Marta Majer

Data Analyst: Portfolio

info & projects:

LinkedIn

GitHub

Tableau

📞 +49 178 9141 396
📧 hello.marta.majer [at] gmail.com

Projects:

1. CDC Influenza Season - Predictive Workforce Allocation Strategy
2. Instacart - Exploratory Data Analysis, Buying Habits and Customer Profiling
3. Rockbuster Stealth - Market Analysis for Strategic Milestone Launch
4. GameCo. - Global Sales Market Analysis
5. Pig E. Bank - Data Mining for Customer Churn Prevention

CDC Influenza Season Project: Background

Motivation: During the influenza season, hospitals and clinics in the United States experience increased patient volumes, especially among vulnerable populations, who may develop serious complications. To address this, the medical staffing agency provides temporary staff to support healthcare facilities.

Objective: Determine the timing and number of staff to send to each state during the upcoming influenza season.

Scope: The agency covers all hospitals in the 50 states of the U.S.

Key Requirements:

- Provide data to support a staffing plan, focusing on the timing and spatial distribution of medical personnel across the U.S.
- Assess whether influenza is seasonal or year-round, and if seasonal, determine its onset and conclusion across states.
- Identify states with large vulnerable populations and categorize them as low, medium, or high need based on their vulnerable population counts.

Analytical Skills and Tools:

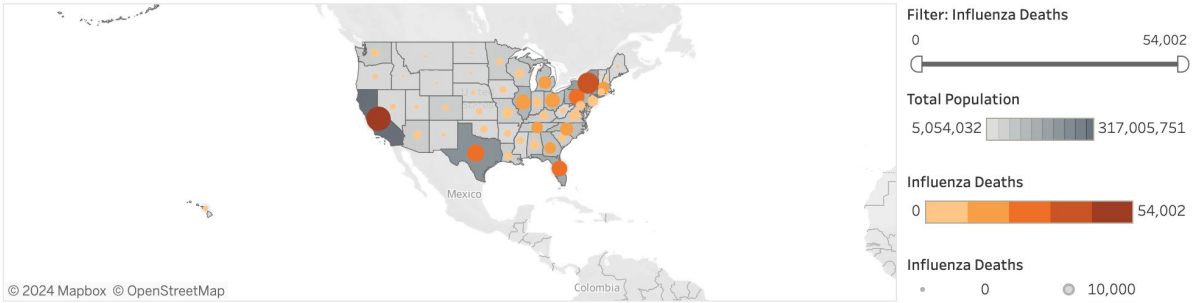
- **Formulate research** questions and **hypotheses** to guide the analysis.
- Assess available datasets for **relevance** and **limitations**, create data profiles covering types, **integrity** issues, and **cleaning** efforts, and **integrate** them into a cohesive dataset.
- Conduct **statistical analyses**, including variance, standard deviation, and **hypothesis testing**, to identify insights.
- Craft a **narrative** for research findings via a **Tableau** Storyboard.

Resources: Access detailed hypothesis testing in a GitHub repository.

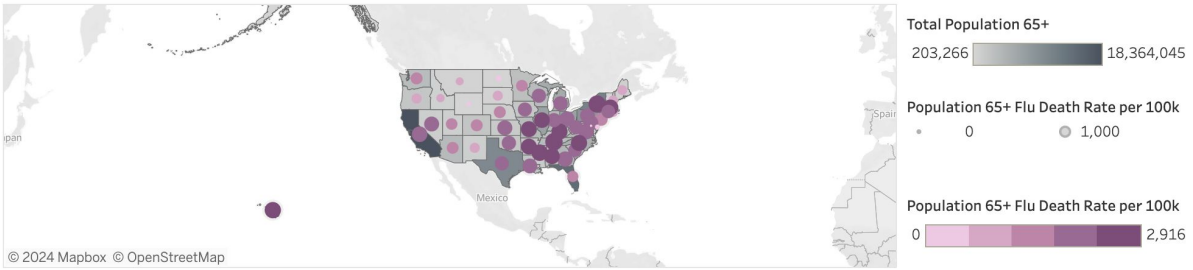
CDC Influenza Season

Influenza Impact: State Population Visualization in Tableau

Population and Influenza Deaths per State (2009-2017)



Population Aged 65+ and Their Influenza Death Rate per 100K by State (2009-2017)



