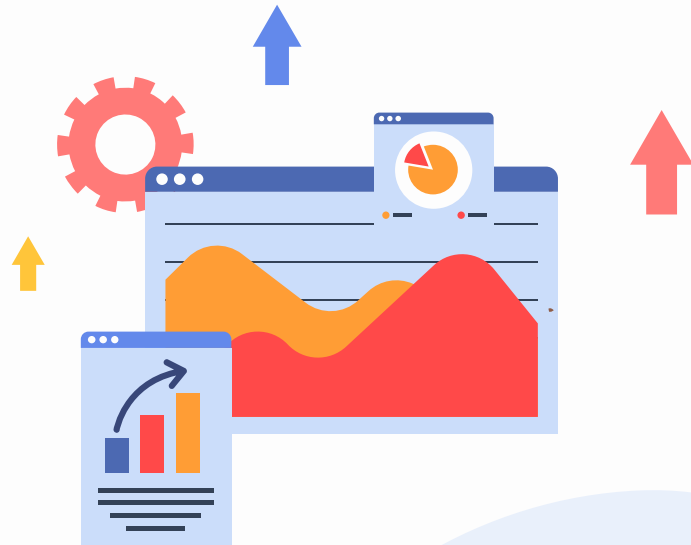


# Haoyu Zhao

## Data Analyst Portfolio



# About Me

Hello, my name is Haoyu Zhao. With a strong background in Business Administration and a passion for mathematics and logical analysis, becoming a data analyst felt like the natural next step for me.

My experience in marketing and procurement has honed my skills in problem-solving, decision-making, and working with data to drive results. I enjoy breaking down complex problems into clear, actionable insights. I'm excited to apply my analytical skills and contribute to the success of your organization. In the next slides, you will see a few projects related to my recent training in Data Analysis.



# Projects

## Game Co

Global video game sales analysis

Skills:



## Influenza

Allocation of medical staff to prepare for the influenza season

Skills:



## Rockbuster

Analysis of customer rental behaviors to support an online streaming service

Skills:



## Instacart

Analysis of customer demographics and purchasing patterns

Skills:



## Pig E. Bank

Customer retention analysis for a global bank

Skills:



## Happiness

Happiness factors worldwide uncover

Skills:





# Game Co

- Game Co is a new video game company, which wants to use data analysis to understand the market development to guide strategies and generate growth in the most sustainable markets.



## Objectives

- Analyse global sales trends over the past years
- Identify top-performing platforms and publishers to understand market competition



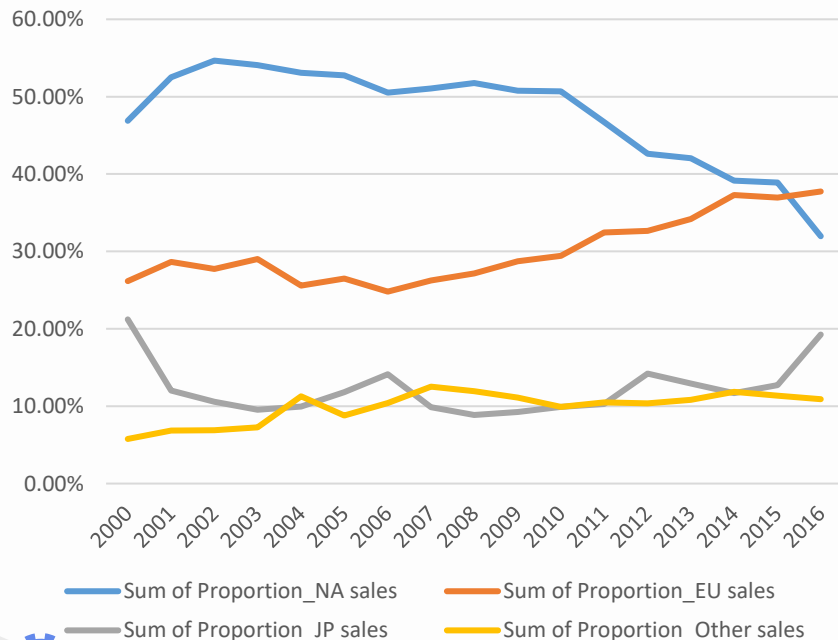
## Methods

- Grouping, Sorting and filtering data in Excel
- Data cleaning
- Descriptive analysis
- Data visualization



# Game Co / Analysis

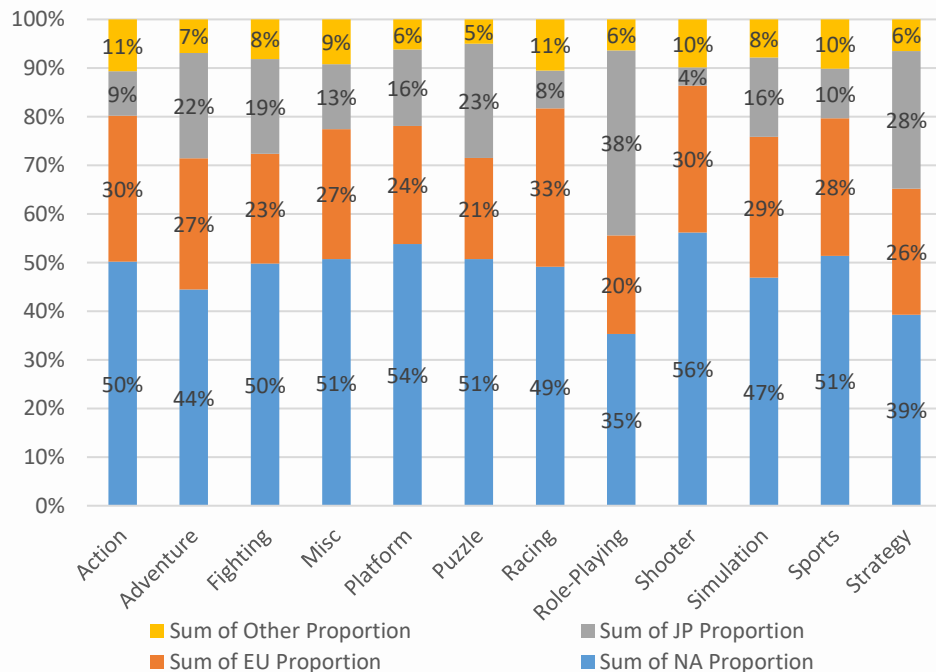
## Percentage of Regional sales by year



The data shows a significant shift in regional game sales: North America's share is declining, while Europe's is steadily increasing, and Japan's is seeing a notable rise.

