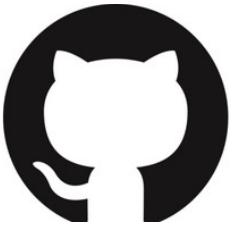


# RENU BALAJI

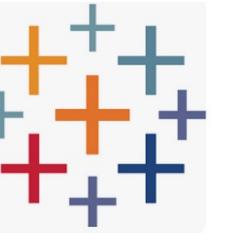
## DATA ANALYTICS PORTFOLIO



[GitHub](#)



[LinkedIn](#)

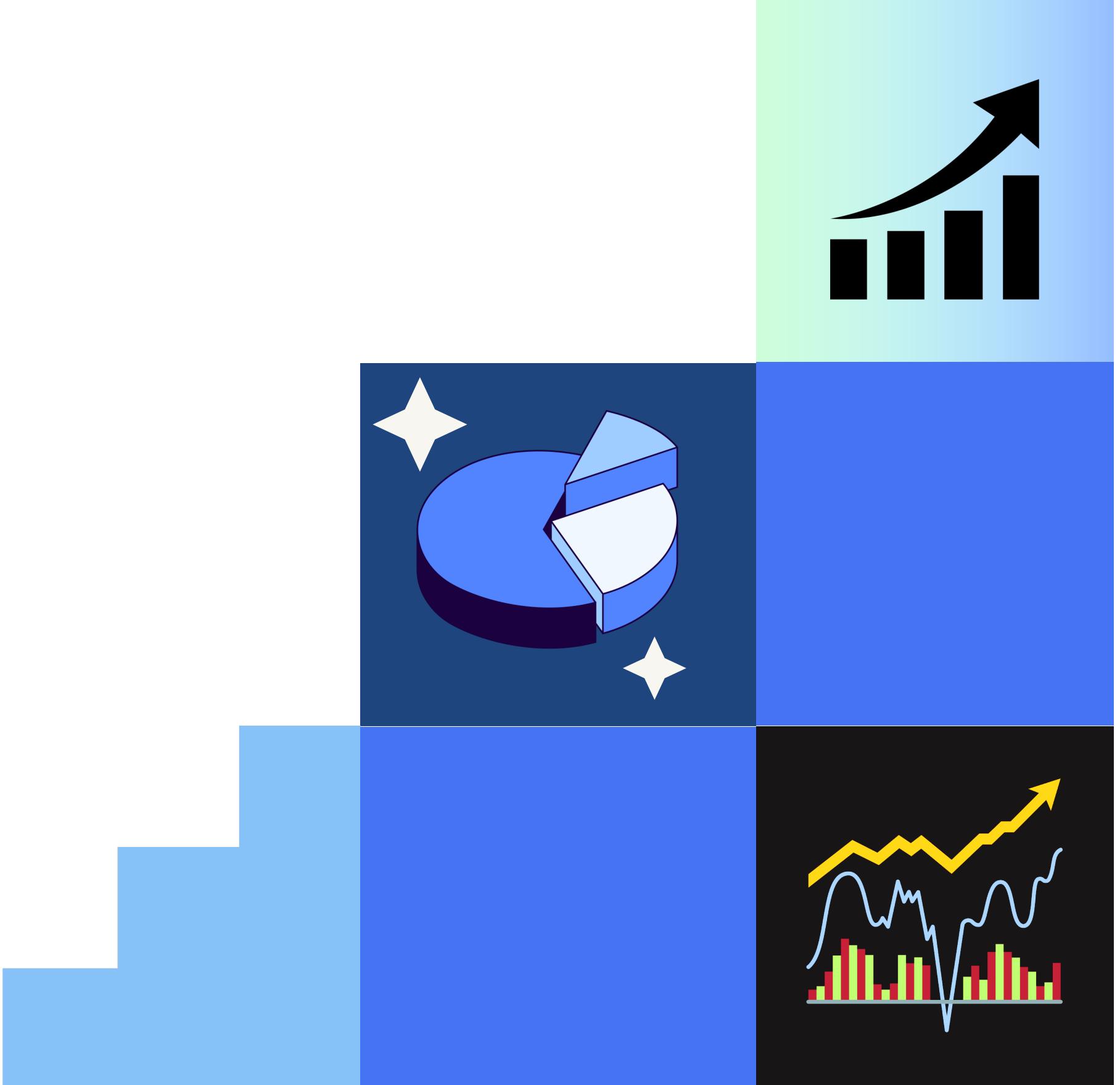


[Tableau](#)



12balaji.renu@gmail.com

matchSpace



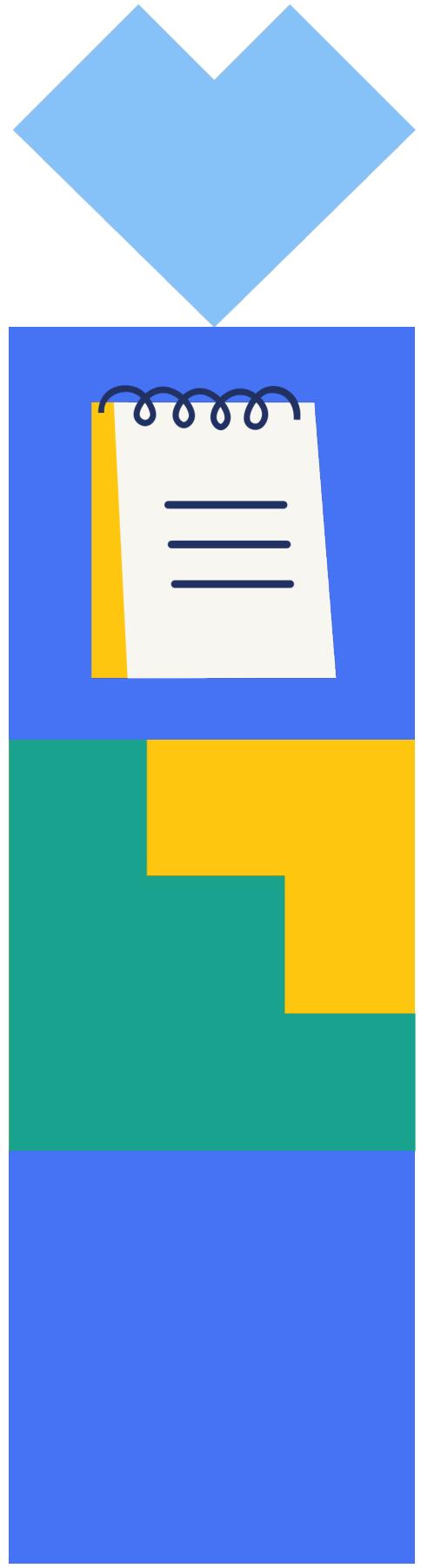


## ABOUT ME:

I am a data-driven professional with a background in molecular biology, data analysis, and clinical research. I excel at transforming complex data into actionable insights that drive strategic business decisions. With a client-focused approach and a knack for problem-solving, I leverage data to optimize processes, improve outcomes, and uncover growth opportunities. I am passionate about using data to make a tangible impact and eager to apply my expertise to help businesses thrive.

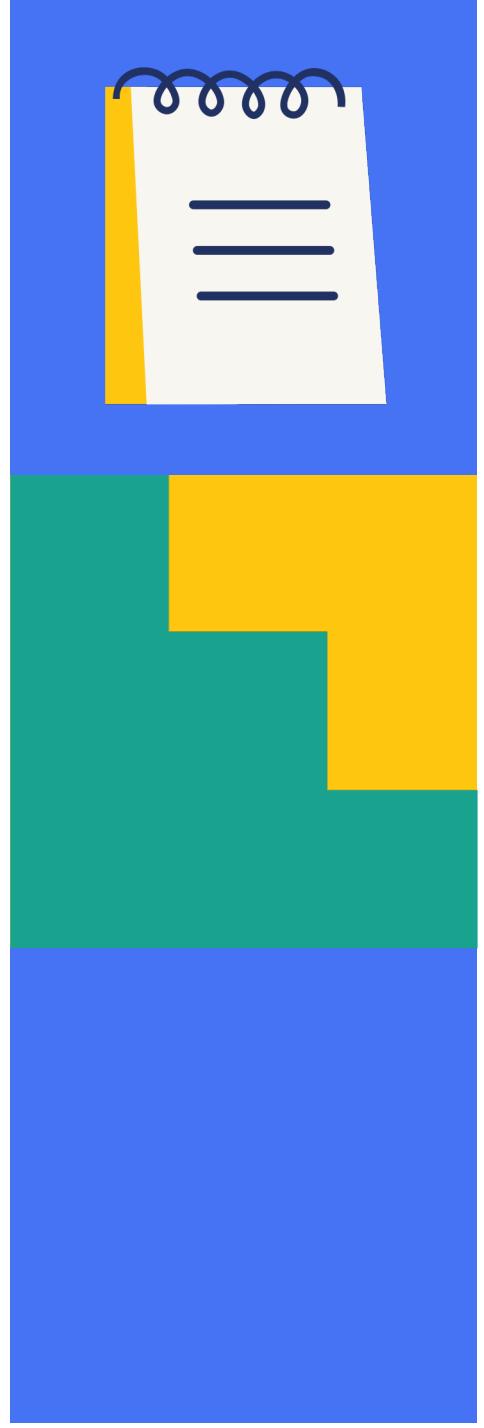


# PROJECTS



## 01. GameCo

Data analysis of Sales across the Global Market



## 02. Preparing for Influenza Season

Planning of staffing across regions during the influenza season

## 03. Rockbuster Stealth LLC

Analysing marketing patterns to launch online video rentals

## 04. InstaCart Shopping

Discerning Sales patterns of the online app

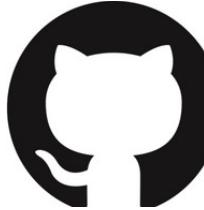
## 05. Global Banking (Pig E. Bank)

To analyse customer attrition

## 06. World Economic Indicators

To analyse country-wise economic indices

# TOOLS



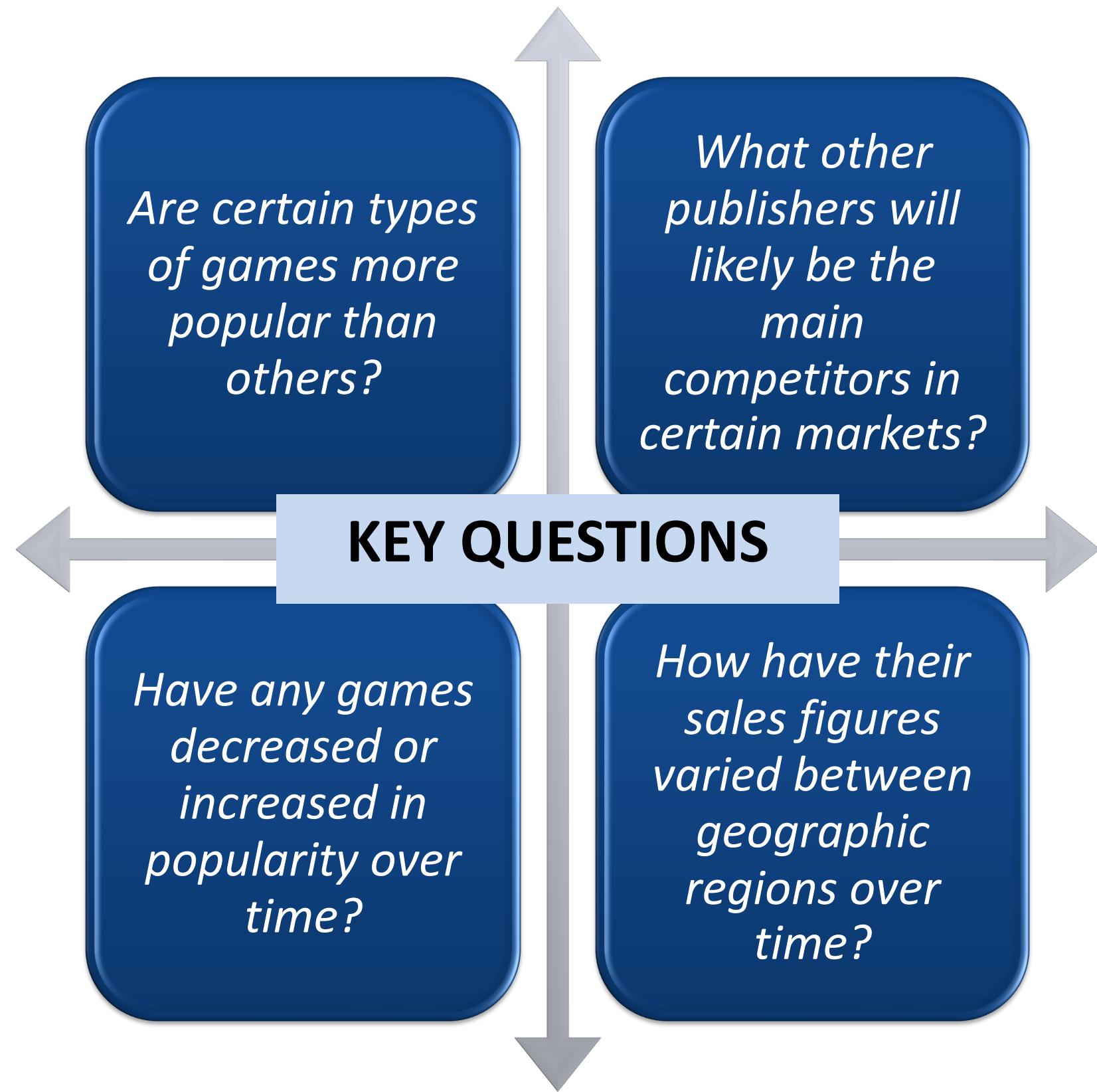
# 01. GameCo

Understanding the global market and sales in an online video game company by planning the marketing budget for the upcoming year

## Objectives

The primary objective was to conduct an end-to-end analysis of the sales data. This includes outlining the business problem, grouping and summarizing the data, formulating hypotheses, using data to prove the hypotheses, visualizing and presenting insights obtained.





