

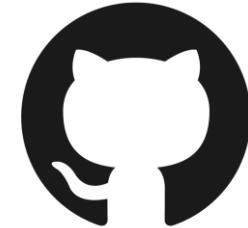
Hanadi Salim

Data Analytic Portfolio

Technical Skills “Tech Stack”



Contact me if you have
questions, ideas, or would
like to work with me



GitHub

Projects List

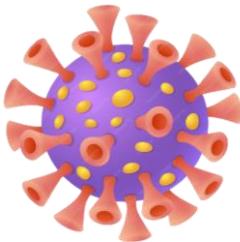
1



GameCo

Analyzing global video game sales

2



Influenza

Preparing for flu season in the U.S.

3



Rockbuster

Answering business questions for an online video rental company

4



Instacart

Marketing strategy for an online grocery store

5



Global Bank

Anti-money laundering projects at a global bank

6



Sustainable Development Goals

Analysis of SDG report

1. Game Co.

Analyzing global video game sales

1. GameCo

Project Overview



Objective

Perform a descriptive analysis of a video game data set to foster a better understanding of how GameCo's new games might fare in the market.



Data Used

[Video game sales data set](#)

Source: [VGChartz](#)



Tools Used

Microsoft Excel

Microsoft PowerPoint



Key Skills

Data cleaning

Data grouping and summarizing

Descriptive analysis

Developing insights

Visualization

Storytelling with data

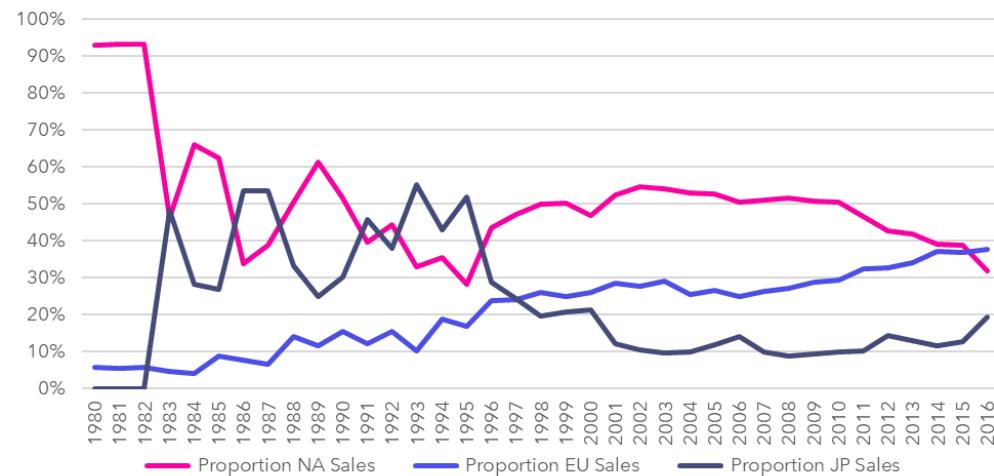
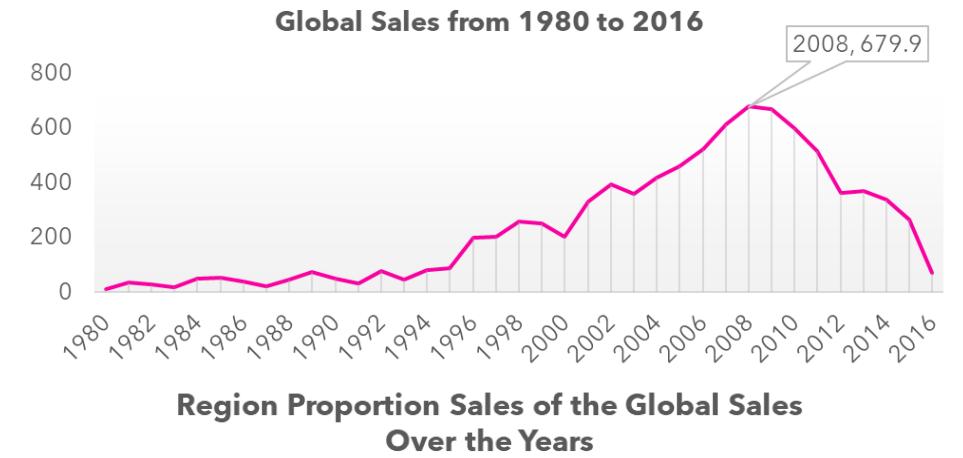
Project Deliverable:
[Presentation](#)



Did GameCo Sales stay the same?

GameCo assumed the sales for the various geographic regions have stayed the same over time.

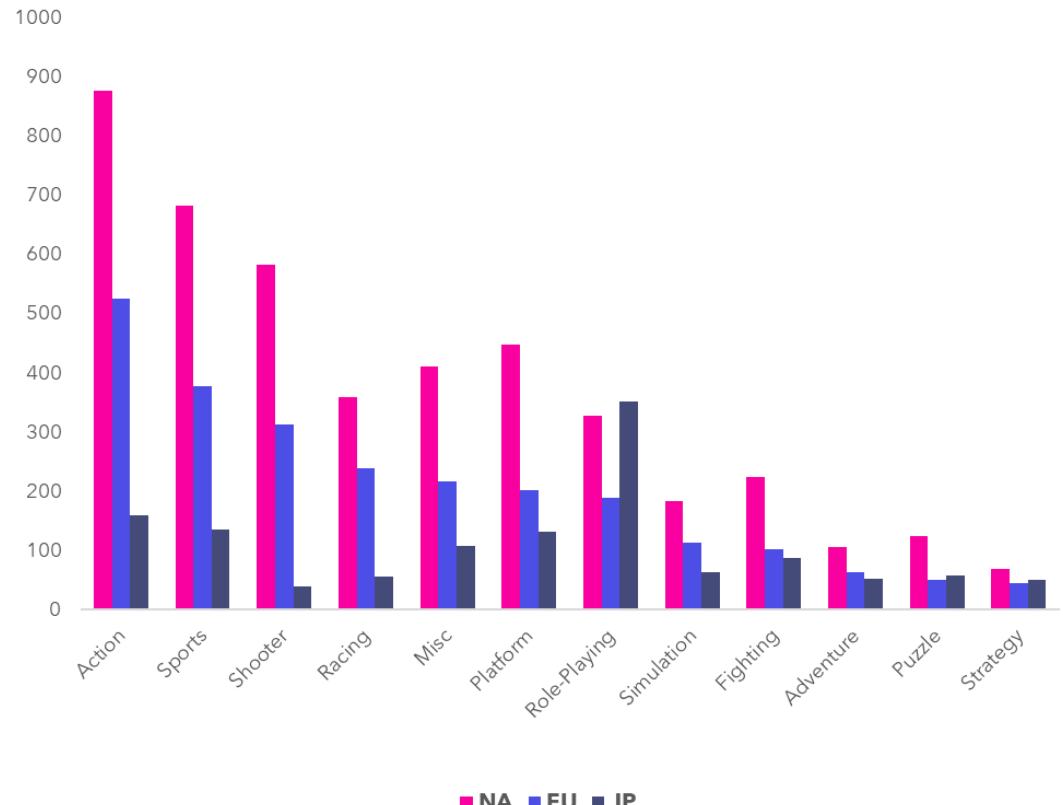
- The Sales have increased over the years. However, by the year 2008 which is the peak of the sales, we started to see a noticeable decline for the forthcoming years.
- North America was the leading region for many years. However, Europe took over North America in 2016 with its continuous rise of sales.



