Data Science Final Project

- **Project Instructions:**
  - ➢ **Each project group should consist of 2-3 students who share the same lab slot time.**

  - ➢ **Project discussions will commence on May 12-13, with the deadline set for May 19-20, exclusively during the designated lab sessions.**

  - ➢ **It is strictly prohibited for any group to engage in discussions outside of the allocated time slots.**

- **Overview:**

  In this project, you will work on a real-world dataset to perform data preprocessing, create insightful visualizations using Power BI, and apply various machine learning algorithms for both classification and prediction tasks. The project aims to provide you with hands-on experience in data preparation, data visualization, model building, and evaluation.

- **Data Set Resources:**

  https://github.com/datasets/five-thirty-eight-datasets/tree/master/datasets

- **Project Components:**
  1. **Data Preprocessing:**
     - ➢ **Objective: Clean and prepare the dataset for analysis.**
     - ➢ **Tasks:**
       - o **Handle missing values.**
       - o **Remove duplicates.**
       - o **Encode categorical variables.**
       - o **Normalize numerical features.**

Data Science Final Project

2. Power BI Visualization

➤ **Objective: Create interactive and insightful visualizations to explore the dataset.**

➤ **Tasks:**

- o **Import the cleaned dataset into Power BI.**
- o **Create various charts and graphs to visualize the data distribution, relationships, and patterns.**
- o **Design a dashboard to present the key insights derived from the visualizations.**

3. Machine Learning Algorithms

➤ **Objective: Build and evaluate machine learning models for classification and prediction tasks.**

➤ **Algorithms:**

Classification	Prediction
Naive Bayes	Regression (Linear Regression, Polynomial Regression, etc.)
Decision Tree Classifier	Decision Tree Predictor
K-Nearest Neighbors (KNN)	K-Nearest Neighbors (KNN) Regressor

➤ **Tasks:**

- o **Split the dataset into training and testing sets.**
- o **Train the machine learning models using the training set.**
- o **Evaluate the models using appropriate metrics (accuracy, precision, recall, F1-score for classification; RMSE, MAE for regression).**
- o **Compare the performance of different algorithms and select the best-performing model for each task.**