**ITECH2303 Mid Semester Assignment**

**Student ID/ Name:**

Requirements: Your chosen CSV file must have at least five columns/attributes, of which at least two must be numeric

|  |  |  |
| --- | --- | --- |
|  | Tasks | Marks |
|  | Read the Data |  |
| 1 | # TASK: load the CSV file from Google Drive into Google Colab Notebook (In a comment provide the URL source of your CSV file) | ……/1 |
| 2 | # TASK: load the data into Pandas dataframe | ……/1 |
| 3 | # TASK: show the first 5 rows | ……/1 |
| 4 | # TASK: show the last 5 rows | ……/1 |
|  | Add Headers |  |
| 5 | # TASK: Create header list and add the header list to dataframe | ……/1 |
| 6 | # TASK: Rename a column | ……/1 |
|  | Basic Insight of Dataset |  |
| 7 | # TASK: Check the datatypes of the dataframe | ……/1 |
| 8 | # TASK: Change the datatype of one of the columns | ……/1 |
| 9 | # TASK: Call the functions ‘describe’ and ‘info’ on your dataframe | ……/1 |
| 10 | # TASK: Choose selections of data using ‘loc’ and ‘iloc’ and display these selections | ……/1 |
| 11 | # TASK: Write a function [2 marks] that calculates the logarithm of the mean value of a column and ‘apply’ this to your columns [1 mark], returning both the mean value and its logarithm. Display your results. | ……/3 |
|  | Correlation and Covariance |  |
| 12 | # TASK: Determine the correlation between two of the numeric attributes | ……/1 |
| 13 | # TASK: Determine the covariance between two of the numeric attributes | ……/1 |
| 14 | # TASK: Explain the difference (in a Text cell) between correlation and covariance | ……/2 |
| 15 | # TASK: Create a new column with random values | ……/1 |
| 16 | # TASK: Determine the correlation and covariance between this new column and your previous two columns | ……/1 |
| 17 | # TASK: Export your notebook as a PDF. Upload both PDF and ipynb to Moodle | ……/1 |
|  |  |  |
|  | TOTAL: | ……/20 |