**CS5610 Advanced R for Data Science**

**Exploratory Data Analysis on the Bike Buyers Dataset**

1. **Motivation and Overview**: This project shows the maintenance of the bikes and their cleaning efficiency and also the identification of different ways to reach to the destination.

* There is a multitude of related factors that come along with the bike purchase.
* A willing buyer has several factors by which he or she can be determined as a true buyer.
* These factors are essential for not only the proposition of the work but also when determining the maintenance of the vehicle and even playing an important role.

1. **Related Work**: Currently, there are over 500 bike-sharing programs around the world. Such systems usually **aim to reduce congestion, noise, and air pollution** by providing free/affordable access to bicycles for short-distance trips. Machine Learning also makes research on the exploratory data analysis of bike buyers to understand the correlated factors and along with that the prediction of bike buyers on the basis of this correlated attributes.
2. **Initial Questions**: How far can the exploratory data analysis can help us to decide whether a customer owns a bicycle or not?

* By plotting according to the annual income of various features can we able to find out the interest of the buyer?

1. **Data**: The collected data is known as Bike Buyers Dataset which is obtained from Kaggle Repository. In this, we have used the dplyr method and performed the data wrangling. The dataset contains 13 attributes and 1001 instances. Among all there is no null values presented in it.
2. **Exploratory Data Analysis**: We have performed EDA for this project using Bar Plot by the help of slider panel.For this suggested EDA job, a significant amount of effort is required, in which we will first examine the information and then identify distinct elements and process them for analysis depending on our observations.

Then, in order to have a clear knowledge of the profound knowledge of the overall goal of the project, a quantitative assessment will be carried out. The suggested work might take between three and four-month to finish.

Based upon the given features EDA is done

1)Marital Status

2)Gender

3)Occupation

4)Region

5)Education

1. **Data Analysis**: We have used dplyr method for data manipulation to derive the features from the large data set using the functions like filter().

Data analysis played an important role while dealing and retrieving the data. Mainly we are analysing the information based on the user choice data and make a data filter based on that.

1. **Narrative and Summary**: The design brings over the performance of a detained time of exploratory data analysis.

* R programming is being used and the programming environment has brought about several insights correctly.
* Many correlated factors are also to be understood along with the objectives that are determined for the proposed work.
* The limitations we have faced in this project was, since there are many columns and rows in the data set it was a bit hard to analyse the data though it did not have any null values.

GitHub Link: https://github.com/wmusurya/Exploratory-Data-Analysis-on-the-Bike-Buyers-Dataset