

# *Jinal Pandya*

(Data Analyst)



# Projects

## **1. GameCo**

(Analysing global video game sales.)

## **2. Influenza**

(Preparing for influenza season by providing info on staffing.)

## **3. Rockbuster Stealth**

(Answering business questions for an online video game rental company.)

## **4. Instacart**

(Marketing strategy for an online grocery store.)

## **5. Pig E. Bank**

(Anti-money laundering project for a global bank.)

# 1. GameCo

---

Market trend analysis for video game development and sales.

A fictitious video game company interested in descriptive analysis on market data to successfully develop new games.



# Project Overview

## **Objective**

To perform a descriptive analysis of video game data to support new game development by providing better understanding of how they would fare in the market.

## **Goal**

- Determine genre popularity trends
- Determine largest publisher competitors
- Analyze market trends to determine video game popularity over time
- Uncover geographic sales differences

## **Data**

The data set provided through VGChartz, can be found [here](#)

## **Skills**

Excel based analysis

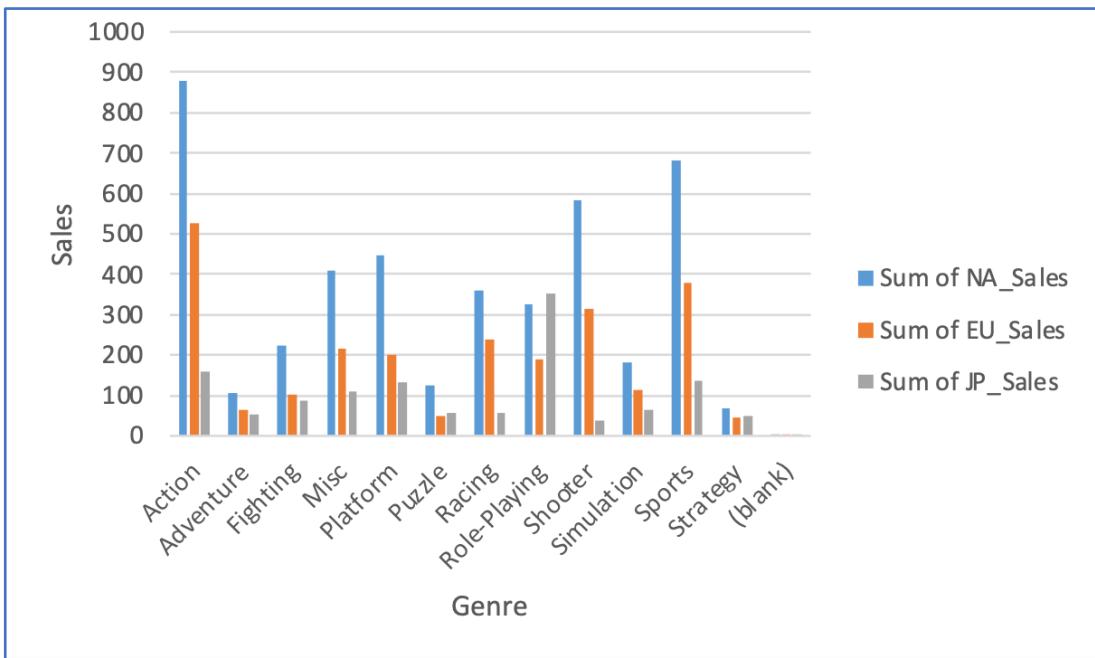
- Data cleaning techniques
- Grouping and summarizing data through pivots tables and filtering
- Conducting descriptive analysis
- Visualizing insights through scatterplots, box and whisker plots, and bar and column graphs

## **Tools**

Microsoft Office: Excel, PowerPoint & Word

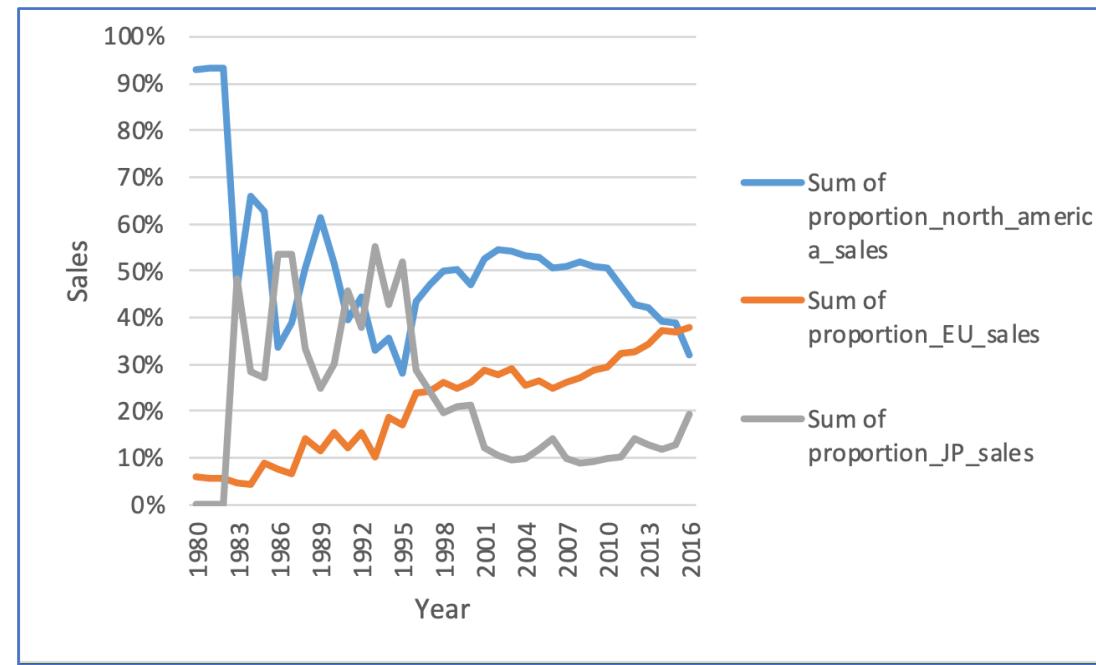
# Analysis

Top 3 genre in different geological location



Top 3 genre in North America and Europe are Action, Sport & Shooter. Top 3 genre of Japan are Role-playing, Action & Sports.

Regional sales proportion on global market



EU Sales have grown steadily over time.  
JP Sales are on the rise.  
NA Sales no longer have the highest market share.

# Recommendations

GameCo has to focus on Europe market while planning the budget for 2017. Since EU sales are taking over North American sales in 2016, focusing on this region will enhance the progress further.

The most popular genre has to be taken into consideration. It could help understanding the market and distribution can be planned as per demand in that particular geological region.

## **Final Reports**



[Project  
Reflections](#)



[Presentation](#)

## 2. Preparing for Influenza Season

Investigating trends to assist in staffing agency needs.

Assist a medical staffing agency with planning the disbursement of temporary workers to clinics and hospitals during the flu season throughout the United States



# Project Overview

## **Objective**

Determine when to send staff, and how many, to each state.

## **Goal**

To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis for influenza season.