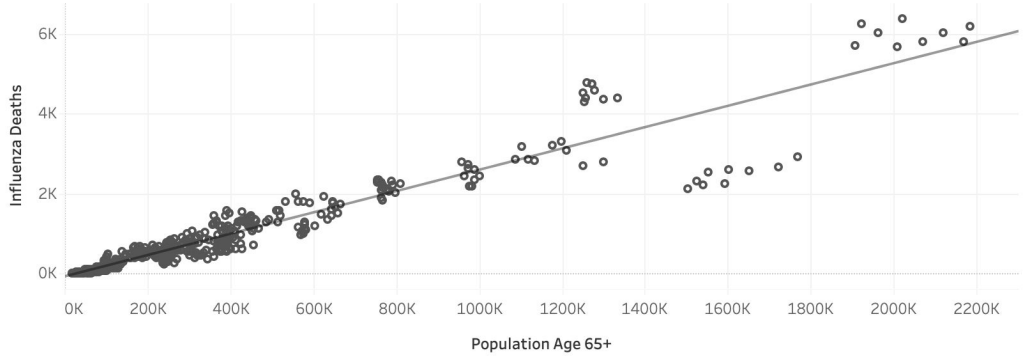
First Graph: Shows state populations with darker gray indicating larger populations. Circle size and color represent influenza death counts. Key findings: California, New York, Pennsylvania, and Texas have the highest mortality. The East Coast, Texas, and California are the most affected, with New York having significant deaths despite its population size.

Second Graph: Displays states with the largest populations aged 65+, in gray. Circle size and color represent death rates per 100,000 citizens. States with smaller vulnerable populations may show higher death rates, indicating a need for targeted healthcare. Notable states include Massachusetts, Tennessee, and Pennsylvania, while Texas and Florida, despite large vulnerable populations, do not have the highest rates.

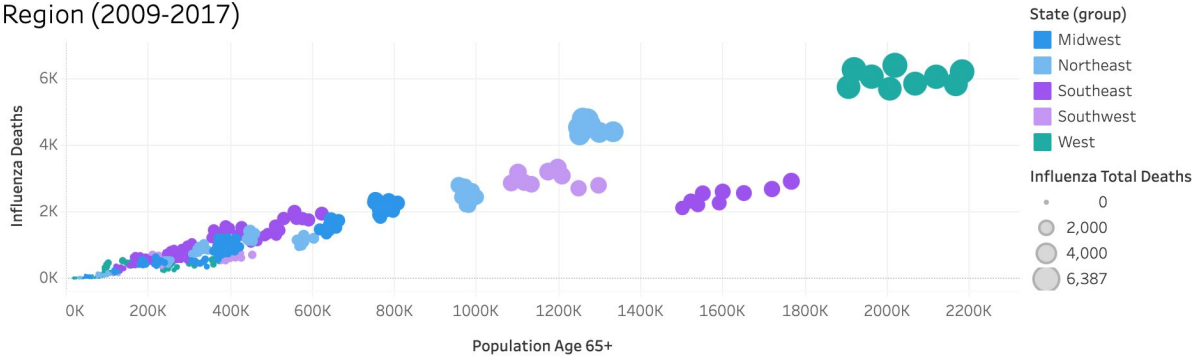
CDC Influenza Season

Understanding the Link: Elderly Population and Influenza Fatalities

Correlation Between 65+ Population Size and Influenza Death Count (2009-2017)



Correlation Between 65+ Population Size and Influenza Death Counts by State Region (2009-2017)



The correlation between the population aged 65+ and influenza death counts is significant, with an R-squared value of 0.9. This indicates a robust relationship, suggesting that increases in the 65+ population are strongly associated with higher influenza death count. This strong correlation underscores the importance of targeted healthcare resources for the elderly to mitigate influenza-related mortality.

Resources: Access full Tableau Storyboard here.

CDC Influenza Season
Key Insights and Recommendations

Key Insights

• Demographics: Statistically significant link between 65+ population and higher influenza deaths.

• Geography: High mortality in California, New York, Texas, and Florida. States like Massachusetts, Tennessee, and Pennsylvania have elevated death rates despite smaller 65+ populations, while Texas and Florida, despite large vulnerable populations, do not have the highest rates.

• Seasonality: Winter sees peak influenza deaths, with a notable drop in summer.

Recommendations

• Resource Allocation: Focus staffing and resources on states with large 65+ populations and high mortality. • Public Campaigns: Target flu prevention and vaccinations, prioritizing elderly and high-risk groups.

• Seasonal Strategy: Increase staffing in winter; use summer for training and preparation.

Next Steps

• Trend Monitoring: Continue tracking flu trends, demographics, and vaccination rates, and adjust plans based on data trends.

• Surveys: Conduct surveys with patients, medical staff, and hospitals to gather insights on staffing needs, patient care, and preparedness.

New Analysis Idea

Comprehensive Vaccination Impact Across Age Groups with Focus on Child Mortality

To assess the impact of vaccination status on influenza outcomes across all age groups, with a specific focus on evaluating flu-related mortality rates among children aged 0-5. The aim is to identify gaps and opportunities for improving vaccination strategies and reducing flu-related deaths in young children.

