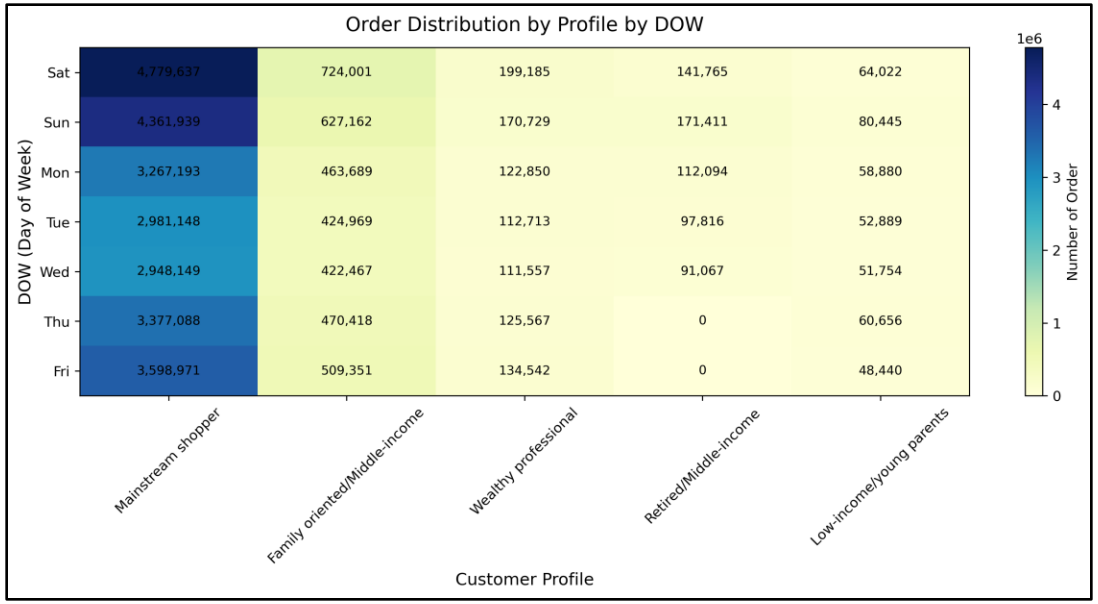
- Order frequency begins climbing steadily after 6:00 AM and reached the **morning peak** between 10:00 and 11:00 AM. The trend drops slightly and **remains stable until** 03:00 PM. Then the frequency of orders start to drop after 04:00 PM to late evening. As expected and normal, orders are at their minimum during early morning starting from late night.

Correlation between Age and Income



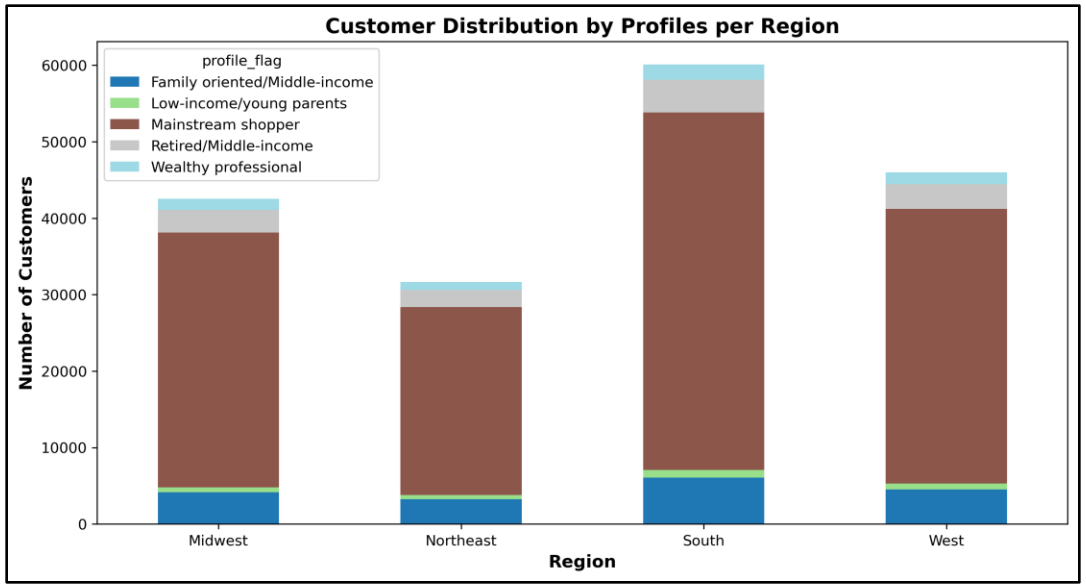
- Income shows weak positive correlation with age, with most customers earning under $200K regardless of age, though those over 40 demonstrate slightly higher average spending power.

Order frequency by DOW by Profile



- Mainstream and family-oriented shoppers drive weekend sales, particularly for produce and dairy, while retired and young parent profiles show strong Sunday purchasing patterns with consistent demand across snacks and beverages.

Customer by profiles per region



- Order distributions are slgihtly higher in South and West, where the "Mainstream shopper" makes a largest number of orders in all region.

Recommendations

- To maximize ad visibility and avoid competing with peak shopping times, ads should be scheduled on Tuesdays–Thursdays and during off-peak hours (before 7 AM or after 8 PM).
- Instacart can promote premium items in early/late hours and essentials in late mornings when spending is lowest.
- Target over-40s with premium products while keeping affordability for younger groups, using age-based personalized promotions.
- Launch weekend promotions for high-value segments, bundle top departments, and tailor Sunday-focused planning kits.
- Prioritize promotions around Produce, Dairy & Eggs, Snacks, and Beverages, while designing targeted campaigns to boost interest in underrepresented departments.

GitHub Repository

Project 5: Bank Churn Prediction

Background

To increase customer retention, Pig E. Bank sales team wants to identify the leading indicators that a customer will leave the bank.

Objectives

- Identify the leading indicators that a customer leaving the bank
- Perform a predictive analysis to identify the main risk factors contributing to client loss

Methods

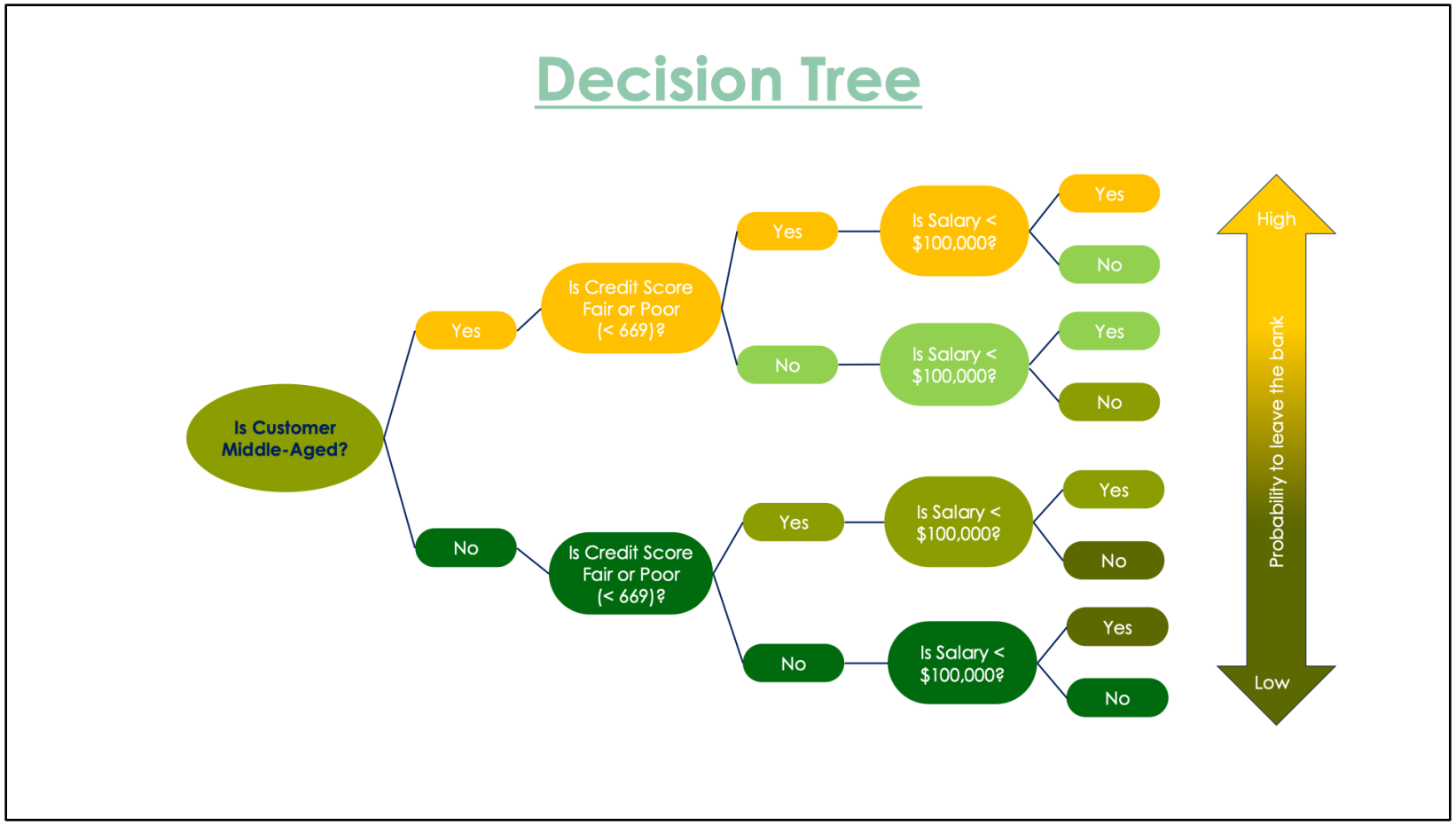
- Data Mining & Predictive Analysis conducted in Microsoft Excel
- Time-series forecasting

