



# Jagpreet Batth M.D.

## Data Analyst Portfolio





# Hello!

## **MY NAME IS JAG BATTH**

I am here because I love DATA, MEDICINE & TECHNOLOGY!

With my expertise in all these fields, I will ensure that any company I work with is on the leading edge in every aspect!

You can find me at



[https://github.com/  
jagpreetbatth](https://github.com/jagpreetbatth)



[https://public.tableau.com/app/  
profile/jagpreet.batth](https://public.tableau.com/app/profile/jagpreet.batth)

# Showcase Projects



- Slides 5-9



- Slides 10-14



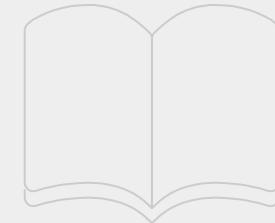
- Slides 27-31



- Slides 15-20



- Slides 21-26



# Professional Project Tools



**GAMECO**



**COVID-19  
RESPONSE**



# GamesCO.

## Background

A fictional video game entertainment company that needs marketing plan implementation based on data analysis that answers key questions in video game preferences and sales.

## Skills & Tools Utilized

- Data Cleaning
- Grouping & Summarizing Data
- Descriptive Analysis Pivot Tables
- Visualizations
- Organization of visuals, graphs, tables and presentation in an understandable format using gifs, animations and word art.

Click link below to view final presentation.



## Key Questions

Are certain types of games more popular than others?

- Are certain video game genres preferred when compared to others ?
- Have video game sales decreased or increased over time?
- Do certain video game platforms perform better than others in terms of sales?

## The Data

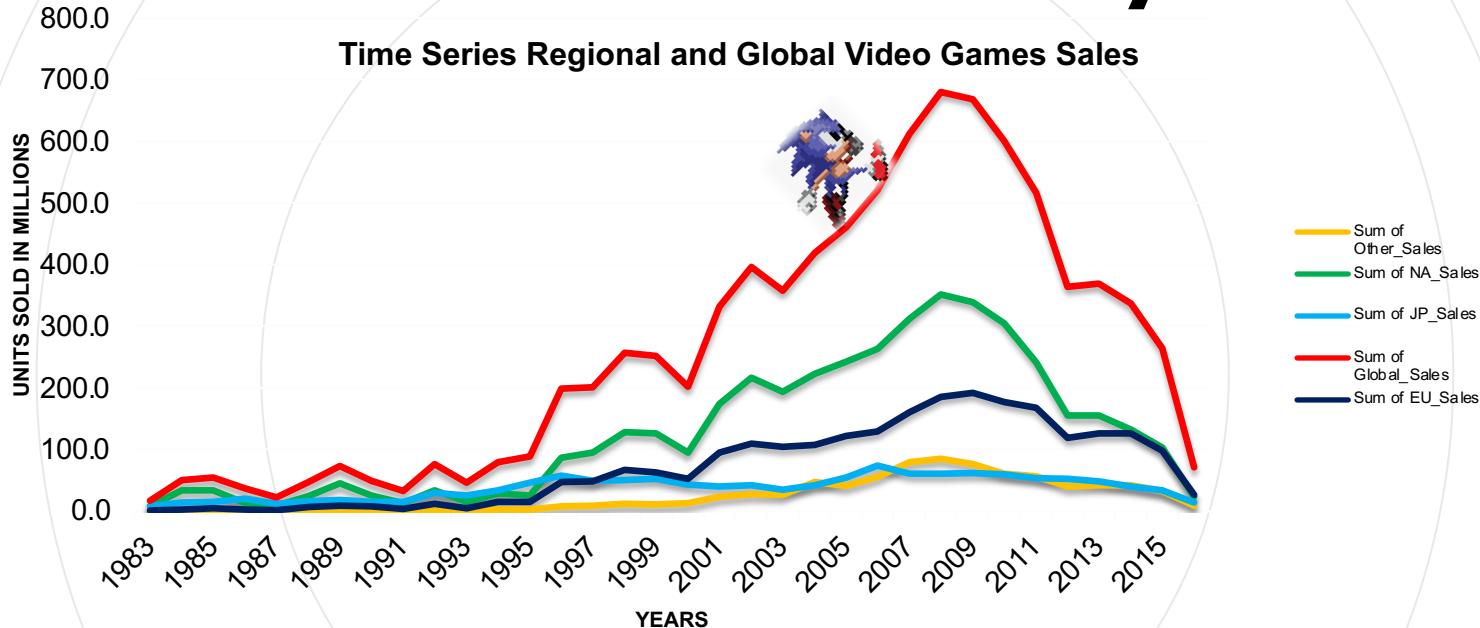
Data values used for this project include video game sales from 1980–2016 and units that sold over 100,000 copies. Variables include multiple platforms, genres, publishing studios spanning the EU, NA, Japan, and Others.

A circular graphic with a black background and a white border. Inside the circle, the word "GAMECO" is written in a bold, red, sans-serif font.



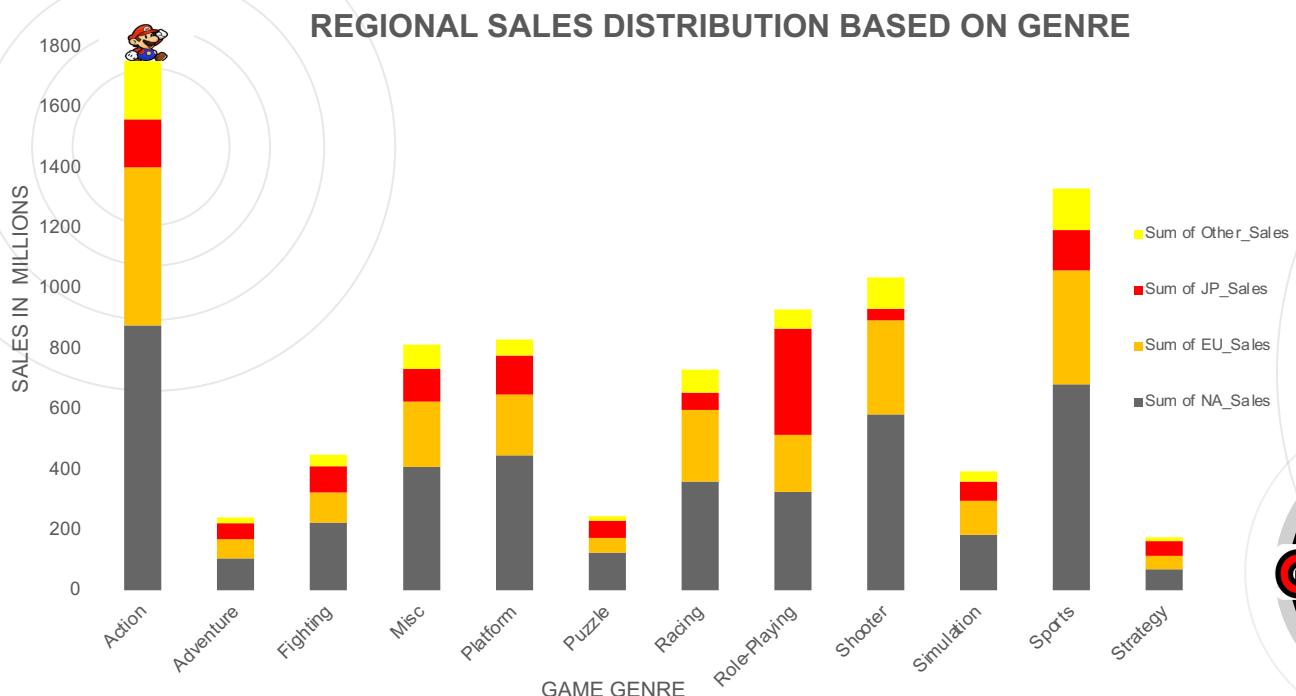


# GamesCO. Data Analysis



- Video game sales overall show variable patterns with both upward and downward trends, contradicting original market expectations that sales are static in nature.
- Although there was an overall upward trend starting around 1995, the peak was attained at 680 million units globally in 2008.
- The downward trajectory noted 2009 and beyond may reflect an online purchase trend as well as the 2008 financial crisis.

