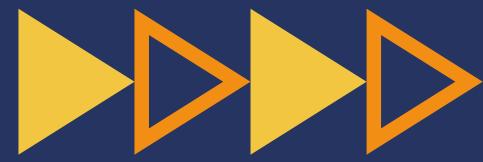




# MALVIKA MALL

## DATA ANALYST





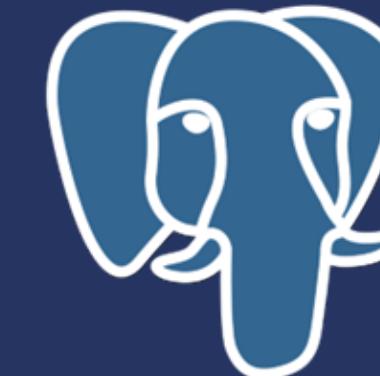
# ABOUT ME

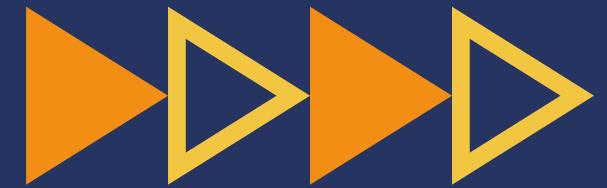
I am a data analyst specializing in legal and compliance, assisting businesses in scaling through data-driven insights. With strong analytical skills and a curious mindset, I uncover valuable connections within complex datasets. I provide tailored solutions considering risks and compliance, enabling informed business decisions.

My expertise includes data collection, cleaning, manipulation, forecasting business risks, analyzing patterns, and projecting growth. I have a track record of successfully delivering projects and offering data-driven insights in staffing distribution, ecommerce buying patterns, and market research.



# TOOLS FOR ANALYTICS & VISUALIZATION





# PORTFOLIO PROJECTS

Welcome to my data analytics portfolio! I am passionate about leveraging data to uncover valuable insights and drive informed decision-making. In this portfolio, you will find a collection of projects that showcase my skills in data analysis, visualization, and predictive modeling.

**01**

## GAMECO

Analysing Marketing Strategy for launch of new video games.

**02**

## PREPARING FOR INFLUENZA SEASON

Analysing and forecasting Staffing Distribution for upcoming flu season.

**03**

## ROCKBUSTER STEALTH LLC

Analysing market trends of an online movie rental company to stay competitive in the market.

**04**

## INSTACART

Analysing Customers Buying Pattern for an e-commerce website.

# 01. GAMECO



# PROJECT OVERVIEW



## OBJECTIVE

- GameCo seeks to analyze its marketing strategy for the launch of new video games next year.
- To evaluate and optimize the marketing approach for upcoming game releases.

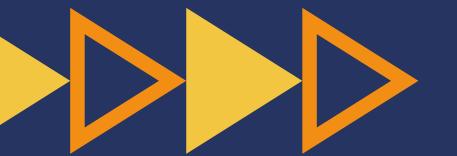
## KEY QUESTIONS

- GameCo stakeholders seek insights on game popularity, competition, and sales across markets and regions.
- Key areas of interest include popular game types, competitor analysis, game popularity trends, and regional sales data.

## DATASET

- Project data set includes historical sales of video games (over 100,000 copies) across platforms, genres, and studios.
- Data sourced from VGChartz website. Download the [video game sales data set](#).

# ANALYSIS STEPS



## 1 DATA PREPARATION

- The dataset was accessed, sorted, filtered, and cleaned using Excel.
- Pivot tables were used for data transformation, grouping, deriving new variables, and extracting insights.

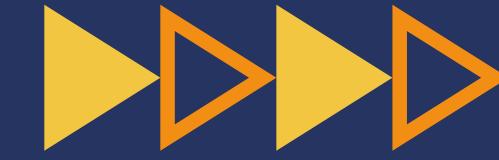
## 2 ANALYSIS

- Analyzed the business scenario and derive insights for marketing budget allocation.
- Visualizations created to understand video game sales variations and explore relationships between variables.

