

# DATA ANALYTICS PORTFOLIO BY SWAPNALI SONAWANE

---



<https://github.com/swapnali-deshpande1511>



[Profile – swapnali.sonawane | Tableau Public](#)



- Gaming CO
- Influenza Forecast
- Rockbuster Stealth LLC
- Instacart
- Pig E.Bank
- World Happiness Report

Global market analysis of video game sales

Staff deployment for Flu season

Movie rental market analysis

Market segmentation analysis

Customer attrition

Discover the worldwide happiness Index

# GAME CO.

---

Global Market Analysis of Video Game Sales





# GAME CO. PROJECT OVERVIEW

---



- For this project, the main objective was to uncover significant trends and patterns by analysing historical data from a video game dataset. The aim is to glean insights into the potential market reception for upcoming games.



## Key Steps

- Data cleaning : find missing values and duplicates; handled missing values by using imputation
- Exploratory Data Analysis (EDA): to calculate for Mean, Median Mode
- Data grouping, filtering, and summarizing using Pivot Table
- Data Charts: Create a column chart of total global sales by publisher; Line chart of the average North American sales by year and so on
- Decide on what type of analysis: Descriptive analysis, Diagnostic, Predictive or Prescriptive
- Interpret results and summarize findings

EXCEL FUNCTIONS USED: FIND & SELECT, REPLACE WITH, REPLACE ALL, REMOVE DUPLICATES; used PivotTables and SUMIFS for summarizing and aggregating data

# ANALYSIS AND INSIGHTS

- Game Co. assumed that regional sales remained constant over time . To test this hypothesis, I created a time series for each region's sales proportion . The variations in these series disproved the assumption .

These are my insights :

- Global sales have declined in the year 2016
- The sales for various region have changed over time
- The sales have declined in the NA region
- The sales are rising steadily in the EU region
- The sales in the JP region have also increased
- The marketing budget to be allocated needs to be revised accordingly



# GAME CO. RECOMMENDATIONS

- The decline in the sales in NA needs to be explored. NA was a good contributor to global sales in the past.
- The NA market has a huge potential due to its big population size.
- One reason explored in this analysis is decrease in the number of games under popular genre like Action and Shooter. Accordingly, marketing gimmicks need to be used to popularize existing games or new genre in the NA.
- Initiatives to be taken in the NA region to re-engage and engage new customers.
- The rise of sales in EU need to be explored to promote steady growth or stabilize higher sales.
- The JP region has also shown growth and could be a promising market in future.
- New markets in other regions need to be explored.
- Please find the whole presentation here



# SEASONAL INFLUENZA FORECAST

Staff Deployment for Flu Season





# SEASONAL INFLUENZA FORECAST PROJECT OVERVIEW

---



The aim of this project was to help a medical staffing agency plan for staffing needs by establishing optimal deployment schedules and staffing levels across the US during Flu season



## Key Steps

- Conduct Exploratory Data Analysis (EDA), including:
  - Descriptive/Summary Statistics: Central tendency & distribution.
  - Creating charts to understand distribution, relationships, missing values, and outliers.
- Data cleaning (including imputation) and transformation (merging spreadsheets).
- Hypothesis testing (t-test in Excel).
- Interpret results and summarize findings and summarize finding

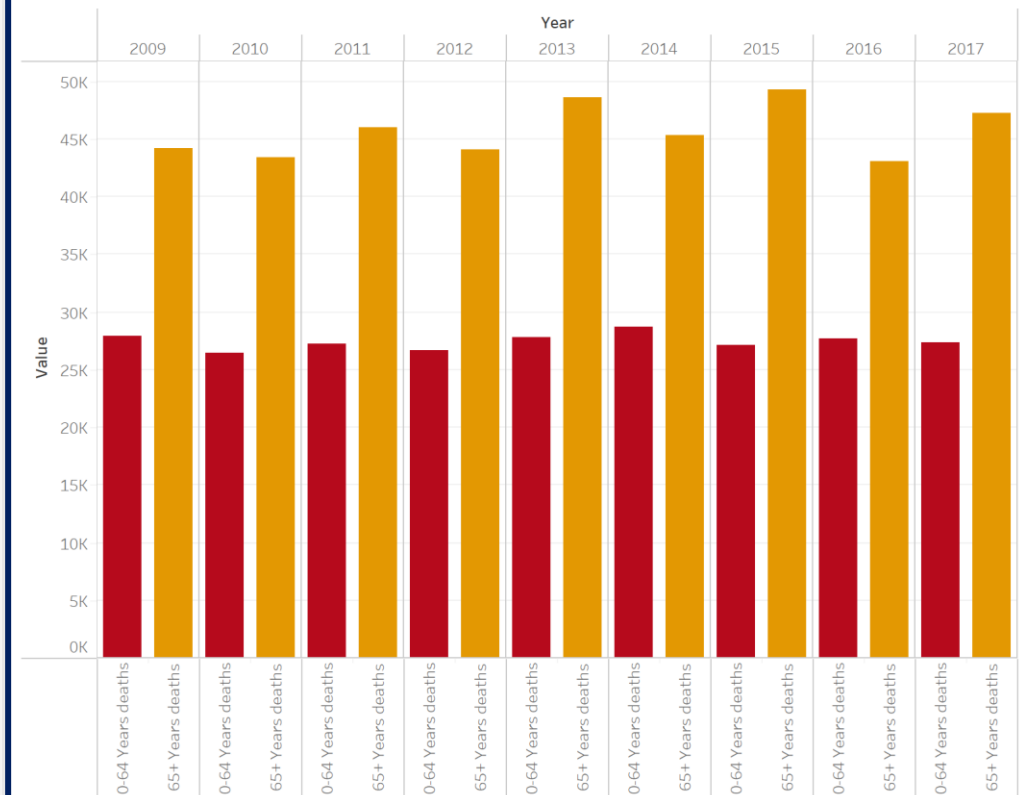


# ANALYSIS AND INSIGHTS



- The aim of the project was to answer three key questions to support staffing agency in making informed decisions:
  - When does the Influenza season begin in the US?
  - How do age groups factor into vulnerability within the population?
  - Which states have higher population belonging to the vulnerable groups?
- In the right and below charts shows the mortality by Influenza in the US, every year since 2009 to 2017. The peak season is January with cases rising from December till March
- From the spatial chart, states can be identified with higher density of vulnerable population