Are certain types of games more popular than others?

- Action, Sports, Shooter games' genre are the top leading games genres. This is due to the sales of these genres in North America and Europe which are the top leading regions.
- For Japan, Role-Playing games genres are the most popular followed by actions games.

Genre Proportion per Region Sales





Recommendations

- **Budgeting:** Increase the budget in Europe since it showed a promising increase in the sales, lower it for North America due to the continuous sales decline, and keep it the same for Japan as there are no measure changes in the sales.
- **Game Allocation:** In each region, allocate the type of game that is popular in that region and introduce the other popular game genre.
- **New Era:** It appears that the current project of the games is at the end of its life cycle which may need a change or a refresher by modifying the game's type to accommodate the current technology (e.g., VR Games) by collaborating with other leading companies or creating the games in-house if the resources are available.



2. Influenza

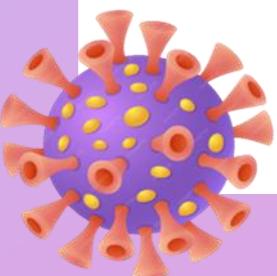
Preparing for flu season in the U.S.

2. Influenza Project Overview



Objective

Help the medical staffing agency determine when to send staff, and how many, to each state during influenza season in The United States since more people than usual suffer from the flu during that time, and Hospitals needs additional staff to accommodate the increase of patients



Data Used

1. [Influenza deaths by geography, time, age, and gender](#)

Source: [CDC](#)

2. [Population data by geography](#)

Source: US Census Bureau



Tools Used

Microsoft Excel
Tableau



Key Skills

Translating business requirements
Data cleaning
Data integration
Data transformation
Statistical hypothesis testing
Visual analysis
Forecasting
Storytelling with Tableau

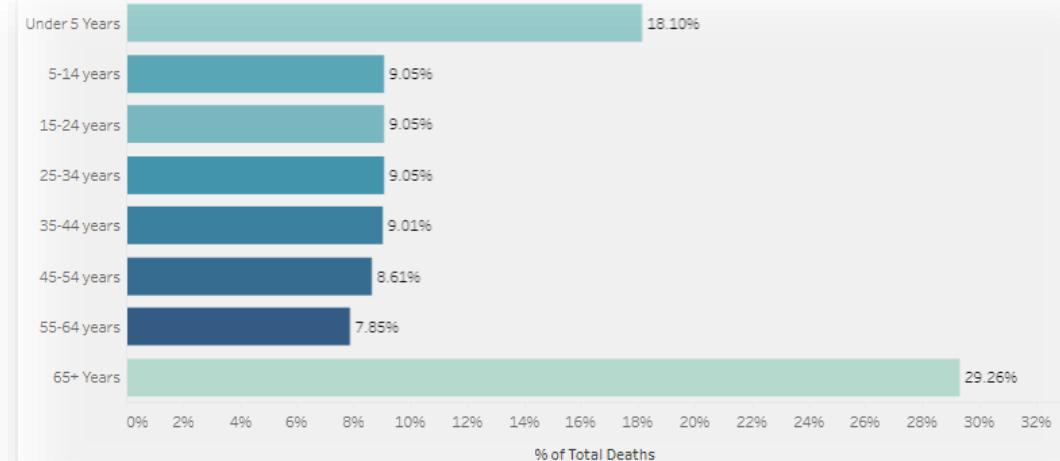
Project Deliverable:
[Presentation Video](#)
[Tableau Storyboard](#)

Who is the vulnerable Population and which state having most of them?

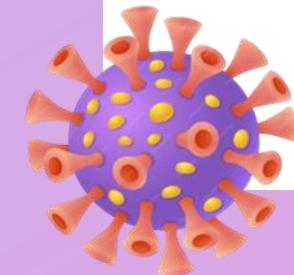
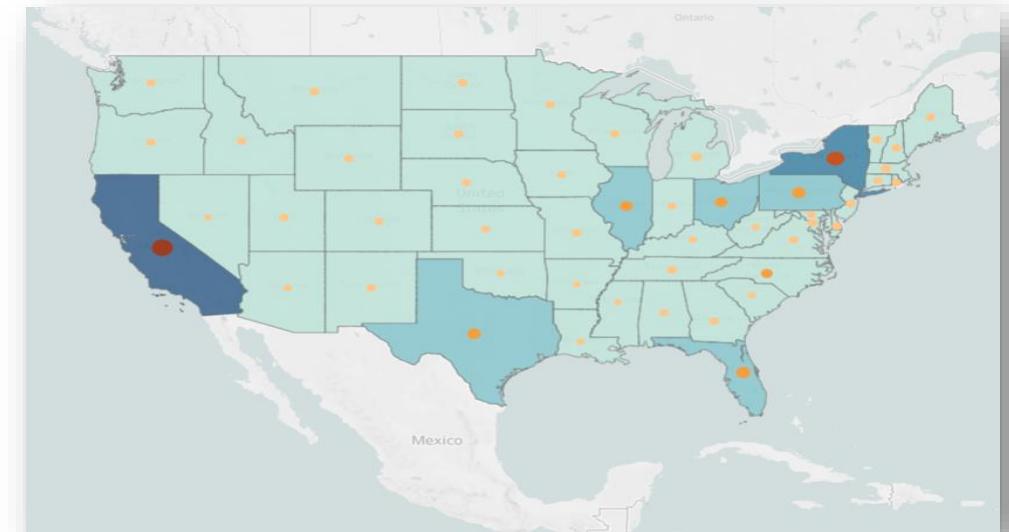
The Project Hypothesis: "If a state has a higher number of 65+-year-old population, then the rate of death due to the flu will be higher."

- **The 65+ Years** old population has a higher death rate than the other age groups, followed by those **Under the 5 Years** old age group.
- **California and New York** have the highest number of 65+ Years old population, which can be due to the large number of their population as a total.

