

SMRITI BHATTARAI

Data Analyst Portfolio



Analytical Skills



- Data Visualization
- Data Cleaning
- Problem Recognition
- Problem-Solving
- Querying
- Statistical Programming

Tools











- Excel
- Power Point
- Tableau
- SQL
- Python

Soft Skills



- Collaborative
 - Communication
 - Analytic
 - Curious
 - Adaptability
-

TABLE OF CONTENT

S.NO	Projects	Slide No.	Tools Used
1	Video Game Sales	21-25	
2	Preparing for Influenza Season	15-20	 
3	Rockbuster Stealth	10-14	  
4	Instacart Basket Analysis	4-9	 
5	Covid-19 Analysis	26-31	 

INSTACART BASKET ANALYSIS



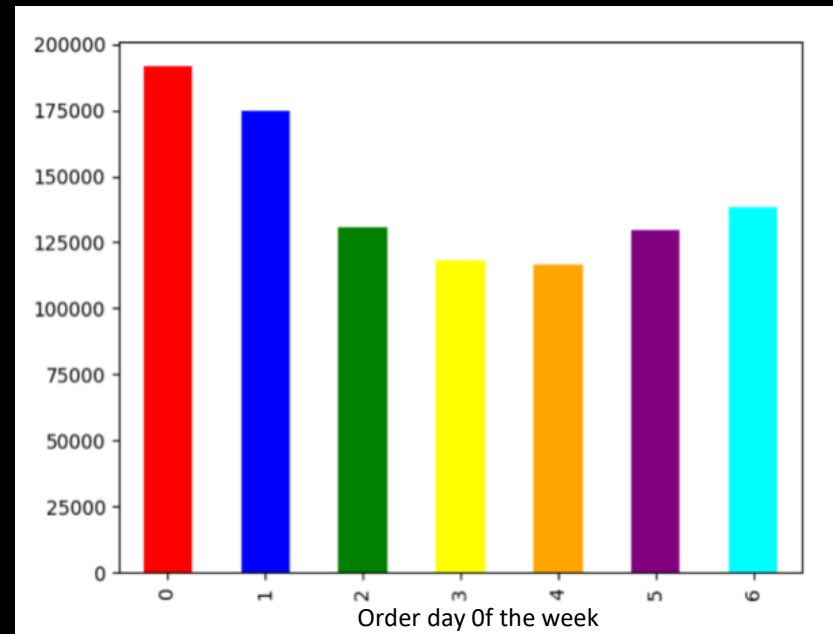
PROJECT OVERVIEW

- **Objective:** Instacart, an online grocery store that operates through an app. Instacart already has very good sales, but they want to uncover more information about their sales patterns. I performed an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria
- **Key Questions:**
 - *What are the busiest days and hours of the week?*
 - *Are there particular times of the day when people spend the most money?*
 - *How can we price range groupings be simplified?*
 - *Are there certain types of products that are more popular than others?*
 - *What different types of customers can be identified and how do their ordering behaviors differ?*

Skills
Data Wrangling
Data Consistency Checks
Combining Data
Grouping & Aggregating Data
Data Visualization with Python

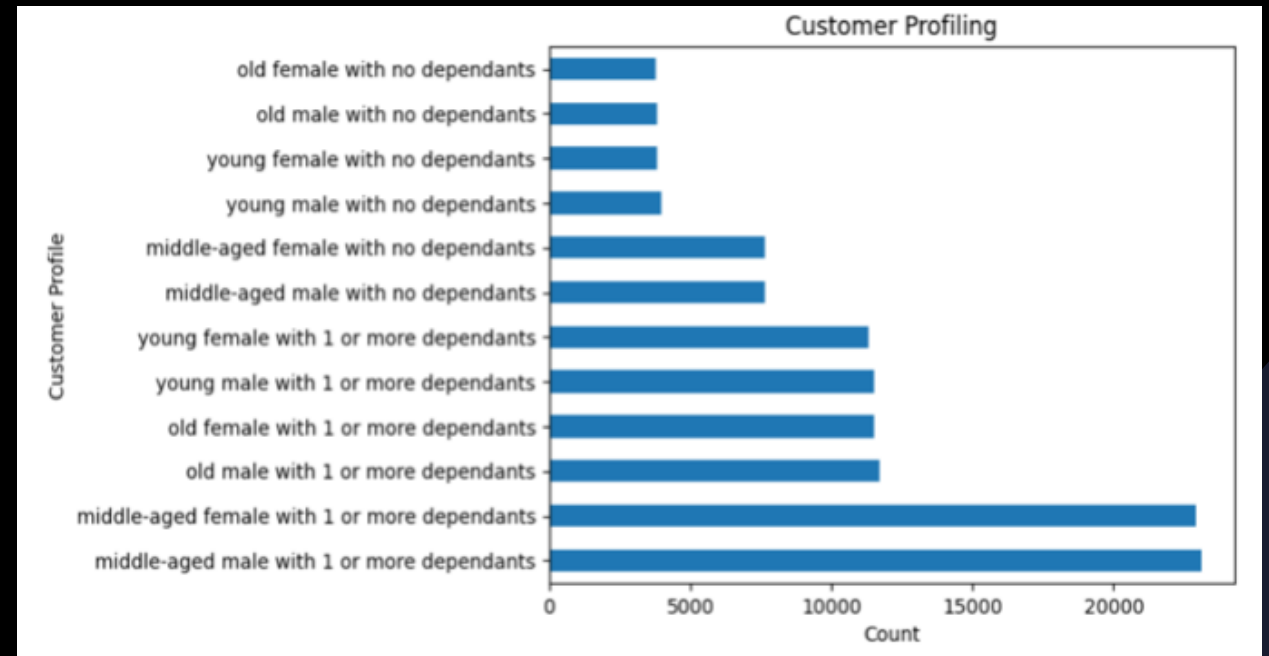
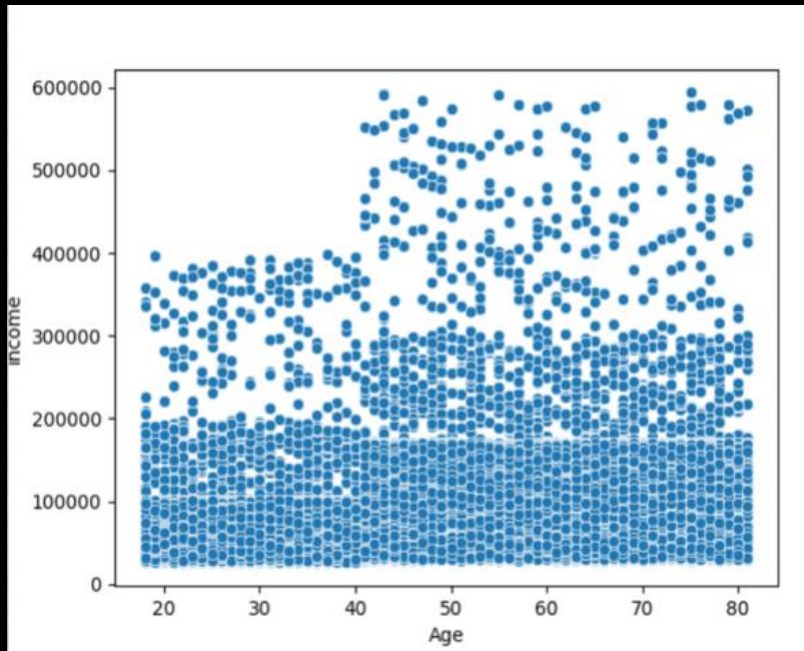
INSTACART BASKET ANALYSIS

