DATA ANALYTICS PORTFOLIO

Brought to you by:

Omar Mohandes

List of Projects

- Game Co.
- Medical Staffing Agency
- Rockbuster Stealth LLC
- Instacart Basket
- Mental Health Disorder

Data analysis tools and techniques

This explicitly mentions both the tools and the analytical techniques used.

python

SQL

pandas

jupyter

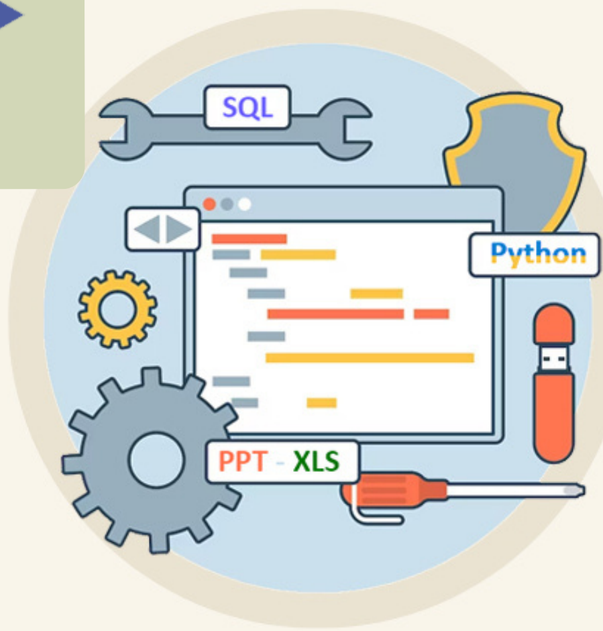
excel

powerpoint

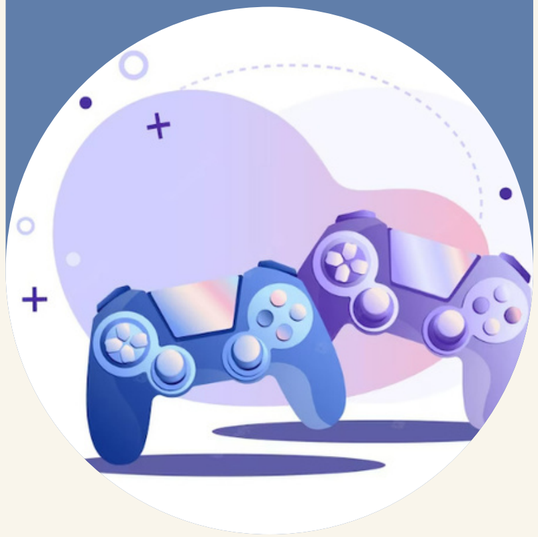
pgAdmin
Management Tools for PostgreSQL

DbVisualizer

photoshop



Game Co.



Project Overview

- **Study introduction:** GameCo is a video game company that aims to use data to inform the development of new games. Therefore, they require a descriptive analysis of a video game data set to gain a better understanding of how their new games could perform in the market.

- **Objectives:** Game Co. is interested in understanding the popularity of different types of games, identifying the main competitors in specific markets, tracking changes in popularity of games over time, and analyzing variations in sales figures across different geographic regions over time

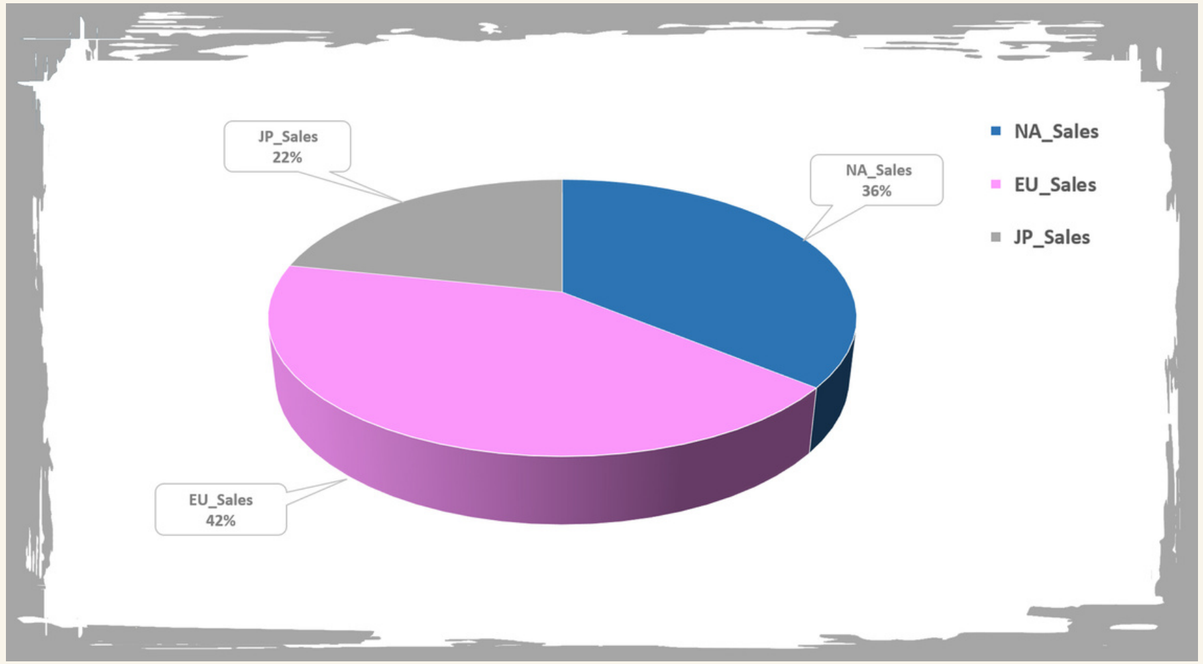
- **Skills:** experience in various skills such as cleaning data, grouping and summarizing data, conducting descriptive analysis, presenting findings, and creating visualizations.

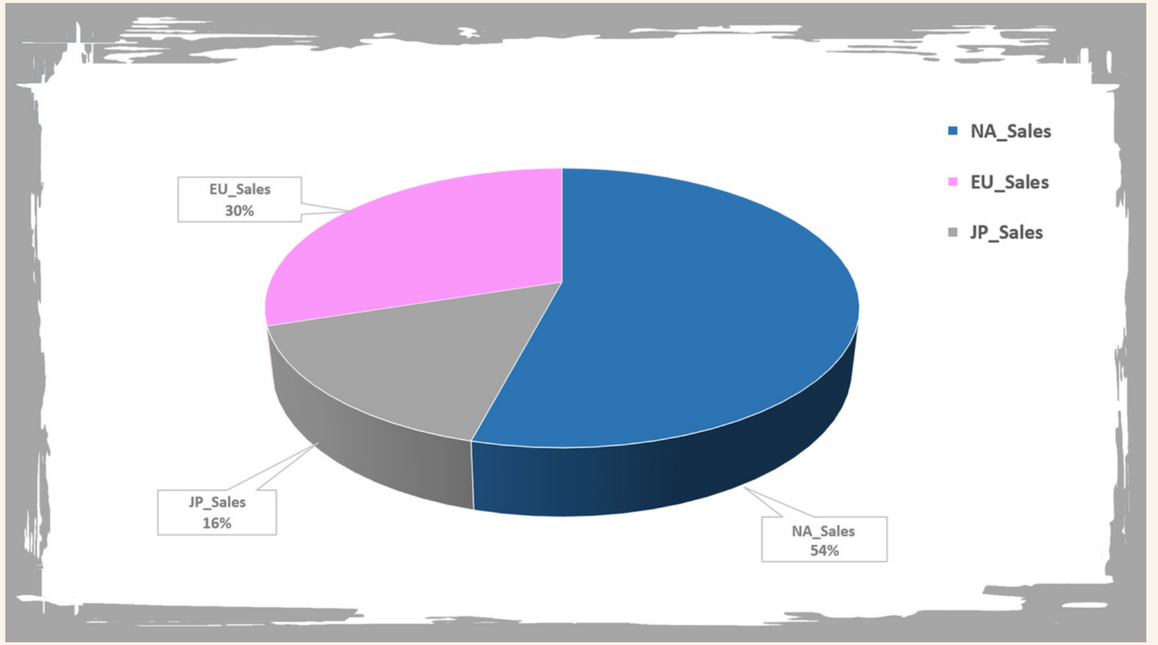
- **Tools** utilized for Study purposes:

Testing the sales percentages in the three major markets and comparing them across different years, including the overall trend and specifically in the year 2016.

Market share percentage in 2016

In 2016, sales in NA and EU were neck and neck, but EU sales saw a slight increase in their favor.





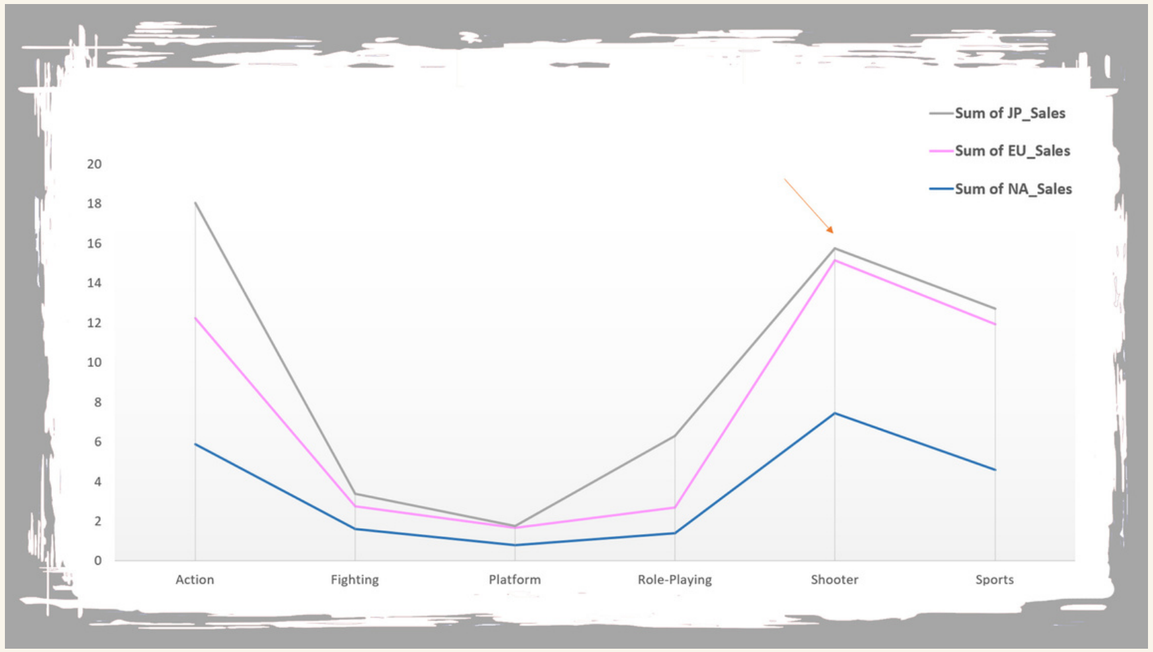
Percentage of market share over the years

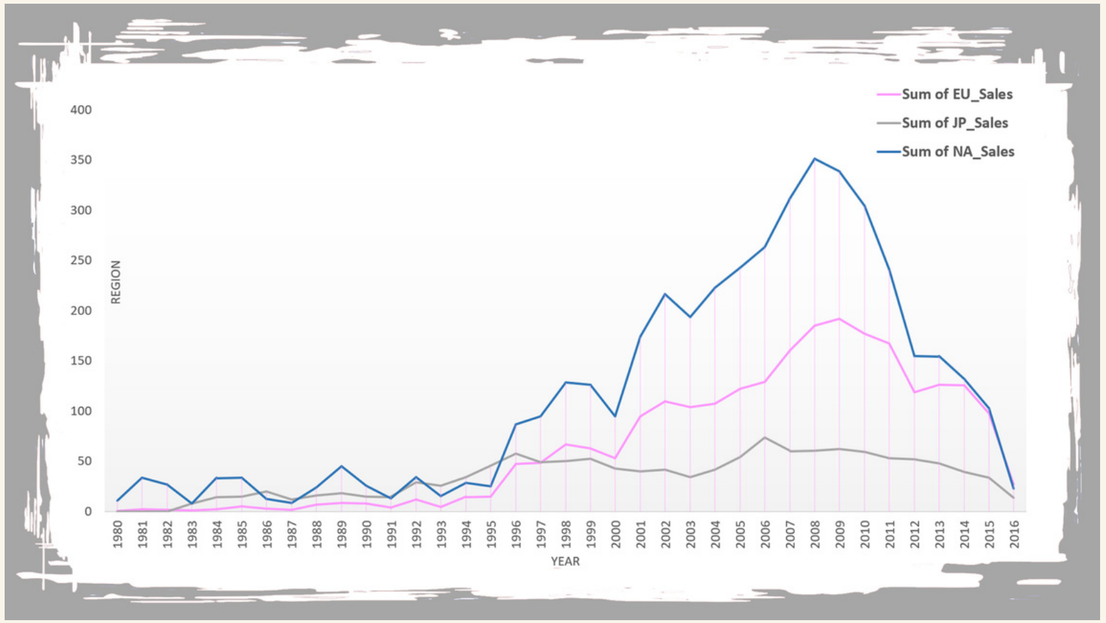
According to the graph, the NA region had the highest sales distribution compared to other regions.