The graph shows North America as the leading market across most genres, with Japan surpassing North American sales only in the Role-Playing genre by a noticeable margin.

## Regional Sales Propotion by Genre



# Game Co / Recommendations



Deliverables

## 01 Europe:

- Highest sales volume and steady growth
- Racing, Strategy and Sports genres are the most popular in EU

## 03 Japan:

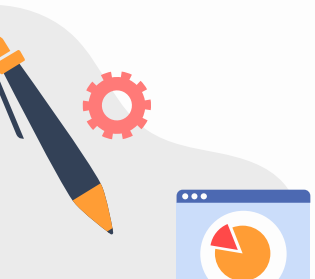
- Cultural affinity for gaming
- High Market Penetration

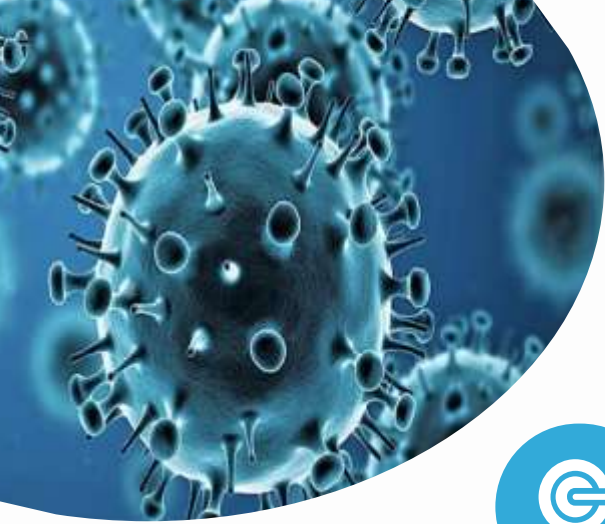
## 02 North America:

- Declined sales, to be the second market
- New marketing strategies and further analysis needed

## 04 Marketing Strategies

- Increase Investment in Europe
- Revise North American Strategy
- Reformulate the marketing strategy and targeted Campaigns in Japan
- Focus on top selling genres: Action, Shooter, Sports, and Role Playing.





# Influenza Season

- The U.S. flu season increases cases among vulnerable groups, driving hospitals to rely on staffing agencies for temporary healthcare personnel.



## Objectives

- Provide a plan for influenza season forecasting additional staff demand.
- Examine trends in influenza and use it to plan staffing needs across the country.



## Methods

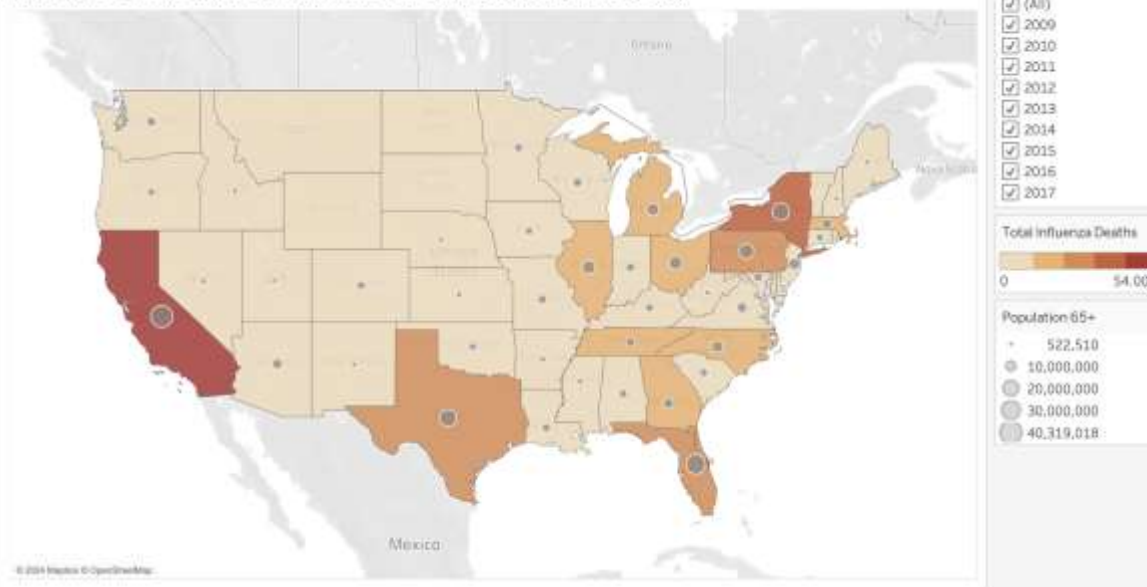
- Grouping, integration, profiling and hypothesis test
- Forecasting
- Data visualization and storytelling





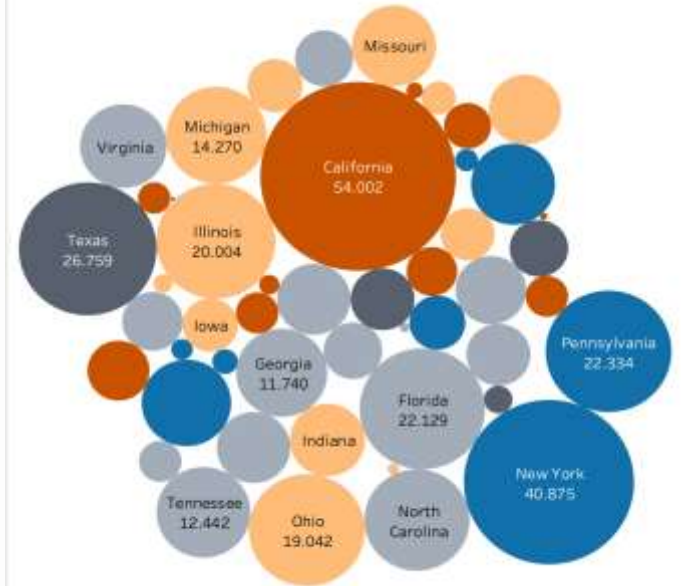
# Influenza / Analysis

Influenza Deaths and Population 65+ in U.S. by State (2009-2017)



The influenza-related deaths are concentrated just in 10 states, mostly of which are from West and Northeast regions.