## **Data**

Influenza deaths by geography, time, age and gender: [CDC](#)

Population data by geography: [US Census Bureau](#)

Project Brief: Preparing for Influenza Season: [Project Brief](#)

## **Skills**

Excel & Tableau based analysis

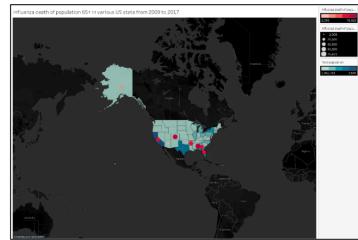
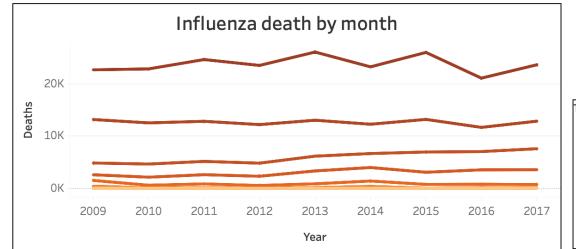
➤ Designing data research projects:

- Formulating [hypothesis](#)
- Data [profiling](#) and [quality](#) measures
- Data [transformation](#) & integration
- Conducting [statistical analysis](#) and hypothesis testing

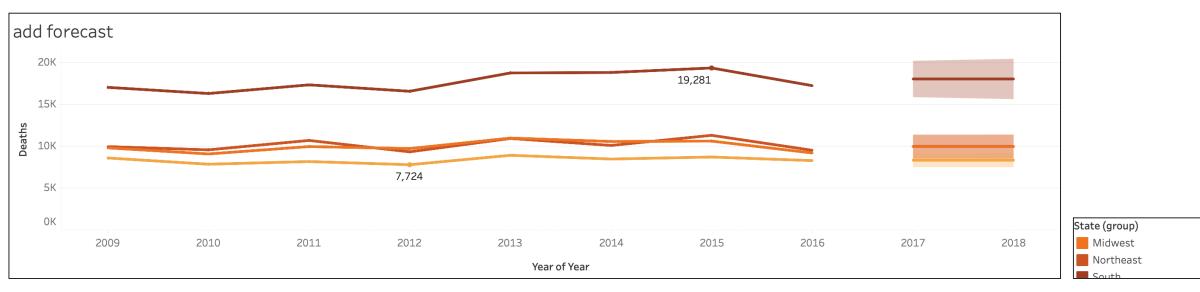
➤ Tableau Visualizations:

- Composition and [comparison](#) charts
- Temporal visualizations
- [Statistical](#) visualizations
- [Spatial](#) analysis
- [Textual](#) analysis

# Analysis



Highest influenza has been recorded in population 65 & over. Therefore, they are considered vulnerable population and are at high risk of dying



The forecast considering the data for 2018 is showing that the highest number of death is expected in south region then in West, Midwest and Northeast region respectively.

# Conclusion & Recommendations

## **Conclusion**

Influenza season is December to March and is same for all the US states.

Highest number of death recorded in population aged 65 & over.

California, New York, Pennsylvania, Florida are highest risk states.

Illinois, Texas, Ohio, North California, Michigan, Tennessee, Massachusetts, Missouri are at medium risk.

## **Recommendation**

Collect the information of medical facilities of all states.

Start preparing for the traditional medical staff before influenza season starts and distribute them as per the state's need considering the provided forecast.

## **Final Reports**



[Report](#)

[Interim Report](#)



[Tableau Presentation](#)

### 3. Rockbuster Stealth

---

Analyze database to develop competitive strategy recommendations against streaming services.

Assist fictitious movie company Rockbuster Stealth in developing a strategy to remain competitive with online streaming services.



# Project Overview

## **Objective**

Provide insights to the Marketing department with a data driven solutions to contribute to the 2020 company strategy.

## **Goal**

o help a fictitious movie rental company launch an online video rental service using its current movie rental licenses.

## **Data**

Information provided by Career Foundry that includes data on film inventory, customers, payments, and other information.

A complete data dictionary can be found [here](#)

## Data Brief

## **Skills**

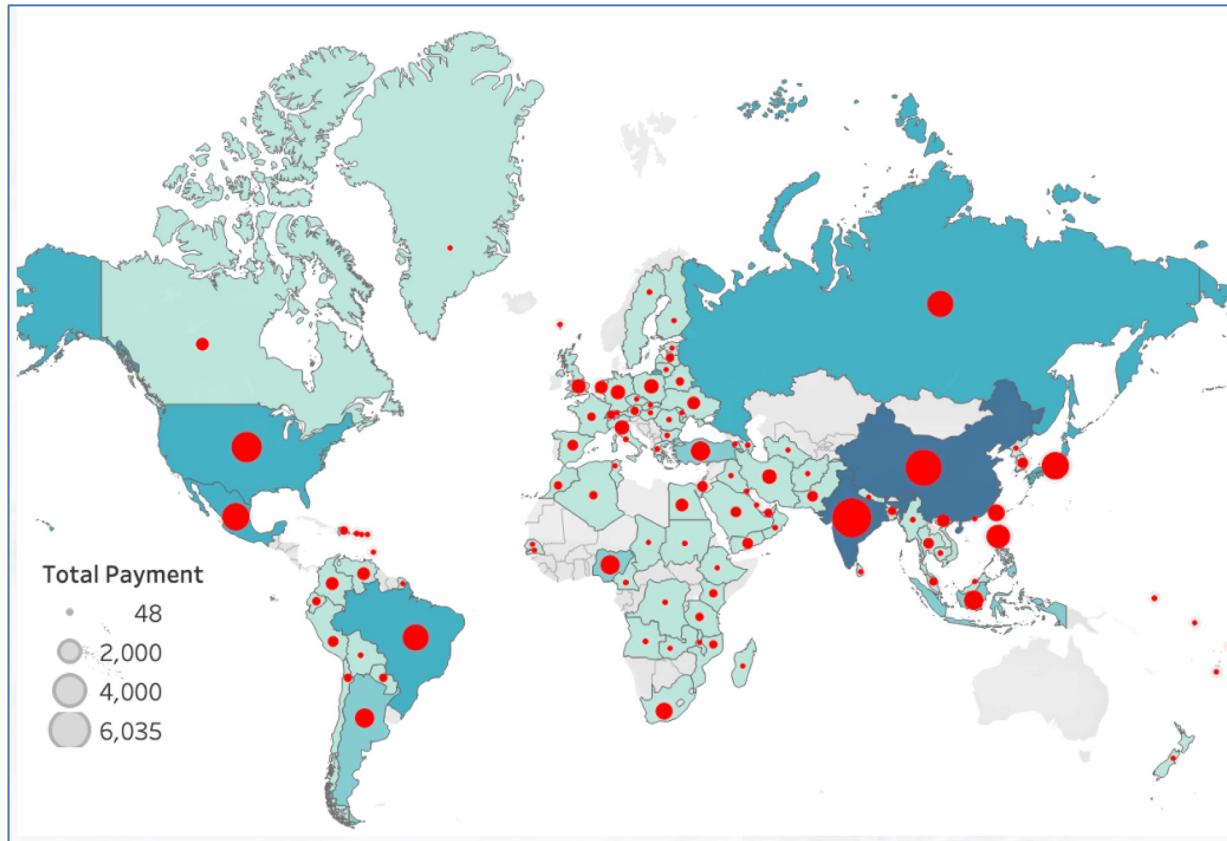
SQL based analysis

- Understand and utilize relational databases
- Query data in SQL by ordering, limiting, and grouping data
- Filter data using WHERE and HAVING clauses
- Identify and clean dirty data
- Join tables
- Perform subqueries and Common Table Expressions
- Present findings

# Analysis

According to the data set, there are 599 customers across 108 countries, with majority residing in Asia and the pacific.

