DATA ANALYTICS PORTFOLIO

PROJECT CASE STUDIES

By Brian Avila

TABLE OF CONTENTS

SKILLSETS

My Skillsets

- Data Analysis
- Tableau Visualizations
- SQL Relational Databases
- PostgreSQL
- Python
- MS Excel (Pivot Tables, VLOOKUP)
- Time Series Analysis and
- Forecasting
- Statistical Analysis
- Data Transformation and integration

INFLUENZA MEDICAL STAFFING PLAN

LET'S DIVE IN

- **GOAL**: To help a medical staffing agency that provides temporary workers to clinics and hospitals on an as-needed basis. The analysis will help plan for influenza season, a time when additional staff are in high demand. The final results will examine trends in influenza and how they can be used to proactively plan for staffing needs across the country.

ANALYSIS

Data Sets

Influenza deaths by geography, time, age, and gender

Source: CDC

Download Data Set

2. Population data by geography

Source: US Census Bureau

Download Data Set

Tools

Microsoft Excel

tableau

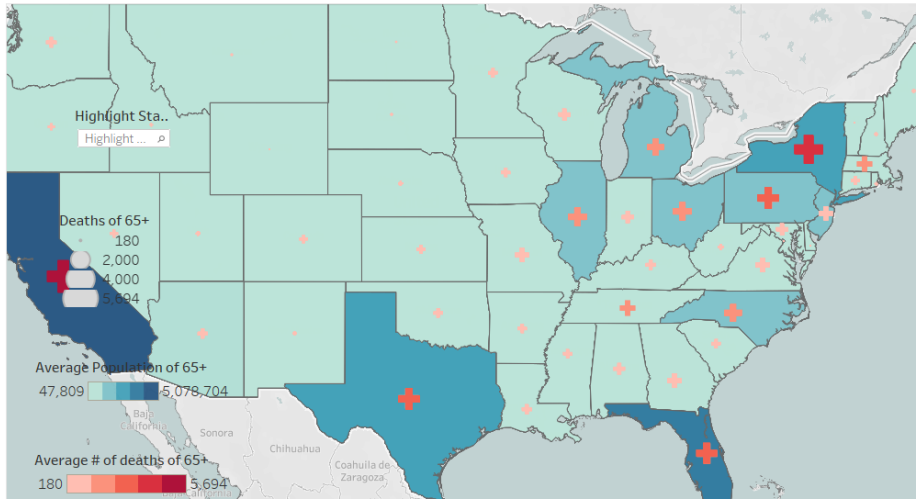
Process

1. Listed the data questions to be answered.
2. Designed a Data Research Project
3. Formulated a research hypothesis
4. Sourced the Right Data
5. Created a data profile
6. Implemented data quality measures
7. Integrated data from 2 sources into 1.
8. Conducted Statistical Analysis by calculating variance and standard deviation
9. Formulated a statistical hypothesis
10. Created an interim report

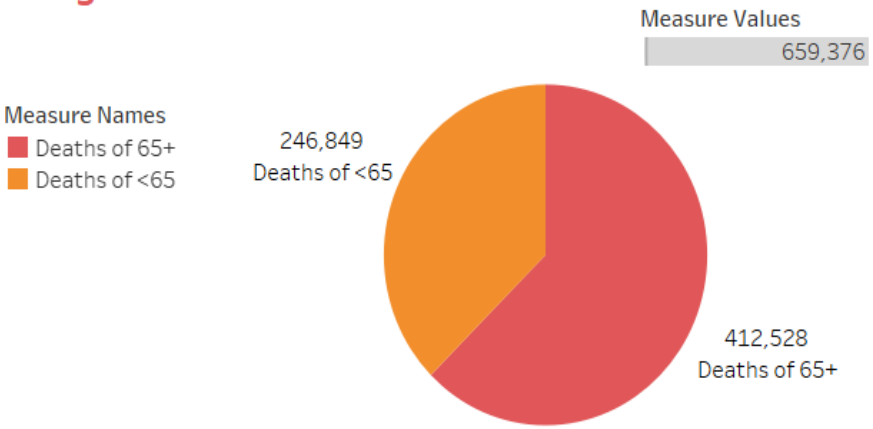
VISUALIZATIONS & STORYTELLING

Map of Influenza Deaths on Average Per Year Between 2009 and 2017 (65+ Yea Age)

-The states with a darker colour fill have a higher population over 65 years of age.
-The red (+) symbol represents the number of deaths. The larger and darker the colour of the (+) symbol, t higher the death rate.