Influenza related Deaths by state in U.S. (2009-2017)



Total Influenza Deaths



Region Group

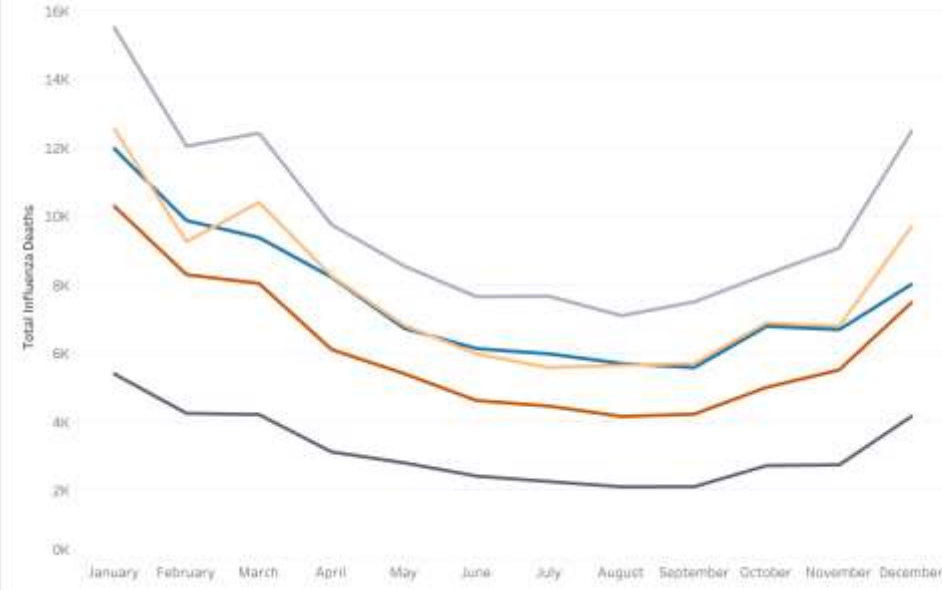






# Influenza / Analysis

Seasonality of Influenza Deaths in U.S. by region (2009-2017)



Region Group

- West
- Midwest
- Northeast
- Southeast
- Southwest



## Comparative Insights:

- We can clearly see the influenza deaths of all the region groups show a clear peak during the winter months, particularly January through March.
- There's a consistent pattern where influenza-related deaths increase during the winter months and decrease in the summer months, reflecting the seasonality of the influenza virus.



# Influenza / Recommendations

## 01 Prioritized Vaccination for Vulnerable Populations

Focus vaccination efforts on populations at the highest risk of influenza complications, particularly individuals over the age of 65 and children under the age of 5.

## 02 Enhanced Resource Allocation in High-Risk Regions

- West and Northeast states show higher deaths counts.
- Efficiently distribute healthcare resources to regions with historically higher influenza-related fatalities, particularly in states with older populations and healthcare shortages.
- Monitor key healthcare resources during the peak months (December to March), ensuring healthcare facilities are prepared to handle increased patient loads.

## 03 Real-Time Monitoring and Regional Support

- Strengthen surveillance systems for timely data collection and share insights with healthcare facilities for quick responses.
- Allocate additional support to regions that experience unexpected surges.



Deliverables



# Rockbuster Stealth

Rockbuster Stealth LLC, a former global movie rental chain, plans to launch an online video rental service to remain competitive against streaming giants like Netflix and Amazon Prime.



## Objectives

- Deliver data-driven insights to help the Rockbuster Stealth LLC management team strategically launch their new online video rental service.



