

MCST1063 : Statistics for Data Science

Project :

This project contribute to 20% of total courseworks.

This project will do stage by state.

Instruction :

- Form a group of 4 or 5 students.

Task 1 :

Find any dataset that consist of :

- Number of columns (> 6) and number of rows should more than 30
- Discrete and continuous data
- Non time series data
- Should have all the measurement scale
 - Nominal (at least 2 variables)
 - Ordinal (at least 2 variables)
 - Ratio/interval (at least 2 variables)

TASK 2 : Perform Descriptive analysis

1) Exploratory data analysis

- Check any missing values, if have-how do you handle it
- Data formatting
- Column rename-if any

2) Data Analysis

a. Categorical data:

- **Graphs:** Pie chart or bar chart (with percent not counts)
- **Numerical:** Proportion and sample size in each category

b. Quantitative data:

- **Graphs:** E.g., histograms or dot plots.
- **Measures of Central Tendency:** E.g., mean, media, boxplots

Interpret and discuss what you can learn from the pictures (e.g., compare shape...) and make sure you report the sample sizes for the survey as well.

Task 3; Inferential statistics

Inference statistical analysis that cover the following analysis based on selected dataset

- Compulsory:
 - Hypothesis testing 1-sample or 2-sample
 - Correlation

- Regression
- ANOVA
- Optional (choose 2):
 - Goodness of fit test
 - Chi Square test of independence
 - Non parametric test

Note : all these task need to be compiled in one report. Maximum pages (not include appendix) is 10 pages.

Submission consist of :

- Report : in pdf format
- Copy of code, R