

Data Analytics Portfolio

Ioana Rusu

ABOUT ME



As a data analyst with a background in support, I specialize in helping businesses transform big data into tailored solutions.

Having a curious mind and a passion for problem-solving, I use connections between complex data sets, to uncover critical insights improving company decision-making and projected growth

Projects



1

AirBnb Berlin

Analysis about AirBnb Rentals in Berlin

2

World Happiness

Insights into the well-being and happiness of people worldwide

3

Instacart

Marketing strategy for an online grocery store

4

Rockbuster

Launch strategy for a new online video service

5

Influenza

Staffing plan for the upcoming flu season

6

Global Bank

Anti-laundering project at a global bank

7

Game Co

Analysis of global video game sales

1. GameCo



PROJECT OVERVIEW

Objective

Game Co is a new video game company that needs data to develop new games. The objective of this project is to perform a descriptive analysis of a video game data set to have a better understanding of Game Co's new games will perform in the market.

Data

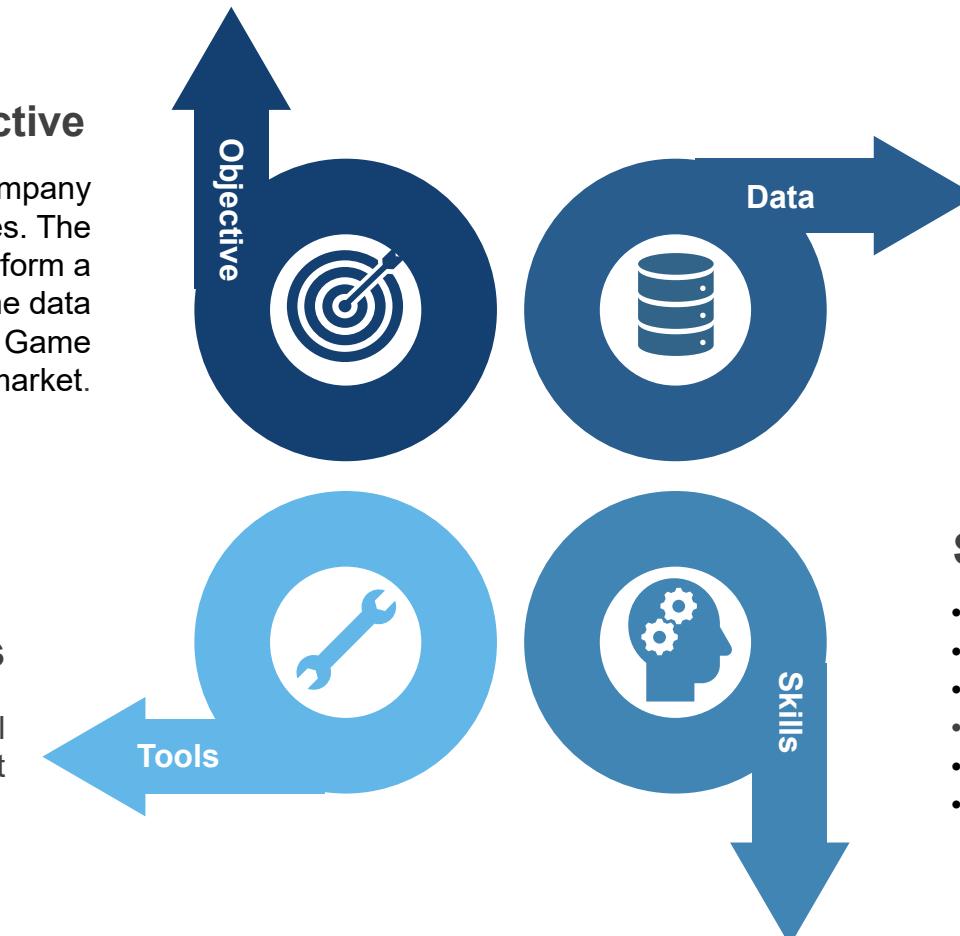
Historical sales of video games (for games that sold more than 100,000 copies) spanning different platforms, genres, and publishing studios.

Tools

- Excel
- Power Point

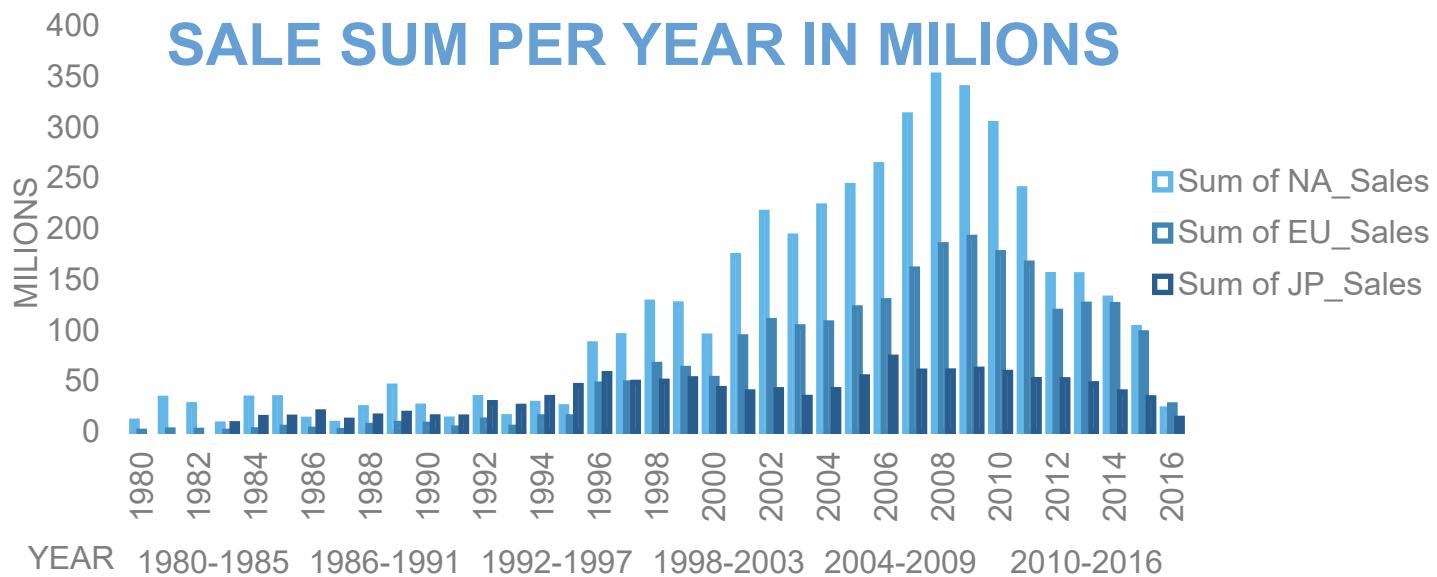
Skills

- Descriptive Analytics
- Data Cleaning
- Data Grouping
- Descriptive Analysis
- Data Visualization
- Storytelling



Hypothesis

Currently, the main understanding is that video game sales behave almost the same across all geographic regions.



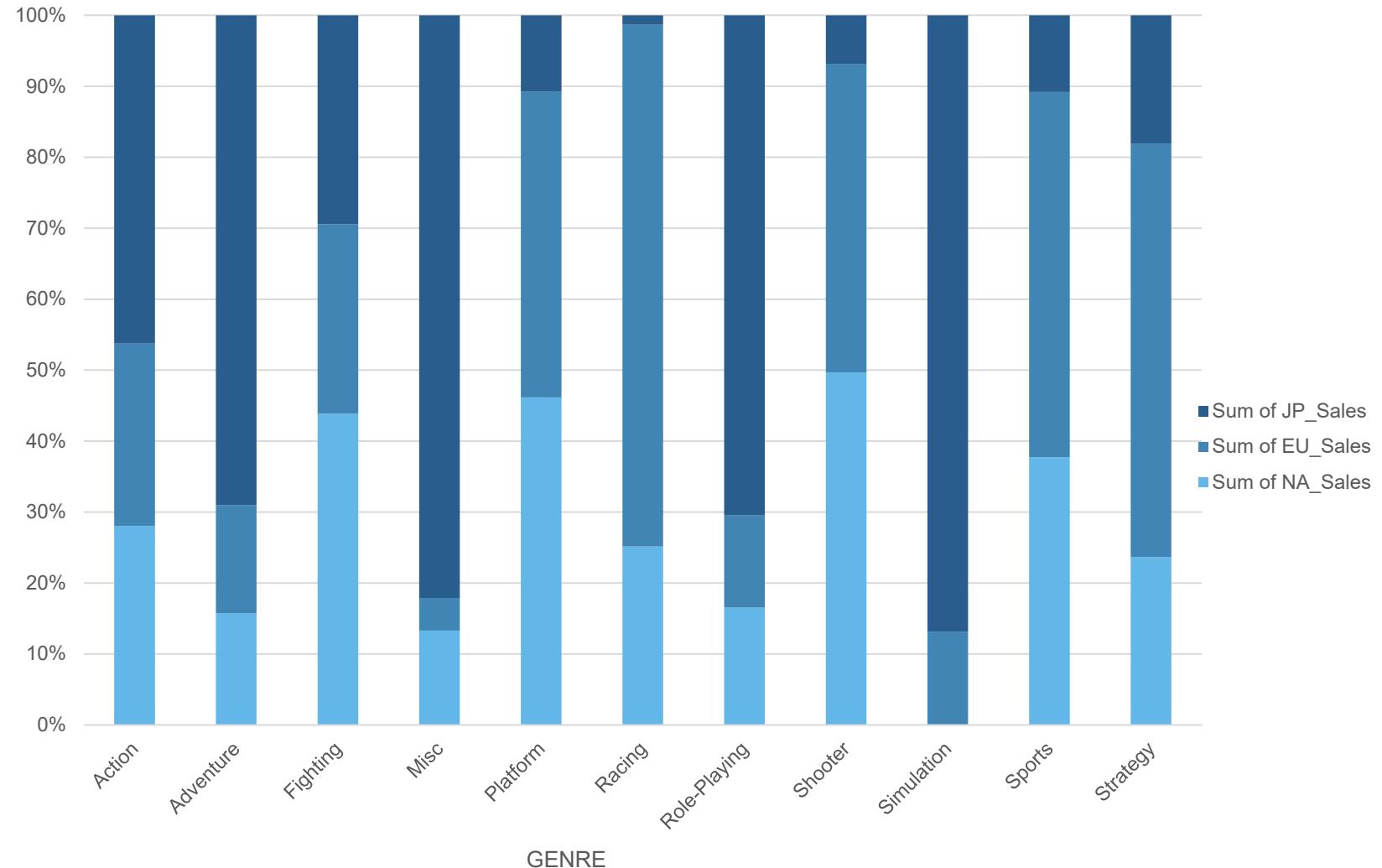
Game sales across each geographic region have not remained the same over time.

North America and Japan are not steady markets, while Europe kept being consistent, uprising steady.

Still, neither markets can compare as sales with NA, being the leader overall of total sold product video games.

All make the hypothesis from the beginning wrong.

2016 Sales by Genre

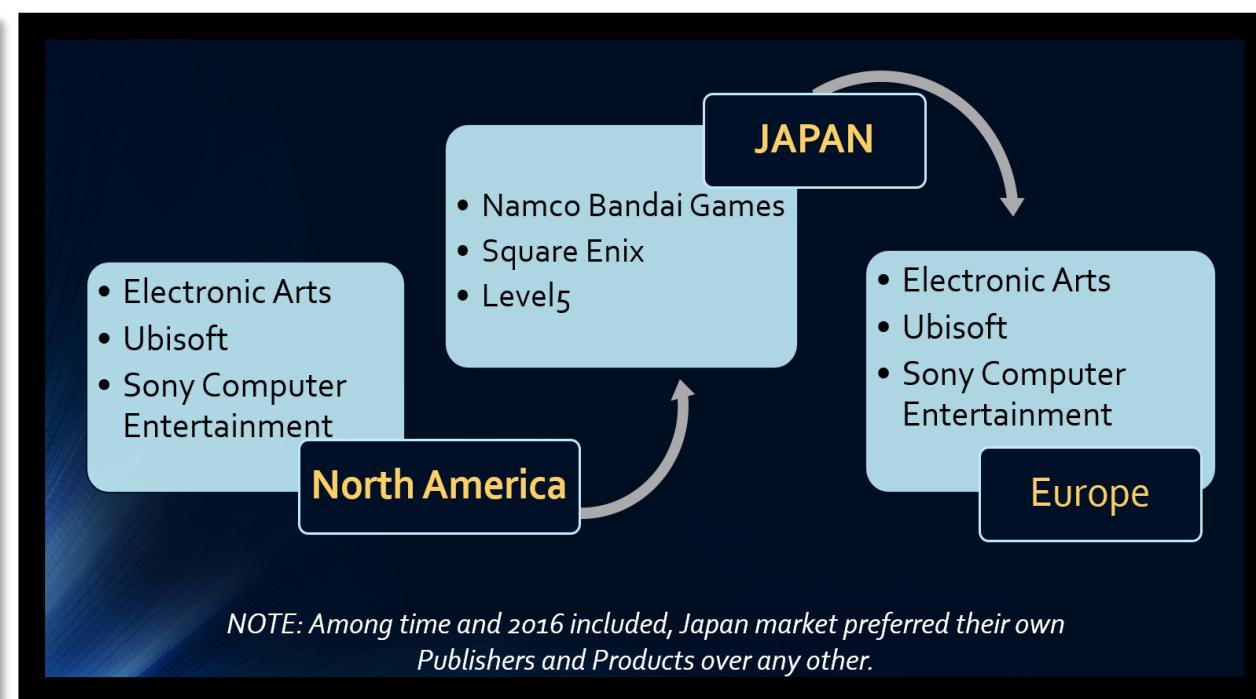
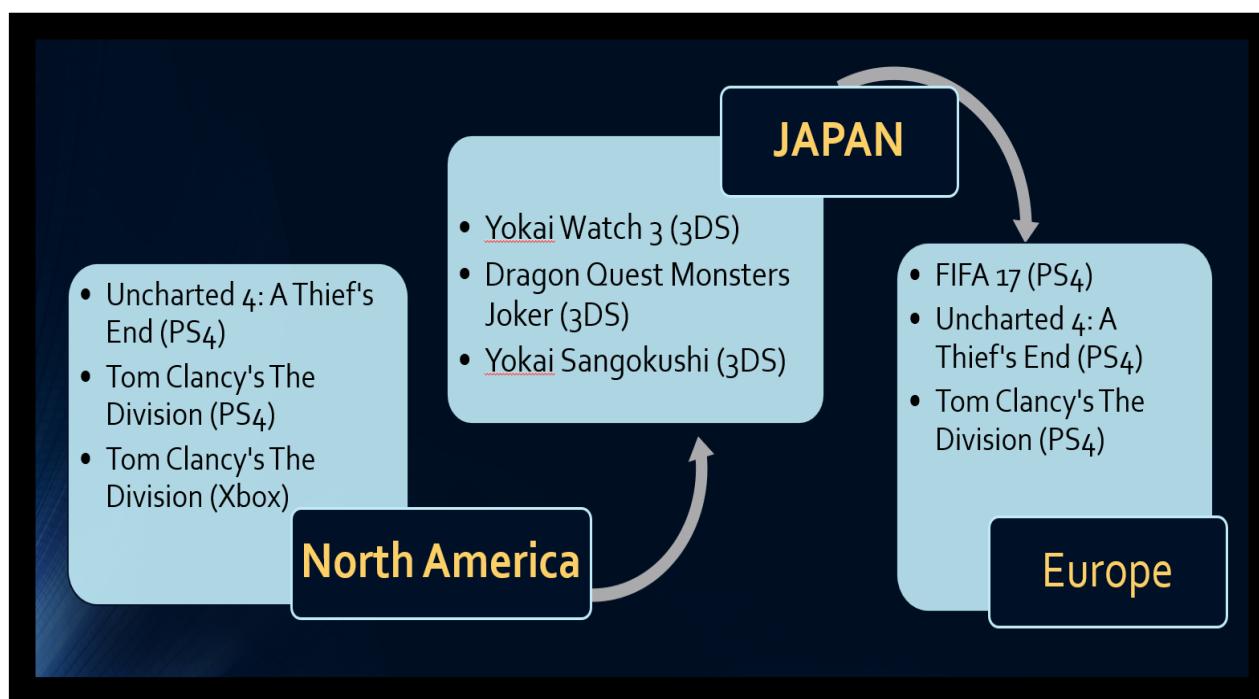


In 2016 Europe lead the market, increasing steadily but constantly. North America was passed by 6%.