Instacart Online Grocery Store Project: Background

Overview: Instacart is an online grocery store operating through an app, seeking to uncover deeper insights into its sales patterns. The task is to conduct initial data analysis to derive insights and suggest segmentation strategies.

Key Questions:

- What are the busiest days and hours for orders to optimize ad scheduling?
- At which times do customers spend the most, informing product advertising strategies?
- How can simpler price range groupings help direct marketing efforts?
- Which product types are most popular, indicating the departments with the highest order frequency?
- What are the different customer types in the system, and how do their ordering behaviors vary?

Analytical Skills and Tools:

- **Jupyter & Python** libraries (`pandas, numpy, os, matplotlib.pyplot, seaborn, scipy`)
- **Data Wrangling & Subsetting**, **Consistency Checks**
- **Combining & Exporting Data**
  Merged multiple dataframes and analyzed merge flag frequencies. Exported results as pickle files.
- **Deriving New Variables**
  Created new columns using conditional logic, user-defined functions, `loc()`, and for-loops.
- **Sampling** data `np.random.rand()`
- **Grouping & Aggregating Data**
  Created flags (e.g., loyalty flags) and summarized data using the `groupby()` function.
- **Data Visualization with Python**
- **Coding Etiquette & Excel Reporting**

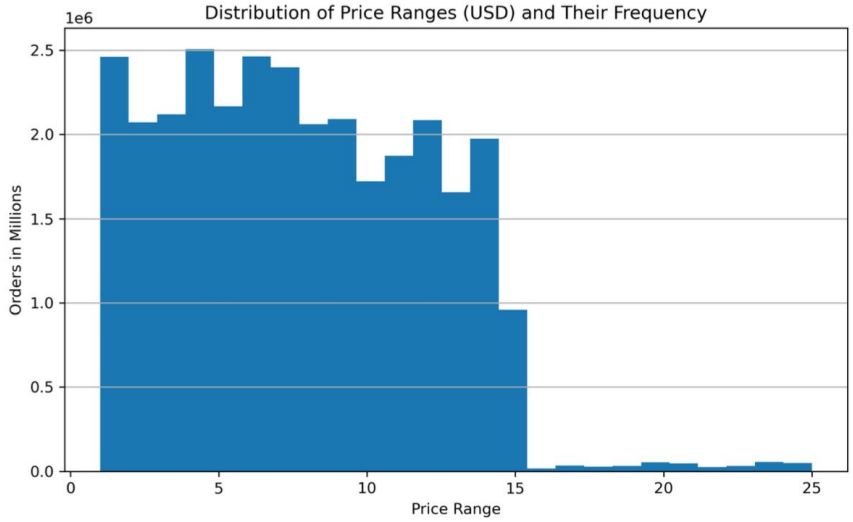
Resources: Access all notebooks and visualisations in a GitHub repository.

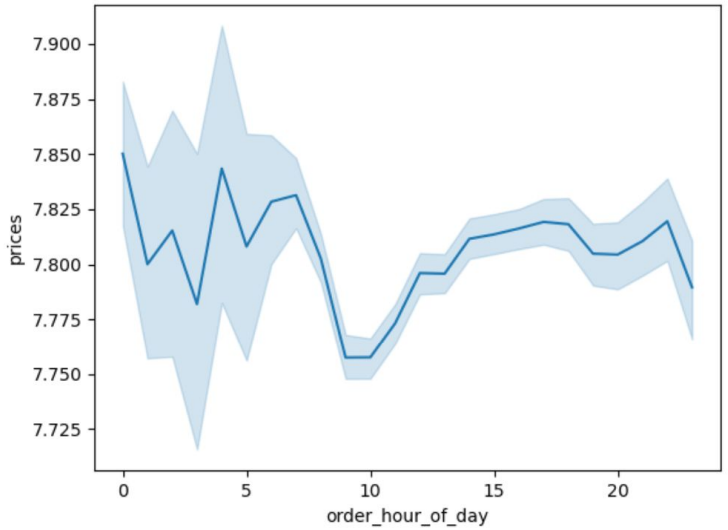
Instacart: Price Ranges and Spending Patterns: Timing and Strategy

Prices fluctuate (or rather are stable on average) between $7.75 and $7.90, peaking in early mornings and evenings. Highest average order prices are at night and in the evening; lowest around 10 AM. Most orders are under $15, with a sharp drop in frequency above this threshold. Instacart offers various products, including premium items, though these see less frequent purchases.