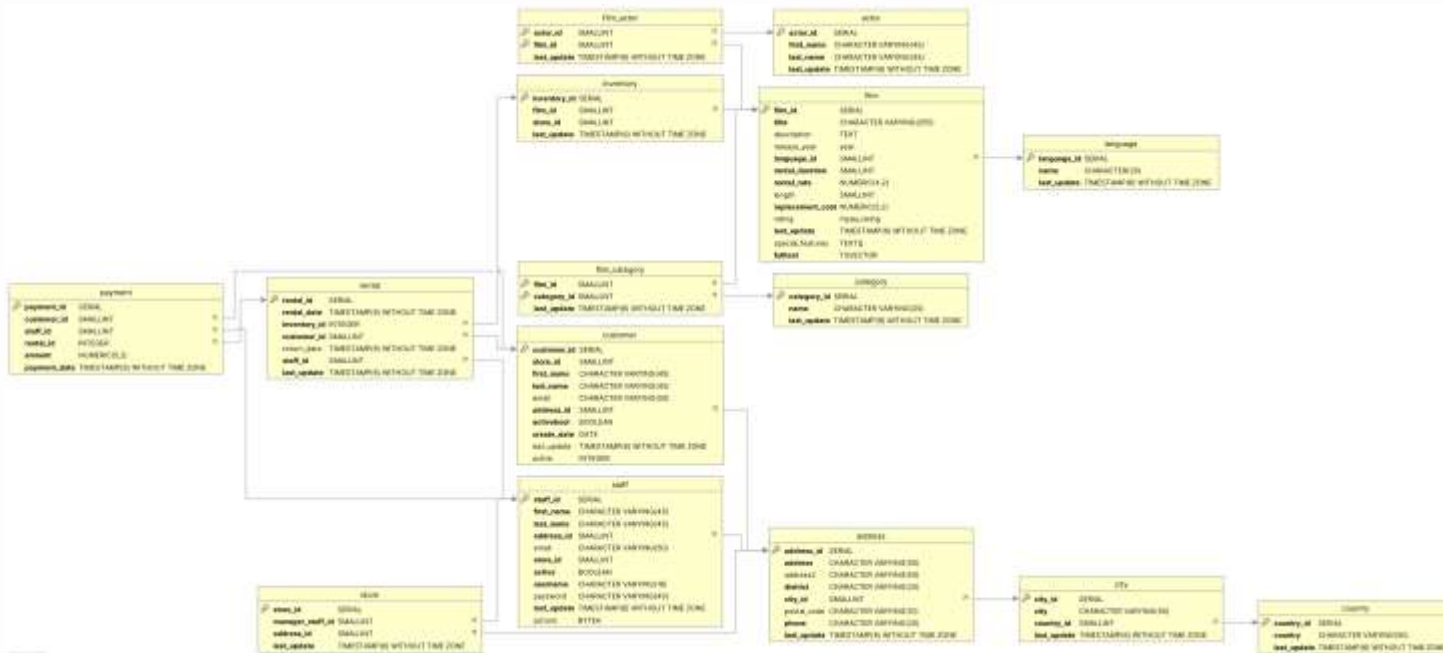
## Methods

- PostgreSQL
- Data Cleaning & Filtering
- Joining Tables
- Subqueries & CTEs
- Data Visualization





# Rockbuster / Analysis



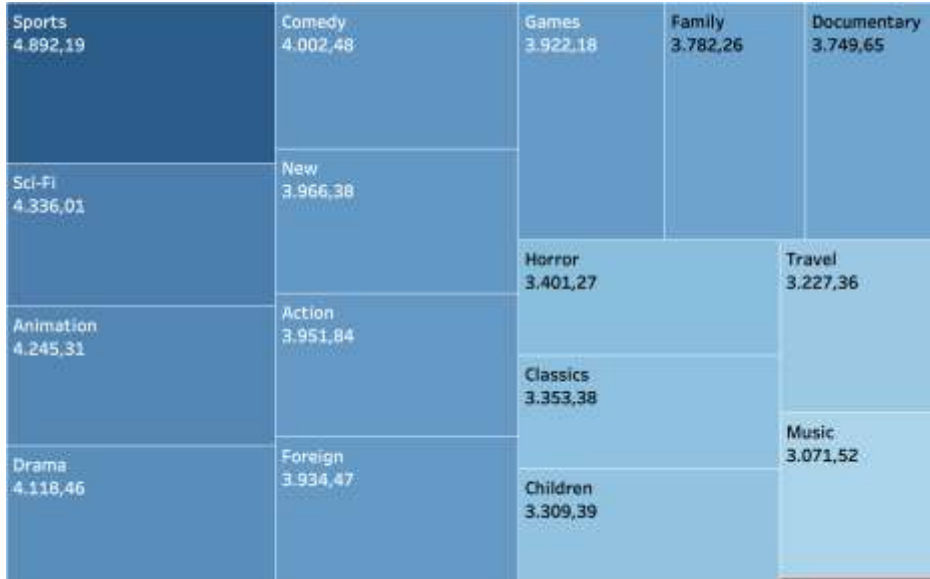
	Rental Duration	Rental Rate	Film Length
Average	5	2,98	115
Minimum	3	0,99	46
Maximum	7	4,99	185

Rental data indicates an average duration of 5 days, with prices ranging from \$0.99 to \$4.99.

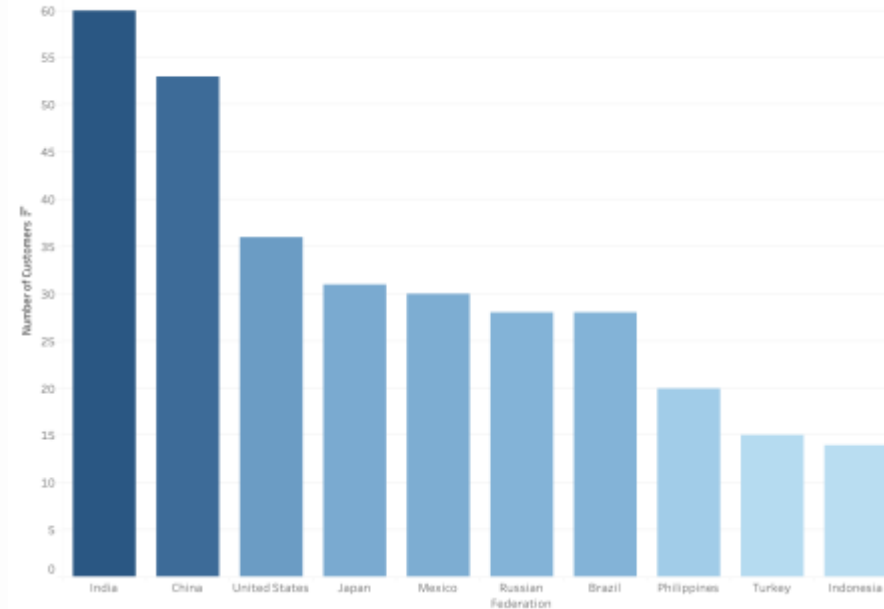


# Rockbuster / Analysis

Revenue by movie genre



Top 10 Countries with most customer numbers



- The most important category for Rockbuster's business is Sports, followed by Sci-Fi, Animation, and Drama.

