EBIS 4043 Big Data Analysis and Applications

Individual Assignment II

The purpose of this assignment is to make sure that **you are picking up the R based analytics skills (Please do not use other tools to generate the answers!)** that have been introduced in this class and check your ability.

As you write your answers, please make sure to follow the instructions below:

* + Use the datasets that were uploaded on iSpace.
  + All your answers including your identity, codes, interpretation should be in **one file: HTML**. Any sort of multiple files will be graded as zero mark.
  + You can discuss the coding with your friends. However, any visible overlap in your interpretation will be considered plagiarism.
  + There can be more than one correct answer to every question. Use any technique that you learned from the classroom.

Dataset on Amazon's Top 50 bestselling books from 2009 to 2019 contains 550 books. data has been categorized into fiction and non-fiction using Goodreads. The name of the dataset is *bestsellers with categories.csv.* No Null Values seen in the dataset. The data dictionary is seen as below.

**Data Dictionary**

* Name - name of the bestselling book (datatype - object)
* Author - author of the bestselling book (datatype - object)
* User Rating - User Rating of the book (datatype - float)
* Reviews - number of reviews for the book (datatype - integer)
* Price - price of the book (datatype - integer)
* Year - Year when it was a bestseller(datatype - integer)
* Genre - Categorised as Fiction and non fiction (datatype - object)

Now you are a data analyst who wants to create a dashboard using **R Flexdashboard** for the dataset. The **requirements** include:

* Create two graphs that show the most 10 popular authors and most 10 popular books. [30 Marks]
* Create a table that shows all the best Books from 2009 to 2019 by users rating and reviews (rating ratings>4.9 and number of reviews > 5,000). [20 Marks]
* Create a trendline that shows the numbers of reviews along with 2009 to 2019. [20 Marsk]
* Create any table or graph of your interest. [10 Marks]
* Ensure the tables and graphs are suitable for effectively presenting the results, displaying a high level of organization and aesthetic appeal. [20 Marks]

**More Tips:**

* Highcharts theme collection: https://jkunst.com/highcharts-themes-collection/
* More information about Flexdashboard: https://rstudio.github.io/flexdashboard/articles/using.html#overview-1