# GamesCO. Data Analysis

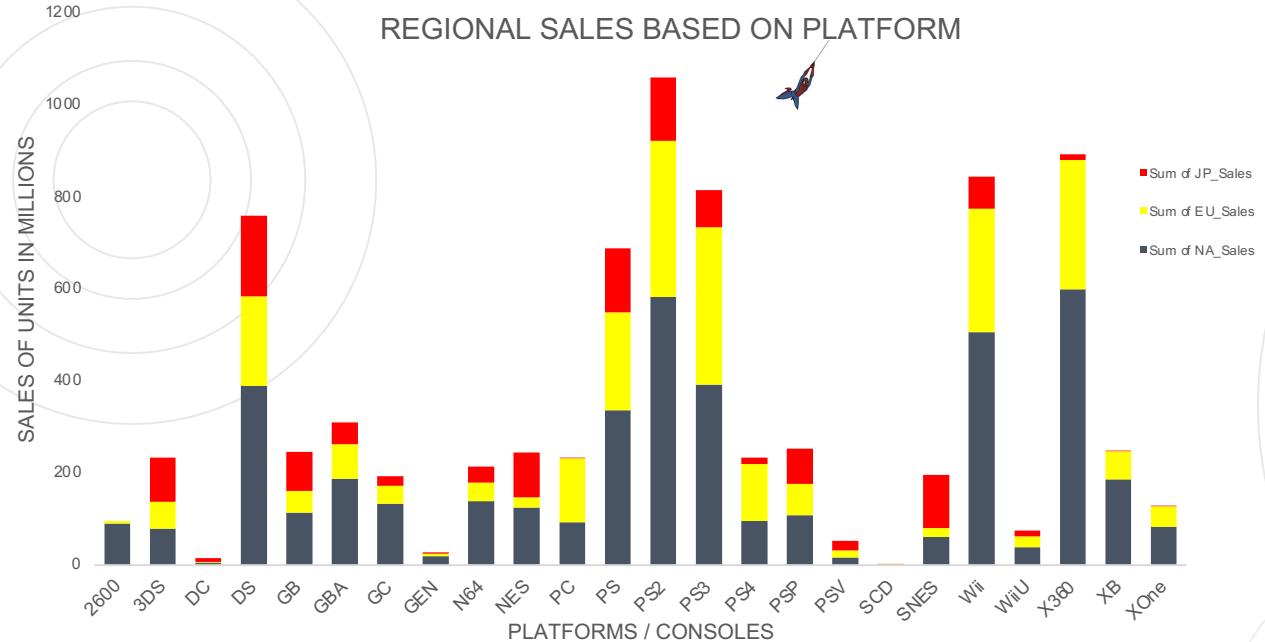


- Based on analysis of the data, we can see that the action genre has the highest proportion of sales overall.
- North America, representing the highest sales market, is in favor of games that are Action, sports and shooting based.
- Note that in Japan, the preferred genre that makes up the highest sales in that region, are Role-Playing games.



# GamesCO. Data Analysis

REGIONAL SALES BASED ON PLATFORM



- Studying the game sales based on our three prime regions, North America, Europe and Japan, we can see that PS2 garners the highest sales, followed closely by Xbox 360 and Wii consoles.
- Utilizing this information, we can determine marketing allocations and work closely with game publishers to create game compatibility geared towards these top performing platforms for maximum brand penetration based on platform loyalty.

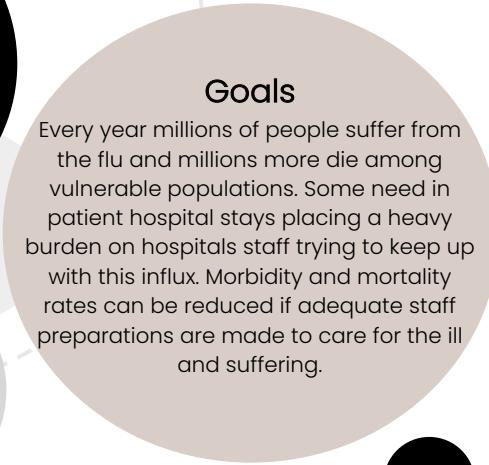
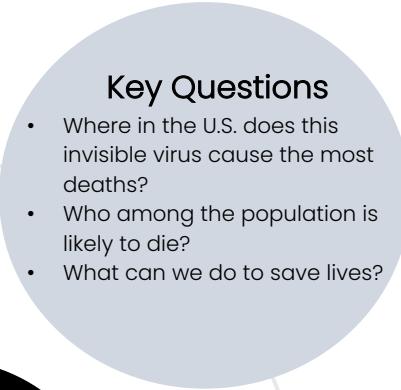
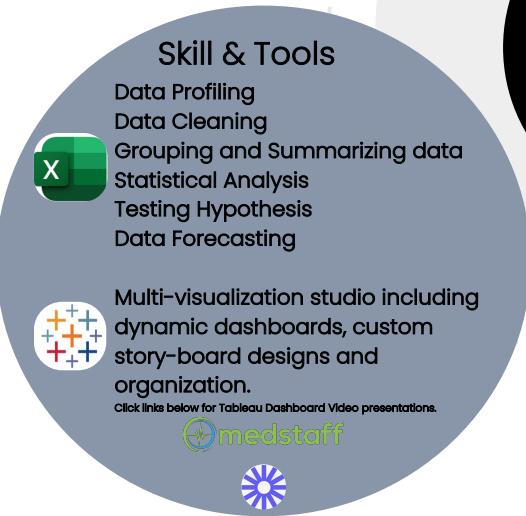


# GamesCO. Data Analysis Reccomendations

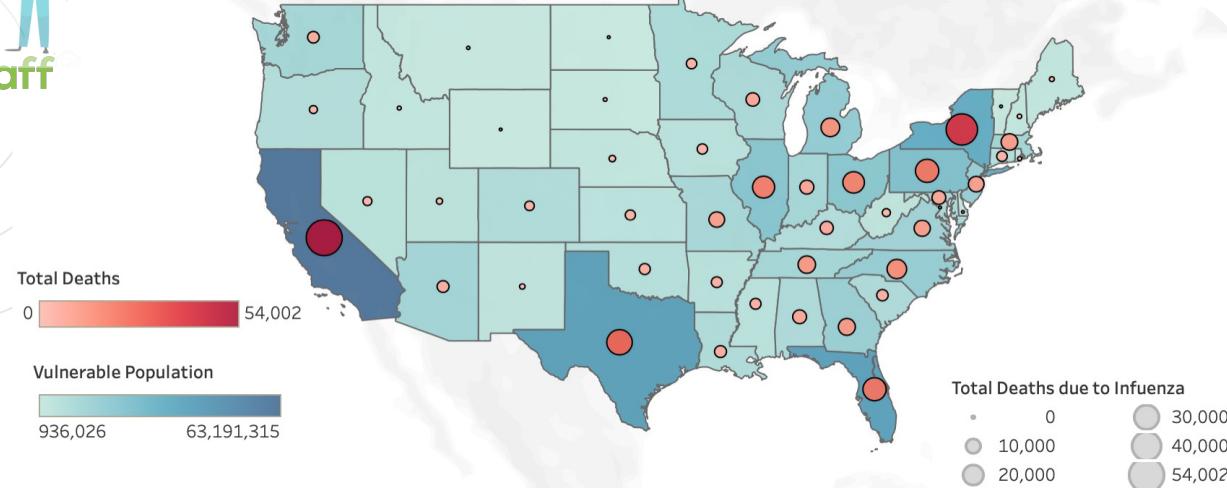
- Reject possible assumptions that video game sales have been static over time. In fact, there have been up-trends and down turns in sales based on our analysis.
- Socioeconomic and geopolitical factors such has financial crises, game purchase method changes and possible pandemics have to be factored into our expectations.
- Marketing allocation should be geared towards Action, Sports and Shooter genre games for increased profits/investment returns.
- In addition to the above mentioned marketing focus on the Action, Sports, and shooter genres, a greater emphasis needs to be placed on the platforms; PS2, XBOX 360 and WII to utilize our analysis for the absolute maximum profitability in the future.
- Based on the regional sales analysis, I recommend a greater budget allocations towards North America followed by Japan and Europe. We saw in our charts of sales over time a downward trend in our most profitable region and by allocating more marketing towards regions on a downward trend, we hope to reverse losses and revert to profits.



# Make data work for you!



# Data that Save Lives!



The interactive Tableau U.S. Influenza death and vulnerable population heat map can help us find out where help is needed the most.

Note the an increased amount of deaths(shown by bubble size) in California, New York and Texas, where there are also increased number of vulnerable individuals (indicated by deeper color).

Using this map and what it can show us, we can save lives by allocating staff to the areas that have been the hardest hit by the influenza virus in the past.

# Power in Numbers!



**334,000**

Estimated Deaths due to Influenza from 2010 - 2020

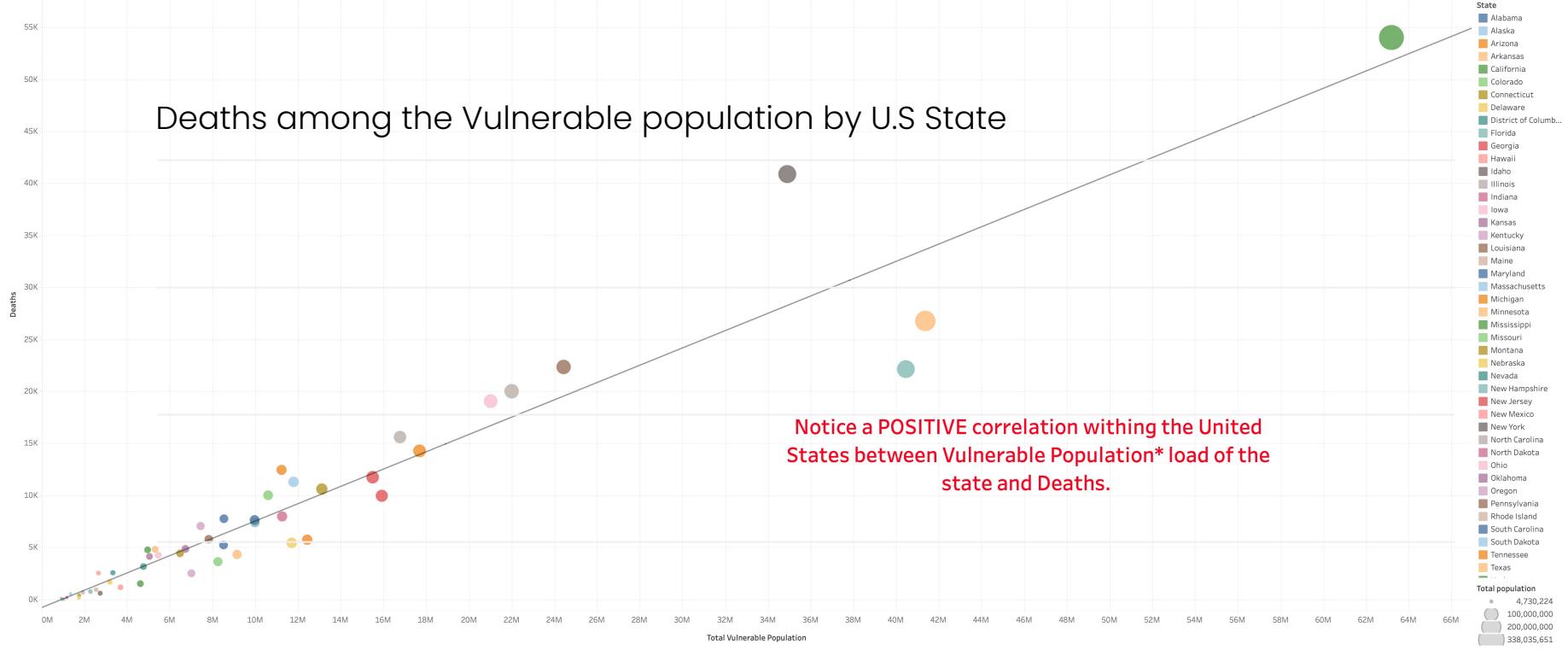


**547,778,514**

Vulnerable population (people more likely to die from the influenza virus) from 2010-2020

**41,000,000**

Flu related illness from 2010-2020 placing care burden on already stressed hospitals



The above scatter plot shows an positive correlation between vulnerable populations and deaths. Increased care and emphasis must be placed on individuals younger than 5 and older than 65 years of age in order to avoid more deaths resulting from the influenza virus.





# DATA=POWER=CHANGE

## DATA KNOWLEDGE

1 The Influenza population data analysis gave the knowledge to makes changes that will save lives in the future, for example, allocating staff and resources where there are an increased number of vulnerable individuals.

## DATA POWER

2 Data analysis of Deaths caused by influenza among various U.S. States gave us the power to prepare for the future to prevent diseases even before they happen.

## DATA TO CHANGE

3 Gaining insight from the data analysis allows staffing companies the power to makes necessary changes and adjust to an ever changing atmosphere in current and future pandemic level diseases.

# RockBuster Stealth

## Background

A fictional video films library company needs data analysis based insight in order to compete with online streaming video platforms such as Amazon, Netflix and Apple and setup an efficient online film library.

## Techniques/Approach

A collaborative effort is needed working with the business intelligence department to design a strategy that will help the online video platform succeed. Utilizing SQL to query customer movie preference, rental patterns, spending routines, and most profitable areas we gain our maximum return on investments goals. Results are then used to answer ad-hoc questions that may arise in the process of project completion.

## Tools/Skills



SQL database queries

Data filtering

Inner/outer/ whole joins

Cleaning/summarizing

Common Table Expressions



Multi-visualization studio including dynamic dashboards, custom story-board designs and organization.



Organization of visuals, graphs, tables and presentation in an understandable format using gifs, animations and word art.

## Key Questions

- Which genres generate the greatest revenue?
- Which movies have generated the greatest income?
- Which countries in the world generate the highest profits?
- Who and where are our highest paying customers?

## The Data

[ROCKBUSTER DATA SET](#)

[PostgreSQL Database](#)

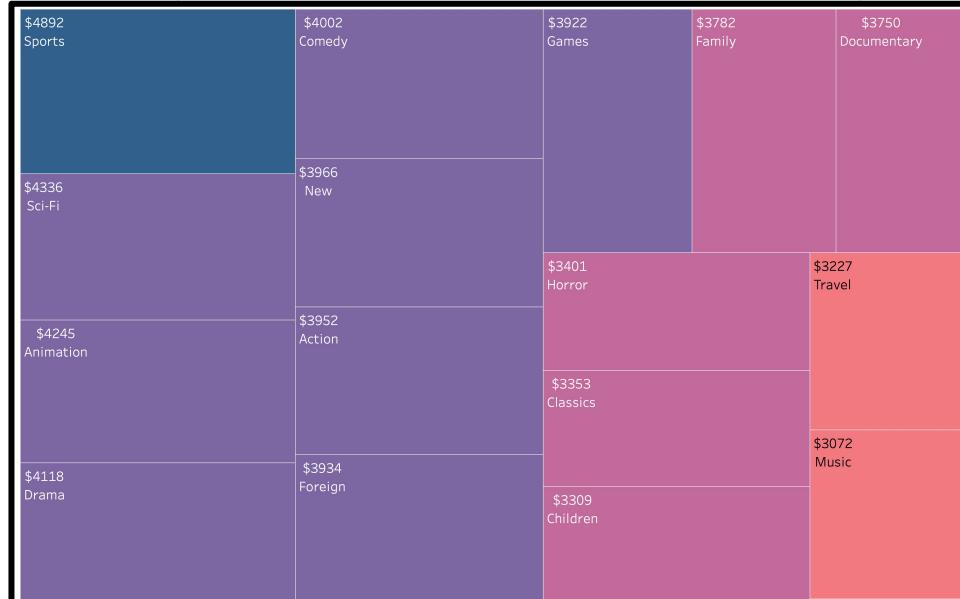
[Tableau Presentation Link](#)



# RockBuster Stealth



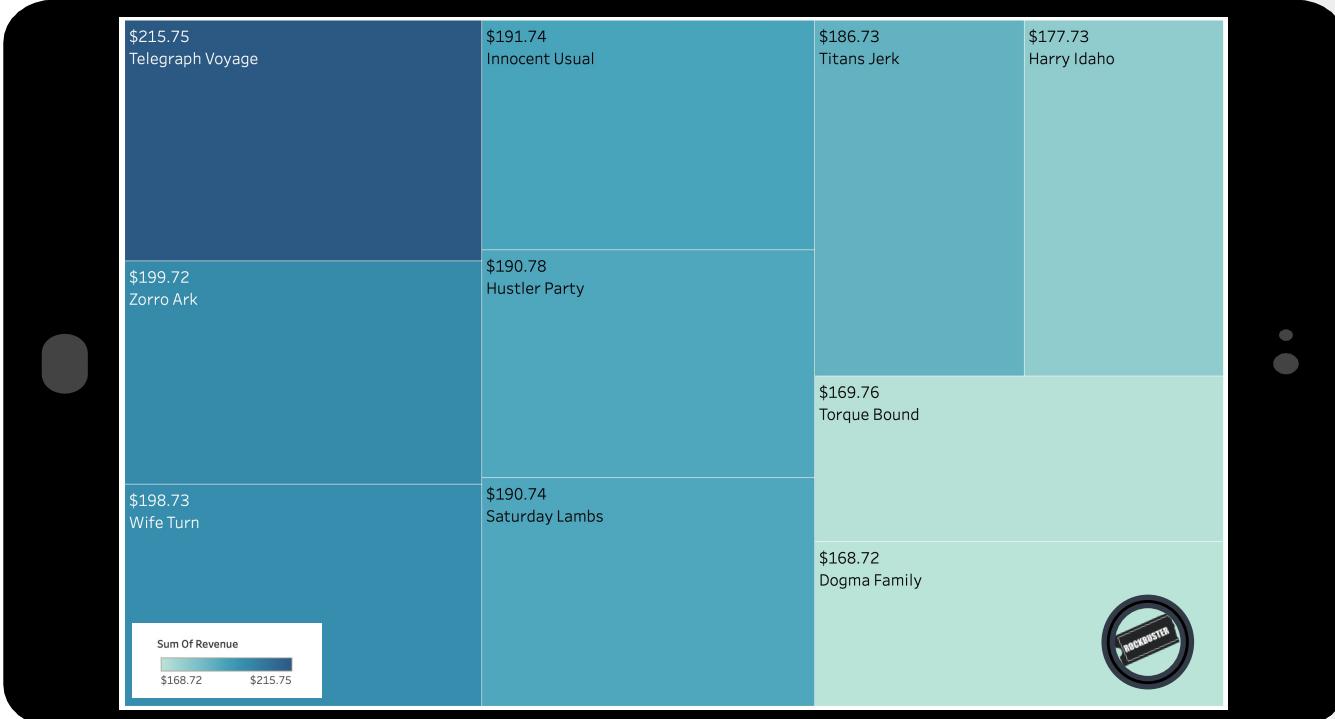
Your  
phone  
Your  
favorite  
movies!



- Sports, Animation & Drama genres are the best Sellers.
- Thriller, Children's and Music are the lowest money makers.
- By adding to the library variety of our highest revenue genres, we can increase profits!

Word Map Showing Highest and lowest Revenue Generators for Rockstar Stealth LLC

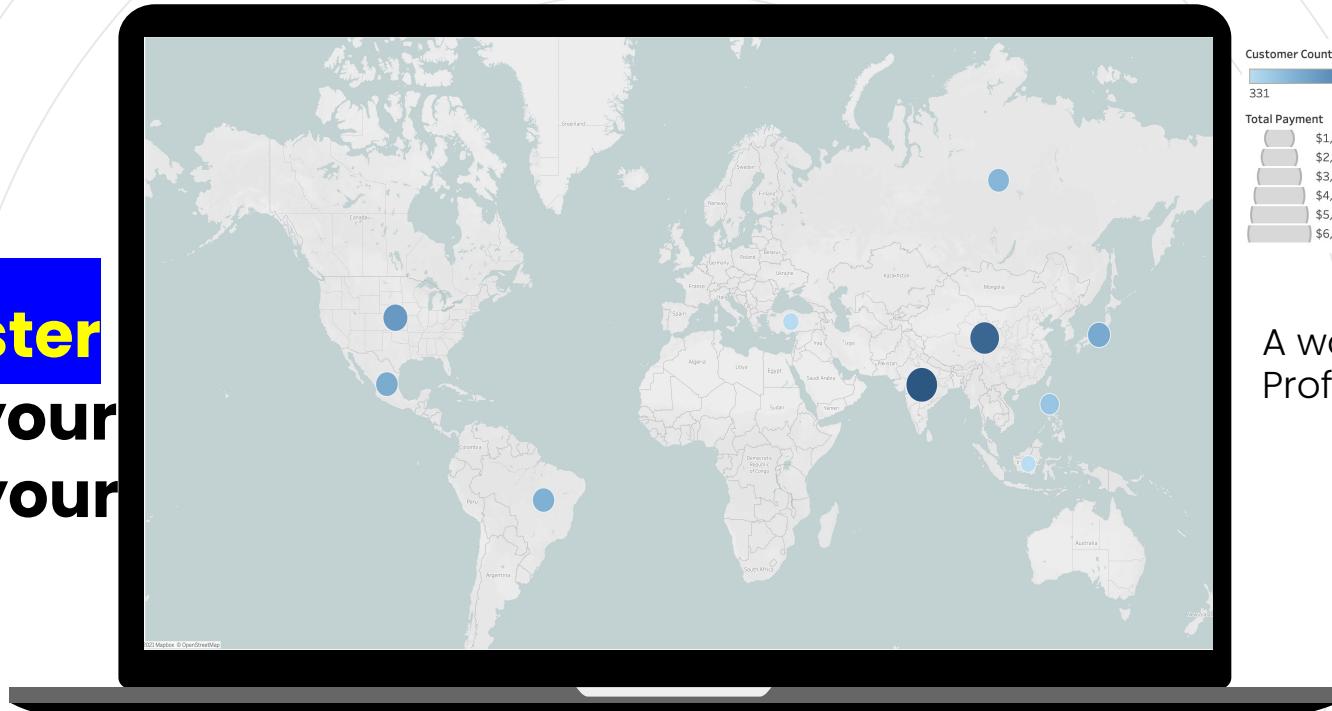
**Rock**Buster****  
gives you the  
Movies you  
want, when  
you want,  
even on your  
tablet!