Data

- Internal Pig E. Bank client dataset containing demographics, account usage, salary, country and transaction details

Analysis Tools

Predictive Analysis



What are the leading risk factors to customer churn?

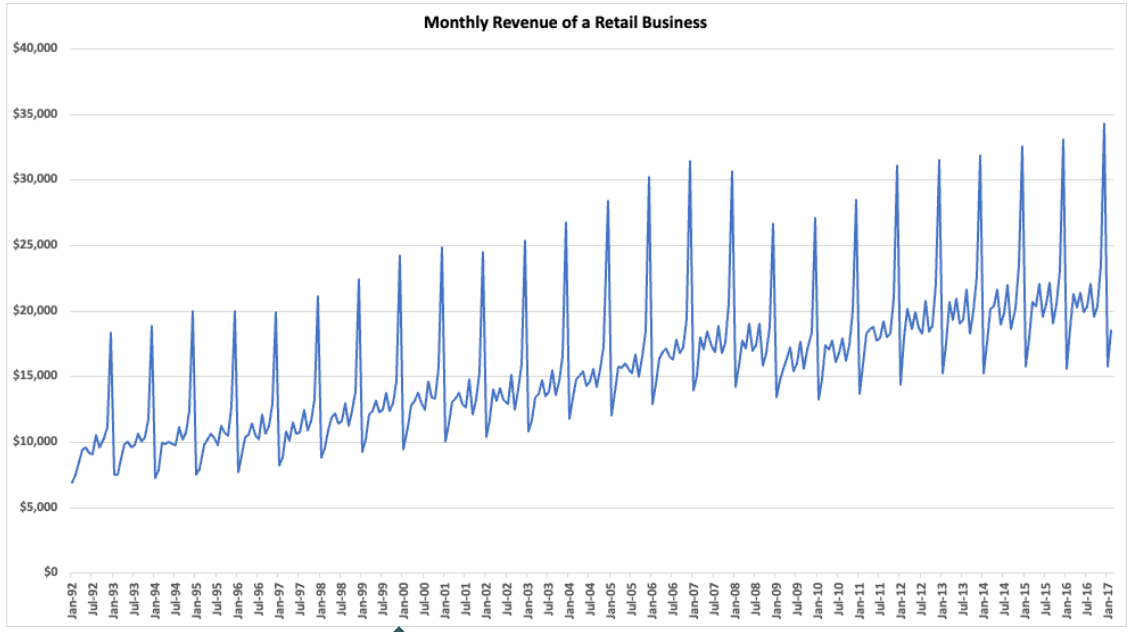
To increase customer retention, Pig E. Bank sales team wants to identify the leading indicators that a customer will leave the bank.