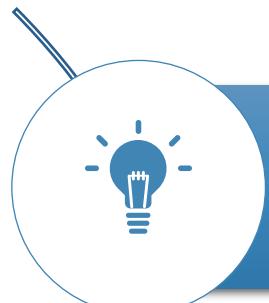
For Japan, the leading genres are action and Role-Playing, while in North America and Europe Shooter, Sports or Action has the most success.

TOP 3 GAMES SOLD in 2016

TOP 3 PUBLISHERS in 2016



RECOMMENDATIONS



Since Japan is a steady market in terms of preferences, we should focus mostly on their products and preferences: own games, most liked genres and 3DS platform.



Europe has a big potential and it is in constant growth. We should invest more into this market and keep in mind the primary liked genres: shooter, sport, action.



North America most liked genres are: shooter, action, sport. We can promote these.

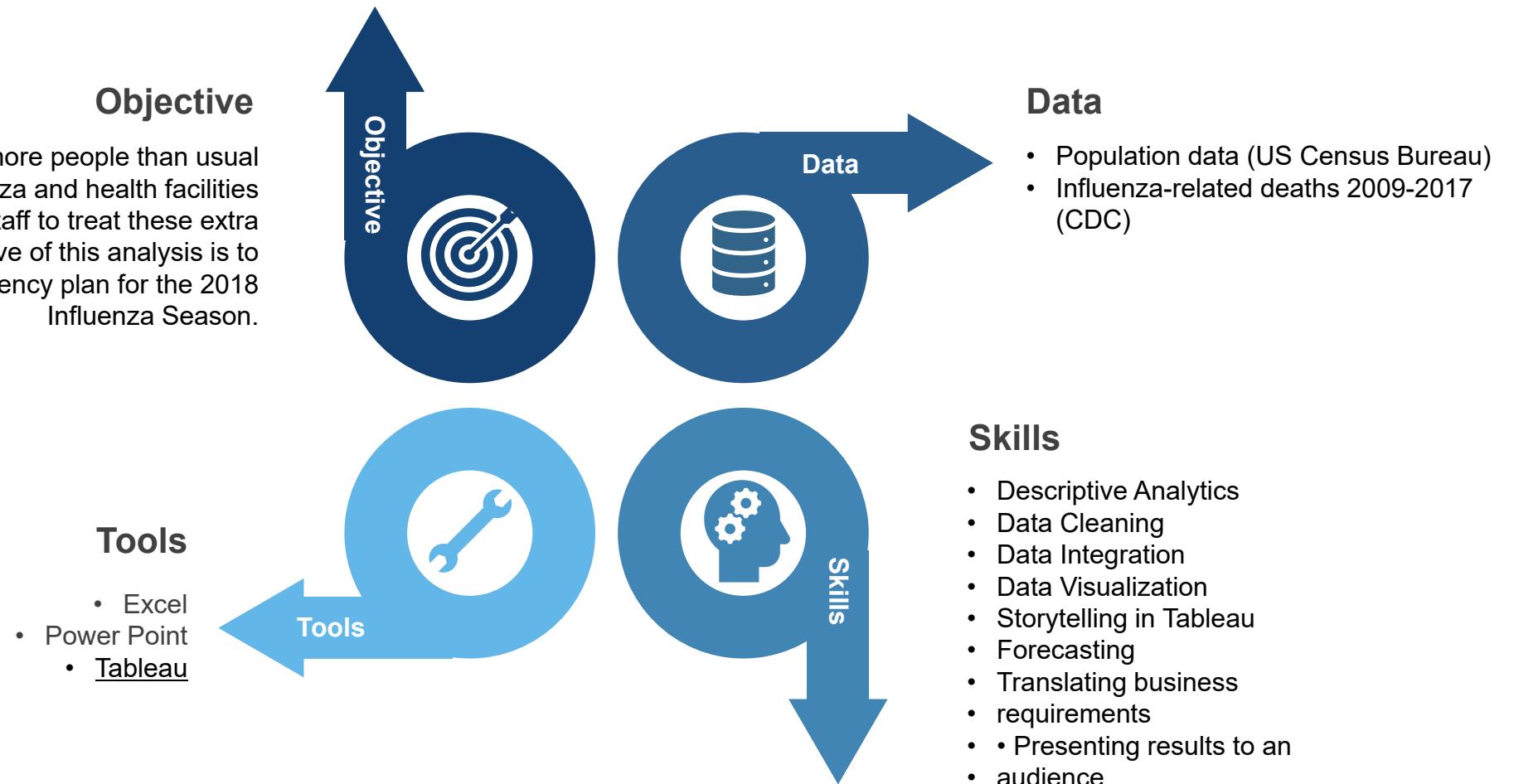
2. Influenza



PROJECT OVERVIEW

Objective

During flu season more people than usual suffer from influenza and health facilities need additional staff to treat these extra patients. The objective of this analysis is to help a staffing agency plan for the 2018 Influenza Season.

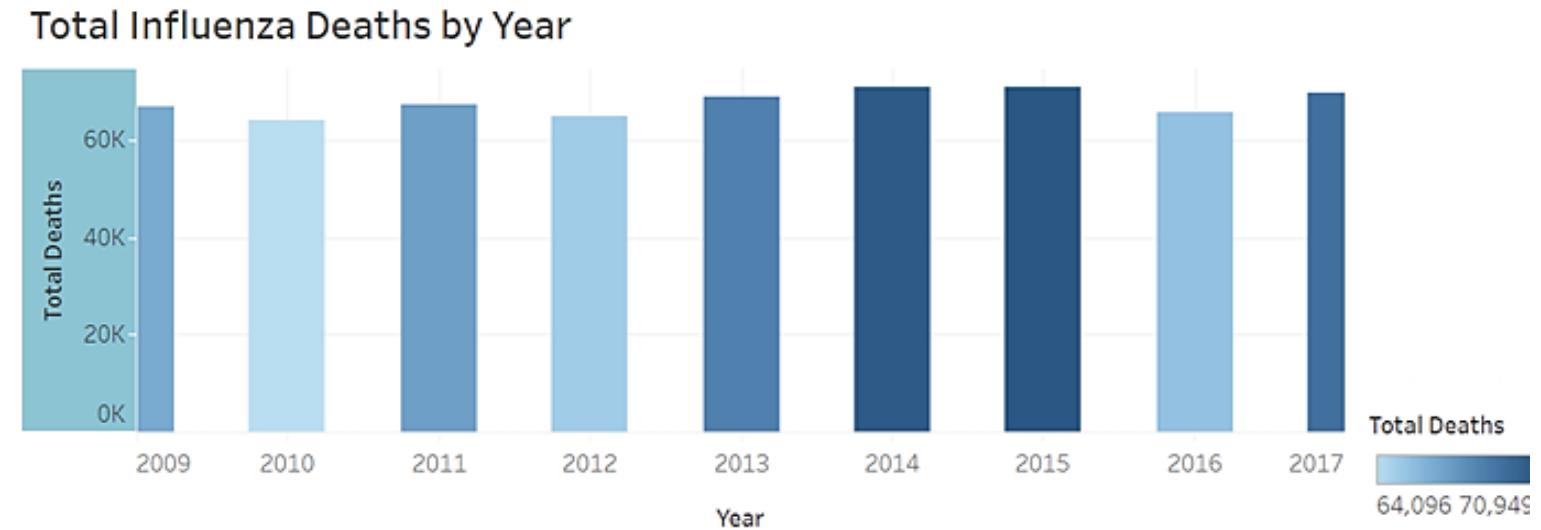


Hypothesis

Vulnerable populations suffer the most-severe impacts from the flu and are the most likely to end up in the hospital. Every year the death rates are the same.



Mortality Rate Count for Influenza Season (2009-2017)



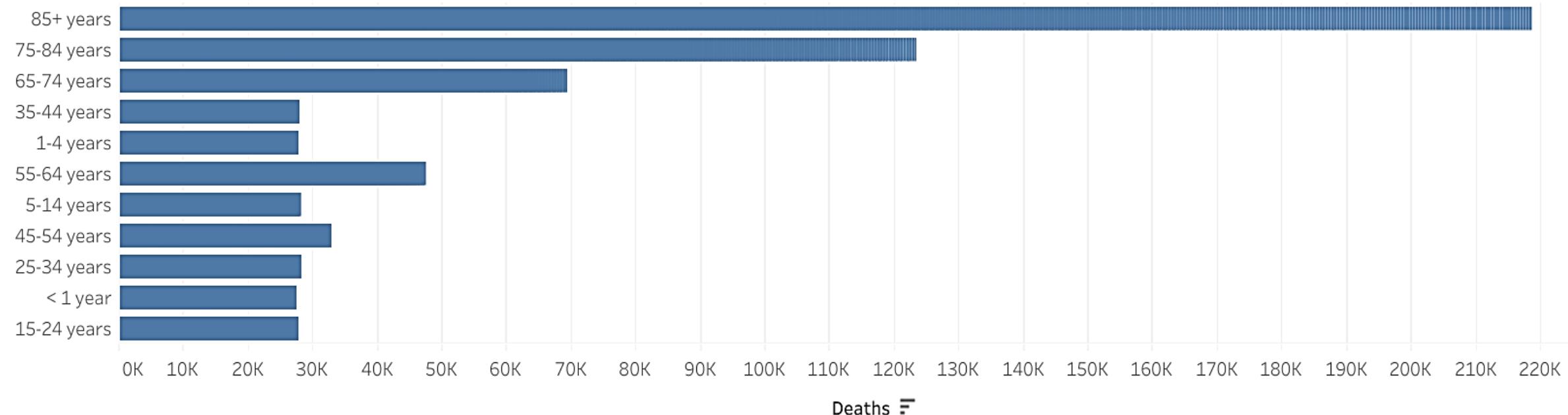
The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics experience increased traffic and patients during influenza season and end up needing additional staff to adequately treat extra patients.

What category of population is most at risk?

Vulnerable population represents those patients likely to develop flu complications requiring additional care.

These are the patients that should receive immediate care, among those with other known conditions ,low immunity system, children under 5 years, and pregnant women

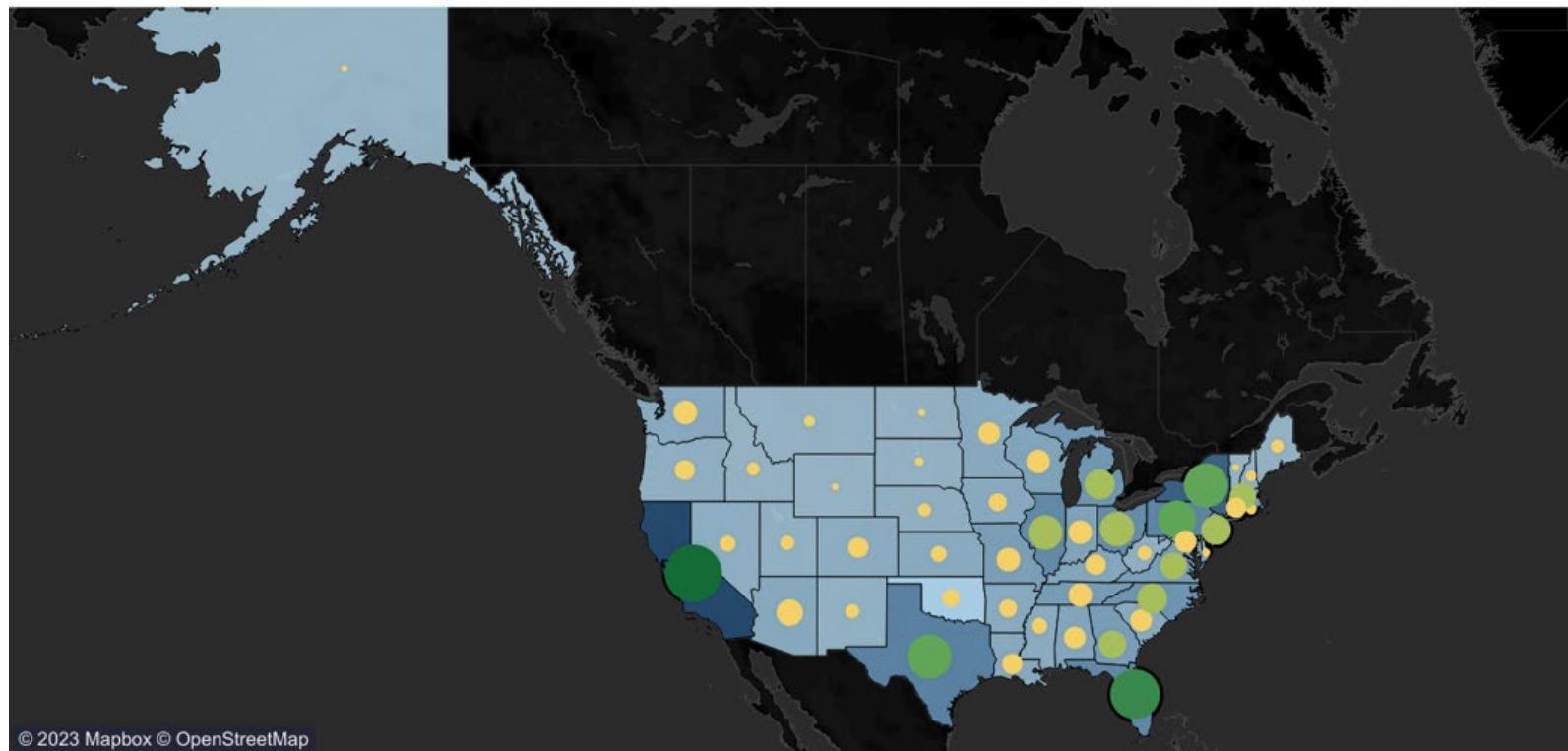
Influenza Deaths by age groups



What states in U.S are most affected by the virus?

