

Angitha KJ

Data Analyst



Projects

Project 1
Game Co

Video Game
Sales Analysis

Project 2
Influenza

Flu Season
Preparation
Analysis

Project 3
Rockbuster

Video Rental
Market Analysis

Project 4
Instacart

Online
Grocery Store
Market Analysis

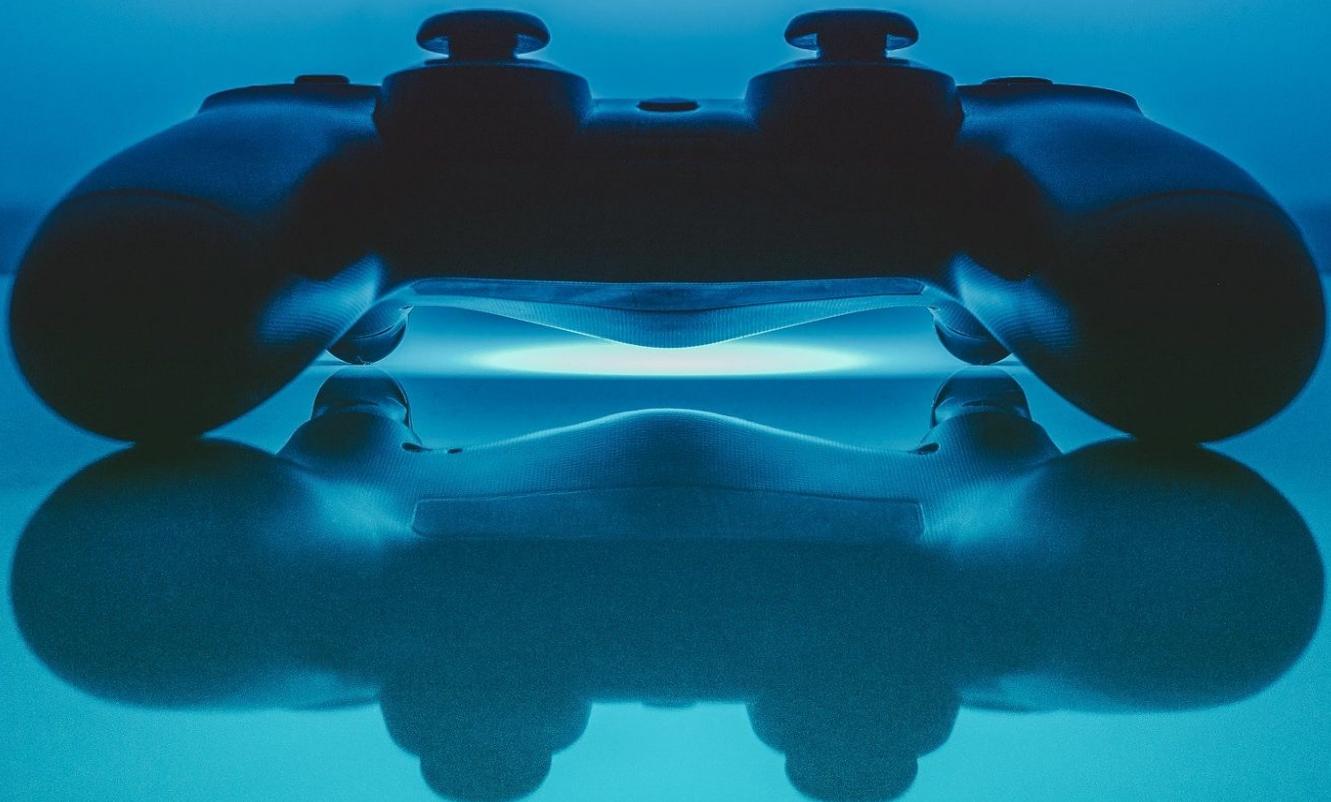
Project 5
Pig.E Bank

Online
Bank
Anti-money
Laundering

Project 6
COVID
Cases in
U.S.

Trend Analysis

Game Co.