Word Map Showing Highest Revenue Generating Movies for Rockstar Stealth LLC

- Telegraph Voyage, Zorro Ark & Wife Turn were the highest revenue generators among movie titles.
- Adding similar movie types to our digital library will maximize profits and brand loyalty!

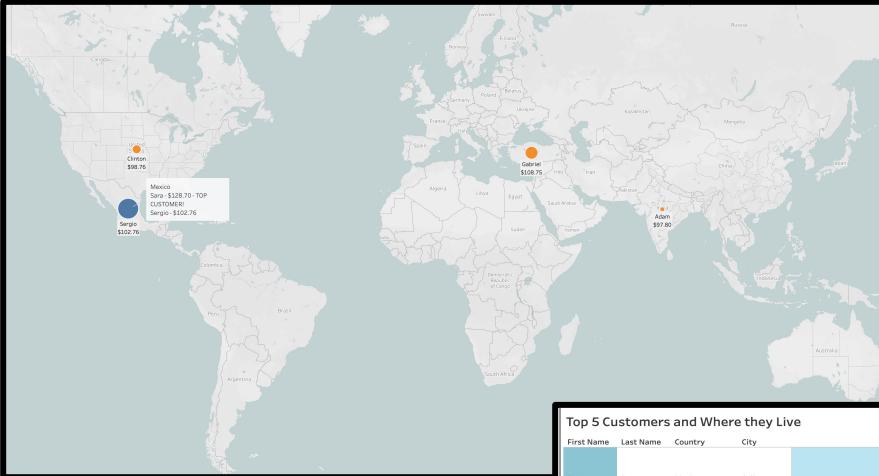
**RockBuster**  
makes your  
Laptop your  
a home  
theater



World Map where darker circle indicates increased payments and larger circle indicates higher customer count.

- Note Asian Countries India and China followed closely by the US make up the highest Paying customers as well as highest customer counts

# Best of the Best!

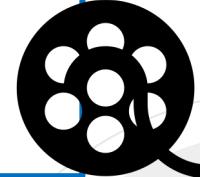


Top 5 Customers and Where they Live				
First Name	Last Name	Country	City	
Sara	Perry	Mexico	Atlixco	\$128.70
Gabriel	Harder	Turkey	Sivas	\$108.75
Sergio	Stanfield	Mexico	Celaya	\$102.76
Clinton	Buford	United States	Aurora	\$98.70
Adam	Gooch	India	Adoni	\$97.80

- Rewarding our top customers will encourage brand loyalty.
- Based on the map, we see that Mexico has 2 of our TOP 5 customers.
- Rewards such as free yearly subscriptions or discounts will ensure our loyal customers do not subscribe to the competition.

# RockBuster Stealth

## Recommendations



- Pivoting towards a digital media platform will not only keep the company with current trends, it will save money in replacement costs in the long run.
- Sports, Animation & Drama genres are the best Sellers. By adding to the library variety including the highest revenue genres, we can increase profits.
- Movies similar to Telegraph Voyage, Zorro Ark & Wife Turn, the highest revenue generators, must be added to the digital library to maximize profits and brand loyalty.
- Increased advertising in high Population Asian countries (India and China) followed closely by the U.S. who spend the most on company products will increase revenue even further.
- Rewarding top tier customer with free yearly subscriptions or discounts will ensure our loyal customers do not subscribe to the competition and encourage other potential customers to subscribe to our platform. The sales team needs to know what the busiest days of the week and hours of the day are (i.e., the days and times with the most orders) in order to schedule ads at times

I  
Heart  
**Movies!**

**Safe!**

**Reliable!**

**Convenient!**



**Easy!**

**Fun!**

The data is a combination of multiple open-source data sets from Instacart. A customer data set was created and included for the purpose of this project and does not reflect the actual Instacart company.

[CUSTOMER DATA SET](#)

[DATA DICTIONARY](#)



Data wrangling



Data Merging

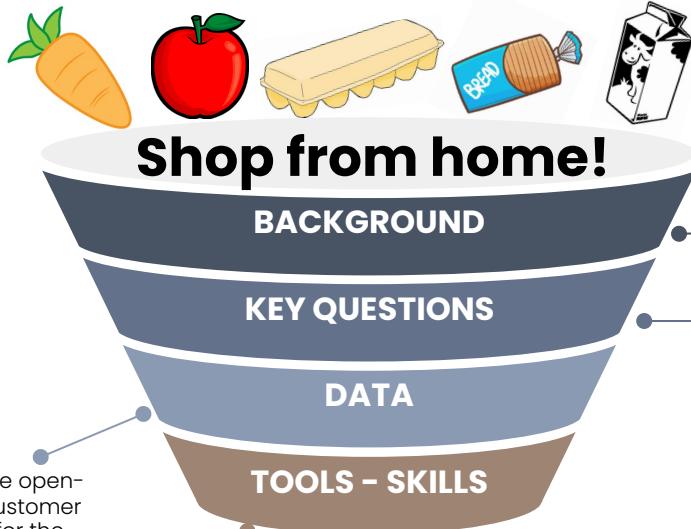
Deriving

Variables

Grouping Data

Aggregating

Data Visualizations



An online grocery company, instacart\* is in need of exploratory data analysis to determine temporal sales patterns, customer purchasing habits, and department trends in order to create a targeted marketing campaign.

+ The sales team needs to know what the busiest days of the week and hours of the day are (i.e., the days and times with the most orders) in order 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.

+ Instacart has a lot of products with different price tags. 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? The marketing and sales teams want to know which departments have the highest frequency of product orders.

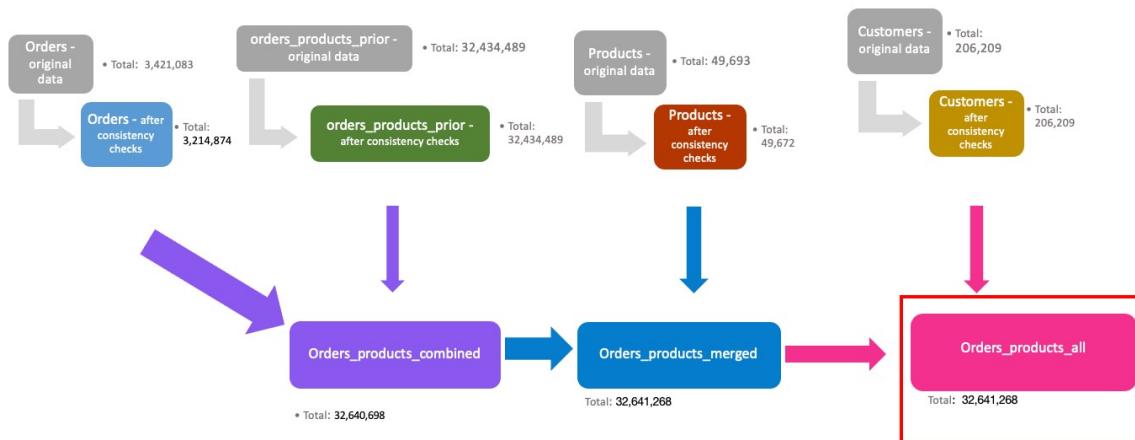
+ The marketing and sales teams are particularly interested in the different types of customers in their system and how their ordering behaviors differ.