Vulnerable States in U.S

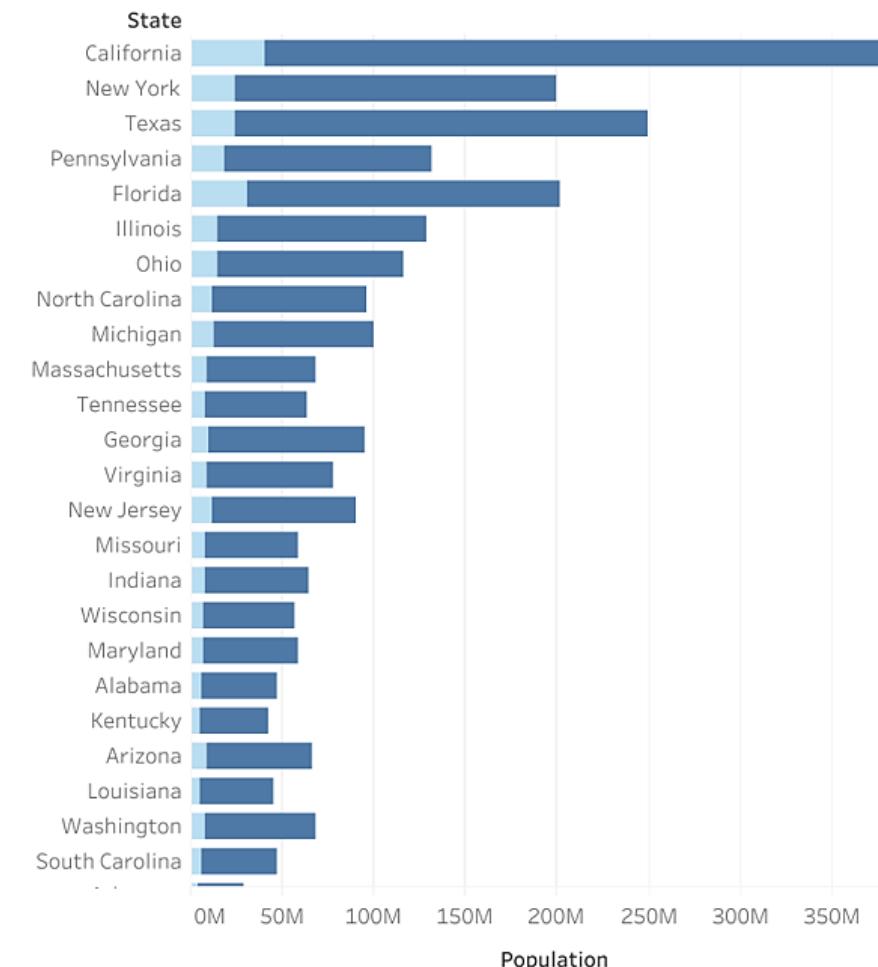
US Population&Deaths for over 65years during Influenza Season



High-Risk States are considered the ones where the Vulnerable Population has the highest number, compared with the Total Population.

These are California, New York, Texas, Pennsylvania, Florida, Illinois and Ohio.

Comparison between Total Population and +65 years

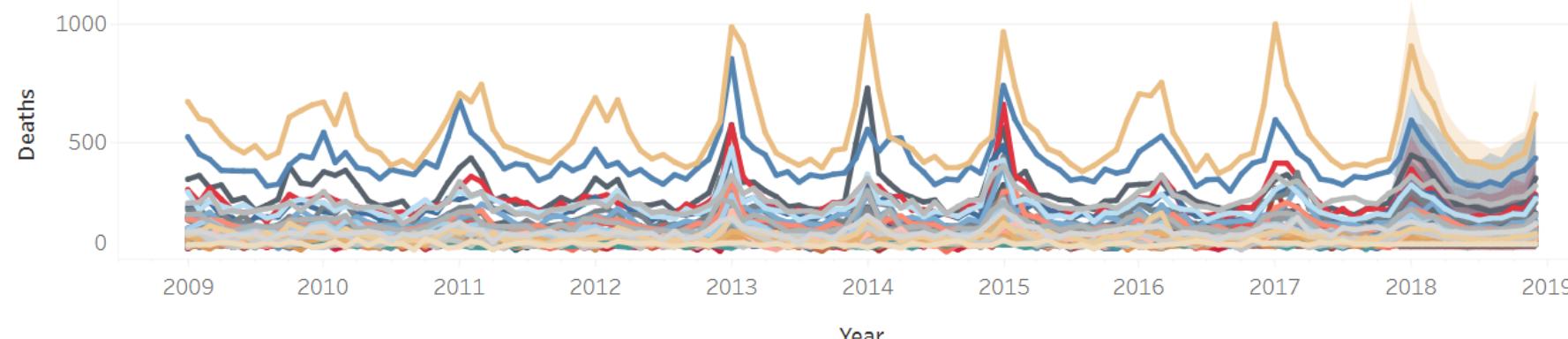


When is Influenza Season?

Vulnerable months

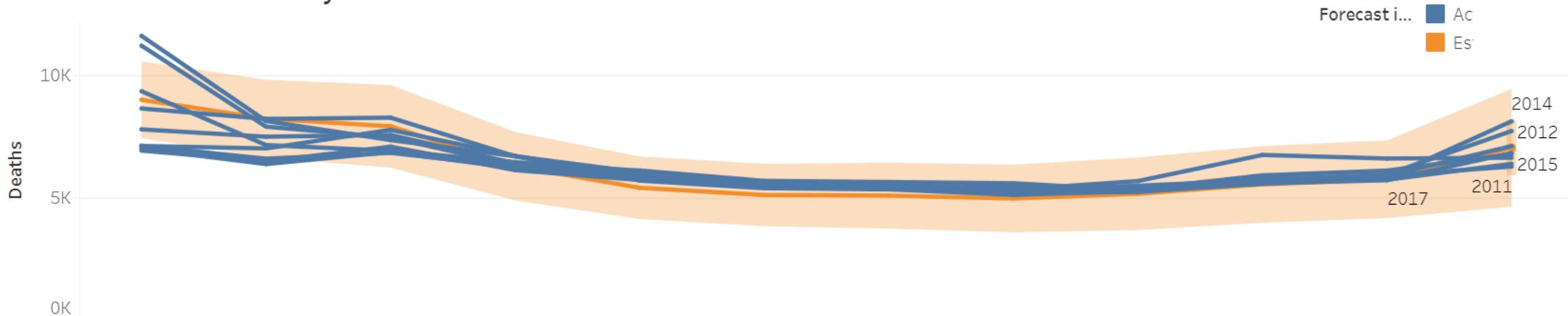
State
Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Col...
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi

Influenza Deaths by State and Year Forecast



In the United States, flu season usually occurs in the fall and winter. While influenza viruses spread year-round, most of the time flu activity peaks between November-March, with high activity during in **December and January**.

Total Influenza Deaths by Months and Years



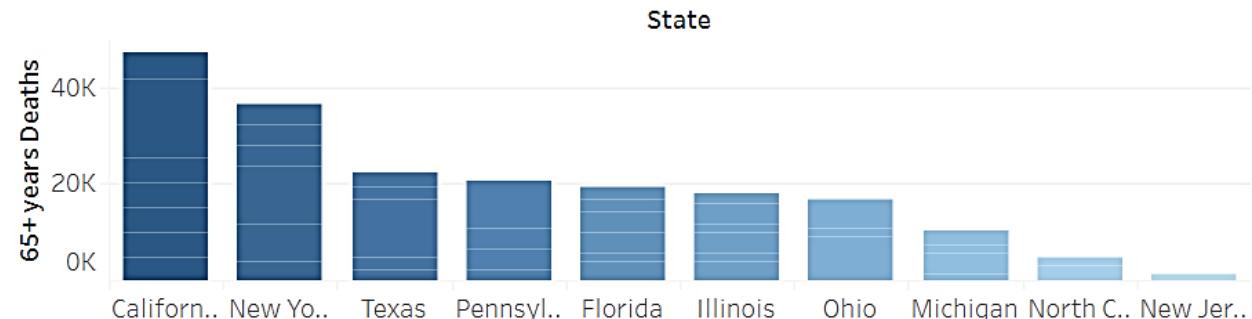
CONCLUSIONS AND RECOMMENDATIONS

Influenza is a virus that affects population throughout all United States, mostly in Southeast, Northeast, Midwest regions. The vulnerable people are those in +65 years of age category. The high season lasts from December to February.

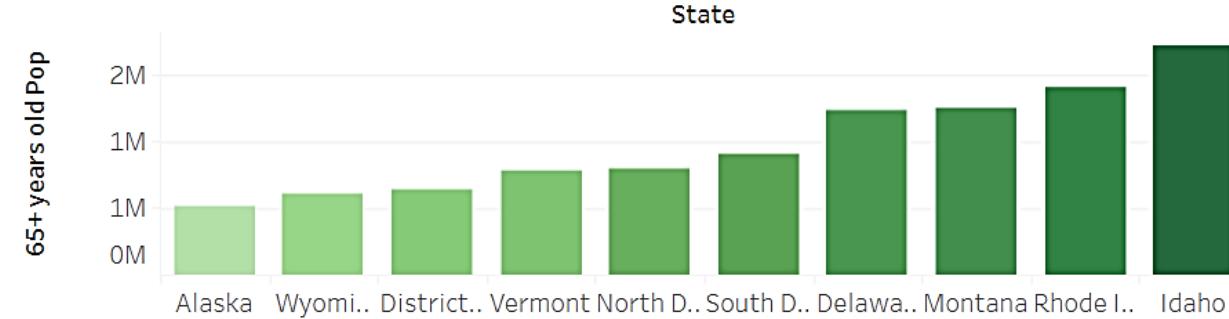
Recommendations:

- Influenza vaccination is the most effective strategy to prevent influenza infections
- Campaigns should be made to inform the population about the risks
- Extra staff should be allocated to hospitals in the most vulnerable top areas mostly during December-February.
- Wearing masks as long as the virus is still said to be active.