Histogram and bar chart showing the prime timeframes for order placement



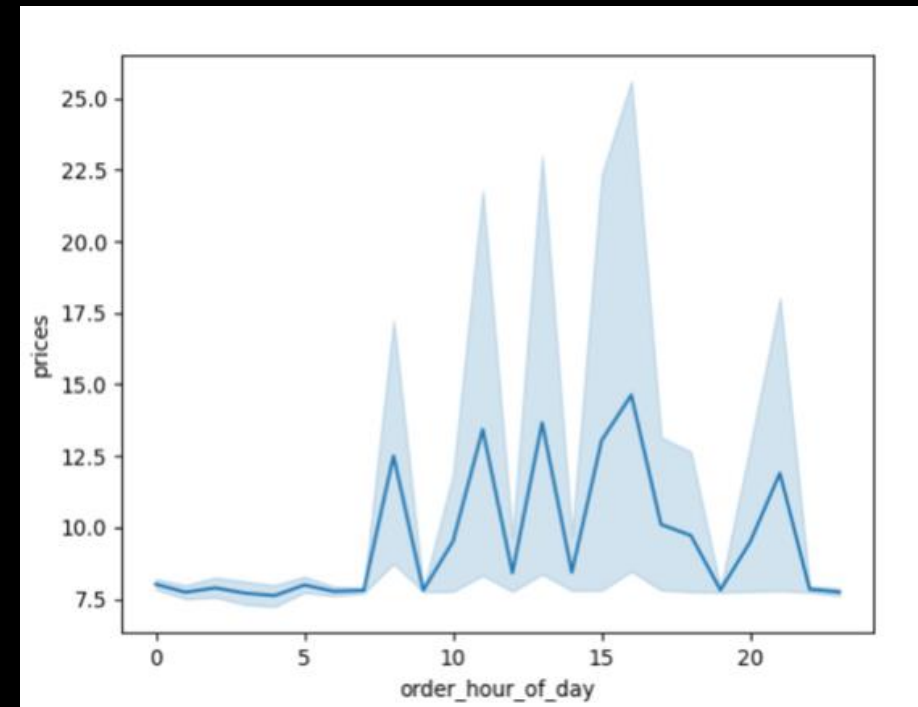
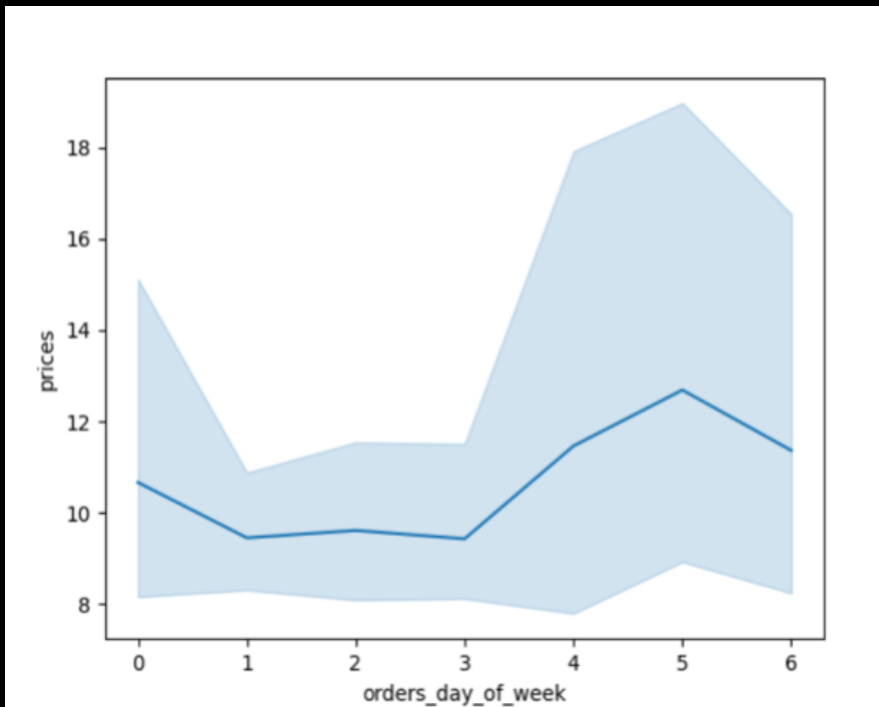
INSTACART BASKET ANALYSIS

Correlation between age and income along with customer profiling



INSTACART BASKET ANALYSIS

Pricing Comparison with orders day of week and order hour of day



RECOMMENDATION

1

The sales team's strategic advertising efforts can be greatly optimized by understanding the peak days and hours of order placement. According to the order_hour_ofday histogram, customers predominantly place orders between 9 AM to 4 PM, emphasizing the importance of targeting ads during these hours. Additionally, the bar chart paints a clear picture: Saturdays and Sundays witness the highest order volumes, showcasing a strong inclination towards weekend shopping among customers.

Armed with this information, the sales team can schedule ads during off-peak hours, maximizing their impact when order placements are comparatively lower, ensuring efficient allocation of resources and a more effective outreach strategy.

2

The most expensive items are predominantly purchased around 3 a.m., marking a unique window for high-value transactions. Notably, aside from these outliers, customer spending remains relatively consistent throughout the rest of the day.

In the early morning hours, it is advisable to present premium and high-value items with targeted advertisements. By aligning product offerings with customer spending habits, the sales team can build targeted campaigns, maximizing revenue potential during these specific time periods.

3

The comprehensive range of products available on Instacart caters to a diverse customer base with varying budgets. The histogram analysis reveals a compelling trend: the majority of products fall within the price range of \$2 to \$15. This insight serves as a valuable guide for marketing and sales teams, indicating that customers are particularly inclined towards affordable and moderately priced items.

4

Recognizing the clear trends in customer preferences, retailers have a unique opportunity to refine their marketing efforts. By concentrating promotional activities around top-selling departments such as Produce, Dairy & Eggs, Snacks, and Frozen items, businesses can create impactful marketing campaigns tailored to meet customer demands effectively.

5

Cultivating loyalty among our current clientele is extremely important. We recommend launching personalized initiatives targeting new customers, enticing them with attractive offers to facilitate their initial purchases and familiarize them with the benefits of our loyalty program. For regular customers, strengthening their loyalty is key; this is achieved through exclusive offers, personalized product offers and a multi-level loyalty scheme that recognizes their continued support. Loyal customers deserve special attention; We continue to enhance their experience with exclusive incentives and early access to promotions, ensuring they feel valued and appreciated.

6

Targeted marketing campaigns: It is necessary to develop marketing campaigns tailored to different age groups. Develop messages and promotions that align with the financial capacity and preferences of individuals in this age group. Highlight products and services that meet their specific needs and desires.



ROCKBUSTER STEALTH ANALYSIS

PROJECT OVERVIEW

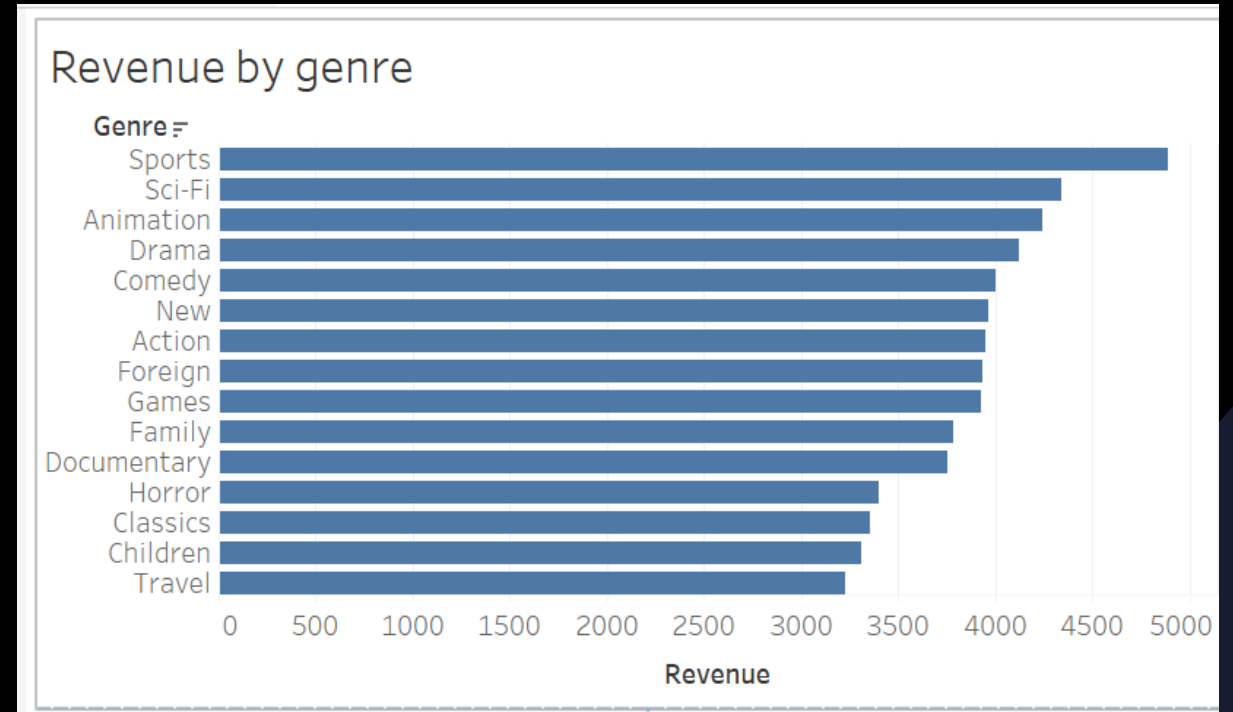
- **Overview:** Rockbuster Stealth LLC is a movie rental company that used to have stores around the world. Facing stiff competition from streaming services such as Netflix and Amazon Prime, the Rockbuster Stealth management team is planning to use its existing movie licenses to launch an online video rental service in order to stay competitive
- **Key Questions and Objectives:**
 - The Rockbuster Stealth Management Board has asked a series of questions, and they expect data-driven answers that they can use for their 2020 company strategy. Here are the main questions:
 - Which movies contributed the most/least to revenue gain?
 - What was the average rental duration for all videos?
 - Which countries are Rockbuster customers based in?
 - Where are customers with a high lifetime value based?
 - Do sales figures vary between geographic regions?

Skills
Database Querying
Joining Tables
Common Table Expressions
Subqueries
Creating a Data Dictionary
CRUD Operations

REVENUE ANALYSIS

Sports, Sci-Fi and Animation are top three popular genre

```
SELECT name, SUM(A.amount) AS "total amount paid"  
FROM payment A  
INNER JOIN rental B on A.rental_id = B.rental_id  
INNER JOIN inventory C on B.inventory_id =  
C.inventory_id  
INNER JOIN film D on C.film_id = D.film_id  
INNER JOIN film_category E on D.film_id= E.film_id  
INNER JOIN category F on E.category_id = F.category_id  
GROUP BY name  
ORDER BY "total amount paid" DESC
```



CUSTOMER ANALYSIS

India, China and U.S. has the highest customer concentration

```
SELECT D.country, COUNT(customer_id) AS  
Total_customers  
FROM Customer A  
INNER JOIN address B ON A.address_id = B.address_id  
INNER JOIN city C ON B.city_id = C.city_id  
INNER JOIN country D ON C.country_ID = D.country_id  
GROUP BY country  
ORDER BY count(customer_id) DESC
```

I also performed subqueries and CTEs in order to answer more complex business questions such as who were the Top 5 customers in any of the top 10 cities who've paid the highest total amounts to Rockbuster



RECOMMENDATION

- 1 Targeting promotions in high concentration countries like India, China, the United States, Mexico, Brazil, and Japan, contributes 39% of total sales.
 - 2 Focusing on genre such as Sports, Sci-Fi, Animation, Drama and Comedy as it contributes 35% of total sales.
 - 3 Prioritizing loyal customers who generate 20% of revenue, often from non-top customer volume customers.
 - 4 Noting successful competition between two stores with different sales strategies, suggesting potential collaboration for mutual enrichment
-