\* Instacart is a real company that has made their data available online. However, the contents of this project have been fabricated for the purpose of this Achievement.



# InstaCart Analysis

## Population flow



The flow chart above represents the original data and the process of exclusion and inclusion of data within the Jupyter notebook. The grey boxes represent original data. The second row of multiple colored boxes represents the data sets post-data checks and additions/subtractions including removal of missing values and duplicates for data integrity. Lastly, the third row is the merging of multiple data sets.

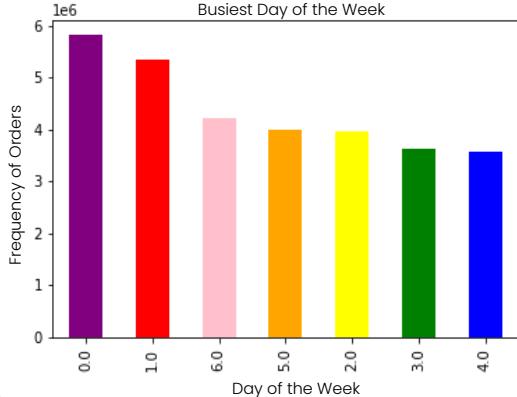
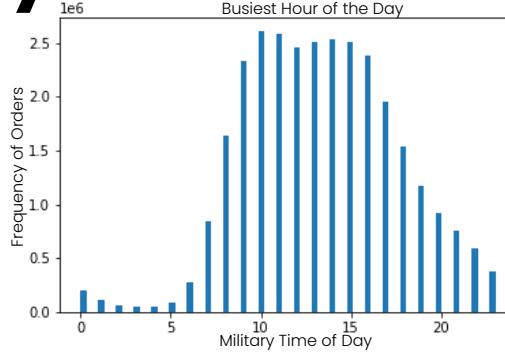




# Analysis



By calculating the busiest hours of the day and busiest days of the week we can place targeted ads for those who shop more during this period and offer deals for those days when it is not so busy. The bar chart on the right indicate that orders increase close to 7 AM and eventual decline close to 7 PM.



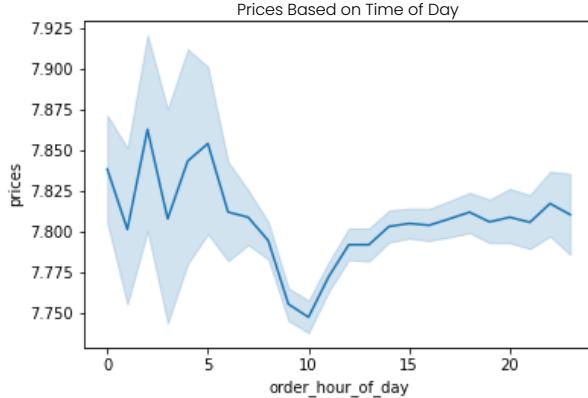
Note the bar chart on the left indicates the busiest days of the week being Saturdays and Sundays. Utilizing this data analysis we can offer greater rewards and offers during less busy days which are shown to be Wednesdays and Tuesdays.

Days Of the Week: 0 = Saturday, 1 = Sunday, 2 = Monday, 3 = Tuesday, 4 = Wednesday, 5 = Thursday, 6 = Friday

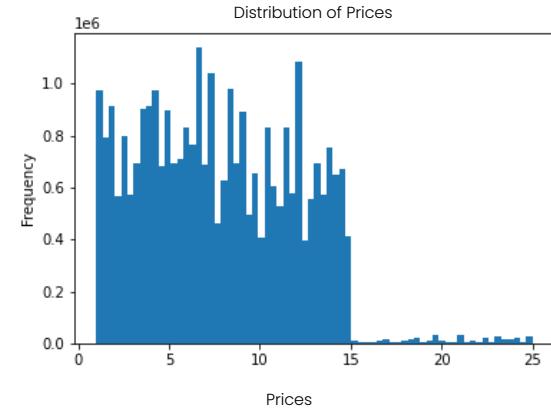




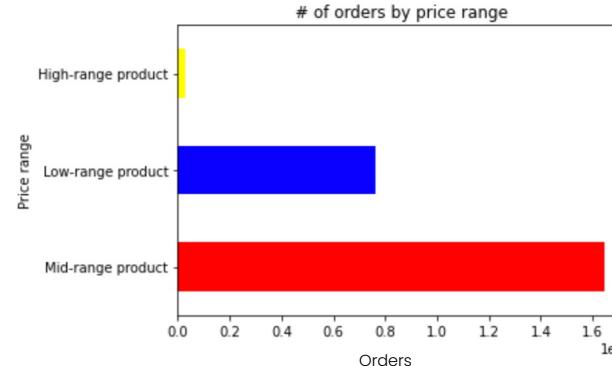
# InstaCart Analysis



The line chart to the left indicates decreased spending at 10 & 11am. Increased spending is noted between 2 & 7am. Expenditures increase again in the late afternoon and evening. Based on the analysis, pop-up sales catering to evening and lunch hours should be offered to encourage sales. Increased sales are occurring during morning hours. Rebates/deals need to be offered for lunch, afternoon and evenings hours.



The histogram to the left indicates that most products bought are priced between \$1 and \$15.

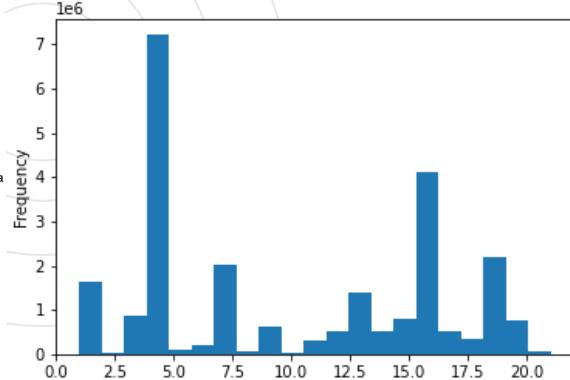


The bar chart to the left indicates that the MID-RANGE (Cost of \$10-\$15) products have an increased amount or orders when compared to high range products, or items costing greater than \$15. High-range products are suffering in sales and need to be coupled with mid-low-range products for combo deals. Taking % rebates, sales off of High-Range products or offering alternatives when buying Mid-range products will increase sales.

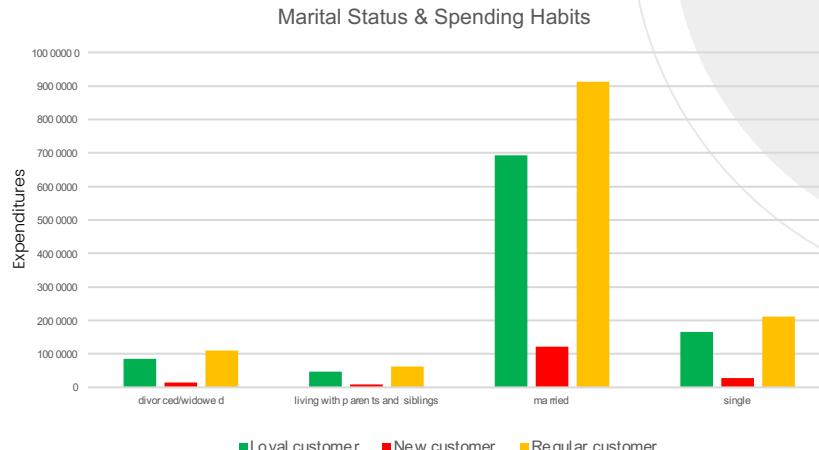


# InstaCart Analysis

Legend
Dept ID
1 frozen
2 other
3 bakery
4 produce
5 alcohol
6 international
7 beverages
8 pets
9 dry goods pasta
10 bulk
11 personal care
12 meat seafood
13 pantry
14 breakfast
15 canned goods
16 dairy eggs
17 household
18 babies
19 snacks
20 deli
21 missing



The histogram at left indicates that most popular department is produce followed closely by dairy eggs. We can also see the least popular departments that include bulk and frozen items. Bulk, pets, alcohol and international item sales are suffering, again coupling rebates when customers buy higher ordered items such as produce and dairy can increase sales and awareness that other items exist in the Instacart product offerings. People will buy items if you make them aware products do exist by offering rebates on coupled lows ordered items.