- **DATA**  
Source: [VGSales](#)  
Period: 1980-2016  
Information provided: title, platforms, year, genre, publisher
- **TOOLS**  
Excel: Data analysis [Github](#)
- **SKILLS**  
Cleaning and data grouping  
Descriptive Statistics and Exploratory Data Analysis (EDA)  
Creating Pivot tables  
Visualising results  
PowerPoint Presentation

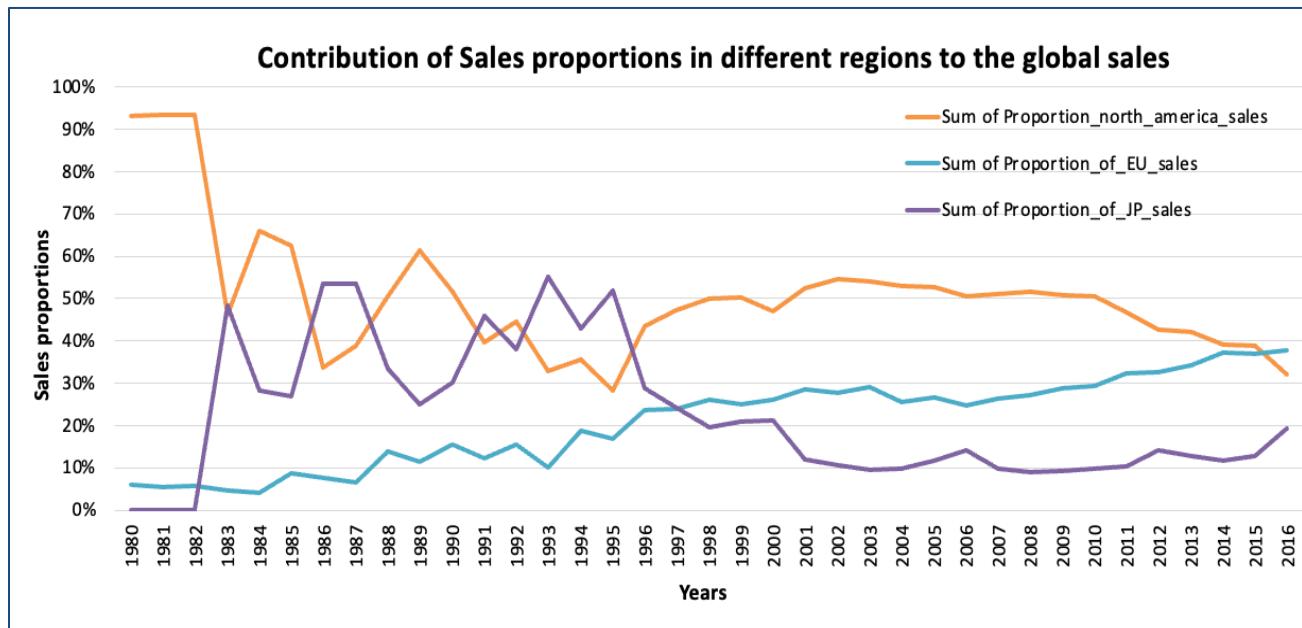
# ANALYSIS

The analysis indicated that the proportion of sales across the years in different locations varied significantly. The sales in Japan varied from 1982 when the sales constituted ~45-50% without a steady change till 1996 after which there was a steady fall in sales till 2001 when it reached about 10%. In 2016, the sales increased slightly from 10% in the previous year to 20%. At the same time, the sales across the EU increased from below 10% to 40% from 1980-2016. The EU sales surpassed the NA sales which was on the top since the last 18 years.

Across the regions, action-based video games sold the most constituting 20% of the total sales.

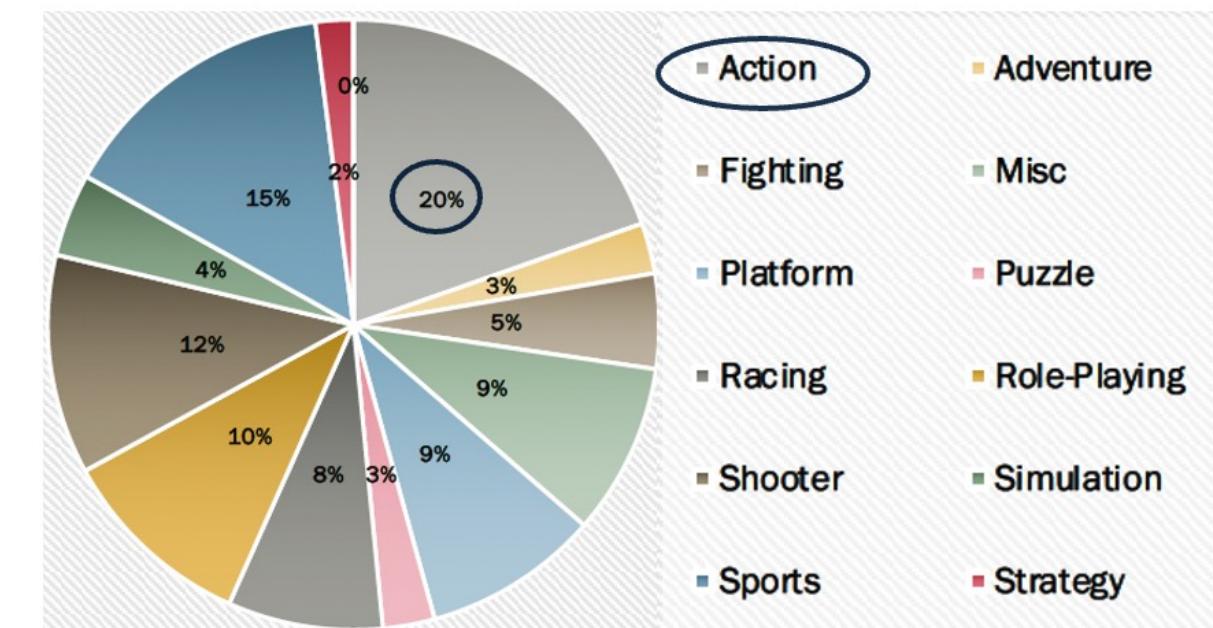
Nintendo sold the highest number of video games (1800 (in millions) per annum global sales) compared to the rest of the publishers.

## TOTAL SALES ACROSS DIFFERENT REGIONS

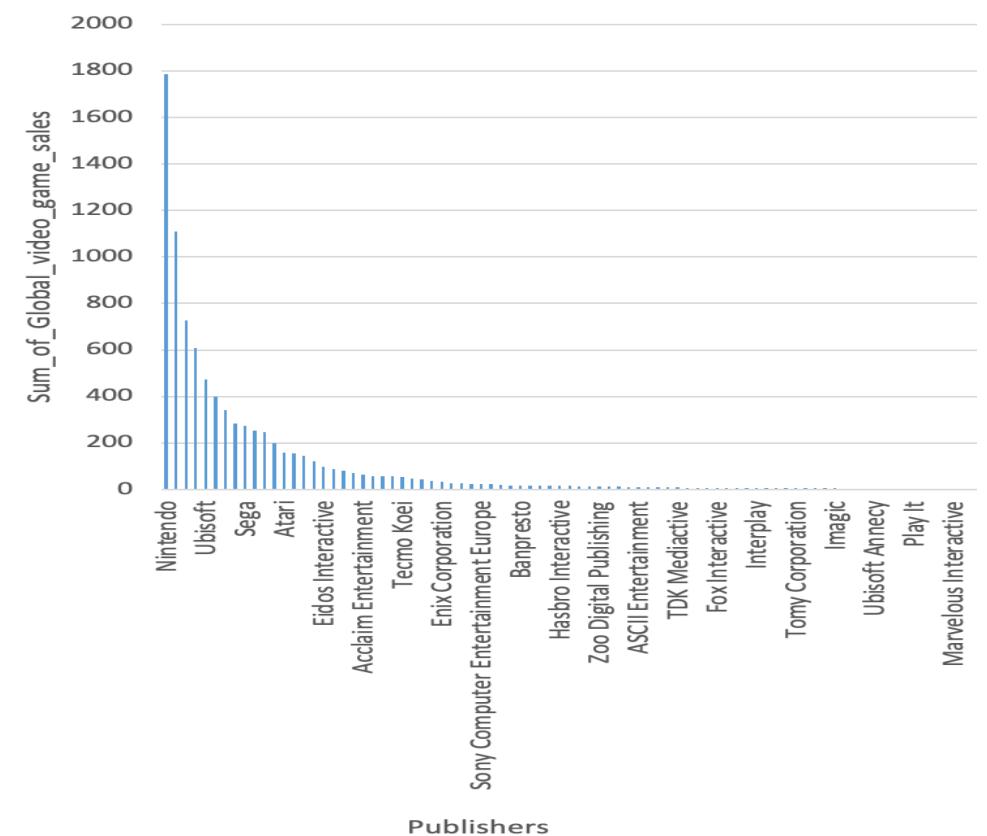


\*Did not consider sales of other countries in the global sales since I wanted to understand the individual contribution of these three countries to the global sales.

## % of video game sales globally GENRE-WISE



## PUBLISHER'S CONTRIBUTION TO SALES



# 01. GameCo

EU sales gained a steady increase over time in its contribution to global sales and may also continue to increase. Therefore, GameCo must focus on EU region for their marketing and continue the marketing strategy that is currently being used since it seems to be working well.

## EUROPE

The marketing strategies for NA and Japan must be refined since both countries have lowered their contribution to global sales.

## NORTH AMERIA AND JAPAN

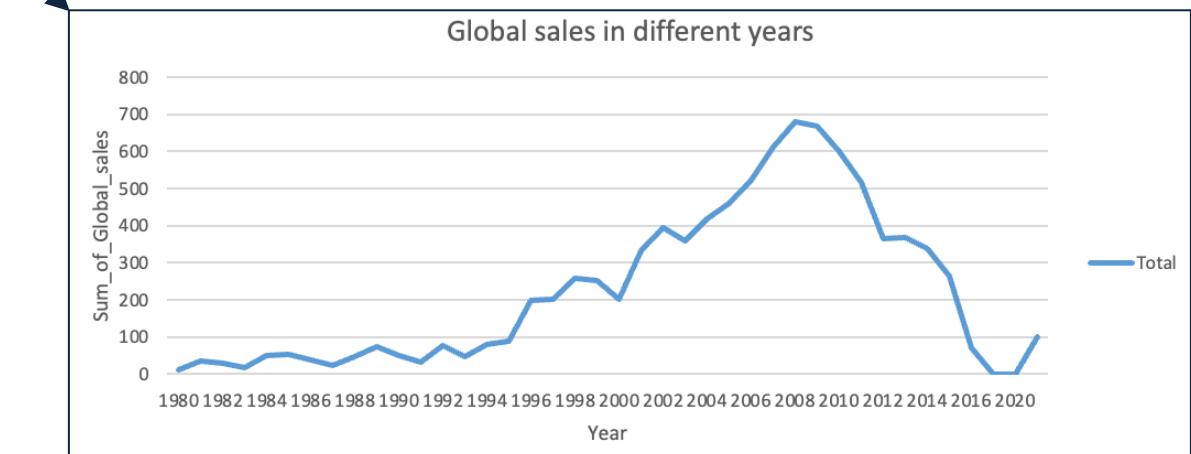
## GENRE

Since action-based games are being sold in higher numbers globally, GameCo must concentrate more on this genre while releasing new games.

## RECOMMENDATIONS

## COMPETITOR ANALYSIS

It also seems like market competition has increased since there has been an overall drop in video game sales from 2008 to 2016 at GameCo. Although, in 2020 there seemed to be a small increase in sales from 0 to 100. Therefore, other advertising and sales strategies must be implemented.

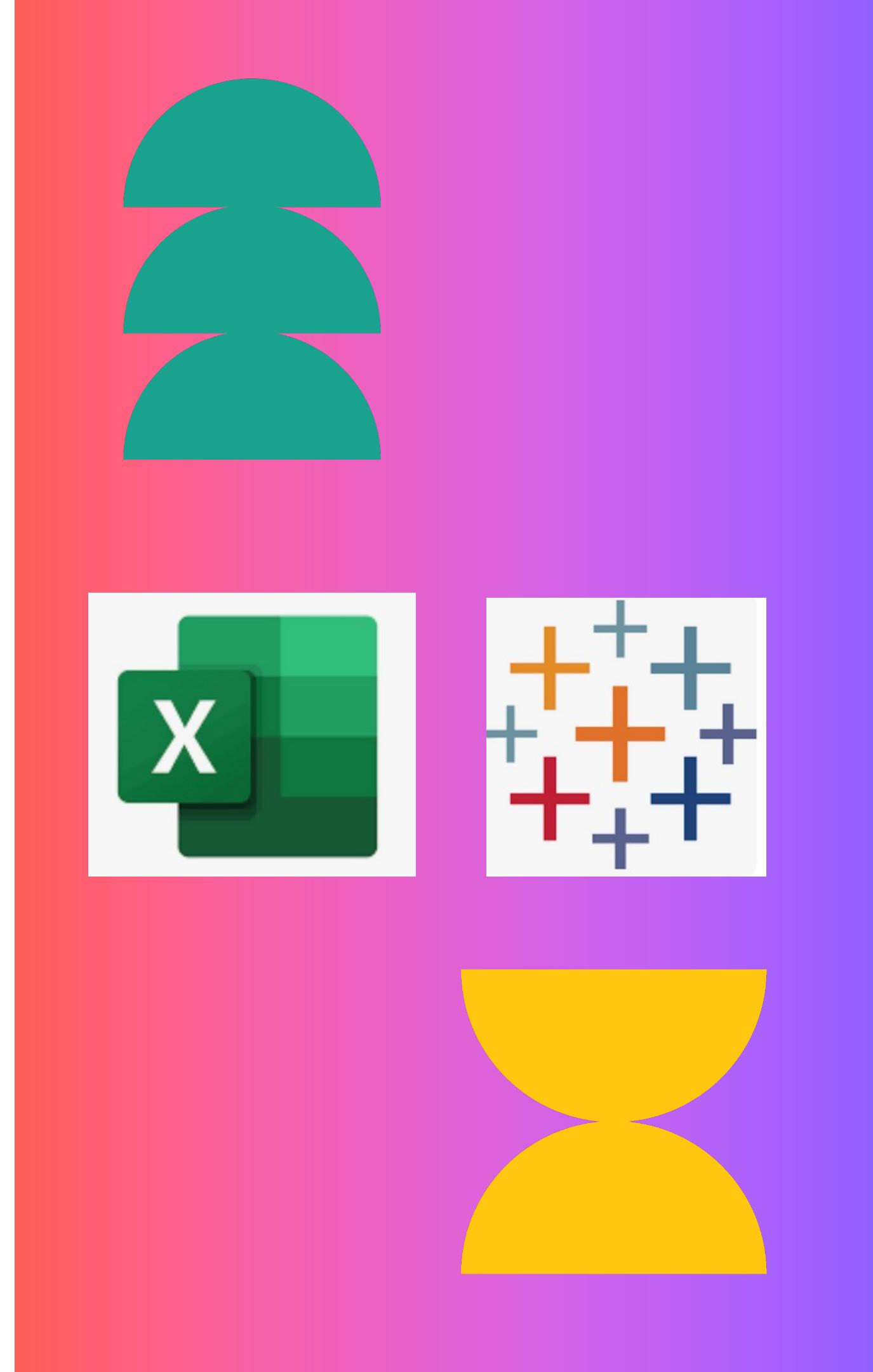


# 02. Preparing for Influenza Season

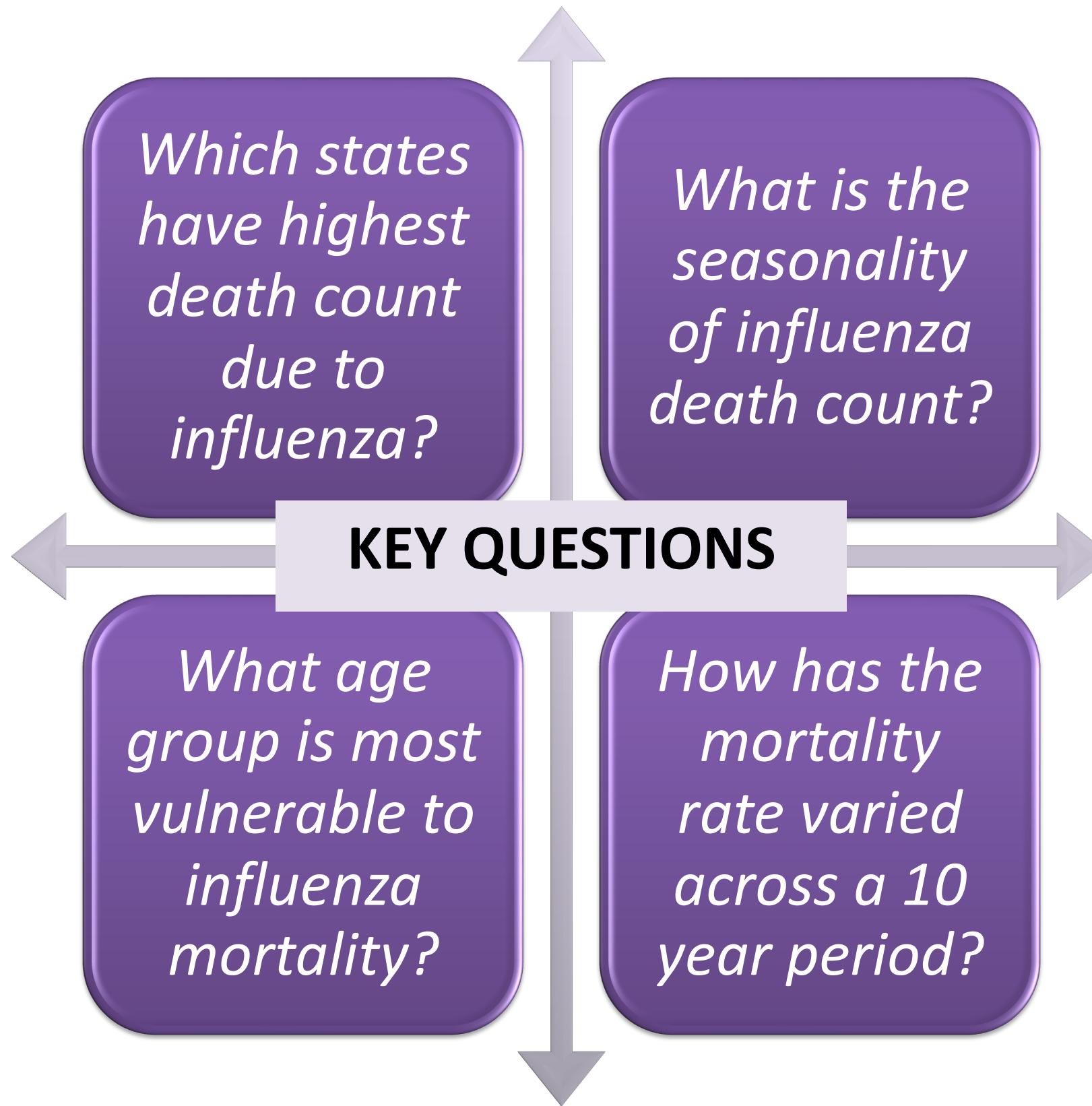
Understanding the staffing requirements during peak influenza season

## Objectives

A medical staffing agency requires data-backed insights to help plan for influenza season, a time when additional staff are in high demand. This includes the examination of trends in influenza and how it can be used to proactively plan staffing needs across the country. The agency covers all hospitals in each of the 50 states of the United States.



