Week 1 Tasks
Data Science Intern

EXPLORATORY DATA ANALYSIS ON A PUBLIC DATASET

Description: Analyze a real-world dataset (such as Titanic, Iris, or any Kaggle dataset). Generate insights through visualizations and summary statistics.

Challenges:

- Handle missing and duplicate data
- Visualize key relationships using Matplotlib/Seaborn
- Write a short report summarizing findings

Tech Stack: Python, Pandas, Matplotlib, Seaborn, Jupyter Notebook

BUILD A SIMPLE MOVIE RECOMMENDATION SYSTEM

Description: Create a basic recommendation system using a dataset like IMDb or TMDb.

Challenges:

- Implement content-based or popularity-based filtering
- Handle text similarity using TF-IDF or cosine similarity
- Display top 5 movie recommendations based on user input

Tech Stack: Python, Pandas, Scikit-learn, Jupyter Notebook

PREDICT HOUSE PRICES USING MACHINE LEARNING

Description: Train a regression model to predict house prices using the classic Boston Housing dataset or a similar dataset from Kaggle.

Challenges:

- Feature scaling and encoding categorical data
- Train/test split and evaluate model using RMSE or R^2
- Visualize model performance and key features

Tech Stack: Python, Pandas, NumPy, Scikit-learn, Matplotlib

*Complete any 1 Task