As the bar chart visualization to the right shows, there is a significant representation of married customers with kids who make up most the orders. Customers who are married with kids have higher expenditures with a higher cost of living due to increased individuals within the family circle. Increased ads need to be directed with offerings that apply to the other brand loyalty representation among the different section/regions of the U.S. market.



# Roadmap To Online Retail

Success 

Pop-Up flash Sales, Instant Rebates and double rewards need to be offered during decreased sales hours and days.

Lower ordered items need more revenue generation, targeting ads with sales on these items will increase sales.

Coupling rebates when customers buy higher ordered items such as produce and dairy can increase sales and awareness that other items that exist in the Instacart product offerings.

Taking % rebates, sales on of High-Range products or offering alternatives when buying Mid-range products will increase sales significantly.

People will buy items if you make them aware products do exist by offering rebates on coupled lower ordered items.

Customers who are married with kids have higher expenditures with a higher cost of living due to increased individuals within the family circle. Increased ads need to be directed with offerings that apply to the other brand loyalty representation among the different section/regions of the U.S. market.



# The Story of COVID-19

## Background

A novel strain of coronavirus – SARS-CoV-2 – was first detected in December 2019 in Wuhan, a city in China's Hubei province with a population of 11 million, after an outbreak of pneumonia without an obvious cause. The virus has now spread to a vast majority of countries and territories across the globe and is characterized as a pandemic by the World Health Organization. COVID-19 took a toll on the world population since the first case was found.

## Techniques/Approach

In the study we look at how the world been effected by and adjusted to the COVID-19 pandemic. From the increased amount of deaths, to the effect the virus has had on the world's economy, we look at all aspects of the disease – Past – present – future to get an idea on what we did wrong and what we need to do right to ensure innocent people do not lose their lives to such a deadly diseases.

## Tools/Skills

- Exploratory analysis through visualizations (scatterplots, correlation heatmaps, pair plots and categorical plots)
- Geospatial analysis using a shapefile - Regression analysis
- Cluster analysis - Time-series analysis
- Multi-visualization studio including dynamic dashboards, custom storyboard designs and organization.
- Organization of visuals, graphs, tables and presentation in an understandable format using gifs, animations and word art.

## Key Questions

- Which Countries have the highest amount of Deaths due to the virus ?
- Where are the highest incidence of COVID-19 populations?
- How has the pandemic effected the world's economy?
- What is the future of the COVID-19 if left un-checked?

## The Data

[COVID-19 Dataset](#)

[Unemployment Data](#)

[GitHub Repository](#)

[Tableau Presentation Link](#)



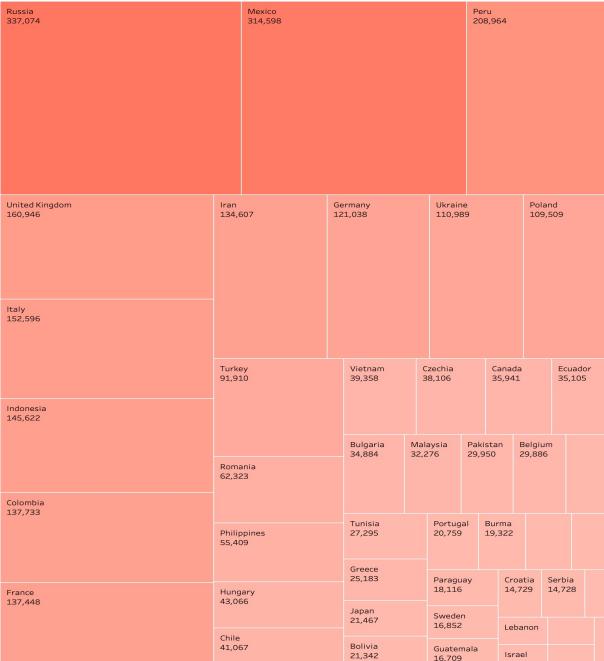


# The Story of COVID-19

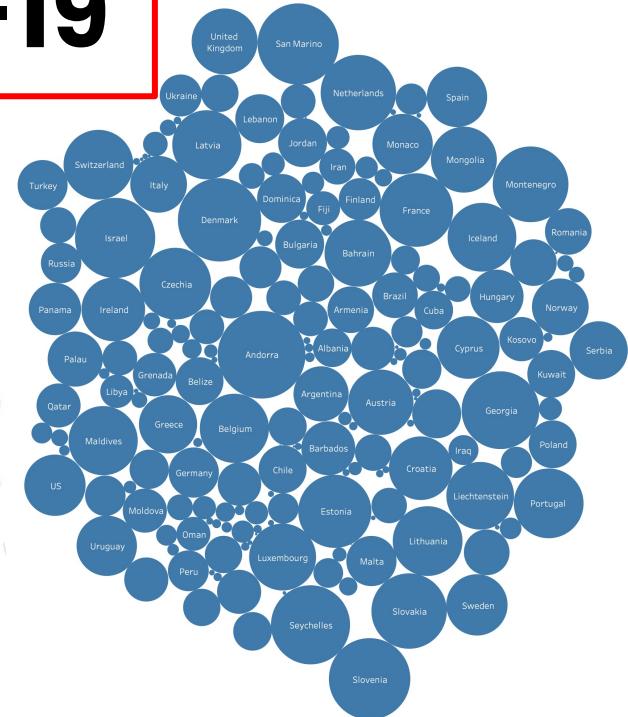
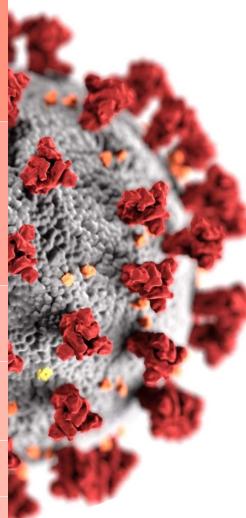
US  
953,513

