

**Department of Computer Science**

**CSC 487: Introduction to Data Science Semester 6, Section A&B (Fall 2023)**

**Assignment No. 01**

**Submission Due Date: 31 October 2023**

**Marks: 05**

**Instructions:**

* The assignment contains ***01 question*** to be answered.
* Before submission, attach provided ***cover sheet*** on top.
* ***Plagiarism is strictly discouraged***. Write in your own words, as similarity will reduce marks.
* Rename file as: **NAME-ENROLLMENTNUMBER.**

***Question 1: [CLO1, PLO2, C2]* [5 Marks]**

**Apply** Data Preprocessing techniques & **Perform** Exploratory Data Analysis on the given dataset.

* Load the dataset in Jupiter notebook using appropriate python library.
* Perform data preprocessing steps (cleaning, handling missing values, outlier detection etc.).
* Generate descriptive and inferential statistics.
* Visualize the data to gain insights into customer churn factors.
* Identify key features that may influence churn.

**Dataset Link:**

[**https://www.kaggle.com/datasets/blastchar/telco-customer-churn/data**](https://www.kaggle.com/datasets/blastchar/telco-customer-churn/data)

Each row represents a customer, each column contains customer’s attributes described on the column Metadata.

**The data set includes information about:**

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents