AIRLINE DATA ANALYTICS PROJECT REPORT

Reducing flight cancellations and delays is essential for both passengers and the airline industry, as these disruptions have far-reaching consequences. They can result in diminished customer satisfaction and financial setbacks for airlines. safeguarding the industry's image, and rebuilding trust among travelers.

**Aim:**

The objective is to examine data spanning from January 2020 to March 2020 and derive significant insights from it. In this undertaking, we scrutinized more than 2 million flights with the aim of addressing the following inquiries:

* Which airports should be focused more on infrastructure development to mitigate the frequency and severity of delays?
* What are the relationships between airline carriers and flight delays/cancellations?
* What are the most common reasons for flight delays based on airports?
* What are the relationships between days in a week or times in a day when there are a lot of cancellations and delays?

**Dataset:**

Original Dataset: The original dataset has over 2m rows and around 109 columns in it.

Link: <https://developer.ibm.com/exchanges/data/all/airline/>

Working dataset: I aim to analyze the data from Jan 2020 to March 2020 and therefore the original dataset is trimmed down in a size. Along with this, Unnecessary columns are also removed. This dataset consists of 9923 rows and 32 columns.

The dataset contains the following columns.

* Month: Month of the flight date.
* DayOfWeek: Day of a Week at which flight operated.
* DayofMonth: Day of a month at which flight operated.
* FlightDate : Flight date.
* Reporting Airline: Operating Airline.
* Origin: Origin airport name.
* OriginCityName: City of origin airport.
* Dest: Destination Airport.
* DestCityName : City of destination airport.
* SchedukedDepTime: Scheduled departure time of a flight.
* ActualDepTime: Actual departure time of a flight.
* Difference\_in\_Dept\_Time: Difference between scheduled and actual departure time.
* DepDelayMinutes: Departure delay in minutes.
* DepTimeBlk: Time slots in which flight is departure.
* ScheduleArrTime : Scheduled arrival time.
* ActualArrTime: Actual arrival time.
* ArrDelay: Difference between scheduled and actual arrival time.
* ArrDelayMinutes: Arrival delay in minutes.
* ArrTimeBlk : Time slots at which flight is arrived at the destination airport.
* Cancelled: Is the flight cancelled or not?
* CancellationCode: Reason for the cancellation.
* Diverted: Is the flight diverted?
* Distance: Total distance travelled by the flight.
* CarrierDelay: Indicates Carrier delay.
* WeatherDelay: Indicates delay due to weather.
* NASDelay: Indicates NAS delay.
* SecurityDelay: Indicates security delay.
* LateAircraftDelay: Indicates late aircraft delay.

**Tools Used:**

* Microsoft Excel
* Python ( Pandas, NumPy, matplotlib)
* Jupyter Notebook
* Power Bi

**Key Takeaways:**

* ATL, DFW, and ORD airports experience higher and more frequent flight delays compared to others. These airports are situated in Atlanta, Dallas/Fort Worth, and Chicago, respectively. Therefore, it is imperative to implement infrastructural modifications at these airports to mitigate delays.
* Southwest Airlines, Delta Airlines, and American Airlines encountered the highest number of both flight delays and cancellations.
* Flight delays at various airports are frequently attributed to weather conditions and security issues.
* The 4th and 5th days of the week experienced a higher number of cancellations compared to any other weekdays, while the 6th day had the fewest cancellations.
* There is a higher likelihood of a flight being canceled when its scheduled departure falls between 5:00 pm and 8:00 pm in the evening.
* Flight arrivals are more prone to delays when scheduled between 7:00 pm and 8:00 pm, and less likely to experience delays when scheduled between 6:00 am and 7:00 am in the morning.