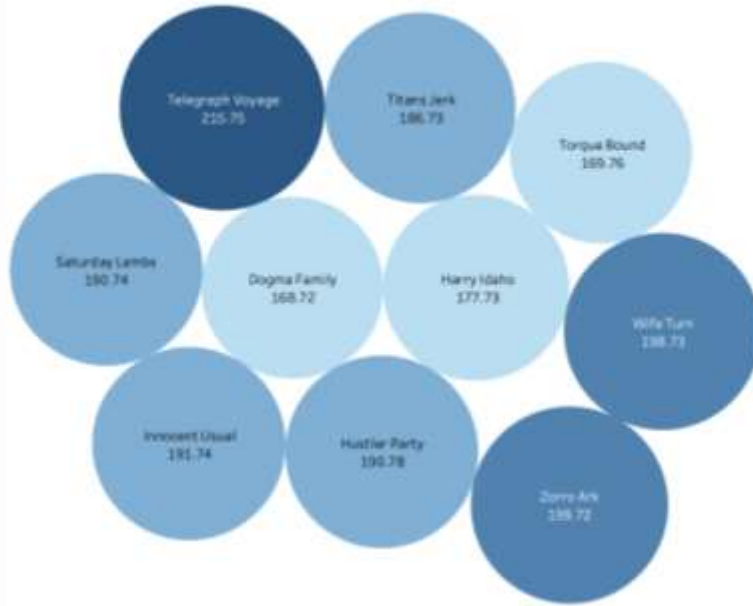
- Generally, the revenue across each genre is fairly homogeneous, except for the Thriller genre, which has only 47,89 revenue.

- The current rental service has the highest number of customers in India and China.
- The rankings of customer numbers are largely consistent with the revenue rankings.

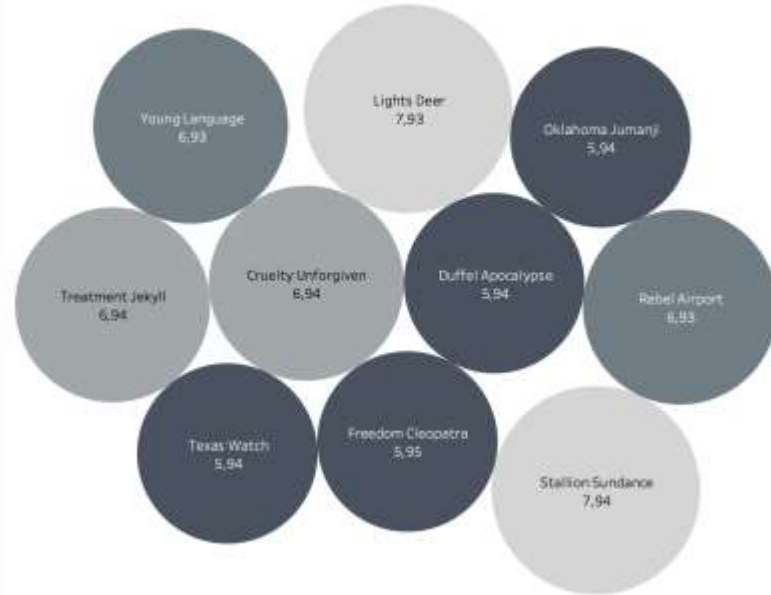


# Rockbuster / Analysis

Top 10 Movies with the Highest Revenue



Flop 10 Movies with the Lowest Revenue



# Rockbuster / Recommendations



## 01 Primary Market:

- ✓ India, China and the United States have the largest customer bases and generate the highest revenue, making them key markets for Rockbuster Stealth to prioritize in future growth strategies.

## 02 Invest in Top Genres:

- ✓ Increase focus on Sports, Sci-Fi, Animation and Drama movies, which demonstrate strong performance in generating revenue. Regularly update the library with popular titles in these categories to attract and retain customers.

## 03 High-Value Customer Programs:

- ✓ Develop loyalty programs and exclusive promotions for high-value customers to encourage sustained engagement.



Project



Presentation



Tableau



GitHub



# Instacart

- This project analyses Instacart's online grocery data using Python to identify key sales patterns and customer behaviors. The analysis offers actionable insights for customer segmentation, peak sales periods, and product popularity, supporting the optimization of marketing and sales strategies.



## Objectives

- Perform an exploratory analysis of the customer purchase history data to uncover patterns that will benefit the marketing and sales departments in marketing campaigns and promotional efforts.



## Methods

- Data wrangling & Data subsetting
- Consistency checks
- Merging data frames & deriving variables
- Grouping & aggregating data
- Data visualization with python.

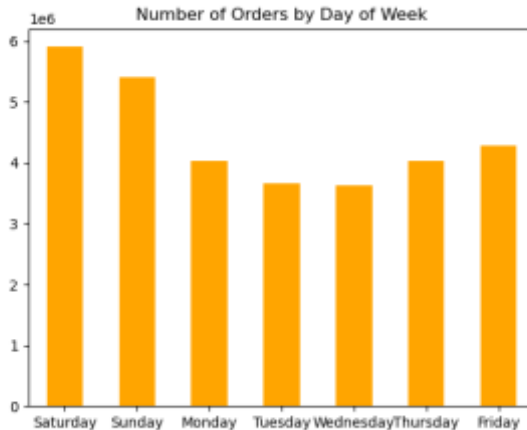




# Instacart / Analysis

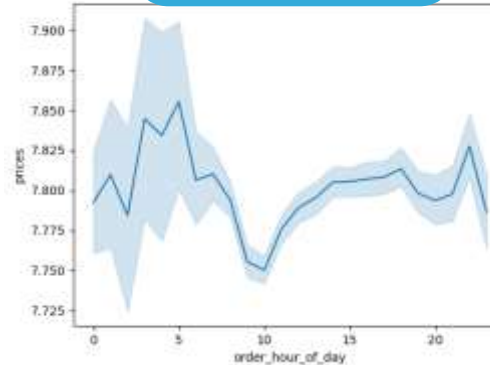


## Population flow



The rush hours of the day is from 9:00 to 17:00.