## 02. Preparing for Influenza season



### ■ DATA

#### Source:

a. [Census Population](#): (US Census Bureau)

Period: 2009-2017

Information provided: country, age and gender

b. [Influenza Mortality count](#) (CDC)

Period: 2009-2017

Information provided: Influenza mortality count across US population and age

### ■ TOOLS [\(GitHub\)](#)

Excel: Data analysis

Visualisation: Tableau

### ■ SKILLS

Outlining business requirements

Building Hypotheses

Data cleaning, Descriptive Statistics

Exploratory Data Analysis (EDA)

Creating Pivot tables

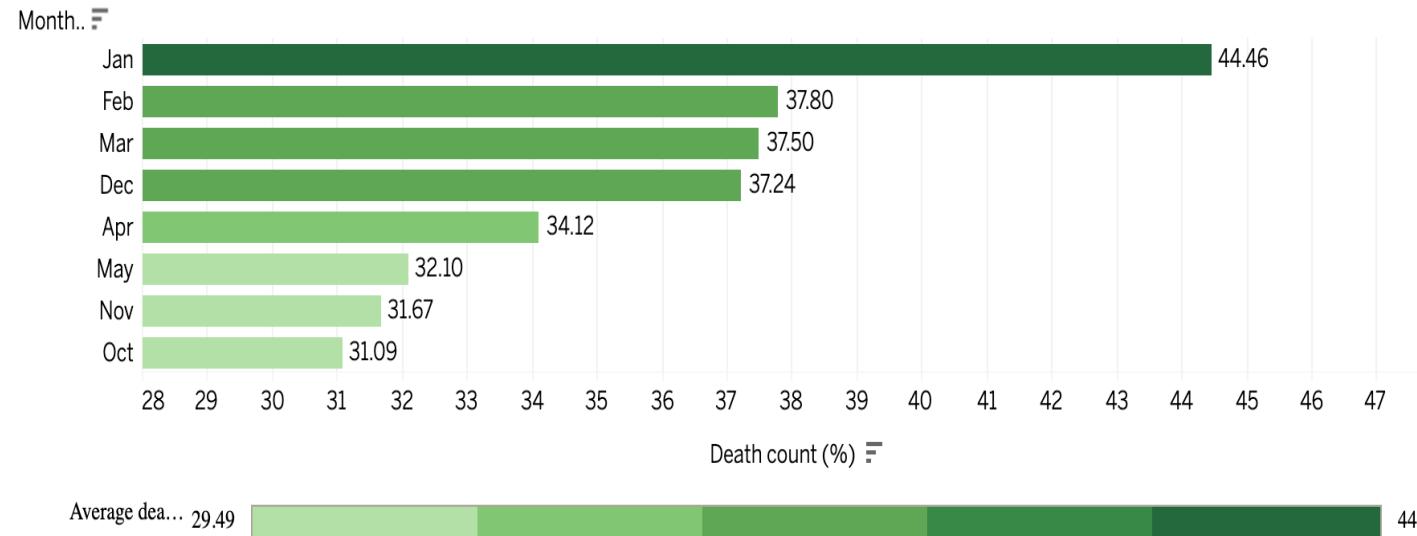
Visualising results using Tableau

Storytelling

Creating visual dashboards

## 02. Preparing for Influenza season

### INFLUENZA ANNUAL SEASONALITY



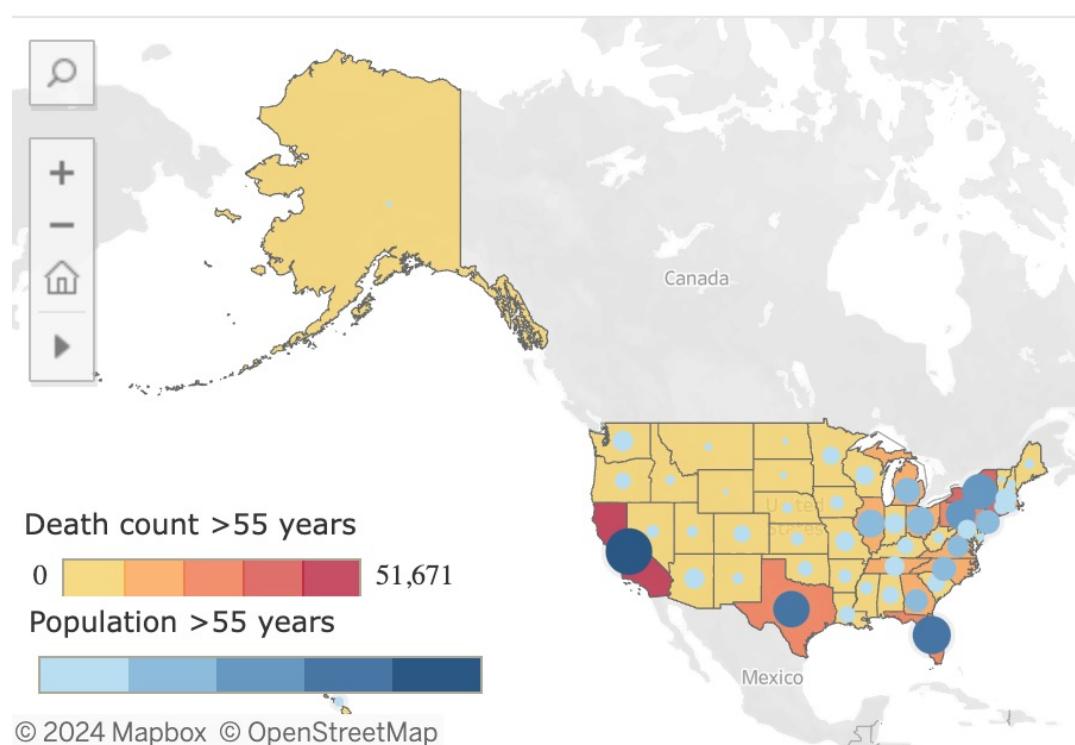
### ANALYSIS

January has the highest (44%) death count and the influenza season begins around December and lasts till April.

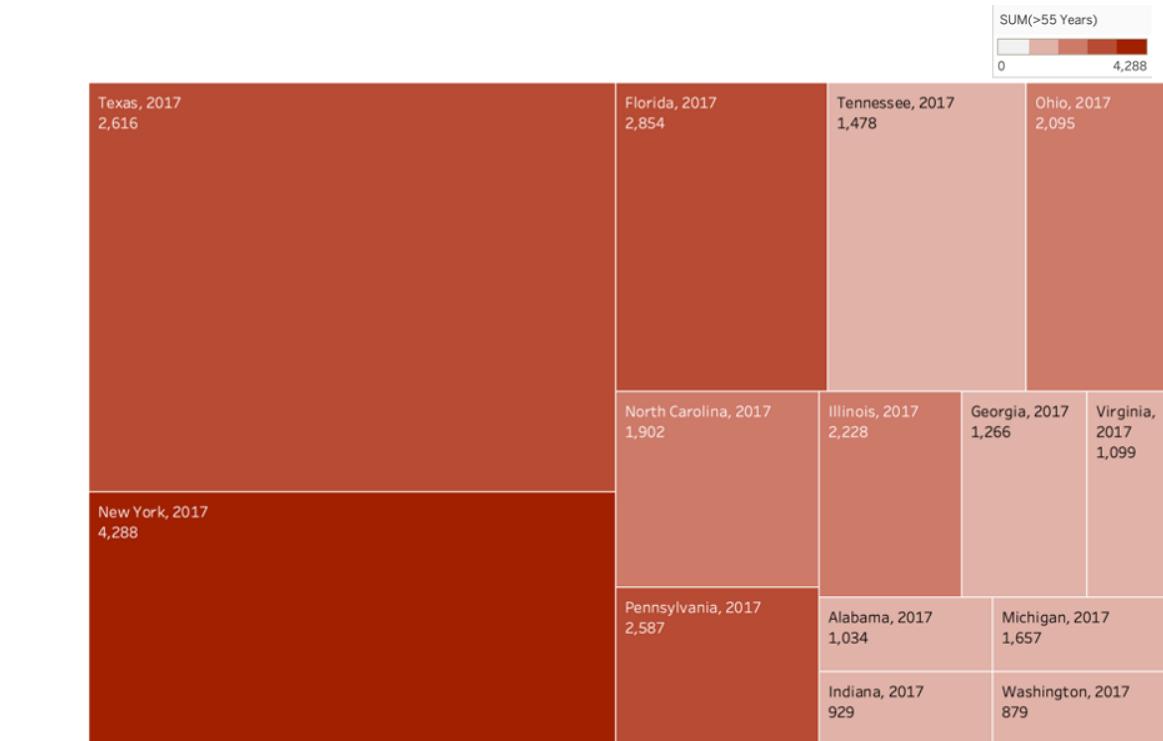
Highest staffing needs during peak season- New York, Texas and Florida

The highest death count was in the year 2015 when there were 49,650 deaths while the year 2010 had the lowest death count of 42,524.

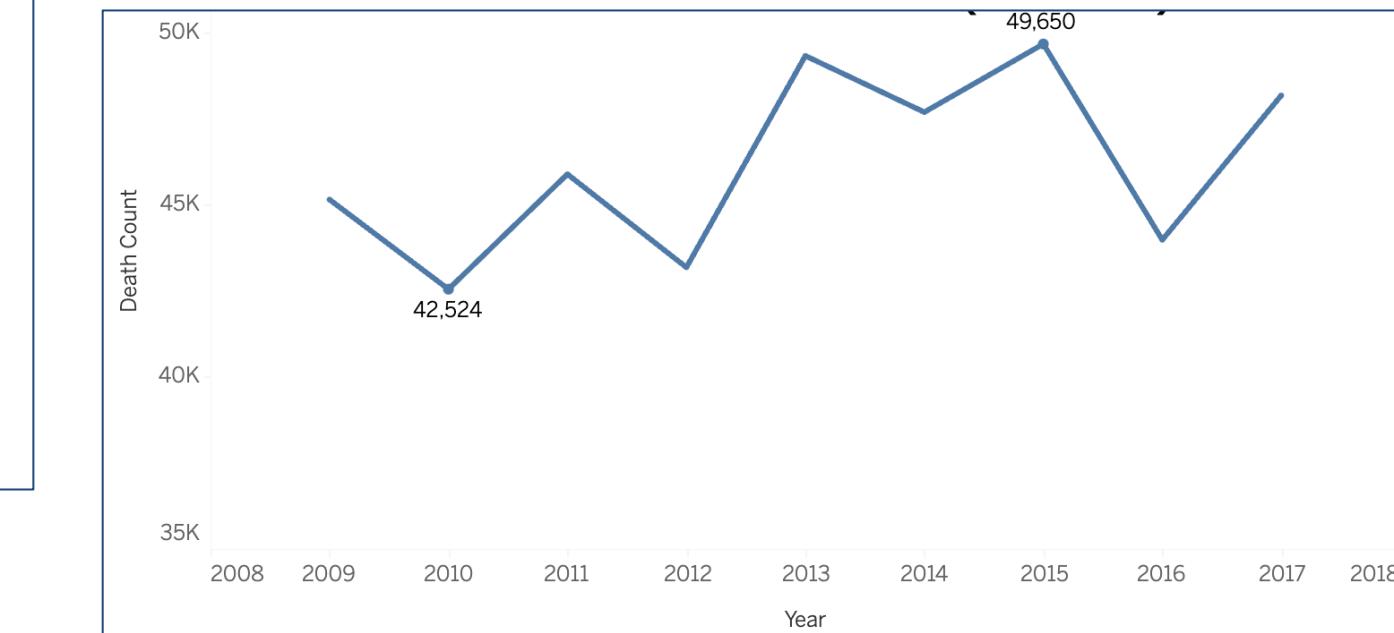
### POPULATION & INFLUENZA DEATH COUNT STATEWISE



### INFLUENZA DEATH COUNT (>55 yrs)



### INFLUENZA DEATH COUNT (2008-2017)



## 02. Preparing for Influenza season

The highest staffing requirement will be from the months December to March

### MONTHLY STAFFING

The states requiring highest staff during peak influenza period would be California, New York, Texas and Pennsylvania

### STATEWISE STAFFING

### AGEWISE STAFFING

People >55yrs of age fall under the vulnerable population category and would require more staff support

### RECOMMENDATIONS

### FUTURE RECOMMENDATION

Vaccination data has to be analysed to understand vaccination rate to death ratio across states and months. This will help in understanding the trend across years which accounts to the spikes and drops in death count.