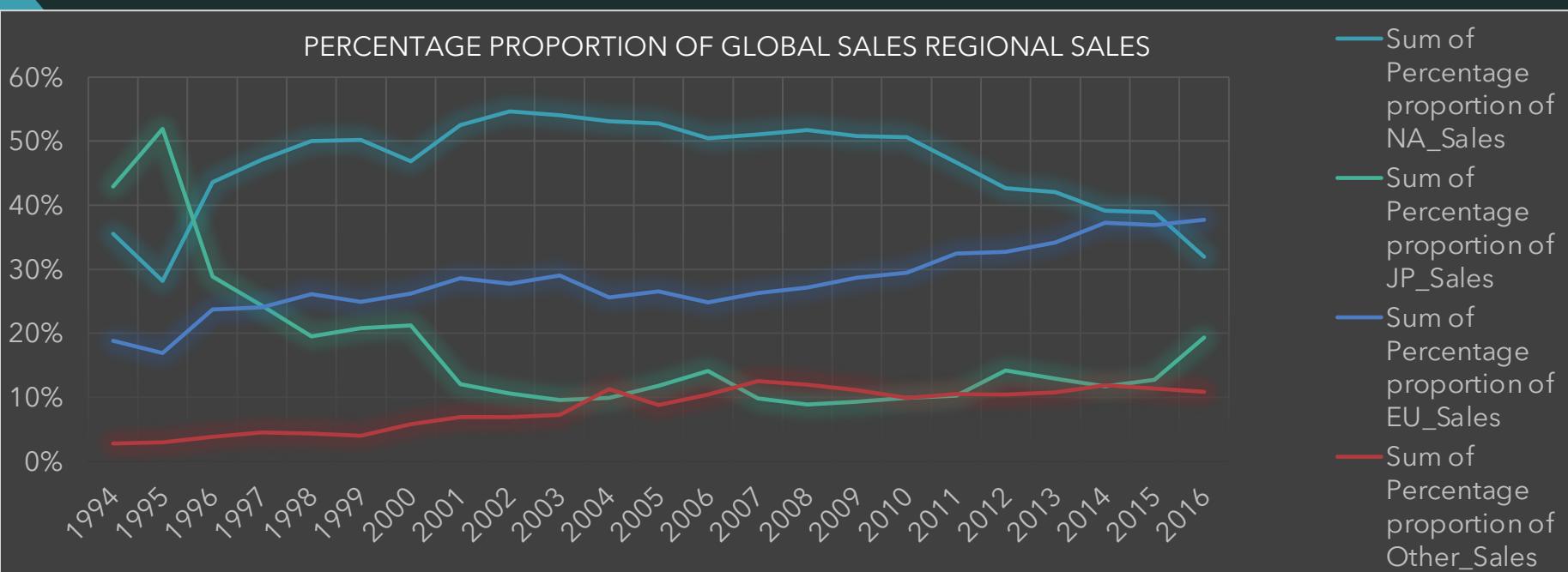


ANALYSIS ON GLOBAL VIDEO GAME SALES AND MARKETING STRATEGY 2017

Game Co Analysis

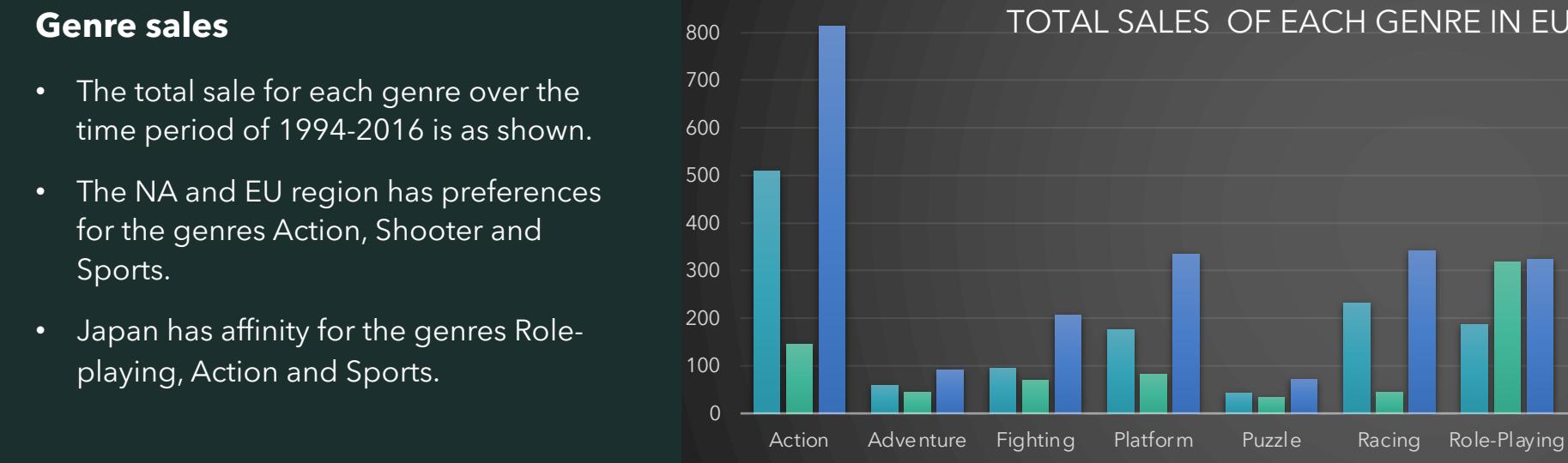
OBJECTIVE	DATASET	KEY SKILLS	TOOLS USED
<p>GameCo is a new video game company, which wants to use data to inform the development of new games.</p> <p>Perform a descriptive analysis of a video game data set to foster better understanding of how GameCo's new games might fare in the market</p>	<p>The data set was provided by the website VGChartz. It covers the historical sales of video games from 1980-2016, that sold more than 100,000 copies - from all platforms & genres.</p>	<ul style="list-style-type: none">• Data Cleaning• Data Grouping• Descriptive Analysis• Developing Insights• Presenting Insights• Visualization	<ul style="list-style-type: none">• Microsoft Excel• Microsoft PowerPoint

Game Co. Analysis



Regional sales

- The graph shows that the sales have not been remaining the same over the time period.
- The N.American region contributes the highest to the sales.
- Japan contributes the lowest but is showing an increase from 2014.
- The trend shown by EU region is promising as it gradually increases from 2014 while in N.America and Other regions the sales decline.



Game Co Analysis

- It would be advisable that the company maintain their budget concentration on NA, EU and Japan.
- Prioritise NA region as it contributes to nearly 50% of the sales or maintain the budget allocation as it is now for the region.
- The trend shows that there is a slight downward curve for the NA region. It would be advisable to revise marketing strategy in the region also include promotions.
- There is a gradual increase in the trend for EU region, increasing the budget and stock would be advisable.
- There is also a slight increase in the trend in Japan region, but it would be better to shrink budget there and concentrate on selling off the stock available.
- It would be advisable to concentrate on genre preferences in different regions.
- Game Genre preference for NA and EU is Shooter, Action and Sports.
Game Genre preference for Japan is Role playing, Action and Sports.

Influenza

Preparing for influenza season &
planning for staffing allocations

Influenza Analysis

OBJECTIVE	DATASET	KEY SKILLS	TOOLS USED
To help a medical staffing agency that provides temporary workers to clinics & hospitals on an as-needed basis. Examine trends in influenza and how they can be proactively be used to plan for staffing needs across the country.	Influenza deaths by geography • Provided by the CDC	• Data Cleaning • Data Grouping • Forecasting • Statistical Hypothesis Testing • Translating Business Requirements • Visualization	• Microsoft Excel • Microsoft PowerPoint • Tableau
	Population data by geography, time, age, and gender: • Provided by the US Census Bureau		
	Counts of influenza lab test results, by state • Provided by the CDC		