INFLUENZA SEASON ANALYSIS

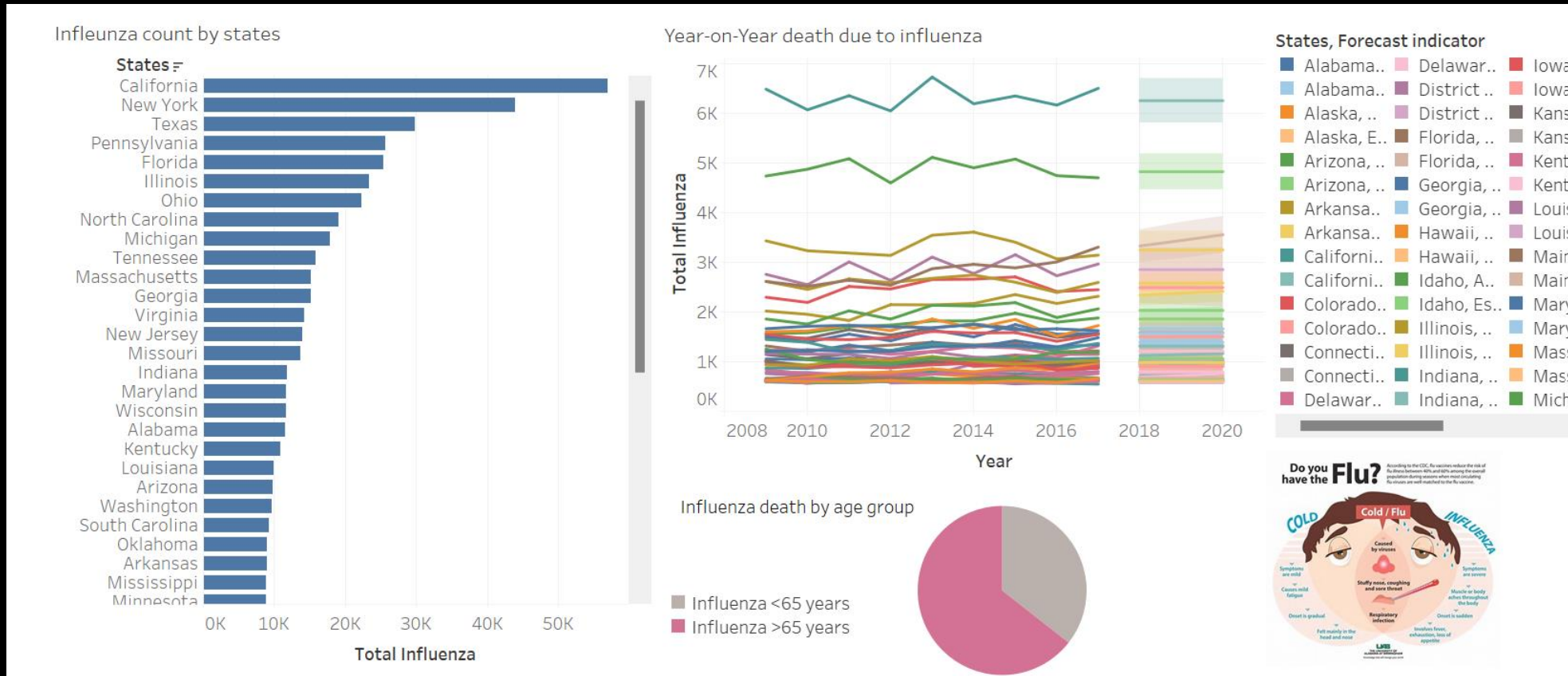
FLU
SEASON

PROJECT OVERVIEW

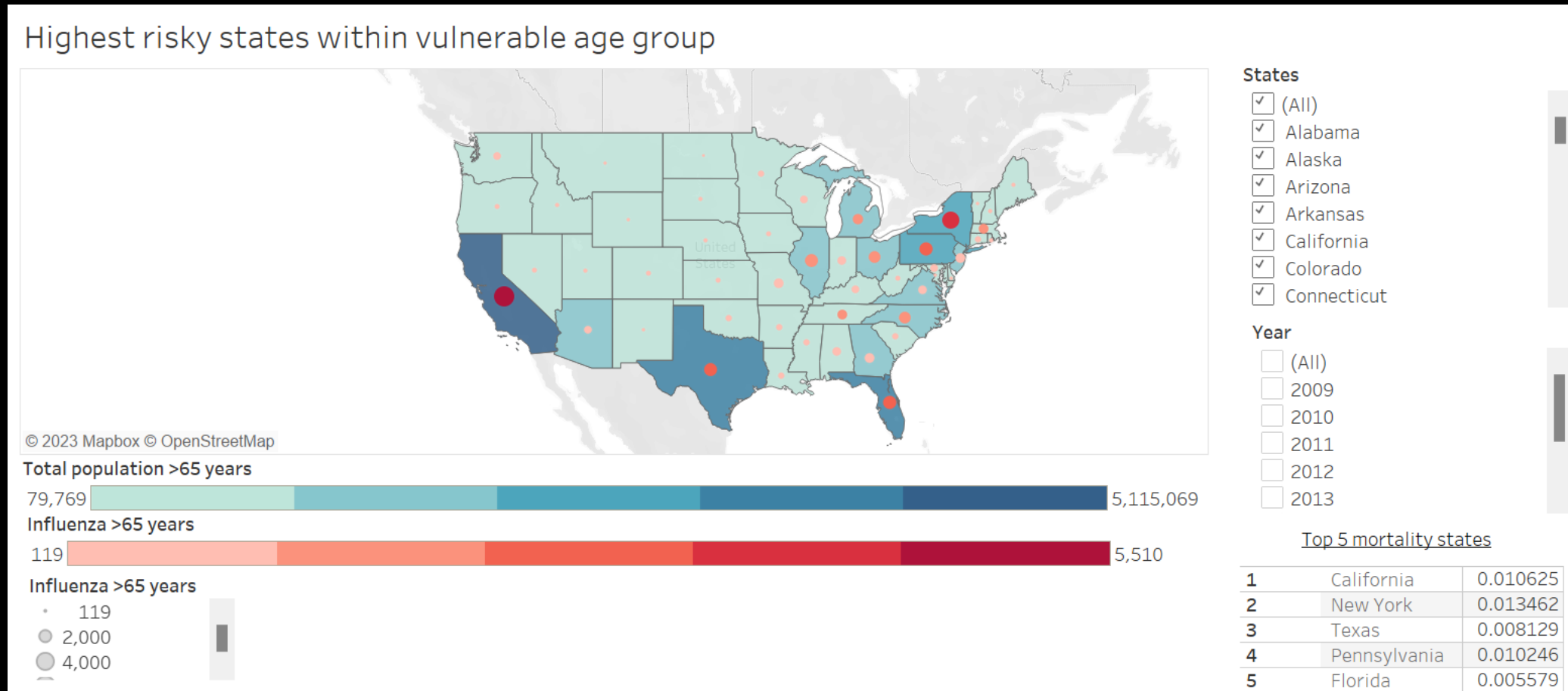
- **Motivation:** The United States has an influenza season where more people than usual suffer from the flu. Some people, particularly those in vulnerable populations, develop serious complications and end up in the hospital. Hospitals and clinics need additional staff to adequately treat these extra patients. The medical staffing agency provides this temporary staff
- **Objective:** Determine when to send staff, and how many, to each state
- **Scope:** The agency covers all hospitals in each of the 50 states of the United States, and the project will plan for the upcoming influenza season

Skills
Data Cleaning
Data Integration
Data Transformation
Statistical Hypothesis Testing
Tableau Visualizations
Presenting Results