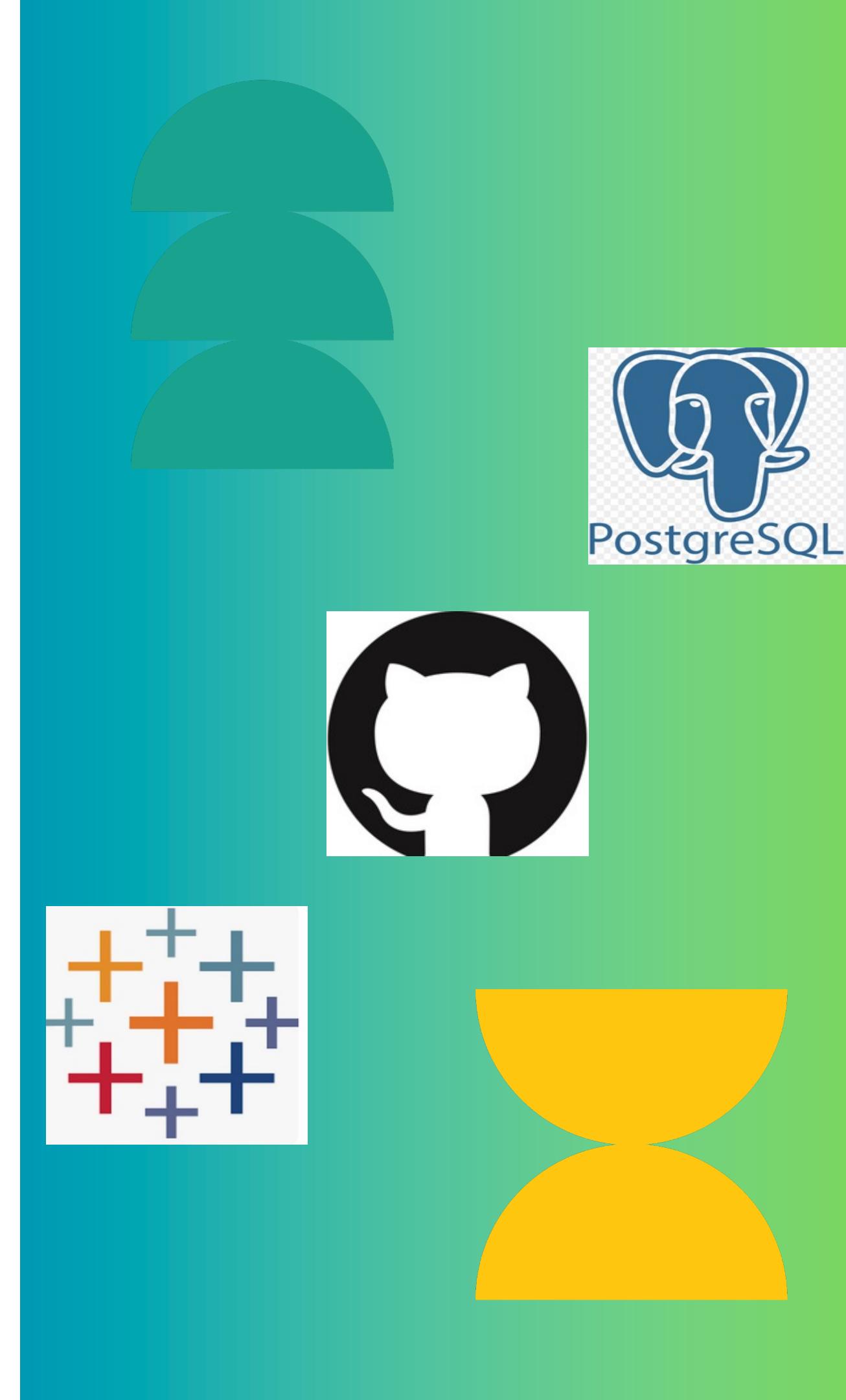


# 03. Rockbuster Stealth LLC

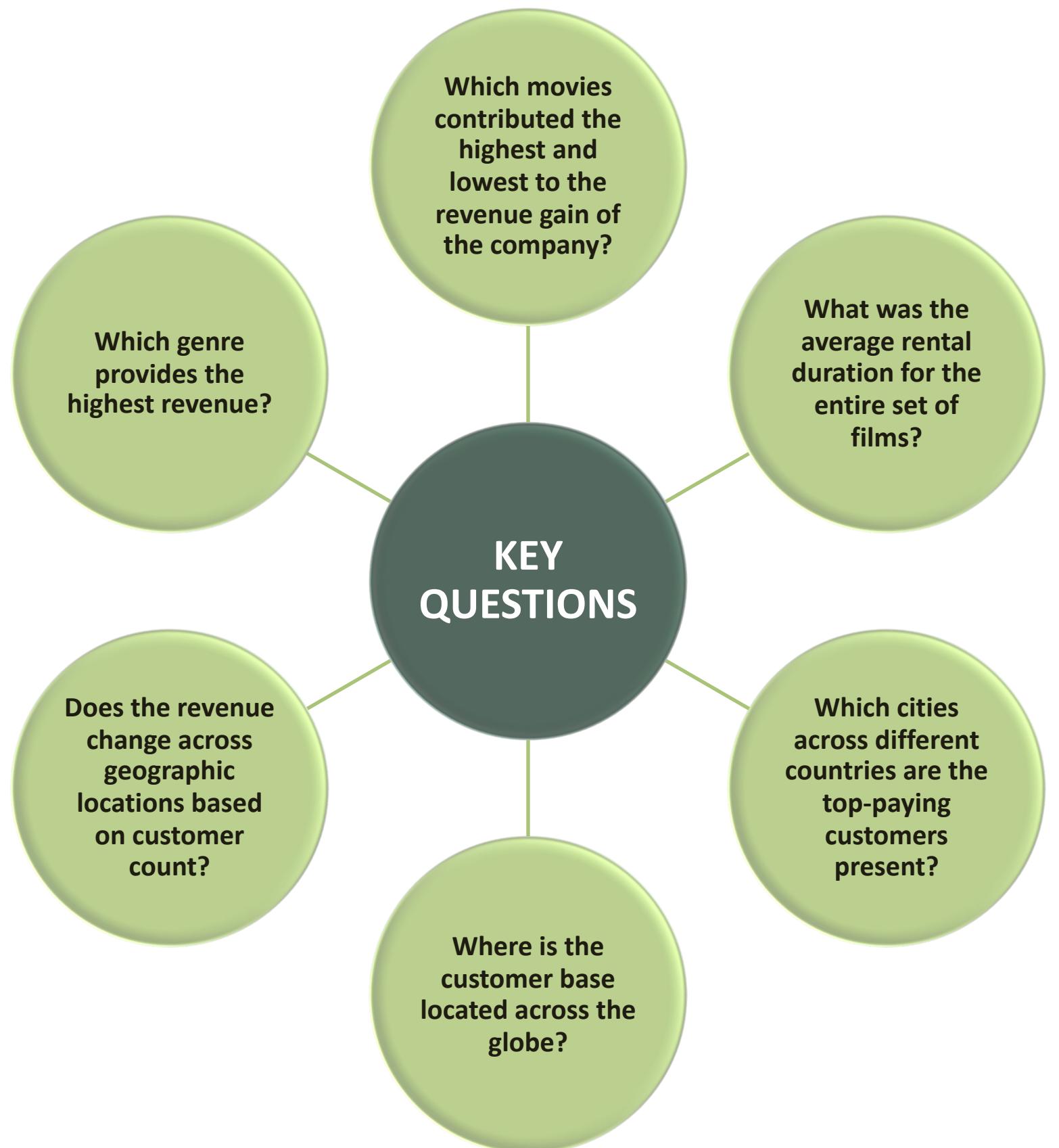
Understanding marketing insights to launch an online video rental platform

## Objectives

Rockbuster Stealth LLC is a movie rental company with stores all around the globe that want to launch an online movie rental service platform. Prior to the launch, there is requirement to analyse the various market needs through data backed insights.



## 03. Rockbuster Stealth LLC



## ■ DATA

COUNT OF CUSTOMERS 599	COUNT OF COUNTRIES 109	Avg. Rental Rate \$2.98	COUNT OF FILMS 1000	COUNT OF LANGUAGES 6 (English, Italian, Japanese, Mandarin, French, German)
COUNT OF FILMS IN ENGLISH 1000	COMMON AGE RATING PG-13	COUNT OF GENRES 17	AVG. FILM LENGTH 115 min.	

## ■ TOOLS

Excel: PostgreSQL

Visualisation: [Tableau](#)

Showcased in: [GitHub](#)

Include link to data

## ■ SKILLS

Relational Databases

Creating a data dictionary

Database querying & Subqueries

Filtering, cleaning and analysing data

Joining Tables

Common table expressions (CTEs)

Storytelling

Creating visual dashboards

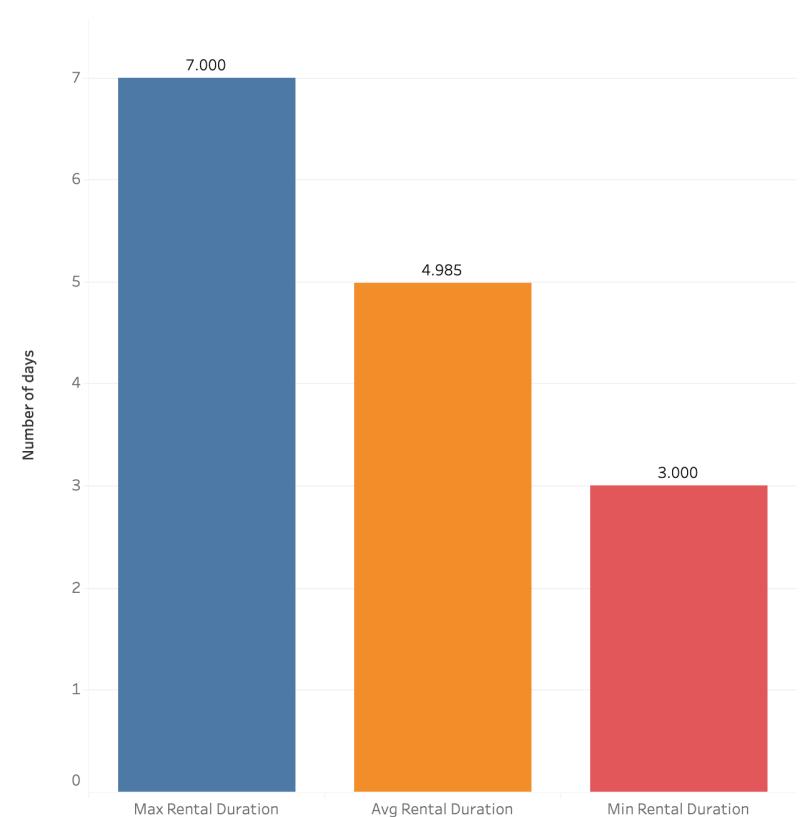
The maximum rental duration of films is 7 days and the minimum duration is 3 days.

Top 3 Countries with the highest customer base are India, China and the USA

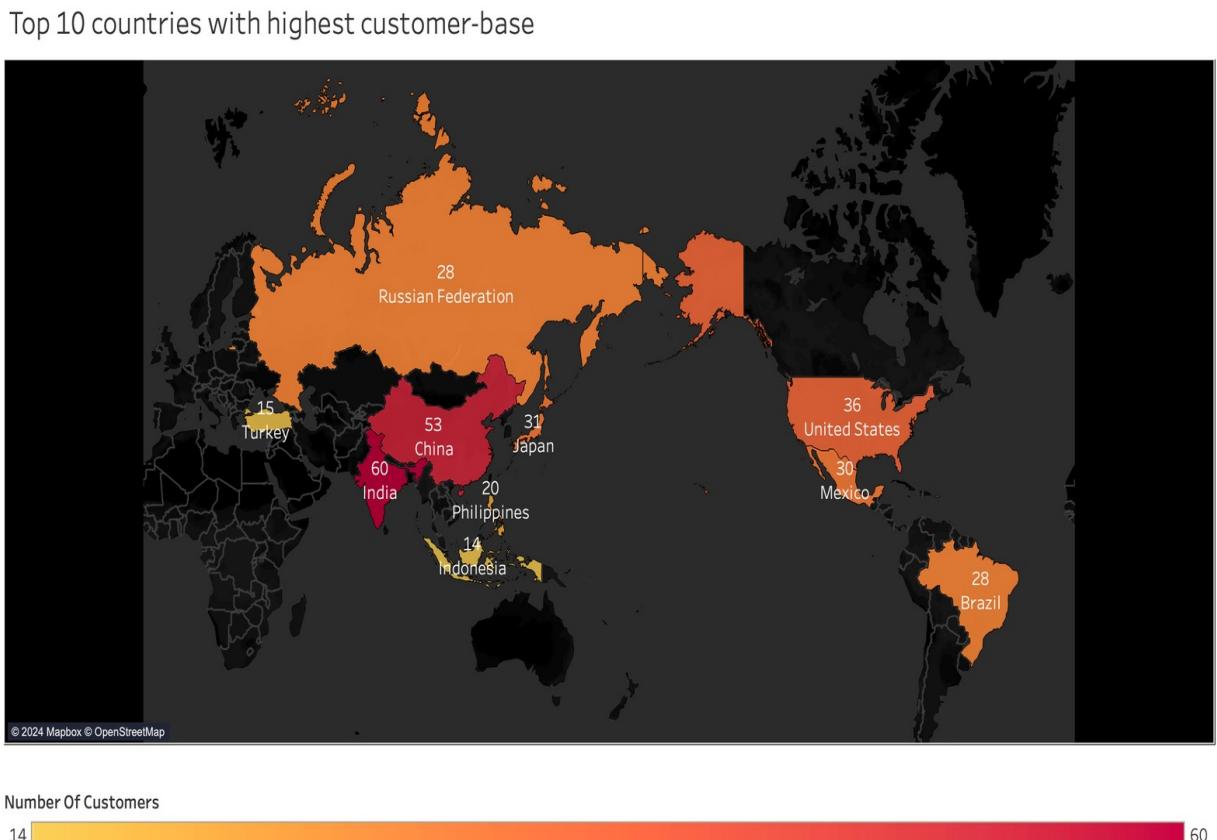
The sports genre contributes to the highest revenue of \$4,892.19 per annum

The top 5 customers from the top 10 cities who contribute the highest in terms of cost are from Turkey, Brazil and Indonesia.