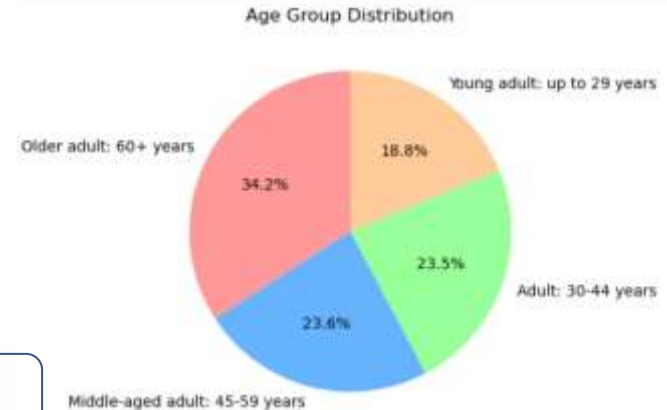
## Sales of the day



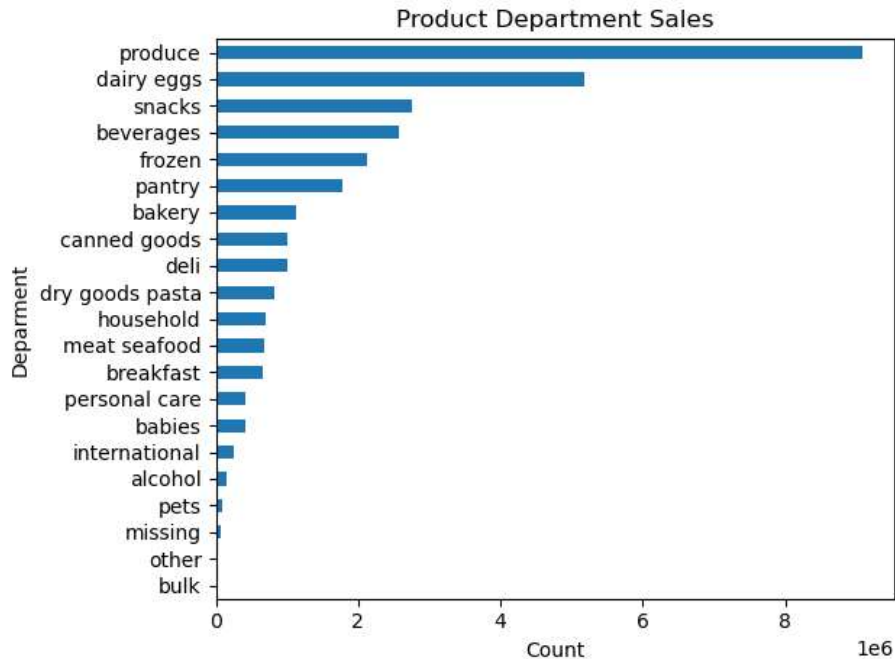
Sales are higher in the early morning, but there is a dip between 9:00 and 10:00.



Saturday and Sunday are the busiest day of the week. Tuesday and Wednesday are the least busy days of the week.



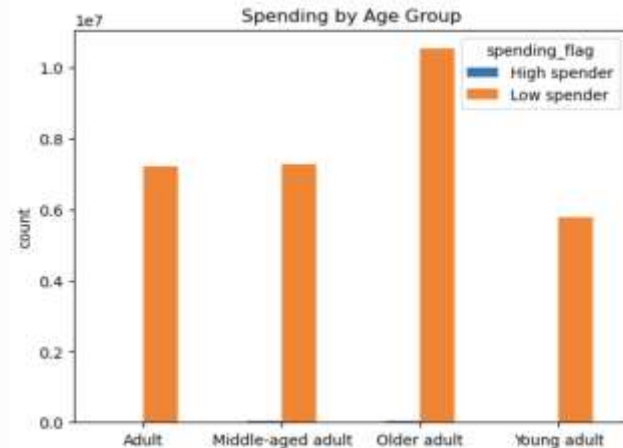
# Instacart / Analysis



The bar chart shows the best seller products come from produce. The next most ordered from departments are dairy/eggs, snacks, beverages, and frozen.



Total Product Orders are more in south and west regions, may be due to high population.



most customers are considered to be low spenders.

# Instacart / Recommendations

## 01 Busiest time:

- ✓ Implement sales and promotions specifically in the middle of week to boost sales activity.
- ✓ Schedule ads during the times 6 am to 8 am and 5pm to 10 pm.

## 03 Customers:

- ✓ Instacart should focus on converting new customers into regular or loyal customers with welcome promotions and incentivize regular customers with discounts to boost loyalty.

## 02 Products:

- ✓ Promotions and discounts on popular categories like Produce and Dairy could encourage customers to place orders more frequently.
- ✓ Similarly, targeted deals on less popular categories such as International, Alcohol, and Bulk items may help stimulate interest and increase sales in those departments.

## 04 Regions:

- ✓ Prioritize advertising in the South while also targeting growth opportunities in the West and Midwest.



Excel Report



GitHub



# Pig E. Bank

- Pig E. Bank is a global financial institution with a dedicated anti-money laundering compliance department.



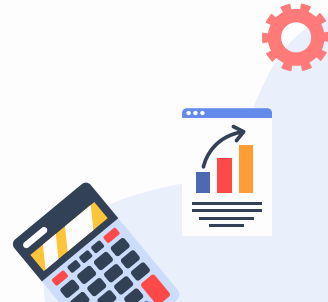
## Objectives

- Determine the primary risk factors contributing to client loss and represent them using a decision tree model
- Provide recommendations to increase customer retention



## Methods

- Exploratory data analysis
- Data mining techniques
- Data cleaning
- Consistency checks
- Data modelling
- Linear regression



# Pig E. Bank / Analysis

## Comparison by Activity

Count of Active Member	Clients Remained	Clients Exited
0	43,77%	70,10%
1	56,23%	29,90%

## Comparison by Gender

Count of Gender	Clients Remained	Clients Exited
Female	43,44%	59,31%
Male	56,56%	40,69%

## Comparison by Number of Products

NO. of Products	Clients Remained	Clients Exited
1	46,82%	69,61%
2	52,54%	15,69%
3	0,64%	13,73%
4	0,00%	0,98%