## Sales distribution by country



| Country       | Amount  |
|---------------|---------|
| India         | \$6,035 |
| China         | \$5,251 |
| United States | \$3,681 |
| Japan         | \$3,123 |
| Mexico        | \$2,985 |
| Brazil        | \$2,919 |
| Russia        | \$2,766 |
| Philippines   | \$2,220 |
| Turkey        | \$1,498 |
| Indonesia     | \$1,353 |

Asia has the highest customer count.  
South and East Asia have the highest concentration of customers.



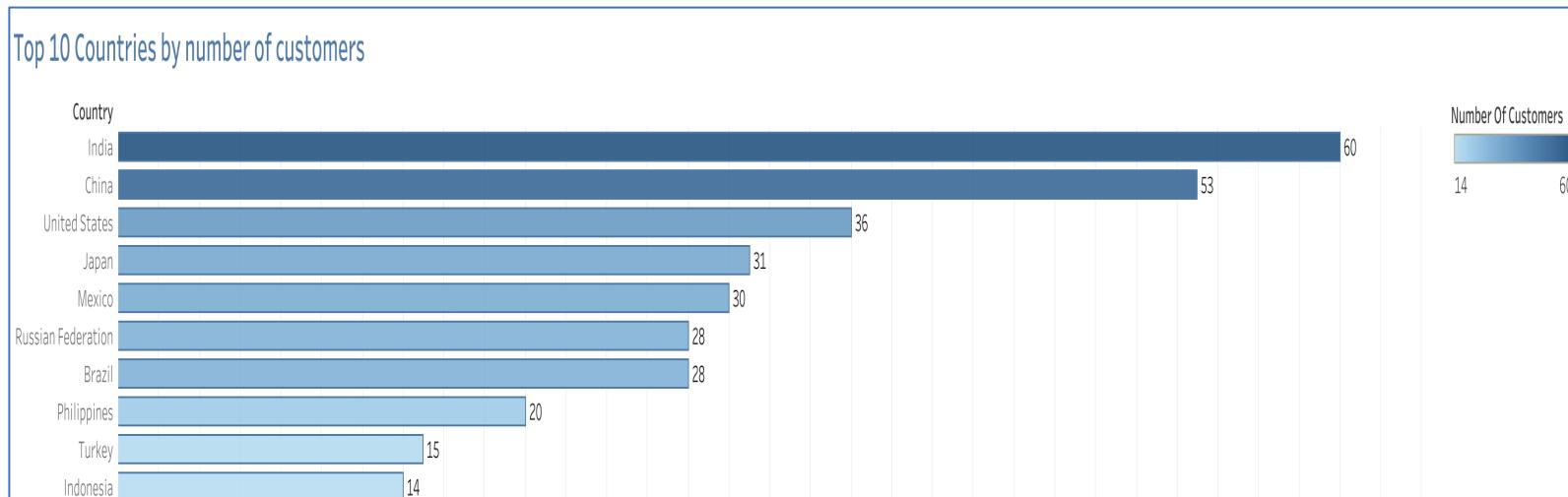
# Analysis

Query    Query History

```
1 SELECT D.country,COUNT(customer_id) AS "number_of_customers"
2 FROM customer A
3 INNER JOIN address B ON A.address_id = B.address_id
4 INNER JOIN city C ON B.city_id = C.city_id
5 INNER JOIN country D ON C.country_id = D.country_id
6 GROUP BY country
7 ORDER BY COUNT(customer_id) DESC
8 LIMIT 10
```

Data Output    Messages    Notifications

|    | country            | character varying (50) | number_of_customers | bigint |
|----|--------------------|------------------------|---------------------|--------|
| 1  | India              |                        | 60                  |        |
| 2  | China              |                        | 53                  |        |
| 3  | United States      |                        | 36                  |        |
| 4  | Japan              |                        | 31                  |        |
| 5  | Mexico             |                        | 30                  |        |
| 6  | Brazil             |                        | 28                  |        |
| 7  | Russian Federation |                        | 28                  |        |
| 8  | Philippines        |                        | 20                  |        |
| 9  | Turkey             |                        | 15                  |        |
| 10 | Indonesia          |                        | 14                  |        |



# Conclusion & Recommendations

## **Conclusion:**

The top 5 revenue making customers are: Arlene Harvey, Kyle Spurlock, Marlene Welch, Glen Talbert, Clinton Buford.

The leading genres for India, United States and Mexico are “Sports”. Whereas in China it’s “Animation”, and in Japan it’s “New”.

The most rented films have PG-13 MPAA rating. Average length of the rented movies is 113 to 116 minutes therefore, it is good to provide more such movies with the similar length.

## **Recommendations:**

To get noticed advertise the upcoming new launch of online video rental service in various platform depending on its popularity in particular country.

## **Final Reports**



[SQL queries and results](#)



[Presentation](#)



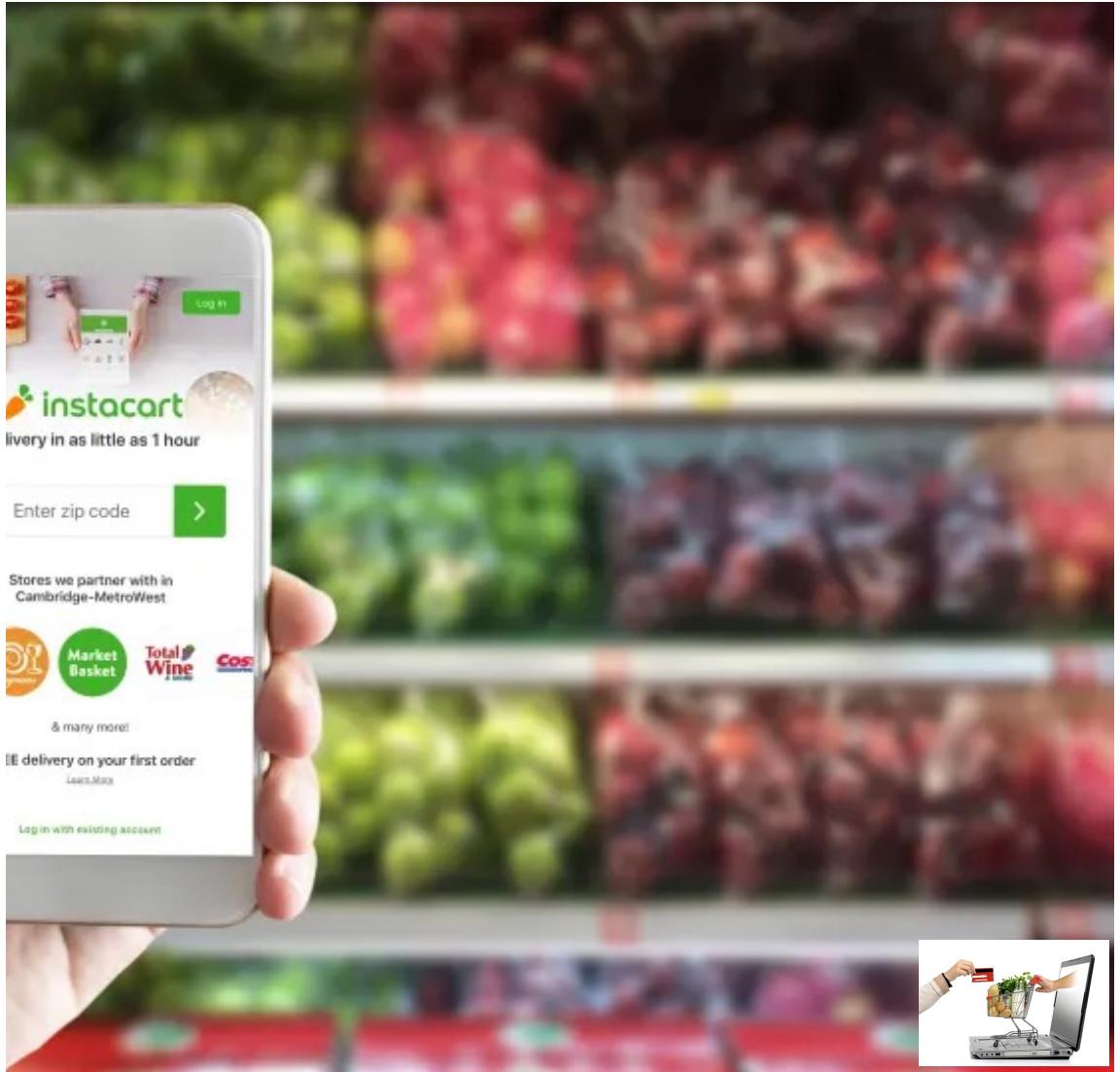
[Data Dictionary](#)