Influenza Mortality Per Age Group



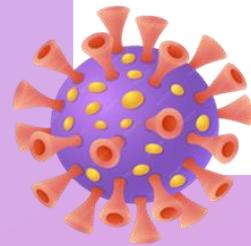
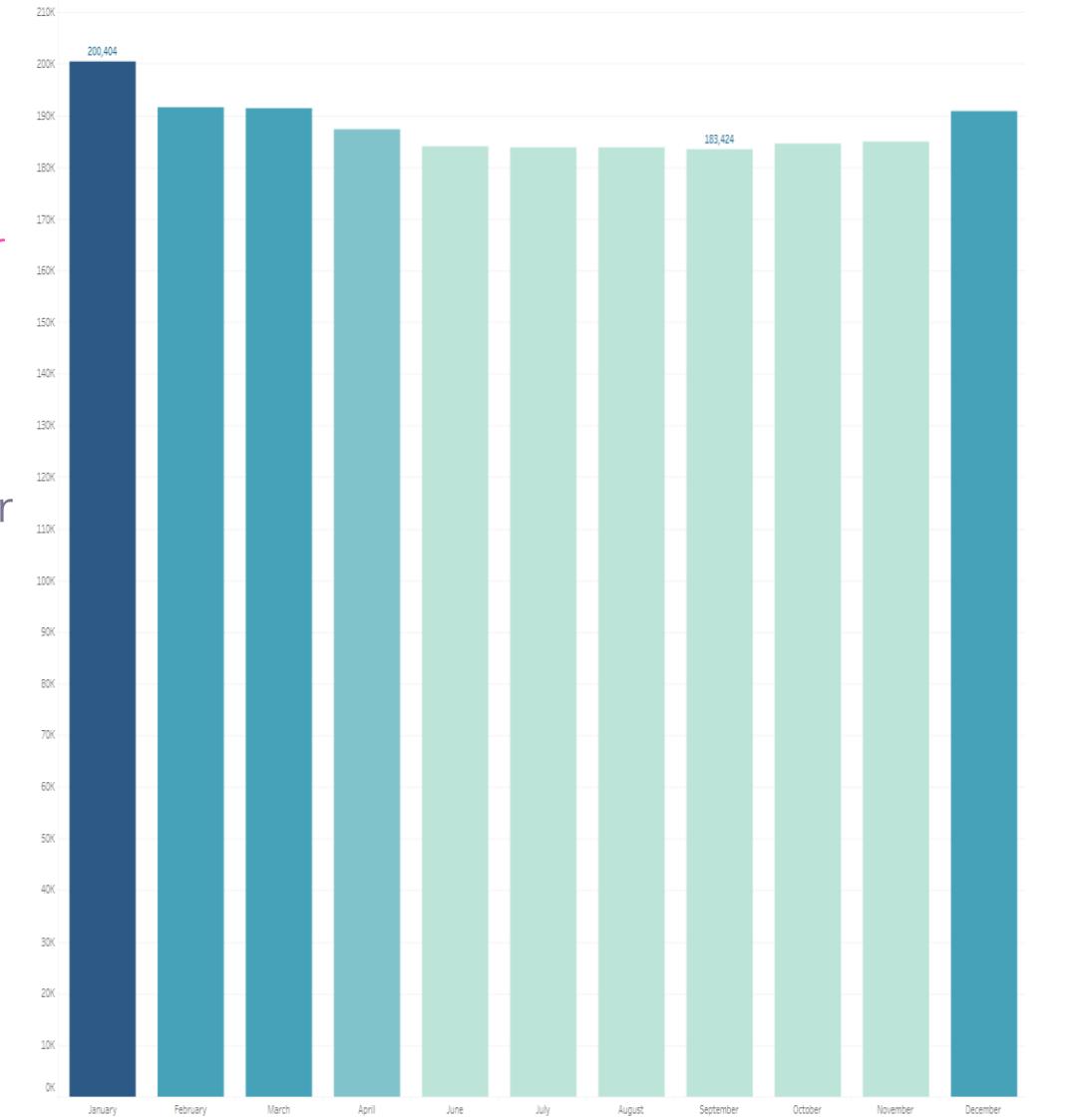
Influenza 65+ Years Mortality Per State



When is the influenza season the highest? And is every state the same?

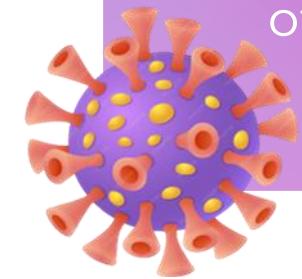
- January has the highest Influenza Mortality number compared to the rest of the year, followed by February, March, and December.

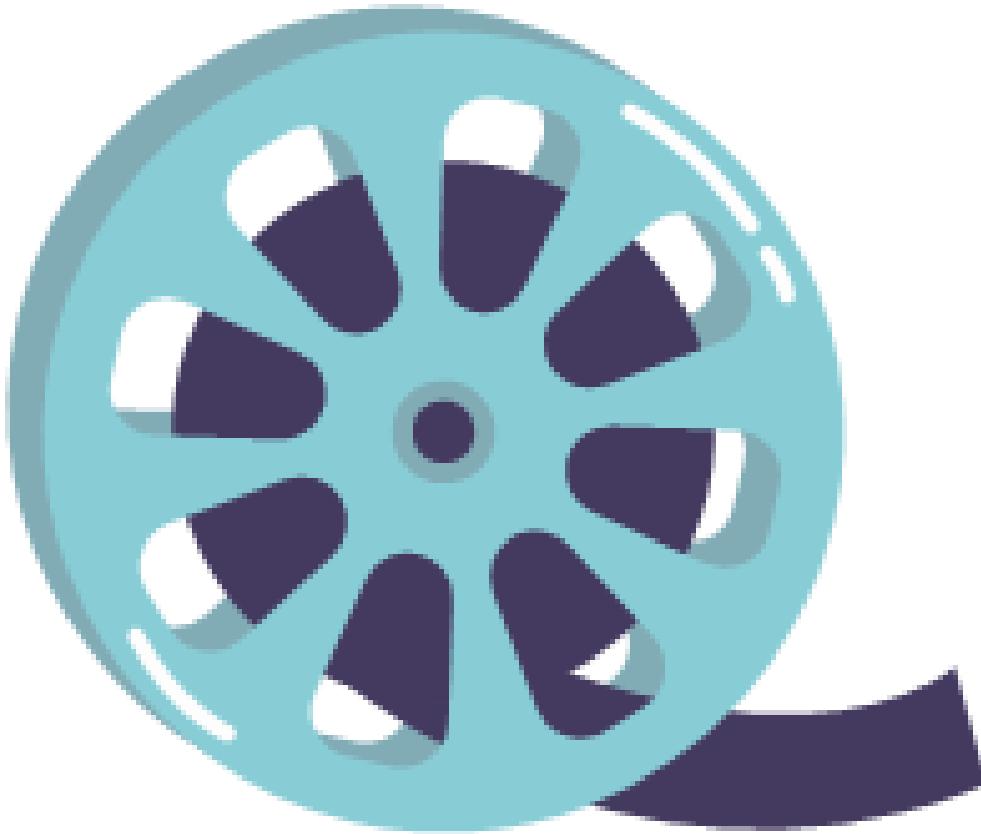
Influenza Mortality per Month



Recommendations

- **Time of staff allocation:** Having a full staff capacity during the rush months (January, February, March, and December.)
- **States to focus on:** California and New York since they have the highest number of 65+ year old population.





3. Rockbuster

Answering business questions for an
online video rental company

3. Rockbuster Project Overview



Objective

Answer business questions to help with the launch strategy for the new online video service by providing insights with data-related queries.