Leading Factors Contributing Client Loss:

- Age: Middle-Aged (13.62%)
- Salary: High, Low
- Credit Score: Fair, Poor

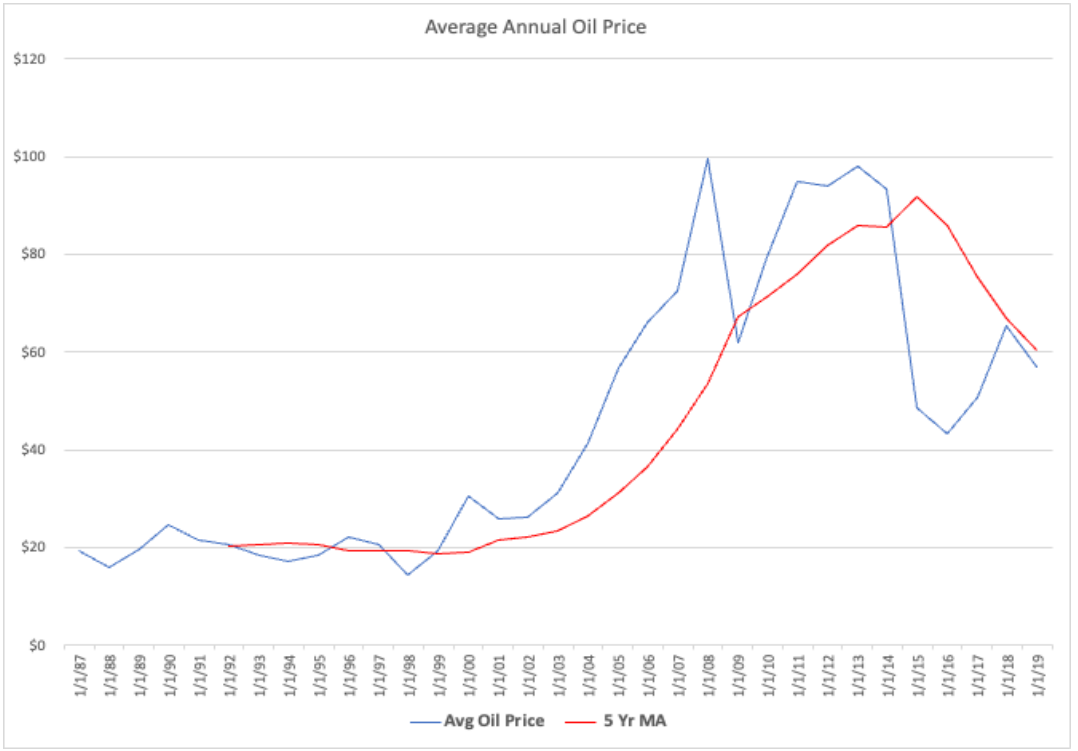
Time-Series Forecasting

Forecasting moving average



Time series (line chart) of monthly revenue

Time series of average annual oil prices with the five-year moving average (red line)



Project 6: US Traffic Accidents Analysis

Background

This project analyzes patterns and trends in U.S. traffic accidents from 2016 to 2023 to identify key factors influencing accident severity and frequency.

Objectives

- To identify major factors contributing to traffic accidents and their severity across different regions and time periods.
- To uncover trends and patterns that can support data-driven decisions for improving road safety and reducing accident rates.

Methods

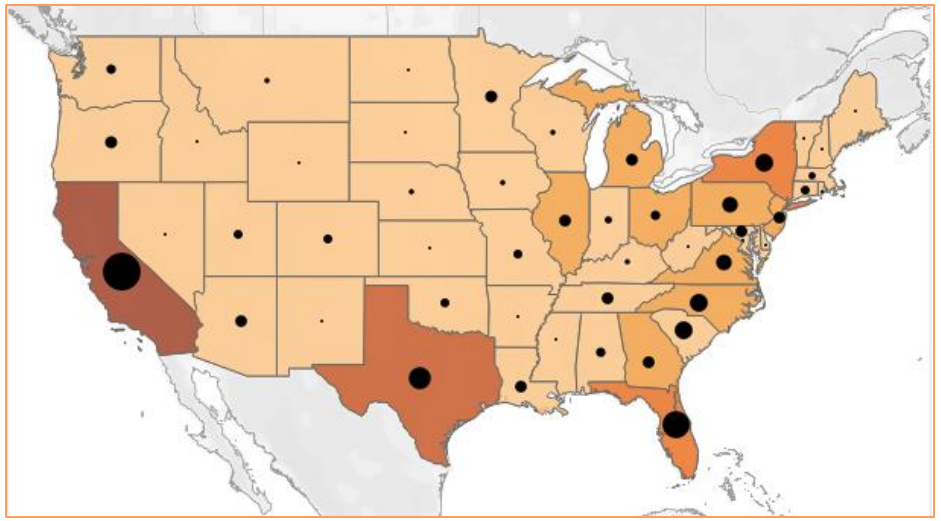
- Exploratory analysis with scatterplots, regression and clustering
- Choropleth map to detect significant accident patterns regarding accident frequency