Top 10 Vulnerable US States

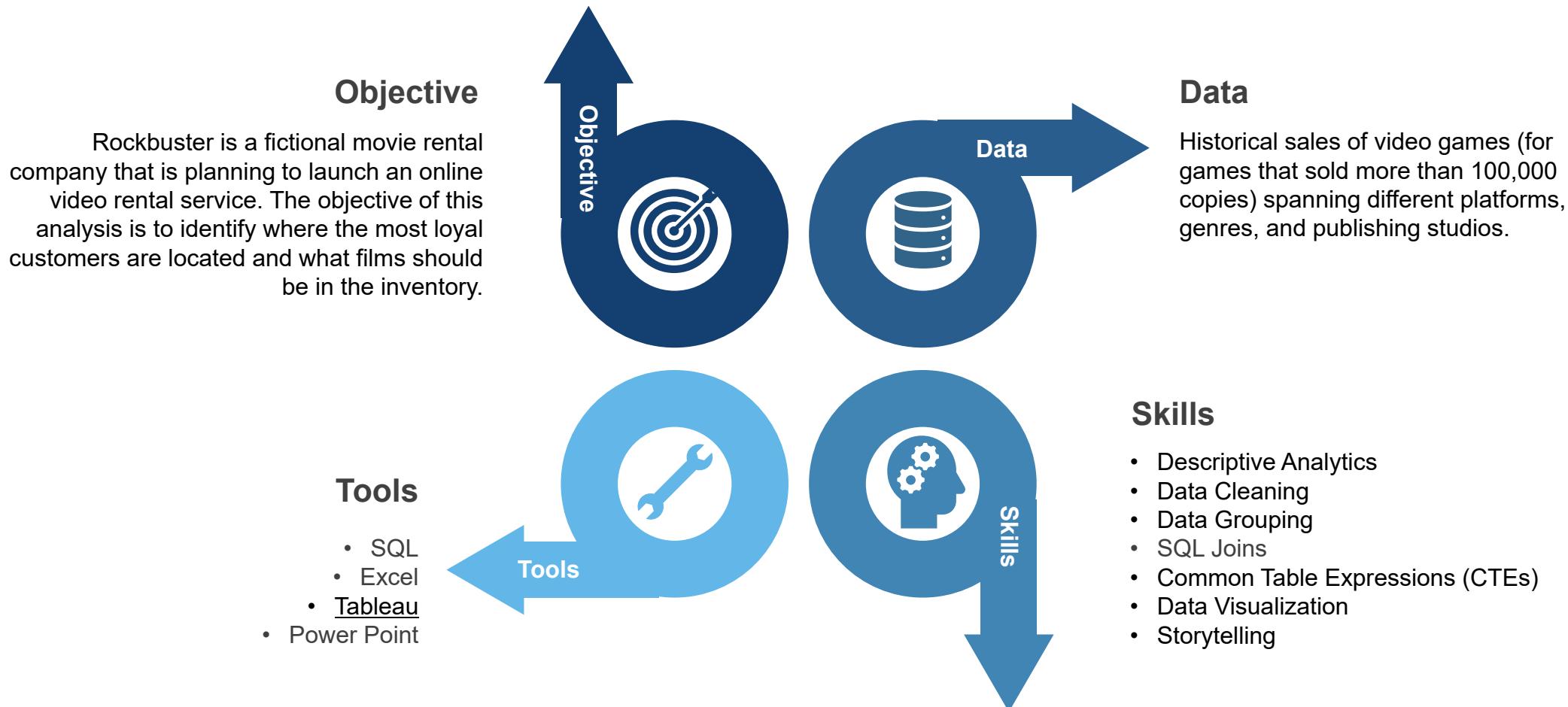


Bottom 10 Vulnerable US States



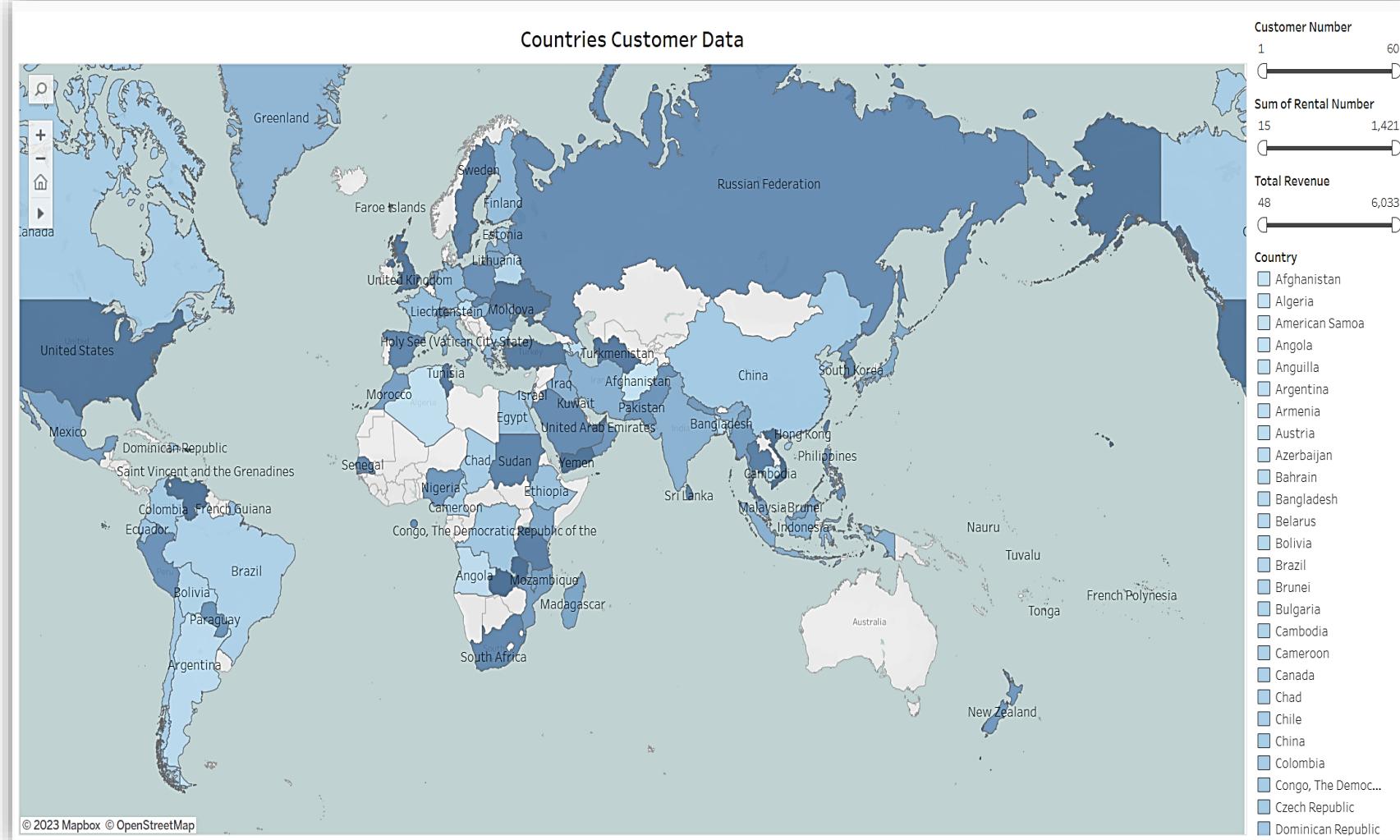
3. Rockbuster

PROJECT OVERVIEW

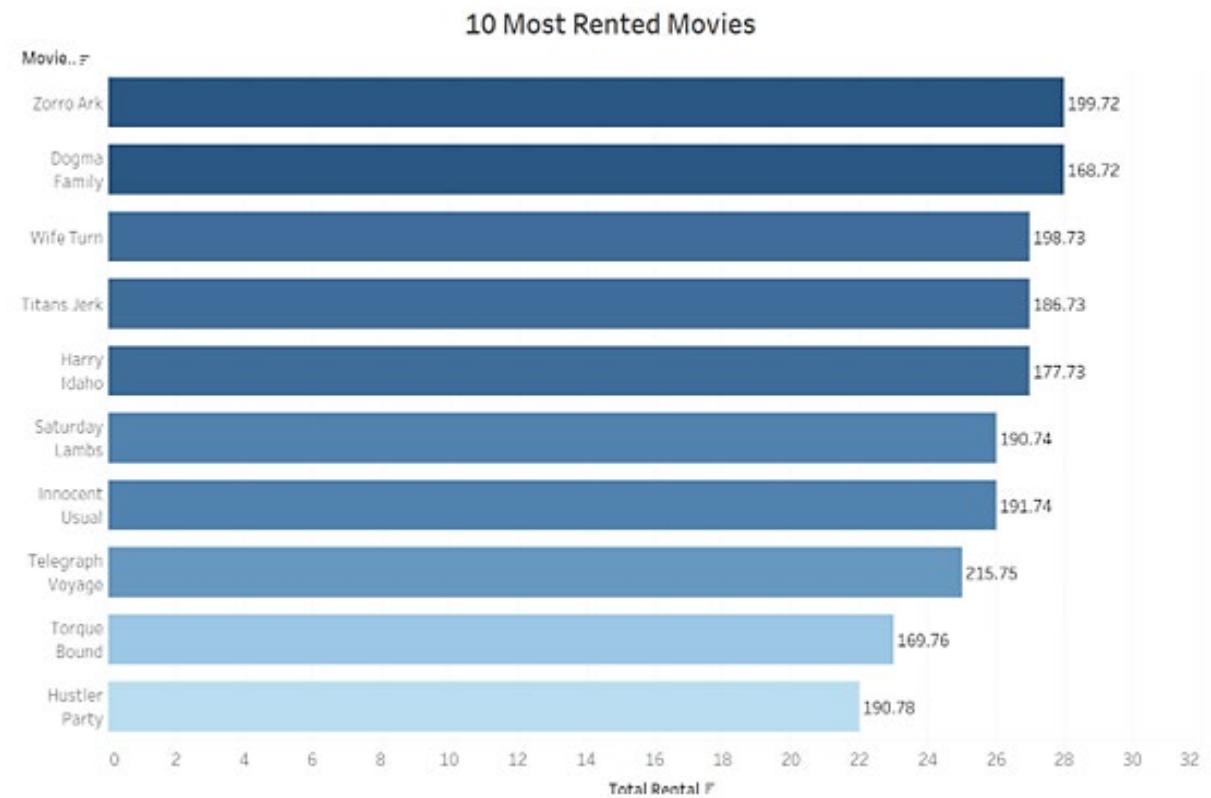
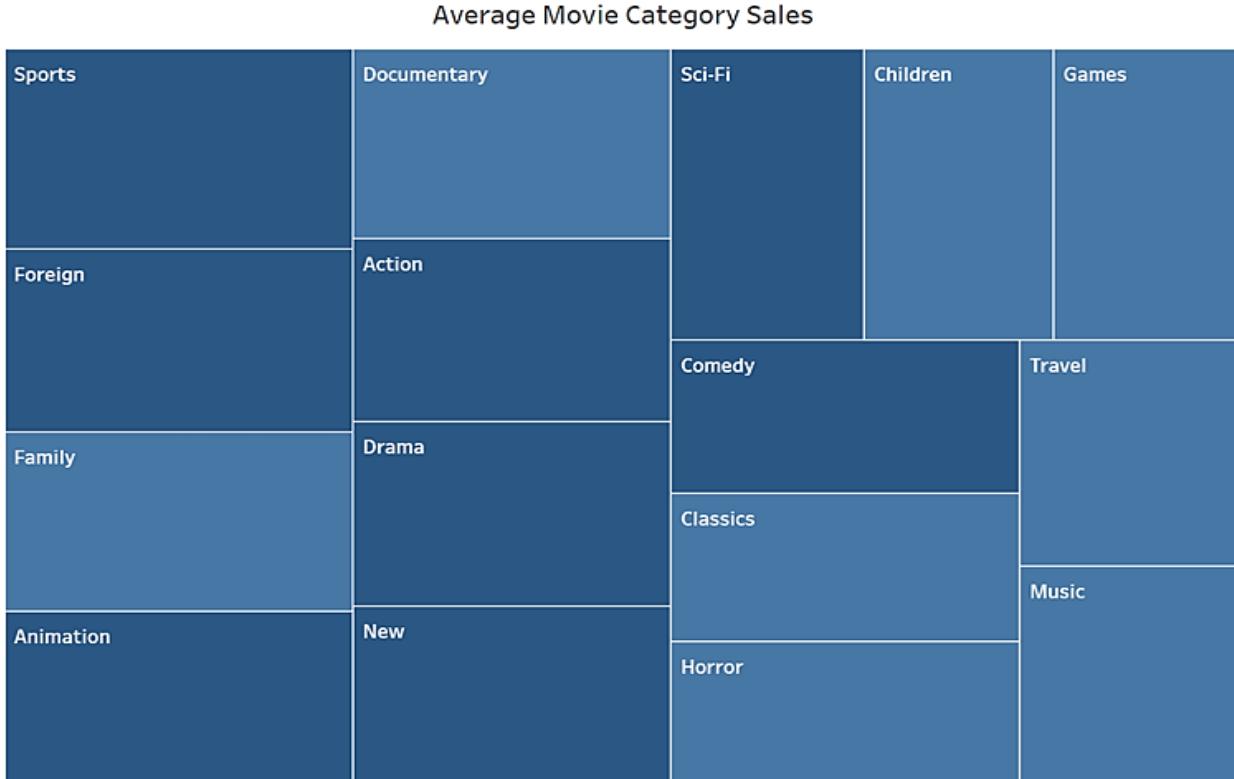


Where are the customers located?

- Rockbuster has 599 customers based into 109 countries around the world.
- Top 10 Countries by customers and Revenue are: India, China, US, Japan, Mexico, Russia, Brazil, Philippines, Turkey and Indonesia. These countries provide more than 50% of the total revenue.
- India, China and United States are the leaders, making themselves 24% of the global Revenue.

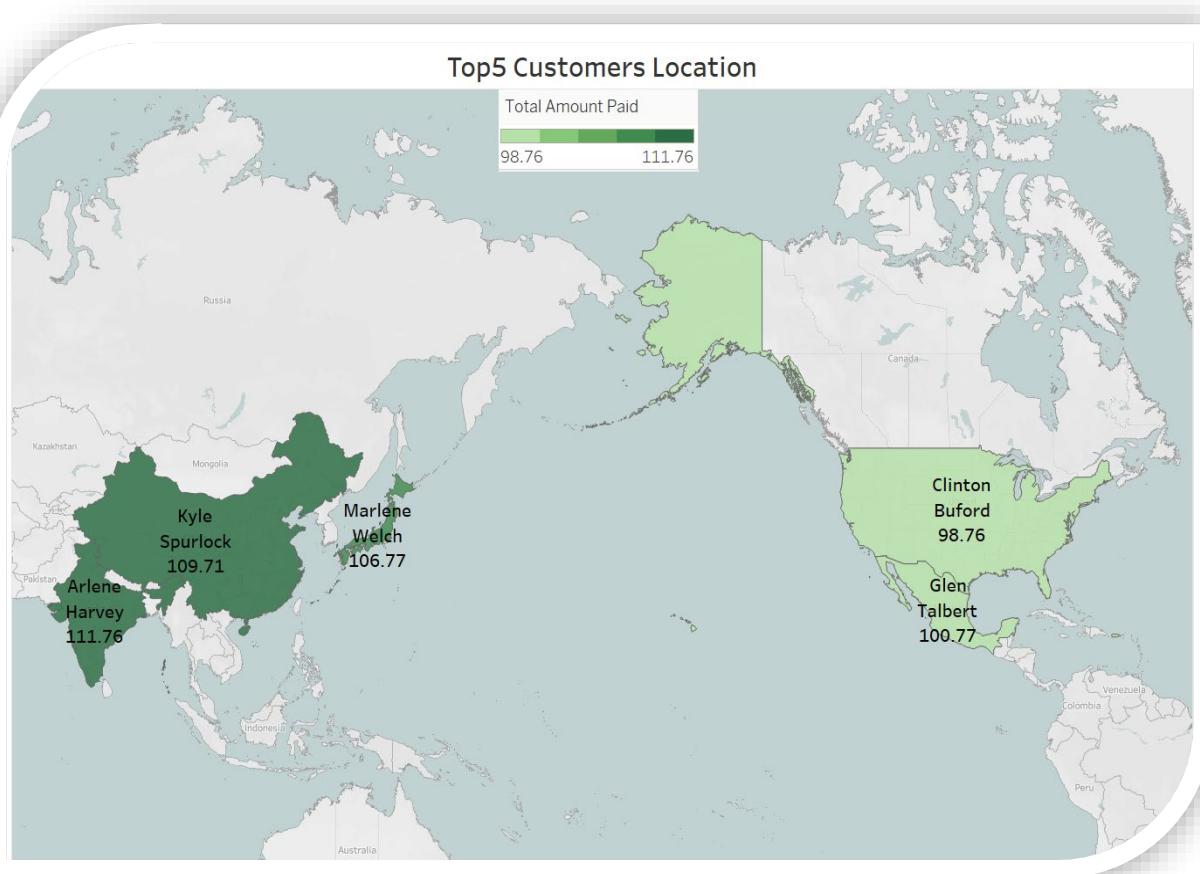


Favorite movies



- Sports is the most popular category of movie, representing 8% of the rentals with the highest revenue of \$4892. The last categories, below \$3401 each, we find Horror, Classics, Children, Travel and Music, representing 27% of the revenue.
- As for number of movies, all categories are close to each other.
- Zorro Ark is the top rented movie for Rockbuster, while Texas Run has the least rental, only 6.

RECOMMENDATIONS & TOP 5 CUSTOMERS



RECOMMENDATIONS:

- More languages should be available, not just English.
- Consider increasing the price for highly rented genre and decreasing it for those with low rentals
- Increase number of movies for PG-13 and N-17 since these are best rented and bought genre.
- Increase market and national content in the best performing markets, such as India and China.
- Investing less in unpopular categories.
- Reward system for loyal customers and collect and measure customer satisfaction along the process.
- Flexible Subscription Plans for different needs of customers.
- Develop a user-friendly website and mobile app that makes browsing, searching, and renting films a seamless experience.

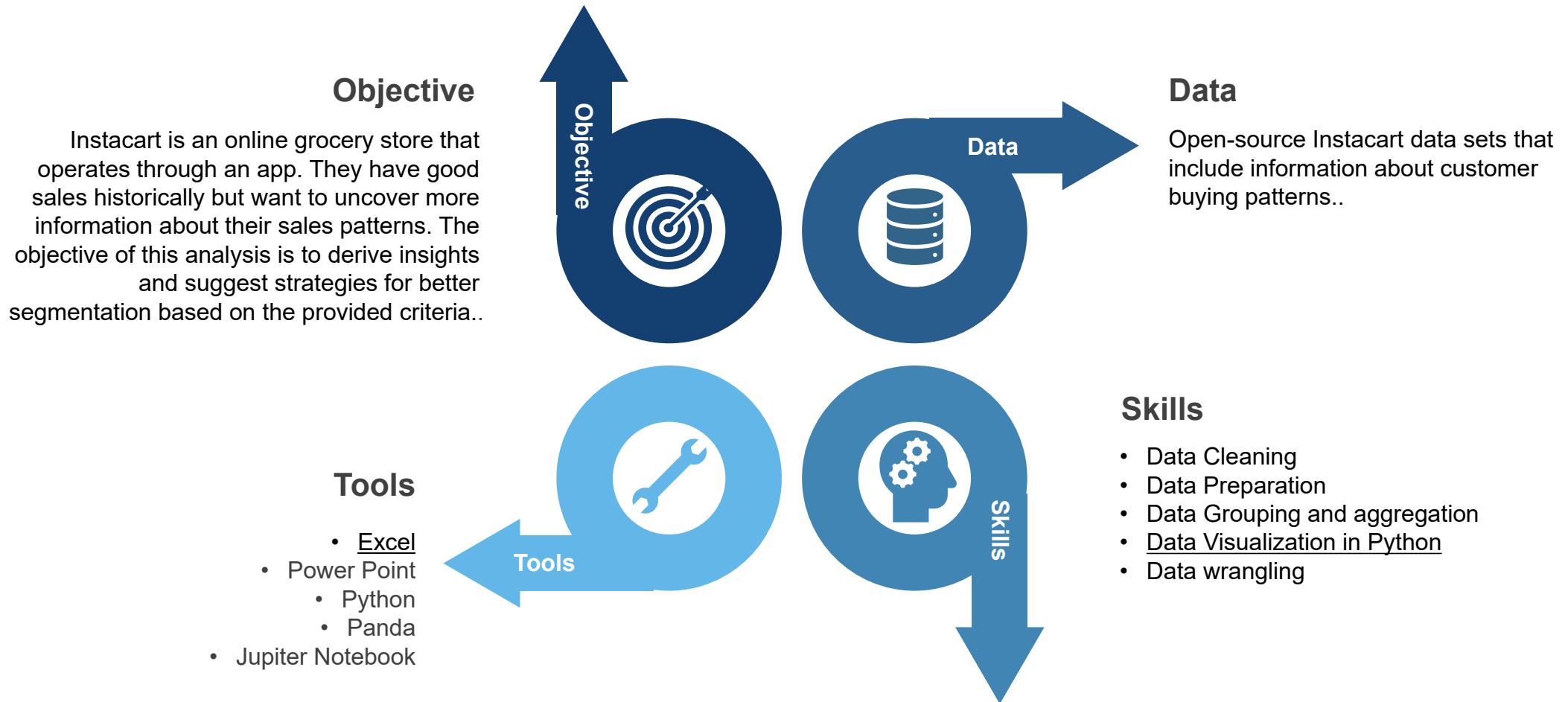
Customers that should be rewarded for their loyalty:

- Karl Seal - Cape Coral, United States
- Marlon Snyder - Santa Barbara D'Oeste, Brazil
- Brandon Huey – Balikesir, Turkey
- Angela Hernandez – Shimonoseki, Japan
- Carrie Porter – Liacheng, China

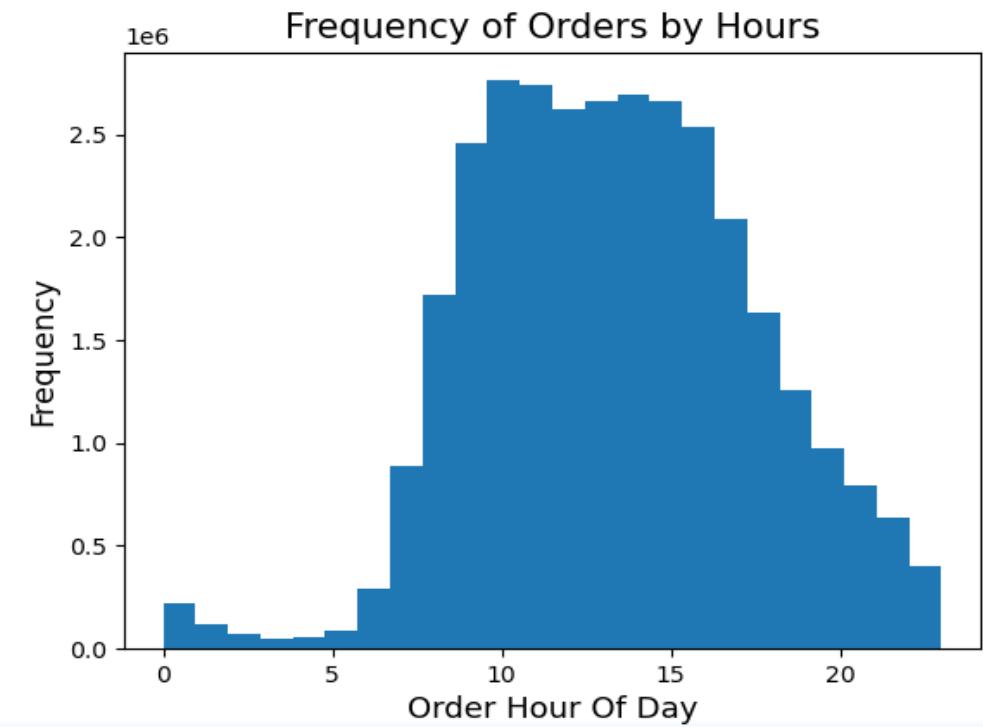
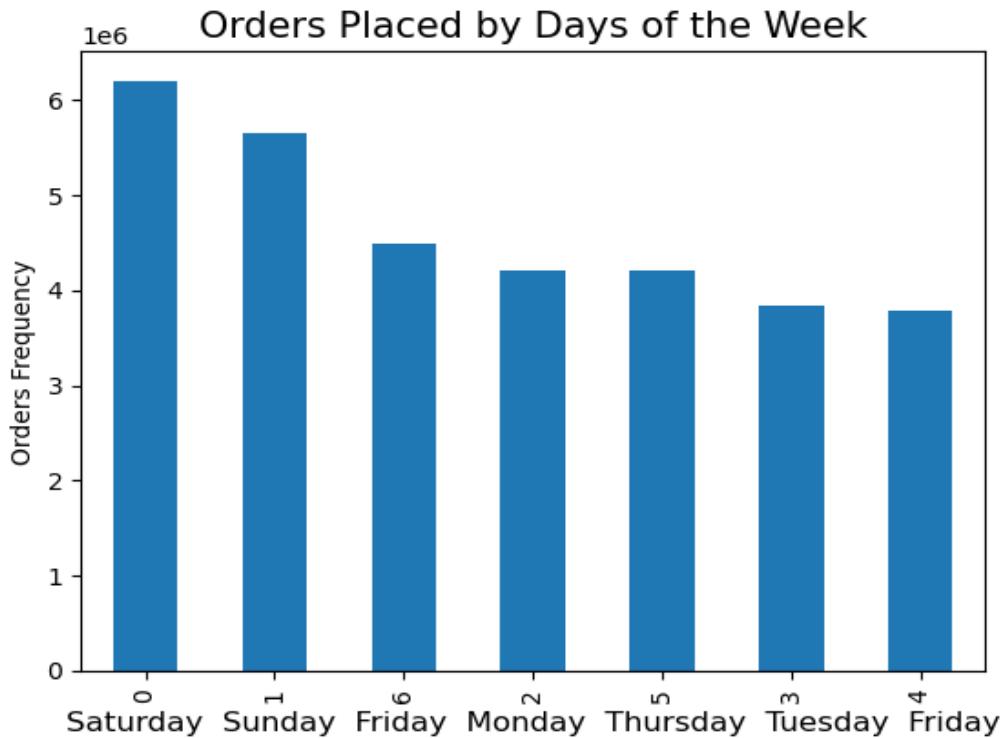
4. Instacart



PROJECT OVERVIEW

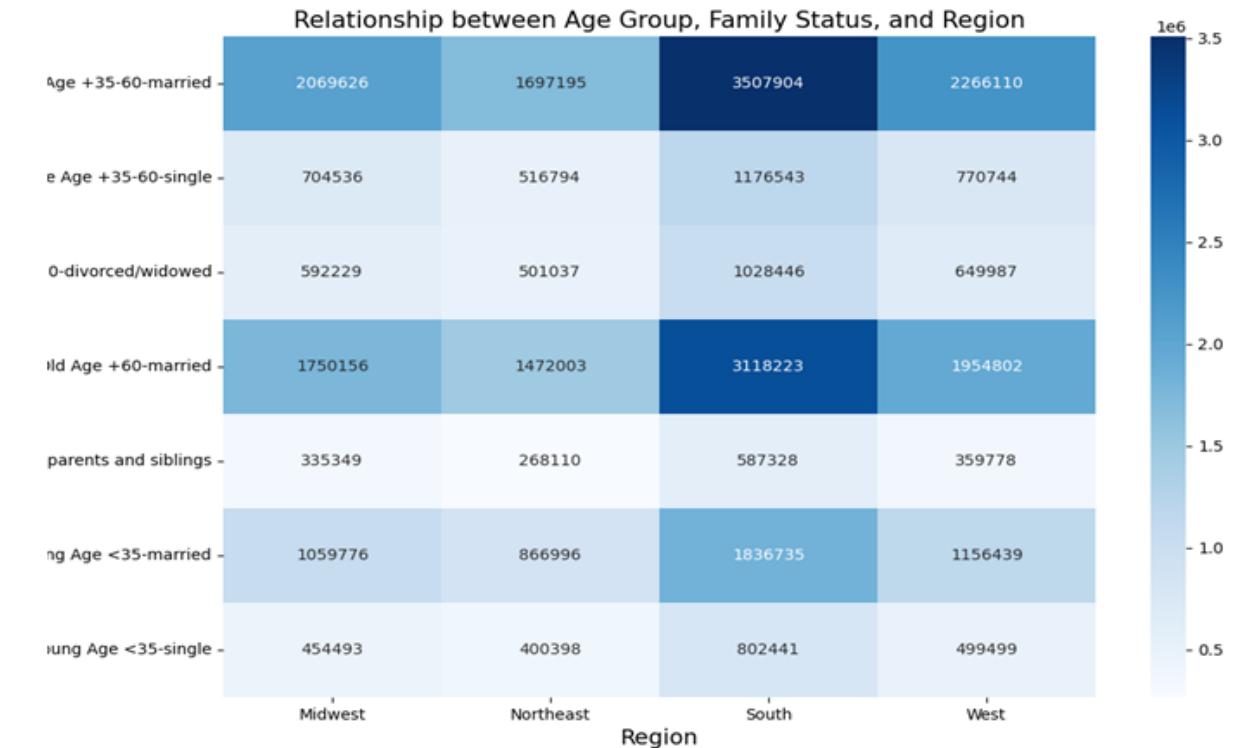
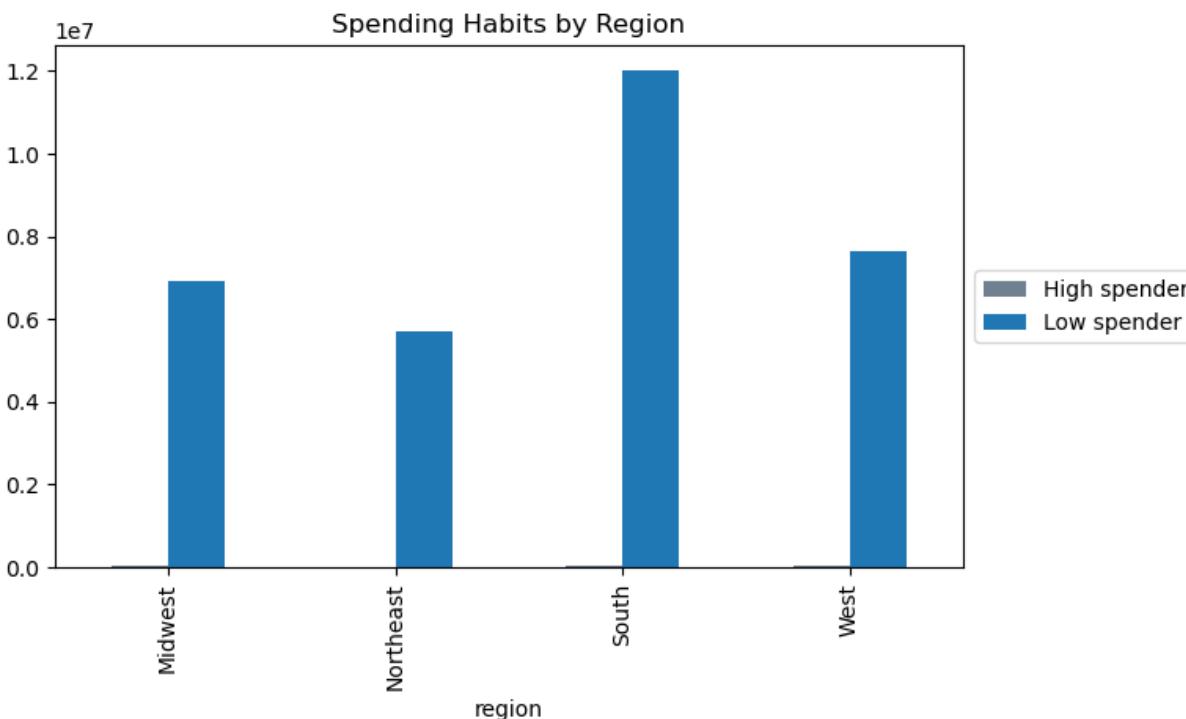


Orders Placed by hour and days



Most orders are being made during weekend: Saturday and Sunday, days when also the prices are increased. Tuesday and Wednesday orders are being made less. Orders are being made mostly in the morning, between 8 AM - 17 PM, with peaks 9-11 AM.