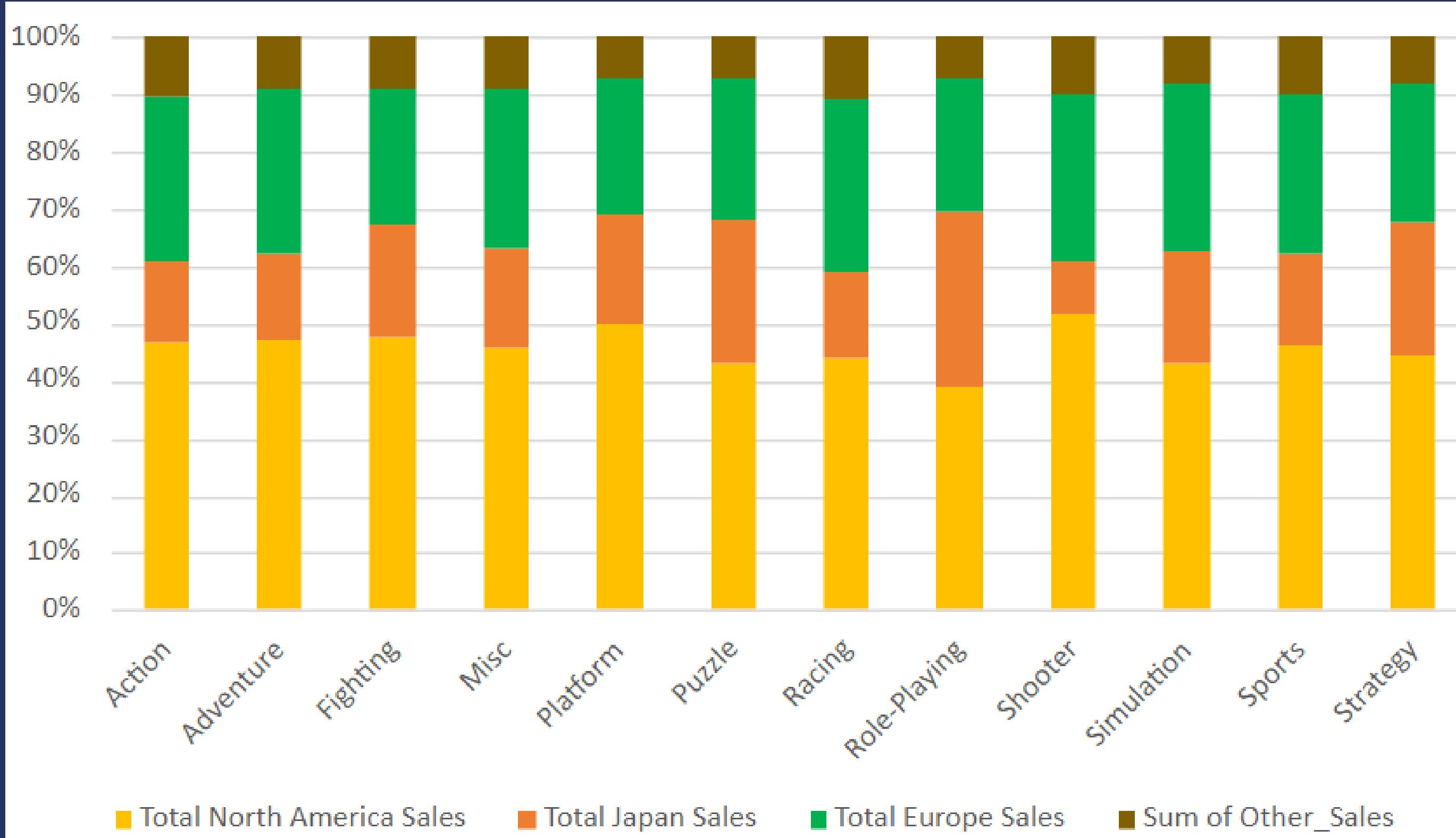
## 3 RECOMMENDATIONS

- GameCo should prioritize the North American market due to its highest sales contribution from 1980 to 2016.
- Focus on Role-Playing, Sports, Miscellaneous, and Action genres.

# VISUALIZATIONS



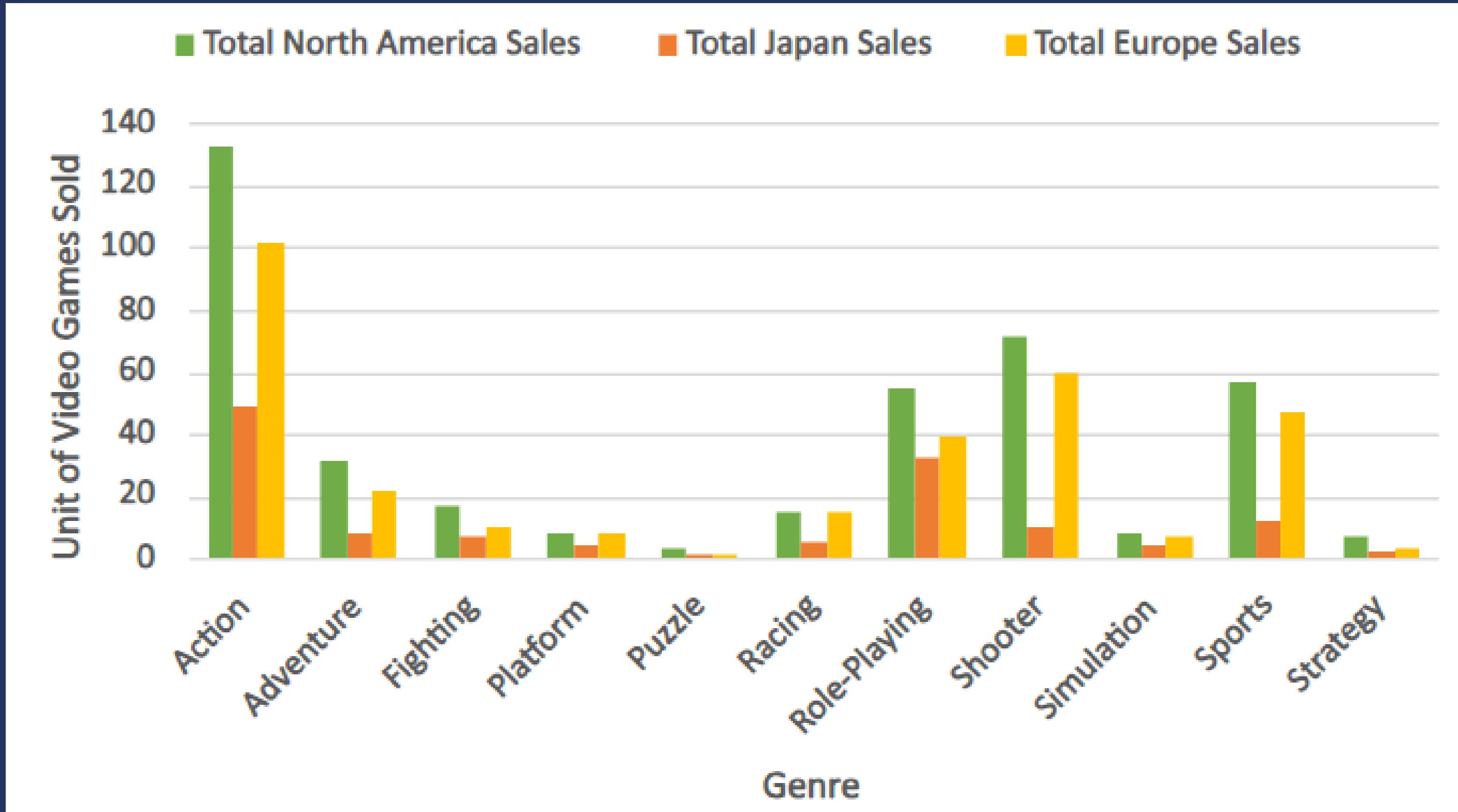
## 100% Stacked Bar Chart by Genre



Europe and Japan sales are increasing consistently however, North America sales is decreasing comparatively.

