**CS5610 Advanced R for Data Science**

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

**What to turn in?**

The final project submission will consist of

1. RMarkdown File: include the following topics. Depending on your project type, the amount of discussion you devote to each of them will vary:
2. **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

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.
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.
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.

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.

1. Your code (.R) along with a dataset, if any. [I expect you to write high-quality and readable R code in your RMarkdown file. I also expect you to document your code.
2. Project Website
3. A 10-15-minutes presentation (in PowerPoint or pdf format) of your project. The presentation will be a prerecorded video.

**Turning in your project:**

Submit your project zip file to Dropbox on eLearning by **8:00 pm** on April 18th.

HANDWRITTEN REPORTS AND LATE SUBMISSION WILL NOT BE ACCEPTED.

**What is next:**

1. All presentations will be available on eLearning to the class on Wednesday, April 20th.
2. Final Project Discussion will be available on eLearning on Wednesday, April 20th. Students should ask questions to the project teams. The projects’ teams must answer all the questions by 11:59 pm Thursday, April 21st.

**Grading**

The final project will be graded in three main parts:

1. **Project proposal and team presentation** (10%). This portion represents the Team Registration, Final Project Proposal, and presentation.
2. **Data Analysis** (50%). This portion will be based on your RMarkdown file in your GitHub repository. This includes the quality of your data analysis and R code, the complexity and level of difficulty of your project, completeness and overall functionality of your analysis.
   * Readability/overall look (10%)
   * Strength of conclusions/findings (10%)
   * Interpretation correctness (10%)
   * Code correctness (20%)
3. **Website** (20%) and **Final Project Presentation** (10%). This portion will be based on the quality of your storytelling and summarization aspects.
4. **Discussion Participation** (10%)
   * Final Project Discussion will be available on eLearning on Wednesday, April 20th. Students should ask questions to the project teams. The projects’ teams must answer all the questions by 11:59 pm, Thursday, April 21st.
5. ***Shiny app*** (**3** **extra points on your project**) if you build a Shiny app for it.
   * The Shiny app should include multiple pages and different functions (interactive: maps, plots, tables, sidebar, drop menus, etc.) to show a reasonable effort. (A BASIC SHINY APP WILL NOT BE CONSIDERED)