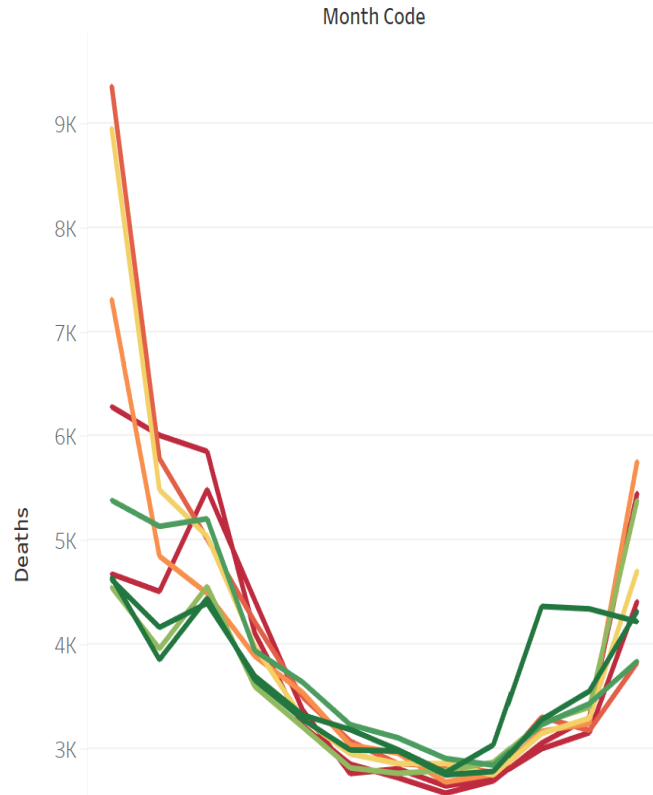
**Influenza deaths in the USA for age groups from 2009 to 2017**



# PROJECT VISUALIZATION

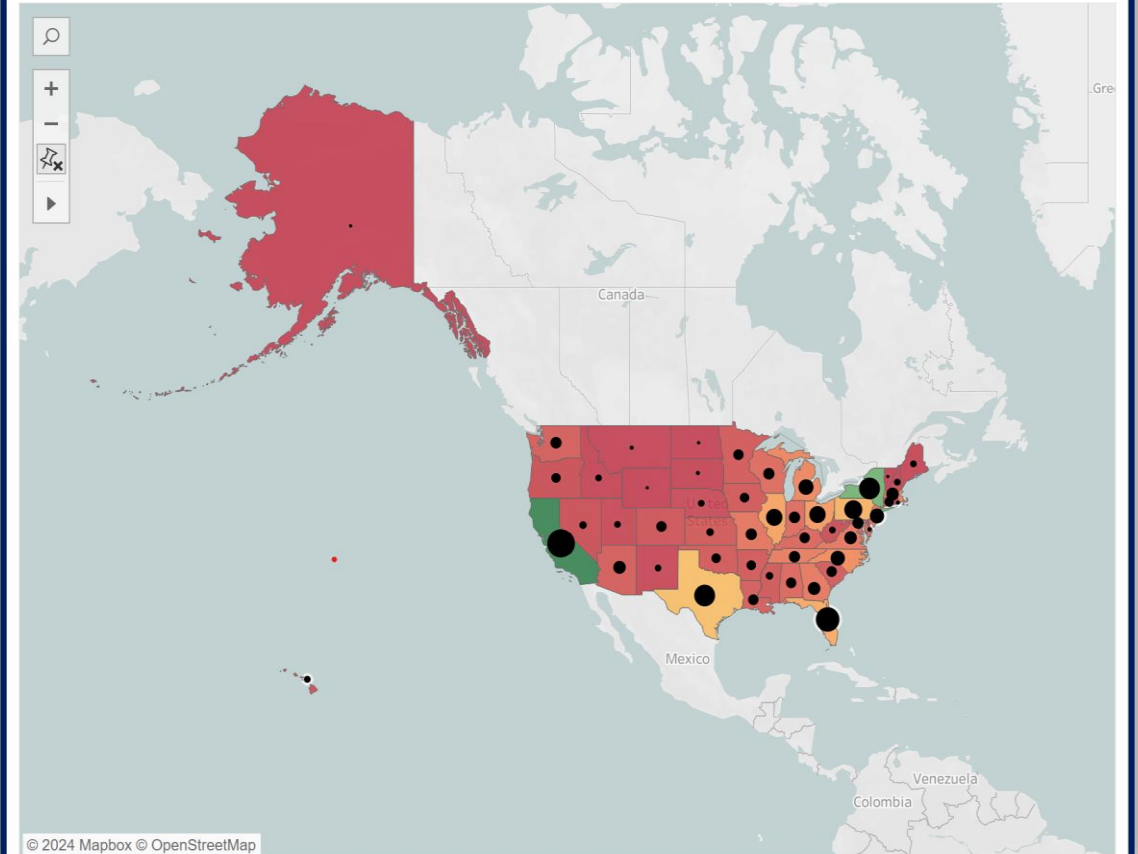


Influenza Season in the USA from 2009 to 2017



According to Influenza Season Line Chart Peak Month of Influenza is January. In the winter Season i.e from November it is increasing and making peak at January and then from February slowly Decreasing.

States with 65+years old Influeza Deaths and 65+ years old Population



# SEASONAL INFLUENZA FORECAST RECOMMENDATIONS

- Additional staffing must be sent between November and March, with peak staffing in held in January
- Staffing assistance should be allocated by state's need priority.

## MONITORING:

To measure the effectiveness of our plan we will compare future and incoming monthly influenza death counts with historical averages to determine the impact of additional staffing.

## LIMITATIONS & NEXT STEPS:

- CDC has suppressed values to safeguard identities where influenza death counts were less than 9. Consequently, we had to impute these values, which may affect data accuracy.
- Next steps involve researching concurrent diseases during influenza season and identifying protective factors in states with highly vulnerable populations but low influenza deaths.







# ROCKBUSTER PROJECT OVERVIEW

---



This project involved analysing Rockbuster movie rental database using CSV and Excel files, primarily utilizing PostgreSQL. The goal was to support the Business Intelligence (BI) department by uncovering valuable insights to aid in the company's launch of a new online video service.



## Key Steps

### Data Extraction:

Used SELECT, JOIN, and WHERE clauses to extract relevant data from multiple tables.

- Data dictionary: Created an Enterprise Relationship Diagram (ERD) using DBVisualizer.
- Extract Enterprise Relationship Diagram (ERD).
- Extracted ERD from Rockbuster's database using DBVisualizer.
- Data Aggregation: Employed GROUP BY and aggregate functions like SUM, AVG, and COUNT to summarize data.
- Data Filtering: Utilized HAVING clauses to filter aggregated results.
- Data Visualization: Created bar charts, line charts, scatter plots, among others