The sales trend over the years and how different game genres performed in 2016.

In 2016, Top six gaming genres and their corresponding behaviors.

The industry exhibits consistent growth, although with a few exceptions. Notably, the shooter genre experienced a significant increase in demand in Japan and the EU, driven by the popularity of three games: Tom Clancy's The Division, Uncharted 4: A Thief's End, and Overwatch.

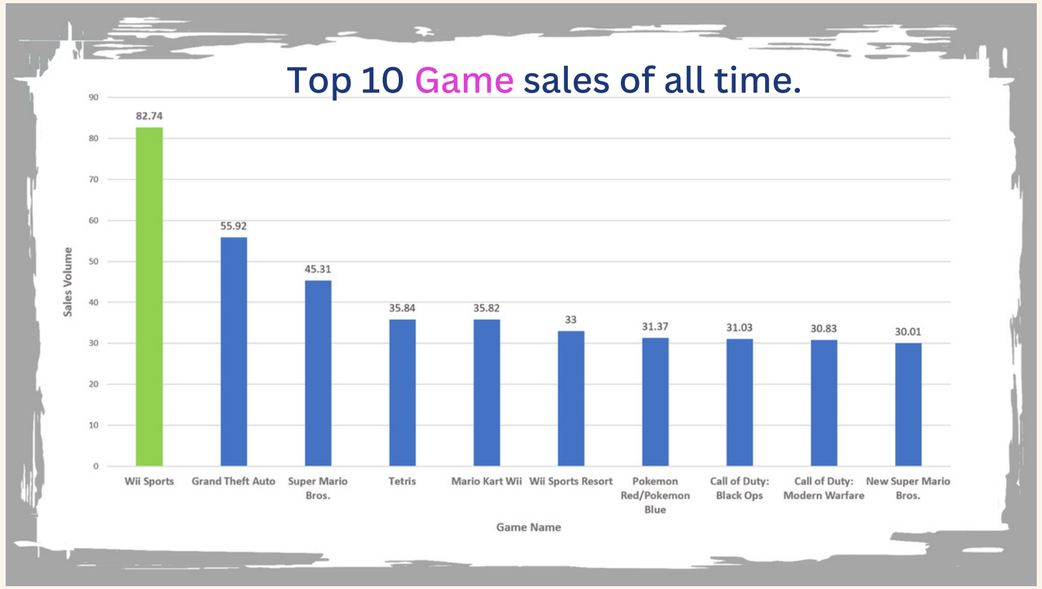


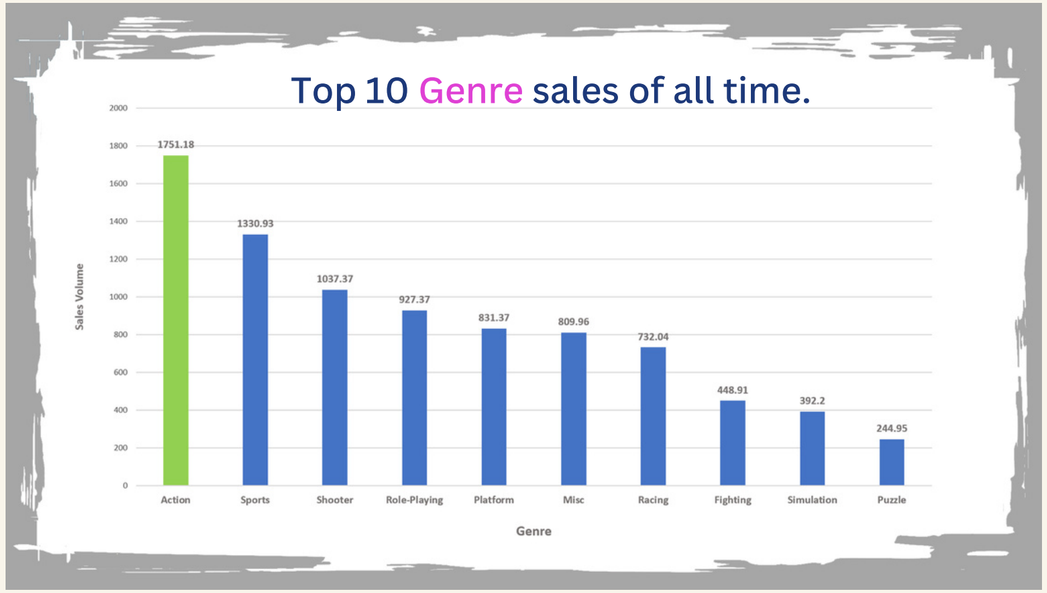


Over the years, there has been a trend in sales.

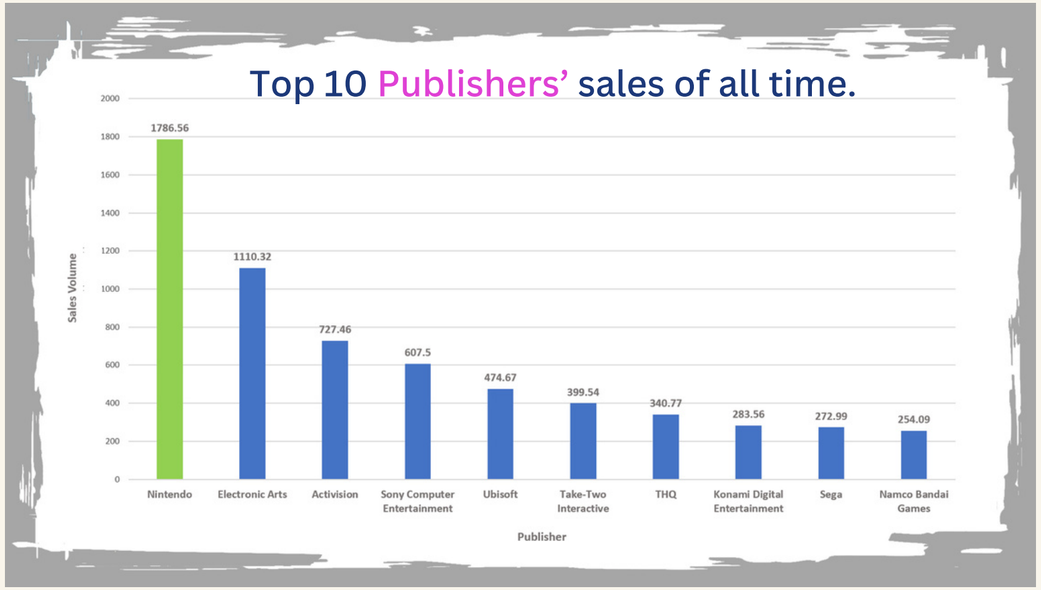
The sales data was organized by region and year. North America had the highest demand, while the other markets did not show any significant changes in demand.

Statistics & Insights

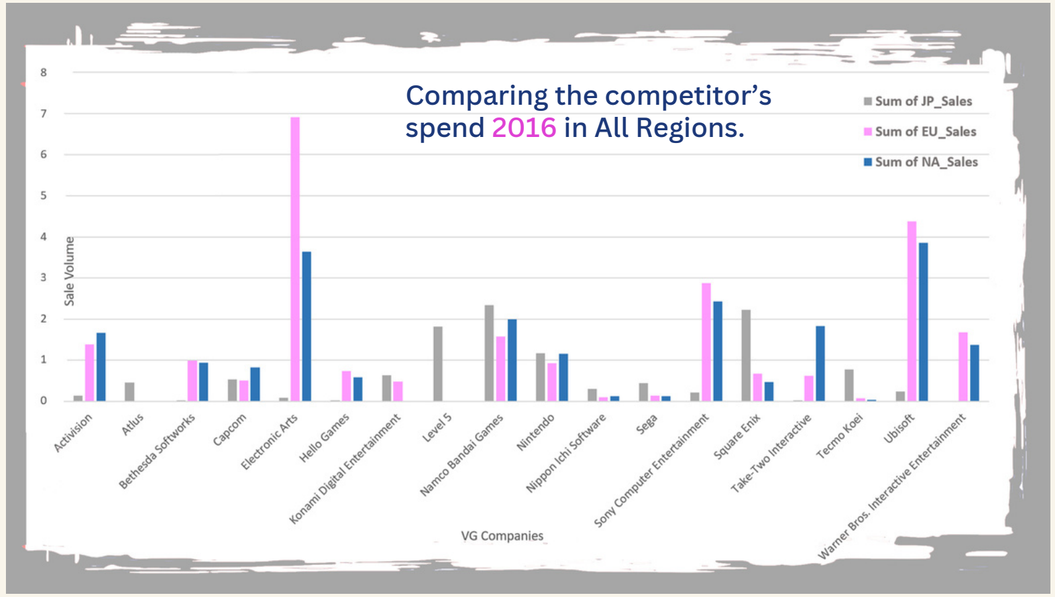


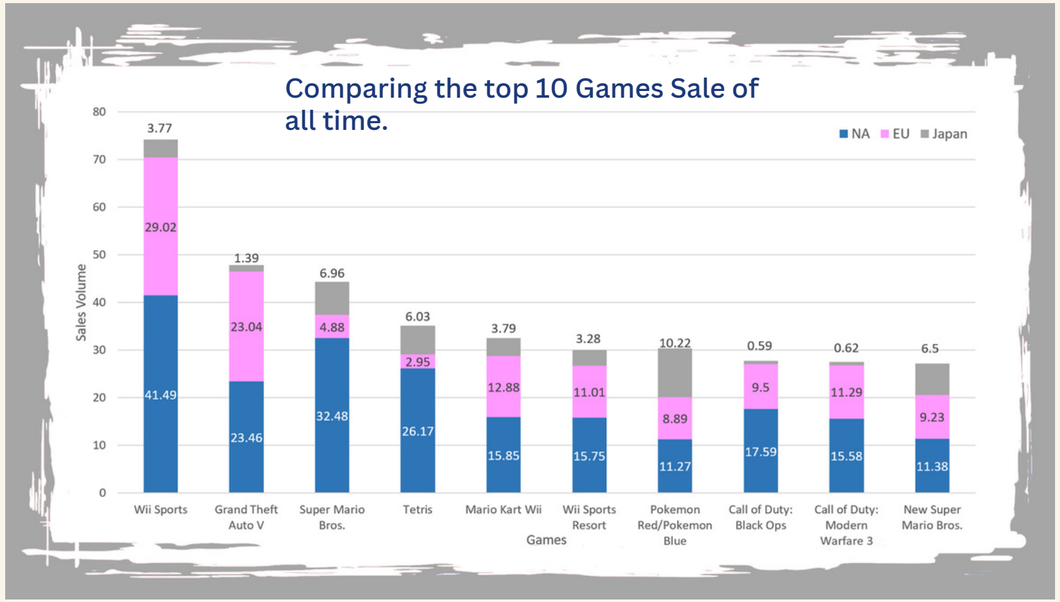


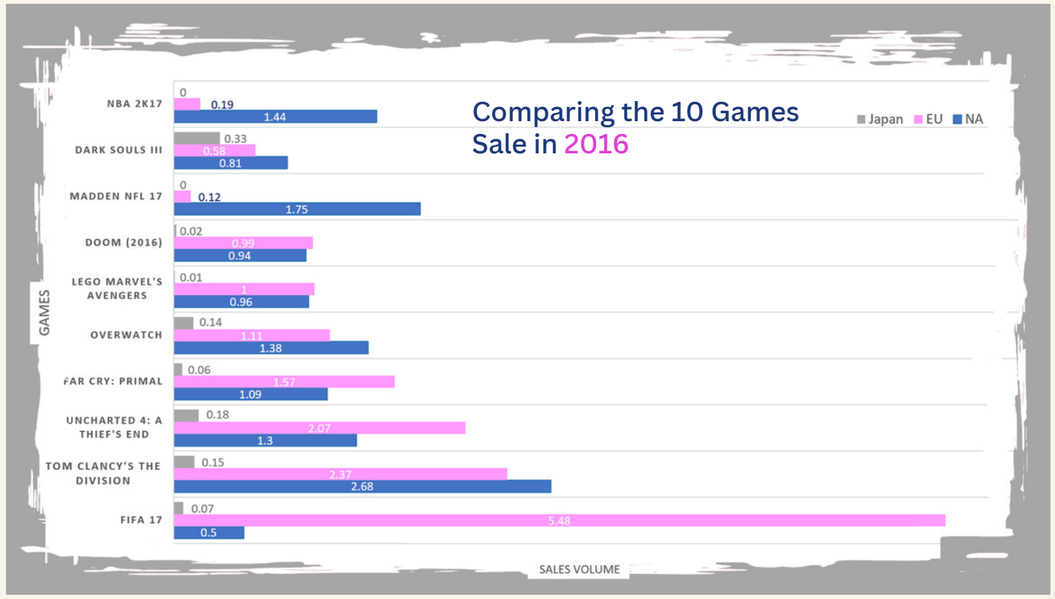
- According to the Top 10 Publishers list, Nintendo is lagging behind its closest competitors.
- The Top 10 Games sales report reveals that the Wii Sports game was the undisputed leader in sales from 1980 to 2016.
- In terms of Game type, Action has dominated all years followed by Sport, as per the Top 10 Game Type report.