INFLUENZA TREND



INFLUENZA IMPACT ON SENIOR CITIZEN



CONCLUSION

Conclusion

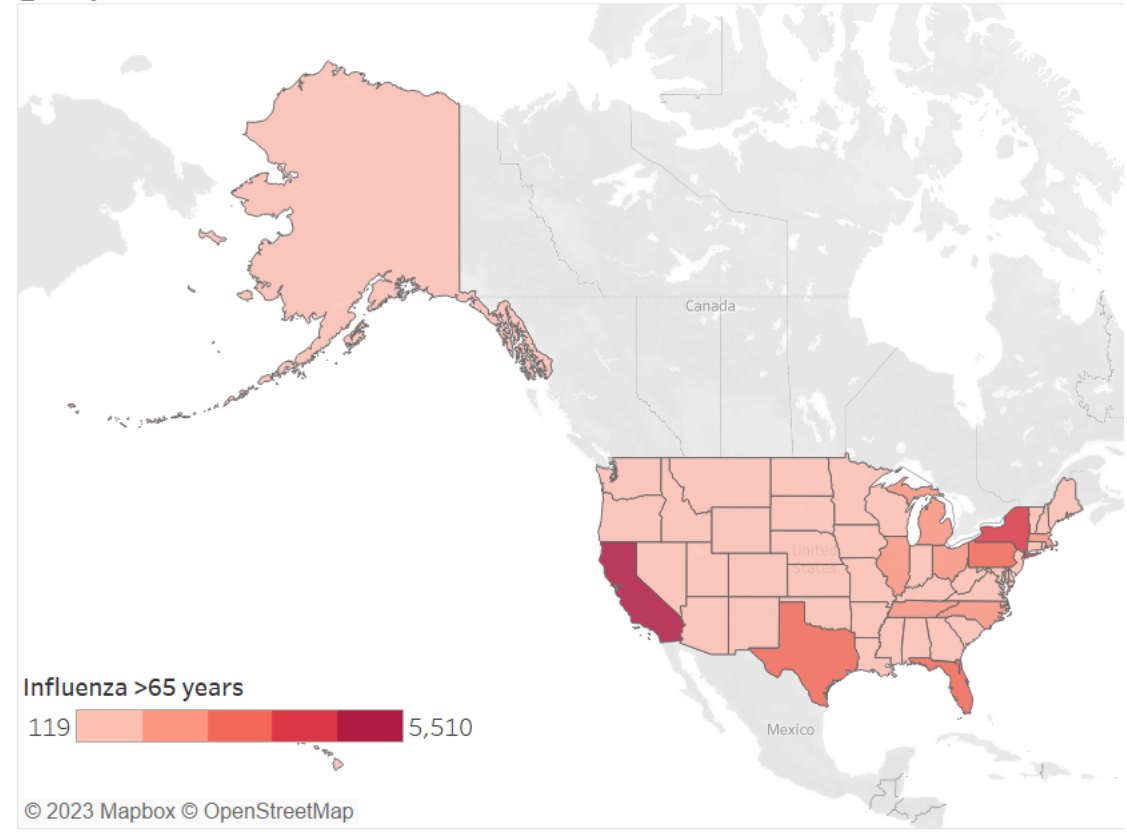
- 1) US states such as California, New York and Texas has the highest influenza death reporting in comparison to other states
- 2) Age group above 65 years are most vulnerable to the influenza season
- 3) And within 65 years age group, states such as California, New York and Texas has the highest death reporting due to influenza, so these states require extra medical staff vs others

Suggestions

The US government should focus on affordable vaccination policy along with enhancing overall patient experience



Medical staff requirement as per death count in vulnerable age group



INFLUENZA SEASON

Limitation:

- I would say, it would have been better if I was able to quantify it and present some fact which at the end would have supported my story or recommendation.
- Some data on staff availability or staff efficiency states wise or hospital wise would have given more better comparison.
- Some data on vaccine effectiveness or vaccination population by each state would have helped build strong story.
- Also, data collected by CDC has chance of potential human error.

Tracking Data: :

- By updated the data yearly would give YOY trend comparison. Also, we would be able to compare it with our projection vs the accurate numbers and see the gap. Further accordingly update the future recommendation
- We could also identify key KPI's which could help monitoring the impact of influenza such as death count and vaccinated population



VIDEO GAME SALES



PROJECT OVERVIEW

- Objectives:

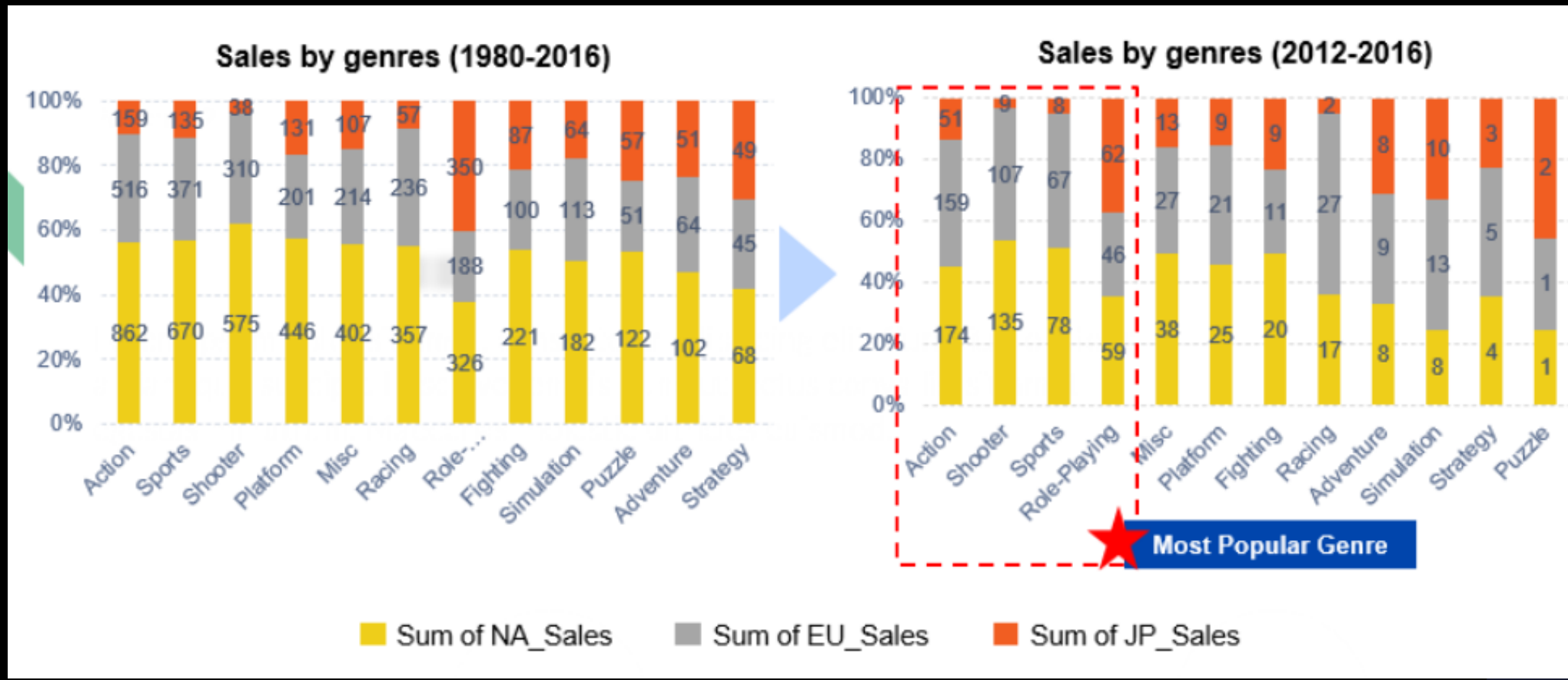
- Perform a descriptive analysis of a video game sales data set to foster a better understanding of how GameCo's (fictional company) new games might fare in the market. Compare assumptions about historical regional sales with the reality of current market conditions
- Compare assumptions about historical regional sales with the reality of current market conditions.