Influenza Analysis

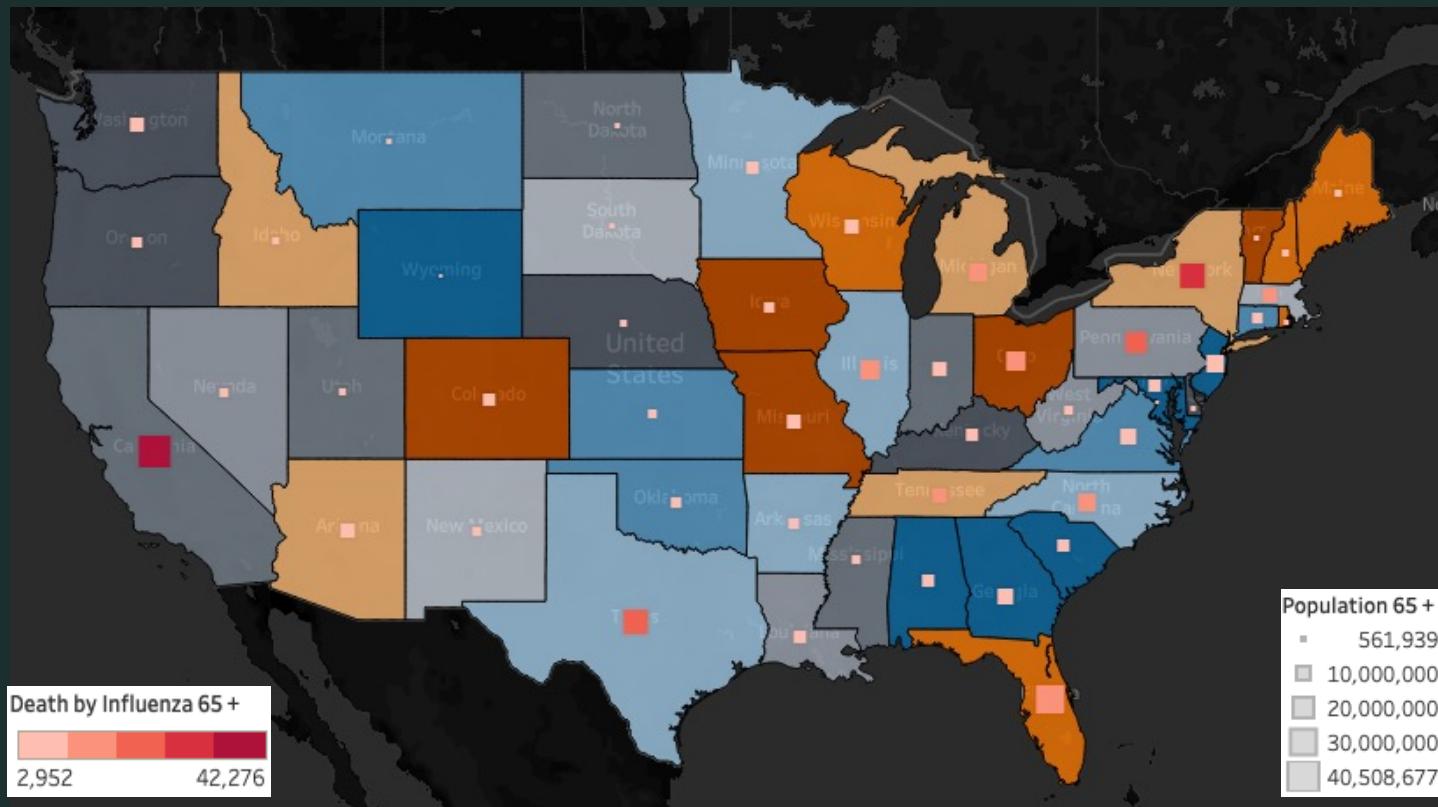


Total deaths by states
Ordered by Top 10

It was verified that the vulnerable population makes up the majority of deaths. This was consistent throughout the years and in all the states.

High population states often have a higher death count, due to population size.

To accommodate for this, we also calculated the total percentage of deaths based on the state's population.



Influenza Analysis

The states that require the most temporary healthcare workers are:

- California, New York & Texas.

They are closely followed by:

- Pennsylvania, Ohio, Florida, Illinois, North Carolina, Tennessee & Virginia.

RockBuster

Expansion into online streaming



RockBuster Analysis

OBJECTIVE	DATASET	KEY SKILLS	TOOLS USED
<p>RockBuster, a movie rental company, would like to expand into the world of online rental market.</p> <p>The management team require answers to business challenges, to develop a 2020 strategic plan.</p>	Expansive database containing customer information, rental, and rental records.	<ul style="list-style-type: none">• Data Cleaning• Data Grouping via Joins and Subqueries• Common Table Expressions (CTE)• Presenting Insights• Story Telling• Visualization	<ul style="list-style-type: none">• Microsoft Excel• Microsoft PowerPoint• Tableau• PostgreSQL• pgAdmin

RockBuster Analysis

MOVIES

1000 MOVIES
17 GENRES
5 MPAA RATINGS
All released in 2006
In English language

CUSTOMERS

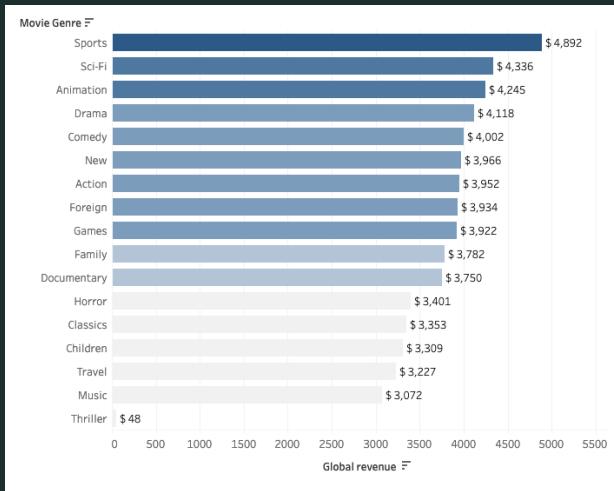
584 ACTIVE
CUSTOMERS
IN
109 COUNTRIES
IN
599 CITIES

