

DATA ENGINEERING ESSENTIALS – LAB ASSIGNMENT

Section: C (AIML)

Course: Data Engineering Essentials

Program: II Year B.Tech CSE (AI & ML)

Duration: 2 Hours

Total Marks: 10

Objective of the Assignment

To evaluate the student's ability to load, inspect, clean, transform real-world datasets, perform feature engineering, and extract meaningful insights through data analysis.

Datasets Provided (Section C)

Customer Churn Dataset

<https://www.kaggle.com/datasets/blastchar/telco-customer-churn>

Spotify Songs Dataset

<https://www.kaggle.com/datasets/lehaknarnauli/spotify-datasets>

Mobile App Reviews Dataset

<https://www.kaggle.com/datasets/lava18/google-play-store-apps>

Credit Card Transactions Dataset

<https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud>

Weather Dataset

<https://www.kaggle.com/datasets/selfishgene/historical-hourly-weather-data>

E-Commerce Orders Dataset

<https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>

Insurance Dataset

<https://www.kaggle.com/datasets/mirichoi0218/insurance>

Flight Delay Dataset

<https://www.kaggle.com/datasets/usdot/flight-delays>

Lab Tasks (Common for All Datasets)

Task 1: Load the dataset and perform initial inspection (head, shape, info, describe).

Task 2: Identify and handle missing values using appropriate strategies with justification.

Task 3: Transform categorical data using encoding techniques.

Task 4: Perform feature engineering by creating at least one new feature.

Task 5: Perform data analysis and write two meaningful insights.