Total Influenza Deaths for Populations under and over 65 Years of Age



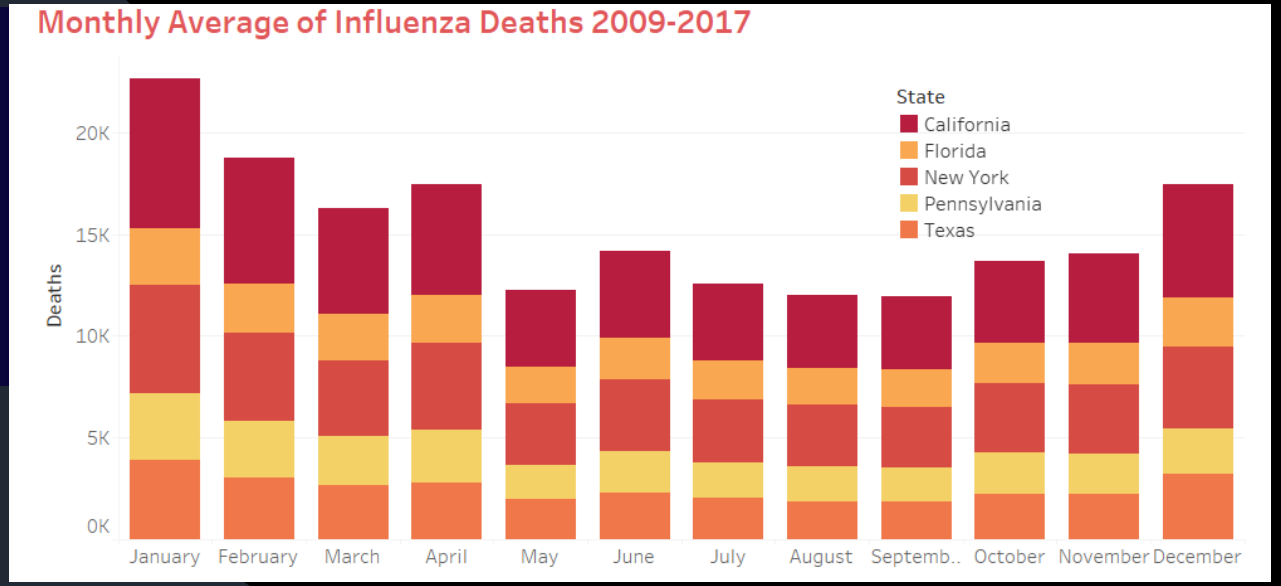
-The pie chart shows that the number of Deaths for individuals over 65 is significantly larger than those under 65 years of age

Influenza Deaths by Age Groups (2009-2017)

The median number of deaths increases considerably for age groups over 65 years of age

CONCLUSION

- The analysis conducted revealed that Influenza poses a threat to populations over 65 years of age.
- The states more populous with this vulnerable age group are California, New York, Texas, Pennsylvania, and Florida.
- The months the Influenza Virus is more prominent are December, January, February, March, and April.
- **The states of California, New York, Texas, Florida, and Pennsylvania have historically high mortality due to Influenza.**
- **It was recommended that medical staff should be sent to these hotspots during the peak months of December through April.**



Project link

ROCKBUSTER STEALTH SALES ANALYSIS

LET'S EXPLORE

Goal: The Rockbuster Stealth Management Board has asked a series of business questions and they expect data-driven answers that they can use for their 2020 company strategy.

- 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?

ANALYSIS

Data Sets

Rockbuster Data Set

Download Data Set

Tools

Microsoft Excel

PostgreSQL

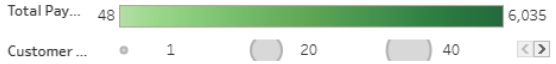
Process

1. Set up a database environment using the PostgreSQL
2. Created a data dictionary
3. Wrote SQL commands in PostgreSQL to answer business questions and organize and sort data
4. Filtered and ordered data using the WHERE and HAVING clauses
5. Created a data profile of summary statistics using SQL
6. Wrote subqueries to answer complex business questions
7. Rewrote subqueries as Common Table Expressions
8. Created a presentation of findings

VISUALIZATIONS OF SQL RESULTS

Customers

Worldwide Distribution of Customers



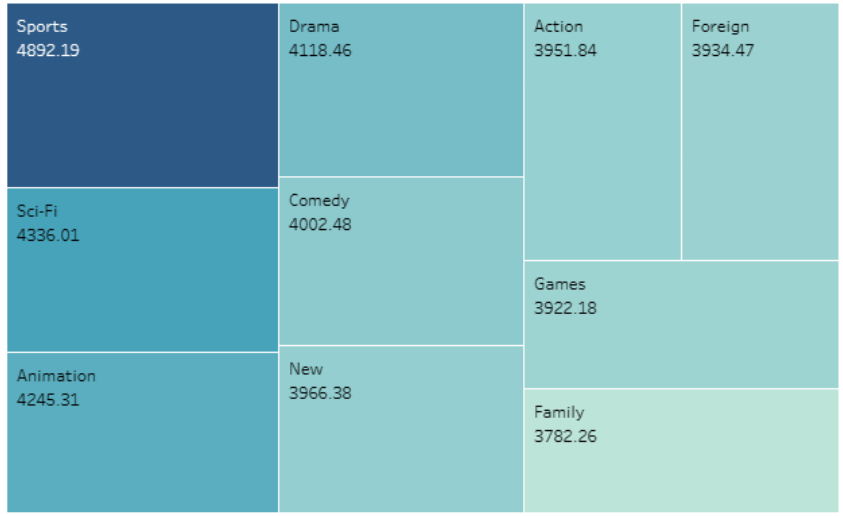
-Rockbuster customers are distributed across the globe.