# VISUALIZATIONS >>>

## Top Performing Genres by No. of Unit Sold



Action, Shooter, Role-Playing and Sports are top revenue generating Genre across all regions.



# PROJECT DELIVERABLES



## EXCEL ANALYSIS

This File contains clean wrangled data along with visualizations.



## STAKEHOLDERS PRESENTATION

This PowerPoint presentation details an analytical story to executives with visualizations and recommendations.



## PROJECT REFLECTION

This document details all the processes and explanation for the selection of statistical technique and appropriate visualizations.



## GITHUB REPOSITORY

Contains complete project folder.

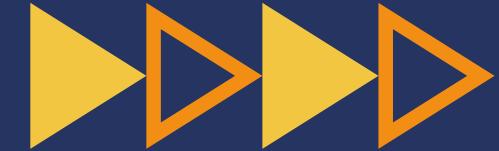


Click each deliverable to read in detail.

## 02. PREPARING FOR INFLUENZA SEASON



# PROJECT OVERVIEW



## OBJECTIVE

Analyze historical data to determine optimal timing and allocation of temporary staff to clinics and hospitals in each state during the upcoming influenza season in the United States.

## KEY QUESTIONS

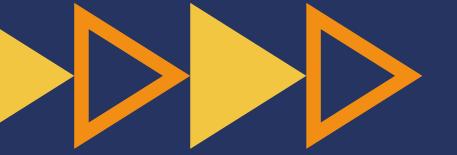
- Collect data on influenza seasonality and duration in the United States to optimize the allocation of medical personnel, with a focus on states with vulnerable populations.
- Determine timing and distribution of medical personnel and prioritize states with highly vulnerable populations.

## DATASET

Influenza deaths by geography, time, age, and gender. Source: [CDC](#)  
[Download Data Set](#)

Population data by geography. Source: US Census Bureau [Download Data Set](#)

# ANALYSIS STEPS



## 1 DATA PREPARATION

- Executed a comprehensive data research project, utilizing Excel and Tableau to address data integrity, ensure quality, and unify disparate datasets.

## 2 ANALYSIS

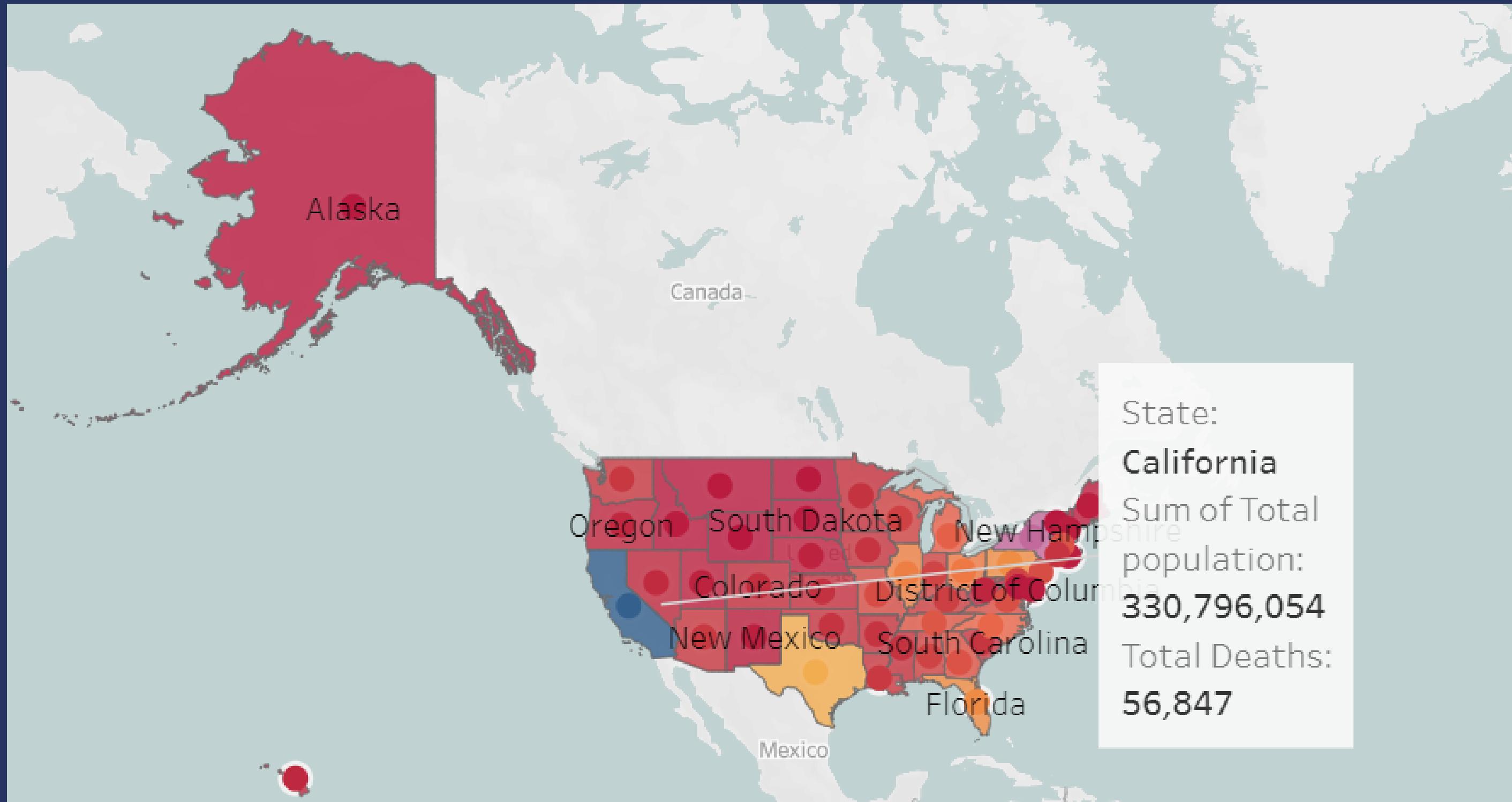
- Performed statistical analyses including variance, standard deviation, and correlation identification.
- Generated comprehensive report summarizing analysis findings with visualizations.

