

Homework 1 (Statistical Regression & R)

Instructions:

- Submit your answer to the task below as a clean report of no more than four pages.

Warning: Only the first four pages will be marked if you submit more than four pages.

- Submit a separate R file containing your syntax/code (the tutors will check your answers by running your R syntax file).

Task:

A retail company wants to understand the customer purchase behaviour (specifically, purchase amount). They have shared a sample of purchase data summary of various customers for selected high volume products from last month. The dataset contains the following columns: *User_ID*, *Gender*, *City_Category*, *Stay_In_Current_City_Years* (this is the length of stay in the current city), *Marital_Status* (coded as 1 for married, 0 otherwise), *Age_num* (this is the approximate age of the user), and *Purchase* (this is the purchase amount).

- 1) Undertake a descriptive analysis of the variables in this dataset.
- 2) Use this dataset to build a linear regression model to predict the purchase amount.

Justify all the choices you make to arrive at your final model.

- 3) Interpret your final model.

- 4) Are all the assumptions about your model met?
- 5) Explain how you can use your final model to help the company improve their sales?
- 6) Suggest possible ways of improving your model.
- 7) Compare your final model to the model built on Gender and Age only.