Big Data project proposal

Team: 6

Name	ID	SEC	BN
Asmaa Adel Abdelhamed kawashty	9202285	1	13
Samaa Hazem Mohamed Abdel-latif	9202660	1	31
Norhan Reda Abdelwahed Ahmed	9203639	2	31
Hoda Gamal Hamouda Ismail	9203673	2	33

Supervisor: Eng. Omar Samir

- **Problem Statement**

Airline Passenger Satisfaction

Airline companies strive to provide the best possible travel experience for their passengers to maintain customer satisfaction and loyalty. Understanding the factors that contribute to passenger satisfaction is crucial for airlines to improve their services, enhance customer experience, and stay competitive in the industry. By accurately predicting passenger satisfaction, airlines can identify areas for improvement and tailor their services to meet customer expectations more effectively. Additionally, satisfied passengers are more likely to become repeat customers and recommend the airline to others, leading to increased revenue and growth.

The problem can be framed as a binary classification task, where the goal is to predict whether a passenger is satisfied or dissatisfied based on the input features. The dataset used for model training and evaluation contains historical data on passenger satisfaction along with corresponding feature values.

- **Dataset(s):**
  https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction

- **Planned approach or Proposed solution:**
  - **Language:** Python.
  - **Framework:** Spark.
  - **EDA:** we may use statistics like (min, max, mean, count, percentages, and different visualizations like histograms and correlation heatmaps.
  - We may use **association rules**, to gain more insights about data and identify correlations between features.
  - **For classification** (predicting satisfaction): we will use machine learning models such as Logistic Regression, Naive Bayes, SVM, Decision Trees, KNN and Random forest.
  - **Map Reduce** in Machine Learning models.