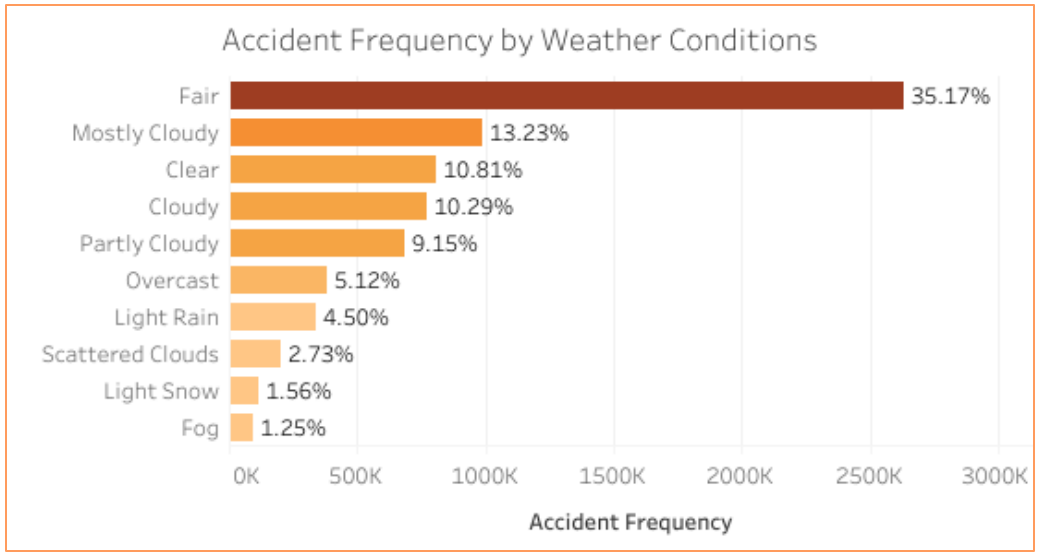
Data

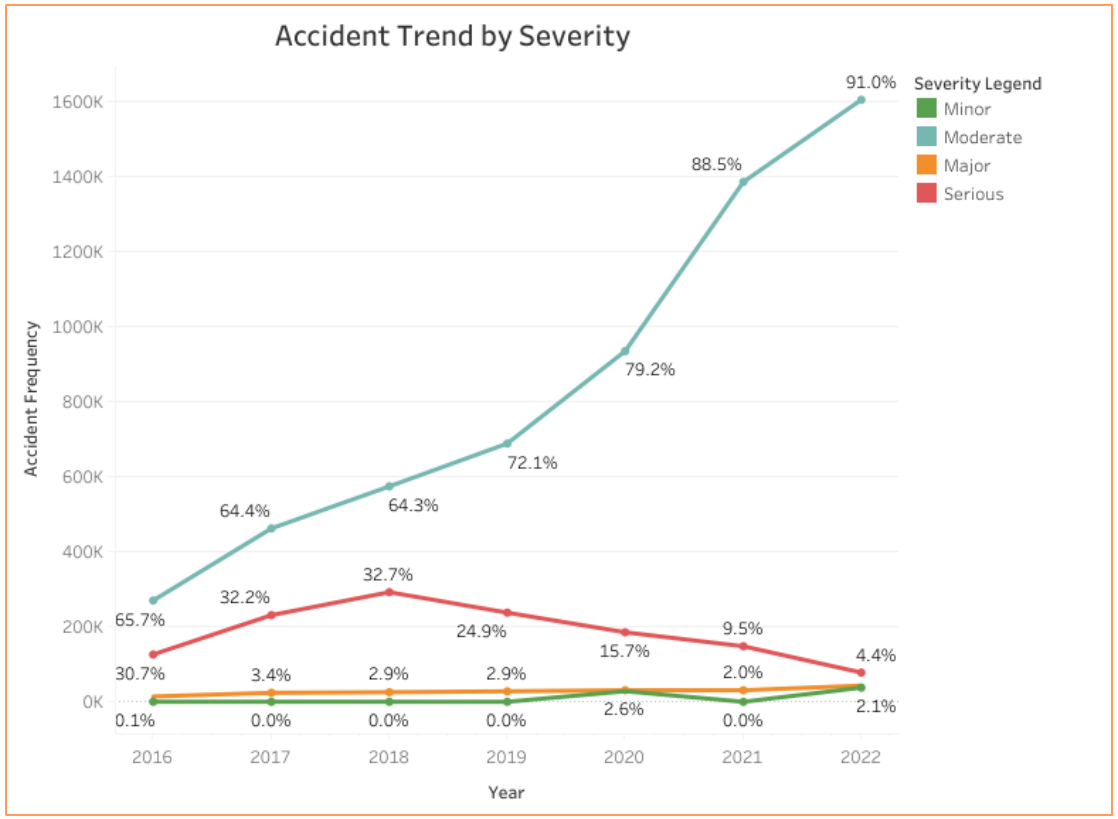
- Internal Pig E. Bank client dataset containing demographics, account usage, salary, country and transaction details