RENTALS

\$61312.04
REVENUE
FROM
16044 TOTAL
RENTAL

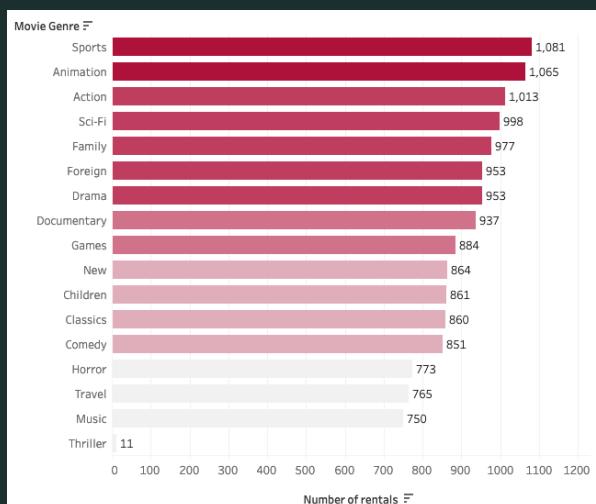
RockBuster Analysis

GENRE POPULARITY BY REVENUE



Top3 popular genre by revenue are Sports, Sci-Fi & Animation

GENRE POPULARITY BY MOST RENTED



Top3 popular most Rented are Sports, Animation & Action

TOP 10 COUNTRIES

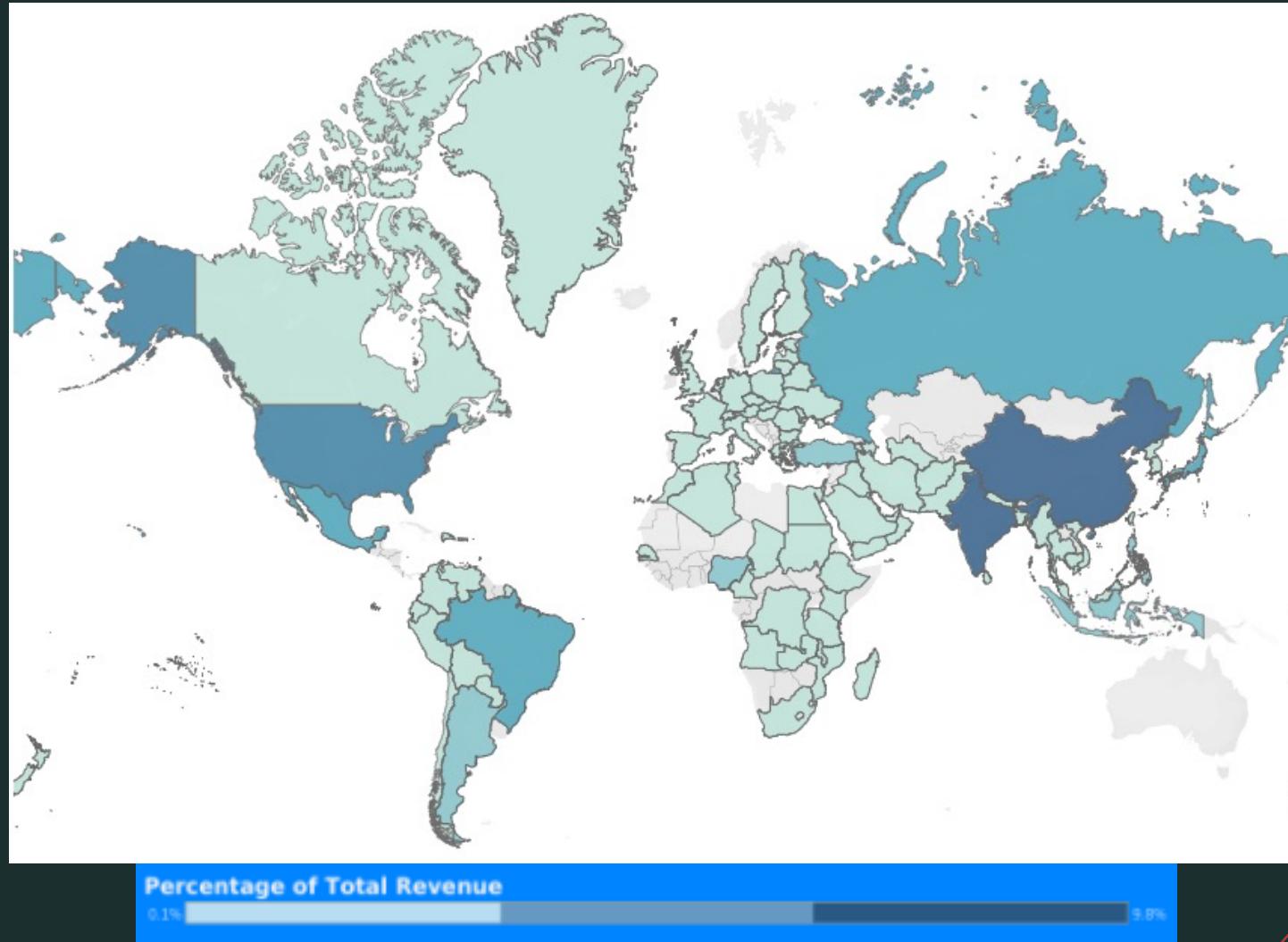
Country	Number Of Customers	Revenue (in USD)
India	60	6,035
China	53	5,251
United States	36	3,685
Japan	31	3,123
Mexico	30	2,985
Brazil	28	2,919
Russian Federation	28	2,766
Philippines	20	2,220
Turkey	15	1,498
Indonesia	14	1,353

TOP 5 CUSTOMERS FROM TOP10 COUNTRIES

Customer Id	Customer Name	Country	City	Total amount paid
225	Arlene Harvey	India	Ambattur	\$111.76
240	Marlene Welch	Japan	Iwaki	\$106.77
424	Kyle Spurlock	China	Shanwei	\$109.71
486	Glen Talbert	Mexico	Acua	\$100.77
537	Clinton Buford	United States	Aurora	\$98.76

Same countries are in top 10 as in the spatial analysis by number of customers and revenue. It shows a strong correlation between customers in country and revenue.

RockBuster Analysis



India & China make up most of our revenue, at 9.8% and 5.6% respectively. We should continue to focus growing here.

The U.S. Mexico, Brazil, and Japan are the medium revenue countries.

We have no presence in Australia and may benefit by marketing more there, due to their tastes being like the U.S. and EU.

RockBuster Analysis

- Marketing should prioritize the top countries which contributes the current highest revenue to attract more customers.
- Inventory must concentrate on the most revenue generating movies. Further analysis needs to be carried, to update the inventory as new releases come up, to hold the market.
- Priority should be given to the genres-Sports,Sci-fi,Animation and Action.
- To generate more profit from movies the proportionality of the rates and replacement rates should be maintained as per demand of a movie at the current situation.
- Rewards should be given to regular customers to promote business, as affiliate programs would be useful.



Instacart

Customer analysis to learn more about
customers behaviors and patterns

Instacart Analysis

OBJECTIVE	DATASET	KEY SKILLS	TOOLS USED
Instacart would like to uncover more information about their sales patterns. Perform initial and exploratory data analysis of some of their orders to derive insights and suggest strategies for better segmentation of their user base.	Customers Data set contains large amounts of user data, including income, marital status, family size, etc. Other data pieces provided by CareerFoundry, also from Kaggle.	<ul style="list-style-type: none">• Data Cleaning• Data Grouping• Data Aggregation• Deriving New Variables• Descriptive Statistics• Exploratory Analysis• Visualization	<ul style="list-style-type: none">• Python Libraries: Pandas, NumPy, SciPy, Seaborn, Matplotlib• Jupyter Notebooks• Microsoft Excel

Instacart Analysis

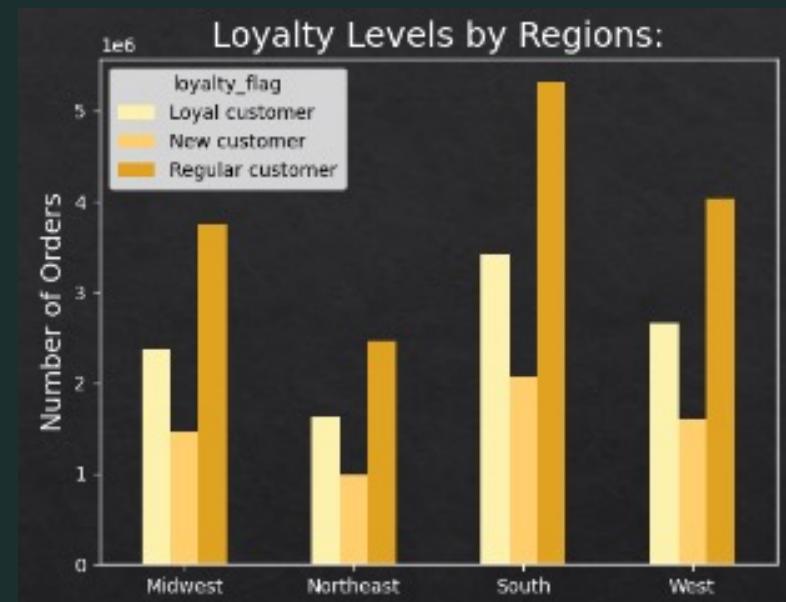


Saturday and Sunday have the highest amount of order activity. Of note, orders are at their lowest in the middle of the week and start picking up after that Wednesday.