# ANALYSIS AND INSIGHTS



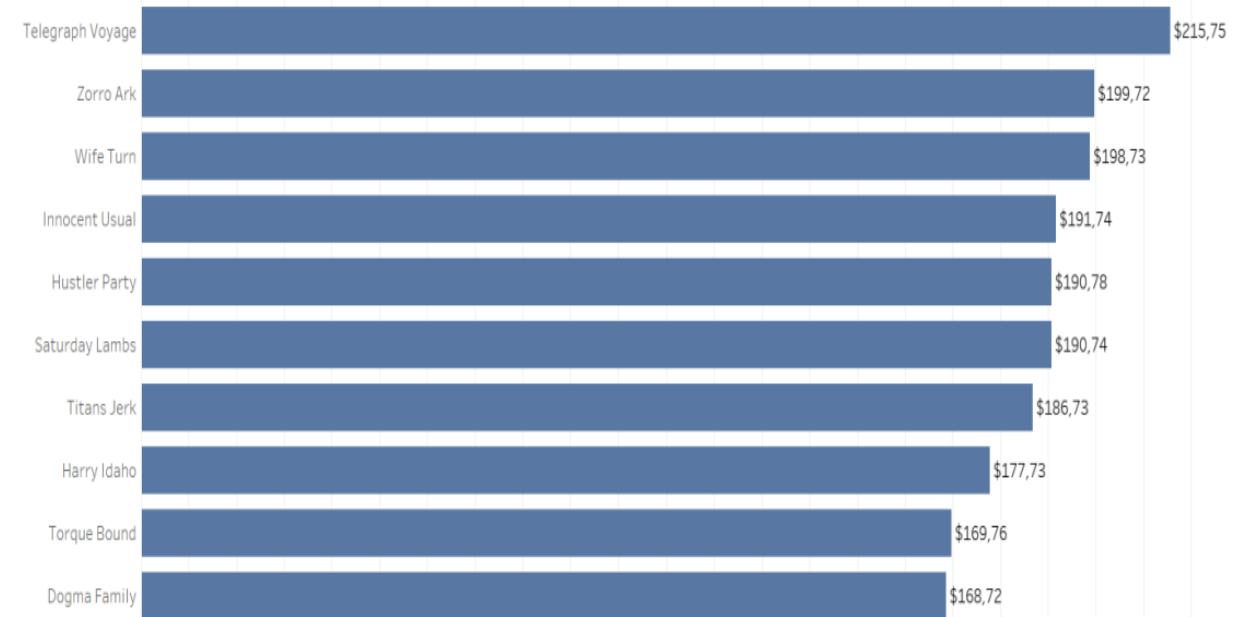
The goal was to support the Business Intelligence (BI) department by uncovering valuable insights to aid in the company's launch of a new online video service.

This was done by answering series of questions.

WHICH MOVIES CONTRIBUTED THE MOST/LEAST TO REVENUE GAIN?

The highest revenue of \$215,75 is generated by Telegraph Voyage, followed by Zorro Ark (\$199,72) and Wife Turn (\$198,73)

Top 10 Movies by Revenue





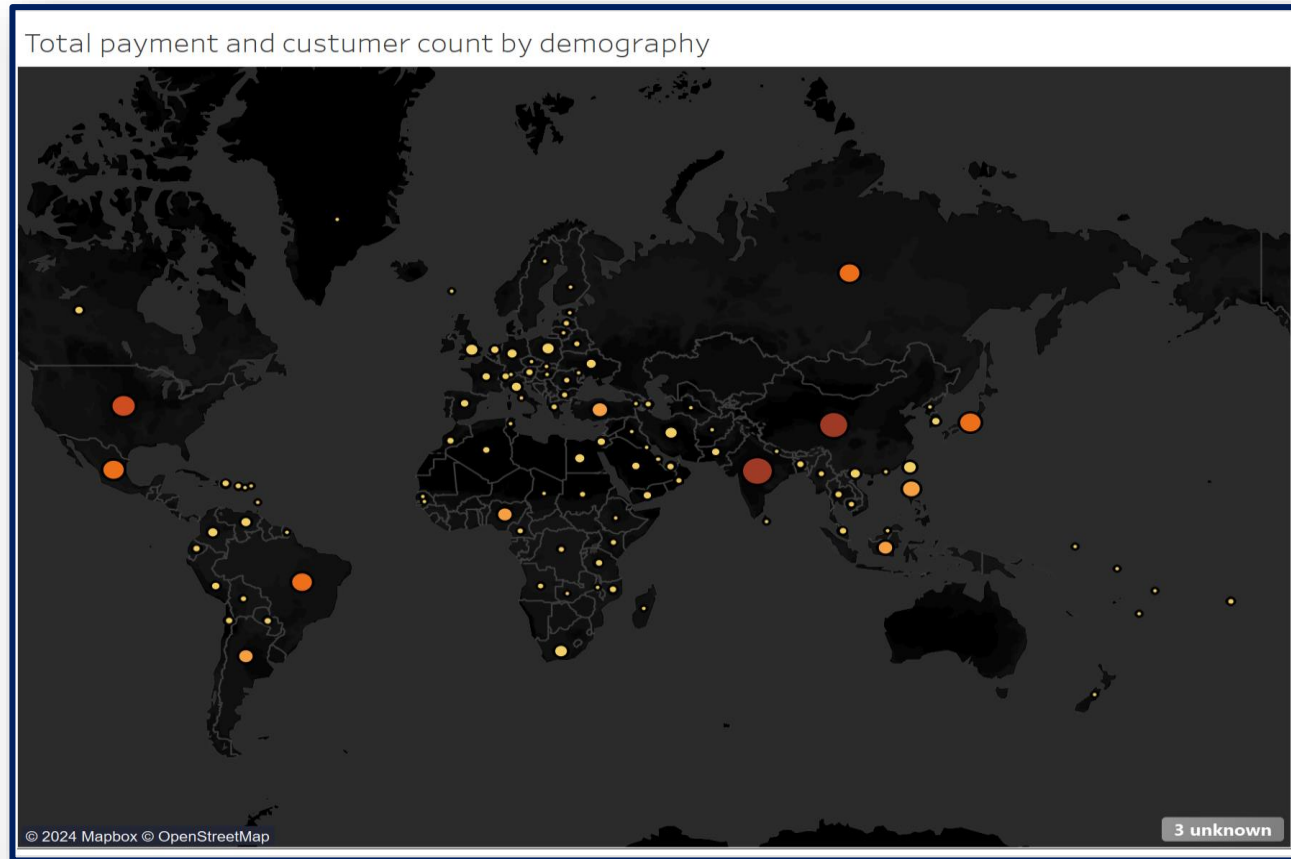
# ANALYSIS AND INSIGHTS



The goal was to support the Business Intelligence (BI) department by uncovering valuable insights to aid in the company's launch of a new online video service. This was done by answering series of questions.

WHICH COUNTRIES ARE ROCKBUSTER CUSTOMERS BASED IN?

The highest number of customers are in India and China



# ROCKBUSTER RECOMMENDATIONS

---

- Maintain niches in the top 10 countries to sustain revenue. With the rest of the world contributing 50% of revenue from just 39% of the customer base, there's significant opportunity to expand customer bases with targeted market strategies
- Maintain inventory of top revenue-generating genres and phase out the Thriller genre upon license expiration due to its lower revenue potential, unless popularity trends shift



# INSTACART BASKET

---

Market Segmentation Analysis





# INSTACART BASKET PROJECT OVERVIEW

---



Instacart wants to uncover more about their sales patterns. The goal is to derive actionable insights and propose segmentation strategies based on provided criteria.