- Key Questions:

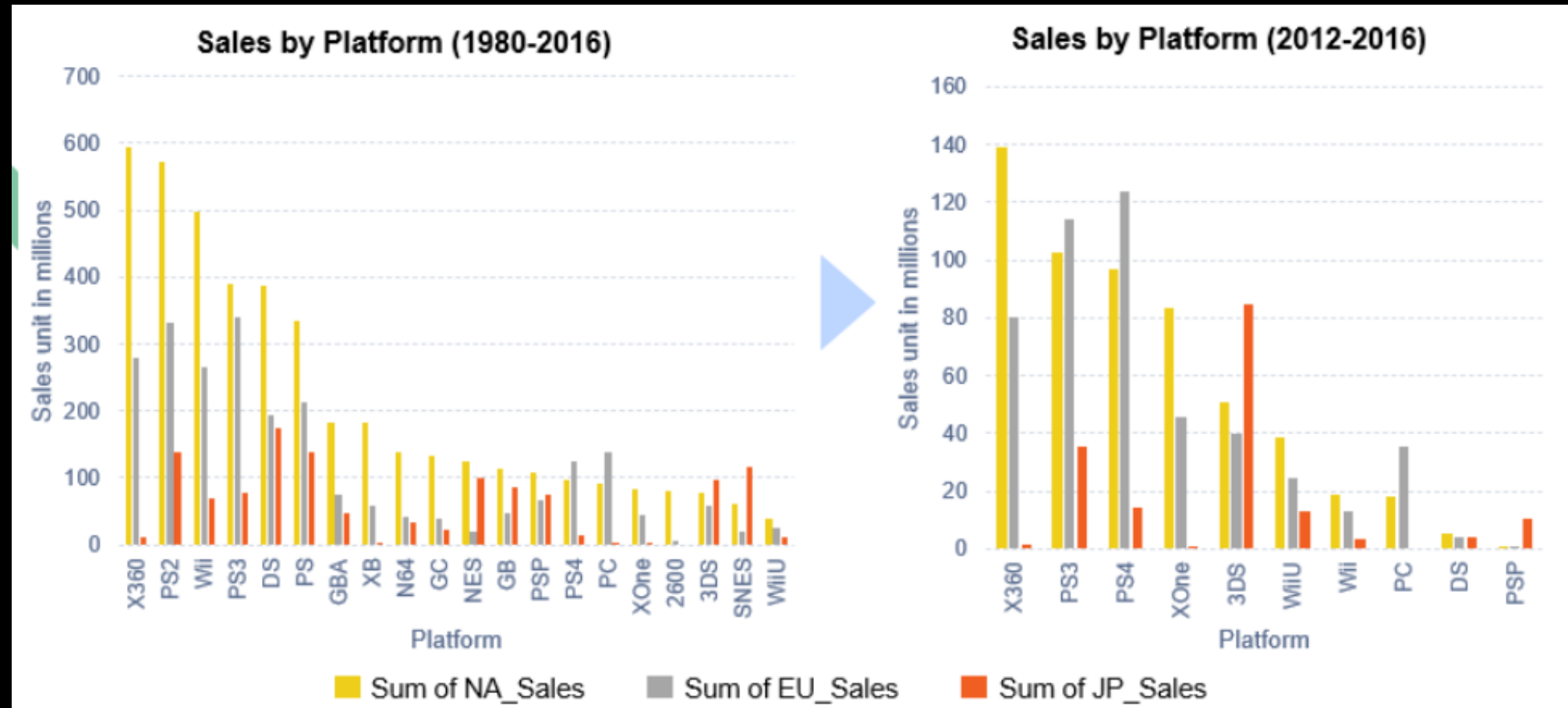
- What are the most popular genres?
- Who are the biggest competitors?
- What are the most popular gaming platforms?

Skills
Descriptive Analysis
Grouping Data
Summarizing Data
Visualizations

What are the most popular genres?



What are the most popular gaming platforms?

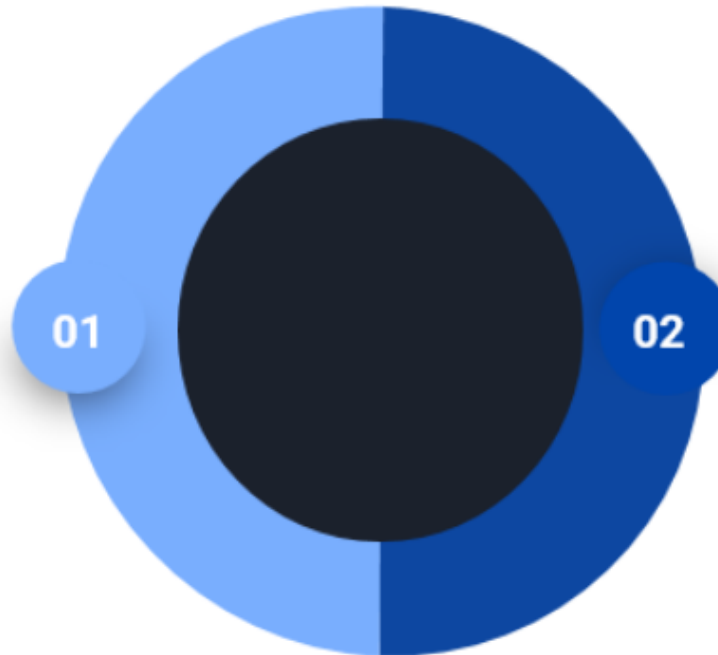


CURRENT SCENARIO AND RECOMMENDTION

Region

Current Scenario: North America is the largest sales contributor and its value has decreased over the years. Whereas EU sales has grown over the years and its contribution has crossed North America contribution. Japan has remained volatile over the years

Recommendation: As sales has not remained same over the years. GameCO will have to redistributed its marketing budget among regions for 2017

