**Batch 4\_Semester VII \_ Year IV\_ Data Science \_ Mid-Semester Exam**

**Duration: 2 Hours Max Marks: 70**

**Note: Kindly upload your R script(.R/.RMD), Dataset and Graphs on google classroom. Do not use the Kaggle, Jupyter notebook and Etc.,**

1. Write a R code to Generate 6X6 matrix (A) using the random numbers between 1 to 100 & generate another 6x6 matrix (B) using random numbers between 500 to 600.

(20 Marks)

1. Find the range of matrices A and B.
2. Find the max value from each row of the matrices A and B.
3. Find the diagonal values of matrices A&B.
4. Multiply matrices A and B
5. Replace the 4th row of Matrix A with diagonal values of matrix B
6. Multiply matrices A & B and store the values on X.
7. Divide the matrix A with matrix B.

2. Create a table using dataframe/tibble with the following categories(a to f) (10 marks)

1. First name,
2. Second Name,
3. Age,
4. Occupation,
5. Place,
6. Random numbers( between 1 to 10)

And input at least 10 samples(observation) and do the following

1. Extract the last column (f)
2. Find the sum, mean and length of the extracted column

3. Create a .txt file (Minimum : 150 words) with the latest news (In English) in cambodia from the official news agency’s website(Note: Include the reference link) and do the following

(20 Marks)

1. Read the .txt file
2. Perform the Tokenization & count the words
3. Identify the most repeated word and replace that word with “KIT”
4. Perform the Sentiment analysis
5. Plot the graph between “words count” and “words” using geom\_col()

e.1. X-axis title =“Number of words”, Size=10, Bold, Text in Blue color, Center, space between graph and title = 20.  
 e.2. Y-axis title =“List of words”, Size=10, Bold, Text in Red color, Center, space between graph and title = 20.  
 e.3. Graph title =”Text Mining using R”,Size = 25, Bold, Center, Text in Yellow color, space between graph and title = 25.

4. Find the assigned Dataset from the list below and perform the following (20 Marks)

1. Import and Clean the dataset.
2. Find the x and y variables(axis) from the assigned dataset and give the reason for selecting the x and y axis.
3. Plot the graph using geom\_col() and perform the linear model in statistics.
4. Perform the facet with the available columns(Choose the any two variables from the columns)
5. Use the theme layer to do the following (Note: Based on dataset, name the x and y axis & graph)

e.1. x-axis title =“\_\_\_\_\_\_\_\_”, Size=10, Bold, Text in Green color, Center, space between graph and title = 20.  
 e.2. y-axis title =“\_\_\_\_\_\_\_\_”, Size=10, Bold, Text in Yellow color, Center, space between graph and title = 20.  
 e.3. Graph title =”\_\_\_\_\_\_\_\_\_”,Size = 25, Bold, Center, Text in Blue color, space between graph and title = 25.