Data Used

1. [Rockbuster data set](#)



Tools Used

Microsoft Excel
PowerPoint
SQL
Tableau
Lucid chart



Key Skills

Database querying
Filtering
Cleaning and summarizing
Joining tables
Subqueries
Common table expressions
Create data dictionary

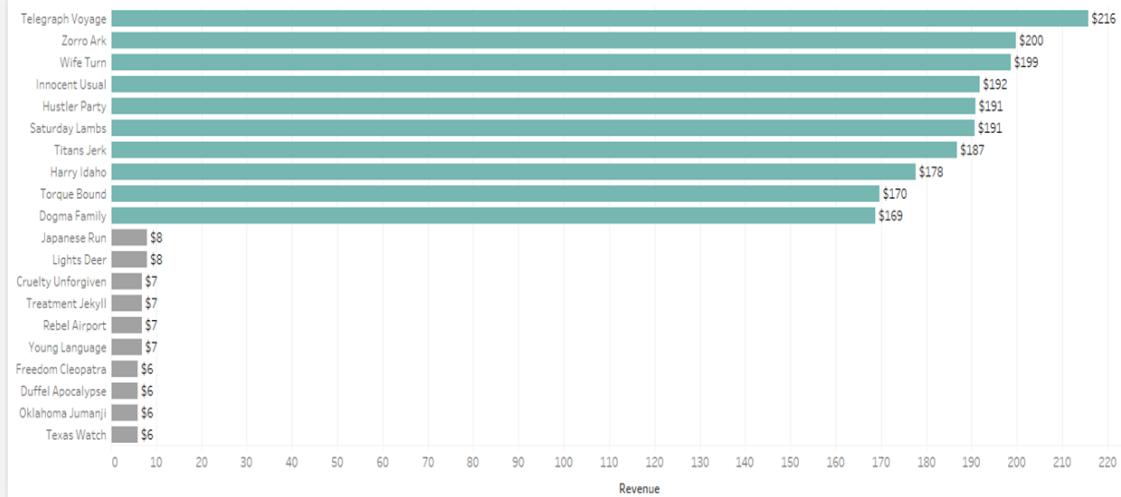
Project Deliverable:
[Presentation](#)
[Data Dictionary](#)
[GitHub](#)



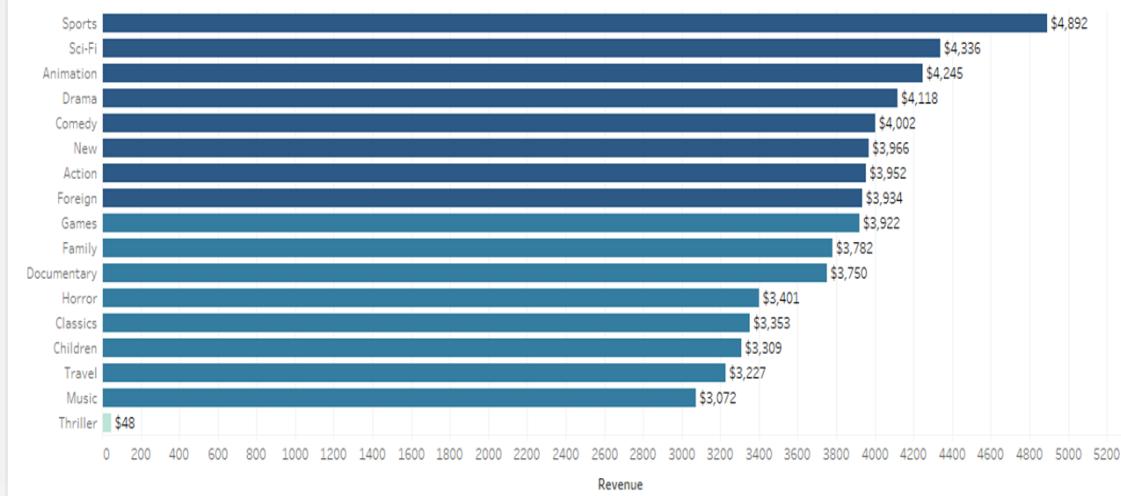
Which movies and genres are the most popular and contribute the most to revenue gain?

- The top three movies, are Telegraph Voyage, Zorro Ark, and Wife Turn.
- For the popular genre, **Sports** movies are the leading category, and the **Thriller** category has the most negligible revenue.

The top & bottom 10 movies title revenue

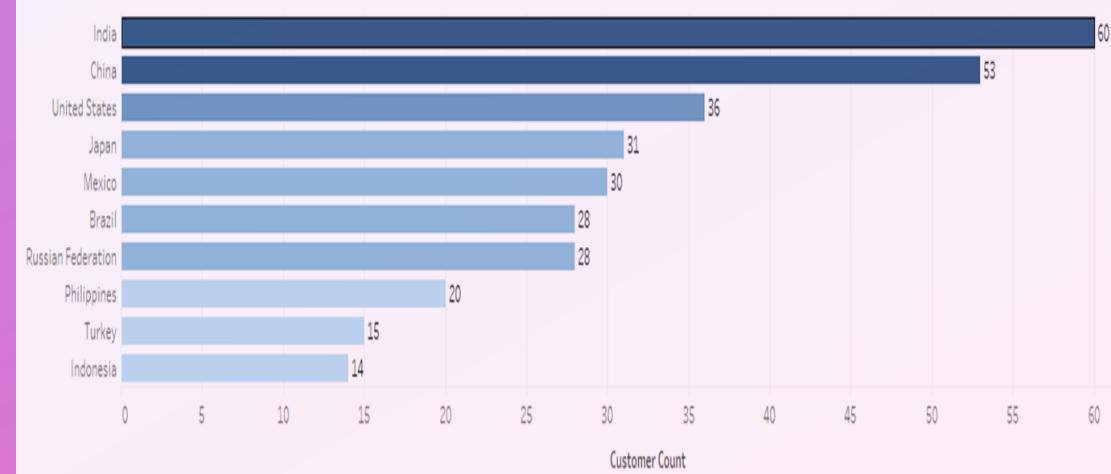


Movies revenue per category



The Map Shows us immediately where are our customers

India and China take the lead in terms of customers' number.



Recommendations

- **Promoting**

- The top 10 movies can be posted on the homepage to get first good impact on the customers.
- When selecting movies for the webpage, it is advised to have Sports movies take the more significant portion than other categories. On the other hand, it is better to avoid having Thriller movies for now.

- **Customer Focus**

- Focus the target of offers and promotions on the top ten countries with the highest number of customers. Additionally, utilize the “word of mouth” advertisement.

- **Next Step**

- After getting stabilized, it is good to explore new countries/customers to expand the business.



Instacart

Marketing strategy for an online grocery store



4. Instacart Project Overview



Objective

perform an initial data and exploratory analysis of some of Instacart data which is a store that operate through an app. This is to derive insights and suggest strategies for better segmentation based on the provided criteria.



