04-MAY-2022 Activity #7: Working with mixed datatypes

1. Read the csv file from the urlhttps://raw.githubusercontent.com/justmarkham/DAT7/master/data/yelp.csv'

2. Create a column describing the number of characters in text column

3. Get the mean values of numeric column

4. Are they correlated?

5. Report the result of correlation using heatmap

6. Create a class that contain the columns of original dataframe but for only the 1 or 5 star review

7. Create the features 'text' column as X and 'stars' as y

8. Demonstrate how to build a raw corpus from multiple files by creating two .docx files, with text contents of your choice.

9. Write the contents of two .docx file to a flat file say .csv.

10. Comment on the significance of using the corpus over flat file (txt, csv etc) with respect to text pre-processing.