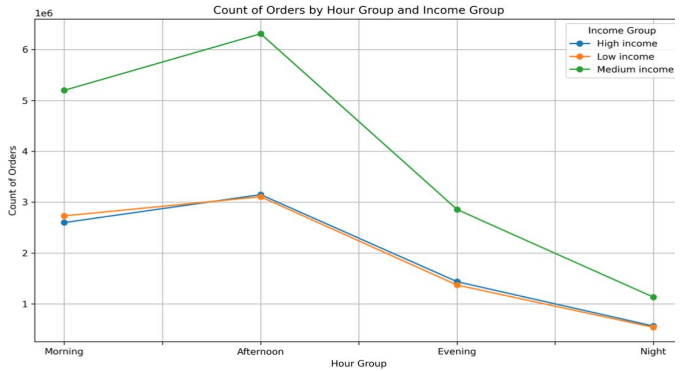
Ideas to think of:

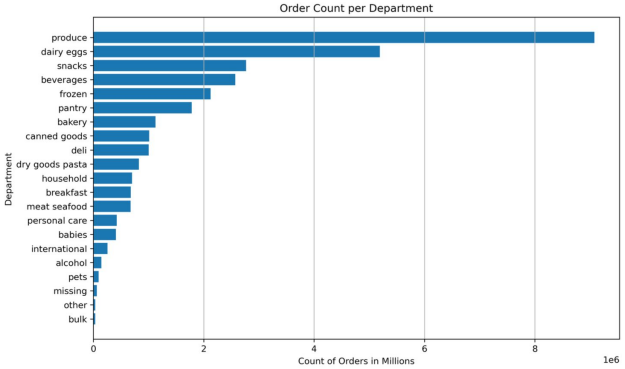
- Bundle products across price ranges to increase order value.
- Use tiered rewards to encourage purchases across all segments.
- Highlight items priced $0-$15 while promoting mid-range and premium products.

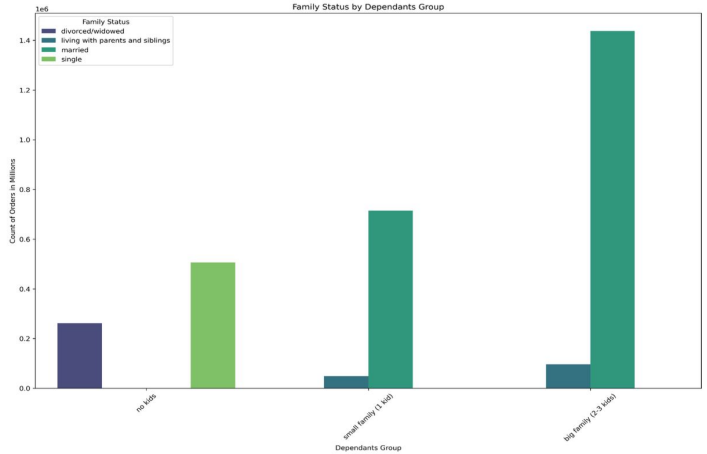


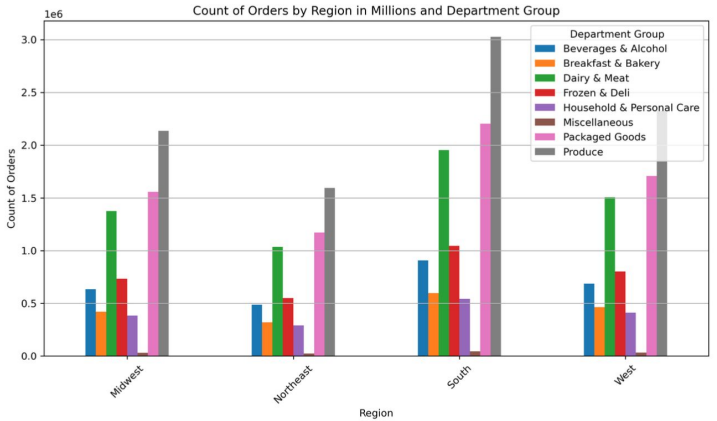


Instacart: Customer profiling based on age, income, purchase habits









Instacart Grocery Basket:
Customer Profiling and Behavioral Insights for Instacart

Key Findings and Insights:

- **Ordering Patterns:** Peak hours are 10 AM to 3 PM, with the highest activity on weekends. Early mornings (midnight to 6 AM) and Wednesdays are the least busy.
- **Spending Behavior:** Average order prices range from $7.75 to $7.90, peaking in the evenings and on weekends. Most orders are under $15, indicating strong price sensitivity.
- **Product Preferences:** Produce and Dairy & Eggs are the top departments, while Snacks and Beverages are moderately popular, suggesting a diverse product range.
- **Customer Segmentation:** Medium-income groups are the most engaged, particularly during afternoons. Families with children are the most active users, while singles and elderly adults prefer convenience items.
- **Marketing Recommendations:** Advertise premium products during evening hours and weekends, and consider bundling low-cost with mid-range items to boost order value.