-As can be observed on the geographical map, most customers are situated in the populous countries of India and China.

Currently there are 0 customers located in Australia. This should be a next target on Rockbuster's list in order to continue expanding its business globally..

Rockbuster Video Sales.twbx

Revenue Based on Genre of Movies Sold

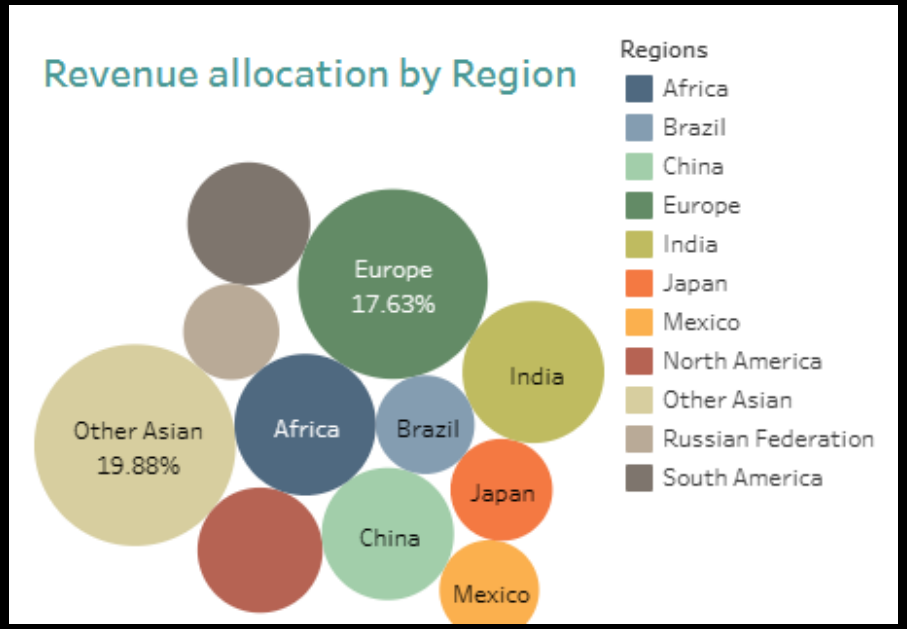


-Movies in the sports category lead in terms of revenue

-Family genre movies generate the least amount of revenue

CONCLUSION

- India and China combined contribute to 1/5th of the total revenue generated from all sales.
- The Asian market makes up to 40% of all revenue and establishes itself as the principal source of revenue from all Rockbuster sales. Continuation of marketing efforts can still potentially increase the sales numbers in the region.
- Also, expansion of marketing campaigns into countries where there are no registered customers, such as Australia, can generate more sources of revenue.



Project Link

PIG E BANK

LOOKING AHEAD

Goal: To increase customer retention, the sales team wants to identify the leading indicators that a customer will leave the bank

