DATA ANALYSIS PROJECTS

## AMAZON

****

In the vast realm of e-commerce, Amazon, the titan of online retail, sought to enhance its recommendation system. The objective of this project is to analyze user behavior in Amazon and determine which products can be recommended to users based on their buying patterns. By leveraging data from a SQLite database, we aim to identify key insights that can help Amazon make personalized product recommendations. This project will also investigate the differences in behavior between frequent and non-frequent viewers, as well as examine the verbosity of frequent users. Furthermore, sentiment analysis will be conducted to gain insights into user sentiments towards the products.

1. Reading Data from SQLite Database:

* Accessed the SQLite database containing Amazon user data.
* Extracted relevant information such as user profiles, product details, and user reviews.

1. Data Preparation:

* Performed data cleaning and preprocessing to ensure the accuracy and consistency of the data.
* Handled missing values, removed duplicates, and corrected data types as necessary.

1. Analyzing User Behavior for Product Recommendations:

* Explored user buying patterns to identify users who are more likely to buy products.
* Utilized conversion rates to determine users who have a higher likelihood of making a purchase.
* Prepared the data specifically to address the problem of recommending products to users.

1. Identifying Products with a High Number of Reviews:

* Analyzed the dataset to identify which products have received a significant number of reviews.
* This information can help Amazon identify popular products that may be recommended to users.

1. Comparing Behavior of Frequent and Non-Frequent Viewers:

* Segmented the users based on their viewing frequency.
* Investigated differences in behavior, such as purchase history, browsing patterns, and preferences, between frequent and non-frequent viewers.
* These insights can aid Amazon in tailoring recommendations to different user segments.

1. Analyzing Verbosity of Frequent Users:

* Examined the level of engagement of frequent users by analyzing their communication patterns, such as posting reviews or participating in discussions.
* Determined whether frequent users tend to be more verbose and contribute more to the Amazon community.

1. Sentiment Analysis:

* Conducted sentiment analysis on user reviews to gain insights into user sentiments towards products.
* Examined positive and negative sentiments expressed by users to understand their preferences and potential improvements for products.

Conclusion: Through this comprehensive analysis of user behavior in Amazon, we have gained valuable insights that can inform personalized product recommendations. By understanding the differences between frequent and non-frequent viewers, Amazon can tailor its recommendations to suit different user segments. The identification of products with a high number of reviews provides further guidance for product recommendations. Additionally, analyzing the verbosity of frequent users can help Amazon understand user engagement levels and foster a vibrant community. Lastly, sentiment analysis enables Amazon to gauge user sentiments towards products, facilitating improvements and enhancing customer satisfaction.

## 2. BITCOIN



In the vast landscape of digital currencies, Bitcoin stands as a true pioneer and a force to be reckoned with. With its revolutionary concept and skyrocketing popularity, Bitcoin has captured the attention of investors, technologists, and enthusiasts alike. The purpose of this project is to analyze historical price data of Bitcoin, one of the most prominent cryptocurrencies in the world. By examining the price movements and trends, we aim to gain valuable insights into the behavior of Bitcoin and provide a comprehensive analysis for investors, researchers, and enthusiasts.

1. Let's Read Data & Perform Basic Analysis:

* Acquired a dataset containing historical Bitcoin price data.
* Explored the dataset's structure, size, and overall layout.
* Familiarized ourselves with the information at hand to set the stage for deeper analysis.

1. Data Pre-processing:

* Checked data types to ensure correctness and accuracy.
* Examined the dataset for missing values and employed appropriate techniques to handle them, ensuring reliable analysis.
* Conducted a thorough search for duplicate entries, removing redundant data points.

1. Analyzing Change in Price of Bitcoin Over Time:

* Visualized the price data on a line chart to observe trends, patterns, and fluctuations in Bitcoin's price over time.
* Analyzed the historical performance to identify periods of growth, decline, and stability.

1. Analyzing Open, High, Low, Close Values of Bitcoin:

* Examined the open, high, low, and close values associated with each date.
* Employed candlestick or OHLC charts to visualize the trading range and price dynamics of Bitcoin.
* Identified bullish or bearish trends and patterns in the price movement.

1. Analyzing Closing Price on Normal Scale & Log-Scale:

* Plotted the closing price on both a normal scale and a logarithmic scale.
* Normal scale visualization provided insights into absolute price values, highlighting drastic changes and outliers.
* Logarithmic scale visualization assessed percentage changes, identifying exponential growth or decline.

1. Analyzing Closing Price on Yearly, Quarterly, and Monthly Basis:

* Calculated the average closing price for each year, quarter, and month.
* Analyzed long-term and short-term trends to identify seasonality, cyclicality, and irregular patterns in Bitcoin's price.

1. Analyzing Daily Change in Closing Price of Bitcoin:

* Calculated the difference between the closing price of each day and the previous day.
* Visualized the daily change in closing price to assess volatility and magnitude of price fluctuations.
* Identified periods of significant price changes and evaluated overall stability or volatility.