## Key Steps

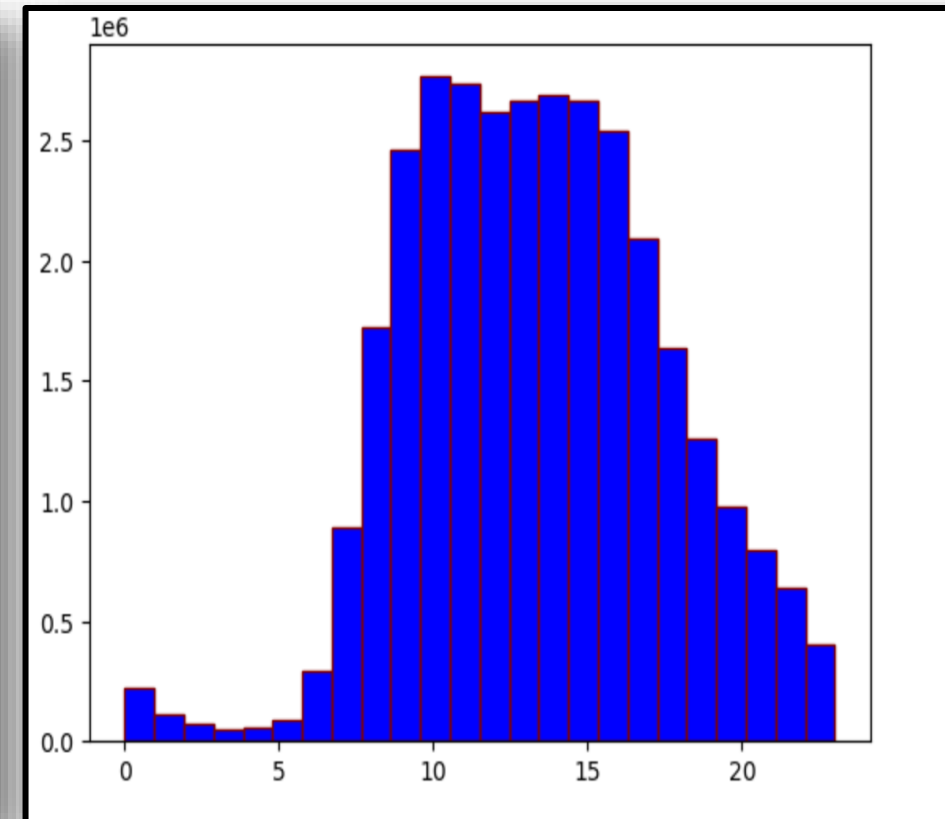
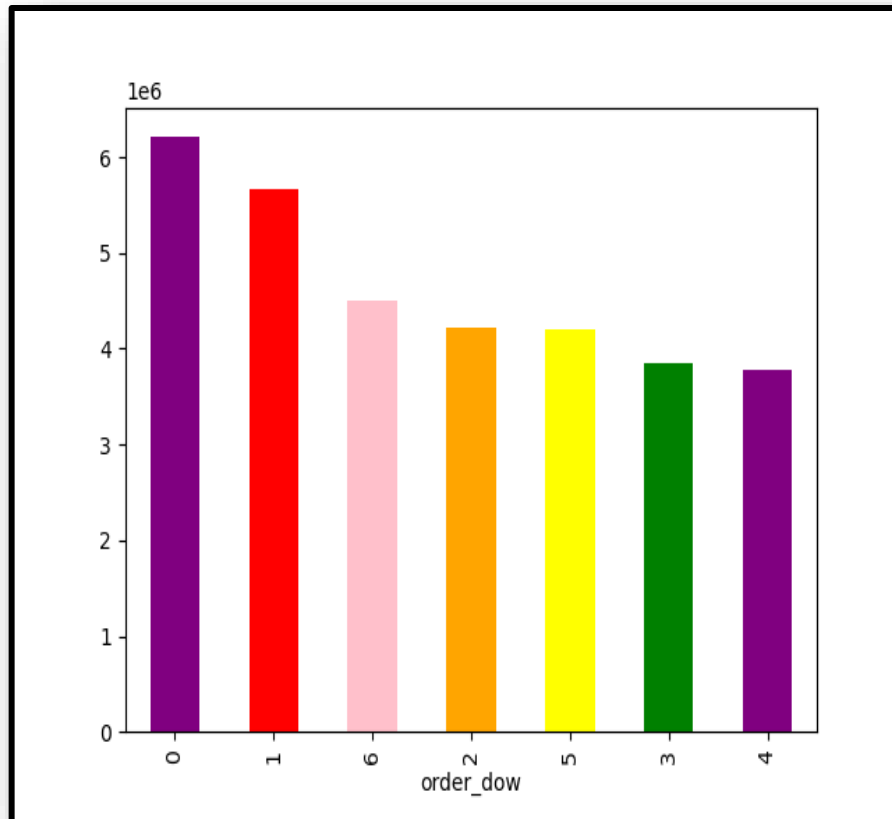
- Write Python Scripts: Used Jupyter Notebook to write Python scripts and document the analysis.
- Import: Imported datasets and necessary Python libraries.
- Data cleaning and data wrangling: Utilized Pandas and NumPy for data cleaning, transformation, and preparation.
- Functions used: "(df.drop(columns = ['variable'])); df.rename(); df.T function to transpose; df.value\_counts() to calculate the frequency of each unique; df.to\_dict('index') to create data dictionary; performed data quality and consistency checks, including frequency counts.
- Exploratory Data Analysis (EDA): Derived new variables, data manipulation, grouping, and aggregating.
- Visualization: Created graphs and visualizations using Matplotlib, Seaborn, and Scipy.
- Export data frames: Exported cleaned and processed data frames for reporting.

# ANALYSIS AND INSIGHTS



The days and times with the most orders to schedule advertisements. From the analysis the days and times identified are as below:

- Busiest days are Saturdays and Sundays
- Busy hours are between 9am to 5pm with the busiest hour being 10am