Data Used

1. [Customers data set](#) (this data is fabricated by CareerFoundry for educational purposes.)
2. The Instacart Online Grocery Shopping Dataset 2017: [Kaggle](#)



Tools Used

Microsoft Excel
Python
Jupyter Notebook



Key Skills

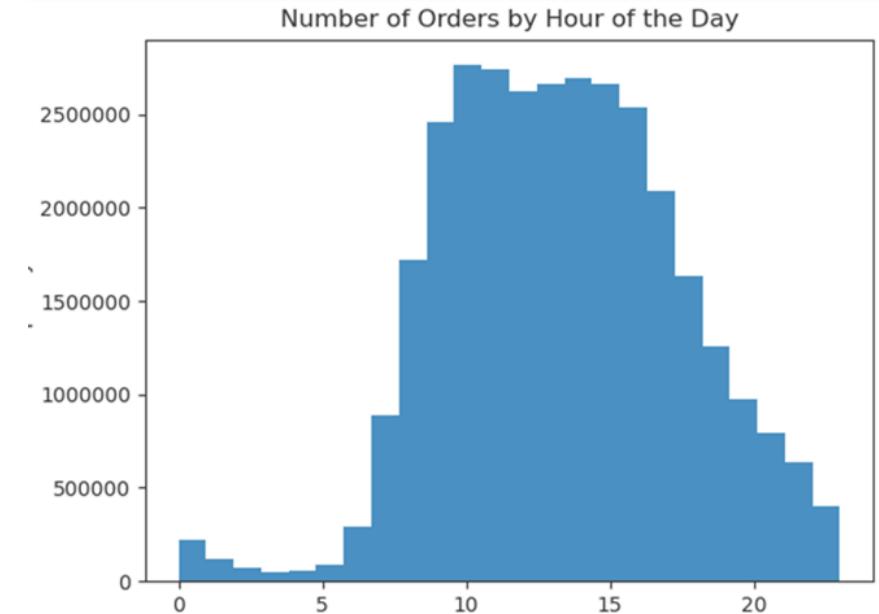
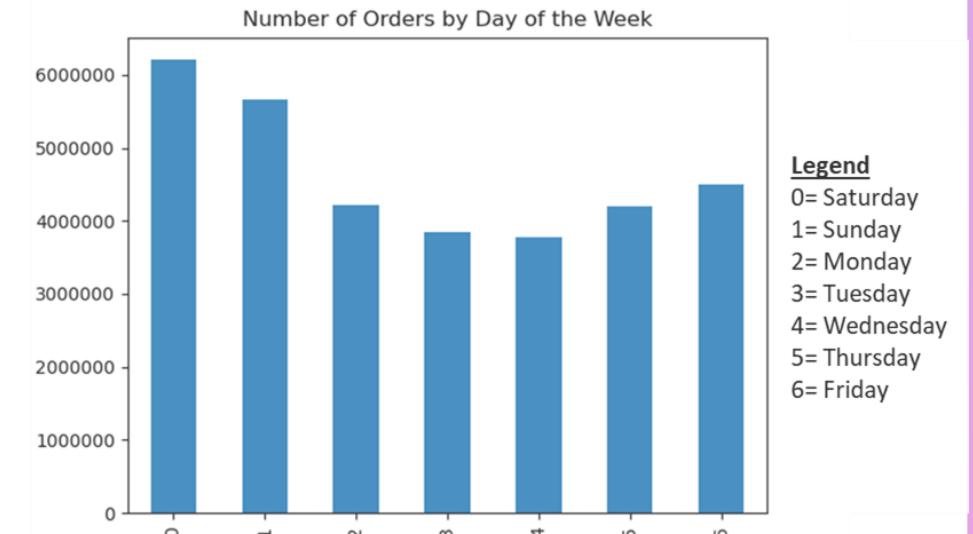
Data wrangling
Data merging
Deriving variables
Grouping data
Aggregating data
Visualization in Python
Reporting in Excel

Project Deliverable:

[Report](#)
[GitHub](#)

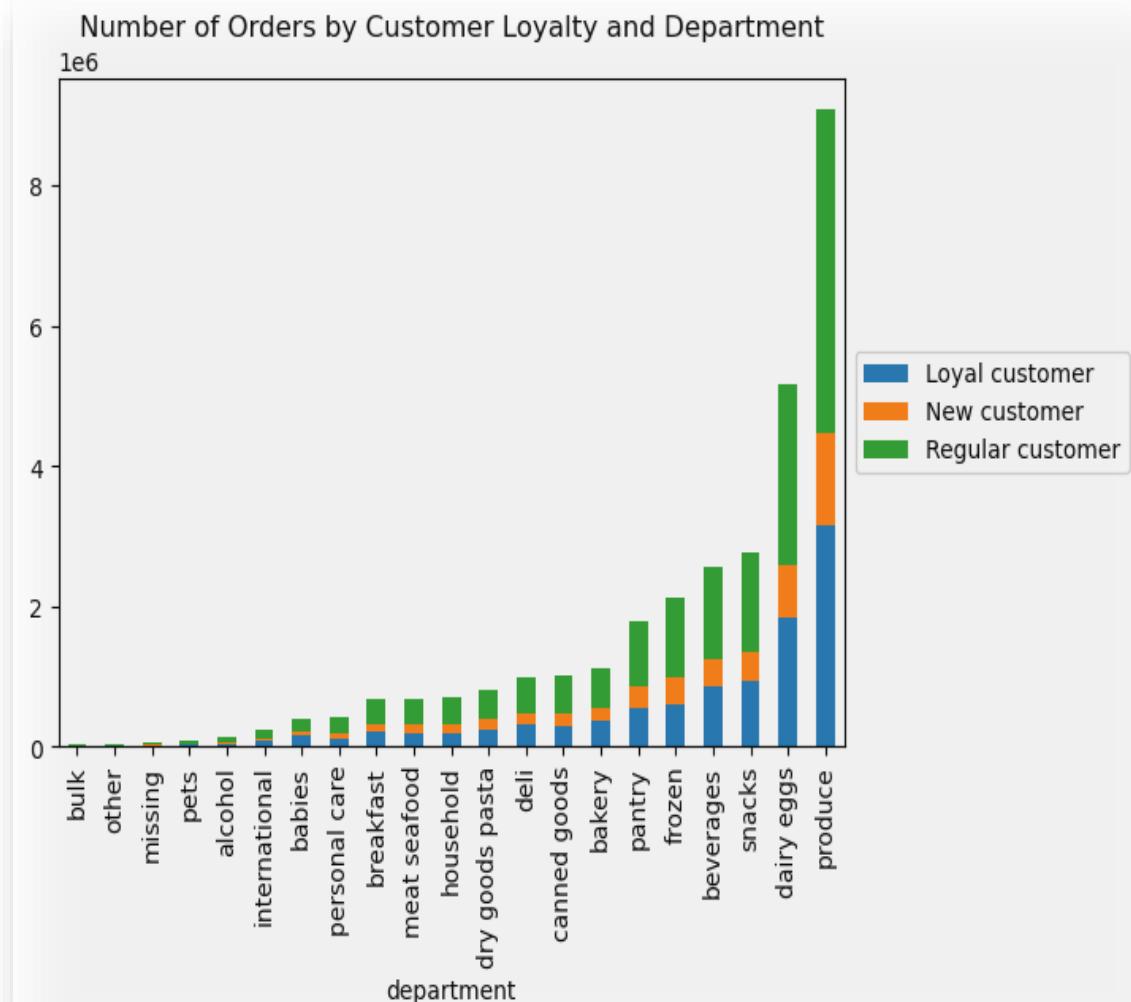
What the busiest days of the week and hours of the day are?

- the busiest days are Saturday and Sunday and as for the busiest hours it is the period from 10:00 to 16:00 hrs.



Which departments have the highest frequency of product orders? And are there differences in ordering habits based on a customer's loyalty status?

- Produce and dairy eggs departments has the highest order number followed by snacks and beverages.
- Regular customers make the highest orders followed by Loyal customers. on the other hand, New customers make the least orders in all departments.