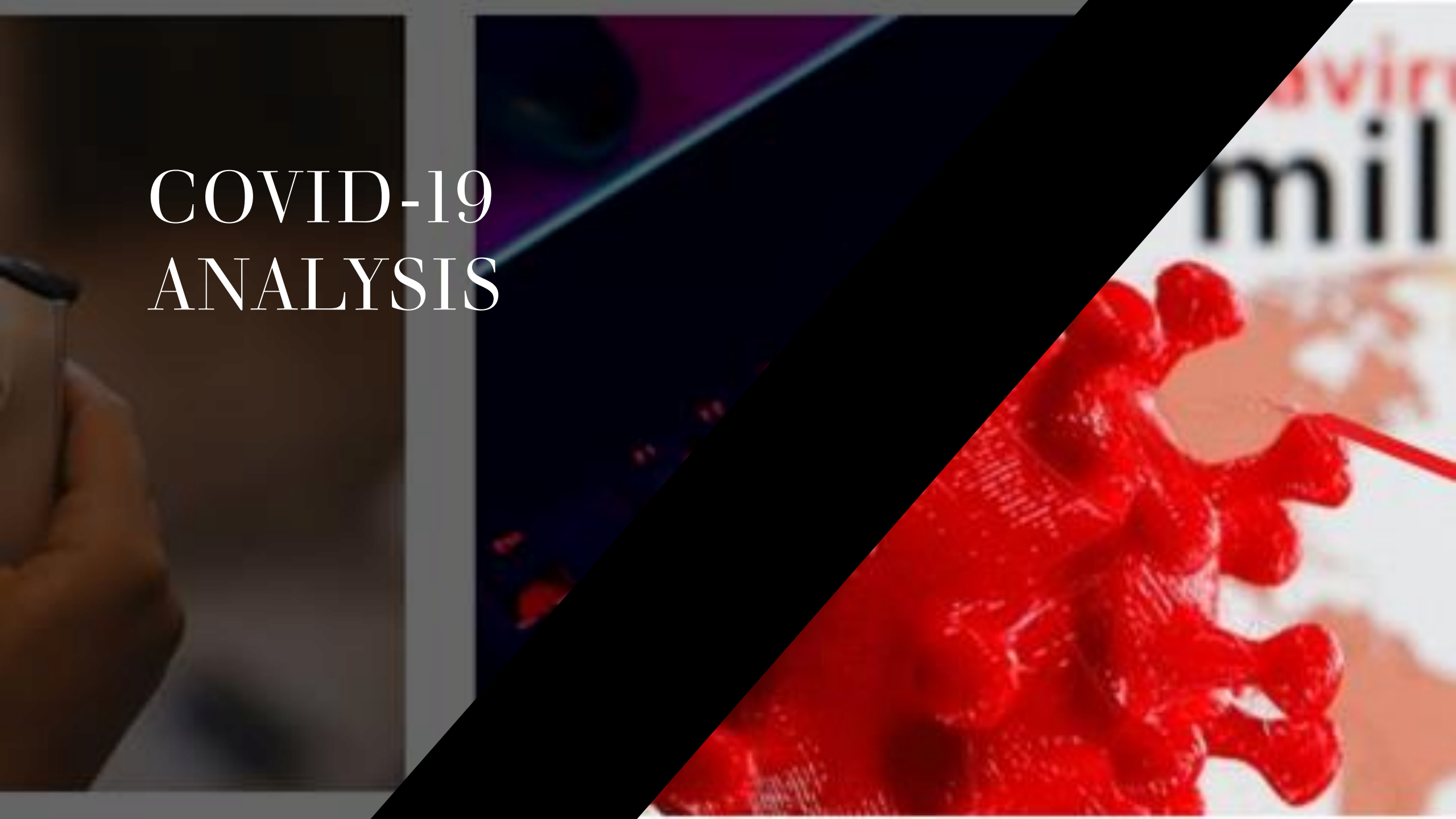


Genre

Current Scenario: Since 1980's Sports, Action and Shooter has been flourishing the gaming industry. Recently, Role-Playing has gain popularity by securing #4 position in the market vs #7 position in 1980's. Trends in North America & EU is similar be it Genre, Platform or Publisher

Recommendation: Sport, Shooter and Role-Playing are popular genres. So GameCo can focus on launching games in these genre as sales will be high compared to other genres. Grand Theft Auto V, Call of Duty, Pokemon & FIFA are high selling games in these popular genres

COVID-19 ANALYSIS



COVID-19 PANDEMIC ANALYSIS

Objectives:

Discover trends, pattern and insights on the COVID-19 pandemic. So that the resource can be allocated

Key Questions:

- Number of total covid cases, deaths, and recovery?
- What continent were most impacted by COVID-19?
- Within continent, what country were most impacted by COVID-19?
- Does population size affect the percentage of cases and deaths?
- Which country had the fastest recovery rate?

Skills:

- Data cleaning and wrangling
- Descriptive analysis
- Geospatial analysis
- Linear analysis
- Time series analysis

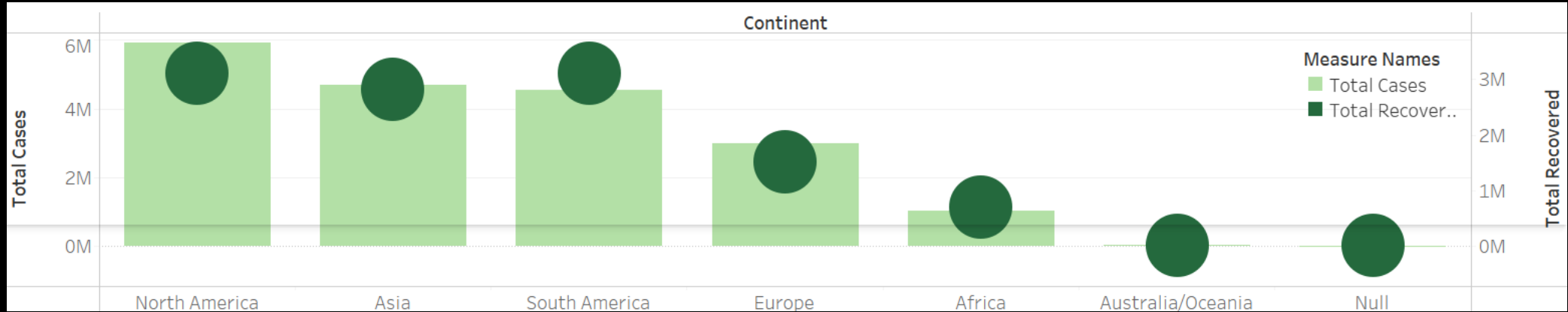
Data:

The source of this data is from Kaggle on a new coronavirus designated 2019-nCoV which was first identified in Wuhan, the capital of China's Hubei province. The data was compiled and scraped from the <https://www.worldometers.info/coronavirus/> website so should be fairly dependable.

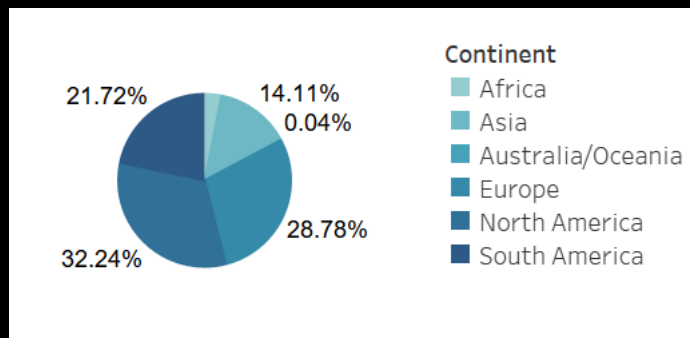
Data Limitation:

Different countries may have varying reporting practices, and there could be delays or discrepancies in the data reported by different sources. Some regions may have more robust reporting systems than others. Also, COVID-19 situation is dynamic, and the status of the pandemic can change rapidly. Therefore, the data available on Worldometer may not always reflect the most current situation.