ANALYSIS

Data Sets

Pig E Bank Clients Data

Download Data Set

Tools

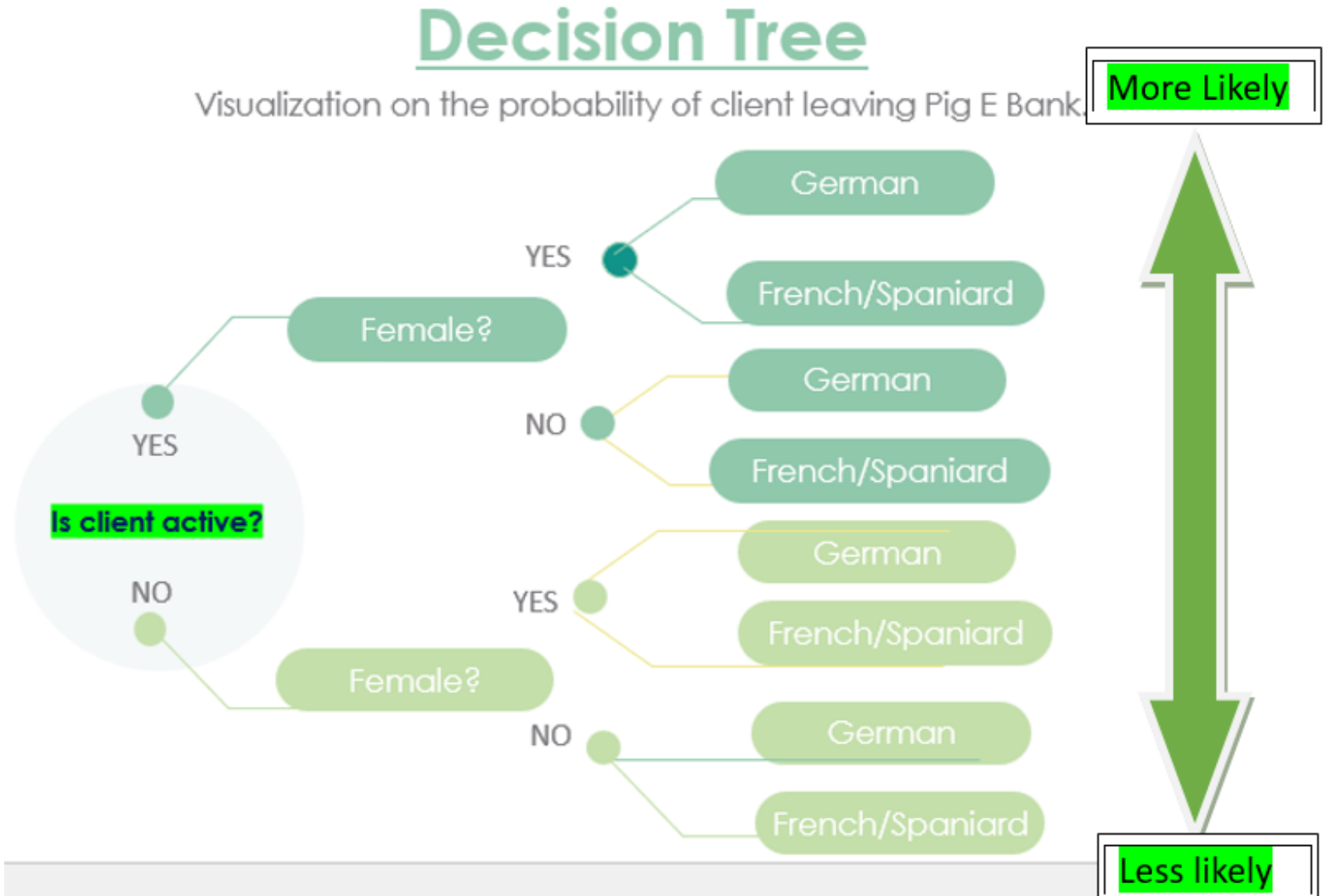
Microsoft Excel

GitHub

Process

1. Researched Software tools for handling big data
2. Suggested ways of controlling for bias and communicating concerns to stakeholders
3. Carried out steps in the data mining process, including data cleaning and descriptive statistics
4. Created a decision tree model to test the outcomes of an analysis
5. Analyzed the output of a linear regression and Identified the correct predictive **model** for different scenarios
6. Created a time series and a simple moving average in Excel
7. Created a GitHub account and repositories

VISUALIZATIONS



CONCLUSION

- Based on the analysis of the data the determining factors for client loss are:
  - Inactivity (inactive clients are more likely to leave Pig E Bank)
  - Gender (Female client loss is disproportionate to males)
  - Nationality, (75 out of 178 German clients left the bank, highest percentage).

Exited from bank			
Row Labels	Count of Gender	Average of Credit Score	Average of Age
Male	83	643	46
Female	121	632	45
Grand Total	204	637	45

Row Labels	Sum of ExitedFromBank?	Sum of IsActiveMember	Average of Age
France	77	21	46
Germany	75	25	45
Spain	52	15	45
Grand Total	204	61	45

Project Link

INSTACART GROCERY BASKET ANALYSIS

LET'S TAKE A LOOK

GOAL: To obtain informational knowledge about who the customers are and their purchasing behaviors. To build an effective marketing strategy we need to analyze customers spending habits.

- The sales team needs to know what the busiest days of the week and hours of the day are to schedule ads at times when there are fewer orders.
- They also want to know whether there are particular times of the day when people spend the most money, as this might inform the type of products they advertise at these times.
- Marketing and sales want to use simpler price range groupings to help direct their efforts.
- Are there certain types of products that are more popular than others?