### RENTAL DURATION OF THE FILMS



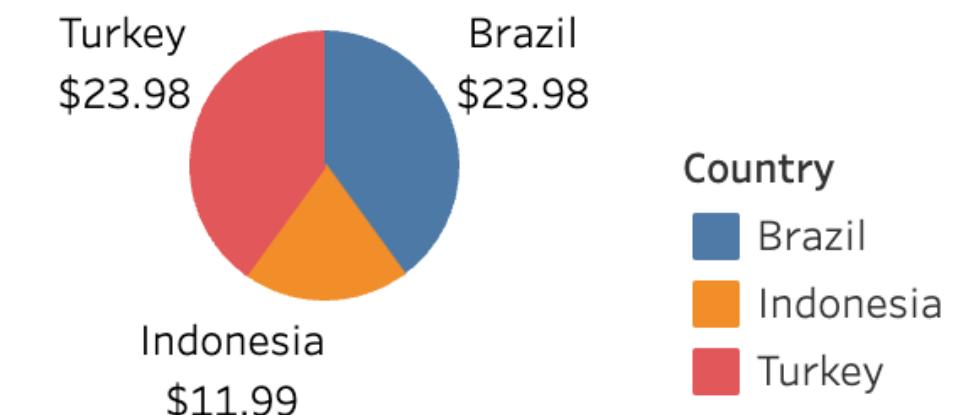
### CUSTOMER BASE ACROSS THE GLOBE



### GENRE CONTRIBUTING TO HIGHEST REVENUE



### TOP 5 CUSTOMERS FROM TOP 10 CITIES WITH HIGHEST CONTRIBUTION



## RECOMMENDATIONS

### LANGUAGE AND GEOGRAPHY

Based on the geographic locations of the customers, since a majority are located in non-English speaking countries, films from other languages must be brought into the sales market

### ADVERTISING STRATEGIES

Since, most of the customers are located in INDIA, CHINA and the USA, more advertising must be done in these locations to increase the customer base

### GENRE BASED REVENUE

The most bought genre of movies are SPORTS, SCI-FI and ANIMATION, thereby focusing on bringing more films in these genres would increase the revenue

### RENTAL DURATION

The average rental duration for a film is about 5 DAYS and an increase in the rental duration may increase the customer interest

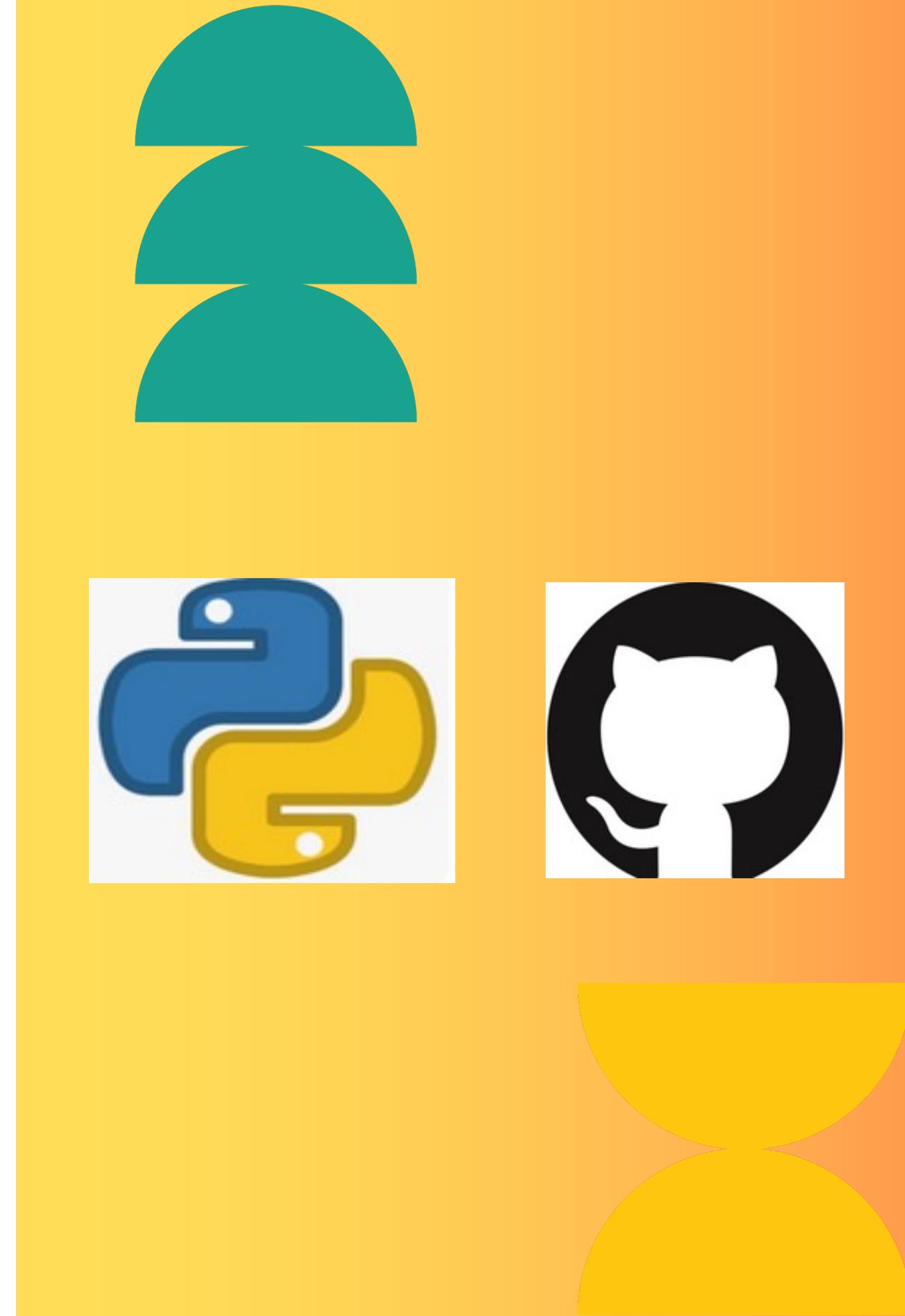


# 04. InstaCart

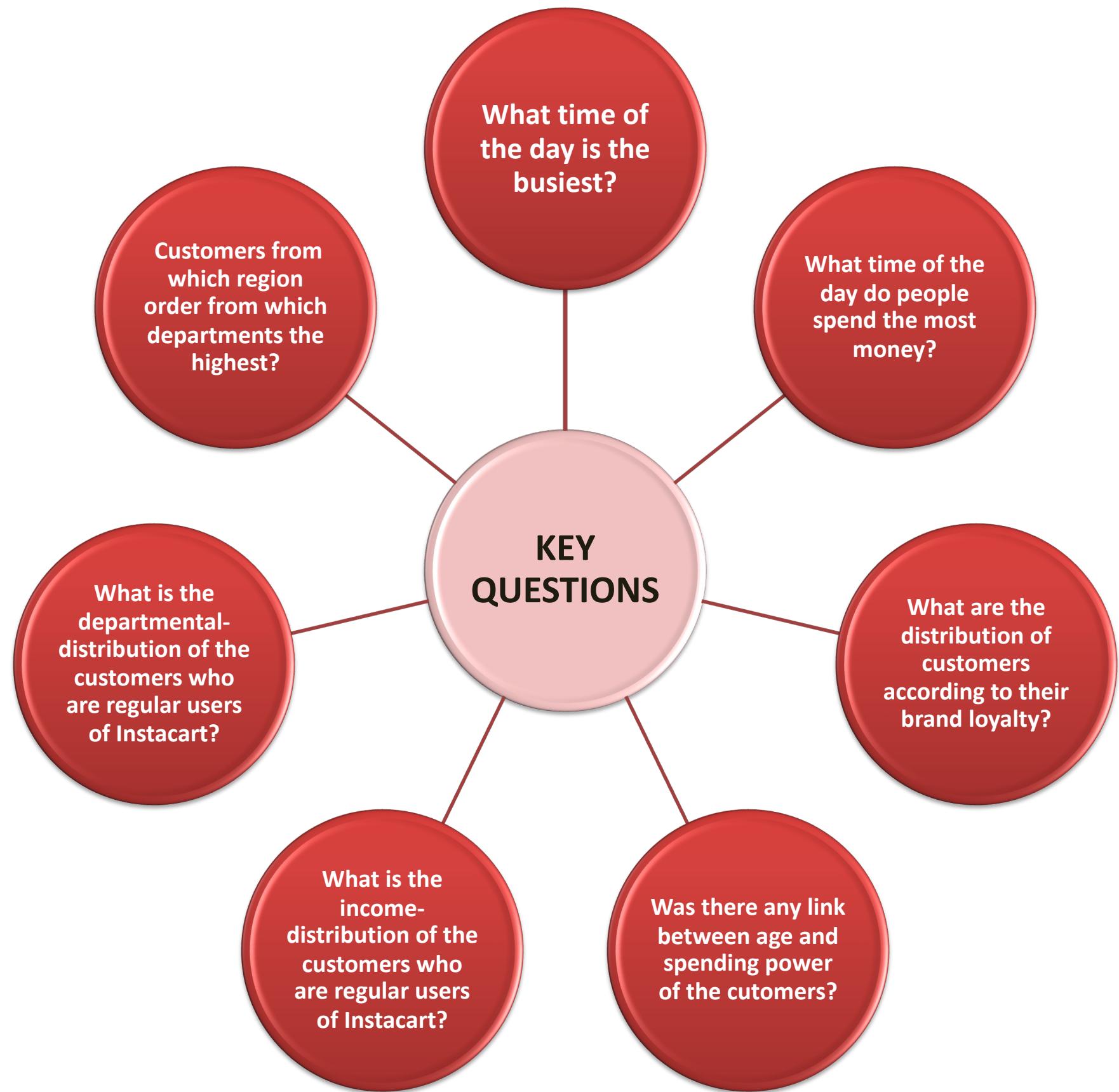
Understanding data insights to boost sales and enhance marketing strategies

## Objectives

Instacart, a leading online grocery store, aimed to increase sales by adopting a more strategic approach to customer targeting and improving segmentation through the analysis of historical data. This would enable customers to order groceries more effectively via their app. The goal of this project was to perform an in-depth analysis of Instacart's sales data to refine their marketing strategy, emphasizing precise customer targeting and effective segmentation to boost sales growth.



## 04. InstaCart



## ■ DATA

### Information provided:

**Customers:** Analyzed to identify purchasing patterns and loyalty.

**Orders:** Examined to determine peak times and spending habits.

**Products:** Categorized to understand popularity and sales impact.

**Departments:** Analyzed for sales volume per department.

## ■ TOOLS [\(GitHub\)](#)

Jupyter Notebook (Python)

Pandas, Numpy: Data analysis

Visualisation: Seaborn, Matplotlib, Scipy

## ■ SKILLS

Population flow

Consistency checks

Data Wrangling

Data Merging

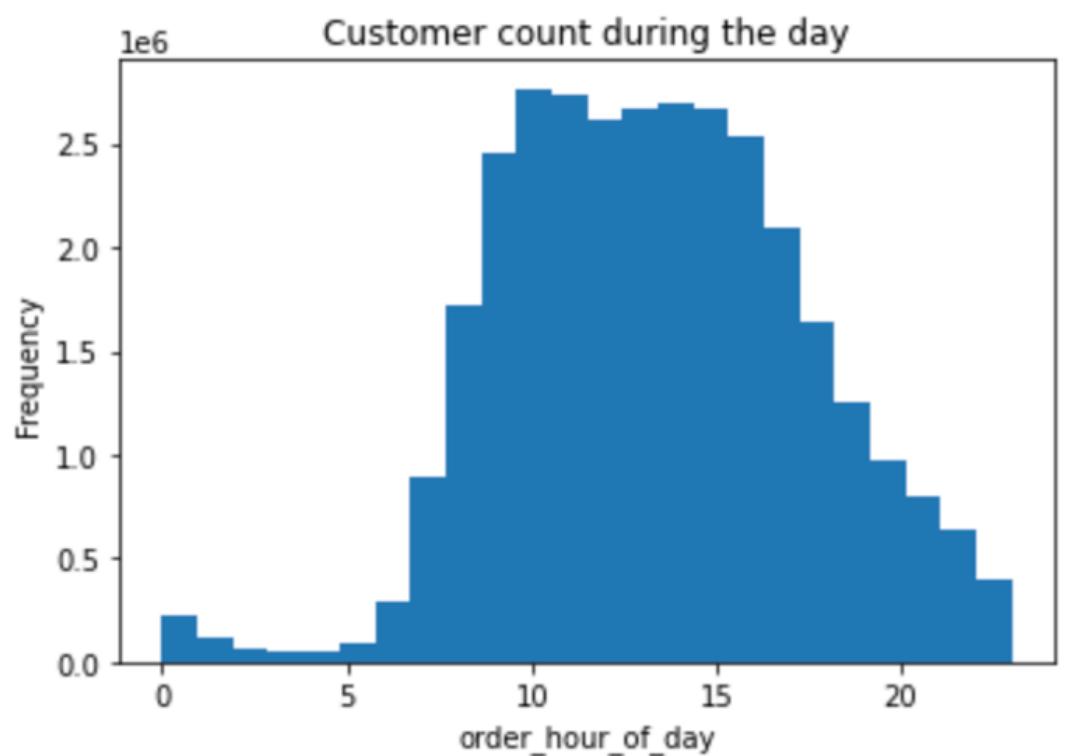
Column Derivations

Creating Pivot tables

Visualising results using Python

## 04. InstaCart

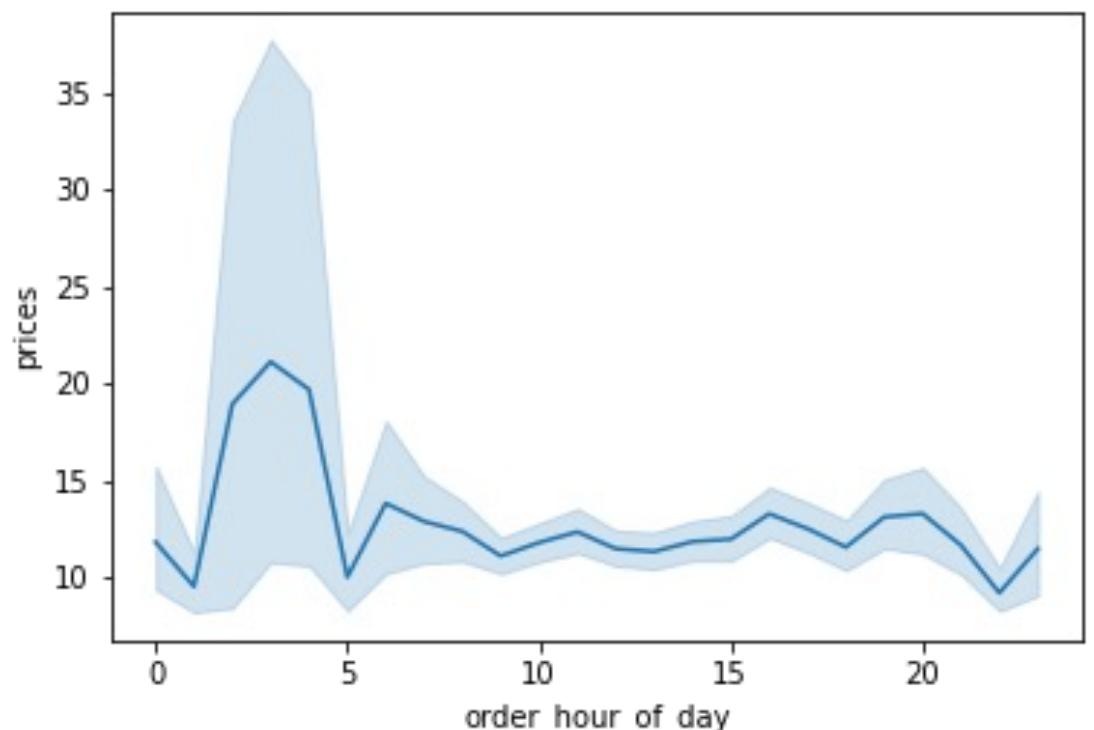
### CUSTOMER ACTIVITY AND REVENUE



The highest frequency of orders is at 10:00 am and the lowest order frequency is at 03:00 am.

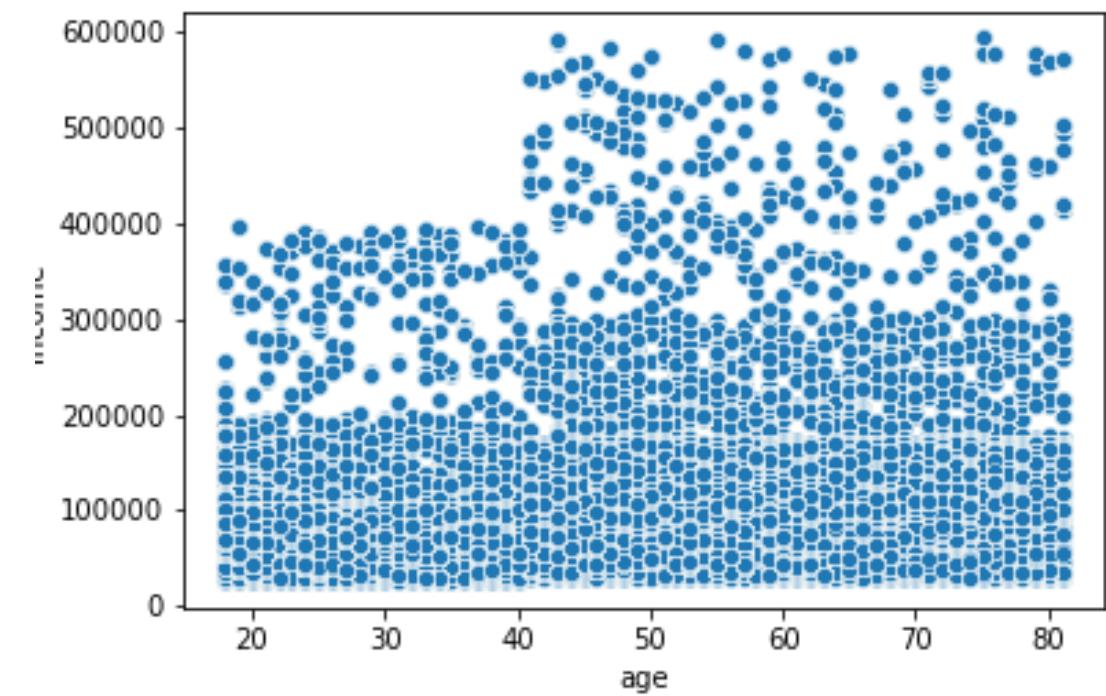
### ANALYSIS

### ORDERING PATTERN



The average is highest during early-morning orders, peaking around 4am then the average stays almost the same.