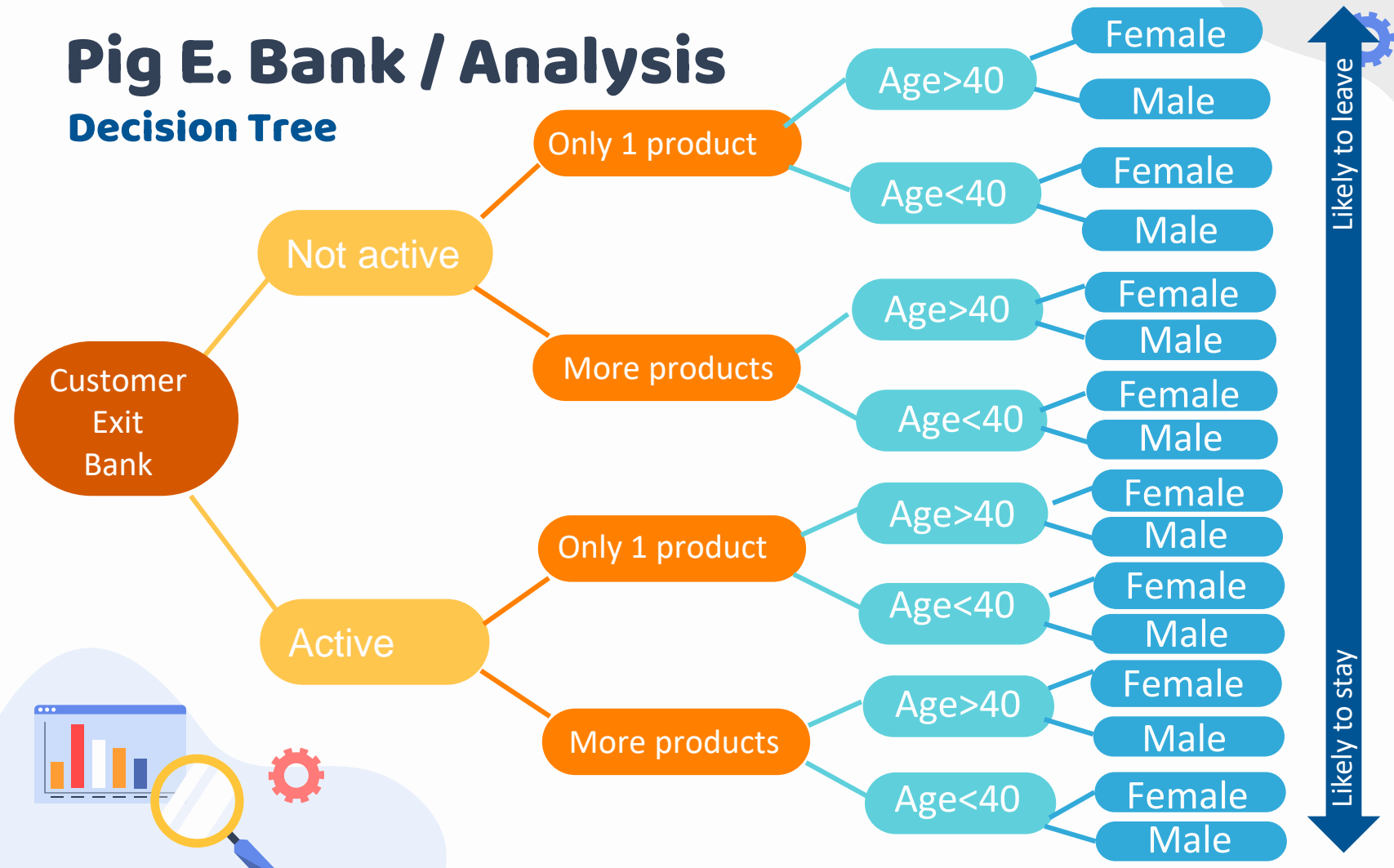
## Comparison by Age

Count of Age	Clients Remained	Clients Exited
18-29	17,56%	4,90%
30-39	50,38%	24,51%
40-49	22,52%	41,67%
50-59	5,60%	18,63%
60-69	2,54%	10,29%
70-82	1,40%	0,00%



# Pig E. Bank / Analysis

## Decision Tree



# Pig E. Bank / Recommendations



## 01 Customer Retention:

- ✓ Focus on building a base of active customers who use multiple products by promoting engagement and account activity.

## 03 Product Promotion:

- ✓ Cross-sell additional products to low-product clients and align offerings with customer interests.

## 02 Customer Insights:

- ✓ Conduct surveys to understand why some customers are inactive and identify changes or new products that could increase their engagement.

## 04 Customer Engagement:

- ✓ Increase engagement with older, high-balance clients through personalized strategies and support German clients with tailored loyalty programs.





# World Happiness Report

- The data for this project comes from the World Happiness Report, a globally recognized survey measuring happiness across nations, spanning from 2015 to 2019.



## Objectives

- To provide a comprehensive view of global happiness trends driving life satisfaction.
- To uncover patterns and relationships between happiness factors.