Recommendations

- **Schedule Ads**
 - It is best to send ads during the weekdays and the beginning of the day since this is the best time to get customers think about what they would like to shop for lunch or dinner, or what to buy for the following weekend since most shopping is pushed to the end of the week.
- **Targeted Customers**
 - Increase advertising for the Instacart to attract more new customers.



Global Bank

Anti-money laundering projects
at a global bank



5. Global Bank Project Overview



Objective

Identify leading factors for customers leaving the bank.



Data Used

1. [Pig E. Bank's client data set](#)



Tools Used

Microsoft Excel
GitHub



Key Skills

Big data
Data ethics
Data mining
Predictive analysis
Time series analysis and forecasting



Project Deliverable:
[Report](#)

What are the risk factors that have contributed to customers leaving the bank?

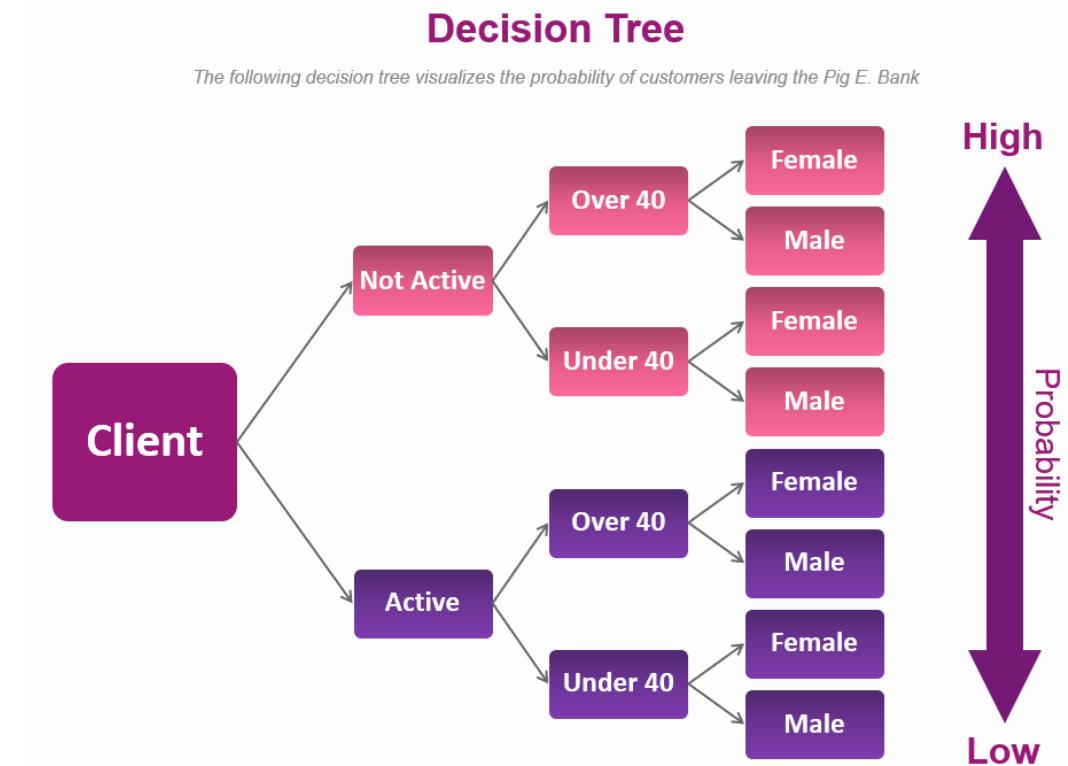
- Created two pivot tables where I filter the “Exited the bank?” column to check the count/average for the all columns to identify the risk factors.
- After checking the outcome, I could see that the age, gender, and member (active or not) have significant differences between the two groups. Therefore, these three factors will be selected as leading factors that contributed to the loss of the client.

	Left the Bank	Count of Gender	Average of Credit Score	Average of Age	Average of HasCrCard?	Average of IsActiveMember	Average of Tenure	Average of NumOfProducts	Average of Balance	Average of Estimated Salary
Row Labels	1									
Female	121	1	632	45	0.65	0.31	5	2	\$87,729	\$99,925
Male	83	1	643	46	0.78	0.29	5	1	\$93,899	\$93,118
	Stay at the Bank	Count of Gender	Average of Credit Score	Average of Age	Average of HasCrCard?	Average of IsActiveMember	Average of Tenure	Average of NumOfProducts	Average of Balance	Average of Estimated Salary
Row Labels	0									
Female	341	1	652	38	0.71	0.55	5	2	\$78,166	\$98,162
Male	445	1	652	37	0.70	0.57	5	2	\$72,443	\$99,591



What are the probability of customers leaving the bank?

- **Clients with following criteria are more likely to leave the bank:**
 - Not Active client
 - Client who are over 40 years old
 - Female client





Analysis of SDG report from 2000 to 2023

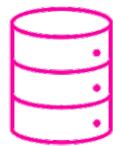
Sustainable Development Goals

6. Sustainable Development Goals Project Overview



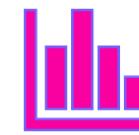
Objective

1. Communicate important information regarding the connections between SDG Scores.
2. Demonstrate change over time.
3. Identify Top and Bottom SDG Score Countries.



Data Used

1. [2000-2023 Sustainable Development Report data sets](#)
2. [world-countries.json](#)



Tools Used

Microsoft Excel
PowerPoint
Python
Tableau
GitHub



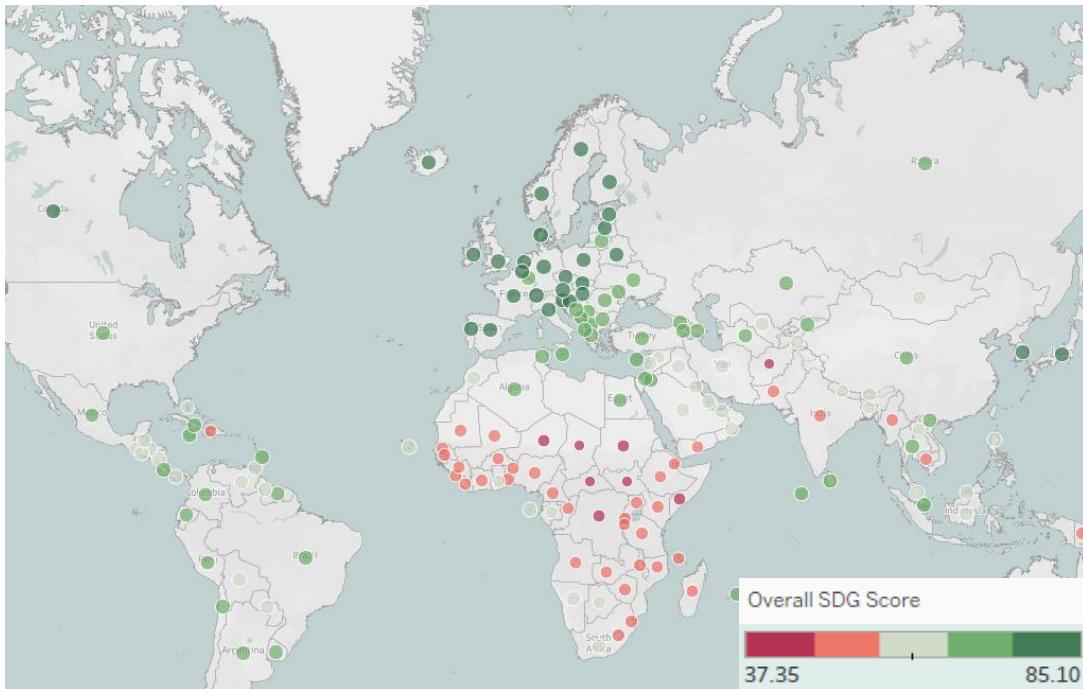
Key Skills

Sourcing Open data
Explore Relationships
Geographical Visualization
Regression analysis
Clustering Analysis
Time series Analysis
Storytelling with Tableau

