**Name:** Chandana Ramesh Galgali

**Batch:** FDS-2

**Roll No:** 16010422234

**FDS Expt 1**

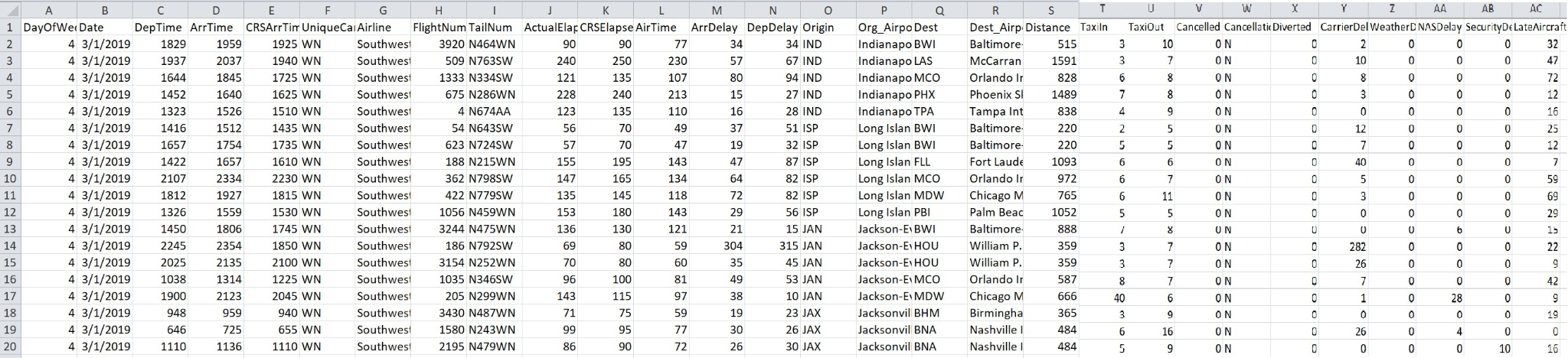
1. **Download a large dataset for the purpose of exploration and ensure that the dataset has a variety of attributes; number of attributes must be at least 25.**

Source of data: Kaggle

Dataset: Flight Delay and Causes

No. of attributes: 28

Screenshots:



1. **Identify the category of each attribute from the dataset which you have created.**
   1. DayOfWeek: Ordinal (1-7 represents the days of the week)
   2. Date: Interval (represents a specific date)
   3. DepTime: Ratio (actual departure time in hours and minutes)
   4. ArrTime: Ratio (actual arrival time in hours and minutes)
   5. CRSArrTime: Ratio (scheduled arrival time in hours and minutes)
   6. UniqueCarrier: Nominal (unique carrier code)
   7. Airline: Nominal (airline company)
   8. FlightNum: Nominal (flight number)
   9. TailNum: Nominal (plane tail number)
   10. ActualElapsedTime: Ratio (actual time spent in the air in minutes)
   11. CRSElapsedTime: Ratio (estimated elapsed time of flight in minutes)
   12. AirTime: Ratio (flight time in minutes)
   13. ArrDelay: Ratio (difference in minutes between scheduled and actual arrival time)
   14. Origin: Nominal (origin IATA airport code)
   15. Org\_Airport: Nominal (origin airport name)
   16. Dest: Nominal (destination IATA code)
   17. Dest\_Airport: Nominal (destination airport name)
   18. Distance: Ratio (distance between airports in miles)
   19. TaxiIn: Ratio (time taken to taxi from wheels down to arrival at the gate in minutes)
   20. TaxiOut: Ratio (time taken to taxi from departure gate to wheels off in minutes)
   21. Cancelled: Binary (1 = flight canceled, 0 = not canceled)
   22. CancellationCode: Nominal (reason for cancellation)
   23. Diverted: Binary (1 = flight diverted, 0 = not diverted)
   24. CarrierDelay: Ratio (flight delay due to carrier in minutes)
   25. WeatherDelay: Ratio (flight delay due to weather in minutes)
   26. NASDelay: Ratio (flight delay due to National Aviation System in minutes)
   27. SecurityDelay: Ratio (flight delay due to security reasons in minutes)
   28. LateAircraftDelay: Ratio (flight delay due to late aircraft arrival in minutes)
2. **Identify the attributes which can provide any kind of useful information either collectively or as an individual. Also, discuss the information provided by the attribute and how it will be computed?**

Here are some attributes that can provide useful information individually or collectively:

1. DayOfWeek: This attribute provides information about the day of the week when the flight occurred. It can be used to analyze patterns in flight schedules, such as peak travel days or days with higher delays. The computation is straightforward as it represents the day of the week on a numerical scale (1-7).
2. Date: The date attribute provides information about the specific date of the flight. It can be used to analyze trends over time, such as seasonal variations in flight delays or cancellations. The computation involves representing the date in a suitable format for analysis.
3. UniqueCarrier: This attribute represents the unique carrier code for the airline. It provides information about the airline operating the flight. It can be used to analyze the performance of different airlines, compare their on-time records, or identify any specific trends related to particular carriers. The computation involves assigning a unique code to each airline.
4. ActualElapsedTime: This attribute represents the actual time spent in the air by the airplane, including taxi-in and taxi-out times. It provides information about the total duration of the flight. It can be used to analyze flight efficiency, compare actual versus scheduled flight times, or identify any significant deviations. The computation involves calculating the difference between the actual departure and arrival times, including taxi times.
5. Distance: The distance attribute represents the distance between the origin and destination airports in miles. It provides information about the length of the flight route. It can be used to analyze the relationship between flight distance and other factors like flight time, delays, or fuel consumption. The computation involves calculating the distance between the coordinates of the origin and destination airports.
6. Cancelled: This binary attribute indicates whether the flight was canceled or not. It provides information about the flight's operational status. It can be used to analyze the frequency of flight cancellations, identify any patterns or trends, or assess the impact of cancellations on other attributes like delays or diversion. The computation involves assigning a binary value (1 or 0) based on the flight's cancellation status.
7. CarrierDelay, WeatherDelay, NASDelay, SecurityDelay, LateAircraftDelay: These attributes represent different types of delays and provide information about the reasons for flight delays. Each attribute indicates the delay time caused by a specific factor (carrier, weather, National Aviation System, security, or late aircraft arrival). They can be used to analyze the main causes of delays, identify patterns or trends in delay types, or assess the impact of delays on overall flight operations. The computation involves recording the delay time for each specific factor.

The computation of these attributes involves collecting and recording relevant data during the flight operations, such as departure and arrival times, flight distances, delay reasons, and cancellation status. These attributes collectively provide valuable information for analyzing flight operations, identifying trends, and making informed decisions related to scheduling, performance improvement, and customer satisfaction.

**Course outcome:**

Understanding of the Data.

**Conclusion:**

The experiment provided a solid foundation in working with categorical and numerical attributes. We gained practical skills and knowledge that can be applied in various data analysis tasks, such as exploratory data analysis. Understanding the characteristics and nuances of different attribute types is crucial for extracting meaningful insights and making informed decisions in data science and analytics.