## 3 RECOMMENDATIONS

- California, Florida, New York, Texas & Pennsylvania are top 5 hotspots requiring additional medical staff comparatively.
- Flu activity peaks between December and January, as per CDC.

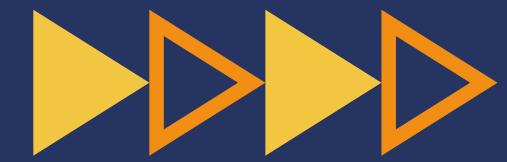
# VISUALIZATIONS >>>

## Influenza Hotspots in United States

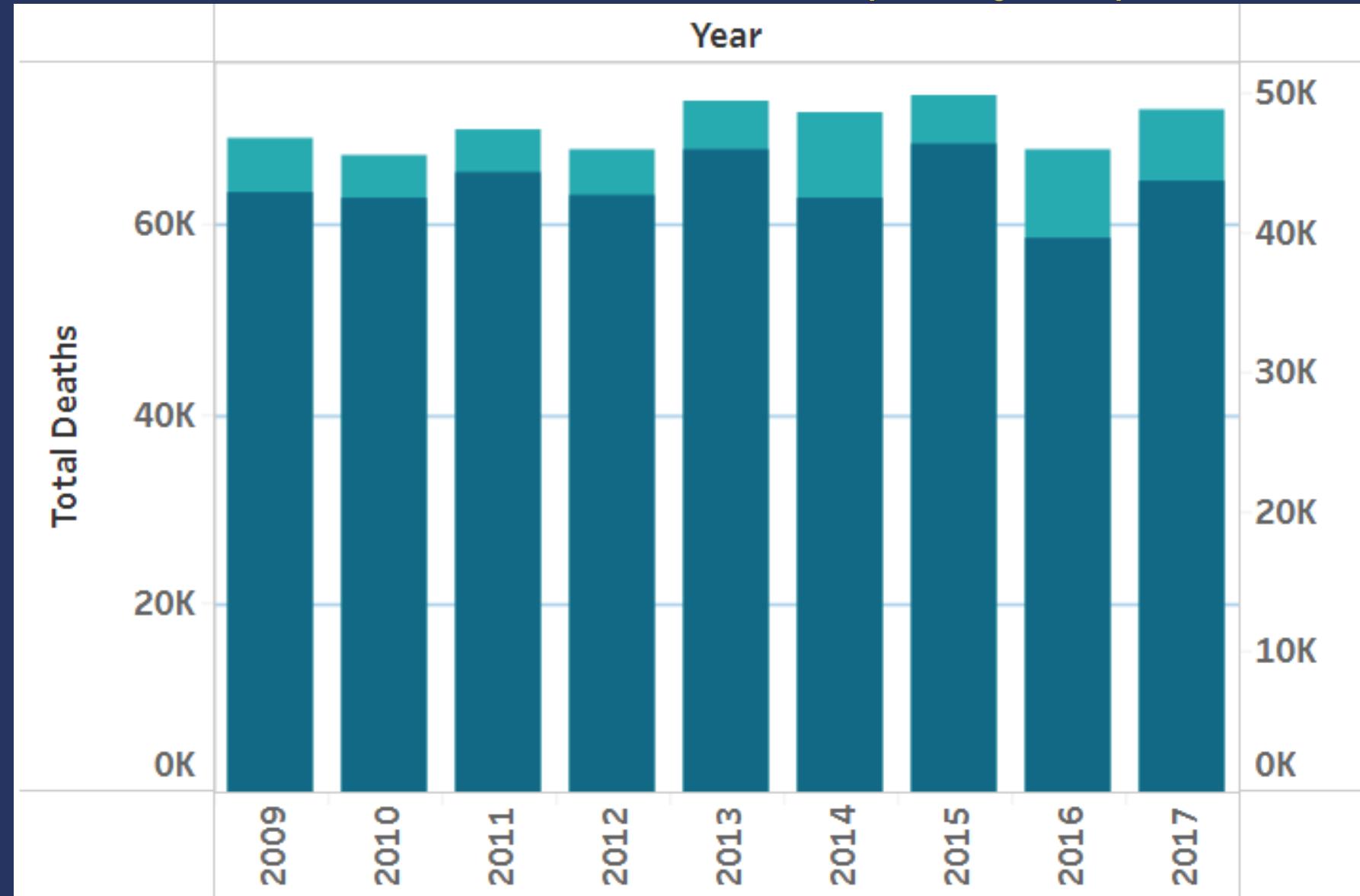


California, Florida, New York, Texas, Pennsylvania are top 5 Hotspots of Influenza in US that would need more medical staff deployed in comparison to other States.