Conclusion: Through this comprehensive analysis of Bitcoin's price data, we have gained valuable insights into its behavior, trends, and patterns. This project provides a thorough understanding of Bitcoin's price movements over time, as well as its open, high, low, and close values. By analyzing the closing price on different scales and time intervals, we have uncovered valuable information for investors and researchers. Additionally, the daily change in closing price analysis helps in assessing the volatility and magnitude of price fluctuations.

## 3 . COVID 19



The COVID-19 pandemic, caused by the novel coronavirus SARS-CoV-2, has had a profound impact on global health, economies, and societies. Since its emergence in late 2019, the virus has spread rapidly across borders, leading to millions of infections and deaths worldwide. The pandemic has caused significant disruptions to healthcare systems, economies, and social well-being. Efforts to combat the virus have included public health measures such as testing, contact tracing, and quarantine, as well as the development and distribution of vaccines. This project aims to analyze global COVID-19 data and provide visual representations to better understand the trends, affected countries, and key statistics related to cases, deaths, recoveries, and active cases. By utilizing various visualizations such as treemaps, bar plots, pie charts, and interactive plots, we can gain valuable insights into the global COVID-19 situation.

1. Analysis of Maximum Total Cases, Deaths, Recovered & Active Cases:

* Identified the country with the highest total cases, deaths, recovered cases, and active cases.
* Created a treemap representation to visualize the distribution of these statistics across countries.

1. Trend of Confirmed Cases, Deaths, Recovered Cases, and Active Cases:

* Examined the trend of these COVID-19 statistics over time to identify patterns and fluctuations.
* Created visualizations, such as line charts or area charts, to represent the trends accurately.

1. Analysis of 20 Most Affected Countries:

* Identified the 20 countries with the highest number of COVID-19 cases.
* Examined the impact of the virus in terms of confirmed cases, deaths, recoveries, and active cases.

1. Bar Plot Representation of Population to Tests Done Ratio:

* Calculated the population to tests done ratio for different countries.
* Created a bar plot to compare the population to tests ratio across countries.

1. Analysis of 20 Worst Affected Countries:

* Identified the 20 countries most severely affected by COVID-19.
* Assessed the impact based on a combination of factors such as total cases, deaths, and active cases.

1. Bar Plot Representation of COVID-19 Cases over Time:

* Visualized the progression of COVID-19 cases across different time periods.
* Created bar plots to show the number of confirmed, active, recovered, and death cases over time.

1. Analysis of Top 20 Countries by Total Confirmed Cases, Recovered Cases, Deaths, and Active Cases:

* Identified the top 20 countries with the highest total confirmed cases, recovered cases, deaths, and active cases.
* Provided insights into the magnitude of the COVID-19 impact in these countries.

1. Bar Chart Representation of Top 20 Countries by Total Confirmed Cases:

* Created a bar chart to represent the total confirmed cases in the top 20 countries.
* Visualized the variations in confirmed cases across these countries.

1. Bar Chart Representation of Top 20 Countries by Total Deaths:

* Created a bar chart to represent the total deaths in the top 20 countries.
* Visualized the variations in deaths across these countries.

1. Bar Chart Representation of Top 20 Countries by Total Active Cases:

* Created a bar chart to represent the total active cases in the top 20 countries.
* Visualized the variations in active cases across these countries.

1. Pie Chart Representation of Statistics for Worst Affected Countries:

* Created pie charts, possibly in a donut shape, to represent the statistics (confirmed cases, deaths, recovered cases, and active cases) for the worst affected countries.

1. Bar Chart Representation of Deaths to Confirmed Cases Ratio:

* Calculated the ratio of deaths to confirmed cases for different countries.
* Created a bar chart to compare the ratio across countries.

1. Bar Chart Representation of Deaths to Recovered Cases Ratio:

* Calculated the ratio of deaths to recovered cases for different countries.
* Created a bar chart to compare the ratio across countries.

1. Bar Chart Representation of Tests to Confirmed Cases Ratio:

* Calculated the ratio of tests conducted to confirmed cases for different countries.
* Created a bar chart to compare the ratio across countries.

1. Bar Chart Representation of Serious to Deaths Ratio:

* Calculated the ratio of serious cases to deaths for different countries.
* Created a bar chart to compare the ratio across countries.

1. Visualization of COVID-19 Cases for a Specific Country:

* Utilized Plotly to create an interactive plot to visualize the confirmed, active, recovered, and death cases for a particular country.

Conclusion: This project provides a comprehensive analysis of global COVID-19 data and presents visualizations to better understand the trends, affected countries, and key statistics related to cases, deaths, recoveries, and active cases. The insights gained from this analysis can help policymakers, researchers, and the public understand the magnitude of the pandemic and make informed decisions to combat its effects. The visual representations serve as valuable tools to communicate complex data effectively and aid in creating awareness about the global COVID-19 situation.