<https://github.com/Jinal166/SQ-L-ROCKBUSTER>

## 4. Instacart

Uncover sales patterns to suggest customer segmentation strategies and advertisement scheduling



# Project Overview

## **Objective**

To uncover more information about the sales patterns of an online grocery store that operates through an app.

## **Goal**

To derive insights and suggest strategies for better customer segmentation of online grocery store.

## **Data**

[customers.csv](#)

[Project Brief](#)

## **Skills**

Python based analysis

- Importing libraries, including [Pandas](#), [Numpy](#), [Matplotlib](#), and [Seaborn](#)

- Conducting descriptive [exploratory](#) tasks
- [Data wrangling](#) and subsetting
- Conducting data [consistency](#) checks

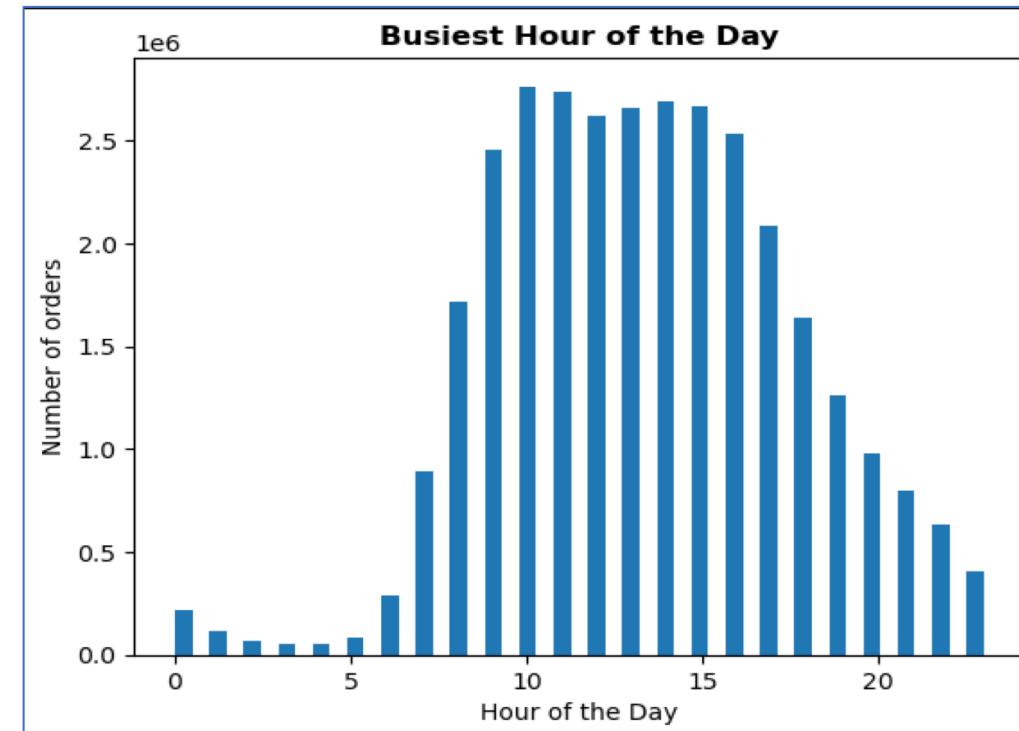
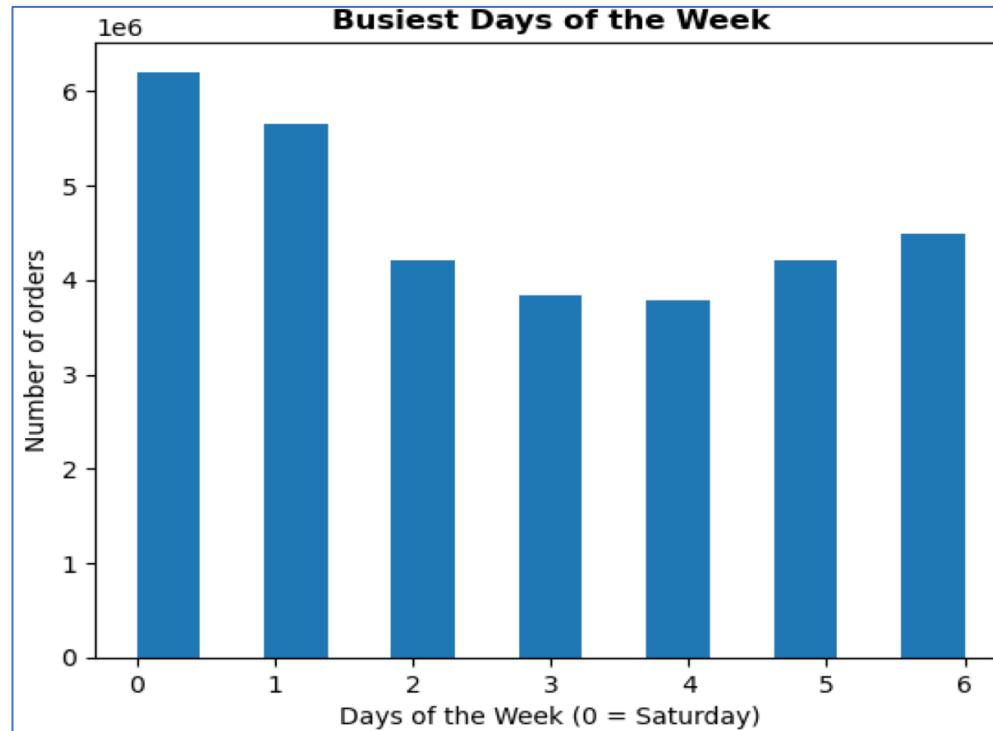
- Combining and exporting data

- Deriving new variables using [conditional logic](#)