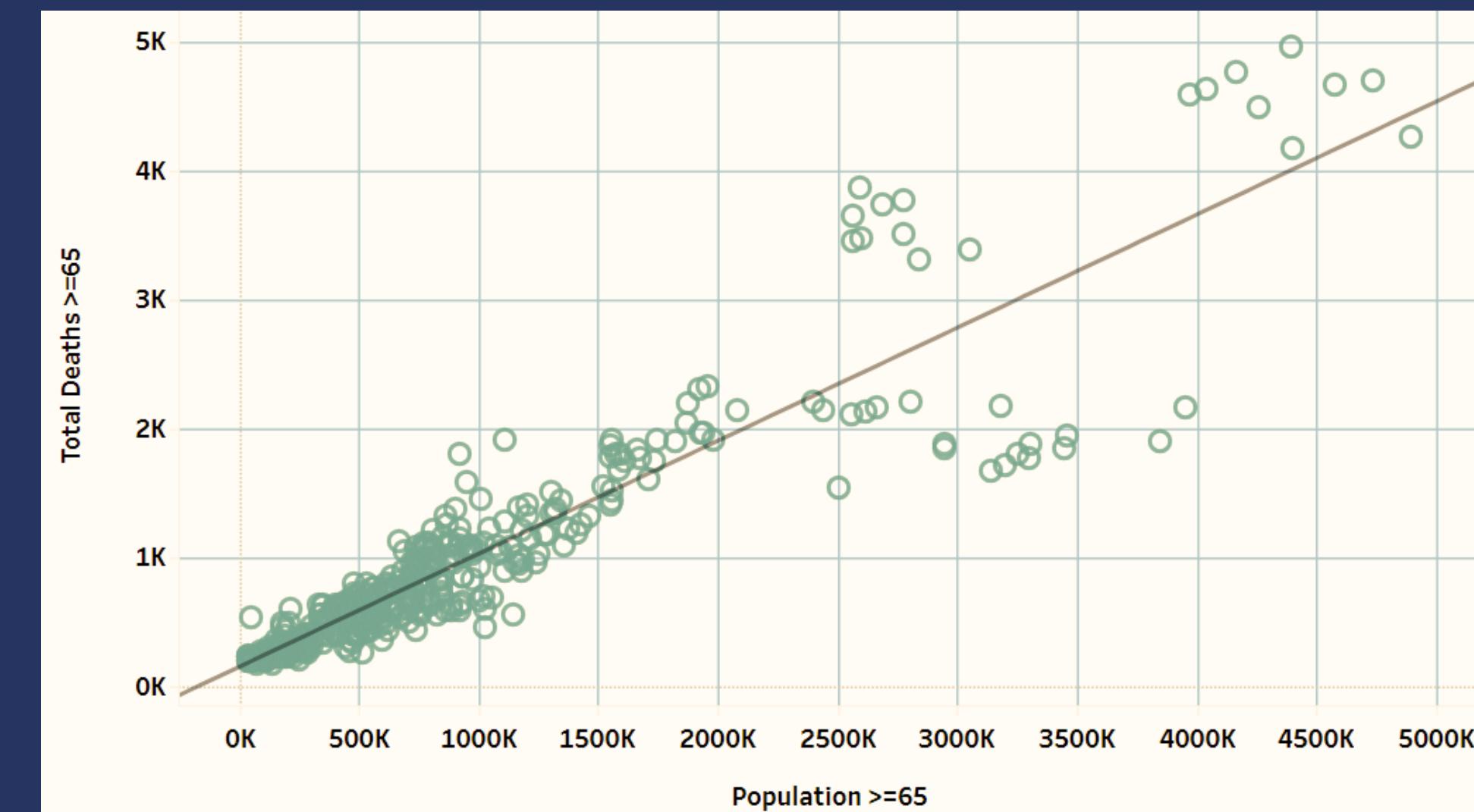
# VISUALIZATIONS



Proportional Distribution of Total Flu Deaths & Vulnerable Patient Flu Deaths (>=65 years)



Correlation between US Population and Influenza Deaths



- Citizens over 65 years of age are highly vulnerable to flu based on historical trends flu deaths.
- There is a strong correlation of 85% between the population of 65+ and flu deaths.
- Hospitals or nursing homes in states with the most vulnerable population should maintain an appropriate staffing ratio.



# PROJECT DELIVERABLES



## EXCEL ANALYSIS

This File contains hypothesis testing and its findings.



## TABLEAU STORYBOARD

This details an analytical story to executives with visualizations and recommendations.



## INTERIM REPORT

This document details consolidated findings of the analysis.

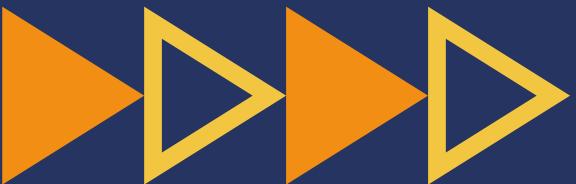


## GITHUB REPOSITORY

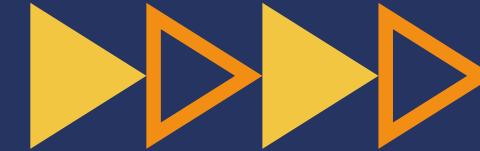
Contains complete project folder.



## 03. ROCKBUSTER STEALTH LLC



# PROJECT OVERVIEW



## OBJECTIVE

Analysing market trends of an online movie rental company to stay competitive in the market.

## KEY QUESTIONS

- Rockbuster's Board seeks data-driven answers for movie revenue, rental duration, customer locations, high lifetime value, and regional sales variations.
- Analysis supports Co.'s strategy to launch an online video rental service by providing insights on key business questions.

## DATASET

This data set that contains information about Rockbuster's film inventory, customers, and payments, among other things.

- [Download Rockbuster Dataset here.](#)
- [Data Dictionary](#)

# ANALYSIS STEPS



## 1 DATA PREPARATION

- Used PostgreSQL for querying, cleaned & summarize, filter and order data using the WHERE and HAVING clauses.
- Used SQL Joins to extract information from multiple tables.