Customer Profiling:

- **Demographics:**
  - Primary customer base includes married adults, particularly those with children.
  - Unique preferences noted among singles, young adults, and elderly customers.
- **Behavioral Insights:**
  - Families with dependents exhibit higher order frequencies, highlighting the importance of family-oriented marketing strategies.

Rockbuster Stealth Movie Rental Project: Background

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.

Key Questions:

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

Analytical Skills and Tools:

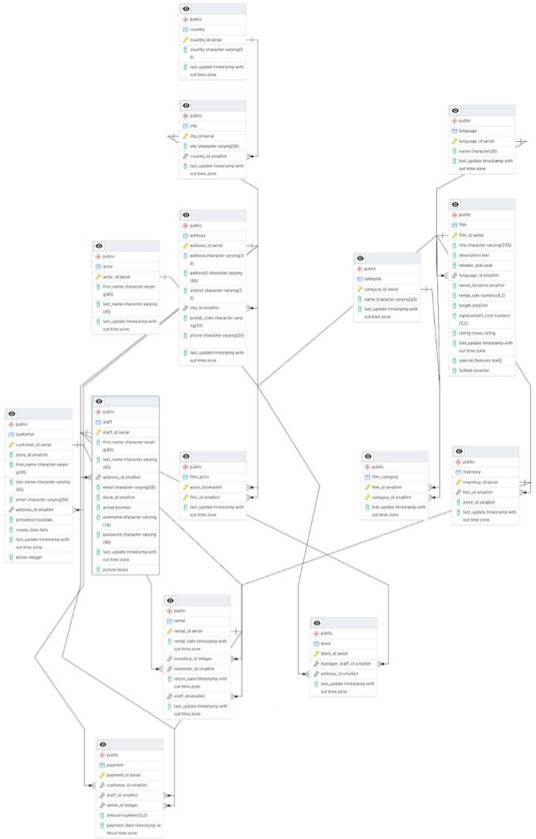
- Relational Databases: Understand OLAP, RDBMS, and PostgreSQL setup.
- Data Structure: Explain data types, schemas, and create data dictionaries.
- SQL Proficiency: Use CRUD operations, write SQL commands for analysis.
- Querying: Organize data with SQL, understand ETL processes.
- Filtering Data: Utilize WHERE, HAVING, and CASE statements.
- Data Cleaning: Identify and clean dirty data, create summary profiles.
- Joining Tables: Perform SQL joins and manage relationships.
- Subqueries & CTEs: Write and rewrite queries effectively.
- Presenting Results: Visualize SQL outcomes and present findings.

Resources: Access all queries and results in a GitHub repository.

Rockbuster Stealth: Data Dictionary Build-Up

Table of Contents

1. Entity Relationship Diagram (ERD)



2. Legend

Primary Key

Foreign Key

3. Tables

2.1 Fact Tables

Payment

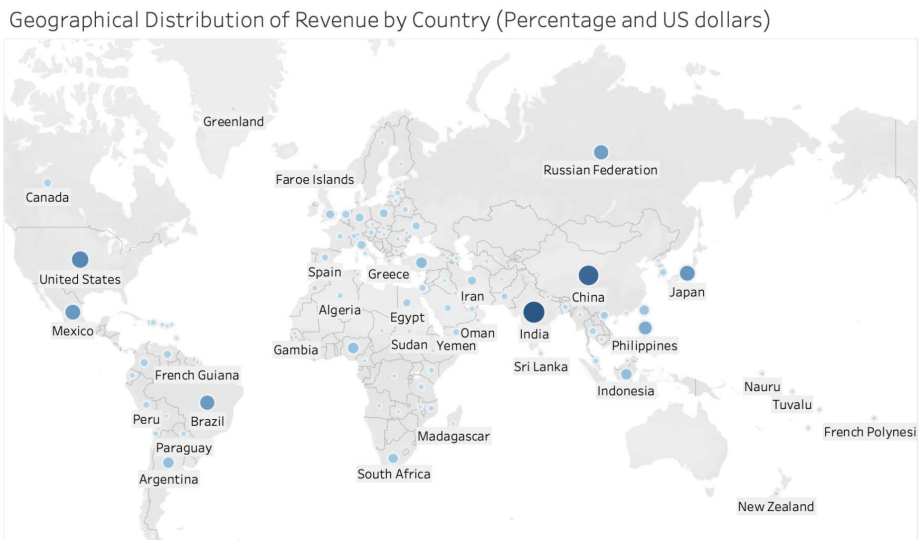
Key Type	Column Name	Data Type	Description
	payment_id	integer	Unique identifier for the payment
	customer_id	smallint	Identifier for the customer (foreign key)
	staff_id	smallint	Identifier for the staff (foreign key)
	rental_id	integer	Identifier for the rental (foreign key)
	amount	numeric	Amount of the payment
	payment_date	timestamp without time zone	Date and time of the payment