- [Grouping](#) and aggregating data

- [Data visualization](#) in Python

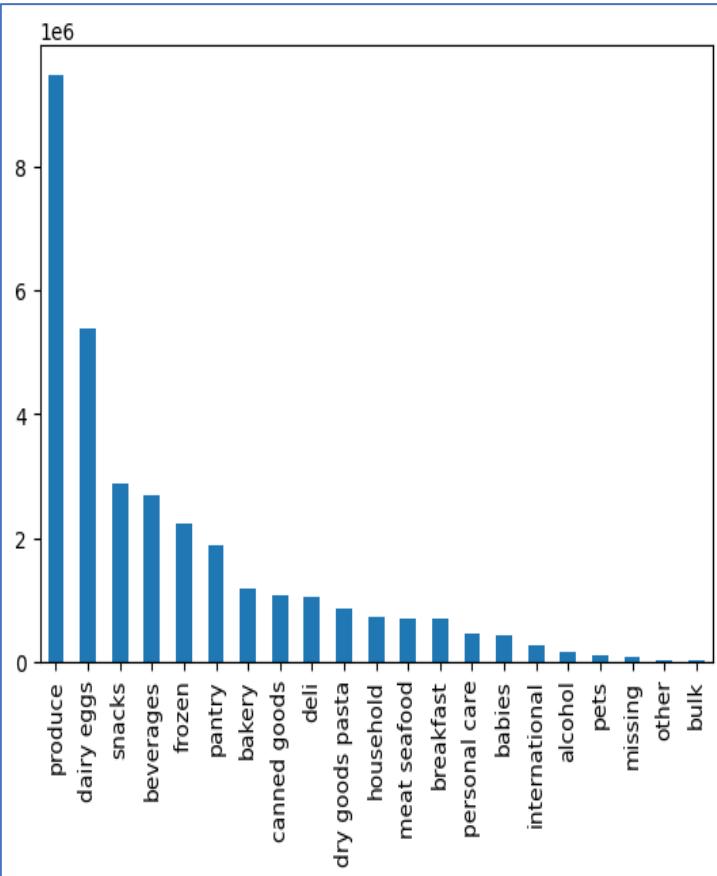
# Analysis



The largest number of orders occur on the weekends, and Instacart receives the most orders between 10:00 AM to 3:00 PM

# Analysis

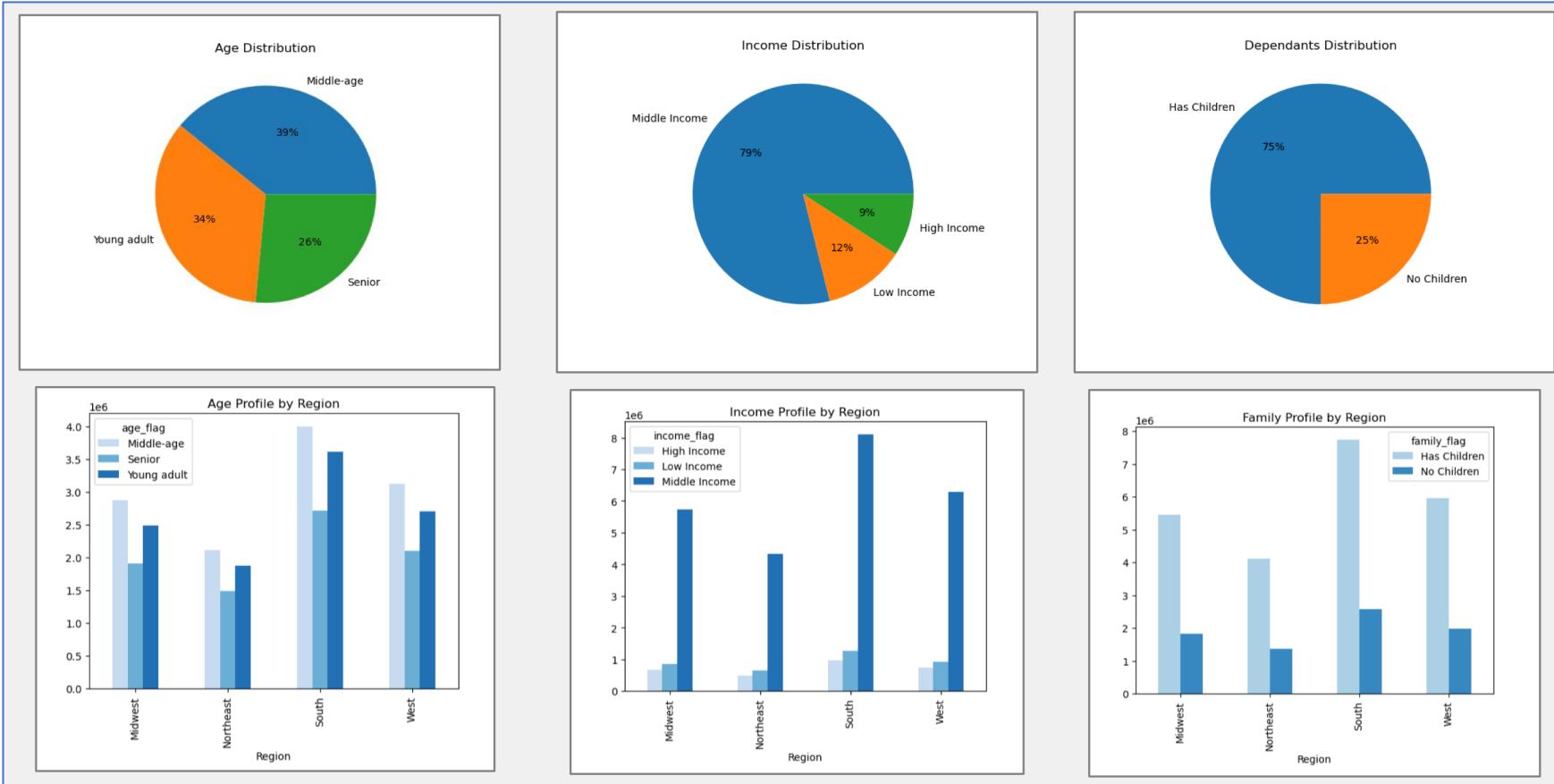
Produce, dairy/eggs, and snacks have the highest number of orders. However, alcohol, products from the baby's department, and bulk products have the fewest number of days between orders, and users spend the most on meat/seafood, dairy/eggs, and bulk items.



| department      | days_since_last_order |     |          |          | prices |      |          |
|-----------------|-----------------------|-----|----------|----------|--------|------|----------|
|                 | min                   | max | mean     | sum      | min    | max  | mean     |
| alcohol         | 0                     | 30  | 9.332472 | 1179302  | 1      | 15   | 8.151446 |
| babies          | 0                     | 30  | 9.335526 | 3133096  | 1      | 15   | 7.634399 |
| bakery          | 0                     | 30  | 10.22586 | 8852225  | 1      | 15   | 7.868462 |
| beverages       | 0                     | 30  | 10.2061  | 19759202 | 1      | 15   | 7.682624 |
| breakfast       | 0                     | 30  | 10.38948 | 5470900  | 1      | 14.9 | 8.068044 |
| bulk            | 0                     | 30  | 9.451048 | 279212.8 | 1.4    | 14.1 | 8.346919 |
| canned goods    | 0                     | 30  | 10.67252 | 7637779  | 1      | 15   | 7.54666  |
| dairy eggs      | 0                     | 30  | 10.15542 | 43363062 | 1      | 15   | 8.352176 |
| deli            | 0                     | 30  | 10.43178 | 7813188  | 1      | 15   | 7.783347 |
| dry goods pasta | 0                     | 30  | 10.77043 | 6042181  | 1      | 15   | 7.34937  |
| frozen          | 0                     | 30  | 10.74014 | 16431751 | 1      | 15   | 7.738627 |
| household       | 0                     | 30  | 10.9688  | 5164912  | 1      | 15   | 7.379943 |
| international   | 0                     | 30  | 10.31608 | 1965703  | 1      | 15   | 7.678797 |
| meat seafood    | 0                     | 30  | 10.50541 | 10999037 | 8      | 25   | 16.30006 |
| missing         | 0                     | 30  | 9.659415 | 560933.3 | 1      | 15   | 8.660655 |
| other           | 0                     | 30  | 10.05684 | 239569.2 | 1.1    | 15   | 6.961995 |
| pantry          | 0                     | 30  | 10.22027 | 14289861 | 1      | 20   | 8.01494  |
| personal care   | 0                     | 30  | 10.60479 | 3394733  | 1      | 15   | 7.998617 |
| pets            | 0                     | 30  | 10.93915 | 734083.7 | 1      | 15   | 7.88786  |
| produce         | 0                     | 30  | 10.02607 | 72455791 | 1      | 15   | 7.980349 |
| snacks          | 0                     | 30  | 10.10842 | 11827920 | 1.6    | 7    | 4.275555 |

# Analysis

Instacart's primary user group is middle-aged, middle-income families.