## 2 ANALYSIS

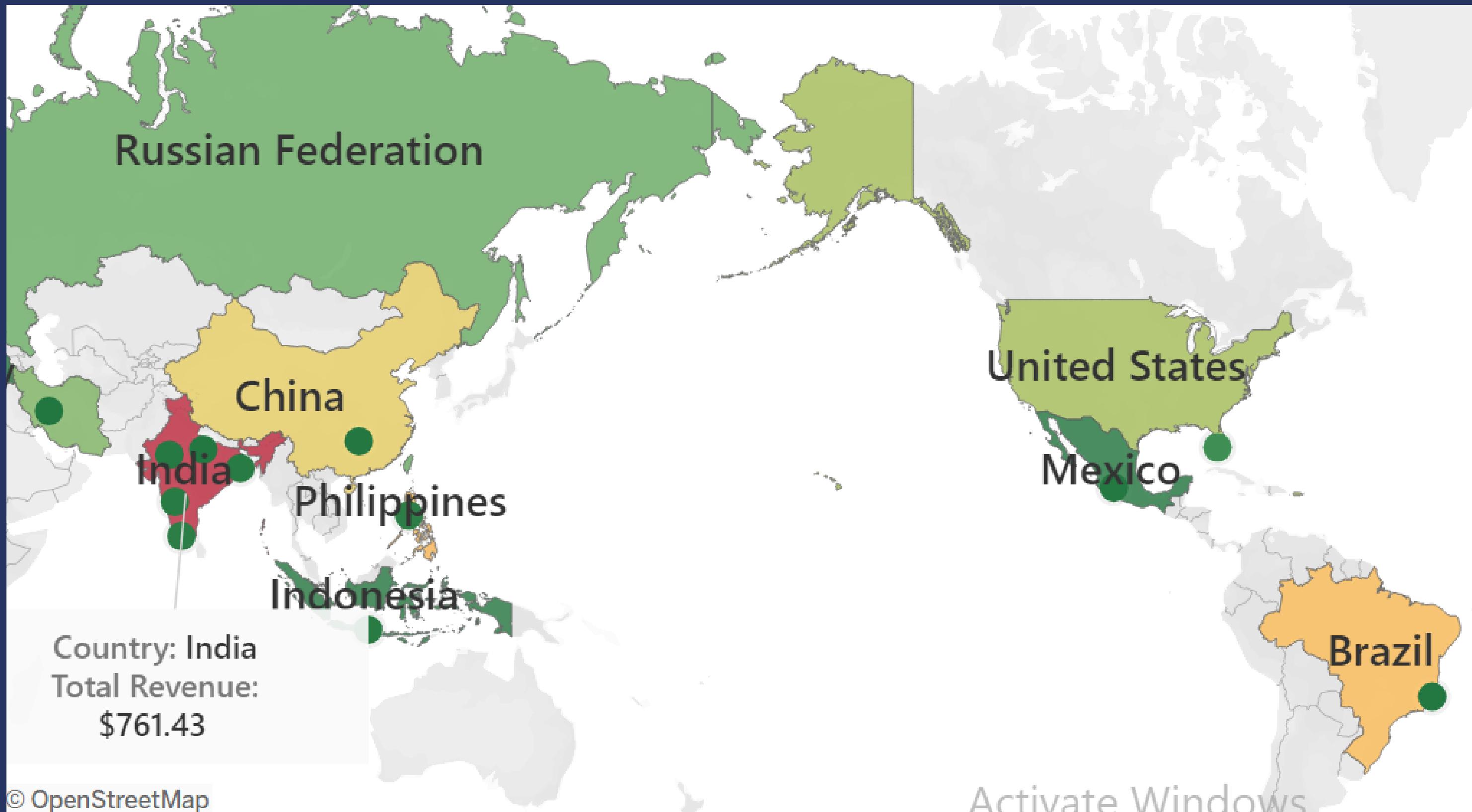
- Developed data profile and summary statistics using JOINS and subqueries.
- Transformed subqueries into common table expressions and SQL results were used for visualizations & deriving insights.

## 3 RECOMMENDATIONS

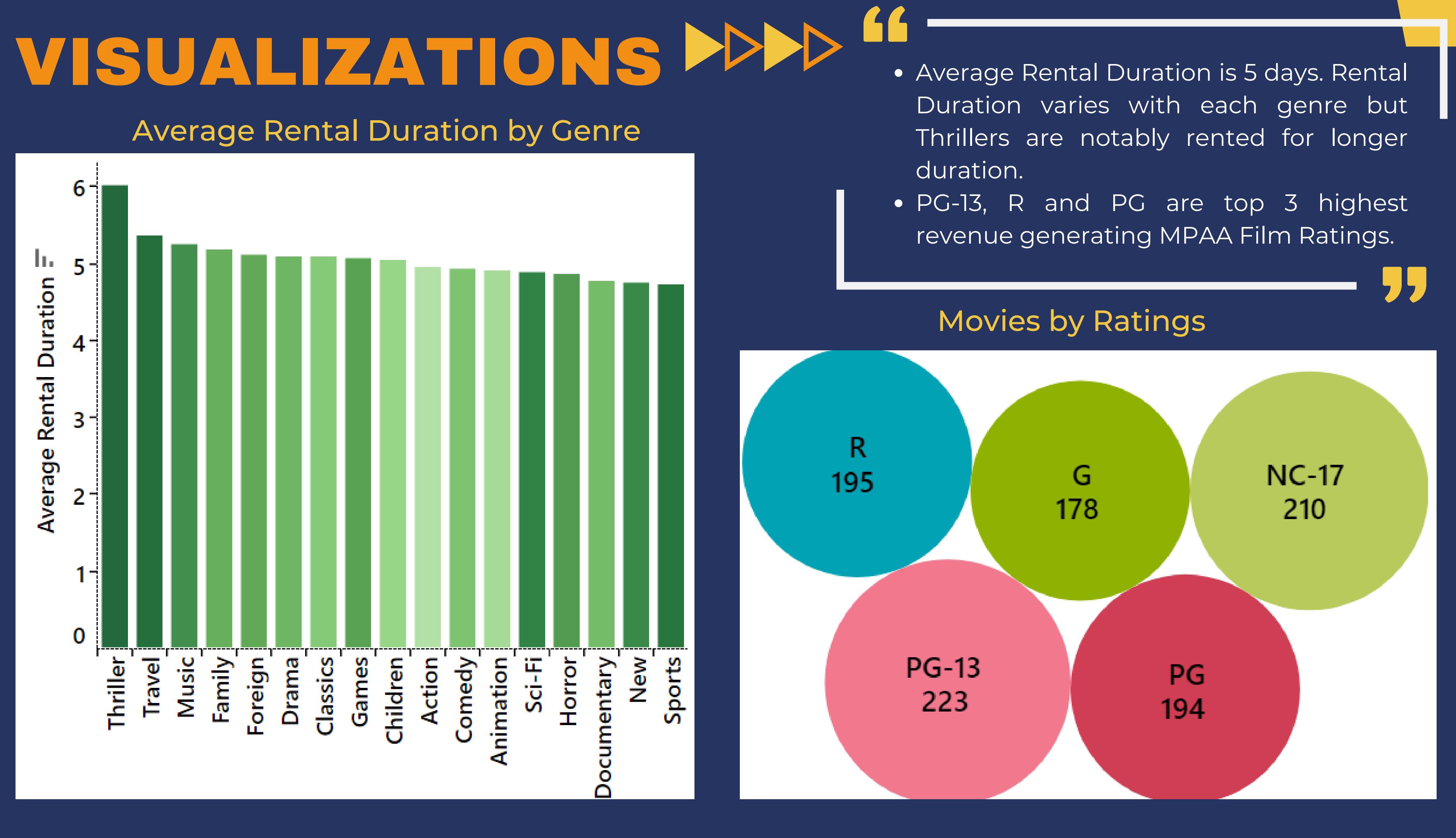
- Rockbuster should prioritize marketing in the top 10 revenue-generating countries.
- Implement loyalty programs or personalized offers to cater its premium users.