Analysis Tools

Exploratory Analysis



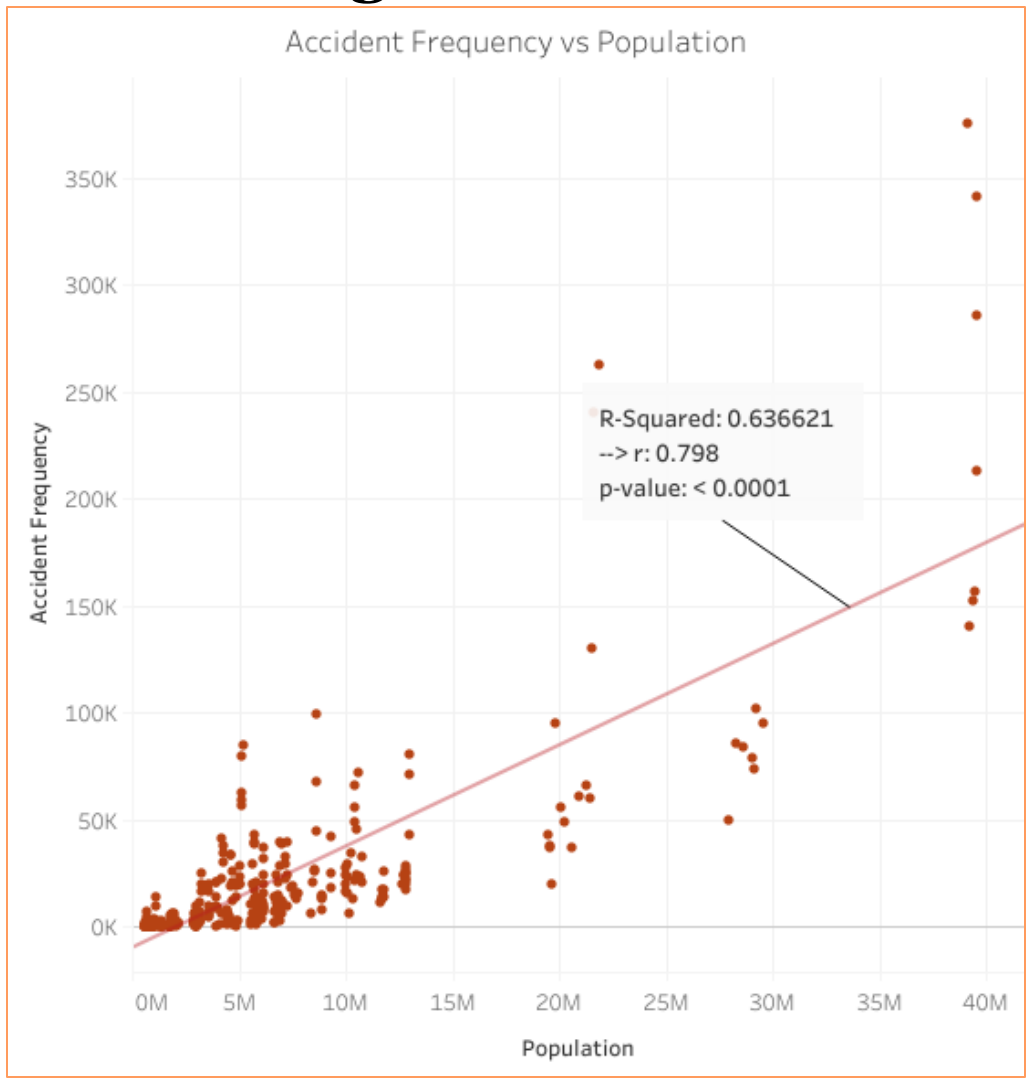


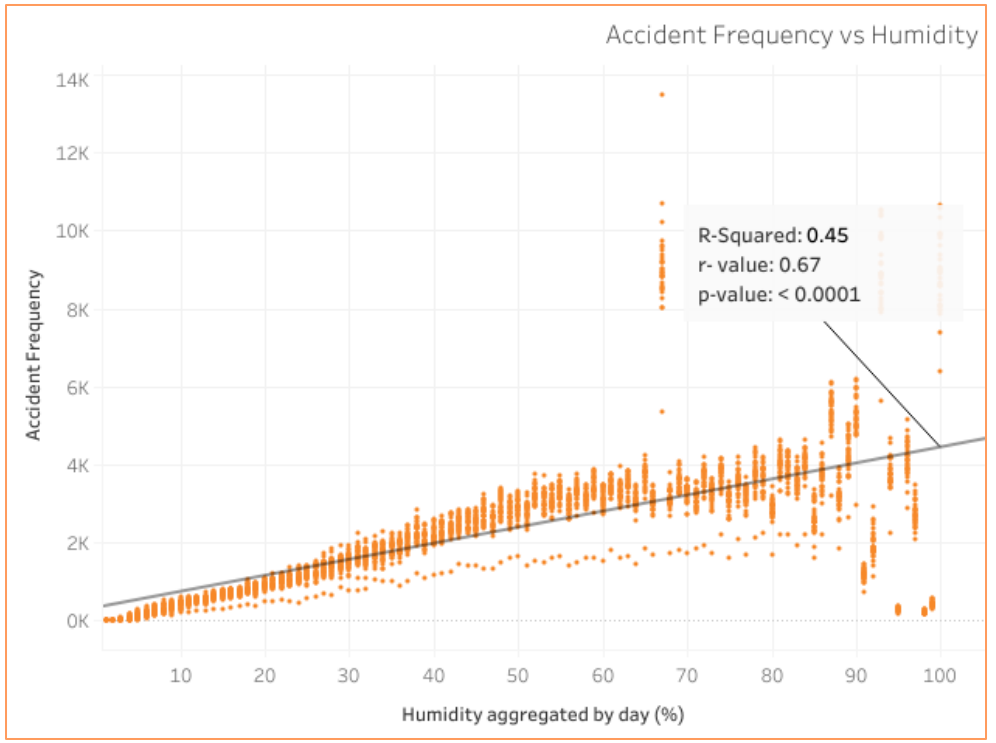


Key Insights:

- Most accident prone states: **California, Florida, Texas**