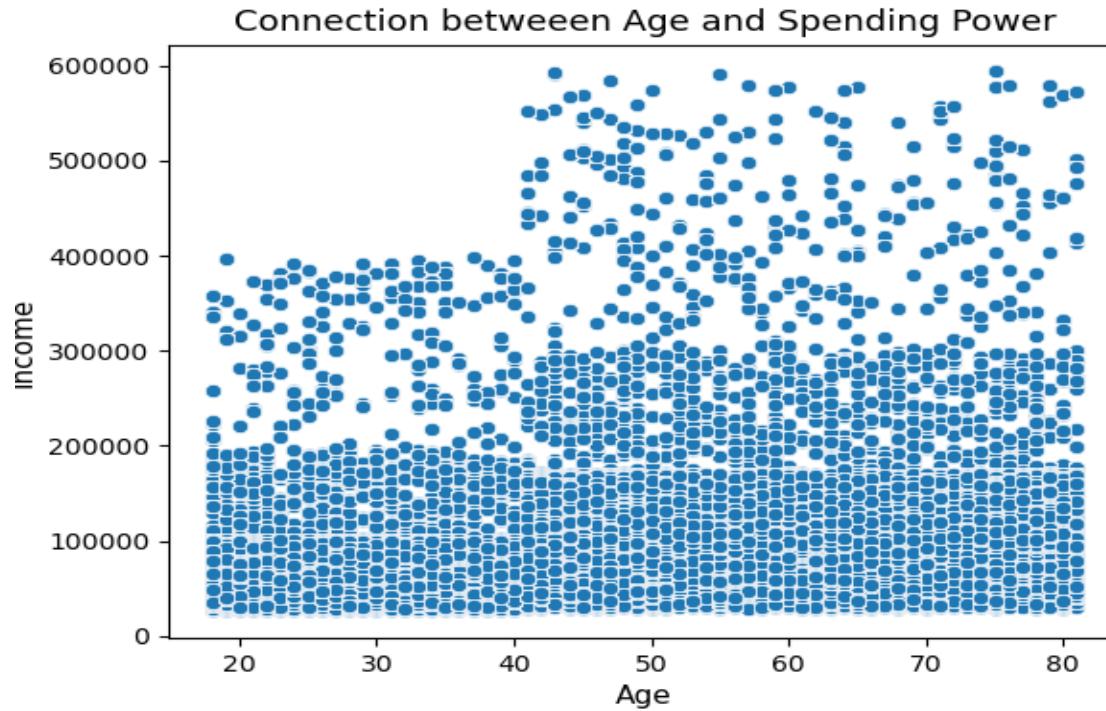
Ordering Habits by Region



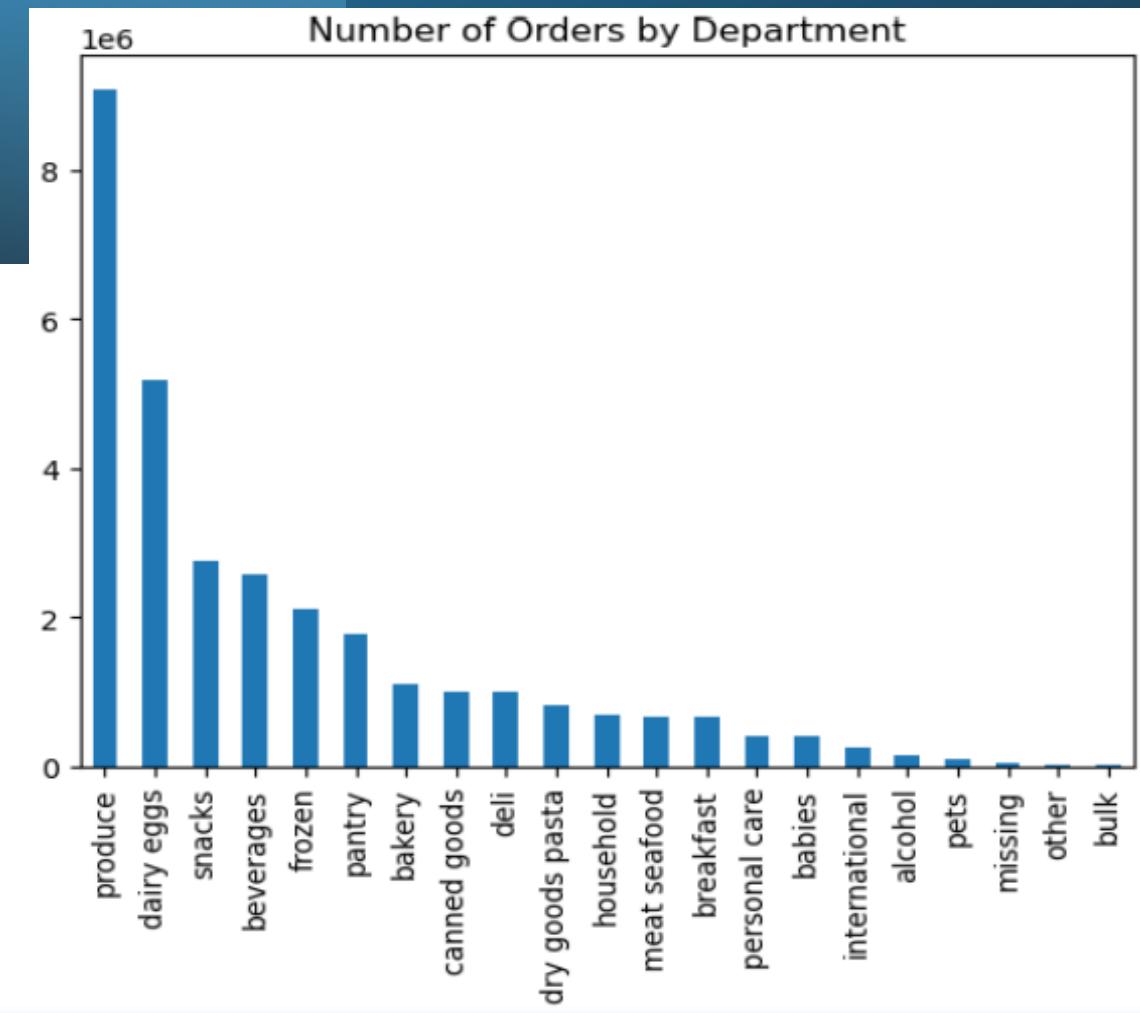
South Region, that also has the highest income, married people and medium income ones, have the most orders.

All married people, no matter the age, represent most of the customers.

Customer Spending Data

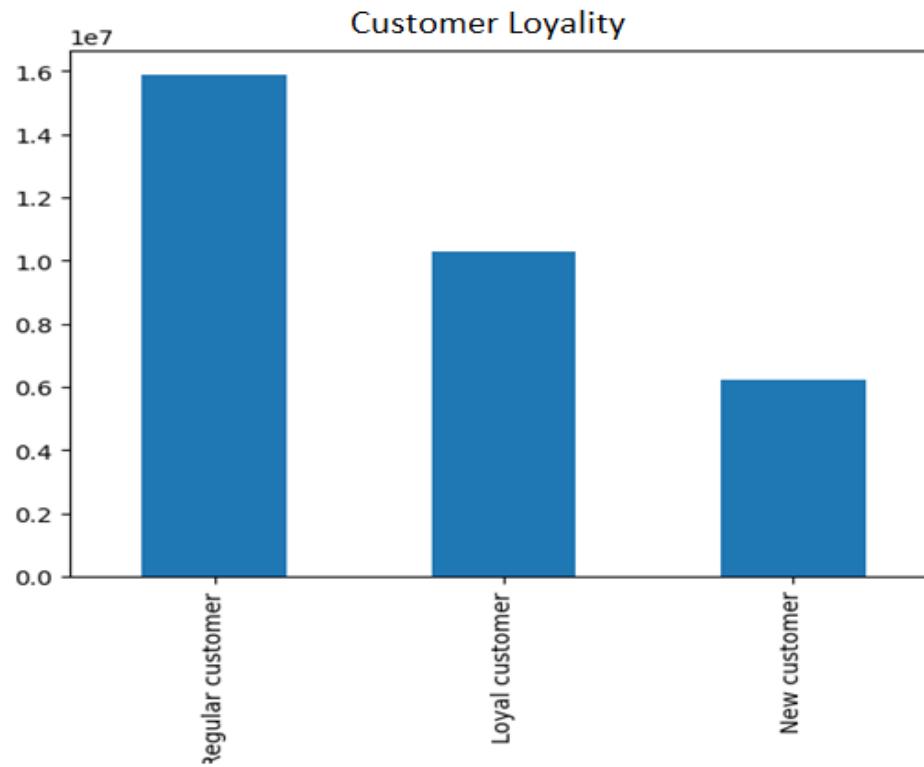


+40 year old have the highest income, thus, the biggest spending power, making the most orders.

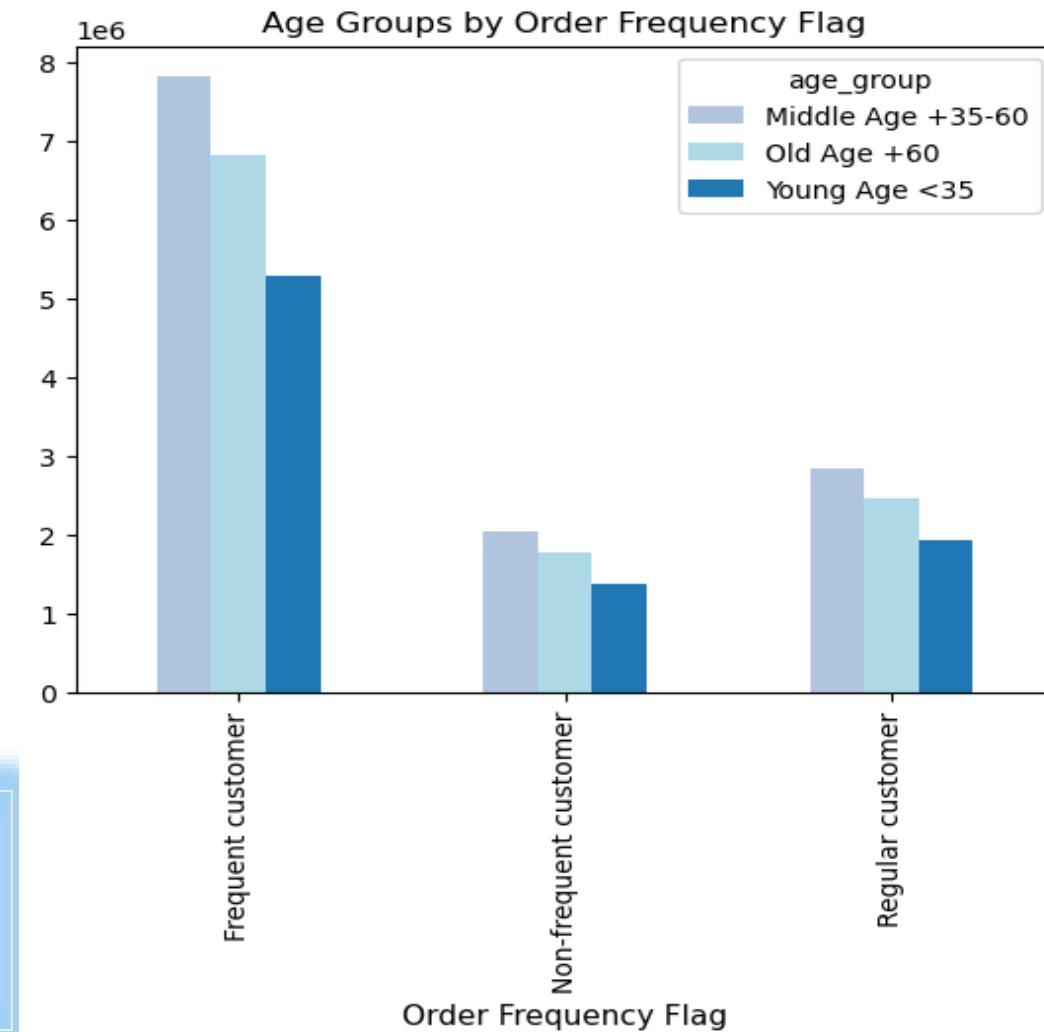


Produce, Dairy/eggs and snacks account for half of the orders. Household and pet products are the ones bought with less frequency while baby products and alcohol are purchased more frequently.

Customer Loyalty



Most of the customers are regular customers, with orders between 10 and 40. Regular customers are making most orders during weekend, same as loyal ones. Married, Middle Age group (35-60 years) represent most of the Frequent Customers



RECOMMENDATIONS



There should be a produce expansion. Knowing the most popular product category, Instacart could focus on broadening the range of products within the category, like adding new brands and using this for promotional communication.



Given the fact that Thursdays people buy more expensive products, the marketing could schedule ads promoting higher-end products specifically on that day to drive more spend.

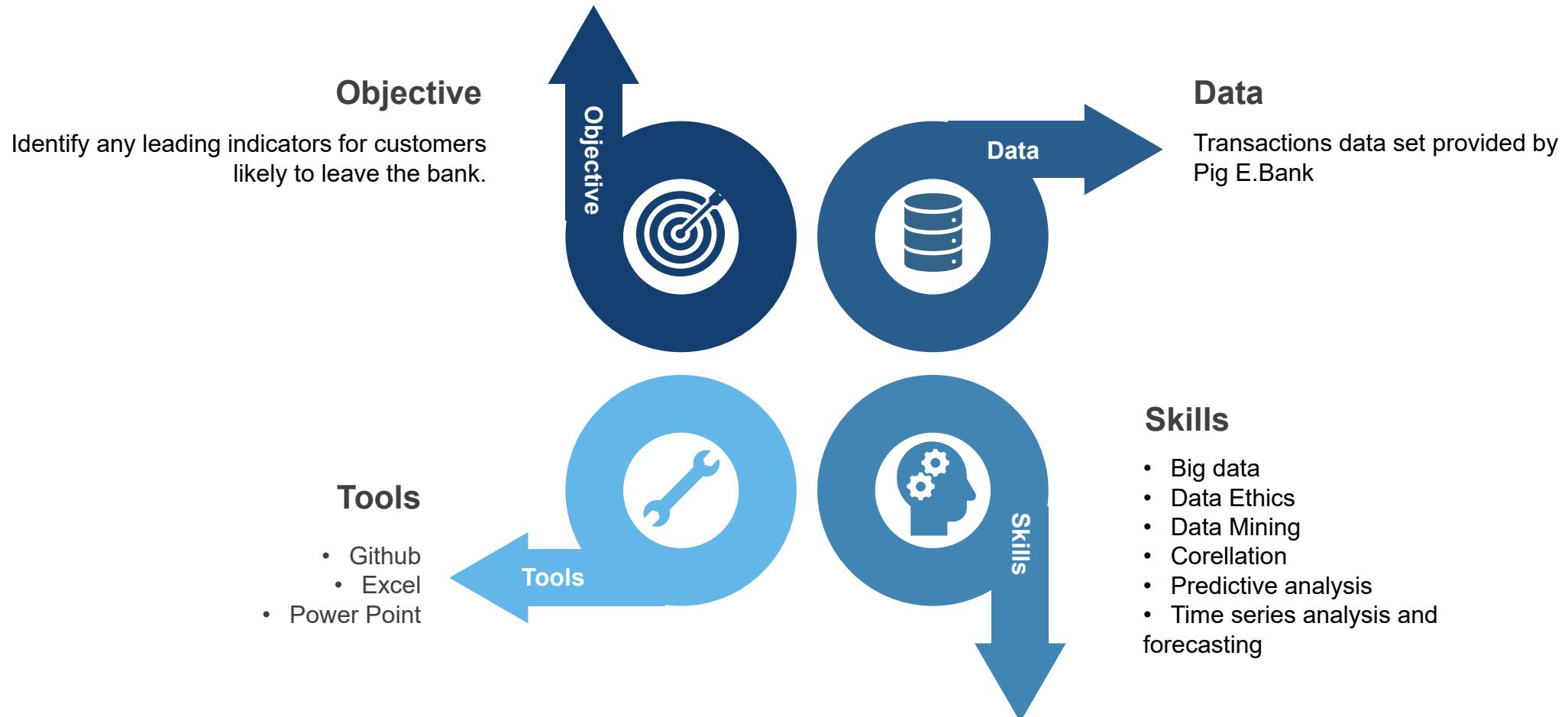


Different income and loyalty profile customers could be targeted with ads of specific goods preferences. Regional ordering habits do not differ significantly, advertising efforts should be handled in a similar manner throughout all four regions

5. Pig E.Bank

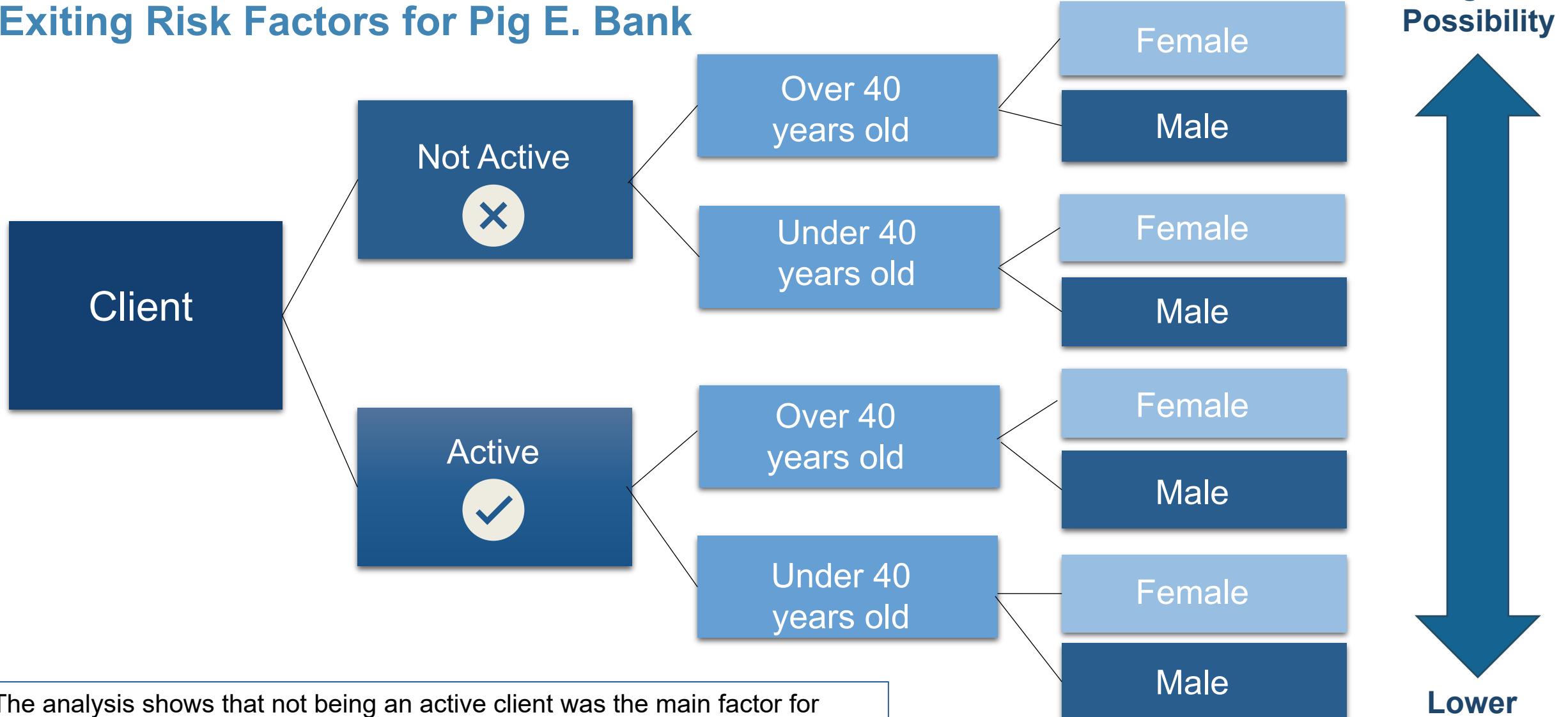


PROJECT OVERVIEW



Decision Tree

Exiting Risk Factors for Pig E. Bank



The analysis shows that not being an active client was the main factor for customers leaving the bank. Age was another main factor. The average age of the customers leaving the bank is higher than the average of those staying. Salary and gender are also factors for clients leaving.