# VISUALIZATIONS >>>

## Top 8 Revenue Generating Countries



India, China, US, Japan, and Mexico are the top 5 countries with the most number of customers with a highest lifetime value.





# PROJECT DELIVERABLES



## EXCEL ANALYSIS

This File contains clean wThis File contains data output, SQL queries and results ranged data along with visualizations.



## STAKEHOLDERS PRESENTATION

This PowerPoint presentation details an analytical story to executives with visualizations and recommendations.



## TABLEAU STORYBOARD

This details an analytical story to executives with visualizations and recommendations.

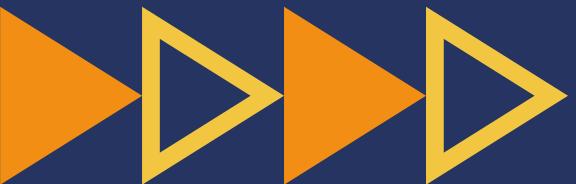


## GITHUB REPOSITORY

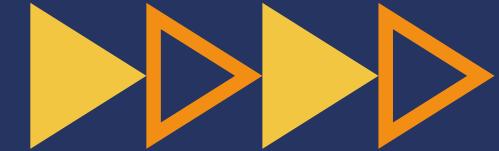
Contains complete project folder.



## 04. INSTACART



# PROJECT OVERVIEW



## OBJECTIVE

- Instacart, an Online grocery store app with a diverse customer database.
- Instacart Stakeholder wants to analyse Customer variety, purchasing behaviors, in order to implement targeted marketing startegy.

## KEY QUESTIONS

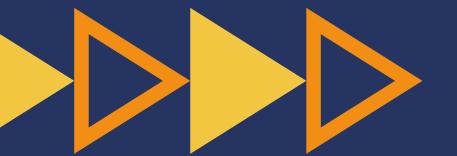
- Instacart stakeholders seek insights on busiest days, peak spending times, and price range groups to optimize operations and maximize revenue.
- Identify customer profiles based on ordering habits, for targeted marketing & personalized experiences.

## DATASET

Open-source data set from Instacart as well as a customer data set created and included for the purpose of this project

- [Customer Dataset](#)
- [Data Dictionary](#)

# ANALYSIS STEPS



## 1 DATA PREPARATION

- Conducted research using to address data profiling, integrity issues, quality and data transformation for a cohesive dataset.
- Transformed data into a consistent and unified format.

## 2 ANALYSIS

- Merged, analyzed, and added new columns with conditional logic, functions, loc, and for-loops.
- Generated summary statistics and visualizations using groupby and Python.

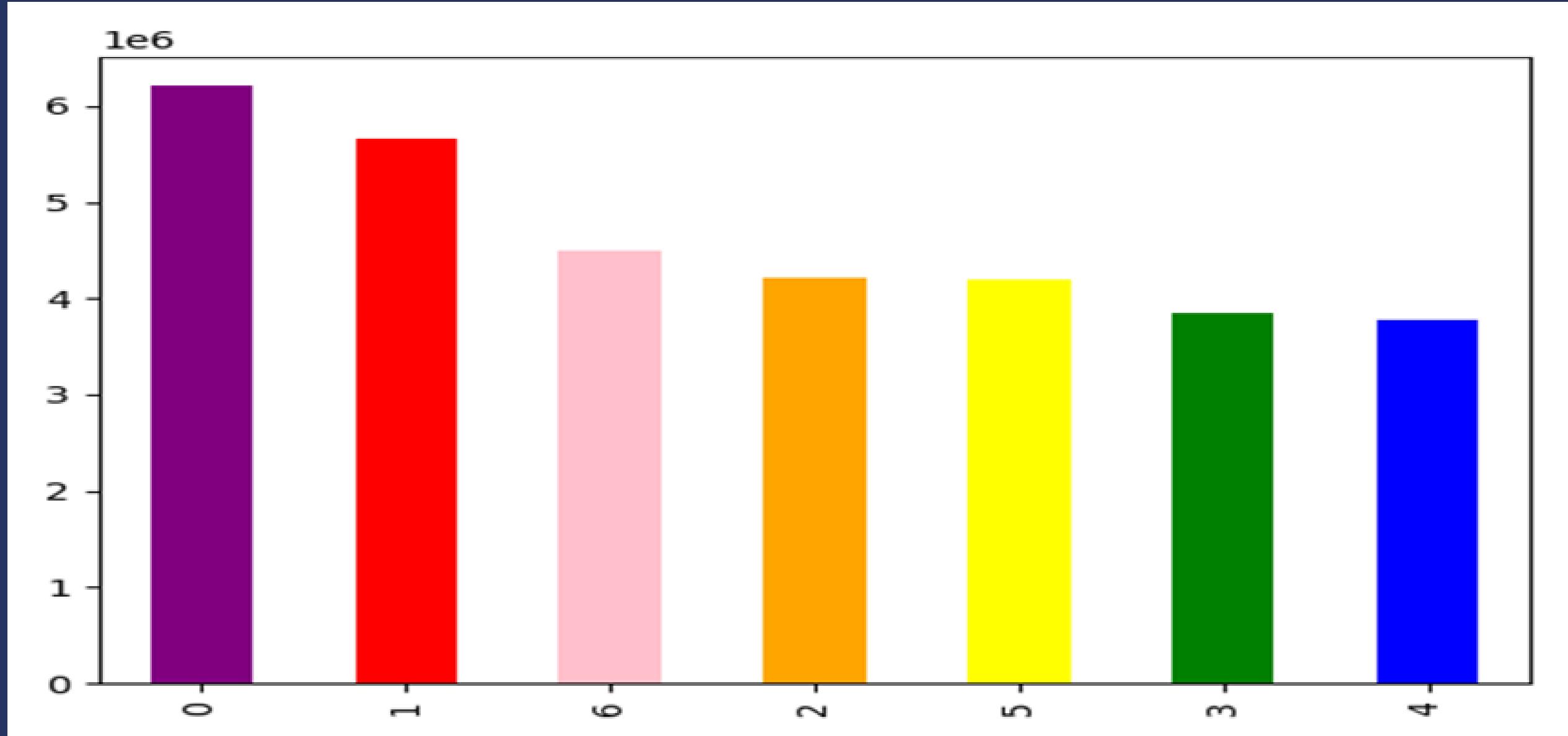
## 3 RECOMMENDATIONS

- Schedule ads on Tuesdays and Wednesdays for increased activity during slower order periods.
- Advertise high price point products between 2am to 3am when customers tend to spend the most money.

# VISUALIZATIONS



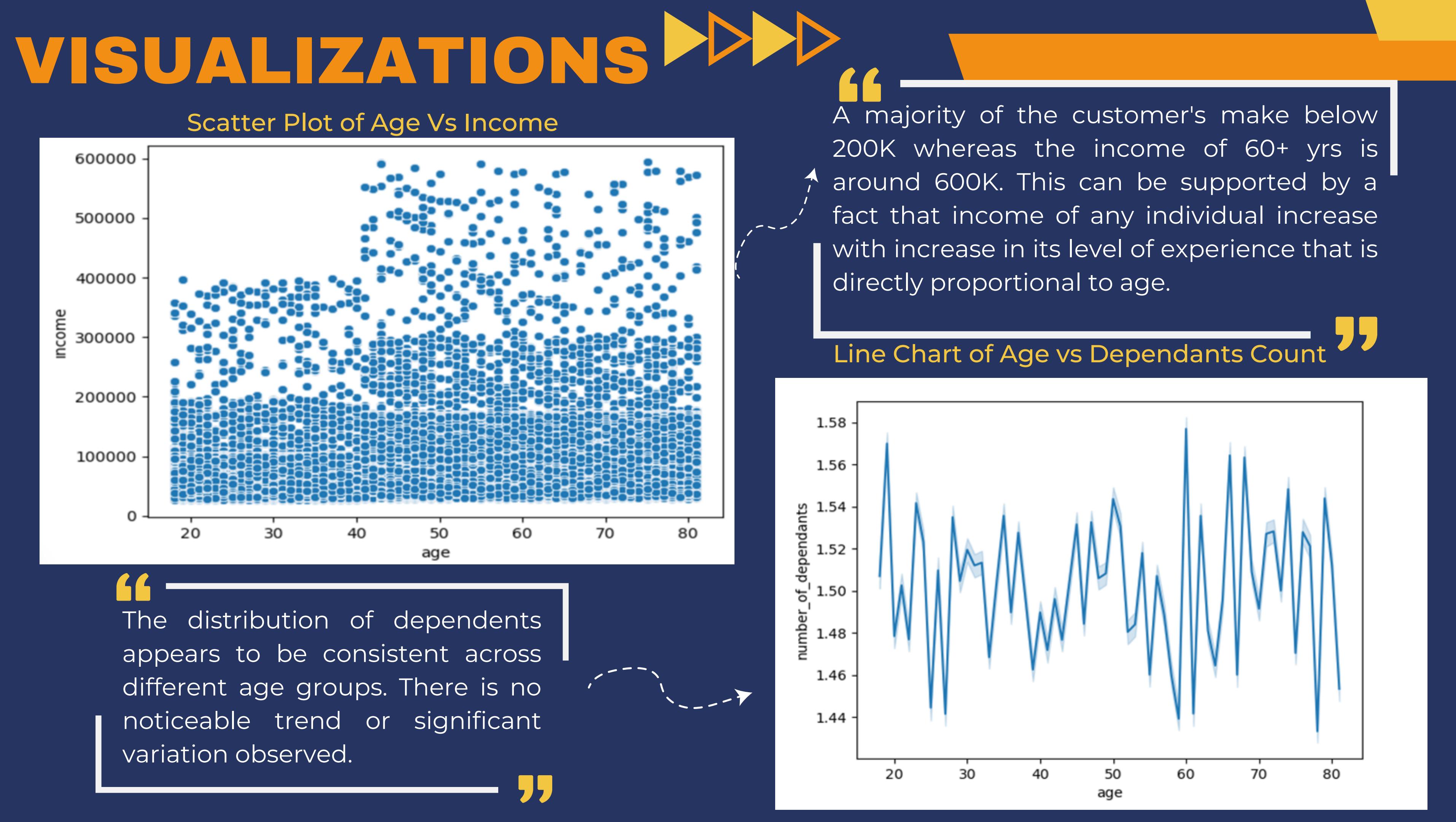
## Orders by Day of Week



0	1	2	3	4	5	6
Saturday	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday

Saturday and Sunday are the busiest day of the week whereas, Tuesday and Wednesday are comparatively slower days. It is quite a known fact that consumers generally go grocery shopping in weekends mostly.







# PROJECT DELIVERABLES



## EXCEL ANALYSIS

This File contains final report on findings and relevant visualizations.



## CROSSTABS

This file contains crosstabs to compute a simple cross tabulation of two (or more) factors.



## JUPYTER NOTEBOOKS

This folder contains all the Jupyter Notebook and scripts.



## GITHUB REPOSITORY

Contains complete project folder.



Click each deliverable to read in detail.

# THANK YOU

Do you have any questions?  
[malvika.nimoria@gmail.com](mailto:malvika.nimoria@gmail.com)  
+49-15212939227