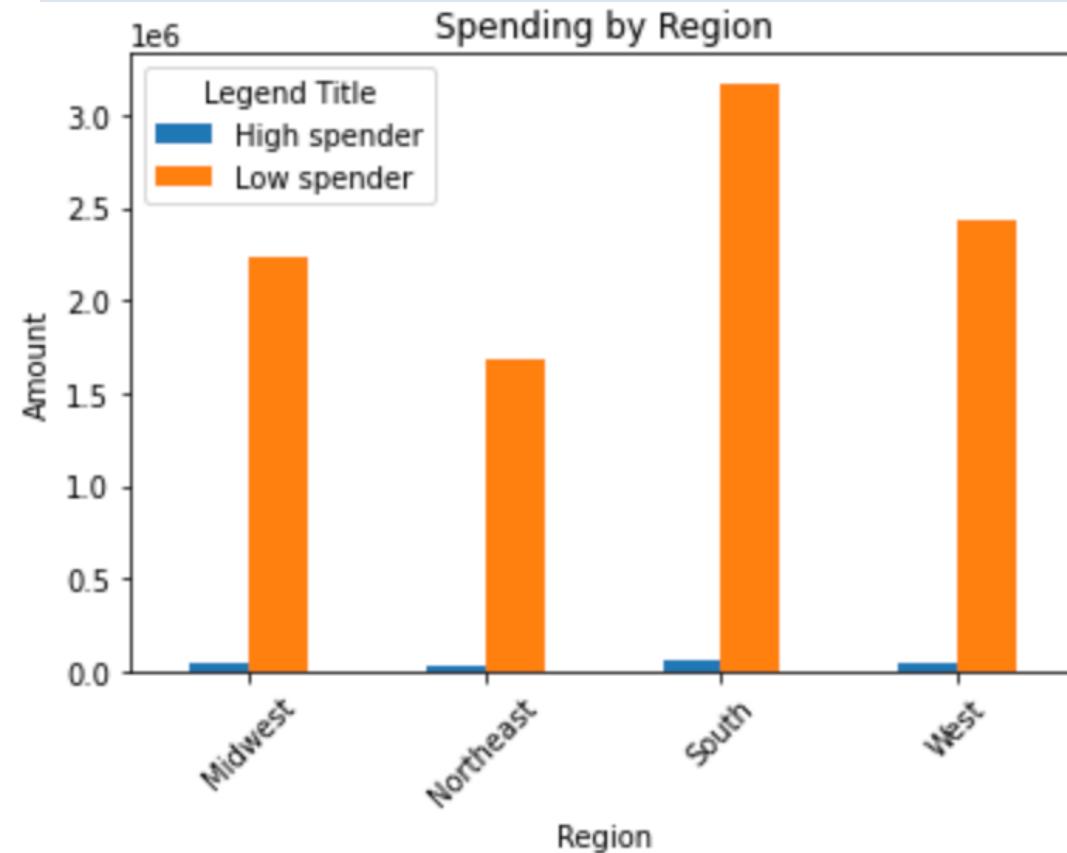
### CUSTOMER PROFILE



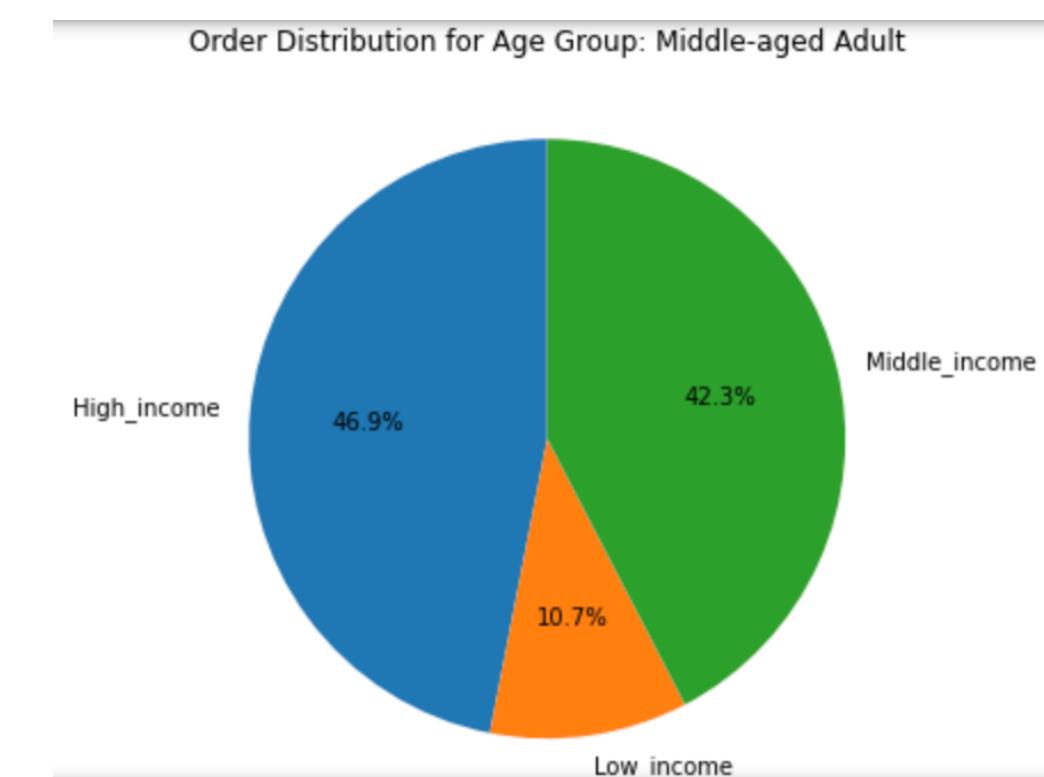
The scatterplot shows that a significant portion of the client base, spanning all age groups, earns up to 200,000, with customers 40 years of age and above earning up to 600,000. Thus, due to their higher income, customers 40 years of age and older may spend more.

## 04. InstaCart

### CUSTOMER ACTIVITY AND REVENUE



### ORDERING PATTERN



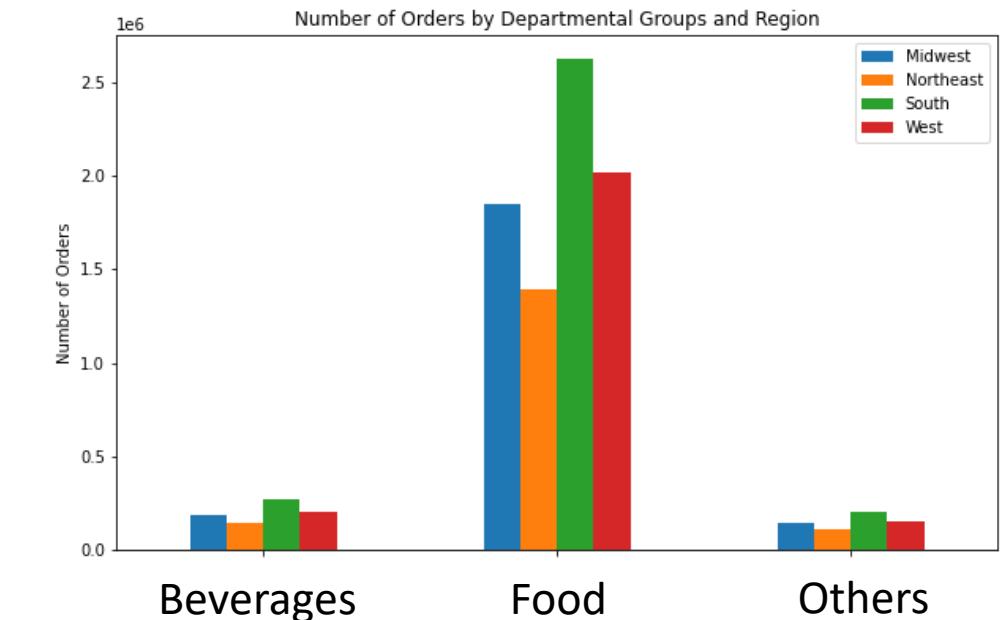
The highest frequency of spenders are from the southern states of the US. Here If avg price < 10 then low spender. If avg price  $\geq 10$  then high spender.

## ANALYSIS

It can be clearly indicated that in the middle-age group, 47% of the total middle-aged customers are from the \*high-income group while 42.3% are from \*middle-income groups.

\*If 'income'  $\leq 50000$  = 'Low\_income'; If 'income' is between 50000 to 100000 = 'Middle\_income'; If 'income'  $\geq 100000$  = 'High\_income'

### DEPARTMENTS AND ORDERS

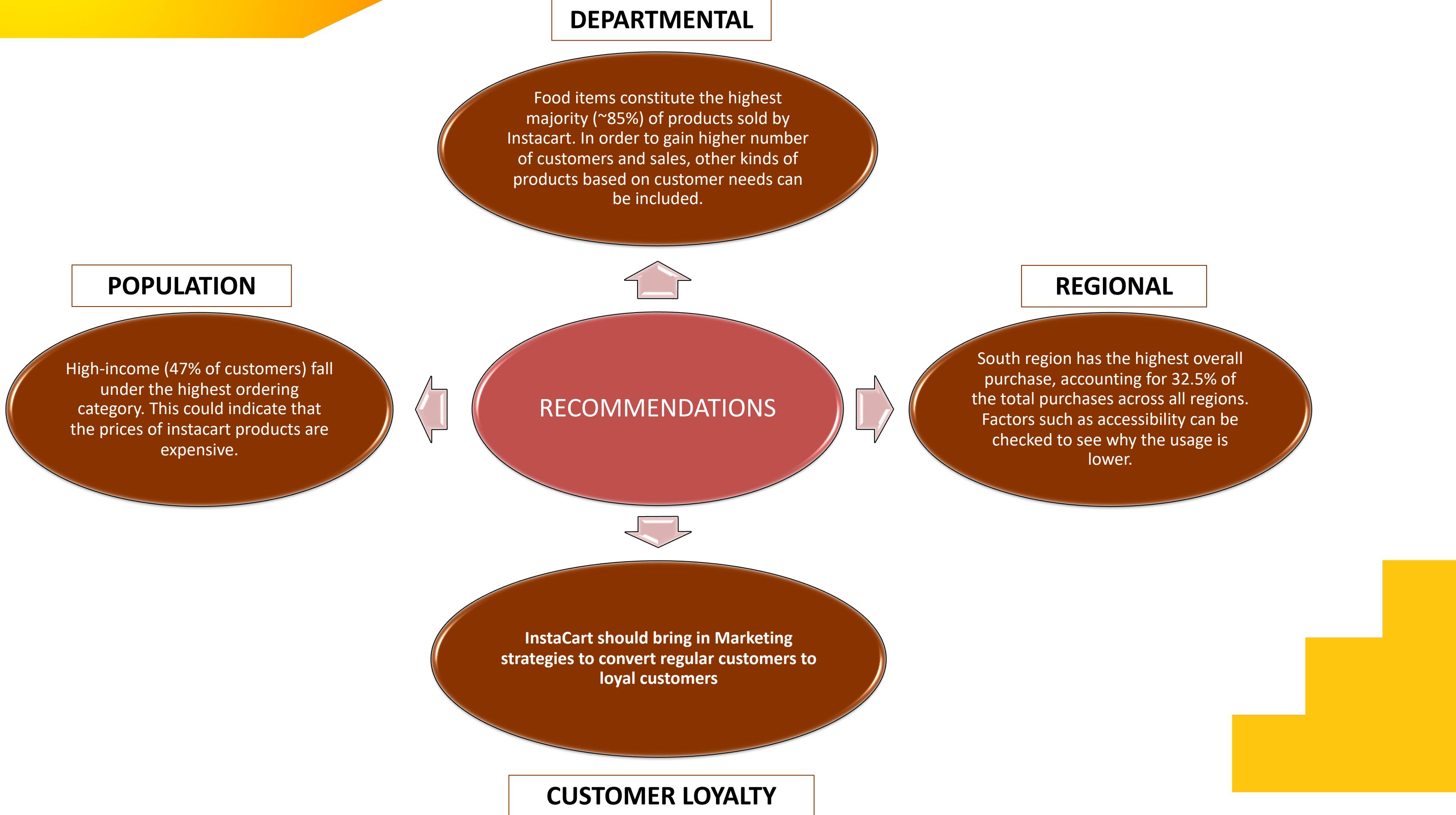


In each region, the food\_items department is the leading contributor, with the following percentages:

- Midwest: 84.34%
- Northeast: 84.35%
- South: 84.67%
- West: 84.71%

These percentages highlight the dominance of the food\_items department in every region.

## 04. InstaCart

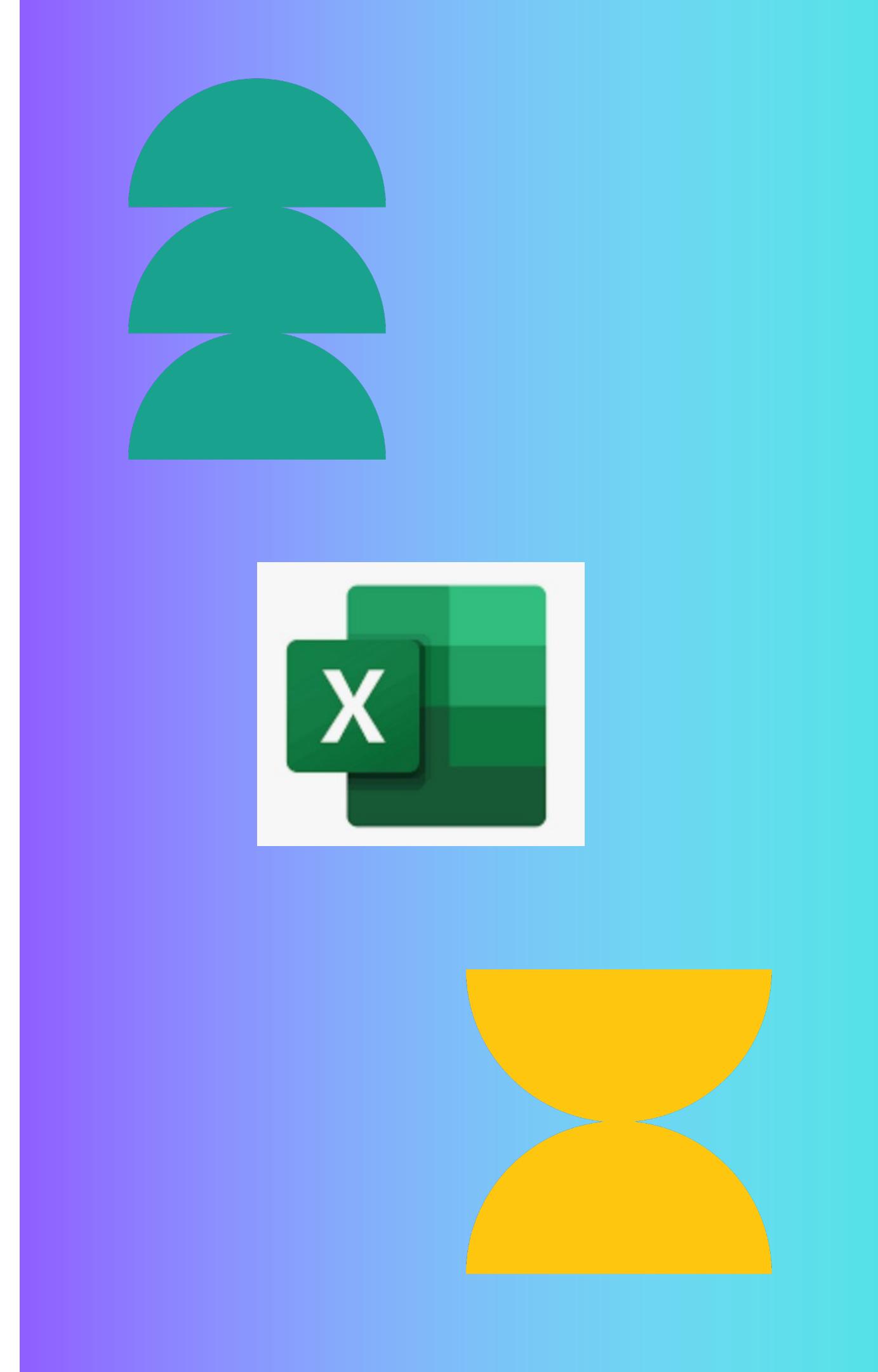


# 05. Global Bank (Pig E.)

A project to help a bank in anti-money laundering activities

## Objectives

Supported a global bank's anti-money laundering compliance department by providing data-backed insights. This analysis aimed to identify factors contributing to customer attrition, with the ultimate goal of developing robust strategies to enhance customer retention.