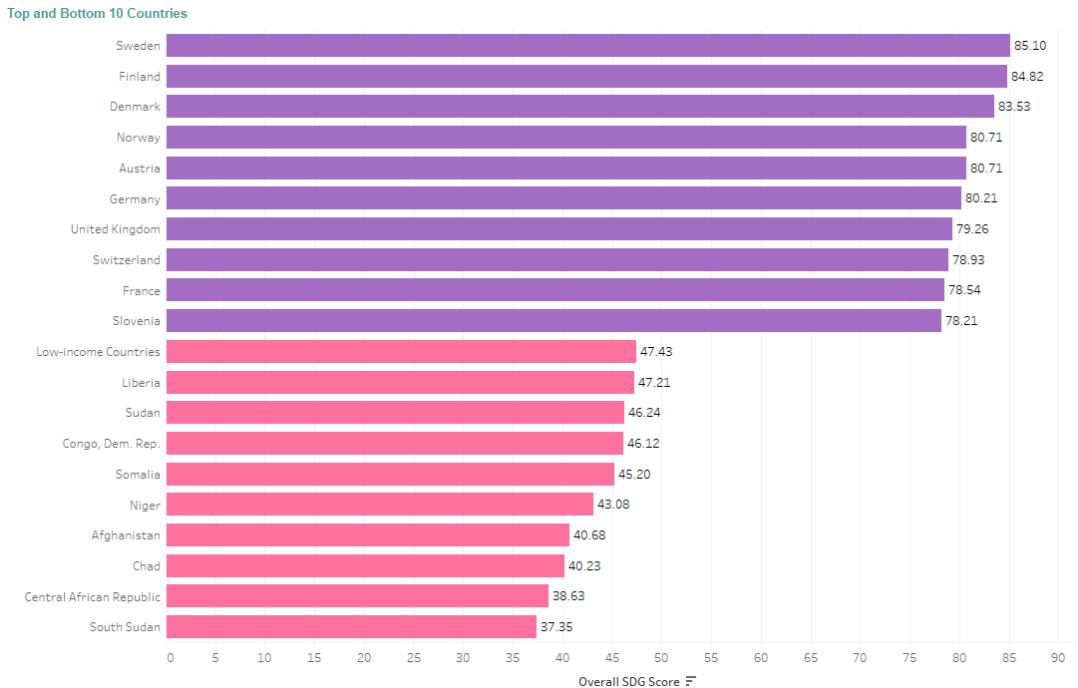
Project Deliverable:
[Tableau Storyboard](#)

Overview

We can see that majority of the countries are on the green side which mean the SDG plans are working although not perfect since the full score is 100.



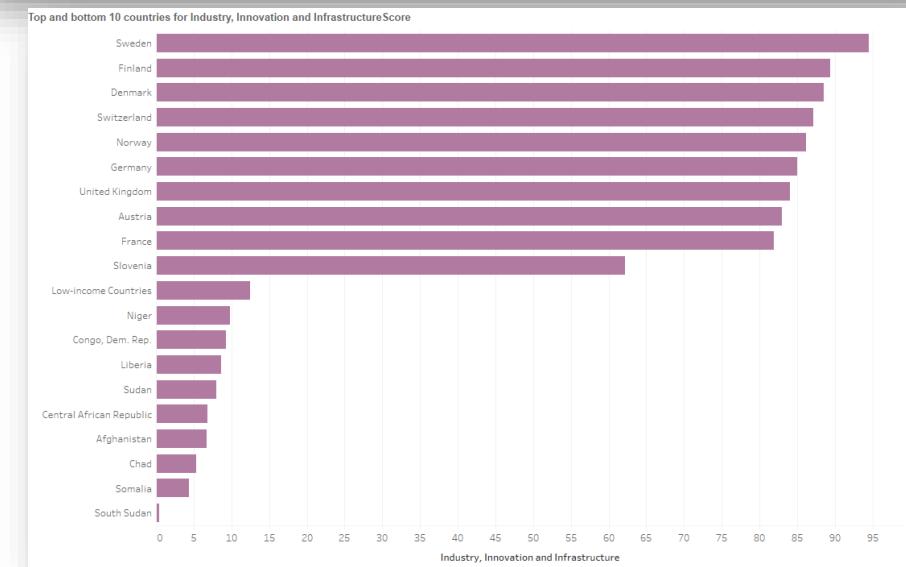
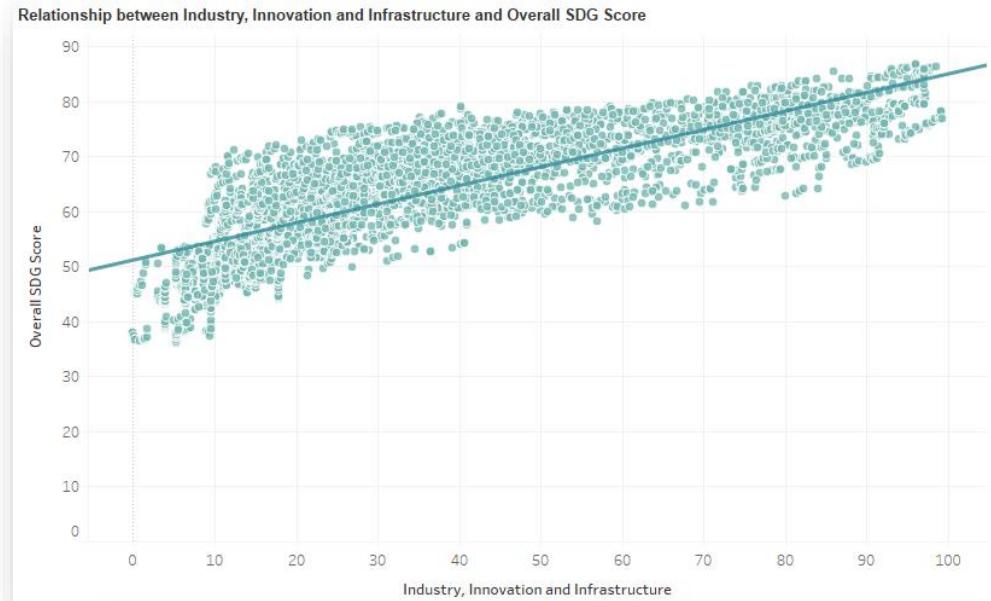
The top countries are known for their economic and general stability. Where the bottom countries do not have strong Infrastructure.



The Project Hypothesis:

“if a Country has high Industry, Innovation and Infrastructure goal, there will be higher chance in developing stabilizers and emergency plans during crises.”