6. AirBnB



PROJECT OVERVIEW

Objective

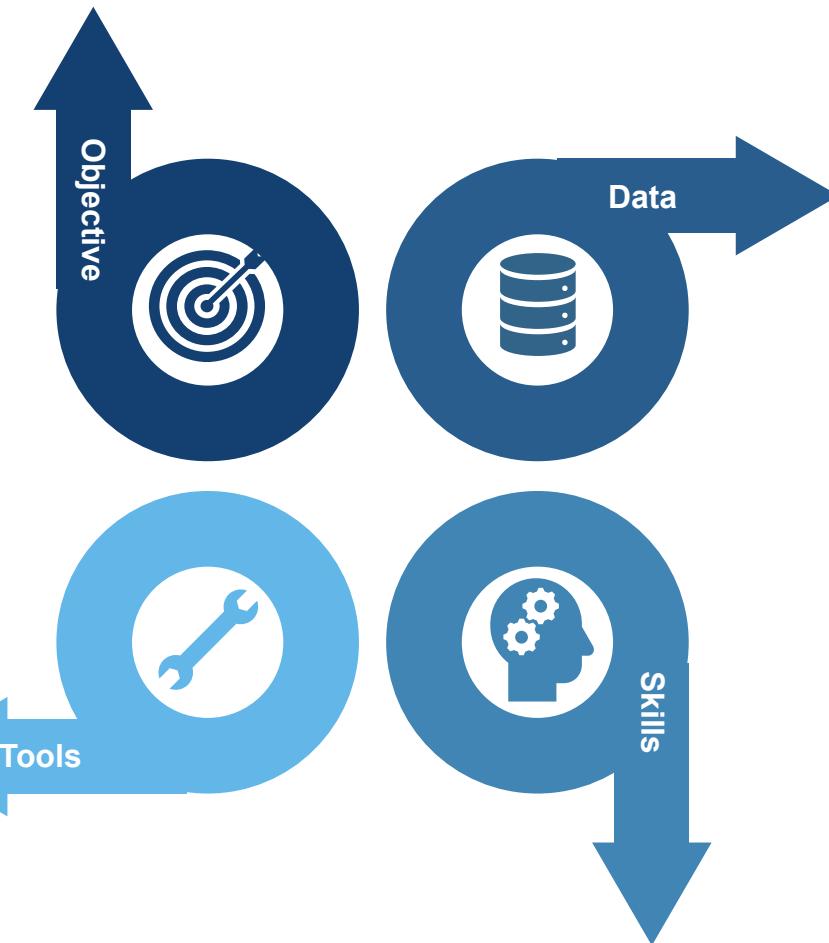
A lot of people like to use AirBnB to find accommodation for holidays. In this analysis, we investigate the connections between rating, prices and various other attributes.

The objective is to explore what can influence the travelers when choosing an accommodation through AirBnb.

Tools

This project excessively uses Python and Tableau. The following tools and libraries were used in Python:

- Anaconda 3
- Jupyter Notebook
 - Pandas
 - Seaborn
- Matplotlib
 - Scipy
 - Numpy
 - Folium
 - Geopy
 - Geoplot
- scikit-learn
- statsmodels.api



Data

This public dataset is part of Airbnb, and the original source can be found here:
<http://insideairbnb.com/get-the-data/>

Skills

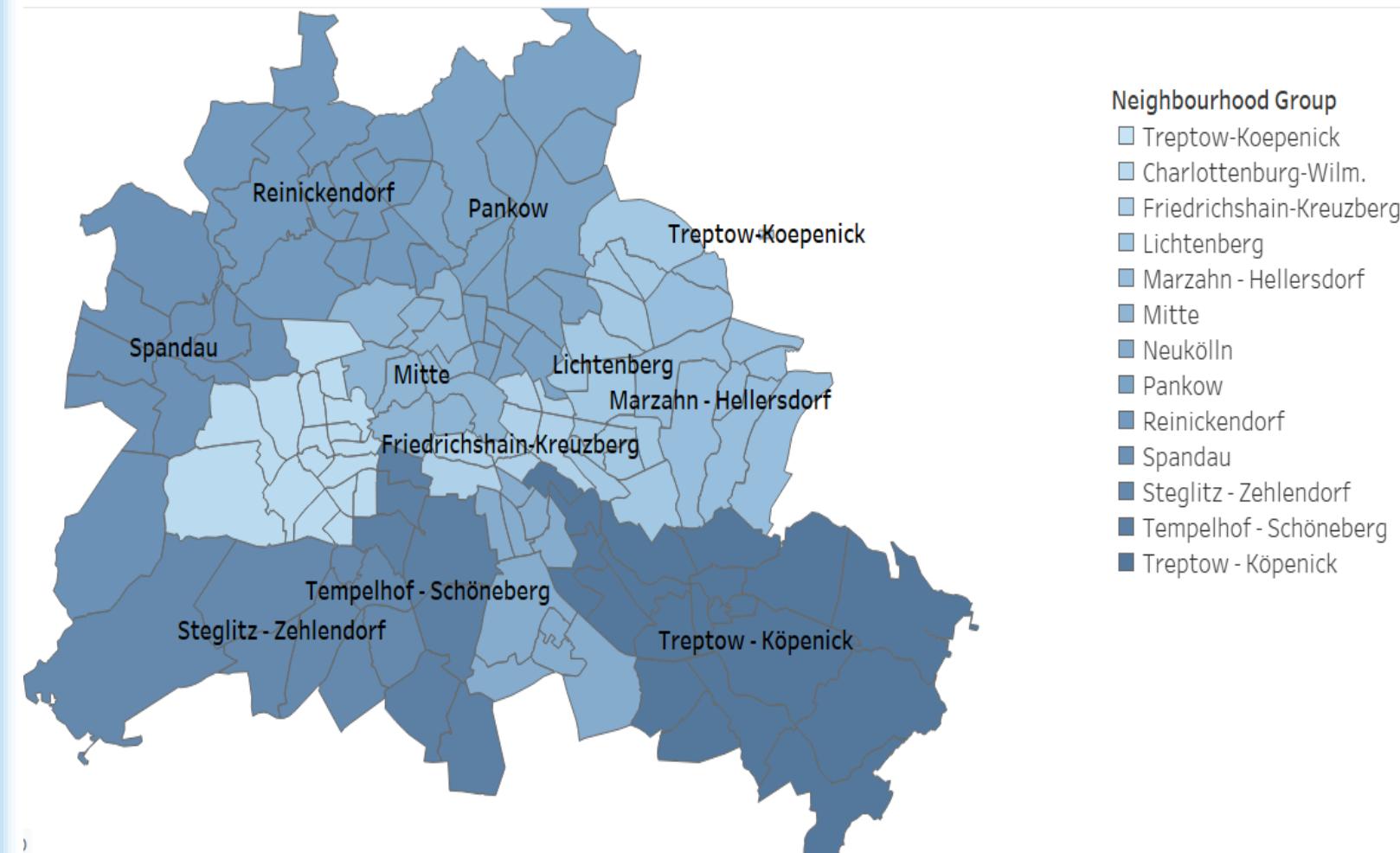
- Data Cleaning
- Data Preparation
- Data Grouping and aggregation
- Data Visualization in Python
- Data wrangling
- Big Data
- Data Ethics
- Correlation
- Predictive
- Time series
- Forecasting

AirBnB Berlin

Berlin is a metropolis of approximately 3.5 million people as of 2022. Since the mid 2000s, the city's population has rapidly grown same as the visitors that come into the city looking for a place to stay.

MOTIVATION: A lot of people like to use AirBnB to find accommodation for holidays. In this analysis, we investigate the connections between rating, prices and various other attributes.

OBJECTIVE: To explore what can influence the travelers when choosing an accommodation through AirBnb.



Neighborhood Prices in Berlin

Top 3 profitable areas are 1. Mitte, 2. Friedrichshain, and 3. Pankow.

Total EUR value spent during this time amounted to approx. EUR 1.3 million.

Still, people tend to spend more nights in less expensive neighborhoods.

The **average costs of stay is most expensive in 1. Mitte, 2. Pankow, and 3. Charlottenburg-Wilmendorf.** The more expensive neighborhoods tend to have somewhat higher averages regarding the number of reviews.



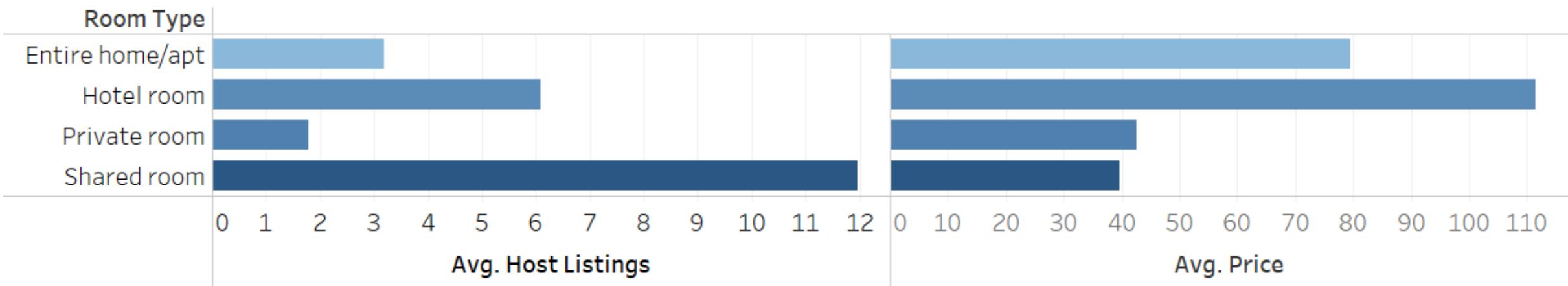
Neighborhood: Avg. Price vs Avg Number of Reviews

Neighborhood Group	Avg. Price	Avg. Nr. of Reviews
Charlottenburg-Wilm.	67.16	24.35
Friedrichshain-Kreuzbe..	63.47	24.02
Lichtenberg	53.90	20.37
Marzahn - Hellersdorf	64.70	26.40
Mitte	70.11	29.01
Neukoelln	52.55	15.69
Pankow	68.38	28.33
Reinickendorf	54.20	26.05
Spandau	58.99	19.30
Steglitz - Zehlendorf	64.21	20.13
Tempelhof - Schoenebe..	61.25	23.67
Treptow - Koepenick	62.71	16.90

Preferred Types of Rooms

Hotel rooms offered via Airbnb are generally more **expensive** than a whole home/apt - gets up to 80 euro per average, compared to a private or shared room (40-42 euro).

This is the most affordable option, but then means of course that there is a lack of privacy and amenities.



Key Findings and Conclusions



Count of Listing: 16264



Count of Hotel Room: 127



Count of Entire Home: 9138



Average Minimum Nights: 6



Count of Private Room: 6793



Count of Shared Room: 206



Average Price: 43 euro



Average Nr. of Reviews: 19

Key Findings:

1. Accommodation prices are generally higher towards the city center.
2. Most profitable areas were near the city center - Mitte, Friedrichshain-Kreuzberg, and Pankow. (*These areas are also known for their tourist attractions, bars, cafes, and clubs*).
3. Shared rooms offer the most affordable stays on average, whereas hotel rooms listed on AirBnB are the most expensive.
4. The most profitable areas are not also the ones where people tend to spend more nights, most probably due to higher prices.

6. World Happiness



PROJECT OVERVIEW

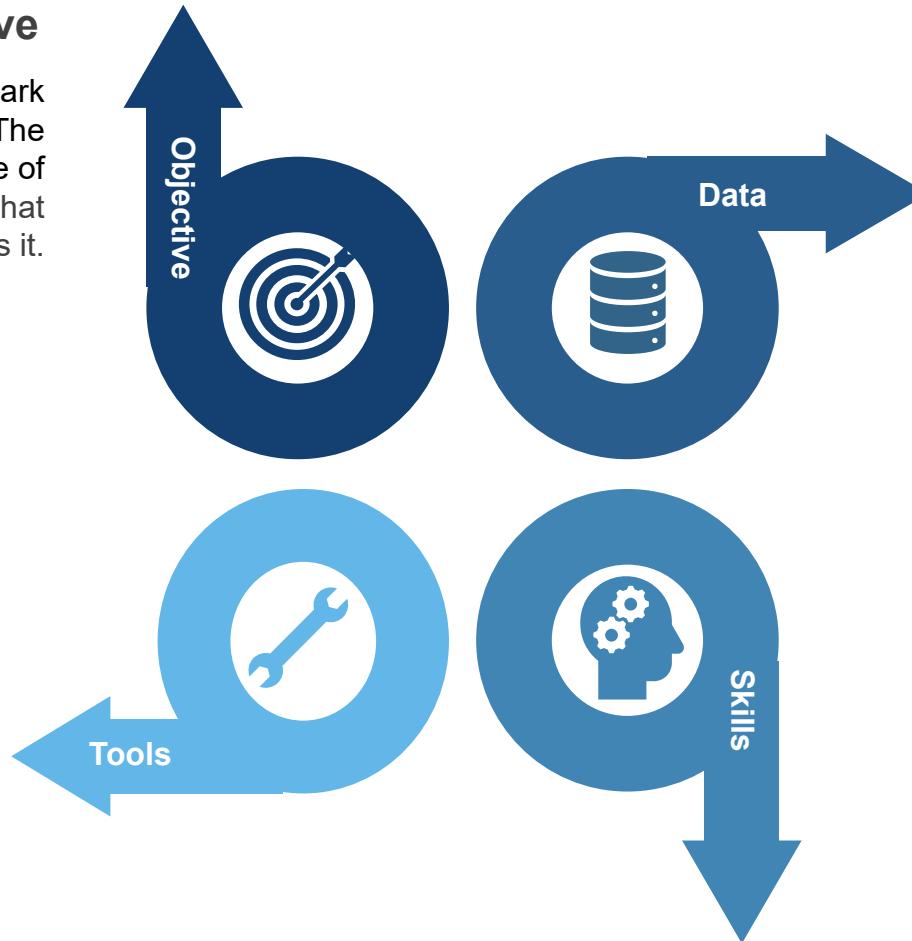
Objective

The World Happiness Report is a landmark survey of the state of global happiness. The objective is to give a clear picture of happiness around the world and what influences it.