## KEY QUESTION

**What are the key risk factors that indicate a customer is likely to churn, and how can identifying these factors help in developing targeted strategies to enhance customer retention?**

- **DATA**

**Source:** CareerFoundry

- **TOOLS**

**Excel:** Data analysis and Visualisation

- **SKILLS**

Data ethics

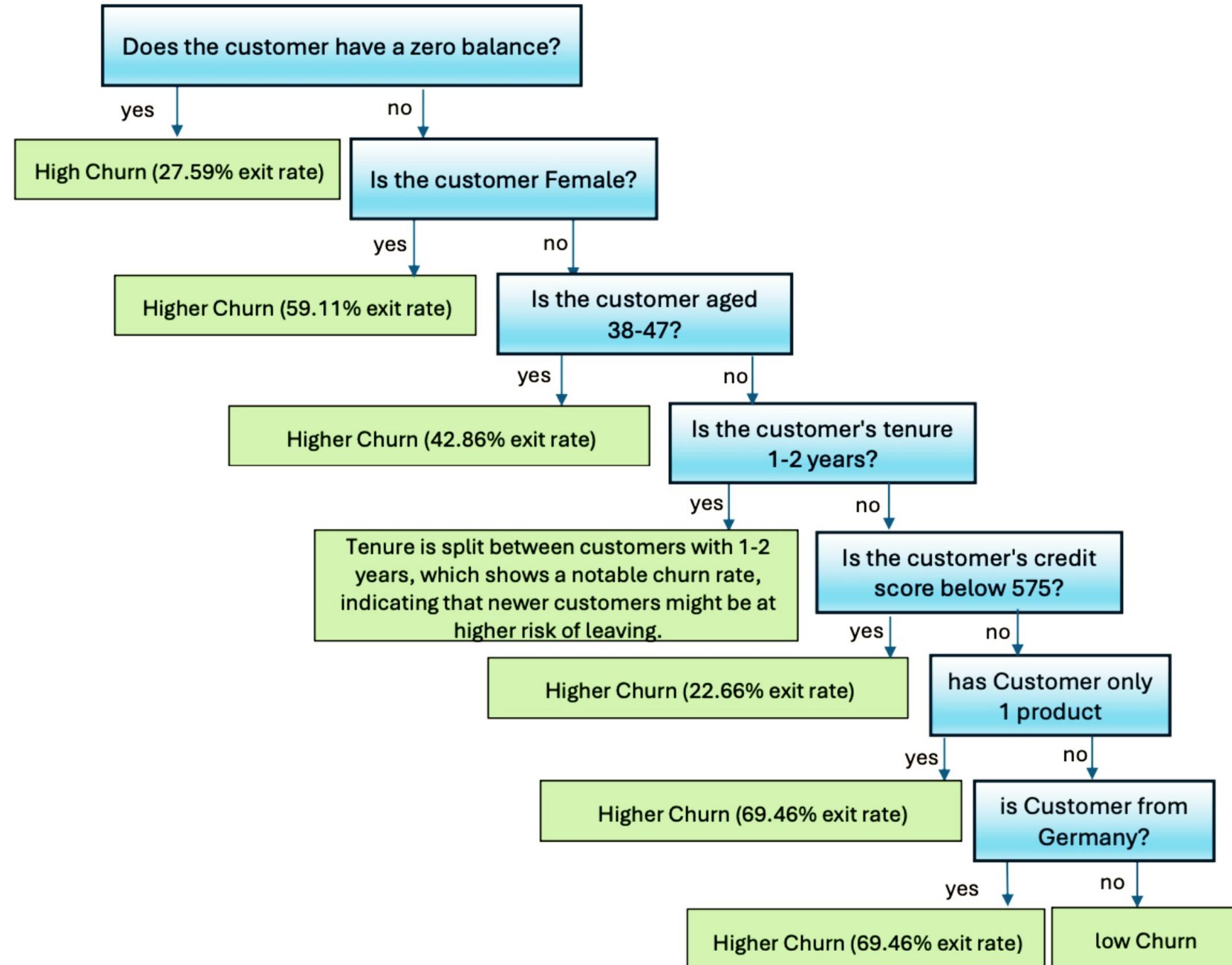
Big Data Analysis

Data mining

Predictive analysis

Time series forecasting

# ANALYSIS



**1. Zero Balance:** Customers with a zero balance have a high churn rate (27.59%). Encourage account activity to reduce churn.

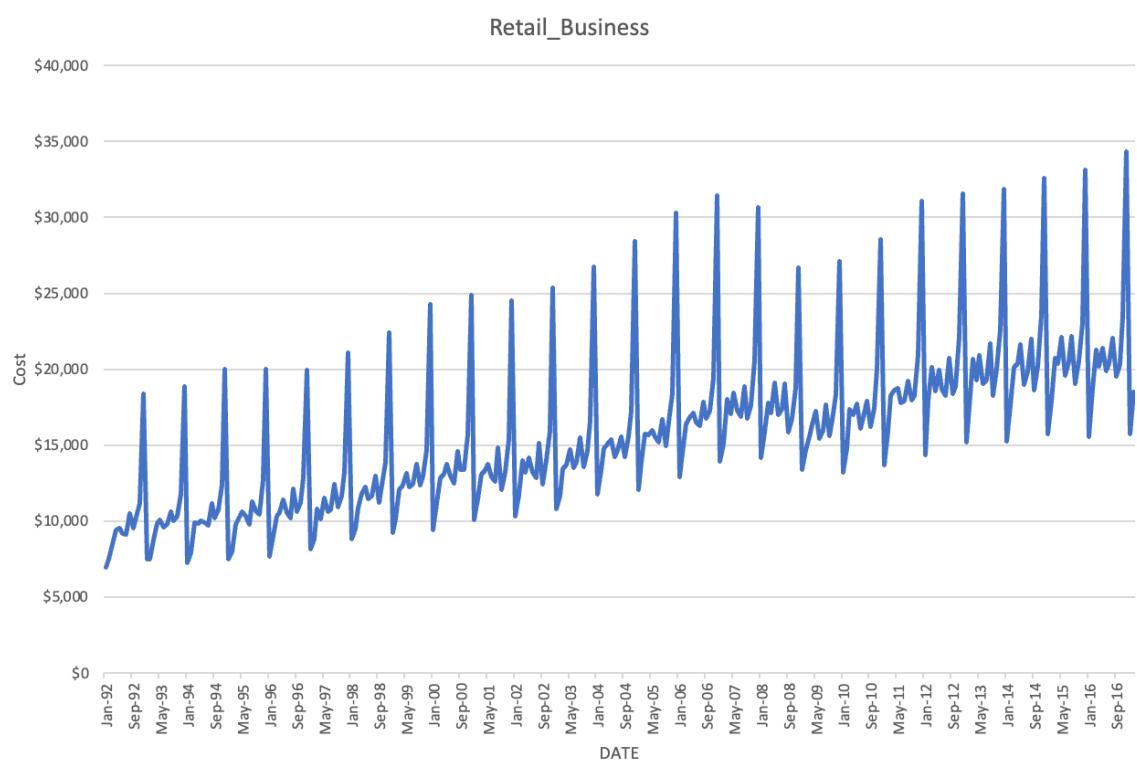
**2. Gender Factor:** Female customers are more likely to churn (59.11%). Consider tailored retention strategies for this group.

**3. Age & Tenure:** Middle-aged customers (38-47) and those with 1-2 years of tenure are at higher churn risk. Implement loyalty programs.

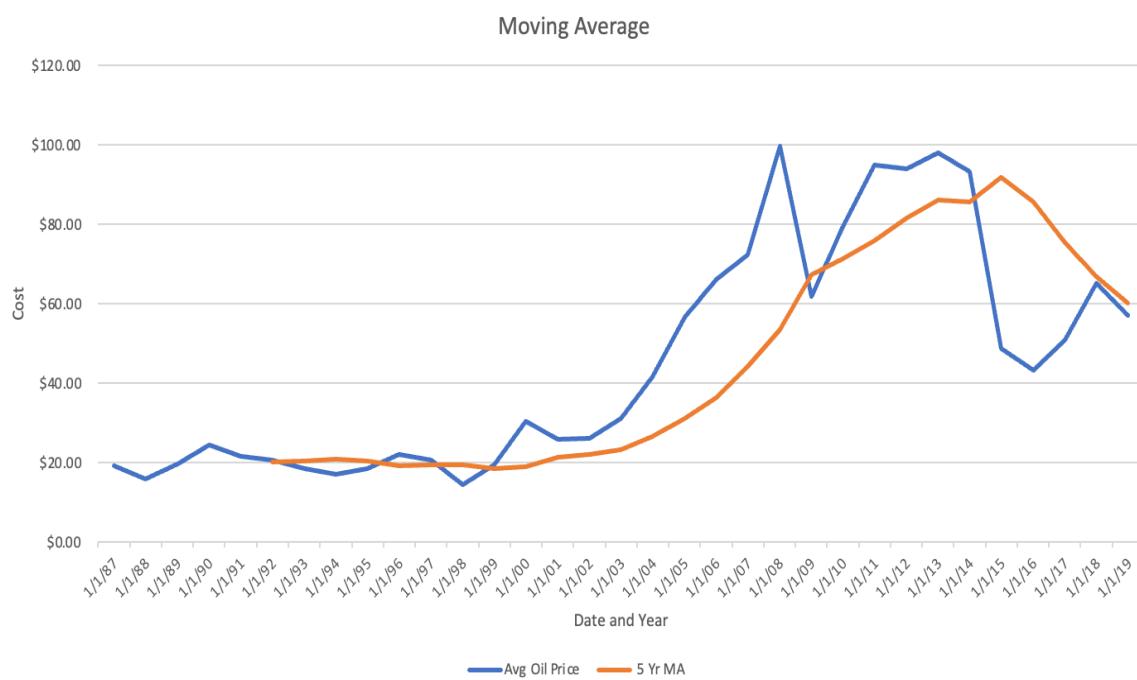
**4. Credit & Engagement:** Customers with low credit scores and only one product have a very high churn risk (69.46%). Focus on cross-selling and credit improvement initiatives.

## 05. A Global Bank (Pig E.)

### TIME SERIES FORECASTING (SEASONAL)



### SIMPLE MOVING AVERAGE



### RECOMMENDATIONS

#### Personalized Engagement

Tailor strategies and loyalty programs for younger clients ( $\text{Age} \leq 42.5$ ).

#### Financial Incentives

Reward higher account balances and promote savings programs.

#### Proactive Retention

Use predictive analytics to identify at-risk clients and offer targeted retention incentives.

# 06. World Economic Indicators

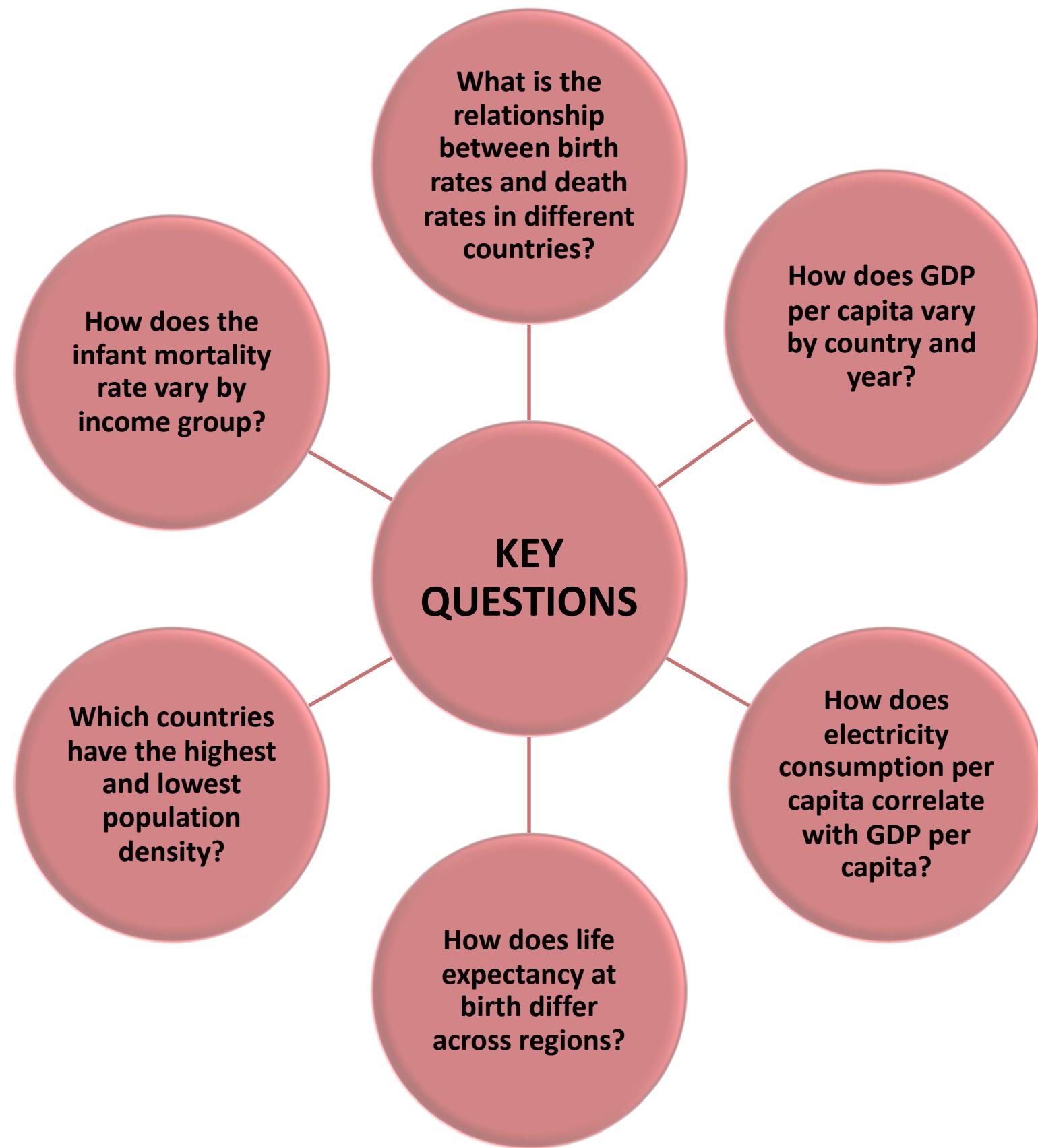
Understanding global trends and patterns of different regions across the world

## Objectives

To conduct meaningful analyses of global trends and the relationships between economic performance and well-being, incorporating factors such as GDP per capita, life expectancy, and the Human Development Index (HDI) is essential.