ANALYSIS

Data Sets

Instacart Market Data

Download Data Sets

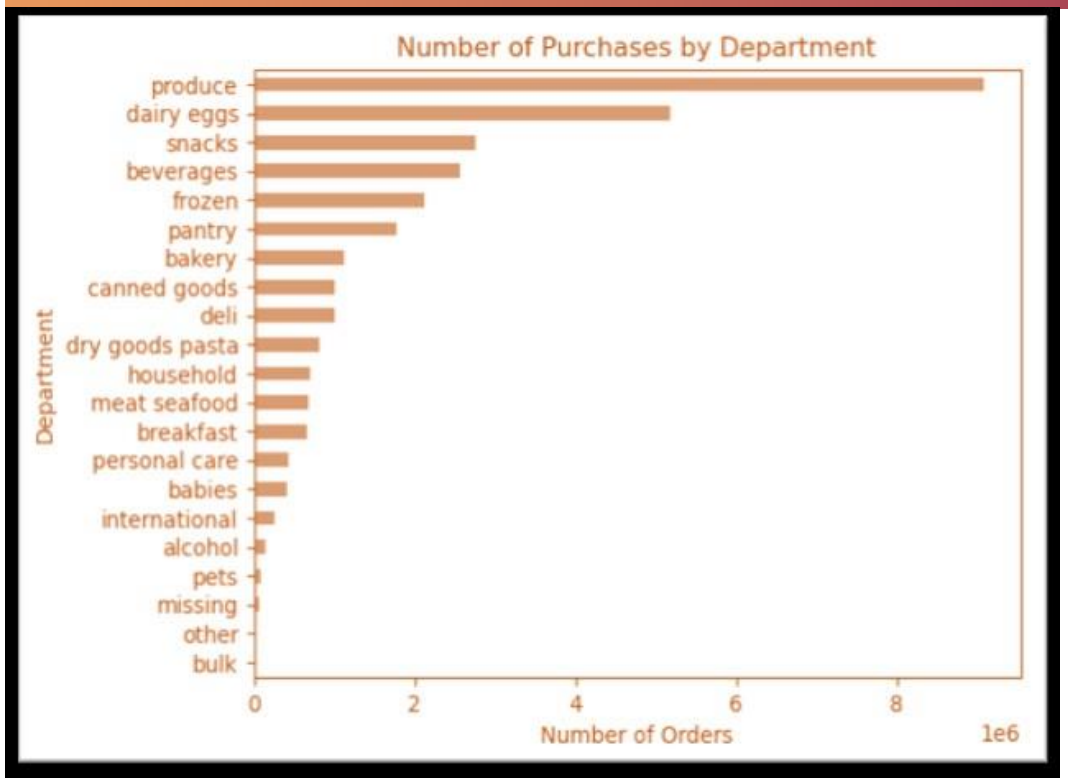
Tools

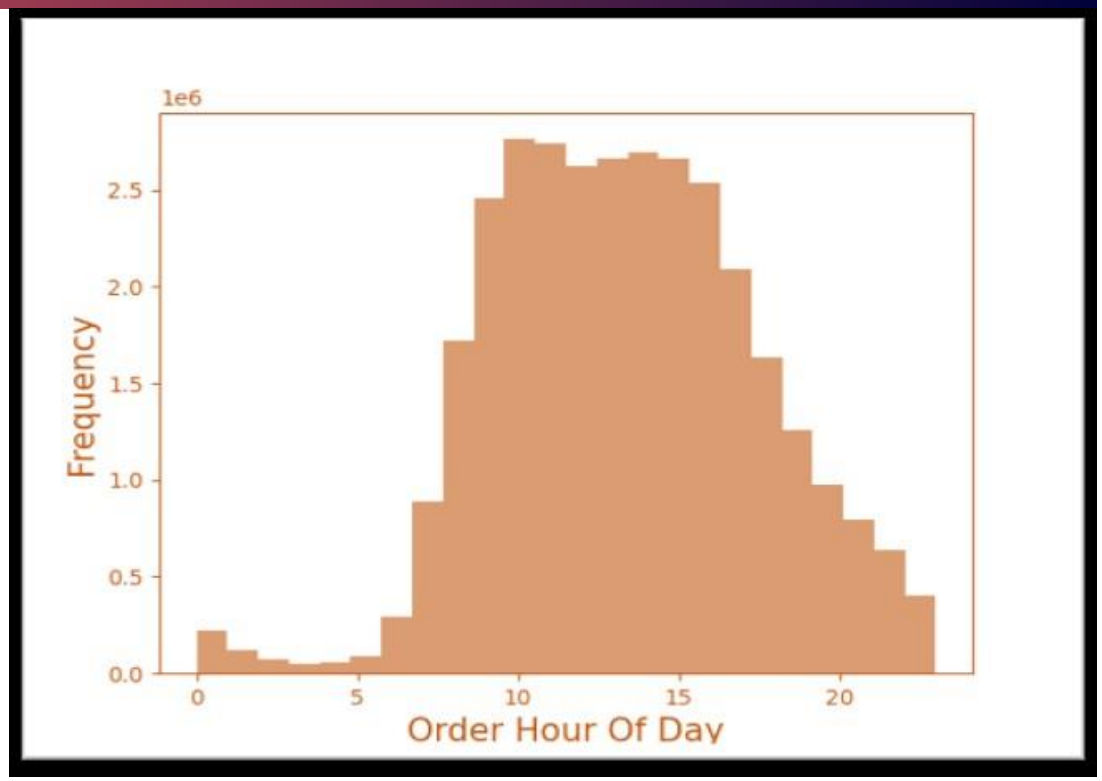
- Microsoft Excel

Process

1. Downloaded data and imported into notebook as a panda's data frame
2. Conducted basic descriptive exploratory tasks
3. Data Wrangling
4. Checked Data for Consistency