Tools

This project excessively uses Python and Tableau. The following tools and libraries were used in Python:

- Anaconda 3
- Jupyter Notebook
- Pandas
- Seaborn
- Matplotlib
- Scipy
- Numpy
- Folium
- Geopy
- Geoplot
- scikit-learn
- statsmodels.api



Data

The happiness scores and rankings use data from the Gallup World Poll. The scores are based on answers to the main life evaluation questions asked in the poll. Happiness score is a national average response to the question: "Please imagine a ladder, with steps numbered from 0 at the bottom to 10 at the top"

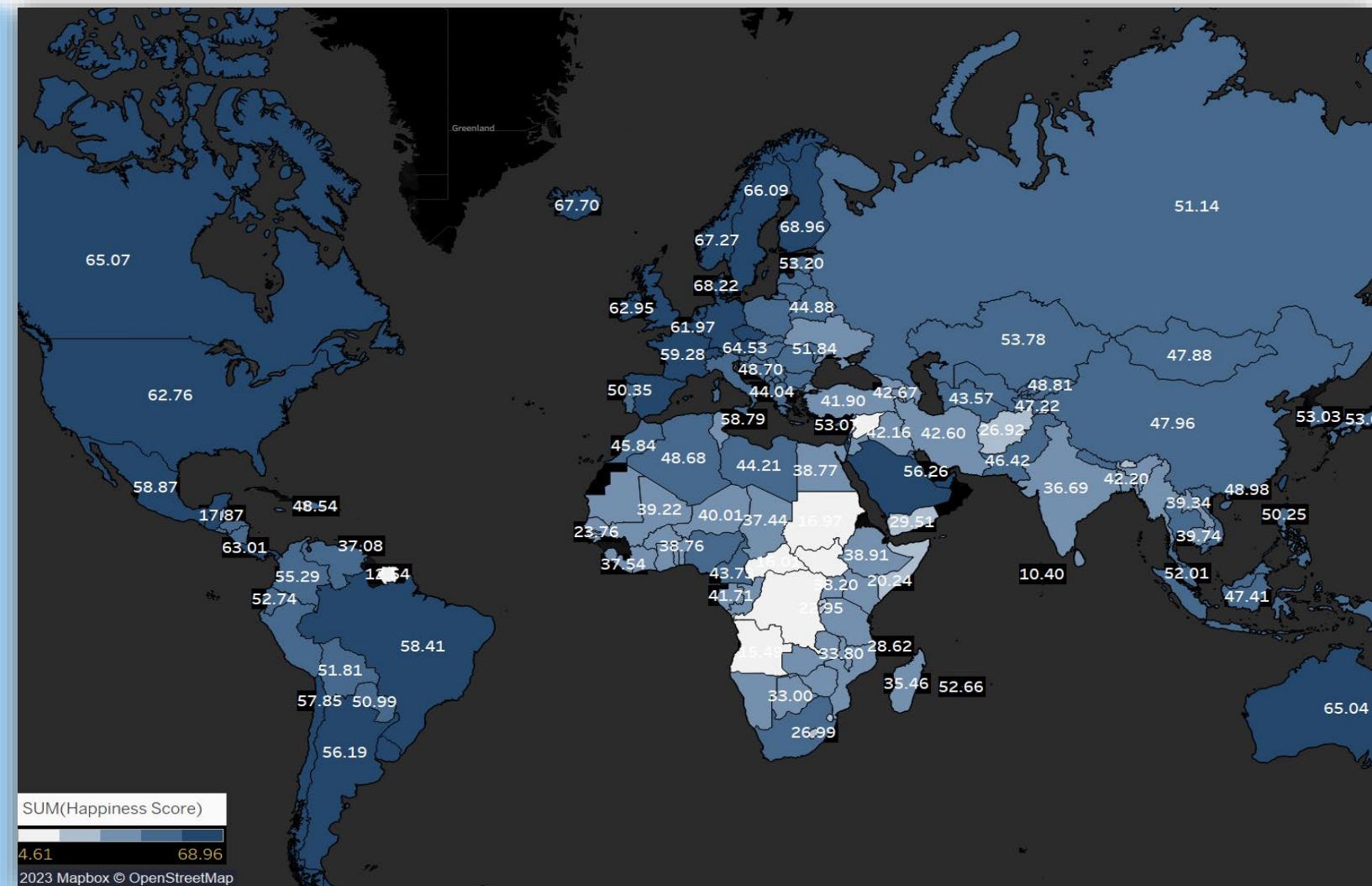
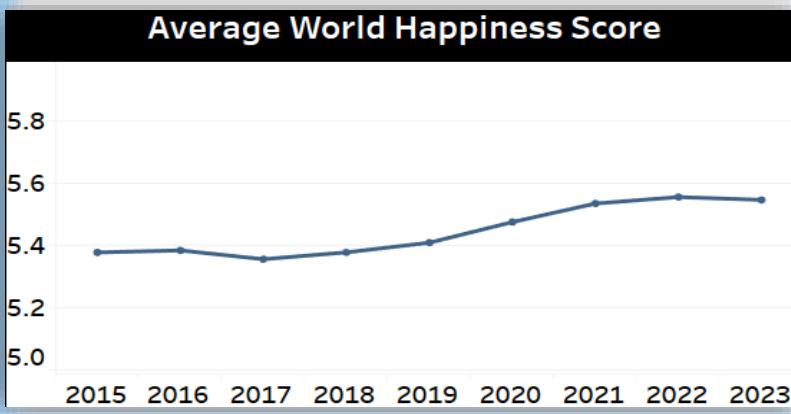
Skills

- Data Cleaning
- Data Preparation
- Data Grouping and aggregation
- Data Visualization in Python
- Data wrangling
- Big Data
- Data Ethics
- Correlation
- Predictive
- Time series
- Forecasting

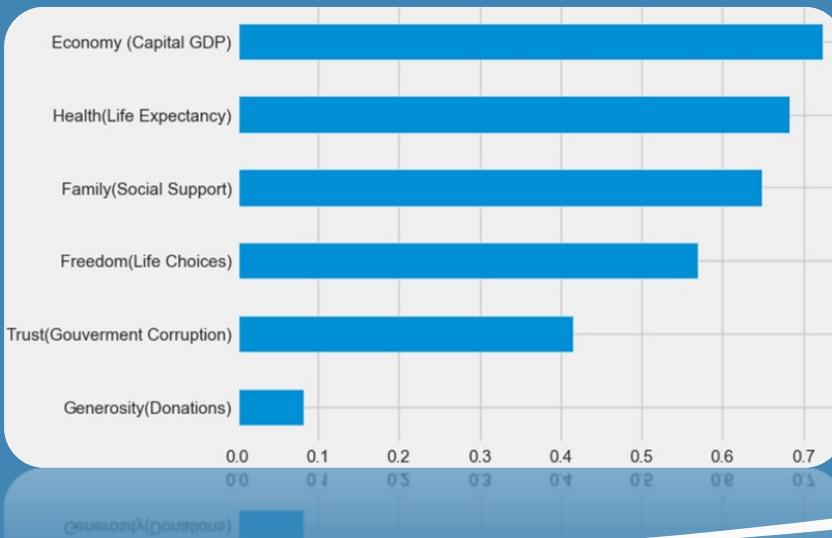
How happy is your country?

Consistently, the happiest countries in the world are those in **Northern Europe** (Finland has been leading for several years), along with Switzerland and the Netherlands. Meanwhile, the countries with the lowest happiness scores are those in **Sub-Saharan Africa** and Afghanistan, which has ranked at the bottom of the list for the past three years.

Over the past few years, most countries have maintained relatively stable Happiness Scores, which align with the trends depicted in the line chart below displaying the Average World Happiness Score.



Happiness Factors



The correlation coefficient is a measure of strength between two variables, ranging from -1 to 1. The closer the number is to 1, the stronger the positive relationship. A strong correlation suggests that these variables tend to move together. However, this doesn't mean that one directly causes the other.

The graph shows different degrees of correlations between **Happiness Score** and the 6 factors.

The World Happiness Report measures happiness and well-being in countries around the world using a range of indicators and questions in their surveys. The 6 factors determining happiness are:

- 1. Corruption Perceptions:** This indicator assesses the level of corruption within a country's government and business sectors.
- 2. Freedom to Make Life Choices:** This indicator measures the extent to which people feel they have the freedom to make life choices that affect their well-being.
- 3. GDP Per Capital:** Questions about income, job satisfaction, and economic stability are often included to gauge the relationship between financial well-being and happiness.
- 4. Generosity:** This indicator assesses charitable behavior, such as whether individuals have donated to a charity in the past month.
- 5. Social Support:** This indicator assesses the availability of social support systems, including questions about having someone to count on in times of need.
- 6. Healthy Life Expectancy:** This factor takes into account the number of years of good health that people can expect to have in their lifetime.

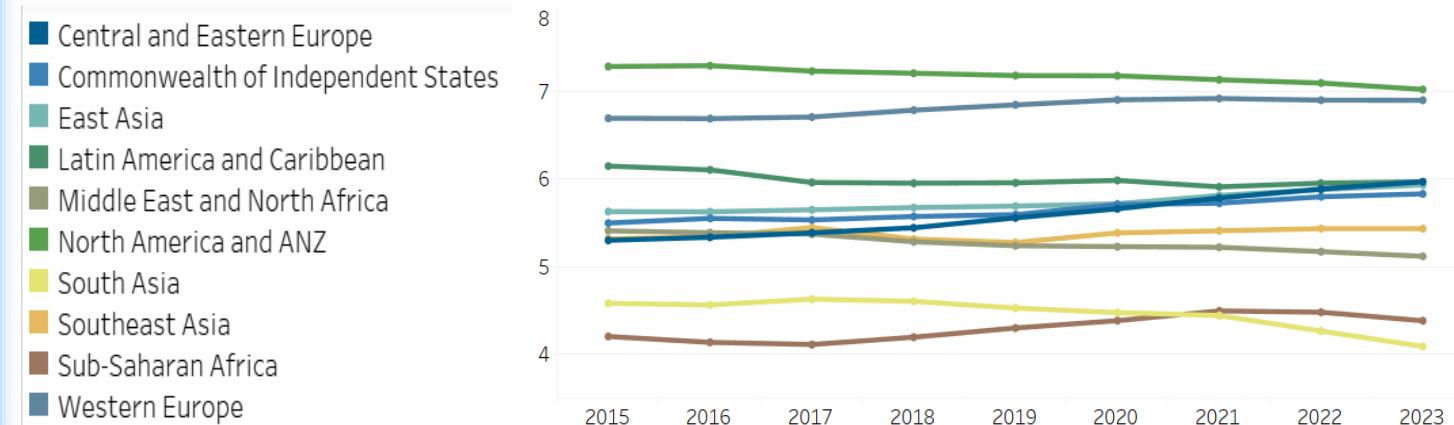
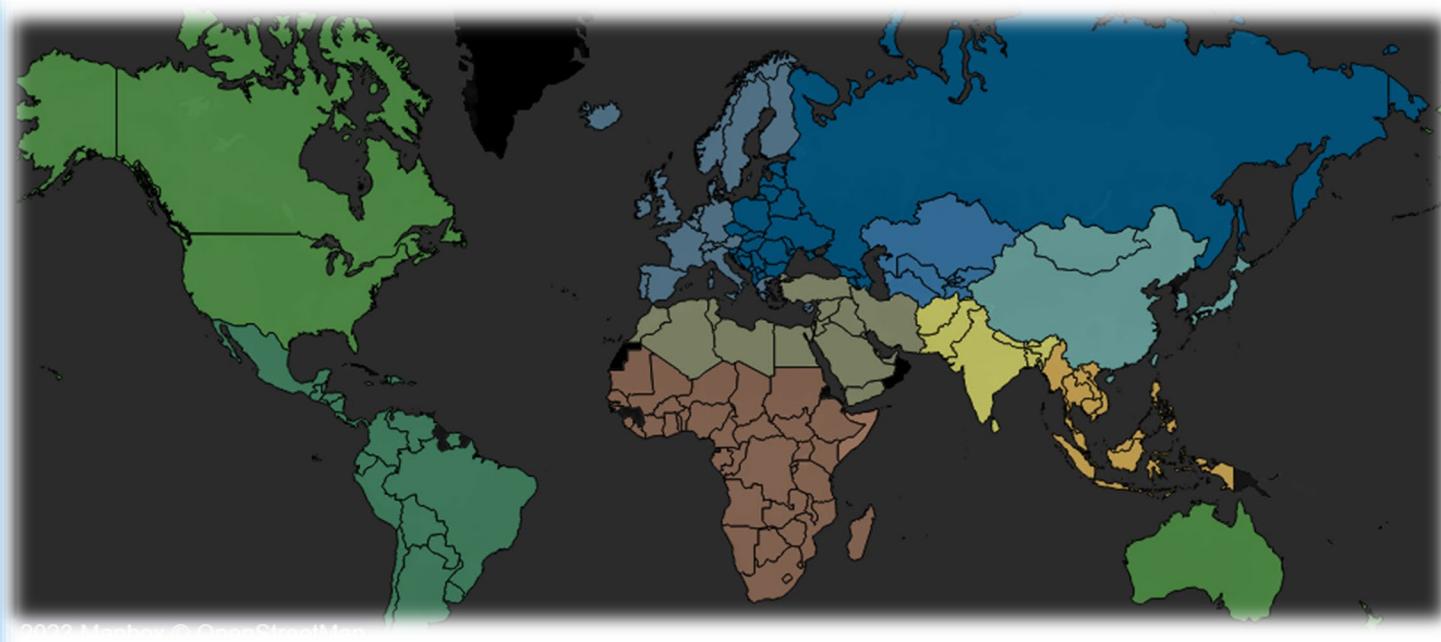
Happiness by Region

In recent years, regions have remained relatively stable. The most significant growth has been recorded in **Central and Eastern Europe**, while the largest decline is attributed to the substantial drop in **Afghanistan's ranking in South Asia**.

The regions most represented in the dataset are Sub-Saharan Africa (22.0%) and Central and Eastern Europe: (19.9%).

Like Australia & New Zealand, North America represent less than 2% of the dataset countries but there is no surprise since they only contain United States and Canada.

Some countries managed to increase their **level of happiness** over the years or keep it almost the same. On the other hand, there are countries that are still unhappy with their way of living or others where the happiness decreased in time due to government, wars, poverty, or other factors in cause.



RECOMMENDATIONS

GDP and Health Life Expectancy have been identified as the most influential factor of happiness level. The following recommendations can help countries prioritize economic well-being to improve overall happiness:



Income Equality: Address income inequality through progressive taxation, social safety nets, and wealth redistribution programs to ensure that the benefits of economic growth are more equitably shared.



Economic Development and Growth: Implement policies and strategies that promote economic growth and development, such as investment in infrastructure, innovation, and entrepreneurship.



Education and Skills Development: Promote education systems that foster lifelong learning, allowing individuals to continuously acquire new skills and knowledge, making them more competitive and adaptable to changing economic demands.



Financial Education Programs: Promote financial education to empower individuals to make informed financial decisions and improve their economic well-being.



Universal Healthcare Access and Preventing Healthcare : Ensure universal access to quality healthcare services, reducing disparities in healthcare outcomes and improving overall health. Focus particularly on regions with limited access to healthcare facilities to prevent issues.



Work-Life Balance and Community and Social Support Programs: Continue to prioritize work-life balance by implementing policies that reduce work-related stress and encourage leisure and family time. Encourage the maintenance and improvement of social safety nets.

THANK YOU

Ioana Rusu - Data Analyst

[Github Profile](#)

[Tableau Profile](#)