Project 1: Flight Delays

Objectives

- Perform exploratory data analysis and visualize the data to identify the causes of flight delays
- Help airport security to understand the effect of weather conditions and other factors on flight delays

Prerequisites

- Exploratory data analysis
- Data manipulation
- Data visualization
- R programming

Industry Relevance

- **Exploratory data analysis:** It is used to find trends and patterns or to check assumptions by analyzing data with visual tools.
- **Data manipulation:** It organizes and changes information to make it more understandable.
- **Data visualization:** It represents data with the use of common graphs, plots, or charts.
- **R programming:** It is used for statistical analysis, graphic representation, and reporting.

Problem Statement

Airport arrival performance and other events are affected by weather, which may result in delays or capacity constraints. Almost half of all airport traffic delays are caused by adverse weather conditions. These uncertainties during airport operations can result in significant delays and inconvenience to passengers. Therefore, the airport authority wants to analyze the flights that are delayed and the effect of weather on the delays.

You have been assigned to visualize the data with the help of histograms, scatter plots, box plots, and pie charts and understand the effect of weather conditions and other factors on flight delays.

Dataset Description

Variable	-	Description
schedtime	-	Scheduled time
Carrier	-	Airline codes
deptime	-	Time of departure
dest	-	Destination of flight
distance	-	Travelling distance
date	-	Date of travel
flightnum	-	Flight number

Dataset Description

Variable	-	Description
origin	-	Airport codes
weather	-	Coded as: 0: ontime 1: delayed
dayweek	-	Coded as: 1: Sunday and Monday, 0: other days
daymonth	-	Number of days in month
tailnu	-	Tail number of flight
delay	-	Delay status

Tasks to Perform

Perform the following tasks on the dataset provided using R

1. Exploratory data analysis:
    - Read the dataset
    - Read the dataset description
    - Understand the data
    - Find out the null values
2. Install the required packages
3. Understand the summary of descriptive statistics
4. Plot the histograms to understand the relationships between scheduled time, carrier, destination, origin, weather, and day of the week
5. Plot the scatter plot for flights on time and delayed

Tasks to Perform

Perform the following tasks on the dataset provided using R:

6. Plot the box plot to understand how many days in a month flights are delayed by what time
7. Define the hours of departure
8. Create a categorical representation of data using a table
9. Redefine the delay variables
10. Understand the summary of major variables
11. Plot histograms of major variables
12. Plot a pie chart to see how many flights were delayed

Project Outcome

- This project is designed to understand how to perform exploratory data analysis and plotting graphs.

- You should be able to analyze the dataset for this project to create a report.

Submission Process

1. Complete the project in the Simplilearn lab
2. Complete each task listed in the problem statement
3. Take screenshots of the results for each question and the corresponding code
4. Save it as a document and submit using the assessment tab
5. Tap the "Submit" button (this will present you with three choices)
6. Attach three files and then click "Submit“

Note: Be sure to include screenshots of the output

Thank You