# Conclusion & Recommendations

To boost sales during lower traffic times, Instacart should schedule advertisements for Tuesdays and Wednesdays after 4:00 PM.

Produce and dairy/eggs are already a part of most people's preferences, it is unlikely that they need to spend a lot of advertising dollars convincing people to purchase them. It may make more sense to increase advertisements of meat and seafood, where prices are higher, and orders are not as high to attract more revenue.

Instacart's largest customer base is middle-age, middle-income families. They should decide whether to target this group further and investigate preferred products and shopping times or choose to attempt to expand their customer base.

## Final Reports



[Final Excel Report](#)



[https://github.com/  
Jinal166/Python-  
Instacart](https://github.com/Jinal166/Python-Instacart)

## 5. Pig E. Bank

---

- Dive into Data Ethics issues and Data Mining
- (Use principals of data ethics to assist Pig E. Bank in navigating challenging issues and begin exploring the use of data mining and predictive analysis.)



# Project Overview

## **Objective**

Identify factors for leaving the bank and analyse statistical information for 2 groups (those who left vs those who stayed) to build a decision tree.

## **Goal**

To build a decision tree for predicting which clients might leave the bank using data mining mechanism.

## **Data**

Pig E. Bank's client data set, found [here](#)

## **Skills**

Explore data ethics issues such as [data bias, security & privacy](#)

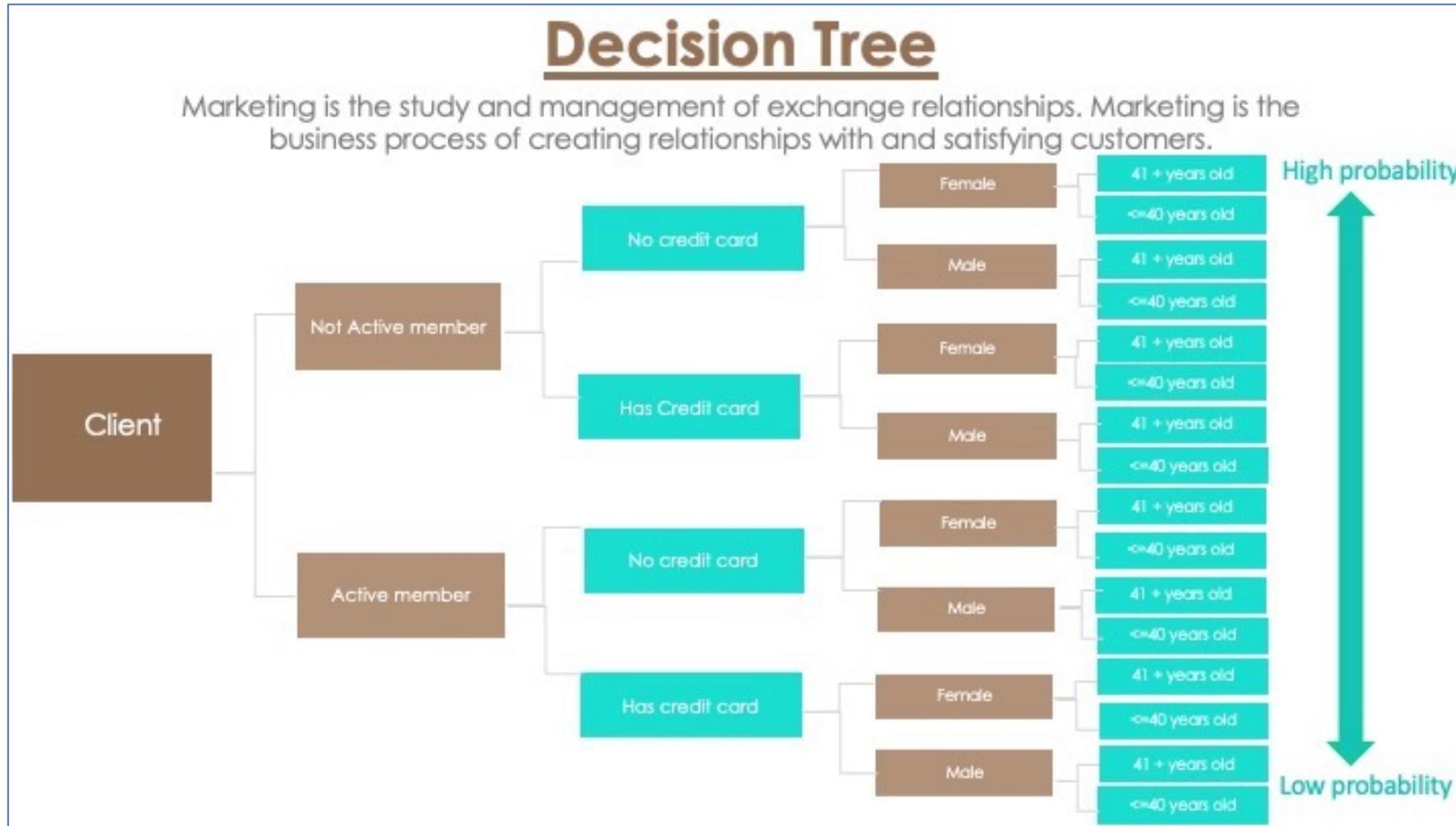
Begin exploring [data mining](#) and usage of decision trees

Understand and utilize [CRISP- DM methodology](#)

Utilize [time-series analysis](#)

# Analysis

According to a descriptive analysis on Pig E. Bank's data, the likelihood of a client leaving can be estimated using the below decision tree algorithm



# Additional Analysis files

Descriptive Analysis of Pig E. Bank's client data [Descriptive Analysis](#)

Time Series Analysis [Time Series Analysis](#)

Stationary vs. non-stationary exploration, moving average forecasting and introduction into Forecasting models

Data Bias Case Study [Data Bias Case Study](#)

Exploration into various data bias types

Data Privacy & Security case study [Data privacy & security](#)

Senario analysis for Pig E. Bank data privacy and security issues



## 5. World Happiness Report

---

- Discover the happiest countries of the world.



# Project Overview

## **Objective**

To analyze the happiest countries of the world and which factors impact the most on the happiest score.

## **Goal**

Discover the happiest countries of the world.

## **Data**

Dataset has been taken from the open-source platform Kaggle.com. and can be found [here](#)

## **Skills**

Python, Tableau Anaconda (Jupiter Notebook) based analysis.

Data Cleaning

Data Manipulation

Exploratory Data Analysis

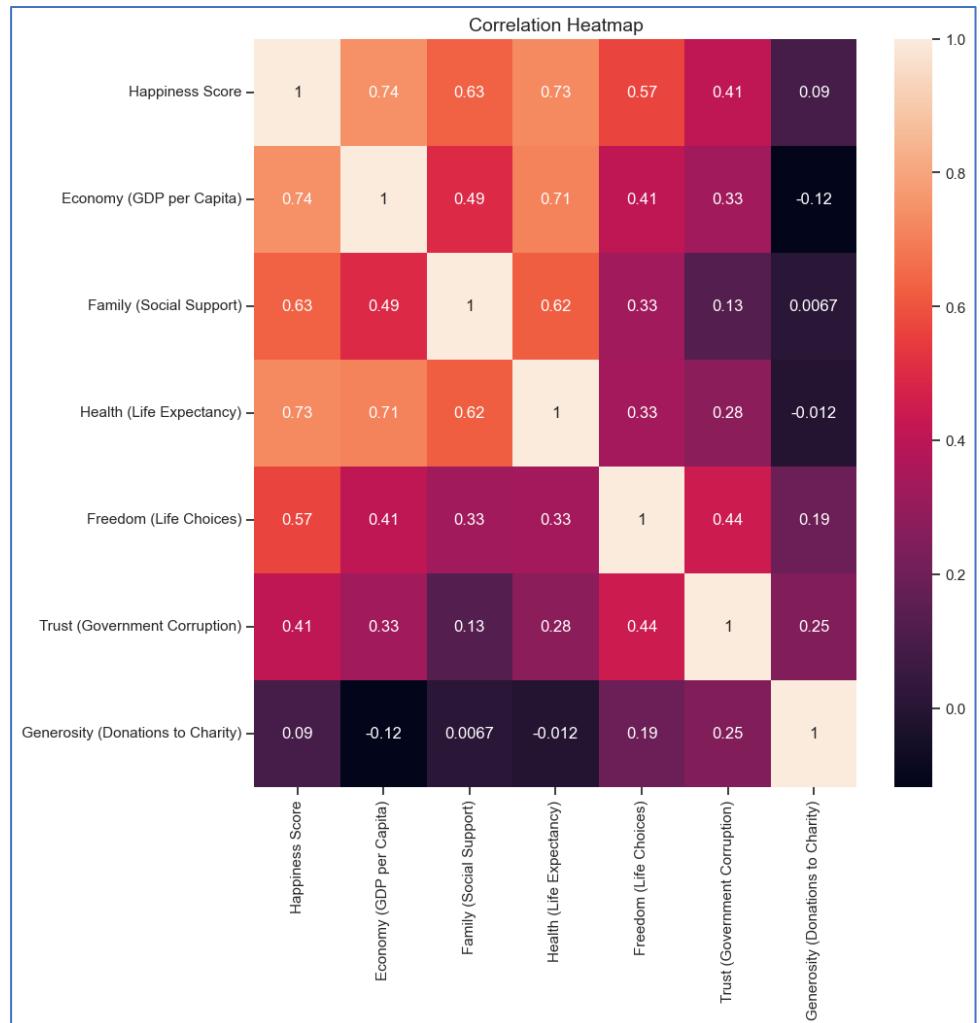
Geospatial Analysis

Supervised & Unsupervised Machine Learning (Linear Regression & Clustering)

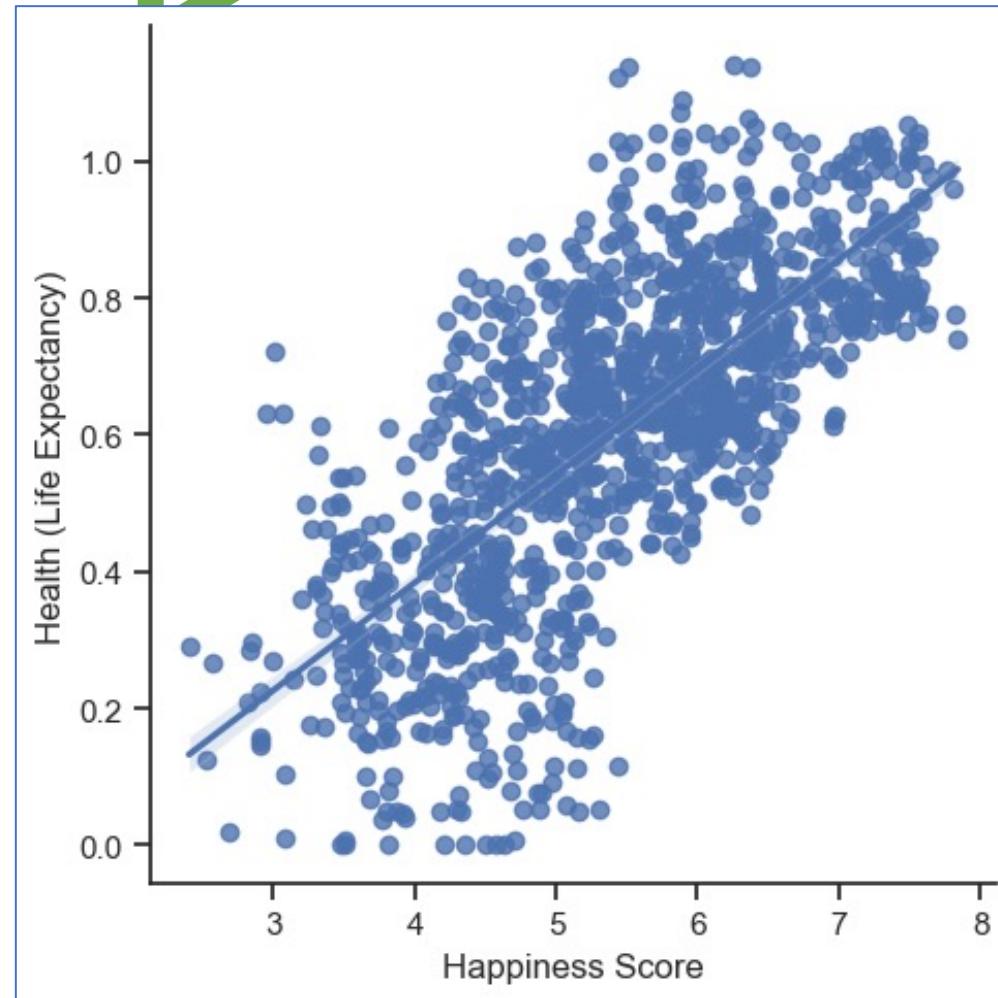
Time-series Analysis

Data Visualization & Storytelling

# Exploratory Data Analysis

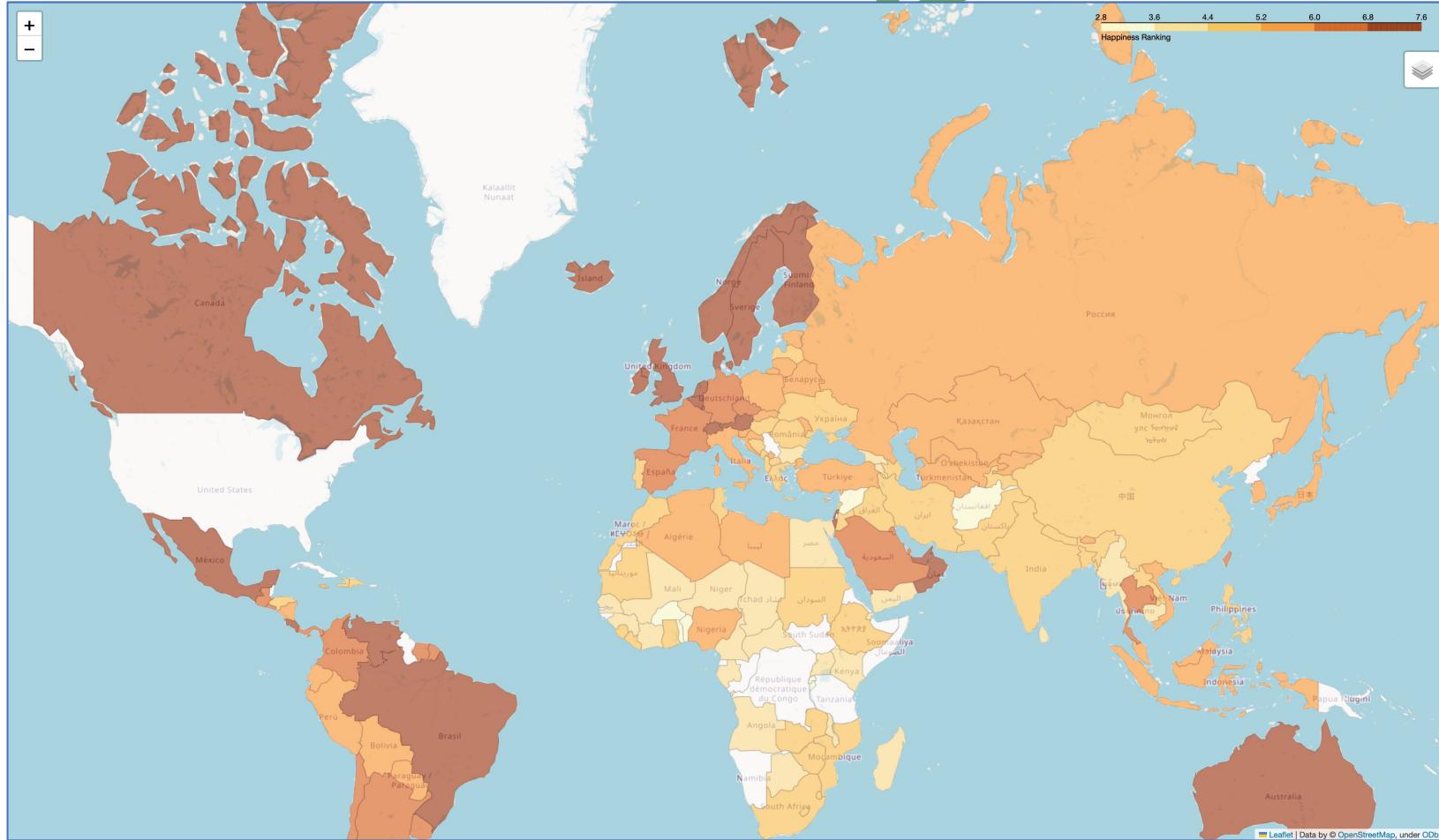


An important part of an analysis is to explore the relationships between key variables. A correlation heat map utilizing Python was created to verify which variables had the strongest relationships.



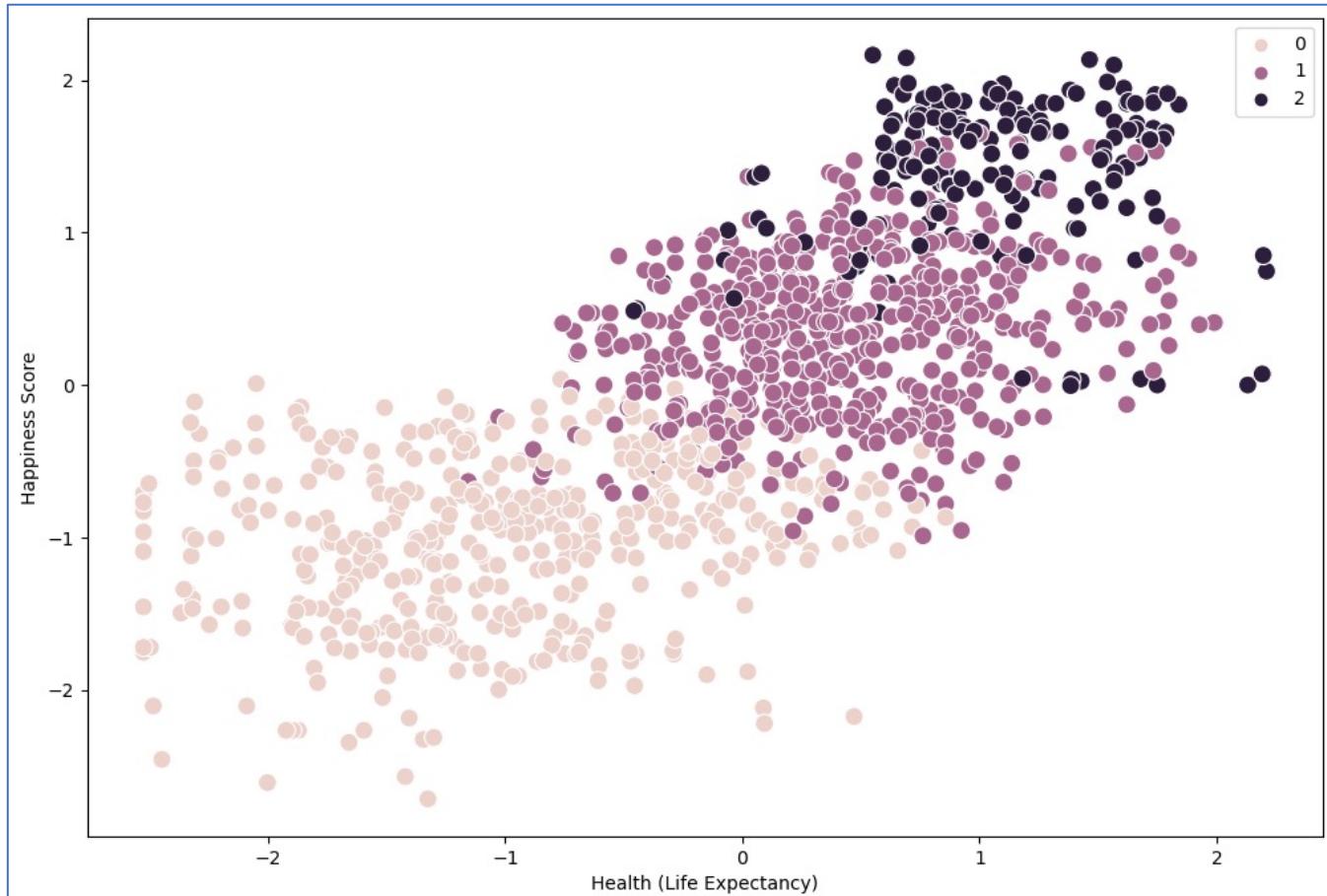
The formation of datapoints in the scatterplot indicates that the two variables are in a strong positive relationship: as one of them increases, the other increases as well.

# Geographical visualization



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

# K-Means Clustering and Linear Regression



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.



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

# Conclusion & 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.

## **Final Reports**



[https://github.com/  
Jinal166/World-  
Happiness-Report](https://github.com/Jinal166/World-Happiness-Report)

|||  
*Thank you*



<https://github.com/Jinal166>



<https://public.tableau.com/app/profile/jinal.pandya>