Statistics & Insights







- According to the findings, Wii Games had strong sales in both North America and Europe, making it the most favored video game.
- In the 2016 graph, the sales force of video games was distributed geographically, with FIFA 2017 ranking as the top game.
- In the North American and European markets, Electronic Arts and Ubisoft were the main competitors, while in Japan, Namco Bandai Games and Square Enix were the top performers.

Recommendations

- To maintain and expand market share, it would be advisable to allocate resources and focus marketing efforts on prioritizing the NA and EU markets. These are currently the top-selling regions for video games.
- To achieve success in the gaming industry, it is important to prioritize the development and promotion of high-quality action and sports games. These genres have proven to be popular across all markets and have a strong demand among consumers.
- It would be wise to focus on the Japanese market and capitalize on the lack of attention from competitors. Developing games that cater to Japanese preferences, like Yokai Watch 3 and Dragon Quest, will give us a competitive edge in this market.
- To gain valuable insights and stay competitive, it is recommended to monitor the successful practices of industry leaders such as Electronic Arts and Ubisoft in the EU and NA markets. Analyzing their strategies, marketing tactics, and game offerings can provide valuable knowledge for adapting and implementing successful practices. Additionally, observing market leaders in Japan, such as Namco Bandai Games, Square Enix, Nintendo, and Level 5, can offer valuable insights and strategies to learn from and potentially apply.

Medical Staffing Agency

Project Overview

Study introduction: During the influenza season, the United States experiences a higher number of individuals suffering from the flu. Sadly, certain groups, such as vulnerable populations, may develop severe complications and require hospitalization. To ensure that these extra patients receive adequate care, hospitals and clinics require additional staff. A medical staffing agency can provide temporary staff to meet this need.

Objectives: To ensure timely medical assistance, it is essential to assess the need for medical staff and determine the number of professionals required in each state. Factors such as urgency and criticality of the situation must be taken into consideration. It is important to note that the staffing agency has a finite number of nurses, physician assistants, and doctors available, and hiring more is not currently feasible due to budget constraints.

Skills: Variety of skills has been used, including cleaning and organizing data, performing hypothesis testing and predictive analysis, creating visualizations, and presenting findings using Tableau.

Tools utilized for Study purposes:

Influenza Season Interim report

Interim report analysis a specific dataset.
The focus was on the number of deaths caused by influenza among individuals who are either 65 years old or older, or 5 years old or younger. Additionally, I conducted the same test on the population data set.
During the analysis, various aspects such as variance, standard deviation, mean, and outliers have been checked.
Furthermore, I evaluated the correlation between the population and the occurrence of influenza deaths.

Interim report findings

Descriptive Analysis

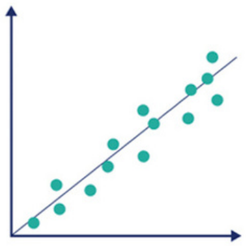
Assuming that the states have a higher proportion of elderly persons (those over 65) and younger people (those under 5 years old). This Age Group will have a higher risk of contracting the flu virus, as well as a higher risk of passing away.

Thus, two variables related to vulnerable individuals will be the focus of the descriptive research. (population Over 65 years and Under 5 years)

	Deaths > 65 Years	Population > 65 Years	Deaths < 5 Years	Population < 5
Data Set Name	Influenza Deaths	Census Population	Influenza Deaths	Census Population
Sample or Population	Sample	Sample	Sample	Sample
Normal Distribution	Normal Distribution	Normal Distribution	Normal Distribution	Normal Distribution
Variance	944,307.02	786,799,499,984.10	Unable to draw any insightful conclusions from this material.	208,782,049,770.34
Standart Deviation	971.75	887,017.19		456,926.74
Mean	896.80	806,988.94		386,282.64
High Outliers	2,840.31	2,581,023.32		1,300,136.13
Low Outliers	-1,046.71	-967,045.45		-527,570.85
Outliers %	3.92%	6.32%		3.92%

Interim report findings

➢ Results & Insights

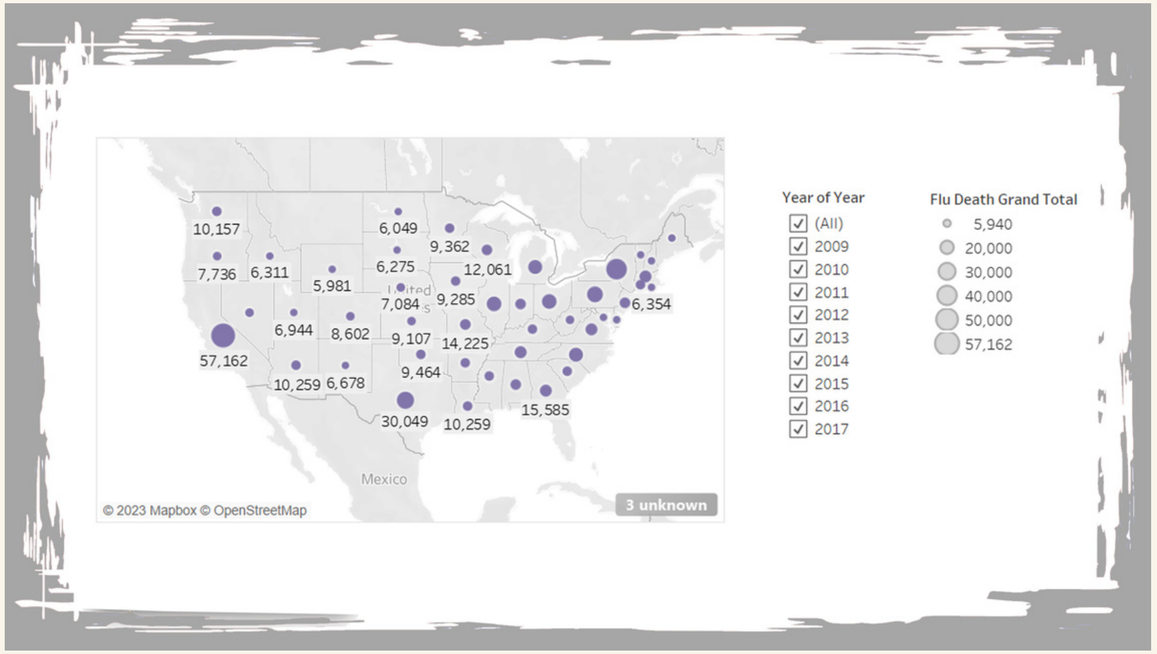


Correlation

Influenza	Population
730	626542.176
764	633101.501
771	644082.431
756	658126.888
792	658993.381
783	646890.235
885	643878.323
757	691297.943
940	719062
180	47808.709
180	48823.283
180	50856.978
180	51376.461
180	56874.692
180	54377.585
180	63707.815
180	70440.734

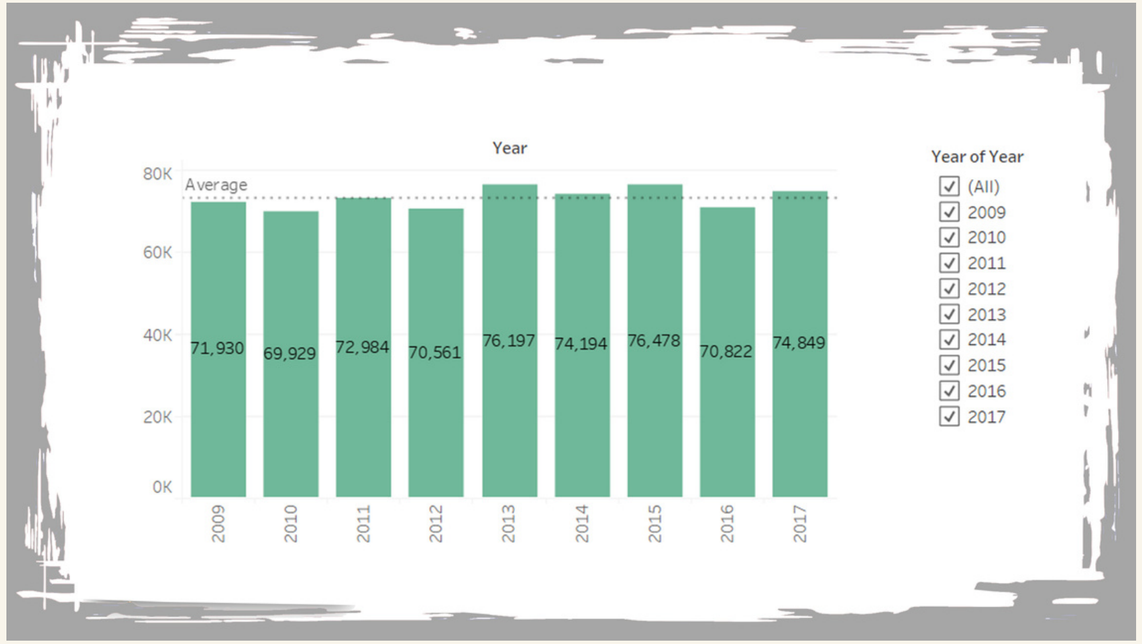
Dependent Variables:	Deaths rate caused by influenza / for the people have 65 years and above
Independent Variables:	Total population / **for the people have 65 years and above**
Null Hypothesis:	Compared to people under 65, people over 65 have the same or lower mortality rates
Alternative Hypothesis:	Compared to people under 65, people over 65 have a higher mortality rates
T-test type	one-tailed test, interested in one direction(if the sample average differs from the overall average in either direction)
Alpha	0.05
P-Value	1.7204E-171
Decision (Significance Level)	P-Value less than 0.05 (the significance level) so the null hypothesis has been rejected People of 65 years and older are at a higher risk of death during influenza season, so the medical staff should be distributed to clinics and hospitals in the states based on their populations over 65.
Next Step	The most vulnerable states will be highlighted in order to distribute medical staff accordingly

Overview Statistics



The flu death rate on a yearly basis

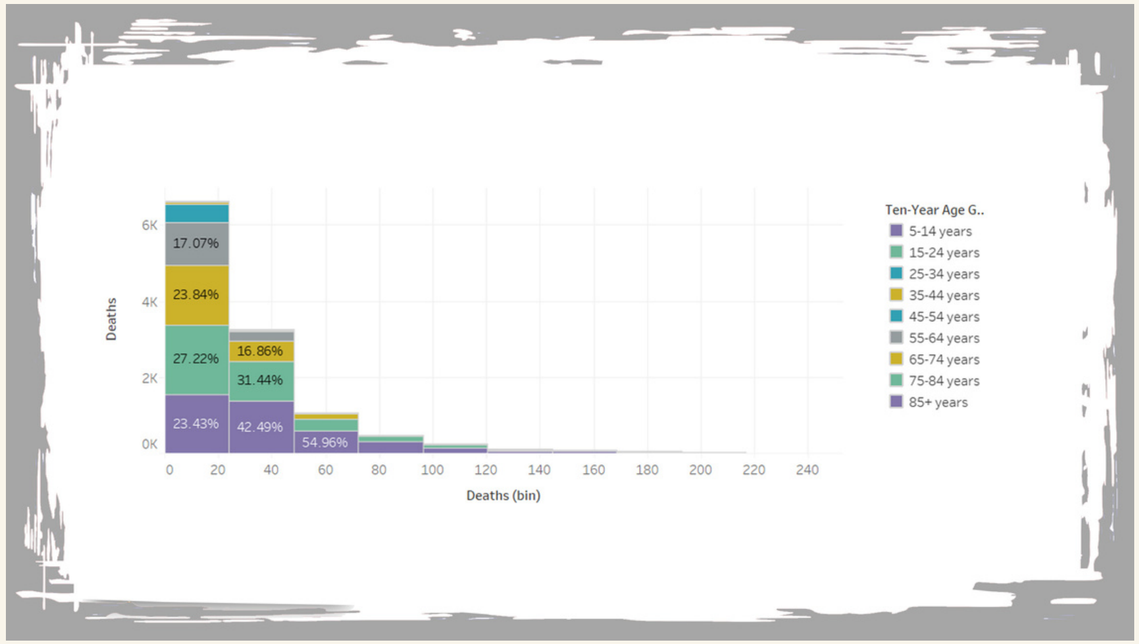
According to the latest estimates, a staggering annual toll of 73,104 lives is attributed to the influenza virus, underlining the significant human cost inflicted by this infectious respiratory illness each year



Flu death distribution over US States

Flu deaths distribution through US States is visually represented, where the size of each circle corresponds to the number of flu deaths in that specific US state, with larger circles indicating a greater magnitude of flu-related fatalities

Vulnerable Group Definition

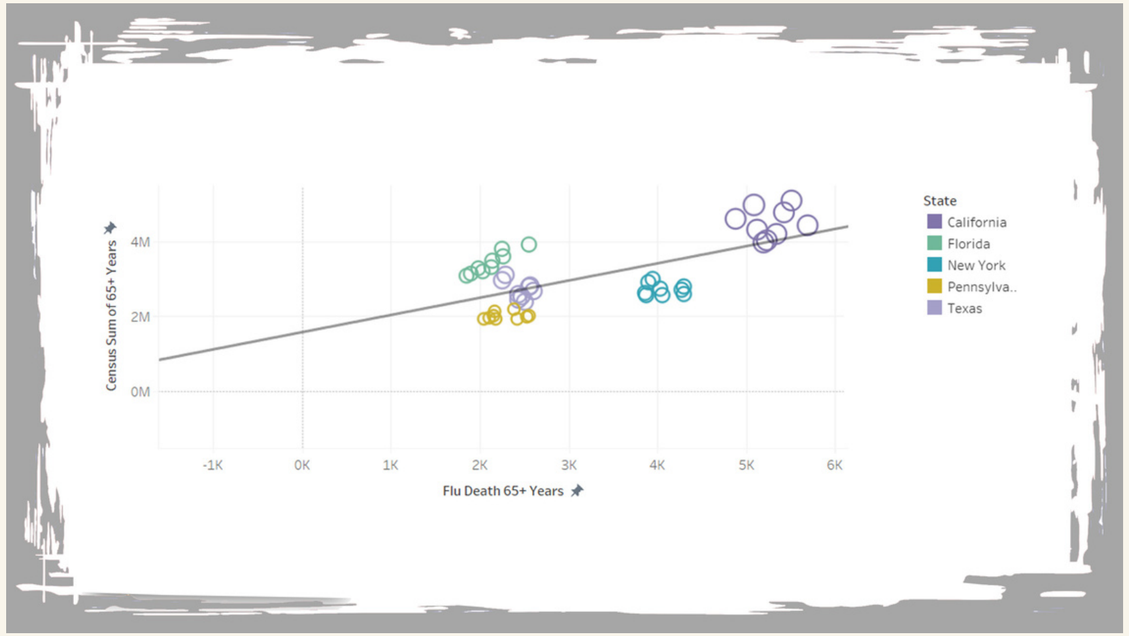


US influenza death rates categorized by age group every ten years between 2009 and 2017.