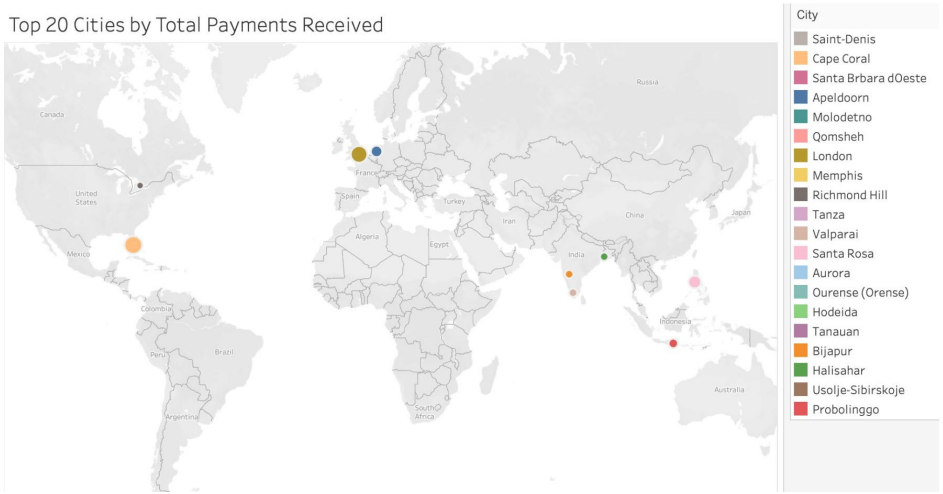
Rockbuster Stealth

Top Countries and High-Value Cities: Mapping Revenue and Customer Count Across Key Regions

The top 10 countries by both customer count and revenue are identical and include the following: India, China, United States, Japan, Mexico, Brazil, Russian Federation, Philippines, Turkey, and Indonesia.

6 out of the top 10 countries also appear in the list of high-value cities. The countries included in the high-value cities list but not in the top 10 countries are Réunion, Netherlands, Belarus, Iran, United Kingdom, Canada, Spain, and Yemen.





Resources: Access full Tableau Storyboard here.

Rockbuster Stealth
Loyalty Program Strategy: Maximizing Revenue & Retention

Key Insights:

- **Top Regions:** Asia (48.1%), North America (13.6%), Europe (13.5%) are key revenue drivers; loyalty programs here enhance retention and spending.
- **High-Value Cities:** Cities like London, Cape Coral, and Richmond Hill offer strong revenue potential; localized loyalty programs can unlock value.
- **Emerging Markets:** South America (11.5%) and Africa (10%) present growth opportunities; loyalty programs can foster customer loyalty and market share.

Recommendations:

- **Tiered Programs:** Implement tiered rewards in high-revenue regions to boost spending and engagement.
- **City-Specific Campaigns:** Launch tailored programs in high-value cities with local rewards and promotions.
- **Referral Programs:** Encourage acquisition in emerging markets through referral-based loyalty schemes.

GameCo. Video Game Project: Background

Objective

Conducting a descriptive analysis of a video game dataset to provide insights into market trends and player preferences, helping GameCo make data-driven decisions for developing new games.

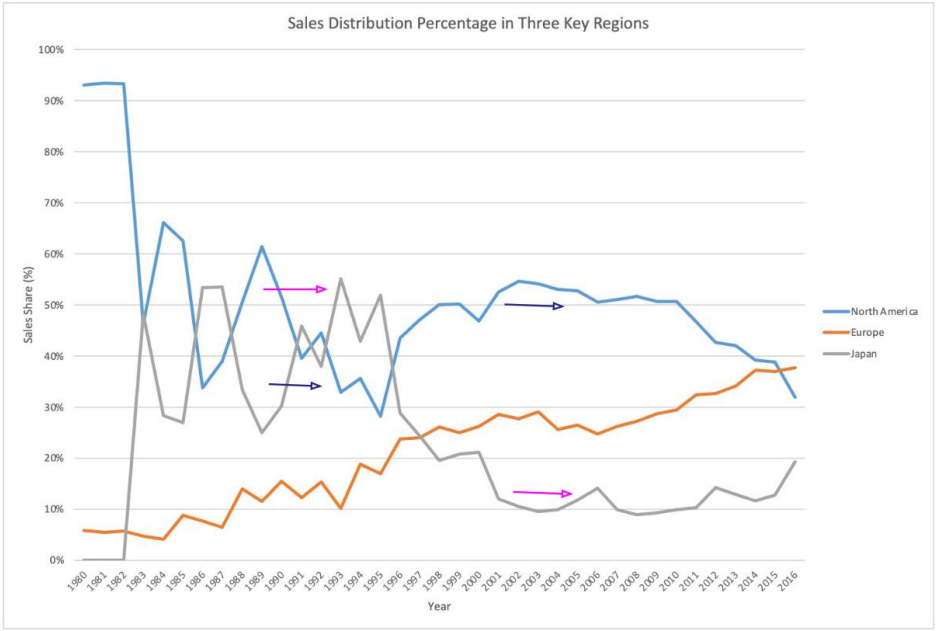
Key Questions:

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

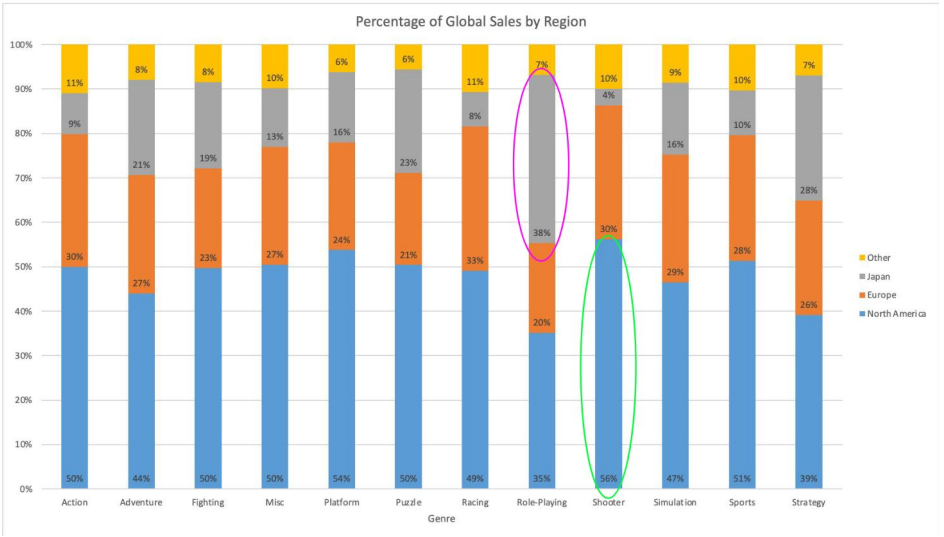
Analytical Skills and Tools:

1. **Data exploration**: accessing, sorting, filtering data (Excel)
2. Data **cleaning** and **preparation**
3. **Pivot** tables: grouping, summarizing, and transforming data (Excel)
4. Data **visualization**: bar/column charts, scatter plots, box plots (Excel)
5. **Descriptive analysis**: understanding distributions and trends
6. **Categorizing** and analyzing data characteristics and potential **biases**
7. Creating **new variables** and **filtering subsegments**
8. Analytical **decision-making**: selecting appropriate analysis for business challenges
9. **Storytelling** with data: presenting insights effectively

GameCo. Market Analysis: Correlations and Growth Opportunities



The analysis showed a negative correlation between North American and Japanese video game sales—when one market's share increases, the other's decreases. This suggests interconnected growth opportunities. Additionally, steady growth in the European market indicates potential for expanding investment to enhance GameCo's overall strategy.



Focusing on top genres in each region can boost GameCo's market penetration.

Role-Playing Games (38% of global sales from Japan) dominate in Japan, indicating a strong regional market.

Shooter Games (56% of global sales from the US) lead in the US, reflecting a significant market presence.

GameCo. Strategic Recommendations: Maximizing Market Reach

1. **Diversified Regional Focus:**

North American and Japanese markets show opposite trends. GameCo should balance focus between them to sustain global sales.

2. **Boost European Investment:**

Europe's steady growth offers a prime opportunity for increased investment and long-term returns.

3. **Genre-Focused Strategies:**

US: Shooter games dominate with 56% of global sales—focus here.
Japan: Role-playing games hold 38% of sales—emphasize this genre.

Pig E. Bank Project: Background

Overview: The objective of this project is to identify the leading indicators that suggest a customer is likely to leave the bank. By analyzing client attributes, the project aims to uncover key risk factors contributing to client attrition and model these factors using a decision tree. This analysis will help the sales team implement strategies to enhance customer retention.

Key Requirements:

- Use pivot tables and other Excel functions to identify the top 3 to 4 factors that lead to clients leaving.
- Gather and analyze statistical information on both groups (e.g., find averages, means).
- Determine the leading factors that contribute to client loss, based on your analysis of the data provided.

Analytical Skills and Tools:

- **Big Data Tools:** Software tools for handling structured and unstructured data (e.g., Hadoop, Spark, NoSQL databases).
- **Data Ethics:** biases
- **Data Mining:** data cleaning, descriptive statistics, proficiency in creating decision tree models.
- **Predictive Analysis & Time Series Analysis:**
  - Knowledge of time series characteristics and ability to create simple moving averages.
  - Differentiation between stationary and non-stationary time series.
  - Research skills in various time forecasting models.