- Most accidents are **moderate severity**
- Most accidents occur in **fair weather conditions**

Linear Regression

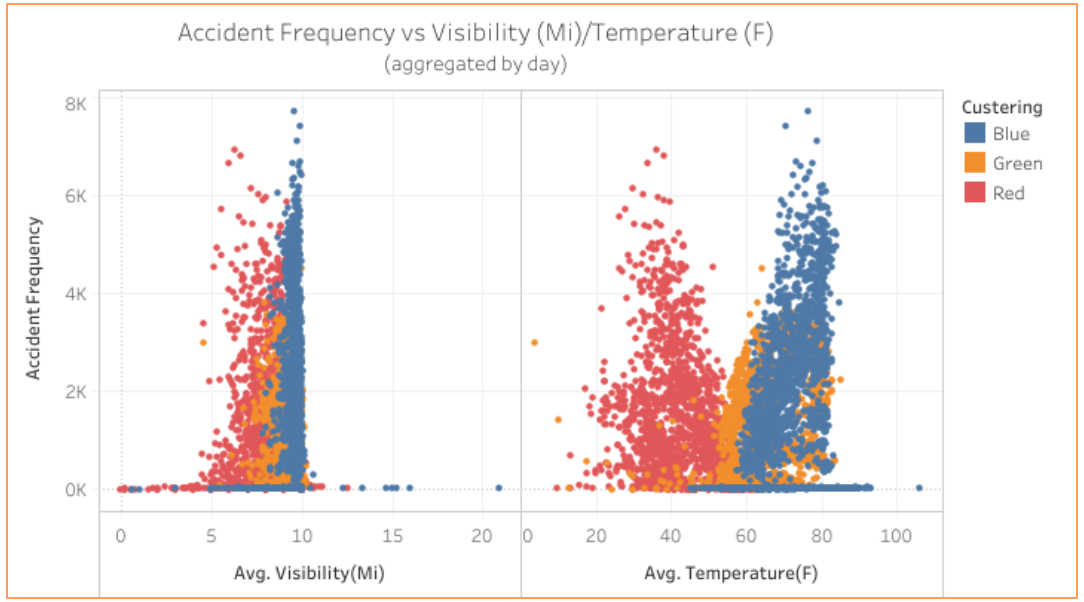




Key Insights:

- **Population → $R^2 \approx 0.64$:** About **64% of the variation** in accident counts can be explained by differences in population → Population strongly correlated with accident rate
- **Humidity → $R^2 = 0.45$**: humidity explains 45% of accident variation → higher humidity might be associated with higher accident frequency, deeper and other factors to be investigated

Cluster Analysis

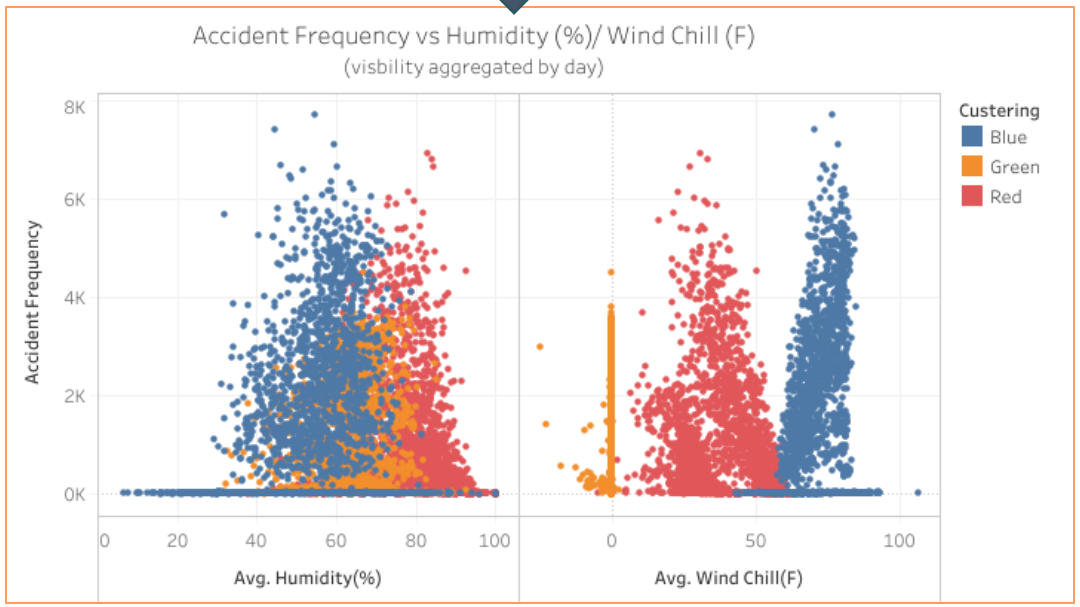


Key Insights:

- **Left Plot:** Periods of moderate visibility (ca. 5-10 miles) are associated with higher accident frequencies
- **Right Plot:** Accidents are significantly higher during both moderate (20-55F) & warm (60-90F) temperature, with relatively fewer accidents during extreme cold or heat.

Key Insights:

- **Left Plot:** Accidents are distributed unevenly across the humidity levels → different humidity levels might lead to different accident frequencies
- **Right Plot:** Accident rates vary distinctly across different wind chill ranges, revealing clear clusters that highlight how environmental conditions influence accident risk.



Key Results & Recommendations

Key Results

- Most accidents occur in **CA, FL, and TX**, which together account for about **41% of all cases**.
- **Morning** and **afternoon rush hours** show the highest **accident peaks**.
- Accidents are most common in **winter months** and under **fair weather conditions**.
- **Moderate visibility** and temperatures between 20–90°F see the highest accident frequencies.
- Accident frequency shows a **strong link to population and a moderate link to humidity**.

Recommendation

- **Targeted Interventions**: Implement specific safety campaigns and infrastructure improvements in CA, FL, and TX
- **Rush Hour Safety**: Focus on traffic management and enforcement during peak weekday commuting hours
- **Winter Preparedness**: Enhance winter road maintenance and driver advisories from November to February
- **Moderate Visibility Awareness**: Educate drivers on risks associated with moderate visibility conditions
- **Broader Environmental Analysis**: Investigate accident causes across diverse temperature and humidity ranges

Tableau Storyboard

GitHub Repository

CONTACT ME

anjanpakhrin@yahoo.com