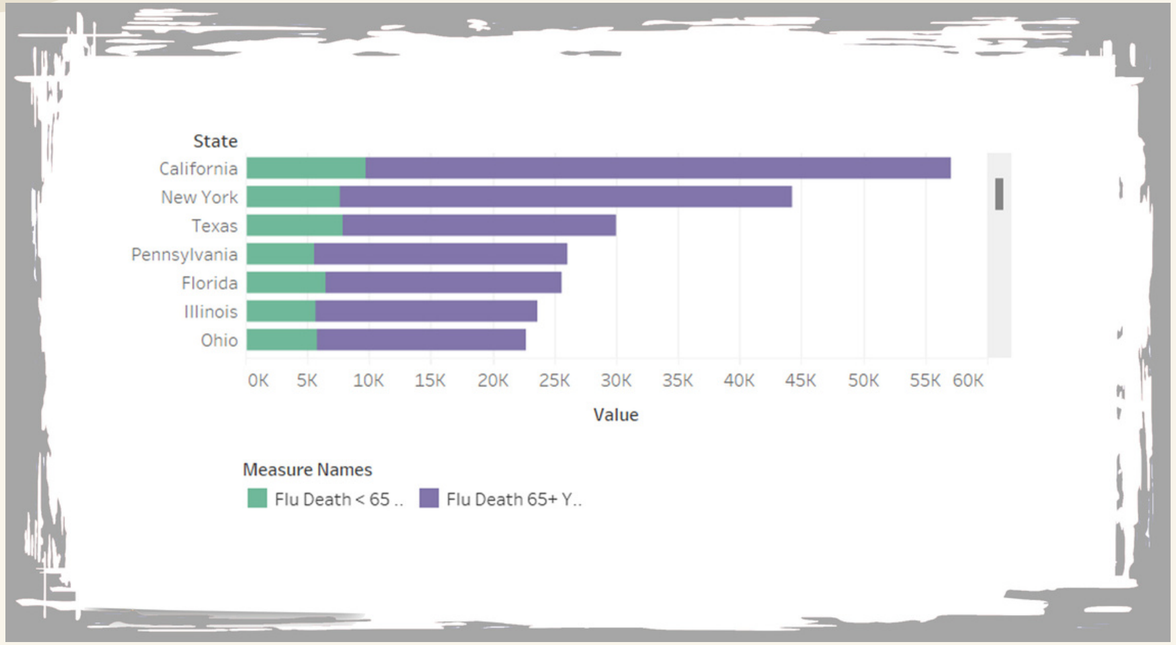
As individuals age, the incidence of influenza-related fatalities and deaths associated with the illness increases within each 10-year age bracket.

Relationship between US population and Flu death for people age 65 and above 2009-2017 / Top 5 States

Flu deaths distribution through US States is visually represented, where the size of each circle corresponds to the number of flu deaths in that specific US state, with larger circles indicating a greater magnitude of flu-related fatalities

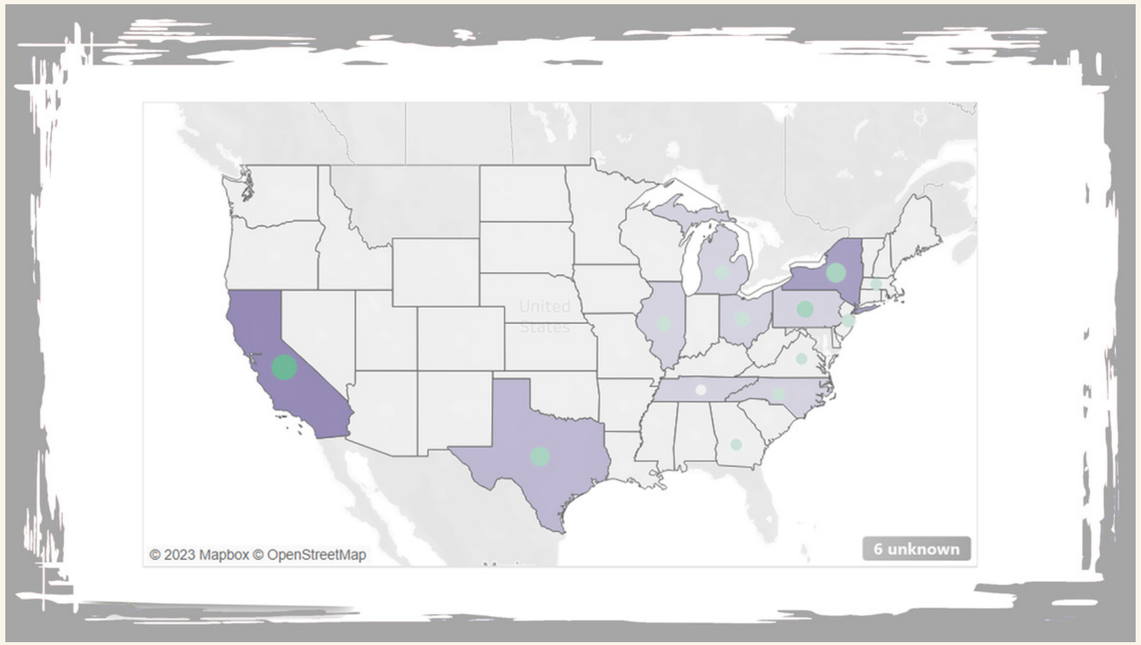


US States with the highest population density of individuals over the age of 65.



US Influenza deaths rates with total population for the people 65 years and above by state 2009-2017

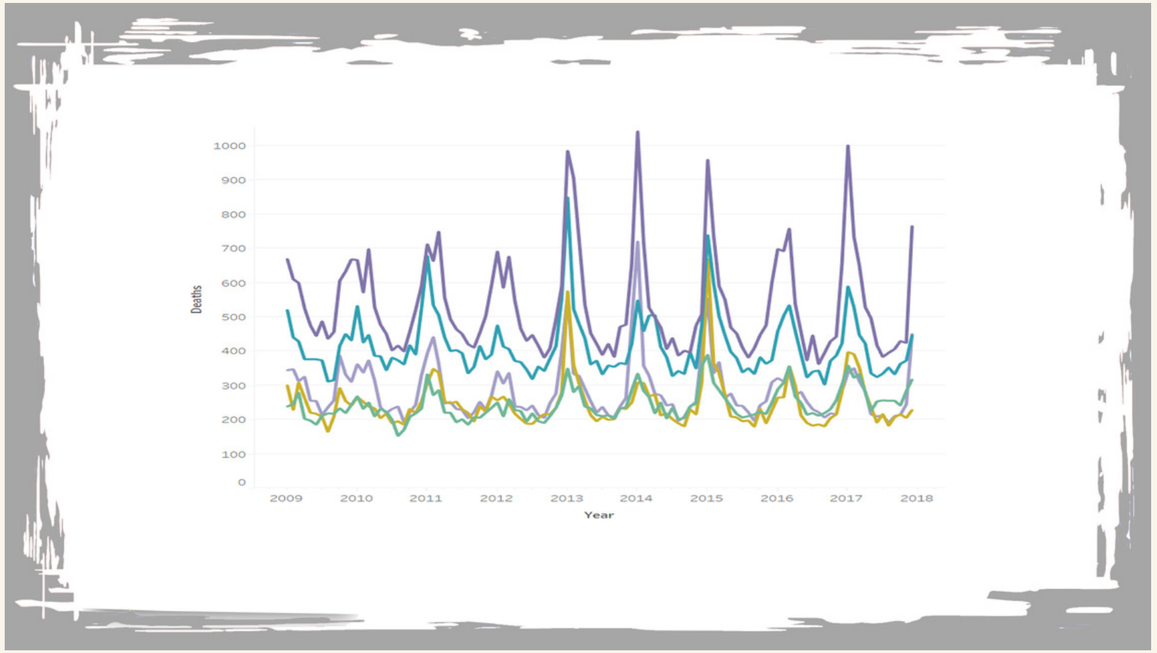
States with a large senior population (65 years and above) have higher rates of flu-related deaths. The intensity of color on the map indicates the severity of the death toll in each state, while the circles represent the number of individuals aged 65 and above in each area.



The percentage of flu-related deaths in the United States between 2009 and 2017, categorized by age group (those over and under 65), and broken down by state.

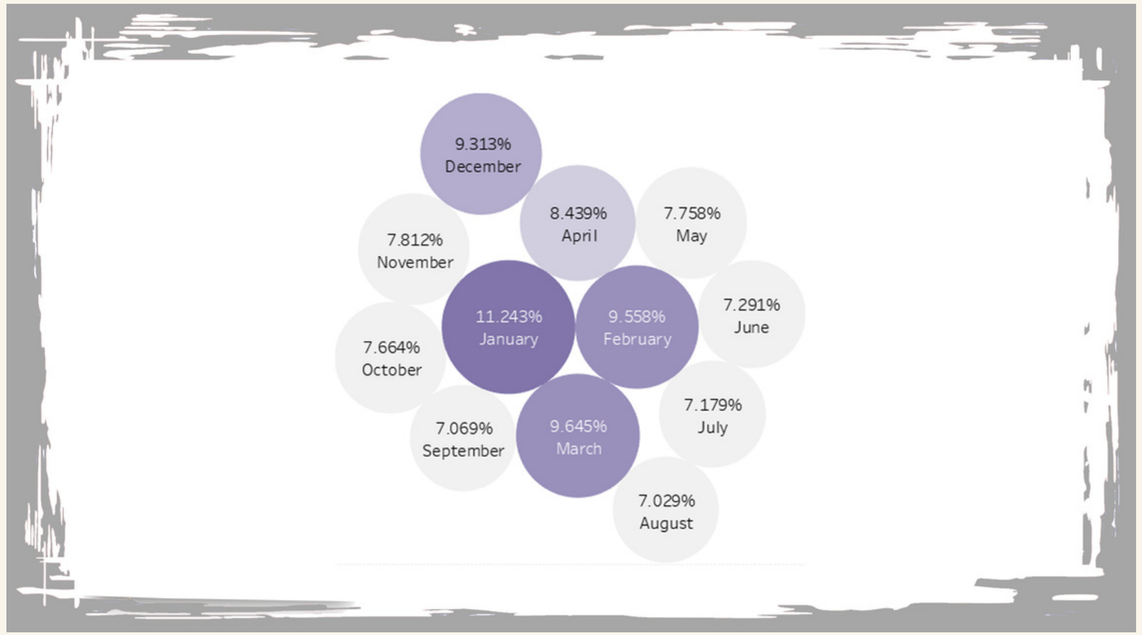
Outlined here are the top five US states boasting the highest concentration of individuals aged 65 years and above, indicating regions where a significant demographic of senior citizens resides

The most crucial time to allocate our resources.



The flu death rate on a monthly basis

This chart displays the percentage of monthly influenza-related fatalities, with darker shades of blue indicating a higher number of deaths.



Top 5 states with the highest number of flu-related deaths in the US.

Flu-related deaths usually peak during the winter months, specifically between December and March.