5. Analyzed results from merge flag frequencies
6. Exported merged data as a pickle
7. Grouped and Aggregated Data
8. Summarized analysis findings
9. Created a report describing analysis results, and recommendations for Instacart stakeholders

VISUALIZATIONS



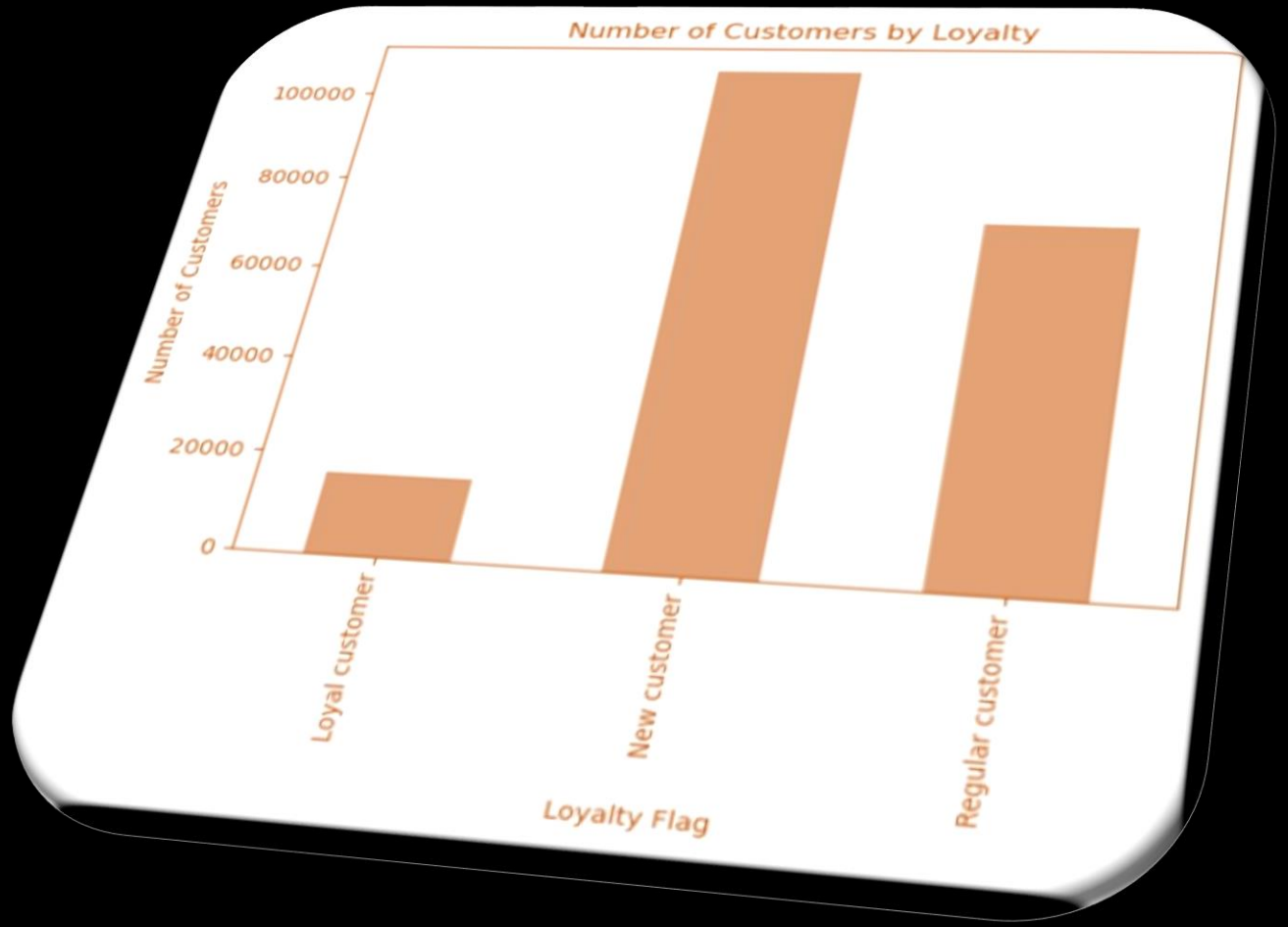


The produce department receives the most orders followed by dairy then snacks.