## Methods

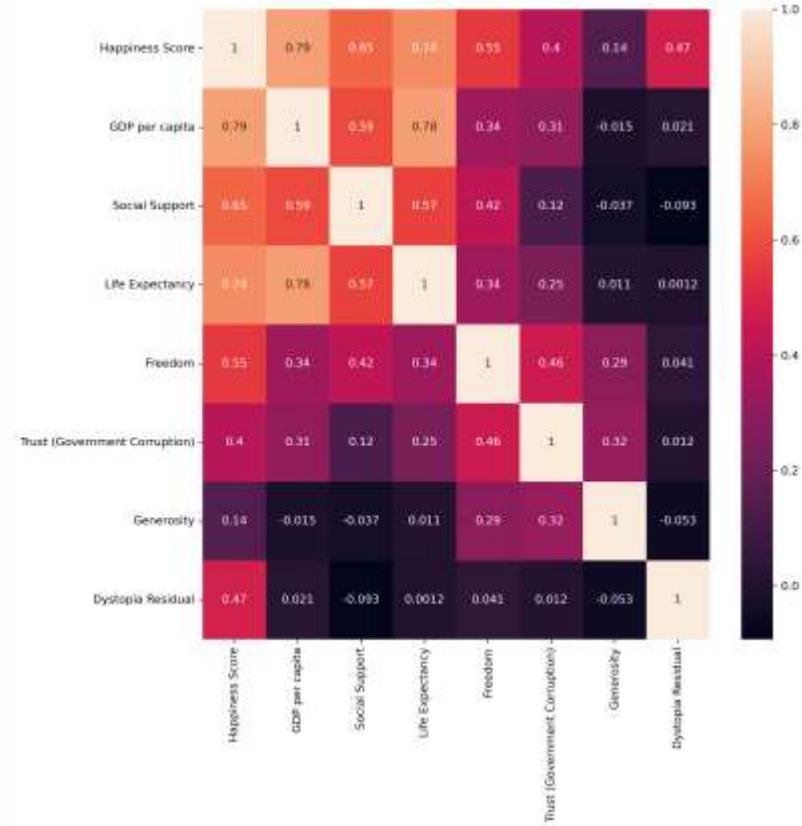
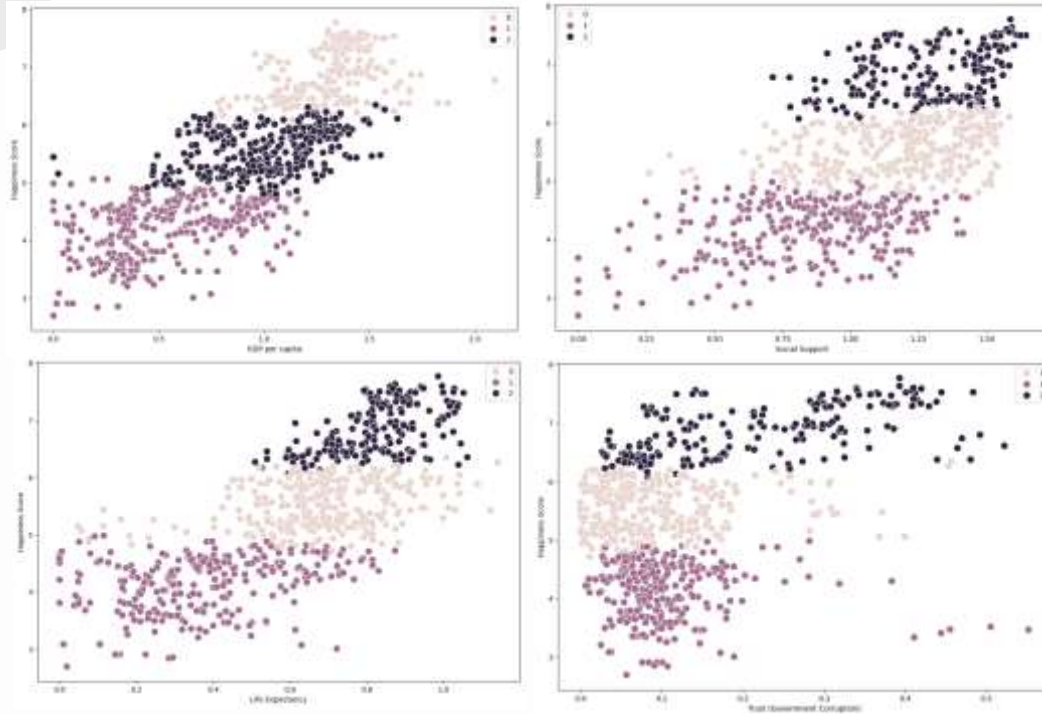
- Data Cleaning, Normalization, and Integration
- Exploratory Data Analysis and Statistical Correlation
- Data Visualization and Storytelling through Tableau



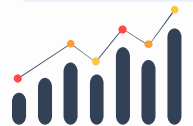




# Analysis



1. GDP per Capita has the highest mean differences between clusters, indicating it is the most significant drivers of happiness.
2. Social Support, Freedom and Life Expectancy also vary noticeably, contributing to the stratification of clusters.
3. Generosity and Trust in Government show smaller variations across clusters, suggesting these factors are less influential in determining happiness levels.





# World Happiness Map



### 3. Geographic Concentrations:

Happiness "Hot Spots" are concentrated in Western Europe, North America, and Oceania.

Regions with smaller circle sizes (low GDP per capita) but darker happiness scores are sparse, indicating that low-income countries rarely achieve high happiness.



### 1. Regional Patterns:

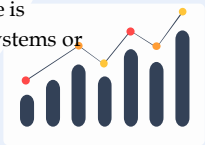
Western Europe and North America dominate the top-right quadrant of the GDP-Happiness relationship, indicating a strong positive correlation between economic prosperity and overall happiness.

Sub-Saharan Africa and war-torn regions in the Middle East fall on the opposite end, emphasizing how economic hardship and instability directly impact happiness levels.

### 2. Notable Outliers:

Some countries like Costa Rica and Puerto Rico have moderate GDP per capita but relatively high happiness scores, potentially reflecting cultural, social, or environmental factors that offset economic limitations.

On the other hand, countries with higher GDP but not so high happiness, such as Hong Kong, China, Singapore and oil-rich Middle Eastern countries, suggest that wealth alone is insufficient without strong social systems or freedom.



# Recommendations

## 01 Prioritize Social Investments:

- ✓ Countries with lower happiness scores should focus on improving access to education, healthcare, and essential services, especially in rural and marginalized communities. Investments in these areas not only improve living standards but also enhance overall well-being.

## 03 Develop Integrated Happiness Metrics:

- ✓ Combining economic, social, and cultural indicators into a unified framework would provide a more holistic understanding of happiness. Leveraging these metrics alongside predictive analytics could help policymakers simulate and optimize the outcomes of targeted interventions.

## 02 Explore Cultural and Regional Contexts:

- ✓ Further research should focus on understanding the unique cultural and regional drivers of happiness, particularly in low-GDP regions, to tailor interventions effectively.





# My Contact



Email:

ivy.zhao1002@gmail.com