## 4) UBER



Uber, the global ride-hailing and transportation network company, has transformed the way people travel within cities. Through its innovative mobile application, Uber disrupted the traditional taxi industry by introducing a peer-to-peer ridesharing model. By leveraging technology, Uber provides a seamless and user-friendly experience for both passengers and drivers. The company has expanded rapidly across countries and cities, overcoming regulatory challenges. Uber offers a diverse range of services, catering to different market segments, and contributing to its overall business strategy. It has had a significant impact on urban mobility, congestion reduction, and transportation accessibility. Uber driver-partners play a role in the gig economy, benefiting from the flexibility and income potential offered by driving for Uber. Technological advancements, such as GPS, real-time mapping, and machine learning algorithms, have enabled Uber to optimize routes and match drivers efficiently.

This project focuses on analyzing Uber pickups in New York City to gain insights into patterns, trends, and hotspots of activity. By leveraging data analysis techniques and visualizations, we aim to uncover valuable information that can be used for business optimization and strategic decision-making.

1. Data Acquisition:

* Obtained the Uber pickups data for New York City, including information such as pickup timestamps, base numbers, and location coordinates.
* Conducted exploratory data analysis to understand the structure and content of the dataset.

1. Data Pre-processing and Cleaning:

* Checked the data types of variables to ensure correct representation.
* Identified missing values and implemented appropriate strategies for handling them.
* Checked for and addressed any duplicated entries in the dataset.

1. Analysis of Monthly Uber Pickups in New York City:

* Determined which month had the highest number of Uber pickups in the city.
* Created a bar chart to visualize the monthly pickups, allowing for easy comparison.

1. Hourly Rush Analysis in New York City:

* Investigated the hourly rush patterns in New York City across all days.
* Utilized a point plot to visualize the number of pickups during each hour, providing insights into peak and off-peak periods.

1. Analysis of Active Vehicles by Base Number:

* Identified the base number that had the highest number of active vehicles.
* Presented the findings using visualization techniques such as box plots, or violin plots.

1. Data Analysis Based on Location:

* Collected and prepared the entire dataset for location-based analysis.
* Considered factors such as latitude, longitude, and pickup density for further examination.

1. Identification of Rush Locations in New York City:

* Utilized the latitude and longitude information from the dataset.
* Employed the Folium library to create a heat map, highlighting areas of high pickup activity in New York City.

1. Analysis of Rush by Hour and Weekday:

* Conducted pair-wise analysis to understand the rush patterns based on the hour of the day and the weekday.
* Created pivot tables and applied background gradients to visualize the variations in rush intensity.

1. Automation of Analysis:

* Explored methods to automate the data analysis process for future updates.
* Considered functions that could automate it rather than writing code every time.

This project offers a comprehensive analysis of Uber pickups in New York City, providing insights into monthly trends, hourly rush patterns, active vehicle distribution, rush locations, and the impact of hour and weekday on pickups. The findings can be utilized by Uber or other stakeholders to optimize operations, allocate resources efficiently, and improve the overall user experience.

## 5) Hotel Booking



Resort hotels are typically located in popular vacation destinations such as beachfront areas, mountain retreats, or tropical islands. They offer a wide range of amenities and recreational facilities to enhance your vacation experience. City hotels are located in urban areas and cater to business and leisure travelers who want convenient access to city attractions, shopping districts, and cultural sites. This project focuses on analyzing hotel bookings and predicting cancellations using the dataset obtained from Kaggle. The dataset contains information about bookings, including customer demographics, booking details, and hotel type (resort or city hotel). By performing exploratory data analysis and predictive modeling, we aim to gain insights into booking patterns, customer origins, and factors influencing cancellations. The findings will help hotel management and marketing teams optimize their operations and improve customer satisfaction.

1. Data Source:

* The dataset used in this project was obtained from Kaggle, titled "EDA of Bookings and ML to Predict Cancellations" by Marcus Wingen.
* It includes information such as booking dates, customer demographics, booking channels, and whether the booking was canceled or not.

1. Data Cleaning:

* Conducted data cleaning procedures to ensure the dataset is accurate and ready for analysis.
* Checked for missing values and implemented appropriate strategies such as imputation or removal.
* Handled outliers, inconsistencies, and data format issues as necessary.

1. Separating Resort and City Hotels:

* Split the dataset into two subsets based on the hotel type: resort hotel and city hotel.
* This separation allows for a focused analysis of each hotel category and comparison between them.

1. Analysis of Guest Origins:

* Explored the geographic origins of hotel guests.
* Analyzed customer nationality data to determine the top countries from which guests originate.
* Created visualizations, such as bar charts or pie charts, to showcase the distribution of guests from different countries.

1. Spatial Analysis of Home Countries:

* Performed spatial analysis using the Folium library to visualize the home countries of hotel guests on a choropleth map.
* Mapped the distribution of guests by country and incorporated color-coding to represent the number of guests from each location.

Conclusion: This project explores hotel bookings and cancellations, focusing on resort and city hotels. By analyzing guest origins and performing spatial analysis, we gain insights into the geographic distribution of hotel guests. The project demonstrates the potential for data-driven decision-making in the hospitality industry and highlights the importance of understanding guest origins and preferences. The findings from this analysis can be used to enhance marketing strategies, tailor services to specific customer segments, and ultimately improve the overall guest experience.