10AM is the hour when the most orders come in. The hours of 10AM - 4PM have relatively similar amounts of orders and comprise most of the activity on the app. Orders start to slow down around 5PM and continue to do so until midnight.

The hours of 12AM - 7AM make up the lowest amount of activity on the app. Orders then start to pick up again around 8AM.

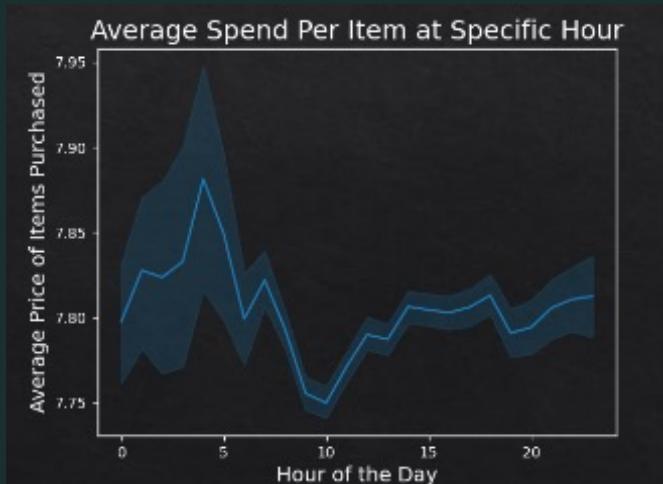


Depending on the region, the number of orders fluctuates, but their patterns among various customer labels are consistent. While there are more orders in the southern region, users retain the same ratio of orders throughout the other regions.

Instacart Analysis



Our most loyal customers make up a very small amount of the userbase. While our new customers make up the greatest number of users, they have the lowest number of orders with us. Our regular customers make the most orders through us and should be our focus. Customers prefer the mid-range and low-range products in terms of pricing



Instacart Analysis

- Weekends are busiest but we can advertise throughout the week. It is ideal to start advertising around 8-9AM, activity picks up around 10AM and a reminder to use our services can be the push they needed.
- Activity dies down during the week. We can offer time limited deals to spur action from users. We can offer deals that end in 3 days, or end of week - or even offer "Happy Hour" time sensitive discounts.
- Users prefer the mid & low-priced items. We can offer a reminder at check out by displaying previously purchased type items, ideally in the low-price range. Perhaps with a discount to incentivize committing.
- For our regular & loyal users, we can offer discounts after a certain number of purchases. These can act as milestones rewards for purchases. These deals can expire within the week to keep their frequency high of usage high.

PIG.E BANK

Anti-money laundering

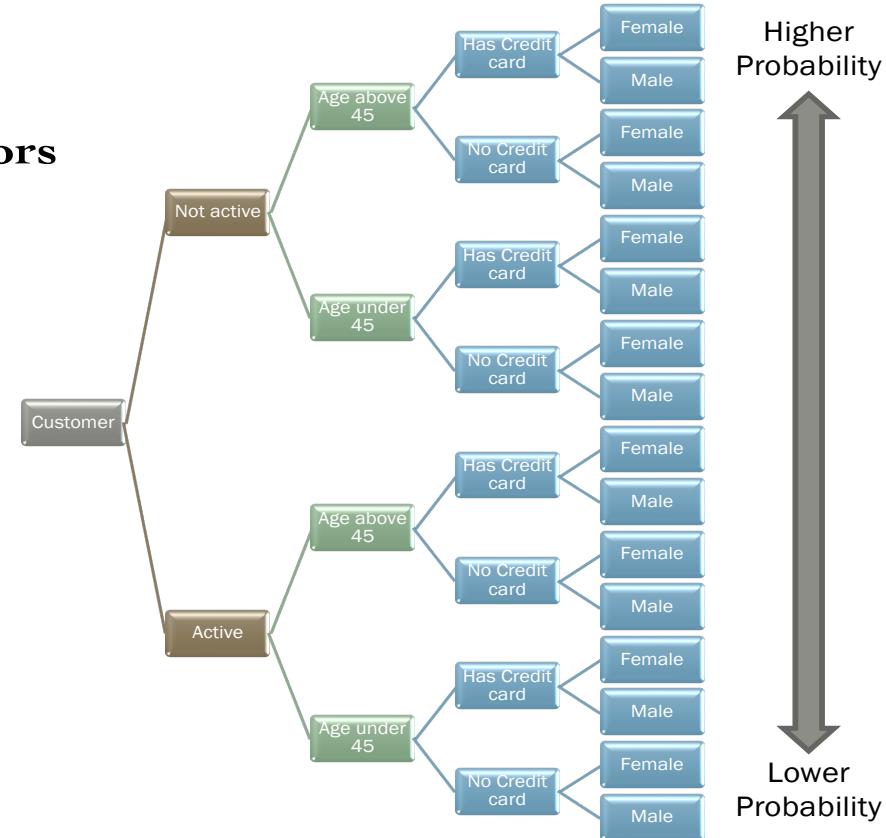


PIG.E BANK – DECISION TREE MODELING

OBJECTIVE	DATASET	KEY SKILLS	TOOLS
To create a decision tree model in order to help Pig E Bank identify the reasons why clients leave the bank	Transactions data set provided by Pig E. Bank bank. Sources: Pig E. Bank' Clients Data	<ul style="list-style-type: none">• Big data• Data ethics• Data mining• Correlation• Predictive analysis• Time series analysis and forecasting	<ul style="list-style-type: none">• Microsoft PowerPoint• Excel• Github

PIG.E BANK – DECISION TREE MODELING

Pig E. Bank Customer Exit Factors Risk Decision Tree



The leading factors that contribute to client loss, based on your analysis of the data provided are Gender, Age, Credit card usage and Active members. The decision tree model is constructed upon those factors.