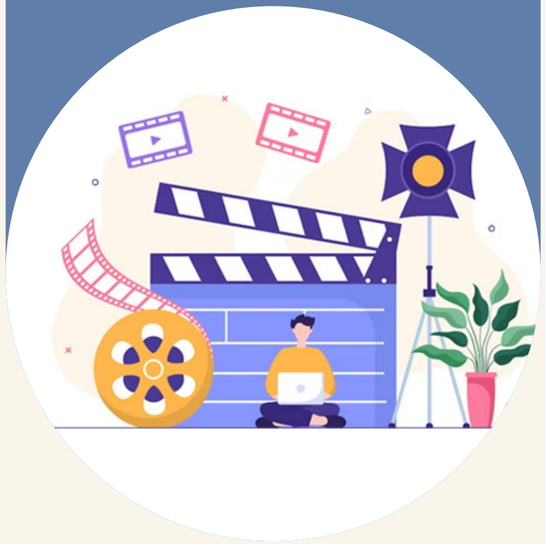
Recommendations

- To address the urgent situation at hand, it is crucial to allocate additional resources to the states of California, New York, Texas, Pennsylvania, and Florida. These particular states have experienced a significant increase in the number of deaths, making it imperative to provide them with the highest amount of support and assistance.

- Among the various demographics, the vulnerable population aged 65 and above requires a greater allocation of additional resources compared to other age groups. Recognizing their heightened susceptibility to health complications, it becomes even more essential to prioritize their needs and ensure they receive the necessary aid and care.

- it is imperative to expedite the delivery of these additional resources. By aiming to have them sent no later than November, we can proactively prepare for the typical peak months of December, January, and February when flu cases tend to surge. By taking prompt action, we can enhance the capacity of healthcare systems, provide adequate support to affected individuals, and minimize the impact of the flu season on the population at large.

Rockbuster Stealth LLC



Project Overview

- **Study introduction:** Rockbuster Stealth LLC, a former global movie rental company, is adapting to the challenges posed by streaming giants like Netflix and Amazon Prime. In response, the management team of Rockbuster Stealth has decided to leverage their current movie licenses and venture into the realm of online video rental services.
  Rockbuster Stealth's business intelligence department needs assistance in devising an effective launch strategy for the new online video service.

- **Objectives:** identifying the movies that had the highest and lowest impact on revenue, determining the average rental duration for all videos, identifying the countries where Rockbuster customers are located, pinpointing the customer base with a high lifetime value, and exploring potential variations in sales figures across different geographic regions.

- **Skills:** using a variety of skills related to managing and analyzing data, including working with relational databases, cleaning and filtering data, summarizing information, joining tables, querying databases, using subqueries, and utilizing common table expressions.

- **Tools** utilized for Study purposes:

- You can find comprehensive information by visiting
  https://github.com/OmarMohandes/

Overview Statistics

Number of countries in which Rockbuster present
108 Country

Total number of customers
599 Customer

Total revenue
$61.312

Total number of movies
1001 Movie

Highest Spend
Eleanor Hunt From **Runion/Africa** spend the most on renting movies with **$ $211.**

Rental Rate	
Minimum	$0.99
Maximum	$4.99
Average	$2.98

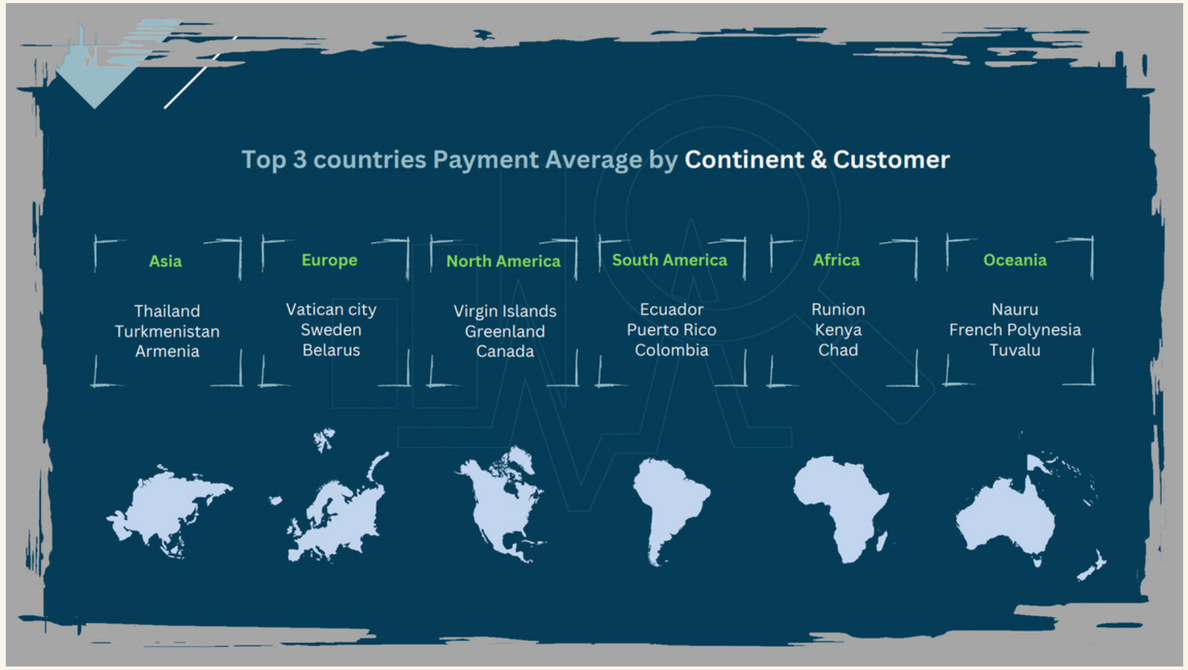
Rental Duration	
Minimum	3 Days
Maximum	7 Days
Average	4.98 Days

Movie Length	
Minimum	46 Min
Maximum	185 Min
Average	115.27 Min

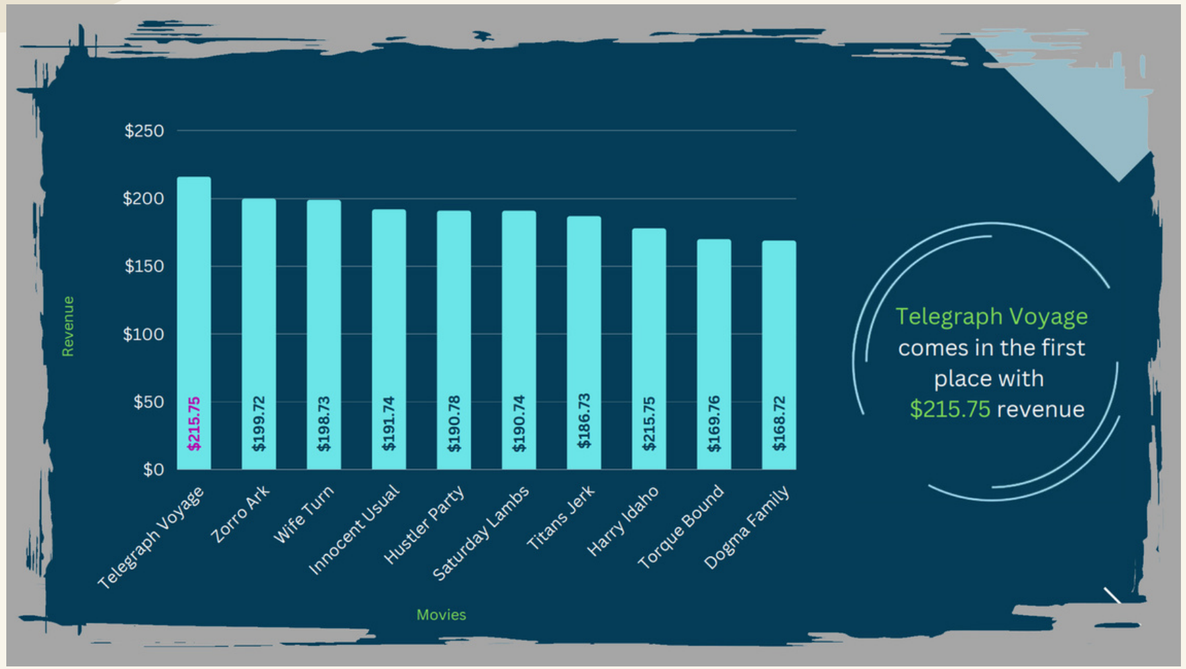
Replacement Cost	
Minimum	$9.99
Maximum	$29.99
Average	$19.98

Breakdown of the average payments made by customers from the top three countries, categorized by continent.

The statistics reveal that Rockbuster Stealth LLC has a presence in multiple countries, resulting in an expansion of its customer base. Additionally, the statistics reflect the company's involvement in a diverse range of movies, along with its association with the highest expenditure on films. This data underscores the company's significant impact and widespread influence in the film industry across different nations.

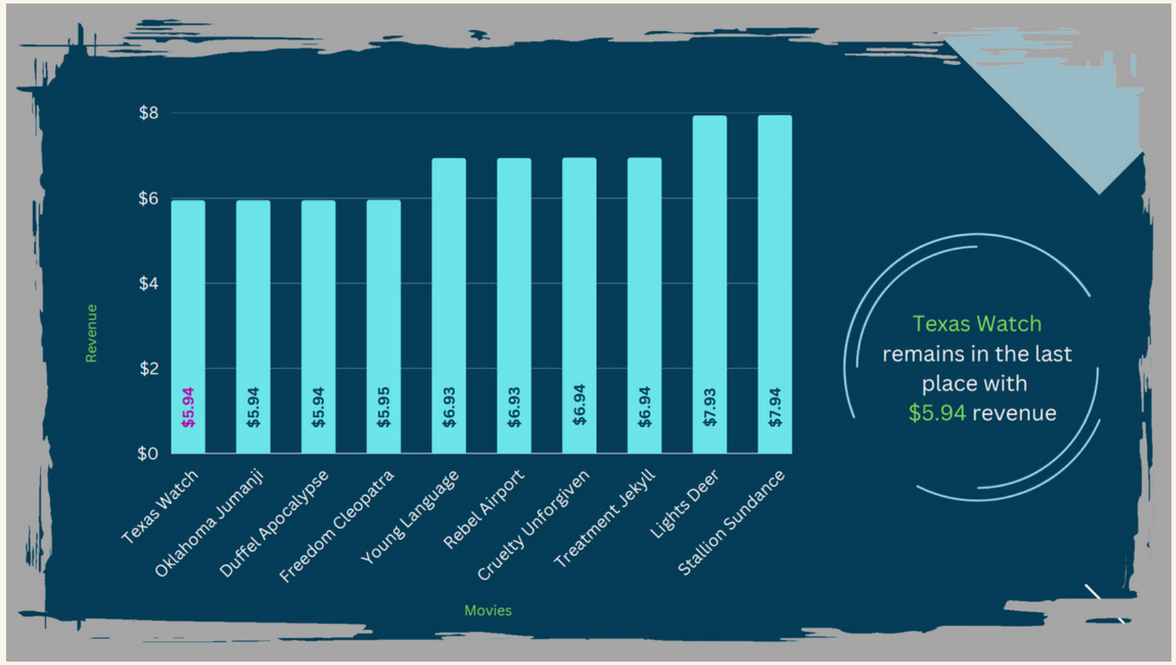


The movies that generated the highest and lowest revenue

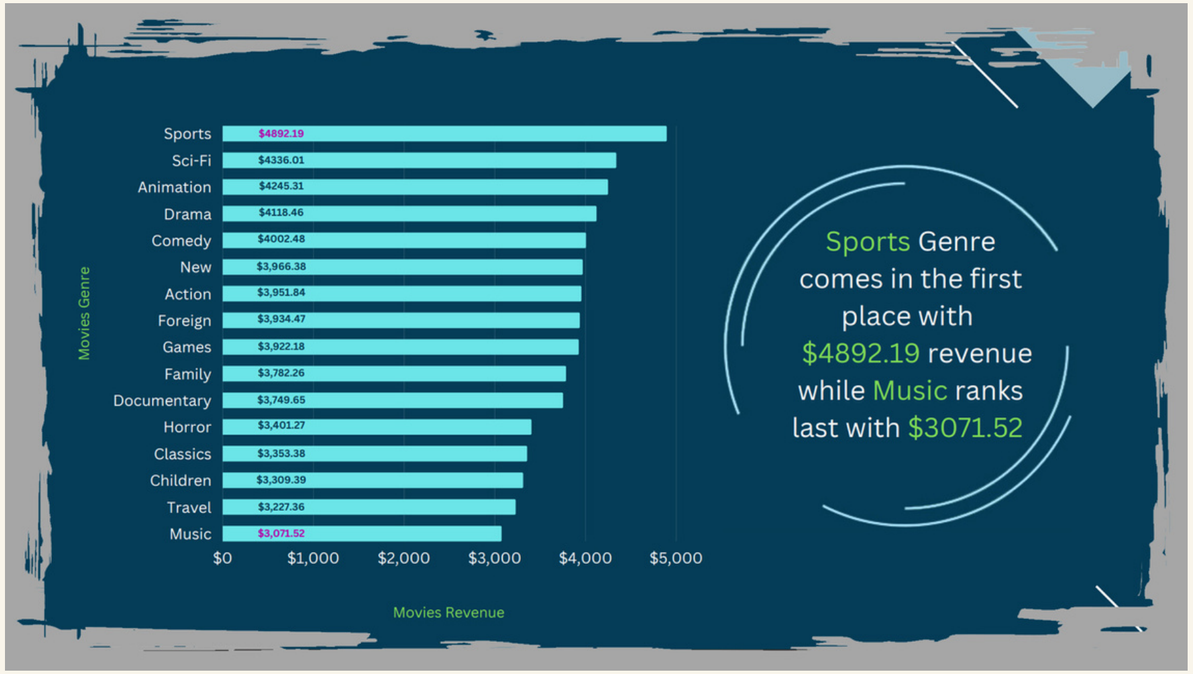


The movies with the highest revenue.