According to the chart above we can observe that most purchases occur between 9am and 4pm

CONCLUSION

- Most orders occur from 9 AM to 4 PM and on weekends. The sales team can schedule ads during off-peak hours to make the most impact when orders are fewer.
- Most products are priced between $2 and $15. This highlights a strong preference among customers for affordable and moderately priced items
- Focusing promotional efforts on high-demand departments like Produce, Dairy & Eggs, and Snacks products will enable us to create effective marketing campaigns that align with customer needs.
- We suggest initiatives to attract new customers, enticing them with appealing offers to make their initial purchases and introduce them to the benefits of loyalty programs.



Project Link

MAIN FACTORS CONTRIBUTING TO HAPPINESS

LET'S DIG DEEPER

GOAL: To obtain informational knowledge about who the customers are and their purchasing behaviors. To build an effective marketing strategy we need to analyze customers spending habits.

- The sales team needs to know what the busiest days of the week and hours of the day are to schedule ads at times when there are fewer orders.
- They also want to know whether there are particular times of the day when people spend the most money, as this might inform the type of products they advertise at these times.
- Marketing and sales want to use simpler price range groupings to help direct their efforts.
- Are there certain types of products that are more popular than others?

ANALYSIS

Data Sets

World Happiness Report

Download Data Sets

Tools

Microsoft Excel

Process

1. Sourced the data
2. Conducted exploratory visual analysis using relevant Python libraries.
3. Wrangled and cleaned the data. Checked Data for Consistency
4. Conducted a geospatial analysis by creating a choropleth map using relevant Python libraries
5. Ran a linear regression on the data and analyzed the model performance statistics.
6. Performed a cluster analysis
7. Created Data Dashboards on Tableau based on the results of the analysis.