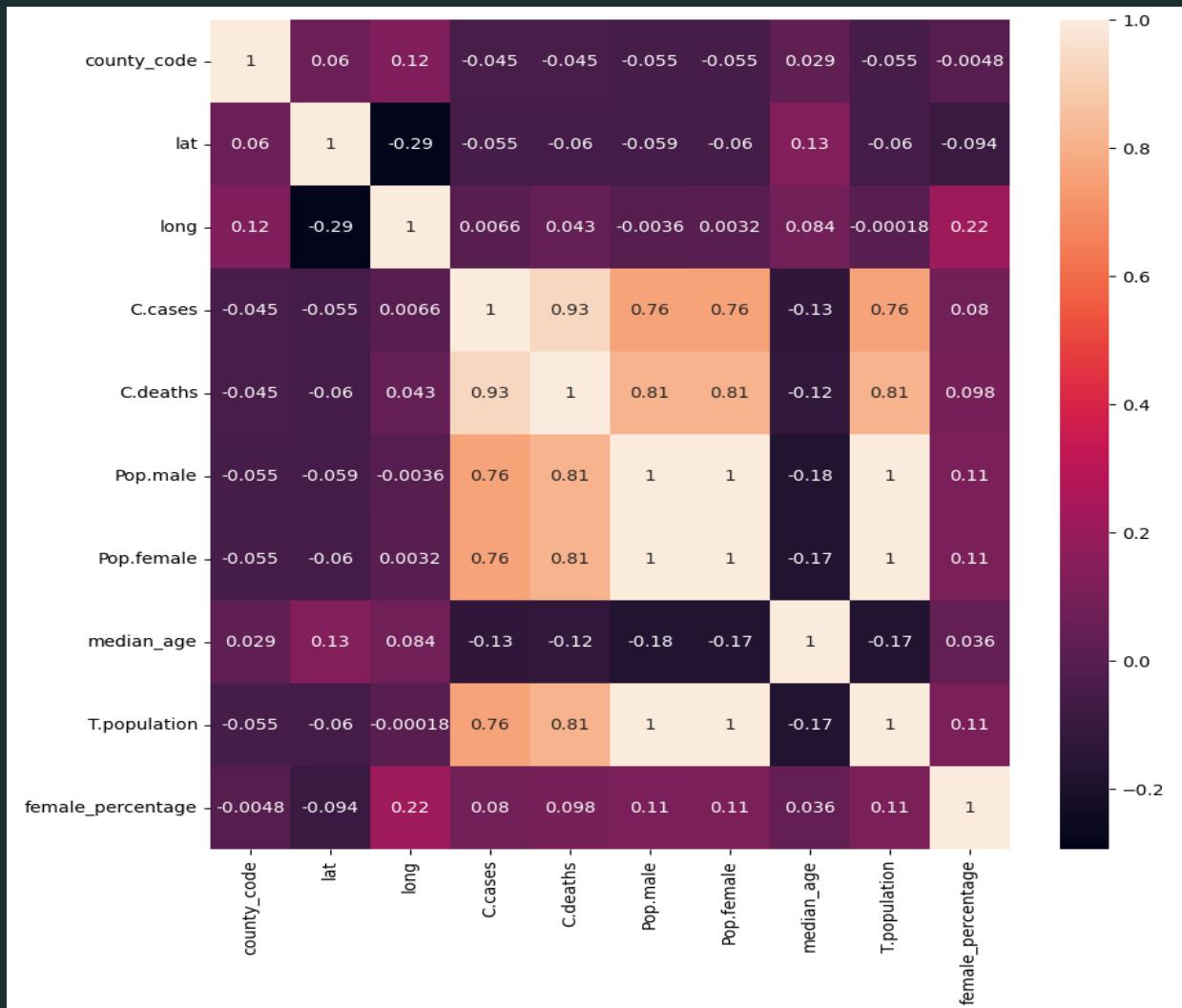
COVID CASES 2020-2023 ANALYSIS

TREND ANALYSIS

COVID CASES 2020-23 TREND ANALYSIS

OBJECTIVE	DATASET	KEY SKILLS	TOOLS
COVID-19 is a novel coronavirus, which caused pandemic in the beginning of 2020. The analysis for the general public. To inform on the trend of COVID virus infection trend.	The dataset to be analyzed, is the cases reported in the United States, is provided by Johns Hopkins University. There are two datasets (covid_us_county.csv and us_county.csv) downloaded from Kaggle, containing information about total COVID-19 cases and death by county and gender till 2023.	<ul style="list-style-type: none">• Data Cleaning• Data Grouping• Data Aggregation• Deriving New Variables• Descriptive Statistics• Exploratory Analysis• Regression Analysis• Clustering Analysis• Spatial Analysis• Time series analysis• Visualization	<ul style="list-style-type: none">• Python Libraries: Pandas, NumPy, SciPy, Seaborn, Matplotlib, Nasdaq• Jupyter Notebooks• Tableau• Github• Microsoft powerpoint• Microsoft Excel