The movie with the lowest revenue



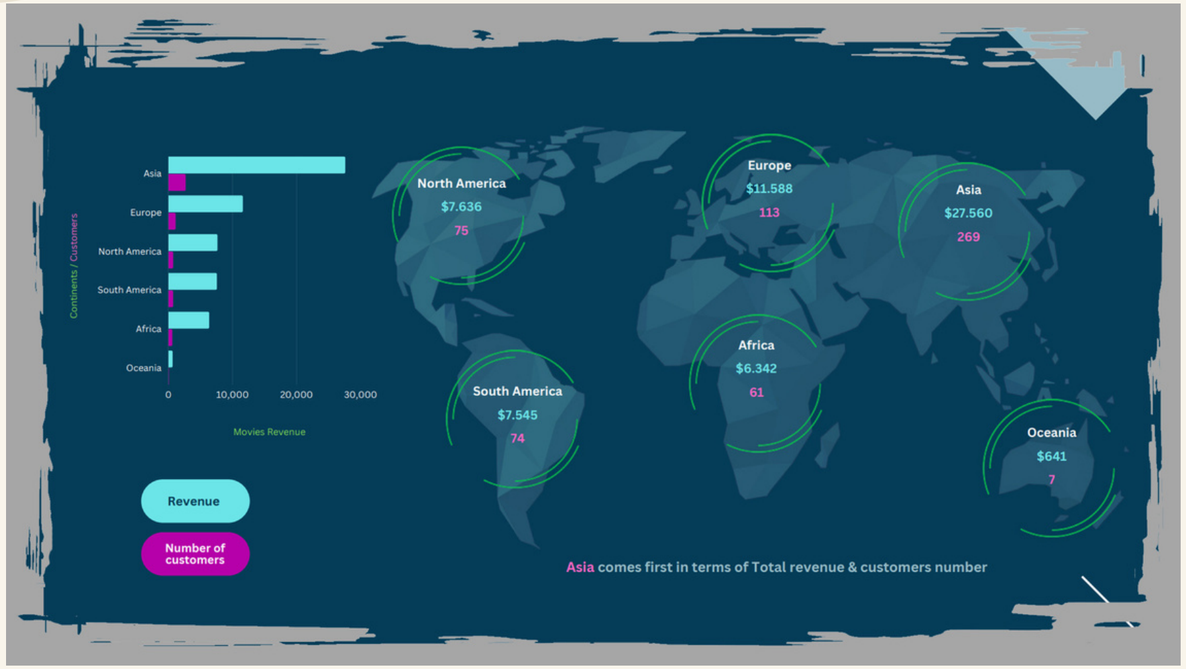
The average rental duration and the highest/lowest revenue generated by movies, categorized by Genre



The highest and lowest revenue-generating movies by Genre.

On average, The duration of movie rentals by Genre

Sales and Customer distribution by geography.



Sales breakdown by continent and customer location.

Distribution of Customers by country.

Recommendations

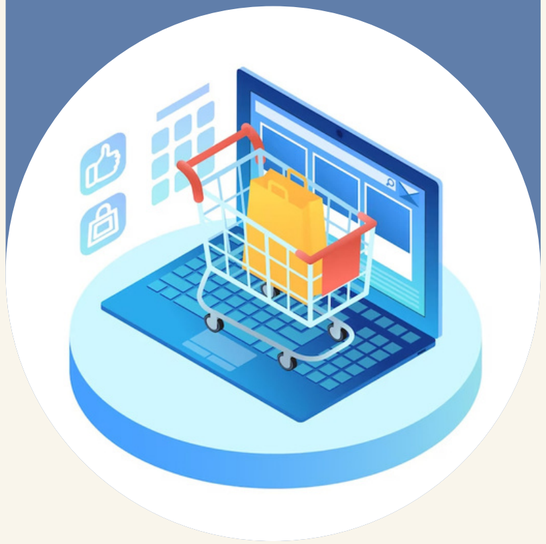
- To maximize annual returns, it is recommended to target markets in the following order: Asia, Europe, North America, South America, Africa, and Oceania. Stick to this strategy for optimal results.

- Our marketing strategy prioritizes films with higher returns, including Sports, Sci-Fi, Animation, Drama, and other genres.

- Let's prioritize our focus and target on the countries with the highest average payment per customer.

- To boost sales, consider sending exclusive offers such as discounts and multi-buys to your most valuable customers. Additionally, prioritize notifying these customers of new releases and promotions to encourage repeat business. It may also be beneficial to optimize your inventory by removing low-revenue titles and focusing on high-revenue ones.

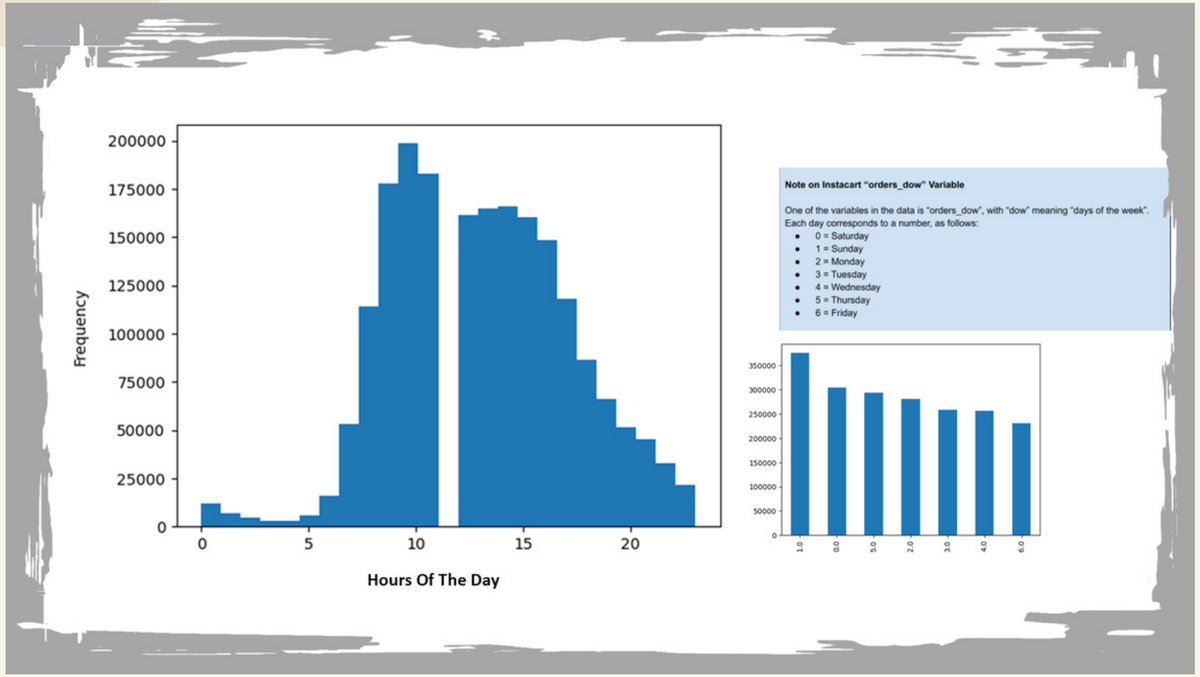
Instacart Basket



Project Overview

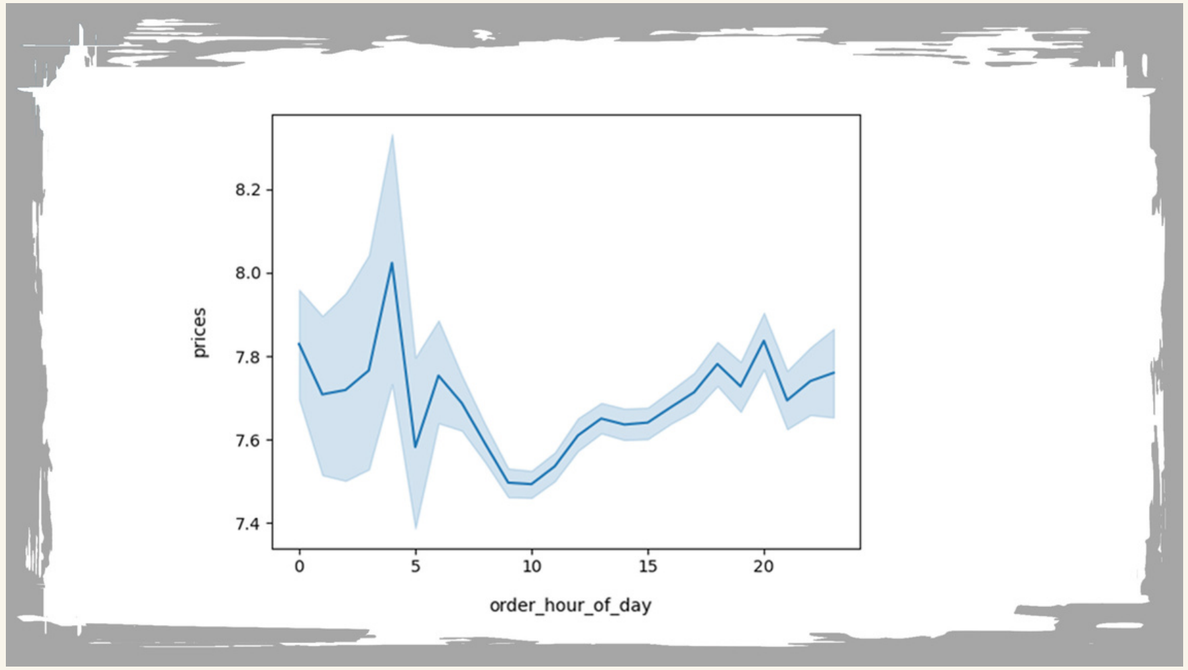
- **Study introduction:** Instacart is an online grocery store that operates through an app. They have impressive sales, but they're looking to uncover more information about their sales patterns. They want to perform an initial data and exploratory analysis of some of their data in order to derive insights and suggest strategies for better segmentation based on the provided criteria.
- **Objectives:** The company requires information on the busiest days and hours during the week, including times with the most orders, to schedule ads during low-order periods. Additionally, they seek insight on peak times when customers spend more, to help direct product advertisements. The marketing and sales teams aim to simplify product price range groupings for better direction of their efforts, considering Instacart's multiple products with varying price tags. They also want to identify the most frequently ordered products and departments for efficient marketing strategies.
- **Skills:** Performing data consistency checks, merging and wrangling data, deriving variables, and grouping and aggregating data using Python libraries. Additionally, Creating visualizations and reports using Excel.
- **Tools** utilized for Study purposes:
- You can find comprehensive information by visiting https://github.com/OmarMohandes/

Answering Key Questions

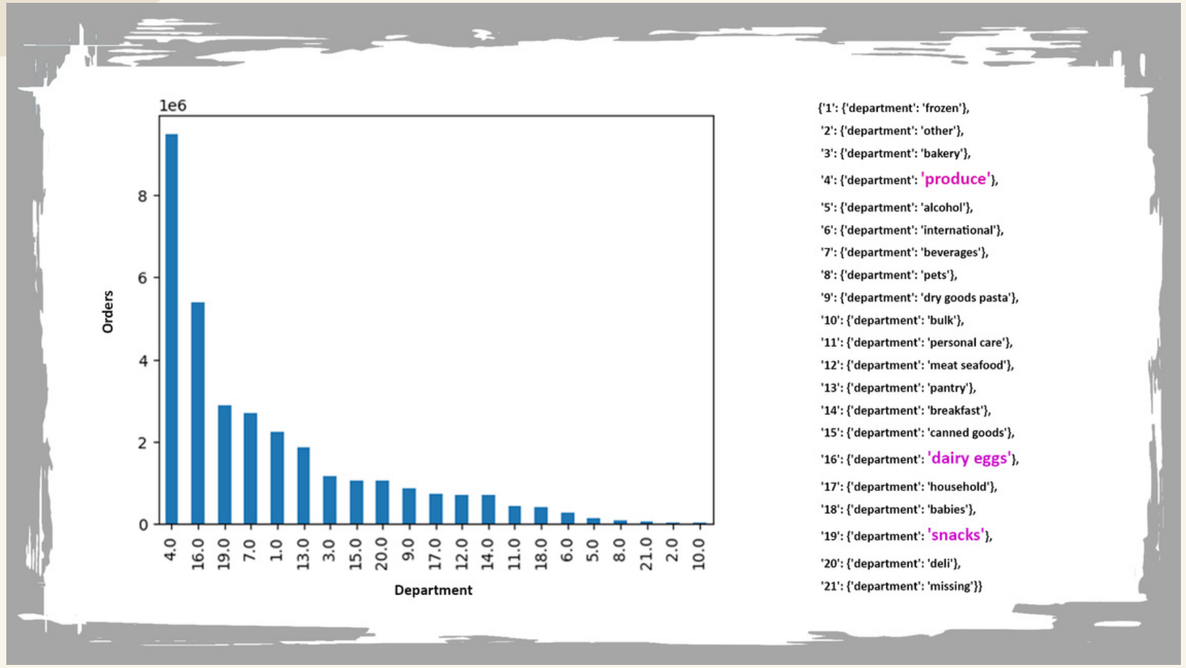


The sales team is interested in identifying the peak hours of the day when customers tend to spend the most money. This information can help them tailor their product advertisements according to those specific times.