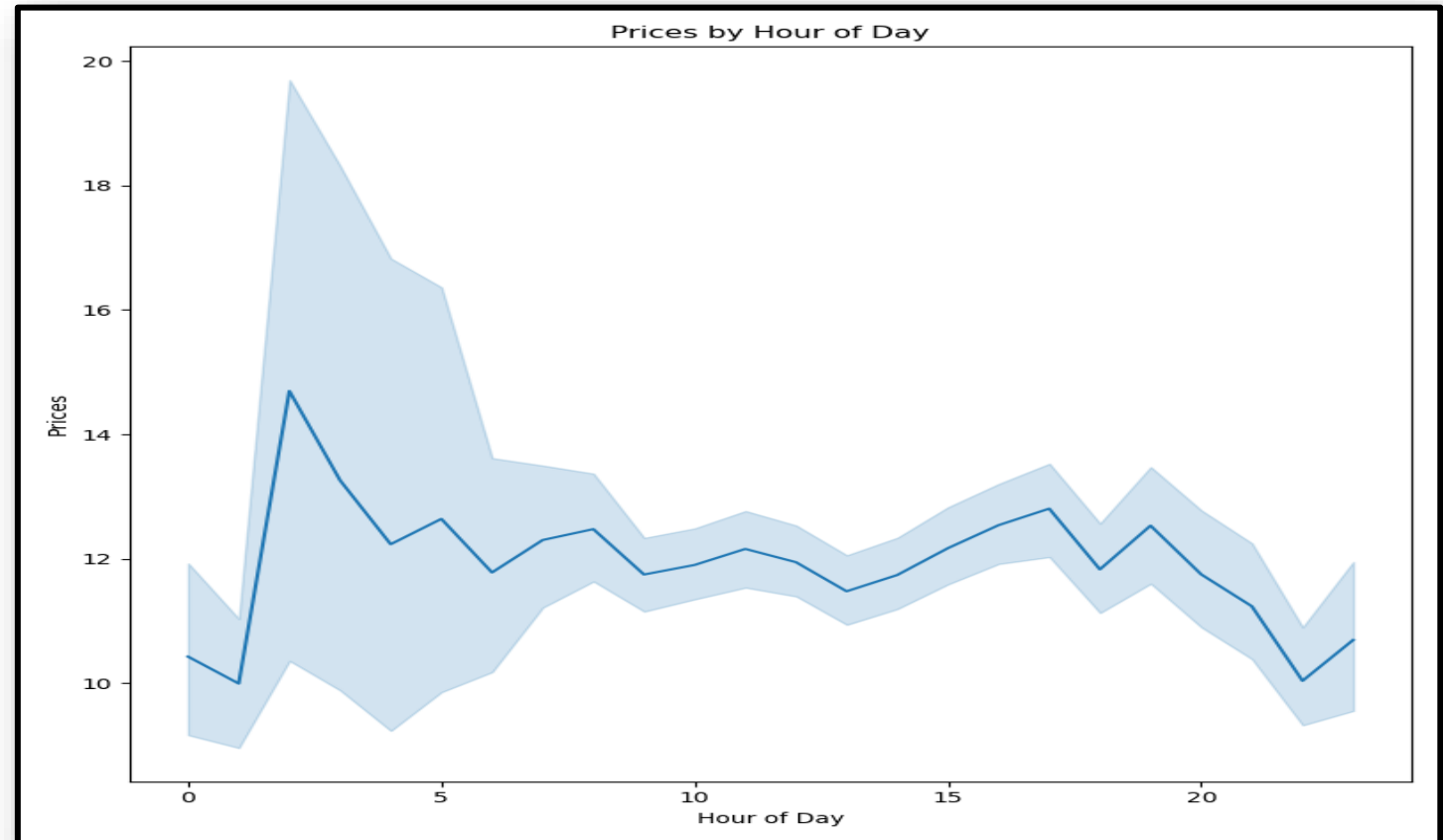


# ANALYSIS AND INSIGHTS



Times of the day when people spend the most money, as this might inform about the type of products being advertised at these times.

Hours of the day people spend most: Between 9am-4pm



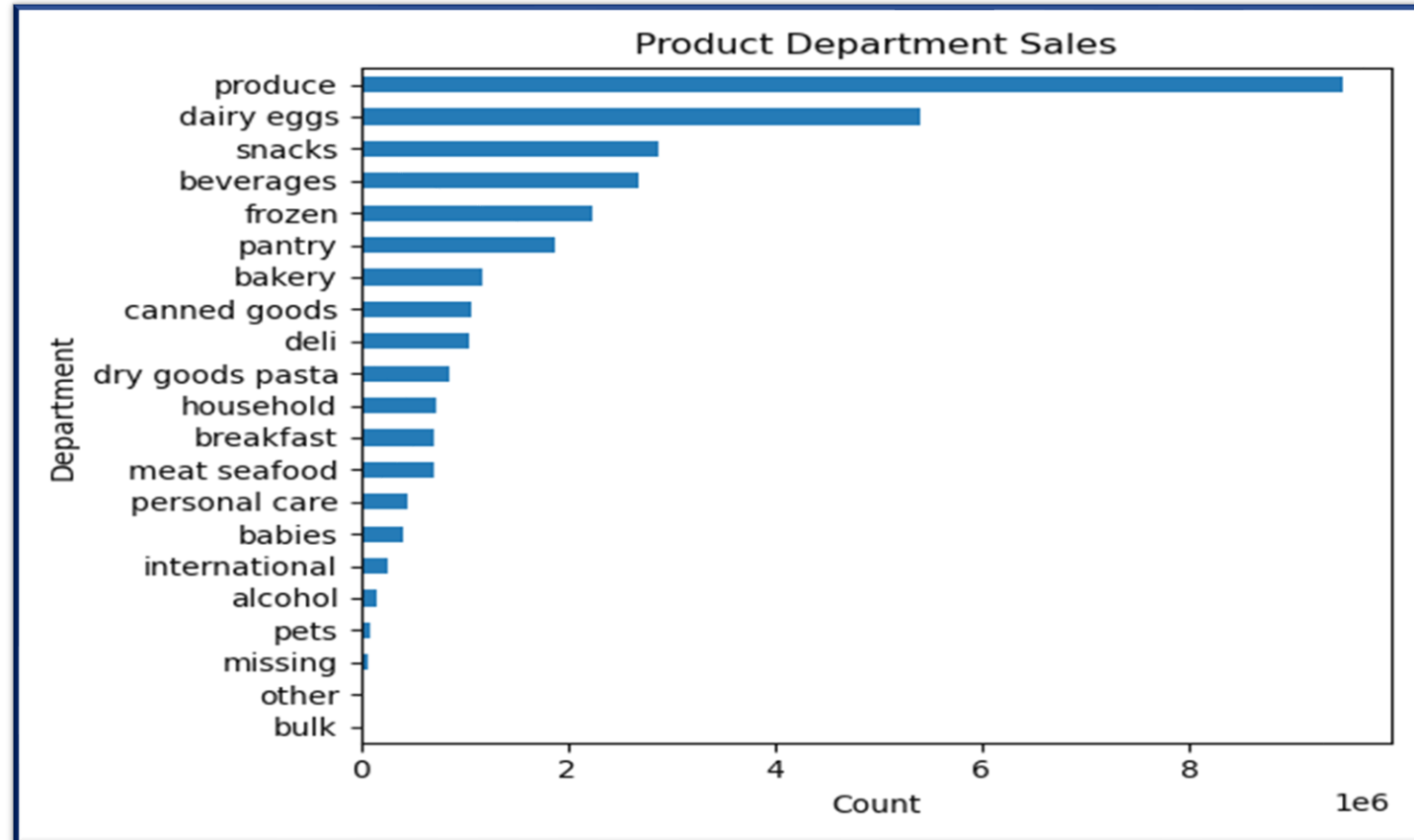


# ANALYSIS AND INSIGHTS



## Most Popular products

The top 5 products are fresh produce, dairy eggs, snacks, beverages and frozen foods.



# INSTACART RECOMMENDATIONS

- Ad Scheduling: Schedule ads during peak hours. People spend the most money during the hours of 8am to 4pm. 9 am is when people spend the most money. Therefore, Instacart should utilize those times to advertise specific products.
- 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 Fresh produce and Dairy & Eggs departments, with additional focus on Snacks, Beverages, and Frozen items, leveraging their convenience appeal.
- Customer Loyalty: There are more regular customers than loyal customers. Therefore, Instacart should offer incentives to reward its regular customers and incentives to non-regular customers (loyal and new) to become regular customers.
- Customer Profiling: Low-spenders are very high compared to high spender in all groups of customers. It will be useful to separately profile high spenders to identify products they spend on and place ads or improve quality of product, express delivery and so on to improve loyalty among high spenders.
- Demographic Focus: Middle-aged single adults are the most frequent shoppers, and parents form 75% of the customer base. Develop familyfocused loyalty programs emphasizing fresh produce, animal-sourced foods, and bulk offerings to drive sales and loyalty



# PIG E. BANK

---

- Customer Attrition



# PIG E. BANK PROJECT OVERVIEW

---



This project involved working with Pig E. Bank's client data set in Excel to support the sales team in identifying key factors that lead to customer attrition. Following the CRISPDM methodology, the project encompassed Business Understanding, Data Understanding, Data Preparation, and Modelling. Evaluation and Deployment were outside the scope of this project. A key feature was the creation of a decision tree to predict the probability of customers leaving the bank



## Key Steps

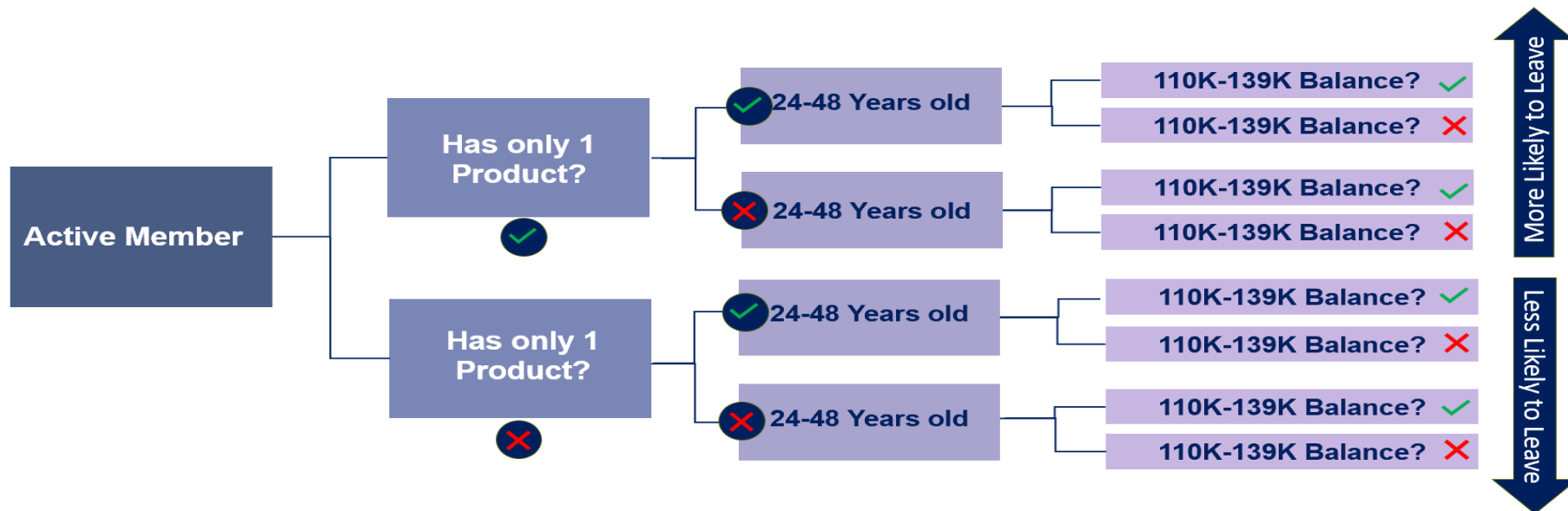
- Business Understanding: Identify business objectives.
- Data Understanding: Perform Exploratory Data Analysis (EDA) and document data-related issues, concerns, and limitations.
- Data Preparation: Involves data manipulation and data transformation, including cleaning, formatting, and data integration.
- Modeling: Use Decision Tree as the modeling technique to identify risk factors.



# ANALYSIS AND INSIGHTS



## How Likely is a client to leave the Bank



# PIG E. BANK RESULTS & RECOMMENDATIONS

## RESULTS :

The highest risk of a customer leaving is between the age group 40-69, having no activity, belonging to the gender female and from Germany.

## RECOMMENDATIONS :

- Personalized Engagement: Reach out to these customers with personalized messages and offers. This can include personalized financial advice, invitations to exclusive events, or updates on new products and services tailored to their needs.
- Activity Incentives: Introduce a reward program that encourages account activity.
- Enhanced Services: Develop financial products specifically designed for this demographic. This can include investment products, retirement planning services, and insurance products tailored to their life stage



# WORLD HAPPINESS REPORT

---

Discover the Happiest Countries in the World



[World Happiness Report | Tableau Public](#)



# WORLD HAPPINESS REPORT PROJECT OVERVIEW

---



To Analyse the happiest countries of the world and various factors associated which contributes and impacts the happiness index .



## Key Steps

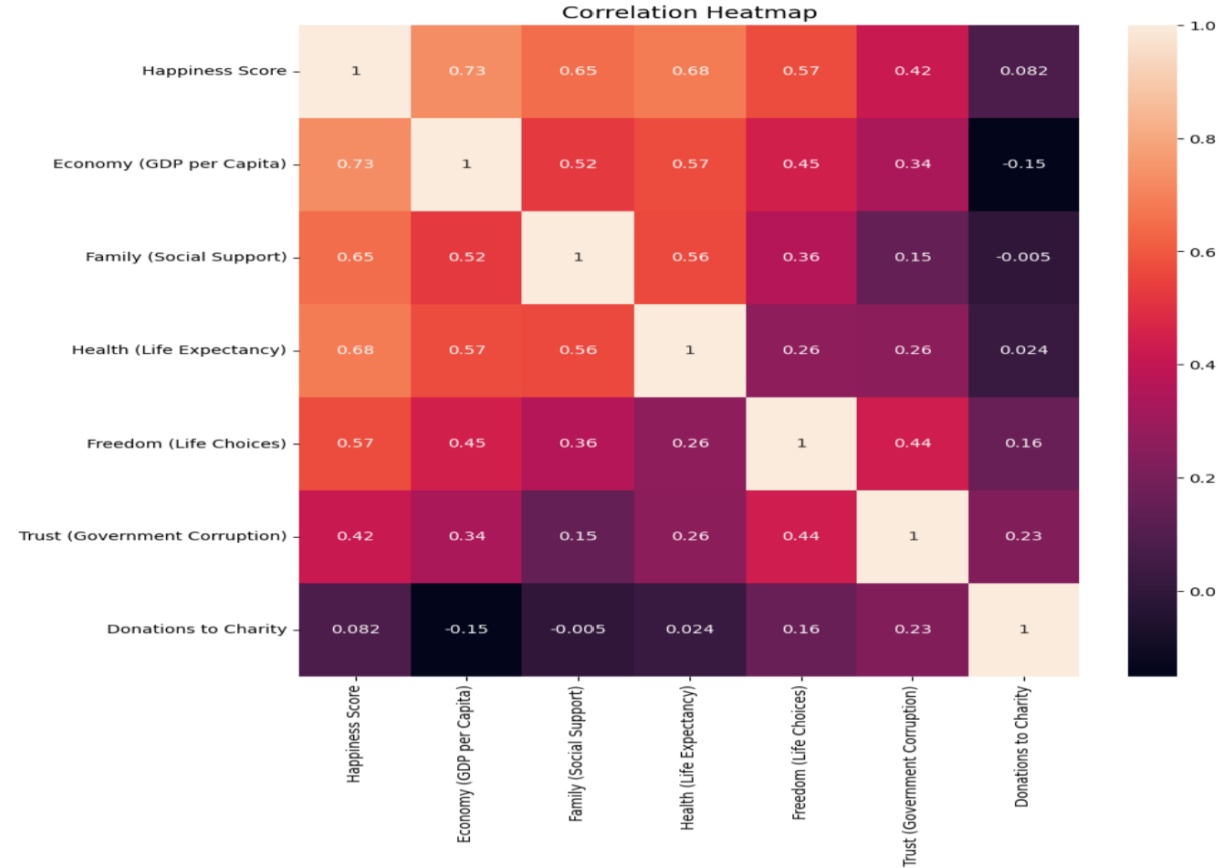
- Write Python Scripts: Used Jupyter Notebook to write Python scripts and document the analysis.
- Import: Imported datasets and necessary Python libraries.
- Data cleaning and data wrangling: Utilized Pandas and NumPy for data cleaning, transformation, and preparation.
- Create Geographical Visualizations with Python & Explore geographic variables in data using choropleth maps
- Supervise machine learning & Conduct a regression analysis in Python
- Conduct a cluster analysis
- Conduct time-series analysis
- Create an interactive dashboard