Total covid cases and recovered cases by continent



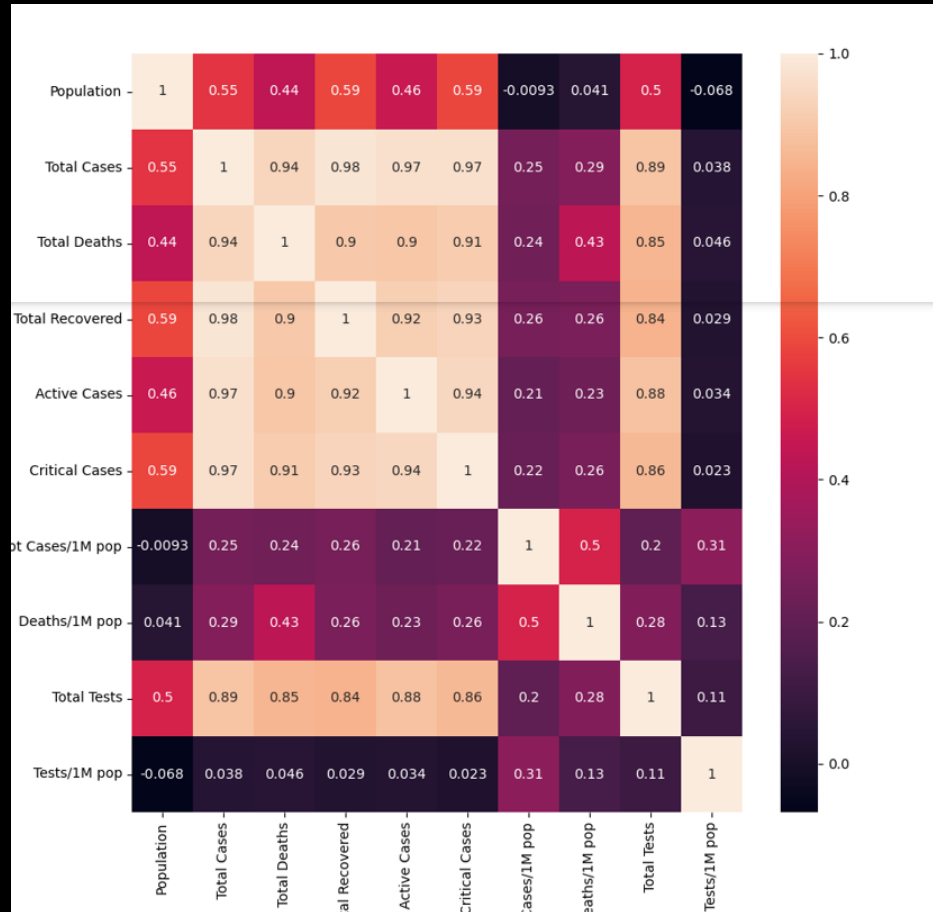
Total death cases by continent



Key Highlights

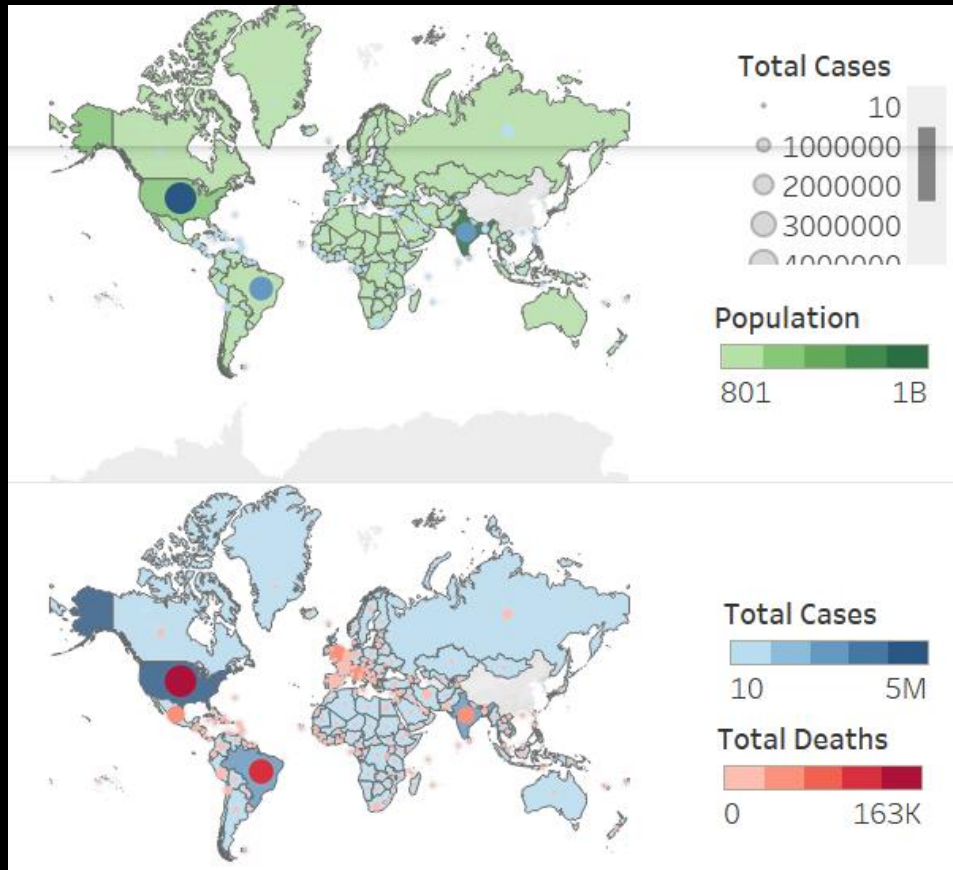
- North America has the highest total COVID cases, followed by Asia and South America. However, the total recovered cases are also high.
- Whereas total death cases wise, North America and Europe are at the top position, and surprisingly, Asia, being the second highest COVID cases, has the second-lowest total death cases.

Correlation between all the metrics



- 1) The total number of COVID cases is not strongly correlated with the population. This essentially disproves my hypothesis that highly populated areas are more contagious to the COVID virus. It means that a country with a high population does not necessarily need to have a high COVID count if they adhere to COVID norms, such as wearing masks, sanitizing regularly, and avoiding social gatherings
- 2) However, total cases are strongly correlated with critical cases, active cases, total deaths, total tests, and total recovered.

Geospatial analysis



Country/Regi..	Population
India	1,381,344,997
USA	331,198,130
Indonesia	273,808,365
Pakistan	221,295,851
Brazil	212,710,692
Nigeria	206,606,300
Bangladesh	164,851,401

Country/Regi..	Total Cases
USA	5,032,179
Brazil	2,917,562
India	2,025,409
Russia	871,894
South Africa	538,184
Mexico	462,690
Peru	455,409

Country/Regi..	Total Deaths
USA	162,804
Brazil	98,644
Mexico	50,517
UK	46,413
India	41,638
Italy	35,187
France	30,312

Country/Regi..	Total Recover..
USA	2,576,668
Brazil	2,047,660
India	1,377,384
Russia	676,357
South Africa	387,316
Chile	340,168
Peru	310,337

Key Highlights:

- India seems to be doing well across all the countries, despite being the highest-populated country, followed by Indonesia.
- Europe's nations, such as the UK, Italy, France, and Spain, come under the top 10 countries by total deaths, despite their low populations and total COVID cases.
- The USA seems to be topping the charts in each category.

Conclusion

- 1) Globally, there are 19,169,166 total COVID cases, 713,007 total death cases and 11,250,837 total recovered cases.*
- 2) The Asian continent has done well in comparison to other continents, and within that, India and Indonesia have outperformed in each category.*
- 3) European countries such as the UK, Italy, France and Spain along with the USA, would need server restriction if they wanted to mitigate the COVID-19 pandemic impact. Also, these nations would need extra medical assistance as well.*
- 4) Data exploration has proved my initial hypothesis wrong: high populated countries have a strong correlation with total COVID cases.*
- 5) Countries such as Macao, New Caledonia, Dominica, Greenland, Falkland Islands, Vatican City, and Cayman Islands have a 100% recovery rate.*
- 6) It would be interesting to analysis the COVID-19 data over time and its impact on key growth metrics.*

SMRITI BHATTARAI

Contact & Links



Smriti.kunwar444@gmail.com



linkedin.com/in/smriti-bhattarai-a773a7146



[smritibhattarai \(github.com\)](https://github.com/smritibhattarai)