The sales team requires information on the peak days and hours, i.e. the ones with the highest number of orders, to plan their ad schedules accordingly and avoid times with low order volume.

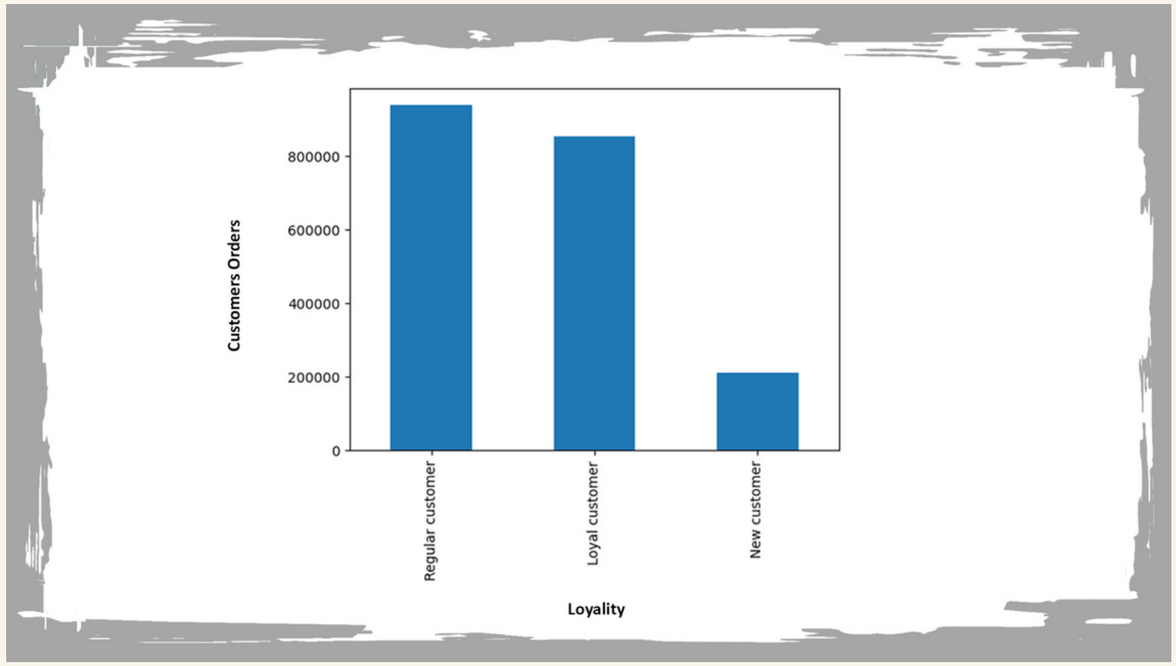


Answering Key Questions

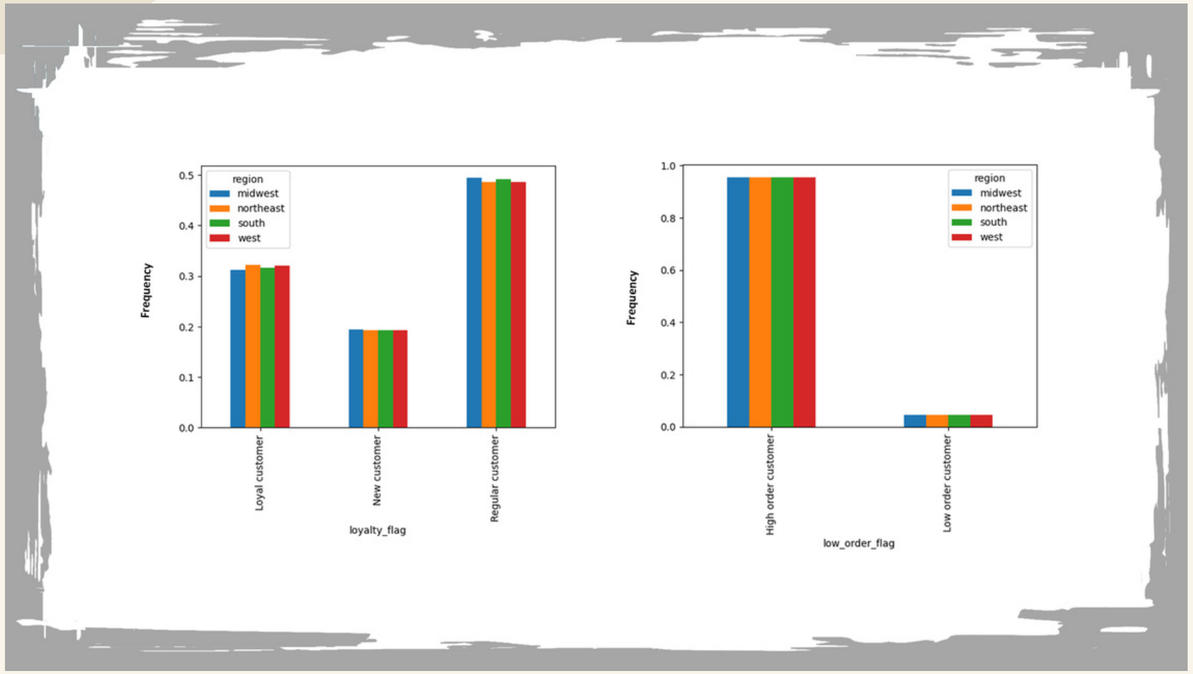


Can you provide information on the frequency at which Instacart users return to the platform, particularly in terms of their brand loyalty?

The marketing and sales teams are seeking information regarding the popularity of certain types of products. They want to identify the departments that receive the most frequent product orders.

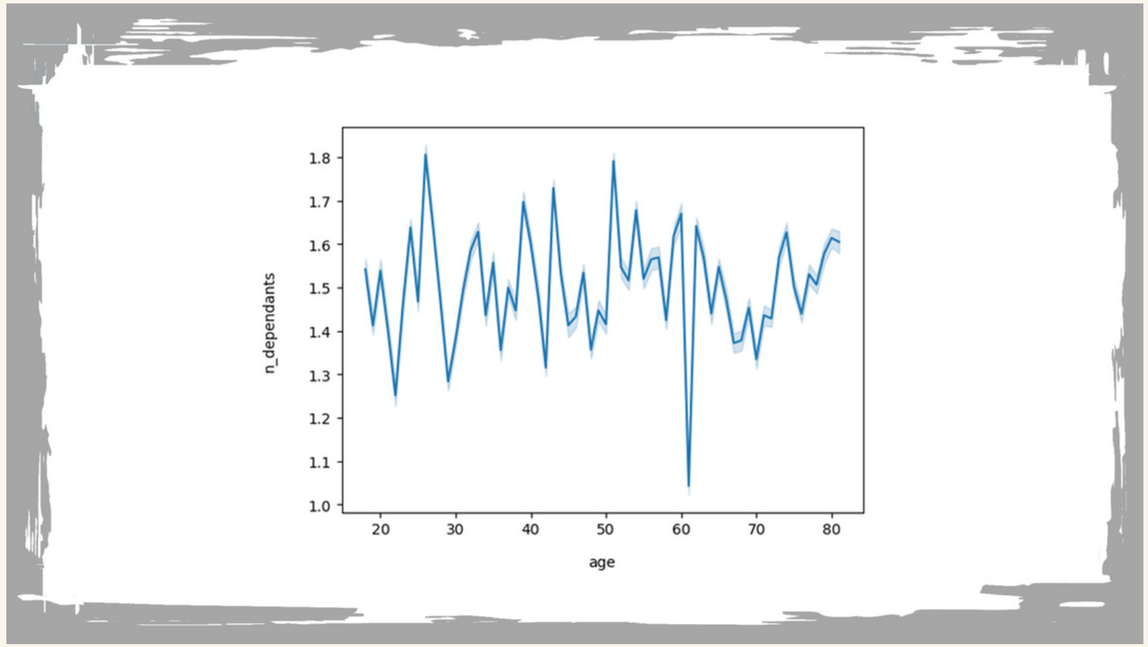


Answering Key Questions

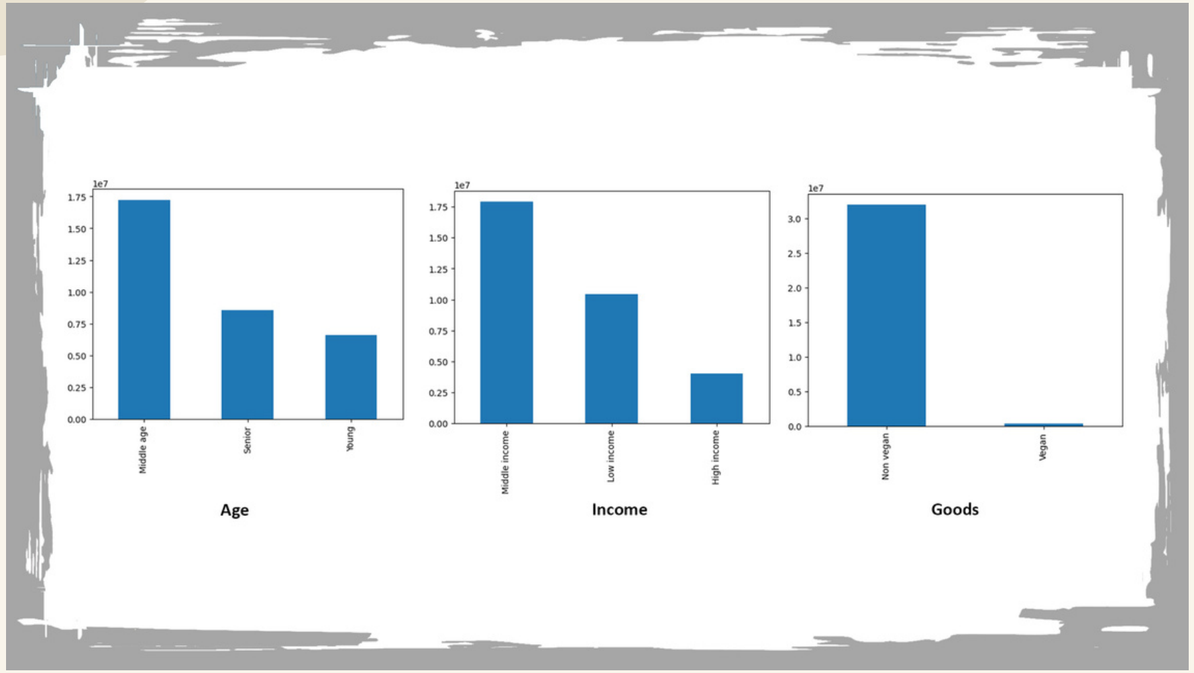


Do ordering habits have any correlation with age and family status?

Is there a variation in ordering patterns depending on a customer's loyalty level and region?

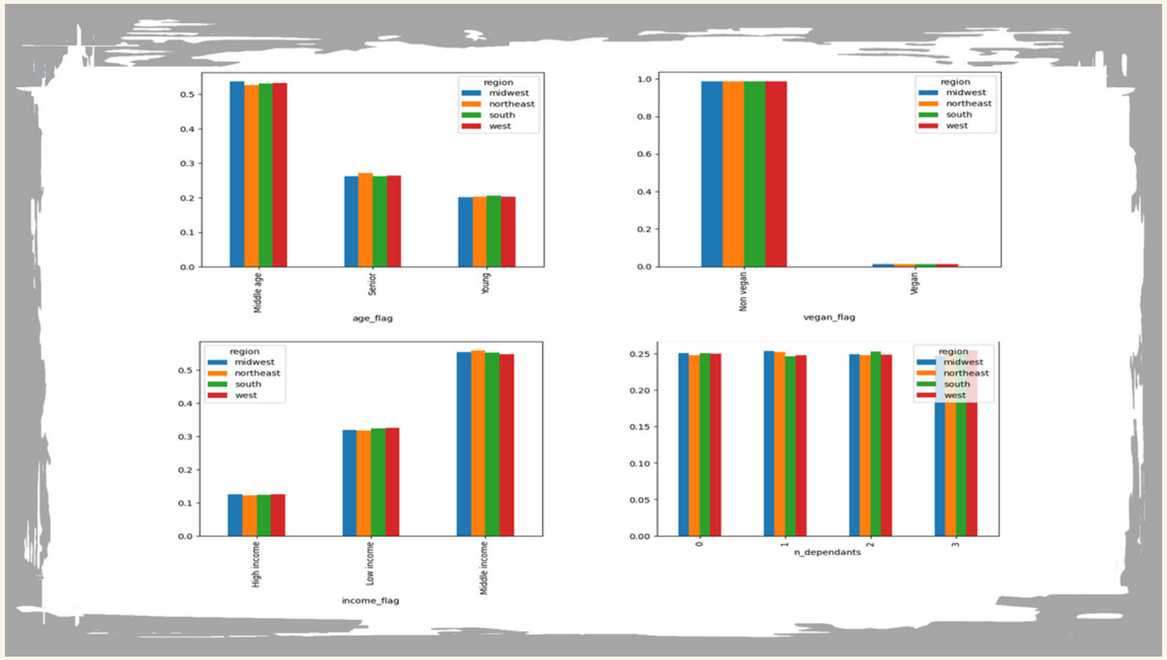


Answering Key Questions



Can you identify variations in the purchasing patterns of individual customer profiles? This includes the amount spent on orders, how often they place orders, and the specific products they tend to purchase.