**Dataset Links**

1. Am Chunny : Airline safety <https://github.com/fivethirtyeight/data/blob/master/airline-safety/airline-safety.csv>
2. Chea Kimmunyvorn : Alcohol Consumption: <https://github.com/fivethirtyeight/data/blob/master/alcohol-consumption/drinks.csv>
3. Chea Nataly: [Antiquities-act](https://github.com/fivethirtyeight/data/tree/master/antiquities-act) : <https://github.com/fivethirtyeight/data/blob/master/antiquities-act/actions_under_antiquities_act.csv>
4. Chhean Pisethpanha : August Senate poll <https://github.com/fivethirtyeight/data/blob/master/august-senate-polls/august_senate_polls.csv>
5. Chhor Menghong: Avengers <https://github.com/fivethirtyeight/data/blob/master/avengers/avengers.csv>
6. Daro Dalya : Bad Drivers <https://github.com/fivethirtyeight/data/blob/master/bad-drivers/bad-drivers.csv>
7. Hann Manich : Bechdel <https://github.com/fivethirtyeight/data/blob/master/bechdel/movies.csv>
8. Heng Senghak : US\_Births\_2000\_2014 <https://github.com/fivethirtyeight/data/blob/master/births/US_births_2000-2014_SSA.csv>
9. Hoeurng Hen: Bob\_ross <https://github.com/fivethirtyeight/data/blob/master/bob-ross/elements-by-episode.csv>
10. Hy Chhayrith : Cabinet\_turnover <https://github.com/fivethirtyeight/data/blob/master/cabinet-turnover/cabinet-turnover.csv>
11. Kaing Sokheng: Candy\_power ranking <https://github.com/fivethirtyeight/data/blob/master/candy-power-ranking/candy-data.csv>
12. Khean Sreythou: Classic Rock: <https://github.com/fivethirtyeight/data/blob/master/classic-rock/classic-rock-song-list.csv>
13. Kim Miratorimoonlight : Recent -grads <https://github.com/fivethirtyeight/data/blob/master/college-majors/recent-grads.csv>
14. Kit Bunrong : Covid - geography <https://github.com/fivethirtyeight/data/blob/master/covid-geography/mmsa-icu-beds.csv>
15. Kruy Chandara: Drug use <https://github.com/fivethirtyeight/data/blob/master/drug-use-by-age/drug-use-by-age.csv>
16. Ley Kimteng : Early senate poll <https://github.com/fivethirtyeight/data/blob/master/early-senate-polls/early-senate-polls.csv>
17. Ly Chandara : Elo\_Blatter <https://github.com/fivethirtyeight/data/blob/master/elo-blatter/elo_blatter.csv>
18. Nhek Pichpanharith: Endorsement <https://github.com/fivethirtyeight/data/blob/master/endorsements-june-30/endorsements-june-30.csv>
19. Pech Sokmeng : Fandango <https://github.com/fivethirtyeight/data/blob/master/fandango/fandango_scrape.csv>
20. Pel Dane : Fifa <https://github.com/fivethirtyeight/data/blob/master/fifa/fifa_countries_audience.csv>
21. Pich Sovannara : Fight songs <https://github.com/fivethirtyeight/data/blob/master/fight-songs/fight-songs.csv>
22. Pong Channy : Food world cup <https://github.com/fivethirtyeight/data/blob/master/food-world-cup/food-world-cup-data.csv>
23. Sang Sonyratt : Forecast methodology <https://github.com/fivethirtyeight/data/blob/master/forecast-methodology/historical-senate-predictions.csv>
24. Sann Chamrouen : Foul balls <https://github.com/fivethirtyeight/data/blob/master/foul-balls/foul-balls.csv>
25. Seng Rathanak : Hate crimes <https://github.com/fivethirtyeight/data/blob/master/hate-crimes/hate_crimes.csv>
26. Set Samnang : Librarians <https://github.com/fivethirtyeight/data/blob/master/librarians/librarians-by-msa.csv>
27. Sim Chhay : Mad-men <https://github.com/fivethirtyeight/data/blob/master/mad-men/show-data.csv>
28. Song Sokhavudthi : March \_madness <https://github.com/fivethirtyeight/data/blob/master/march-madness-predictions-2015/womens/bracket-00.tsv>
29. Sreang Vuthy : Media-mention <https://github.com/fivethirtyeight/data/blob/master/media-mentions-2020/cable_weekly.csv>
30. Sun Chenny : Mlb-allstars <https://github.com/fivethirtyeight/data/blob/master/mlb-allstar-teams/allstar_player_talent.csv>
31. Tan Somnang : Most-common name: <https://github.com/fivethirtyeight/data/blob/master/most-common-name/adjusted-name-combinations-list.csv>
32. Tea Vengtieng : Mueller polls <https://github.com/fivethirtyeight/data/blob/master/mueller-polls/mueller-approval-polls.csv>
33. Teng Seavpor : Nba-draft <https://github.com/fivethirtyeight/data/blob/master/nba-draft-2015/historical_projections.csv>
34. Tha Pithsopheak: Nba-draymond <https://github.com/fivethirtyeight/data/blob/master/nba-draymond/draymond.csv>
35. Thepanom Ratanawan : Ncaa-womens: <https://github.com/fivethirtyeight/data/blob/master/ncaa-womens-basketball-tournament/ncaa-womens-basketball-tournament-history.csv>
36. Thon Lynan : Nfl-favorite <https://github.com/fivethirtyeight/data/blob/master/nfl-favorite-team/team-picking-categories.csv>
37. Ty Leapheng : Nutritions studies <https://github.com/fivethirtyeight/data/blob/master/nutrition-studies/raw_anonymized_data.csv>
38. Vansen Hengmeanrith : Police killing <https://github.com/fivethirtyeight/data/blob/master/police-killings/police_killings.csv>
39. Voeun Thavin : Pollster rating <https://github.com/fivethirtyeight/data/blob/master/pollster-ratings/pollster-ratings.csv>
40. Vorn Kiriratanak : Presidential candidate <https://github.com/fivethirtyeight/data/blob/master/presidential-candidate-favorables-2019/favorability_polls_rv_2019.csv>
41. Youk Sakmonysothea : Pulitzer <https://github.com/fivethirtyeight/data/blob/master/pulitzer/pulitzer-circulation-data.csv>