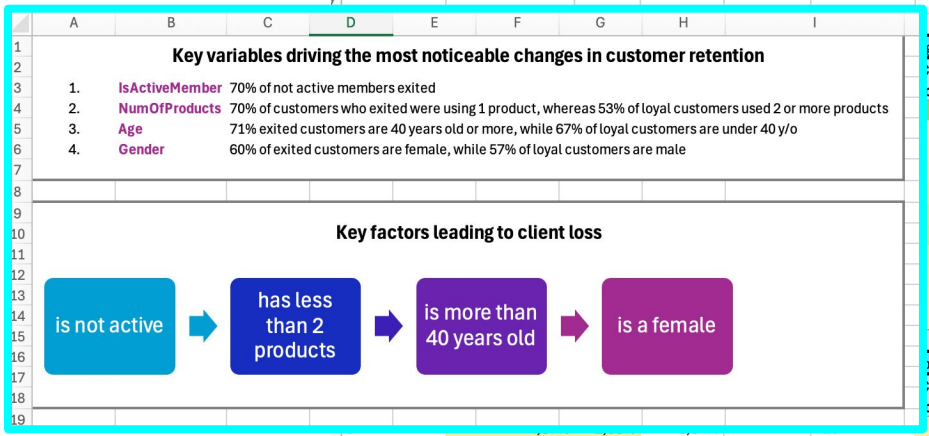
Resources: Access full data mining project on a GitHub repository.

Pig E. Bank:
Key Retention Factors by Customer Attributes

I analyzed various factors contributing to customer churn by using pivot tables. Several key factors, including customer activity level, number of products, age, and gender, stood out as potential drivers of customer loss.

Based on this, I created a decision tree model (next slide) to highlight the most significant factors affecting whether a customer is likely to exit. The tree illustrates how combinations of these factors either increase or decrease the probability of a customer leaving. For instance, inactive customers with fewer products and those over the age of 40 tend to have a higher likelihood of exiting the bank, whereas active customers with multiple products have a lower likelihood of leaving.

The model provides clear insights into which characteristics push customers toward exiting, allowing the bank to target retention efforts effectively.



comparing retention by ACTIVITY

by Activity	Column Labels		
Row Labels	loyal	exited	Grand Total
is not active	43,77%	70,10%	49,19%
is active	56,23%	29,90%	50,81%
Grand Total	100,00%	100,00%	100,00%

comparing retention by CREDIT CA

by Credit Card	Column Labels	
Row Labels	loyal	exited
does not have a	29,27%	29,42%
has a credit car	70,73%	70,58%
Grand Total	100,00%	100,00%

comparing retention by COUNTRY

by Country	Column Labels		
Row Labels	loyal	exited	Grand Total
France	51,27%	37,75%	48,48%
Germany	23,03%	36,76%	25,86%
Spain	25,70%	25,49%	25,66%
Grand Total	100,00%	100,00%	100,00%

comparing retention by GENDER

by Gender	Column Labels	
Row Labels	loyal	exited
Female	43,38%	59,32%
Male	56,49%	40,68%
Other	0,13%	0,00%
Grand Total	100,00%	100,00%

g retention by (account) BAL

Column Labels	
oyal	exited
37,28%	27,45%
0,25%	0,00%
1,91%	1,96%
5,22%	2,94%
9,16%	7,35%
14,38%	21,57%
16,41%	21,57%
8,78%	8,82%
5,09%	5,39%
1,53%	1,47%
0,00%	1,47%
100,00%	100,00%

mparing retention by AGE

Column Labels	
oyal	exited
0,38%	0,00%
5,21%	1,48%
13,23%	3,92%
22,39%	7,84%
26,72%	16,16%
15,64%	19,10%
6,87%	22,55%
2,80%	10,81%
2,80%	7,84%
1,14%	7,86%
1,40%	2,45%
0,76%	0,00%
0,38%	0,00%

1	9,29%	14,73%	10,41%	20-24	
2	10,56%	13,22%	11,11%	25-29	
3	10,31%	9,80%	10,20%	30-34	
4	8,65%	8,33%	8,59%	35-39	
5	10,17%	9,79%	10,09%	40-44	
6	9,29%	10,29%	9,50%	45-49	
7	9,93%	7,36%	9,40%	50-54	
8	11,20%	9,80%	10,91%	55-59	
9	11,06%	10,29%	10,90%	60-64	
10	5,47%	3,43%	5,05%	65-69	
Grand Total	100,00%	100,00%	100,00%	70-74	
				75-79	

Pig E. Bank: Customer Churn Prediction Model