What types of demographic information can be classified such as age, income, specific goods, and family status?



Recommendations

- To improve Instacart's advertising strategy, scheduling ads during times of low order volume, such as Friday at 3 AM, may be ideal. However, this could result in fewer people seeing the ads. Therefore, the most feasible time to schedule ads would be at 3 PM.
- It is recommended that Instacart triggers most of their ads early in the morning since that is when most people make their purchases. To increase sales, it is important for Instacart to advertise in departments 4, 16, and 19.
- Further exploration is necessary to determine what encourages new customers to become regular customers. It appears that most regular customers are from the Midwest. To better understand customer behavior, Instacart should consider incorporating additional variables.
- Although there are some differences in the products ordered by vegan customers, they do not make up a large portion of Instacart's customer base. However, producing ads based on their dietary preferences could still be profitable for Instacart. Overall, Instacart should continue to analyze customer behavior and adapt their advertising strategy accordingly.

Mental Health Disorder

Project Overview

- Study introduction: The dataset provides data on the prevalence of various mental health disorders worldwide, such as schizophrenia, bipolar disorder, eating disorders, anxiety disorders, drug use disorders, depression, and alcohol use disorders. The data is presented in a user-friendly visual format, allowing for a better understanding of how these issues affect people's lives and the implications associated with them.

- Objectives: The objective of this study is to elucidate the factors that exert an influence on mental health outcomes. This entails a comprehensive analysis of diverse determinants, including anxiety, eating disorders, depression, and alcohol use, in order to discern their associations with overall mental well-being. By successfully identifying these contributing factors, it becomes possible to develop targeted interventions and formulate policies that address the underlying causes of mental health concerns, consequently fostering the promotion of positive mental health.

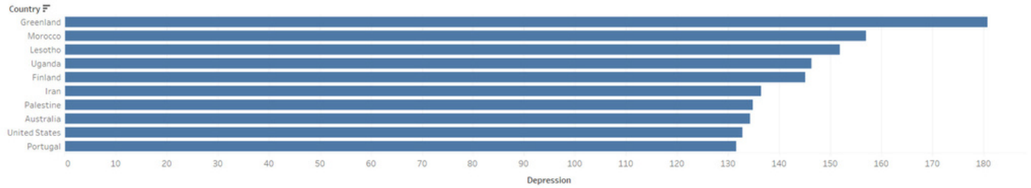
- Skills: Performing data consistency checks, merging and wrangling data, deriving variables, and grouping and aggregating data using Python libraries.

- Tools utilized for Study purposes:

- You can find comprehensive information by visiting https://github.com/OmarMohandes/

Overview Statistics

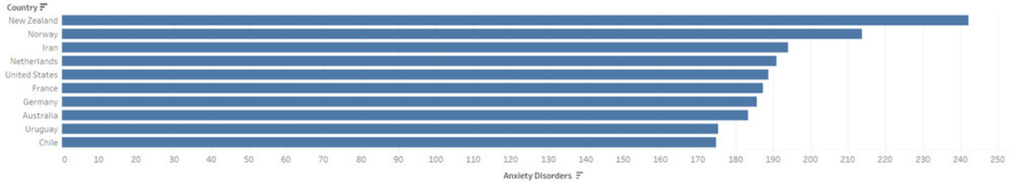
Top 10 countries with the highest depression rate.



The following examination focuses on the top ten countries characterized by the highest rates of depression, with Greenland occupying the foremost position and Portugal being situated at the bottom of this ranking

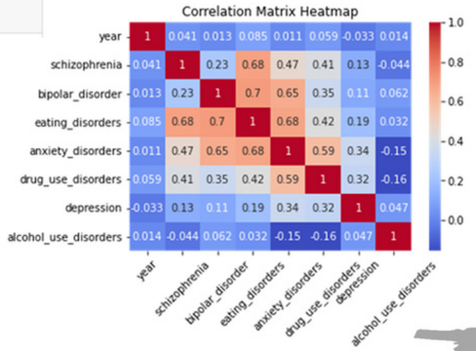
The present analysis highlights the top ten countries exhibiting a high prevalence of anxiety disorders, with New Zealand being identified as the country with the highest incidence, and Chile being positioned at the lowest rank within this list.

Top 10 countries with anxiety disorders.



Correlations

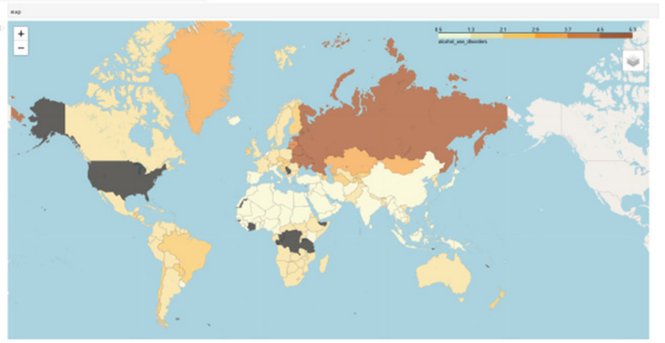
```
# Create a heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')

# Set plot attributes
plt.title('Correlation Matrix Heatmap')
plt.xticks(rotation=45)
plt.yticks(rotation=0)

# Show the plot
plt.show()
```



A discernible correlation can be observed between the act of dietary consumption and the occurrence of anxiety disorders, which is subsequently followed by the connection between food intake and bipolar disorder.

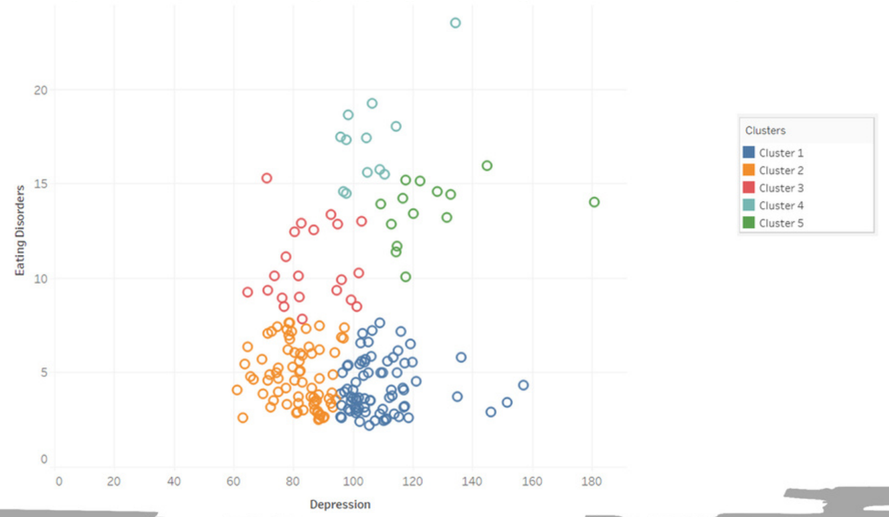
By utilizing this map, we can discern countries exhibiting high levels of alcohol consumption, where alcohol use potentially constitutes one of the primary factors contributing to mental health issues. However, further investigation is required to establish a more comprehensive understanding.

```
folium.Choropleth(
    geo_data = country_file,
    data = data_to_plot,
    columns = ['country', 'alcohol_use_disorders'],
    key_on = 'feature.properties.ADMIN',
    fill_color = 'YlOrBr', fill_opacity=0.6, line_opacity=0.1,
    legend_name = "alcohol_use_disorders").add_to(map)

folium.LayerControl().add_to(map)

map
```



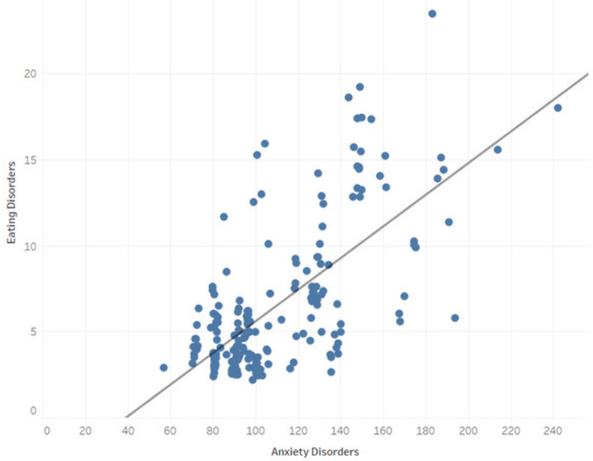
Correlations

Cluster Analysis that shows the grouping of Eating Disorder and Depression



In the realm of predictive modeling, linear regression emerges as a prominent algorithm for forecasting future outcomes when the target variable exhibits a continuous nature. Consequently, a linear regression analysis was employed, allowing for the derivation of a regression line that provides estimates for the predicted values of anxiety.

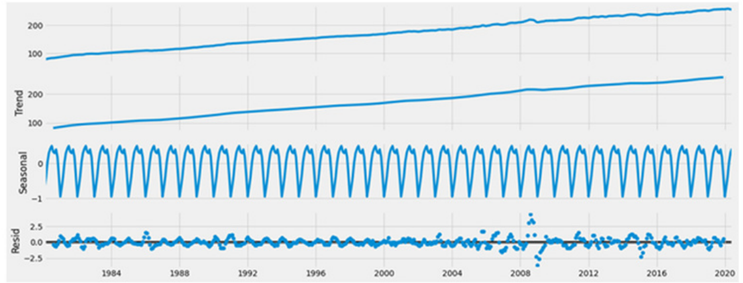
Regression linear and the relasioship between Eating disorder and Anxiety disorder



This analytical approach enables the recognition of unique patterns and shared similarities among individuals, thereby offering valuable insights into diverse manifestations of mental health conditions, associated risk factors, and treatment requirements.

Time Series Analysis

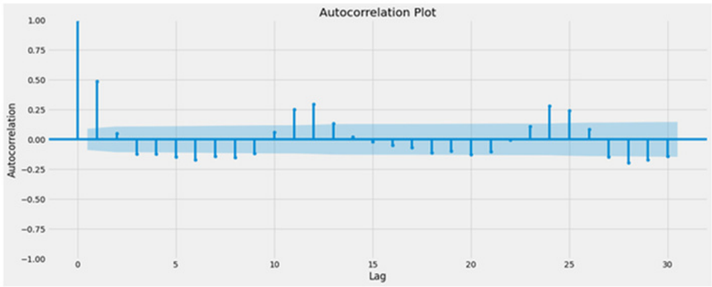
```
import statsmodels.api as sm

decomposition = sm.tsa.seasonal_decompose(data_sub, model='additive')

from pylab import rcParams # This will define a fixed size for all special charts.

rcParams['figure.figsize'] = 18, 7

decomposition.plot()
plt.show()
```



The graph displays the data level across various conditions. The second graph demonstrates a similar trend to the data level, indicating a minimal presence of noise. The trend chart exhibits a consistent upward trajectory in inflation. The third component, seasonality, is evident through periodic fluctuations represented by a spikey curve. In the absence of seasonality, the curve would remain flat. Finally, the last component pertains to residual or noise, which appears to be minimal in the given graph.

Within the depicted visual representation, the vertical lines correspond to the lags present in the series, while the blue area denotes the confidence interval. Notably, when the lines extend beyond the upper threshold of the confidence interval, it indicates a significant correlation between the lags. Moreover, the presence of numerous lags surpassing this interval suggests non-stationarity of the data.

```
# Create the autocorrelation plot
plot_acf(data_diff['Value'], lags=30)

# Add labels and title
plt.xlabel('Lag')
plt.ylabel('Autocorrelation')
plt.title('Autocorrelation Plot')

# Show the plot
plt.show()
```



Data Analysis Insights

Depression and Mental Well-being:

The study demonstrated a positive correlation between depression and mental well-being. Higher levels of depression were consistently linked to diminished mental health, including lower levels of life satisfaction, increased psychological distress, and impaired daily functioning.

Eating Disorders and Mental Well-being:

The findings indicated a positive correlation between eating disorders and mental well-being. Individuals with eating disorders experienced lower levels of mental well-being, characterized by symptoms such as body dissatisfaction, negative self-image, and psychological distress.

Alcohol Use and Mental Well-being:

We can discern countries exhibiting high levels of alcohol consumption, where alcohol use potentially constitutes one of the primary factors contributing to mental health issues. However, further investigation is required to establish a more comprehensive understanding.

Anxiety and Mental Well-being:

The analysis revealed a significant positive correlation between anxiety and decreased mental well-being. Higher levels of anxiety were associated with poorer mental health outcomes, including increased psychological distress, impaired functioning, and reduced overall well-being.

Insights and further information can be found at:

https://github.com/OmarMohandes

https://public.tableau.com/app/profile/omar3388

omar.mohandes@gmail.com

+49 1476 227701

THANK YOU!