## 06. World Economic Indicators



### ■ DATA

Source: [Maven Analytics](#)

Period: 1960-2018

### ■ TOOLS

Python, Tableau, Excel

### ■ SKILLS

Cleaning and data grouping

Descriptive Statistics

Exploratory Data Analysis (EDA)

Consistency checks

Data Wrangling

Data Merging

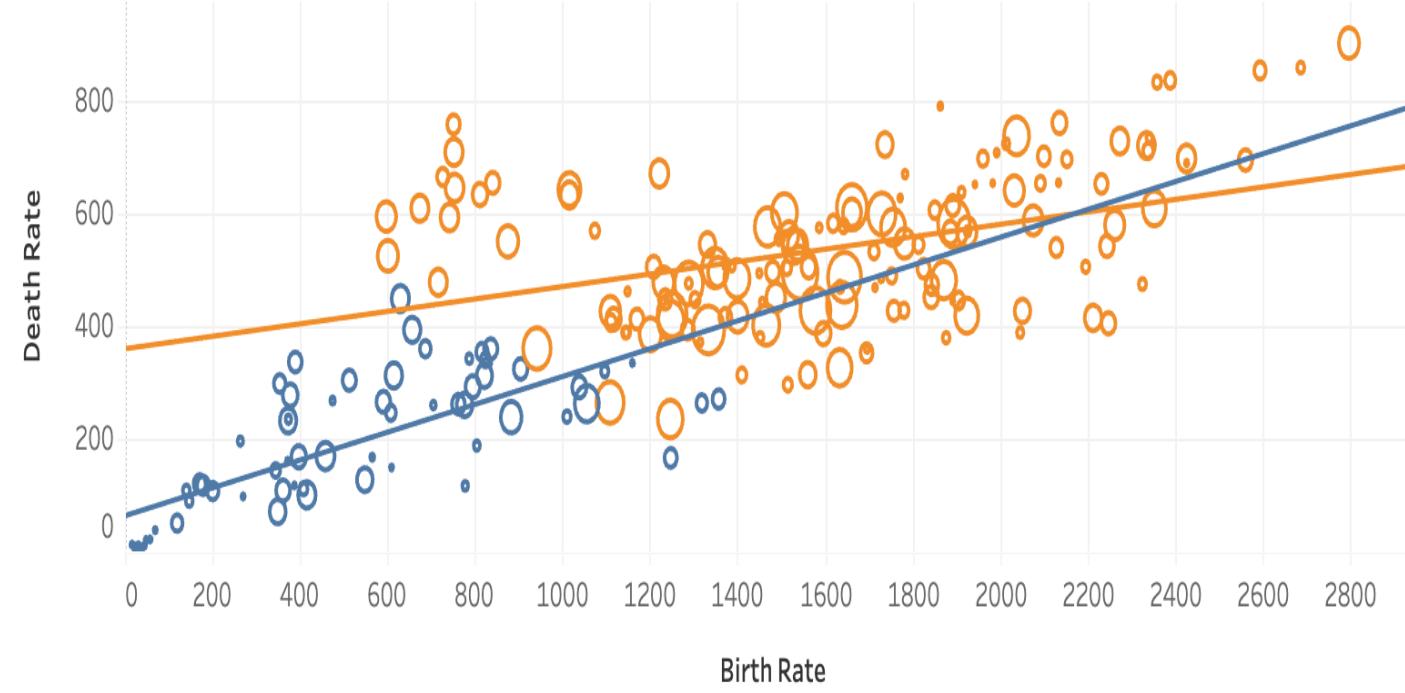
Column Derivations

Creating Pivot tables

Visualising results using Python

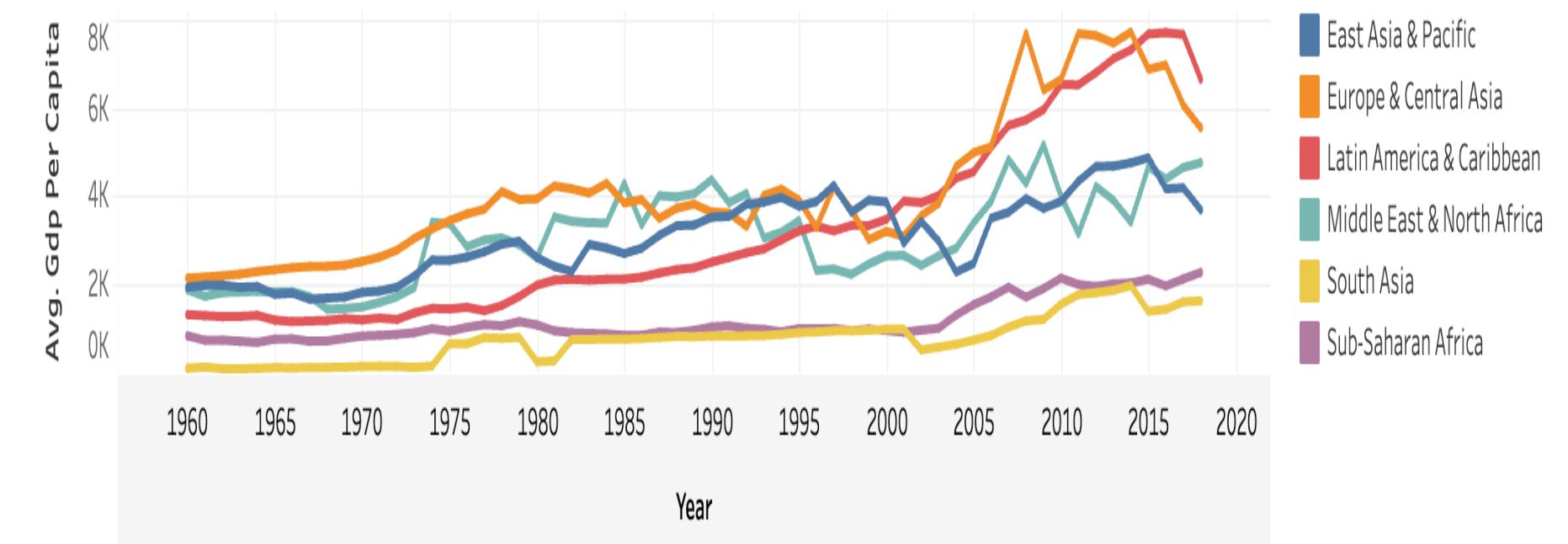
# ANALYSIS

## BIRTH RATE VS DEATH RATE IN DIFFERENT COUNTRIES



Higher birth rates tend to correlate with higher death rates, but this relationship varies between the two groups, with the blue group showing a stronger relationship between birth and death rates than the orange group.

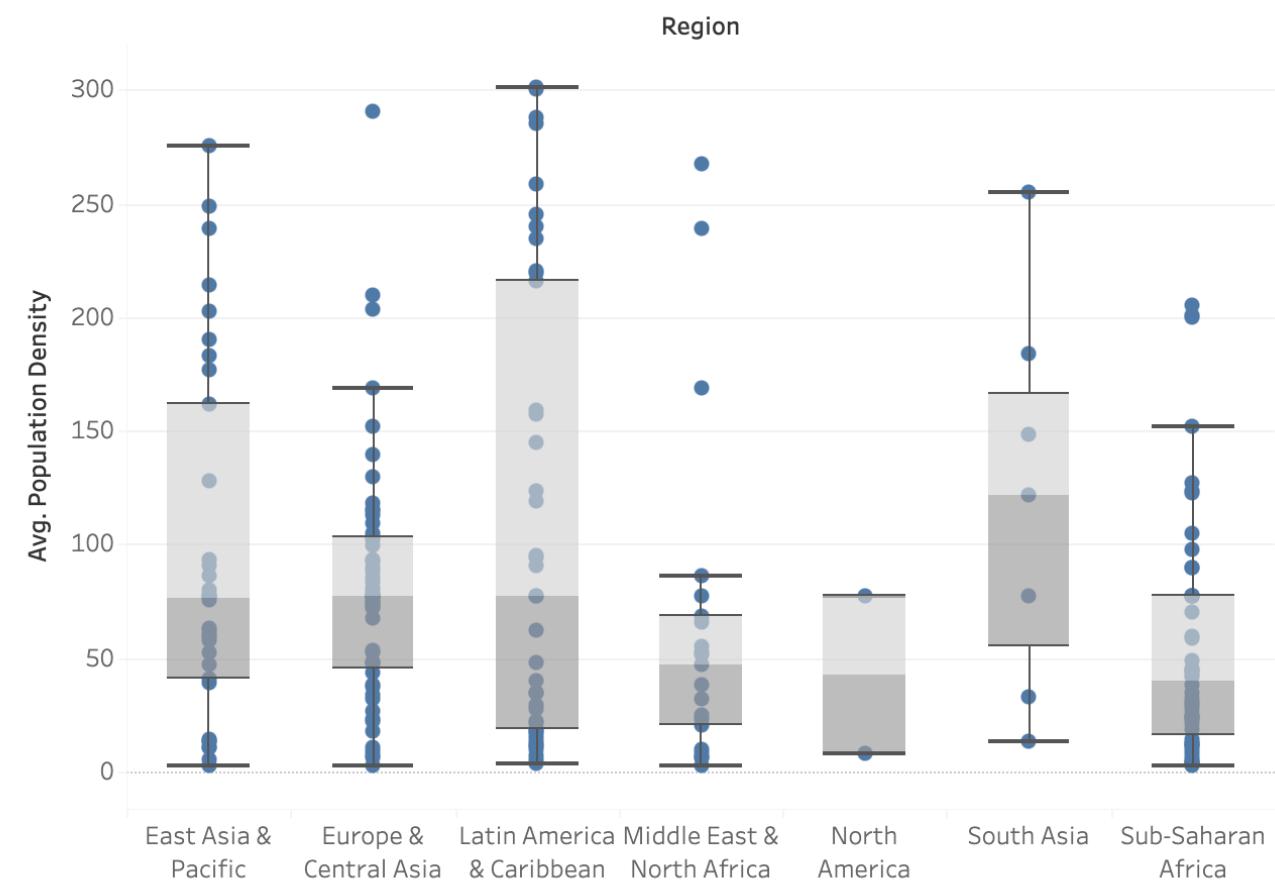
## GDP per capital across regions per year



- Europe & Central Asia:** Highest GDP per capita, with steady growth.
- East Asia & Pacific:** Strong rise, especially post-2000.
- Latin America & Caribbean:** Gradual growth.
- Middle East & North Africa:** Volatile but moderate growth.
- South Asia:** Low, with gradual improvement after the 1990s.
- Sub-Saharan Africa:** Lowest GDP per capita, slow growth.

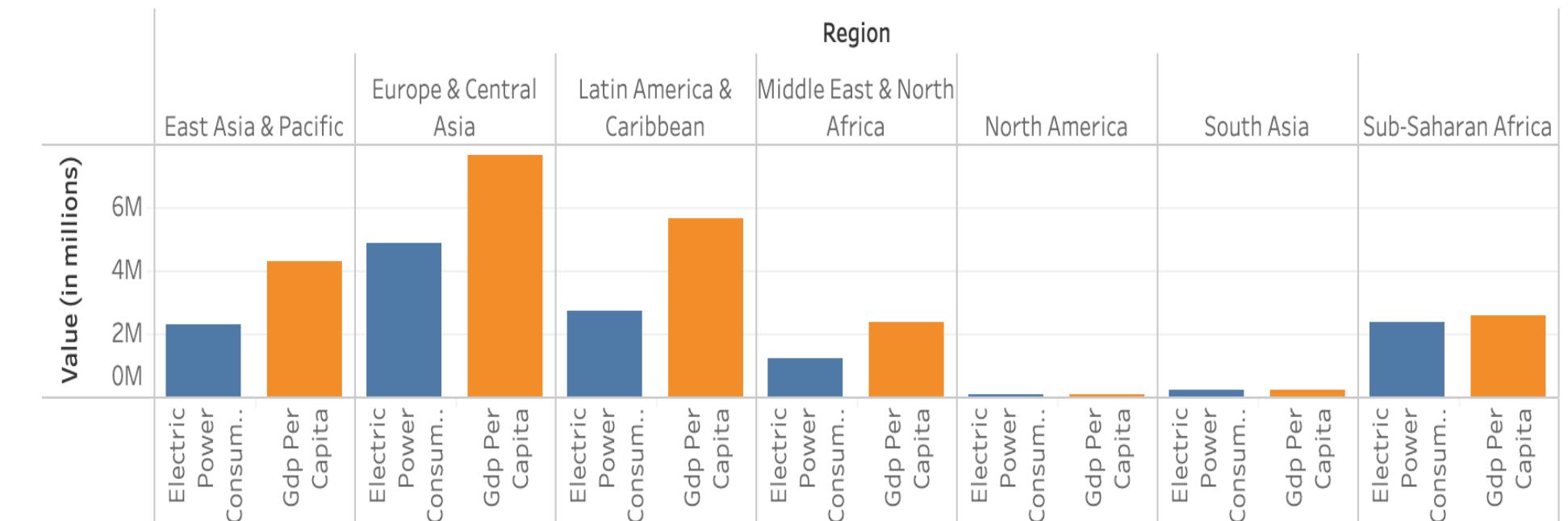
## ANALYSIS

## REGION VS POPULATION DENSITIES



The plot shows population density variability across regions. **Latin America & Caribbean** and **South Asia** have the highest spread with outliers, while **North America** and **Middle East & North Africa** show lower, more consistent densities. **South Asia** has a higher median, and **North America** has one of the lowest.

## ELECTRIC POWER CONSUMPTION VS GDP PER CAPITA ACROSS REGIONS



**Europe & Central Asia** and **East Asia & Pacific** show high values for both indicators.

**Latin America & Caribbean** and **Middle East & North Africa** have moderate GDP per capita with lower power consumption.

**North America** shows very low values for both metrics.

**South Asia** and **Sub-Saharan Africa** have the lowest overall values, reflecting lower economic and energy consumption levels.

# RECOMMENDATIONS

**Focus on Healthcare and Family Planning in High Birth/Death Rate Countries:** these countries may benefit from targeted healthcare improvements and family planning programs to reduce mortality rates and manage population growth effectively.

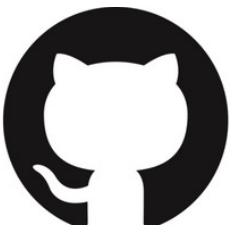
**Expand Energy Access to Boost Economic Growth in Low-GDP Regions:** Invest in electrification and renewable energy to drive economic development in Sub-Saharan Africa and South Asia.

**Improve Infrastructure and Sustainability in High Population Density Regions:** Focus on sustainable urban planning and infrastructure investments to accommodate growing populations, especially in East Asia and South Asia.



# THANK YOU

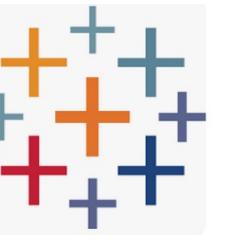
RENU BALAJI



[GitHub](#)



[LinkedIn](#)



[Tableau](#)



12balaji.renu@gmail.com

mathSpace