# ANALYSIS AND INSIGHTS



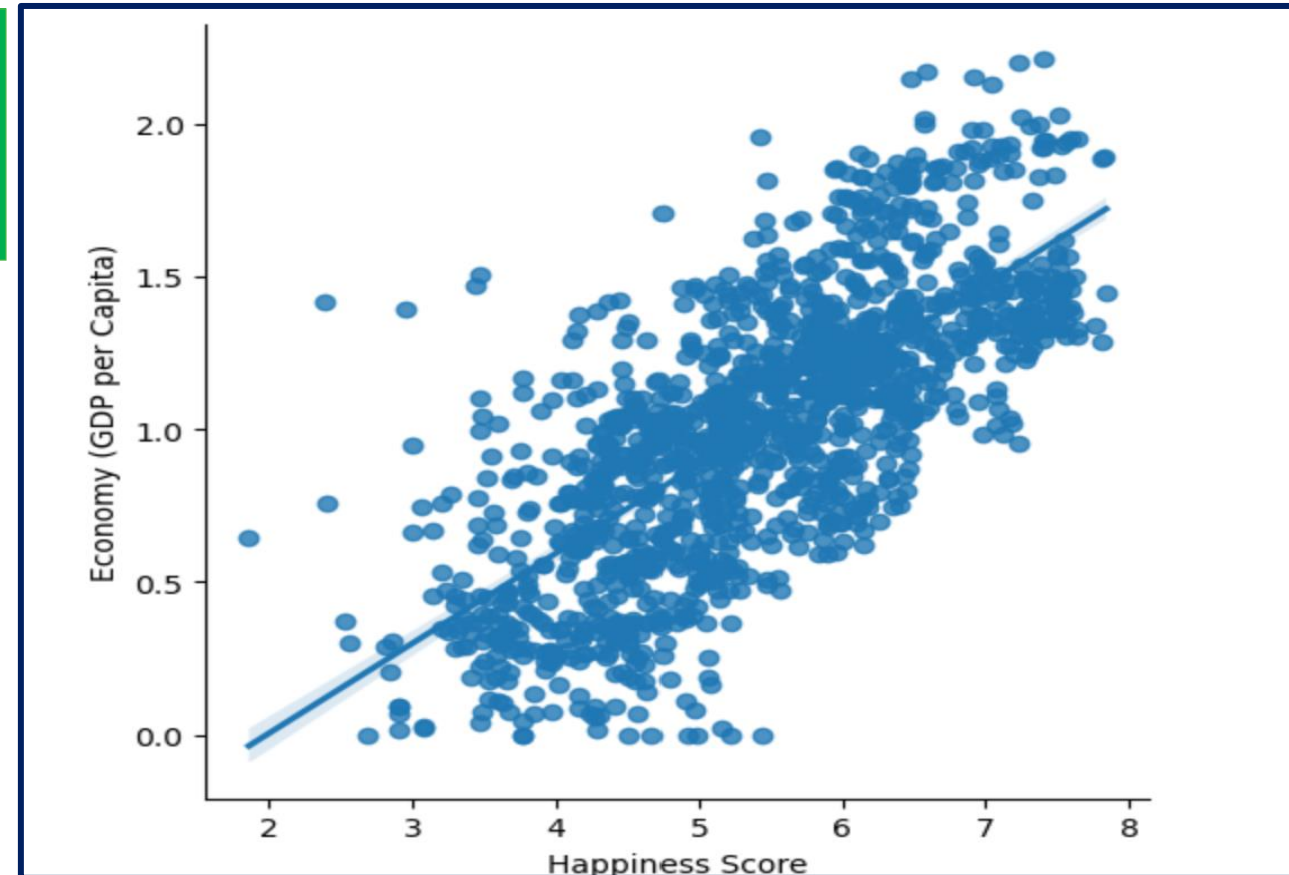
An important part of an analysis is to explore the relationships between key variables. A Correlation heat map utilizing python was created to verify which variables had strongest relationships.



