Project R and RStudio.

|  |  |
| --- | --- |
|  | Install both R and RStudio as described in class and posted on WebCampus. |
|  | Start RStudio. |
|  | Download [data for R project] from WebCampus. |
|  | Extract R Code.zip to a folder. |
|  | Assign a Working Directory as discussed in class. It should be the folder, where you extracted scripts and datasets. |
|  | Now you need to run a specific .RMD script from the folder you downloaded from WebCampus: |
|  | Find out which script to use by using the mapping below. |
|  | Answer the following questions: |
|  | * Which data mining method did your script execute?   + Is this method supervised or unsupervised? |
|  | * What is the name of the dataset(s) the script used? |
|  | * Did you have to download any packages, and if so, which ones? |
|  | * Which libraries does your script use? |

Assignment 2 (for extra credit). Task: by presidential candidate, display bar charts whose tweets get retweeted and whose tweets are retweeted within a positive or negative comment (sentiment analysis).

|  |  |
| --- | --- |
| 1 | Register at kaggle.com |
| 2 | Download a dataset composed of twitter discussion of the Republican Party debates from this link <https://www.kaggle.com/crowdflower/first-gop-debate-twitter-sentiment/version/1>  (you will find a [Download Data 2.4MB] link there |
| 3 | Download R script from here:  <https://www.kaggle.com/bluelight773/d/crowdflower/first-gop-debate-twitter-sentiment/gop-debate-exploration/code> |
| 4 | Run the script from step 3 on the dataset from step 2. |
| 5 | Shortly describe what the script does. |

Assignment 3 (for extra credit). Task: build a word cloud of most popular words used in the Republican Party debates.

|  |  |
| --- | --- |
| 1 | Register at kaggle.com |
| 2 | Download a dataset as discussed in the assignment 2 above. |
| 3 | Download R script from here:  <https://www.kaggle.com/usmankhaliq/d/crowdflower/first-gop-debate-twitter-sentiment/the-donald-cloud/notebook>  (you will need to copy and paste code into R). |
| 4 | Run the script from step 3 on the dataset from step 2. |
| 5 | Shortly describe what the script does. |

Find your user ID below and run the corresponding script:

|  |  |
| --- | --- |
| **User name** | **Script** |
| st1 | 01 Linear Regression.Rmd |
| st2 | 02 Logit Regression.Rmd |
| st3 | 03 Association Rules.Rmd |
| st4 | 04 K-Means Clustering.Rmd |
| st5 | 05 Trees.Rmd |
| st6 | 06 KNN.Rmd |
| st7 | 07 Naive Bayes.Rmd |
| st8 | 01 Linear Regression.Rmd |
| st9 | 02 Logit Regression.Rmd |
| st10 | 03 Association Rules.Rmd |
| st11 | 04 K-Means Clustering.Rmd |
| st12 | 05 Trees.Rmd |
| st13 | 06 KNN.Rmd |
| st14 | 07 Naive Bayes.Rmd |
| st15 | 01 Linear Regression.Rmd |
| st16 | 02 Logit Regression.Rmd |
| st17 | 03 Association Rules.Rmd |
| st18 | 04 K-Means Clustering.Rmd |
| st19 | 05 Trees.Rmd |
| st20 | 06 KNN.Rmd |
| st21 | 07 Naive Bayes.Rmd |
| st22 | 01 Linear Regression.Rmd |
| st23 | 02 Logit Regression.Rmd |
| st24 | 03 Association Rules.Rmd |
| st25 | 04 K-Means Clustering.Rmd |
| st26 | 05 Trees.Rmd |
| st27 | 06 KNN.Rmd |
| st28 | 07 Naive Bayes.Rmd |
| st29 | 01 Linear Regression.Rmd |
| st30 | 02 Logit Regression.Rmd |
| st31 | 03 Association Rules.Rmd |
| st32 | 04 K-Means Clustering.Rmd |
| st33 | 05 Trees.Rmd |
| st34 | 06 KNN.Rmd |
| st35 | 07 Naive Bayes.Rmd |
| st36 | 01 Linear Regression.Rmd |
| st37 | 02 Logit Regression.Rmd |
| st38 | 03 Association Rules.Rmd |
| st39 | 04 K-Means Clustering.Rmd |
| st40 | 05 Trees.Rmd |
| st41 | 06 KNN.Rmd |
| st42 | 07 Naive Bayes.Rmd |
| st43 | 01 Linear Regression.Rmd |
| st44 | 02 Logit Regression.Rmd |
| st45 | 03 Association Rules.Rmd |
| st46 | 04 K-Means Clustering.Rmd |
| st47 | 05 Trees.Rmd |
| st48 | 06 KNN.Rmd |
| st49 | 07 Naive Bayes.Rmd |
| st50 | 01 Linear Regression.Rmd |
| st51 | 02 Logit Regression.Rmd |
| st52 | 03 Association Rules.Rmd |
| st53 | 04 K-Means Clustering.Rmd |
| st54 | 05 Trees.Rmd |
| st55 | 06 KNN.Rmd |
| st56 | 07 Naive Bayes.Rmd |
| st57 | 01 Linear Regression.Rmd |
| st58 | 02 Logit Regression.Rmd |
| st59 | 03 Association Rules.Rmd |
| st60 | 04 K-Means Clustering.Rmd |
| st61 | 05 Trees.Rmd |
| st62 | 06 KNN.Rmd |
| st63 | 07 Naive Bayes.Rmd |
| st64 | 01 Linear Regression.Rmd |
| st65 | 02 Logit Regression.Rmd |