VISUALIZATIONS

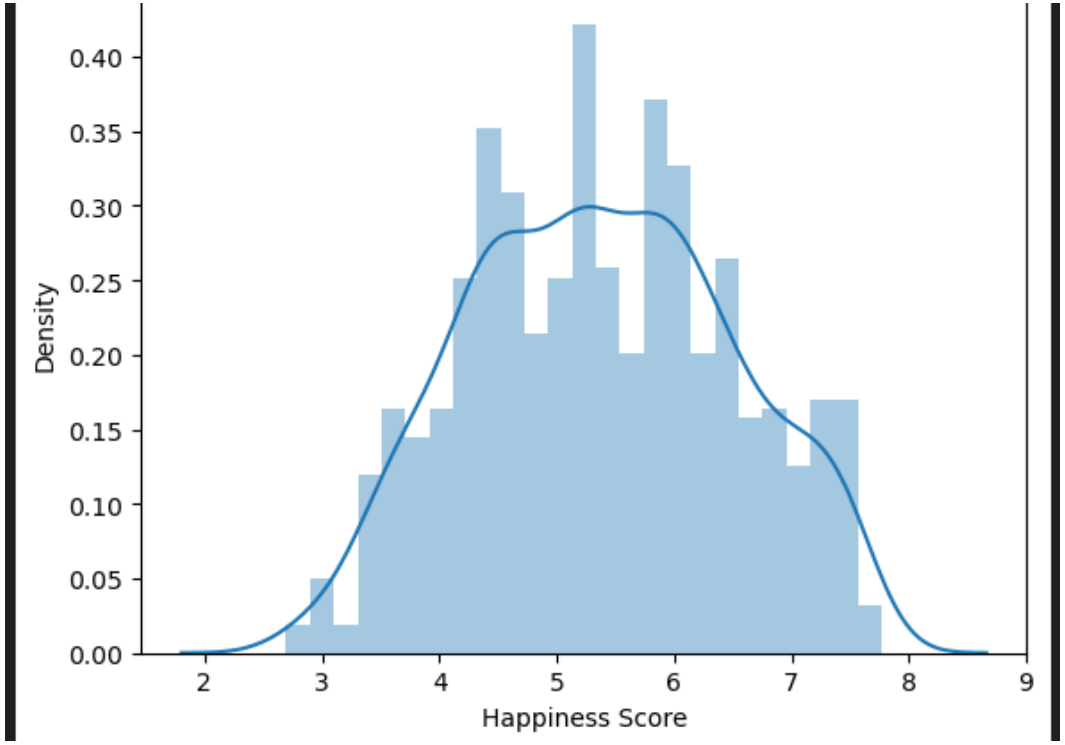
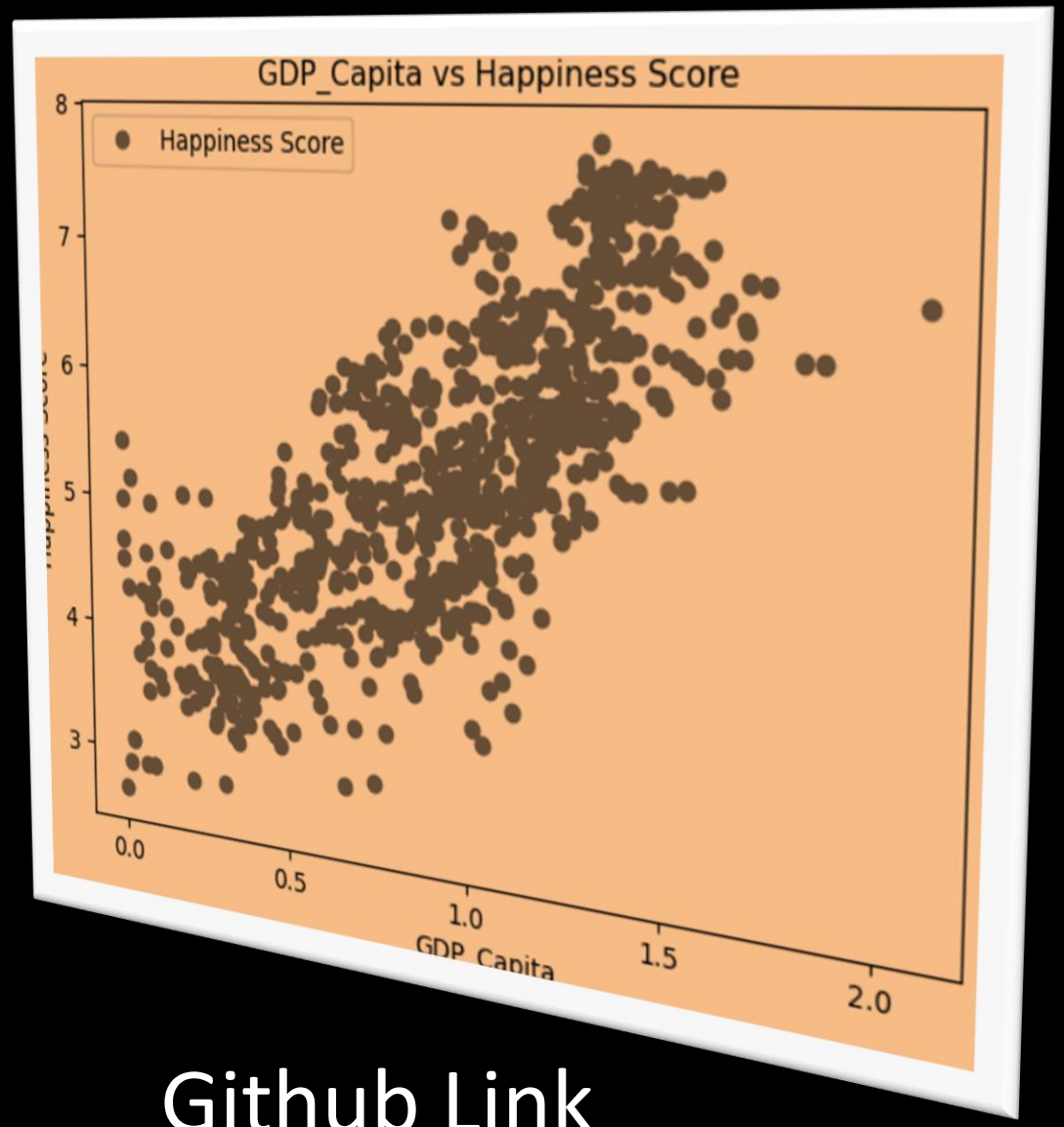


Chart shows the distribution of global happiness scores between 2015 and 2019. The mean happiness score is 5.4.

The world map above visualizes how the happiness scores are distributed globally in terms of high and low.

CONCLUSION

- Based on the analysis of the World Happiness Report we can answer our pre analysis questions:
- The top 5 ranked countries with the highest happiness scores are: Denmark, Finland,
- Iceland. Norway, and Switzerland.
- The top 5 unhappiest countries are: Burundi, Central African Republic, Rwanda, South Sudan and Syria
- The analysis validated the hypothesis that "Countries with a high level of freedom are likely to have a correspondingly elevated happiness score" by revealing a connection between freedom and happiness score, supported by regression analysis.
- In summary, our analysis unveiled strong positive correlations between GDP per capita, Family, and Life Expectancy with happiness scores. Freedom and Generosity also correlated positively, while corruption showed a slightly negative correlation
- Following our results, we determined that the main factor from the report contributing to the happiness level of the population of a country is its **GDP per capita** or economy.



Github Link

Tableau Link

Project links

THANK YOU

CONTACT INFORMATION

(25) Brian Avila | LinkedIn

+1 (508) 905 9144

brianavila819@gmail.com