COVID CASES 2020-23 TREND ANALYSIS

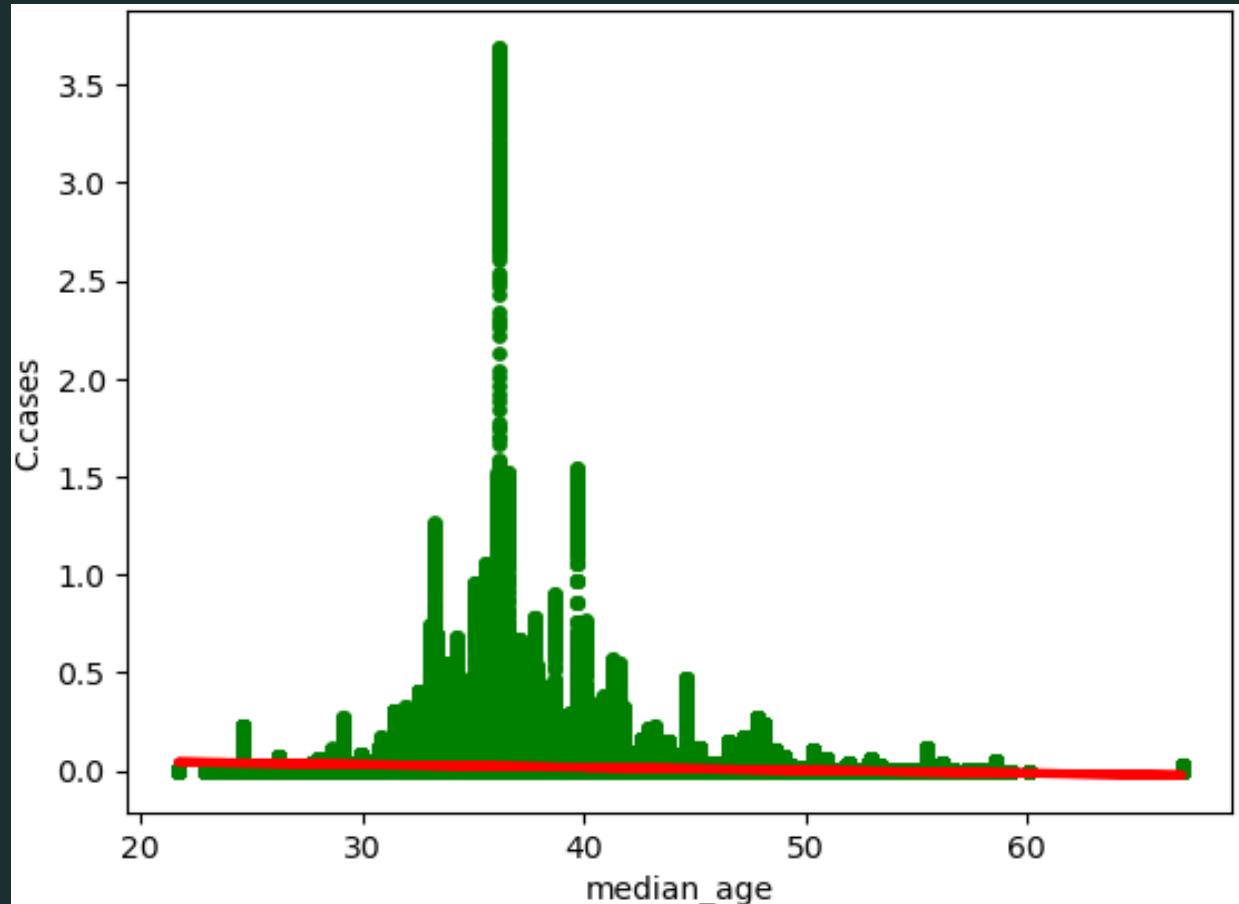


The coefficients shows that the death and cases of COVID are negatively correlated to the median age of the population.

The number of deaths and cases are positively correlated to the total population.

Also there seems to be a slight positive correlation for deaths and COVID cases to percentage of female population.

COVID CASES 2020-23 TREND ANALYSIS



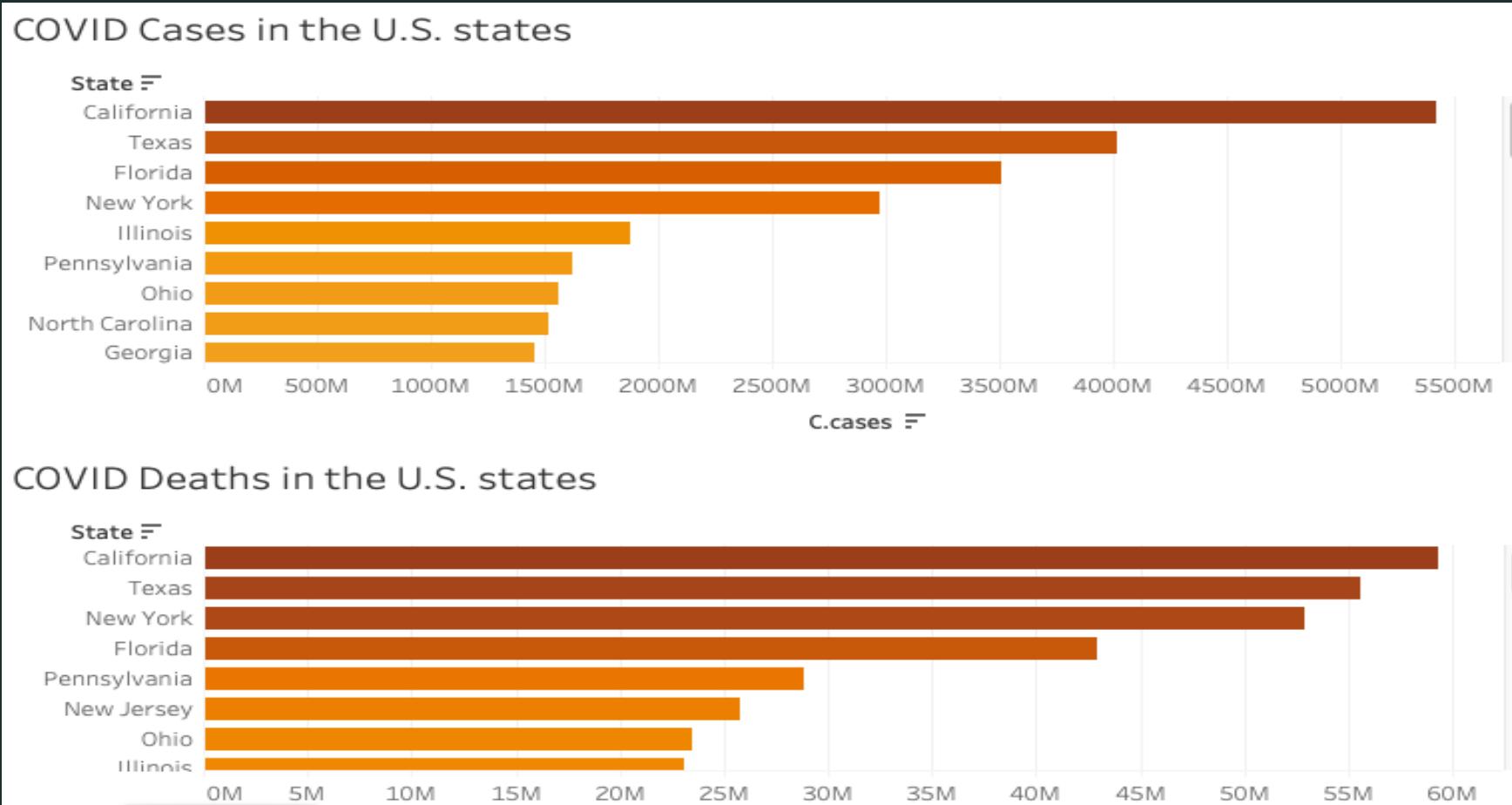
HYPOTHESIS - The population at the age group of 30 to 50 are susceptible to the COVID virus infections.

The scatter plot shows a negative regression, appears to support the hypothesis.

The slope drops slightly as the X value increases marking a decrease in the y value.