- So, from the trend line we can clearly see a strong positive relationship between high Industry, Innovation and Infrastructure score and the overall Sustainability Development score. In which, as the first increase the later increase with it.
- Additionally, Countries who had high overall score were also high in the Industry, Innovation and Infrastructure score. Which could mean if countries with low overall score works on the Industry, Innovation and Infrastructure aspect, they could increase their SDG score.



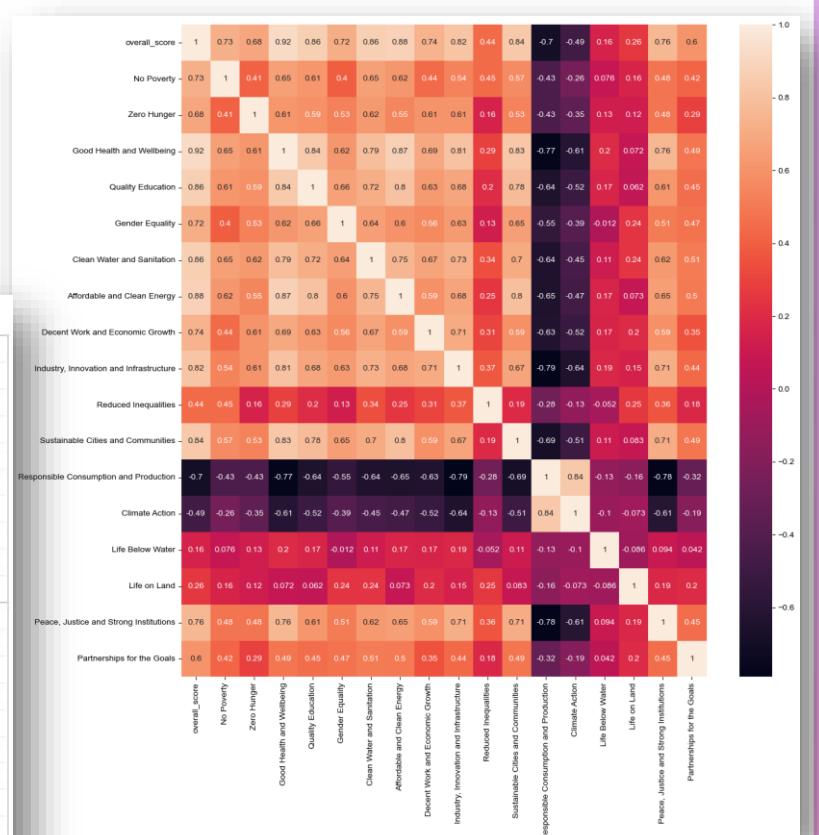
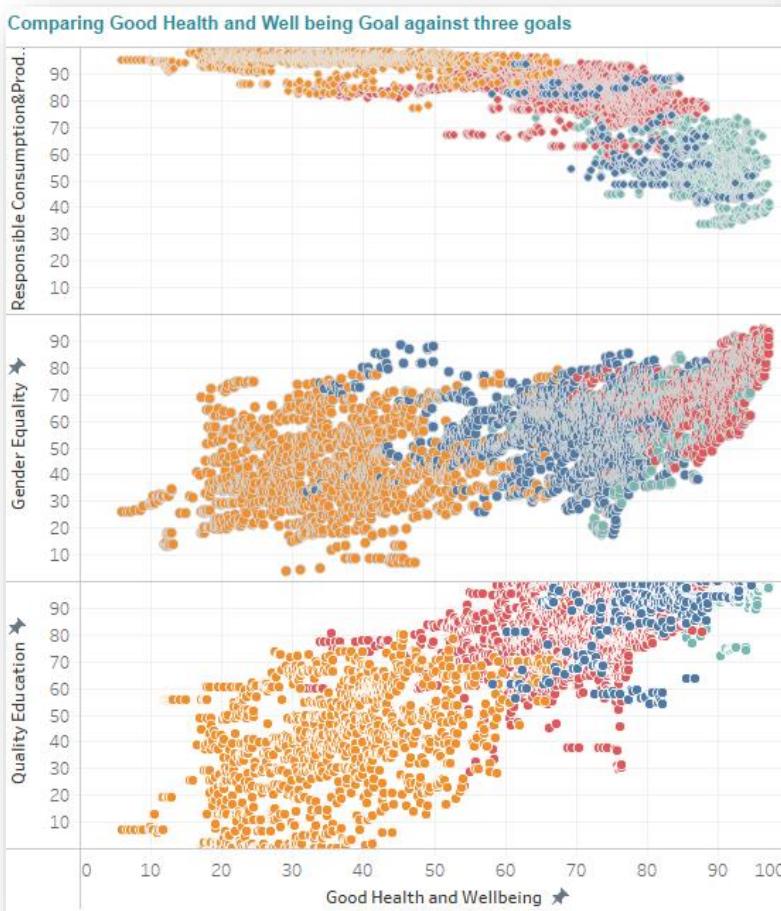
What goals scores have the most impact on the overall score?

- The goals scores with strong positive impact that range from 0.6 to 0.92 coefficient which mean the higher the score of these goals the better the overall score and vice versa. are as follow: No poverty, Zero Hunger, Good Health and Wellbeing, Quality Education, Gender Equality, Clean Water and Sanitation, Affordable and Clean Energy, Decent Work and Economic Growth, Industry, Innovation and Infrastructure, Sustainable Cities and Communities, Peace, Justice and Strong Institutions, Partnerships for the Goals.
- Reduced Inequalities Goal score has a 0.44 coefficient which indicate a moderate positive relationship.
- The goals scores with weak positive impact are Life Below Water and Life on Land.
- Responsible Consumption and Production goal has a - 0.7 coefficient indicate a strong negative correlation which mean the lower the goal's score the higher the overall score is and vice versa.
- Climate Action goal has a - 0.49 coefficient indicate a moderate negative correlation.

Good Health and Wellbeing goal score has the highest impact on the overall score with 0.92 coefficient.

When creating this cluster analysis, it showed us an interesting outcome as follow:

- The higher the "Good Health and Wellbeing" score is, the lower the "Responsible Consumption and Production" score become. I expected the opposite. This also gives us an idea on how some countries have low score on the "Good Health and Wellbeing" goal, but they have high score for "Responsible Consumption and Production" goal.
- When comparing "Good Health and Wellbeing" score with "Gender Equality" and "Quality Education" we can see that there is a positive relationship. On the other hand, the cluster shows us that majority of the countries have low score in regard to "Good Health and Wellbeing" goals. which means if these countries increase this goal's score , then there is a chance they will eventually have high score for quality education and gender equality accordingly.



Final Result

- Limitations
 - When looking at the data points on the map over the years, I noticed that some countries doesn't have data.
 - There weren't enough variables other than the countries name, year, and goals scores. Which limited us from checking if the economical status and population volume have an effect on the overall sustainability Development progress.
- Recommendations:
 - If a country that is not presented in the data wants to evaluate their status, they can check another country score with similar features.
- Next Steps:
 - Incorporate an additional data set with population number for each country to check if it is related to having high score.
 - Investigate if economical status has affect on the countries' sustainability development progress.

Thank you

Hanadi Salim