# ANALYSIS AND INSIGHTS



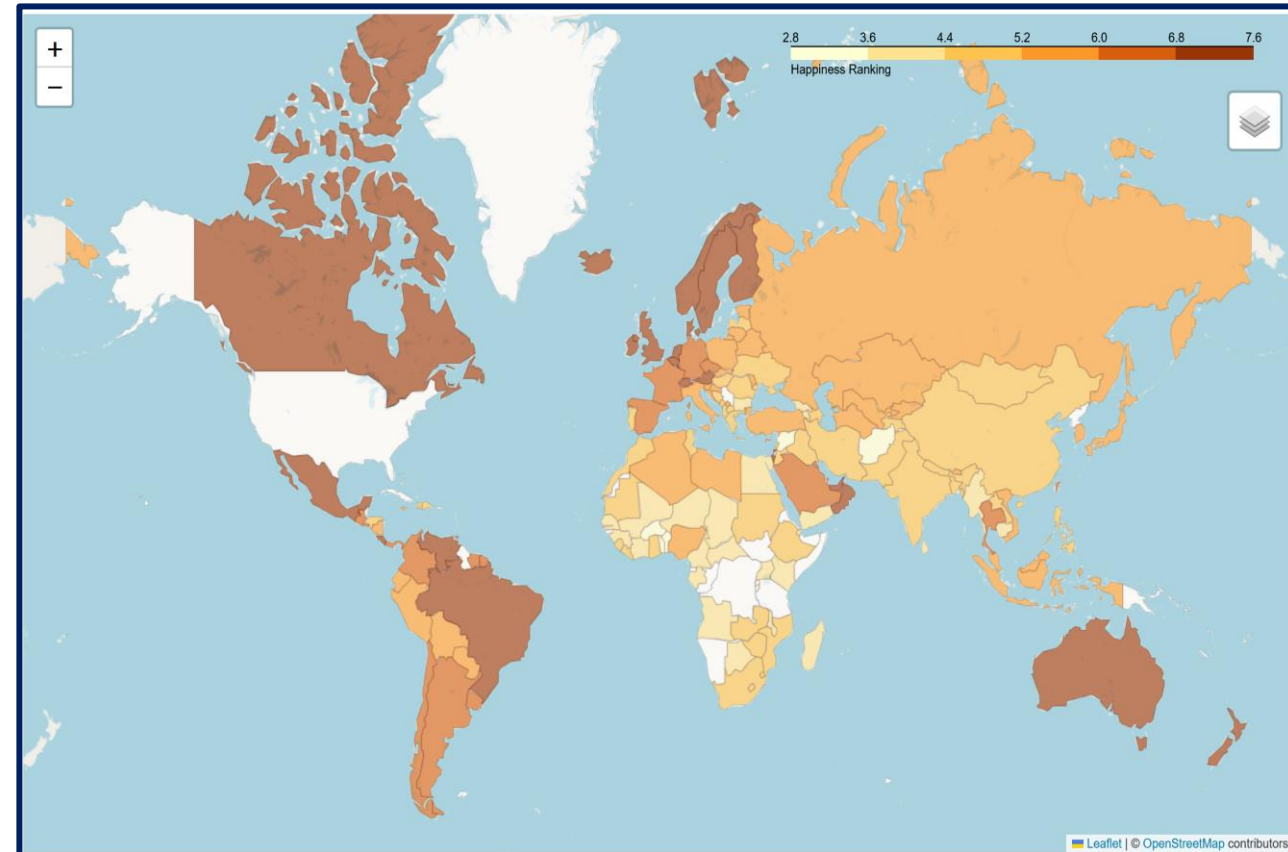
The Formation of datapoints in the scatterplot indicates that the two variables are in strong positive relationship : as one of them increases , the other increases as well.



# ANALYSIS AND INSIGHTS



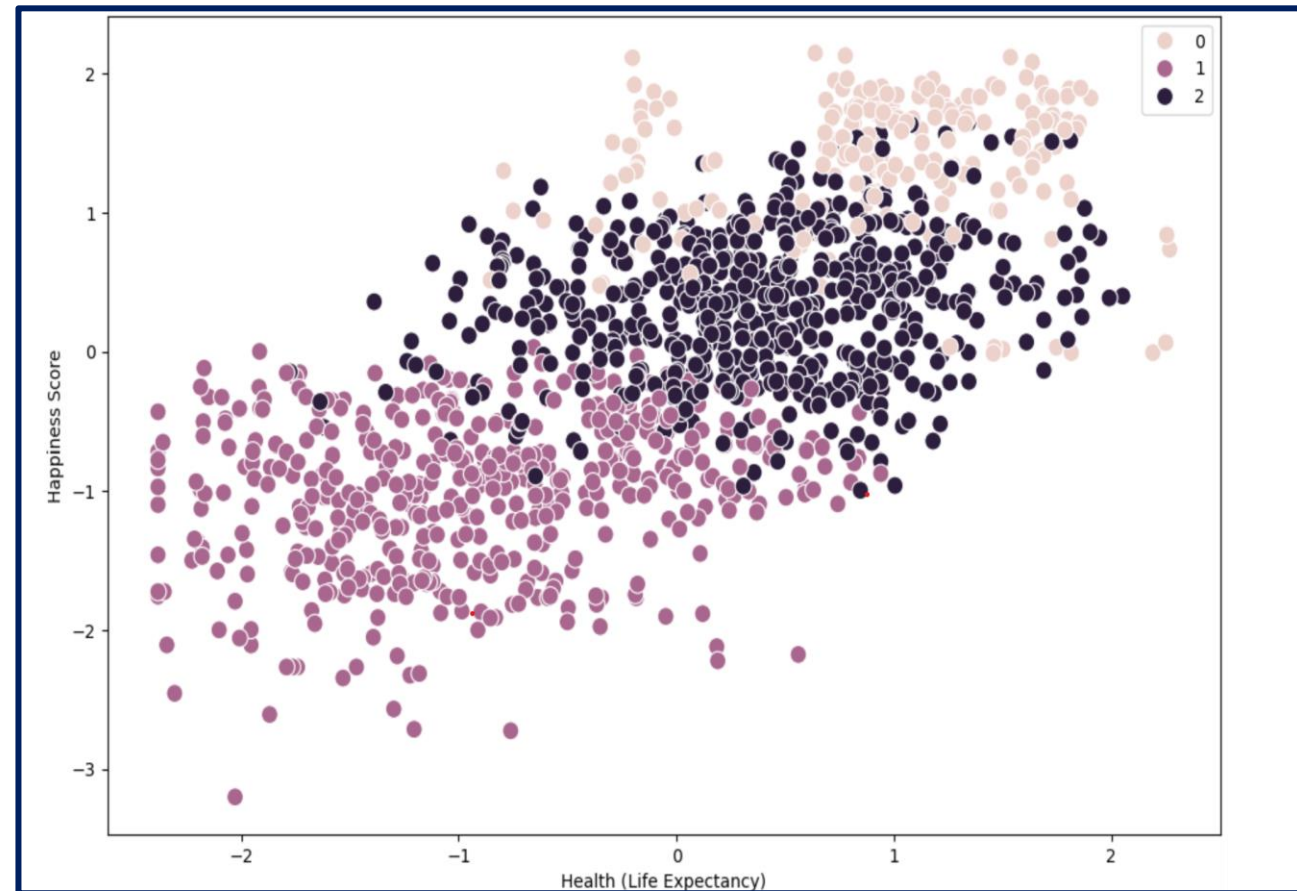
The Continent North America has the countries with the highest happiness score . The lowest happiness score countries are some Afrikan countries and South Asian countries .



# ANALYSIS AND INSIGHTS



The higher the health score the higher the happiness score. The pink cluster shows unhappy country. The purple cluster is showing somewhat happy country. The dark purple cluster shows happy country.





# ANALYSIS AND INSIGHTS



There is a positive relationship between Health ( Life Expectancy ) and a country's Happiness score.



# WORLD HAPPINESS REPORT RECOMMENDATIONS

Economy, health and family are the top three factors in determining the happiness score of the country.

In summary we can say, the happiest countries are mostly wealthy, have social support and have significantly higher GDPs and higher life expectancy.

Not every single country have been considered for this survey for the world happiness report so that could be the limitation for this project.

It would be interesting to see the effect of population on happiness score of the country for the further research.



---

THANK YOU !