Brazil  
643,277

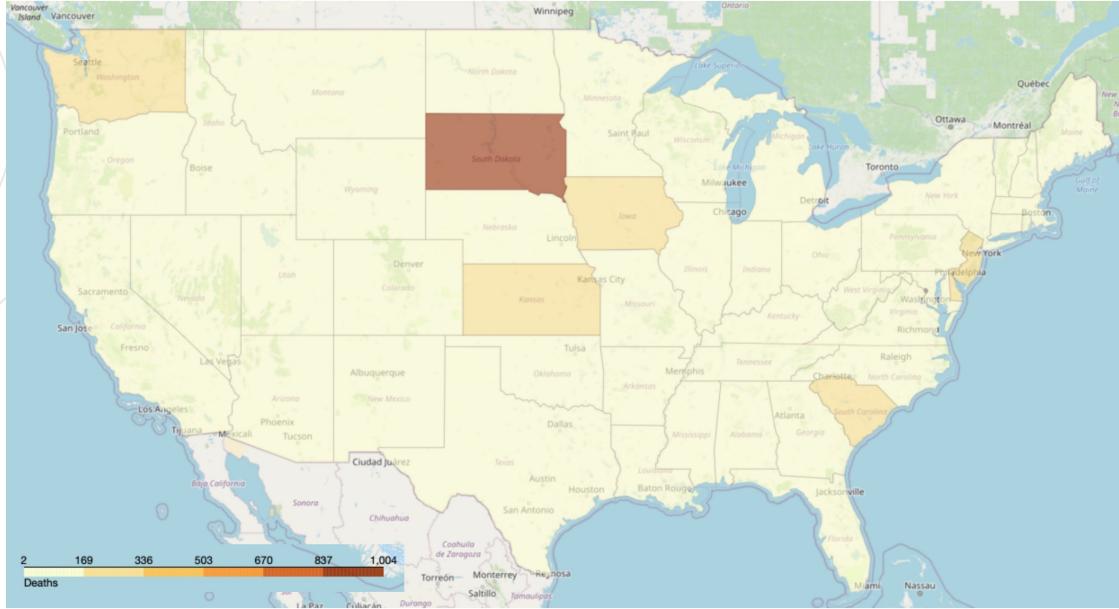
India  
510,905



The word map above shows the countries that have suffered the most in terms of deaths since the start of the COVID-19 pandemic to today. USA, Brazil and India stand out as the highest number of deaths overall and places where economies have suffered just as much as the people.

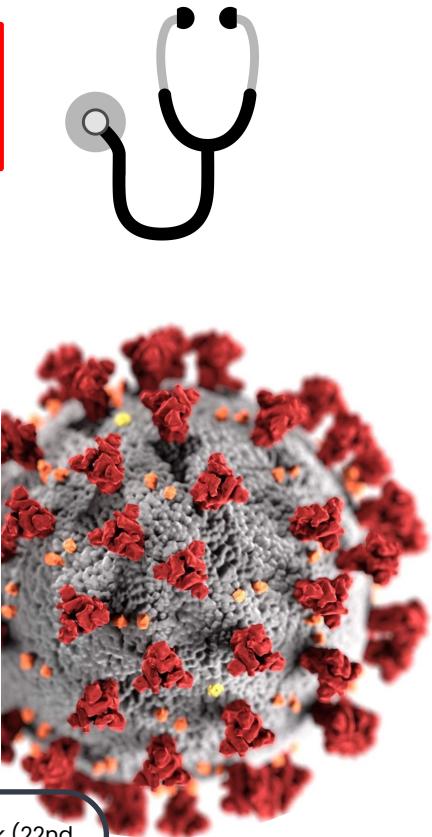


# The Story of COVID-19



The python generated *choropleth* map above indicates that during the early stages of the the COVID-19 outbreak (22nd January to 27th July 2020) within the U.S. where South Dakota, Iowa, Kansas, Washington, New Jersey and South Carolina were among the hardest hit in terms of mortality count.

Based on this information, we see that the U.S. was not severely affected with the Pandemic but the spread of COVID-19 virus was rampant with no safety controls in place.

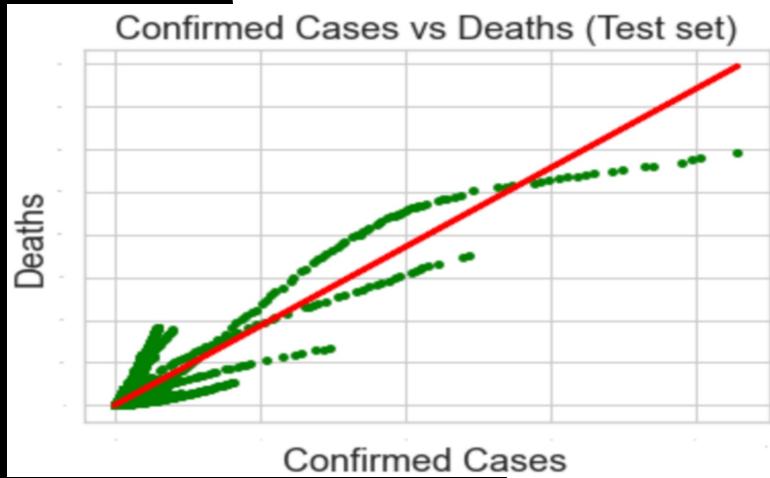
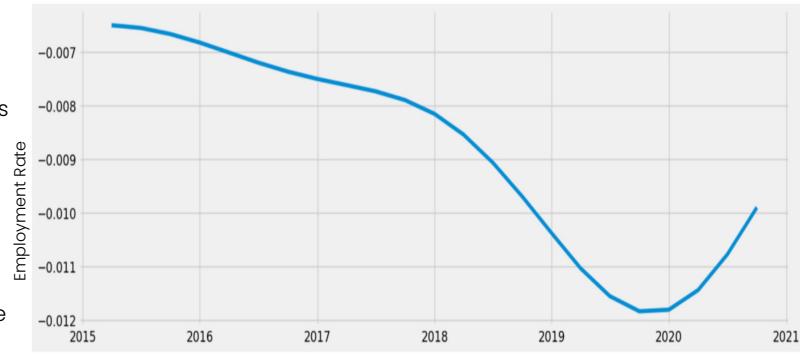


# The Story of COVID-19

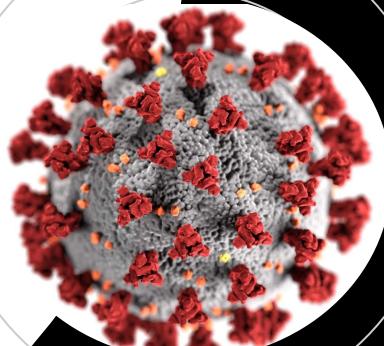


By doing a time series analysis using python, we can study the effect COVID-19 has had on the un-employment rate in the world. The un-employment rate, one of the indicators of economic status among various countries of the world, has seen a upward trend since the start of the pandemic in late 2019/early 2020.

Although the un-employment data was showing a (-) negative trend before the pandemic as seen in the time series plot on the right, we clearly note an upward curve due to the COVID-19 pandemic.

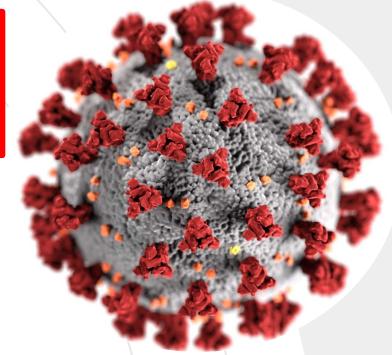


Looking to the future of the pandemic, using a supervised machine learning linear regression model, trained and tested in python, we can see that if enough people get tested and have confirmed (+) COVID-19 tests, and increased number deaths will result. Although this does not prove a cause and effect relationship it does gives us the opportunity to makes the necessary changes such as wearing masks, vaccinating and following safety guideline set by the CDC to "curve the spread" of this deadly killer.





# The Story of COVID-19



The Covid-19 pandemic has taken a heavy toll on the human population overall in the world.

Countries that lack up-to-date medical supplies and technologies as well as hospital access are the ones that have the highest mortality rates.

In the early stages of COVID-19 spread, the U.S. noted near exponential increases in cases leading to deaths.

The economic toll that COVID-19 took on the world economy is significant leading to decreases in jobs, decreased incomes and a overall down-turn of consumer spending.

If left un-checked, the future battle against controlling the COVID-19 pandemic looks bleak in terms of the potential deaths the virus can cause.

Virus awareness and education is still needed in countries that lack proper care and preventative measures to help further "curve the spread" and decreases possible mutations that may "escape" the vaccination protocol.



UNLEASH YOUR **DATA**  
UNLEASH YOUR **POTENTIAL!**

*Thank you!*