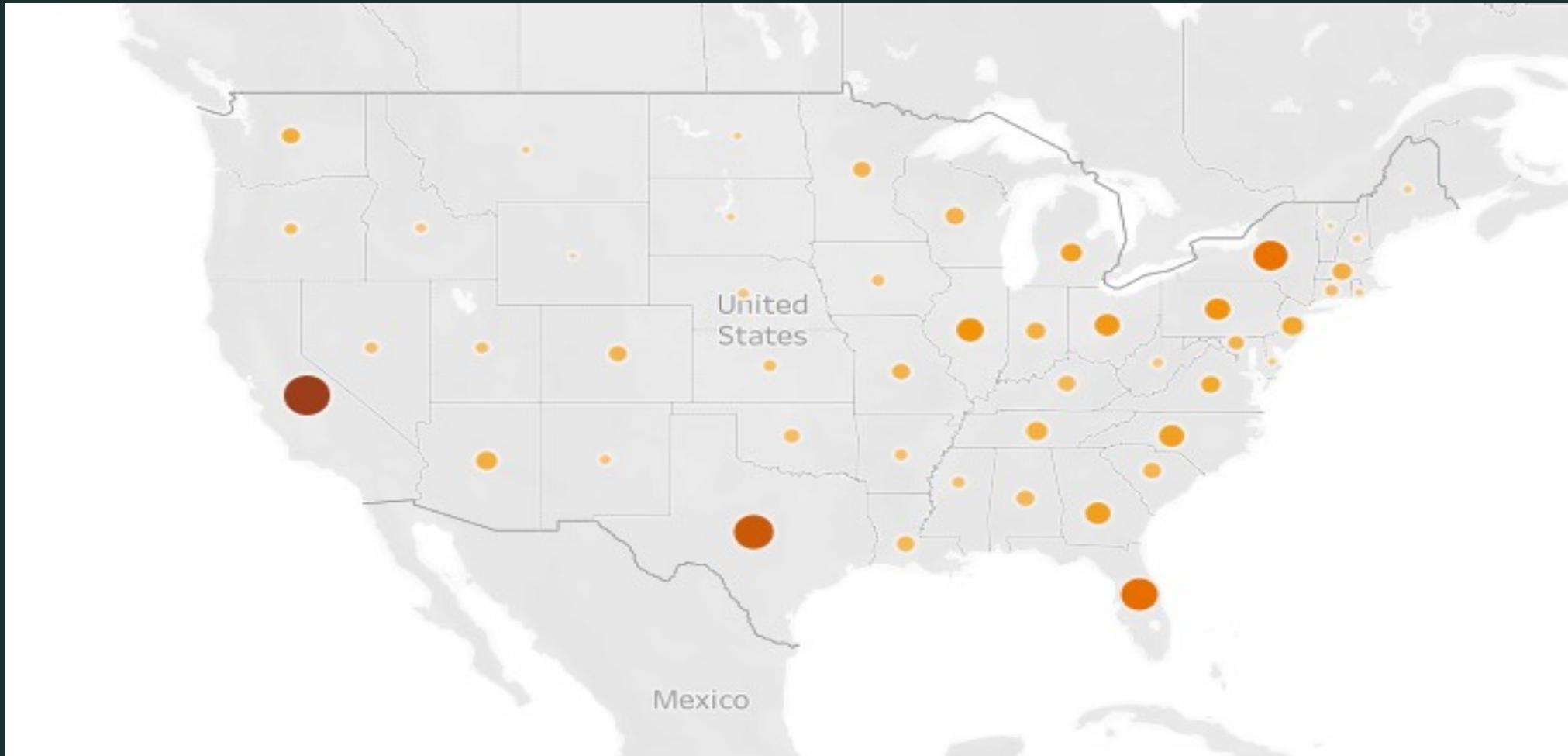
This means that the COVID cases is less among the population with age 50 and above. The cases are more among younger population below 50 years of age.

COVID CASES 2020-23 TREND ANALYSIS



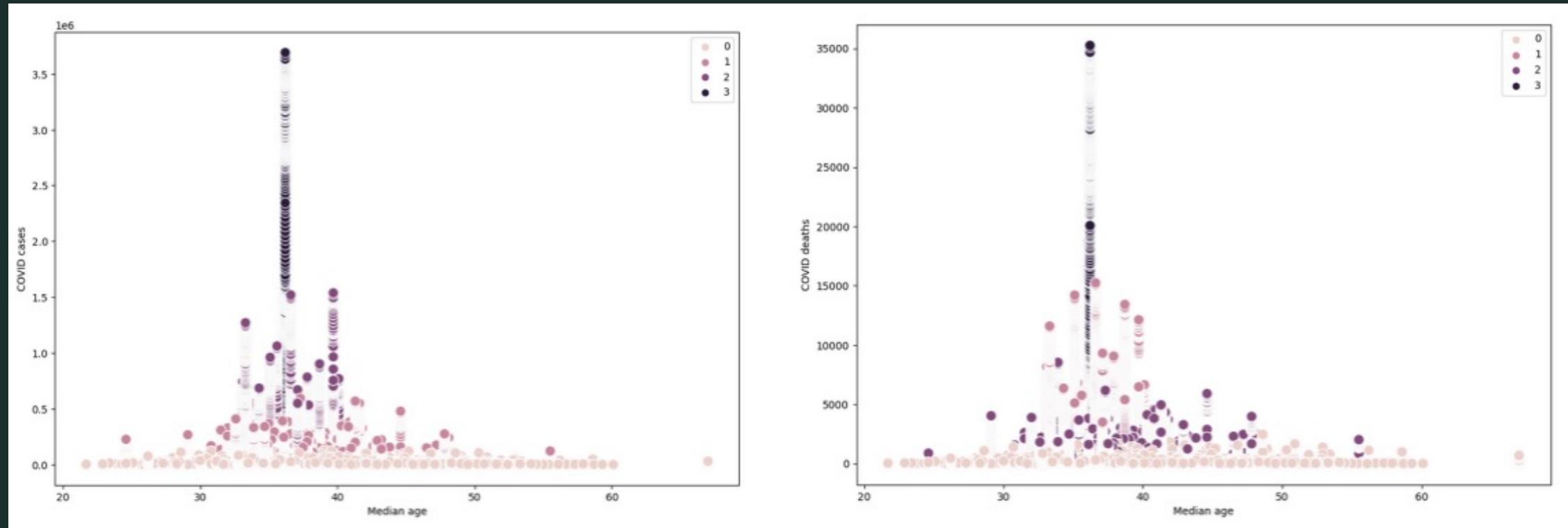
The U.S. states California, Texas, Florida and New York marks more number of COVID cases & COVID related deaths.

COVID CASES 2020-23 TREND ANALYSIS



The population size affect the percentage of COVID cases. The states with highest population and COVID cases are California, Texas, Florida and New York.

COVID CASES 2020-23 TREND ANALYSIS



HYPOTHESIS - The population at the age group of 30 to 50 are susceptible to the COVID virus infections. The Clustering analysis clearly inclines to the hypothesis that the population at the age group of 30 to 50 are susceptible to the COVID virus infections. Cluster analysis is helpful in this dataset as the data has non-linear relationships.

COVID CASES 2020-23 TREND ANALYSIS

- The population at the age group of 30 to 50 are more susceptible to COVID virus infections as they are the active working class in a society.
- The COVID cases and COVID related deaths are at its peak in December every year from 2020 to 2023.
- The higher the population in a state more the cases and death related to COVID. The top states having higher cases and deaths are California, Texas, Florida and New York.
- When considering COVID cases and COVID related deaths for the median age, the population below 30 years of age and above 50 years of age are at low risk, while population between the age of 30-50 are at high risk.
- It's highly recommended to maintain social distancing, take necessary sanitation steps, promote work from home whenever possible to reduce social interactions and wear mask when in public especially in the U.S. states with higher population and COVID cases.

Thank you

[Link to Tableau public](#)

[Link to Github](#)

