Datasets Assignment - 3

Chaitanya Sai Nutakki

AP21110010253

CSE-D

1. Airline Database

The dataset contains information on 5888 airlines as of January 2012. Some of the information is public data and some is contributed by users.

This dataset can be used for a variety of purposes, such as:

- Analyzing the airline industry
- Tracking the performance of individual airlines
- Identifying trends in the airline industry
- Developing new business applications for the airline industry

Here are some specific examples of how the Airline Database could be used:

- An analyst could use the dataset to track the performance of different airlines over time. They could look at factors such as revenue, passenger traffic, and profitability.
- A business consultant could use the dataset to identify potential markets for new airlines. They could look at factors such as population growth, economic development, and airport infrastructure.
- A software developer could use the dataset to develop a new app that helps travelers find the best deals on flights.

Types of analysis that can be done using this particular airline dataset

- ***Descriptive analysis:*** This type of analysis describes the data and identifies trends and patterns. For example,we can find out the following:
    - The average number of passengers per flight
    - The most popular routes
    - The busiest airports
    - The distribution of airfares
    - The number of flights per day

In order to perform this analysis, first we need to clean and prepare the data. Once the data is clean and prepared, we can use statistical tools to describe the data and identify trends and patterns.

- ***Comparative analysis:*** This type of analysis compares different airlines or different time periods. For example, you could use comparative analysis to compare the performance of two airlines over time or to compare the cost of airfare between different cities.

In order to perform this we first need to define the variables that we want to compare. Once we have defined the variables, we can then use statistical tools to compare the values of the variables between the two airlines or time periods.

- ***Predictive analysis:*** This type of analysis uses data to predict future events. For example, you could use predictive analysis to predict the likelihood of an airline going bankrupt or the demand for air travel in a particular market.

In order to perform this, we first need to collect data that is relevant to the event that we would like to predict. Once the data is collected, we can use ML algorithms to train a model that can predict the likelihood of the event occurring.

- ***Prescriptive analysis:*** This type of analysis uses data to recommend actions. For example, you could use prescriptive analysis to recommend which airlines to partner with or which routes to add.

2. Accident Database

The US Accidents dataset on Kaggle is a dataset of information about traffic accidents in the United States. It was created by Sobhan Moosavi and contains information on over 7.7 million accidents that occurred between February 2016 and March 2023.

Link of Database: https://www.kaggle.com/datasets/sobhanmoosavi/us-accidents

The dataset includes the following information for each accident:

- Date: The date of the accident
- Time: The time of the accident
- Location: The latitude and longitude of the accident
- Severity: The severity of the accident, as classified by the police
- Injuries: The number of people injured in the accident
- Fatalities: The number of people killed in the accident
- Weather: The weather conditions at the time of the accident
- Road conditions: The road conditions at the time of the accident
- Light conditions: The light conditions at the time of the accident
- Other: Other information about the accident, such as the number of vehicles involved and the type of vehicles involved

The US Accidents dataset can be used for a variety of purposes, such as:

- Analyzing traffic accidents and identifying trends and patterns.
- Predicting traffic accidents
- Improving traffic safety by identifying and addressing the factors that contribute to accidents.

Types of analysis that can be done using this particular accident dataset

Descriptive analysis:

Find out how many accidents occur each year: You could count the number of accidents in each year and plot the results to see if there is a trend.

Find out when most accidents occur: You could group the accidents by the time of day and plot the results to see if there is a peak time for accidents.

Find out what types of accidents are most common: You could group the accidents by the type of accident and plot the results to see which types of accidents are most common.

Comparative analysis:

Compare the number of accidents that occur in urban areas versus rural areas:

Compare the number of accidents that occur in different seasons: We could group the accidents by the season (winter, spring, summer, or fall) and count the number of accidents in each season.

Predictive analysis:

Predict the number of accidents that are likely to occur in a particular area: By ML algorithms to train a model that predicts the number of accidents that are likely to occur in a particular area.

We can also predict the severity of an accident

Prescriptive analysis:

Recommend ways to improve traffic safety

Allocate resources more effectively

3. Covid Database

The ECDC collects this data from national health authorities in the EU/EEA. The data is used to track the spread of COVID-19 in the EU/EEA and to inform public health interventions.

Link of Database: ***https://www.ecdc.europa.eu/en/covid-19/data***

Some essential ways that the data can be used:

- To track the number of cases and deaths in the EU/EEA
- To identify countries or regions that are experiencing outbreaks
- To assess the effectiveness of public health interventions
- To inform decisions about travel restrictions and quarantine measures
- To develop models to predict the future spread of the virus

Types of analysis that can be done using this particular accident dataset

- ***Descriptive analysis:*** This type of analysis describes the data and identifies trends and patterns. For example, you could use descriptive analysis to find out how many cases and deaths have occurred in the EU/EEA, which countries are experiencing outbreaks, and how the number of cases and deaths has changed over time.
- ***Comparative analysis:*** This type of analysis compares different countries or regions. For example, you could use comparative analysis to compare the number of cases and deaths in different countries or to compare the effectiveness of different public health interventions.
- ***Predictive analysis:*** This type of analysis uses data to predict future events. For example, you could use predictive analysis to predict the number of cases and deaths in the future or to predict the effectiveness of new public health interventions.
- ***Prescriptive analysis:*** This type of analysis uses data to recommend actions. For example, you could use prescriptive analysis to recommend policies or interventions to reduce the spread of COVID-19.