Group Project

SCSP 3223 Data Analytics Programming

Due: 4th February 2021

1. This is a group project with maximum 3 students per group.
2. The objective of this project is to expose students to data analytics process start from data preparation, data wrangling, visualization and machine learning application.
3. This project is divided into several tasks:
    a. Find a dataset which contains enough data to practice data preparation and analysis (at least 1000 rows)
    b. Formulate research question(s) from the dataset. What do you want to present?
    c. Do appropriate data cleaning, preparation and wrangling.
    d. Do appropriate data aggregation and group operations.
    e. Visualize your analysis using appropriate visualization.
    f. Perform appropriate machine learning method to answer your research question in (b).
4. Submission of the project will be via e-learning. Documents to submit:
    a. Report in pdf format which explain in details all the tasks in step 3. Report can be written in Microsoft Word or in Jupyter Notebook using the notebook Markdown.
    b. ipynb file (Python code)
    c. Video presentation (any software of your choice)
    d. Initial dataset (before performing any operations)
5. Please refer to the rubric given to see the assessment method of this project.

Please make sure that all group members' details (name and matric number) is written in all submitted documents (report, ipynb file and video presentation).

MARK SHEET FOR GROUP PROJECT

ITEMS	MARKS		
	1 - 3	**4 - 7**	**8 - 10**
PART A – REPORT			
Section i			
Dataset			
Research Question			
Section Total	(/ 20 × 5) =		
Section ii			
Data Cleaning and Preparation			
Data Aggregation and Group Operations			
Analysis and visualization			
Machine Learning			
Section Total	(/ 40 × 25) =		
Section iii			
Conclusion			
Overall Writing			
Section Total	(/ 20 × 5) =		
TOTAL (PART A)			
PART B – ipynb PYTHON CODE			
Code Review			
TOTAL (PART B)	(/ 20 × 5) =		
PART C – VIDEO PRESENTATION			
Verbal Presentation			
Visual Presentation			
Overall Presentation			
TOTAL (PART C)	(/ 30 × 10) =		
OVERALL TOTAL	**(/ 50 × 25) =**		

RUBRIC FOR GROUP PROJECT

ITEMS	MARKS		
	1 - 3	**4 - 7**	**8 - 10**
PART A - REPORT			
Dataset	Too simple with not enough data to perform data preparation and analysis	Appropriate with enough data to perform data preparation and analysis	Good choice of dataset with large amount of data which can be used to perform complete data preparation and analysis
Research Question	Questions overly simplistic, unrelated, or unmotivated	Questions appropriate, coherent, and motivated	Questions well motivated, interesting, insightful, and novel
Data Cleaning and Preparation	Overly simplistic or incomplete	Appropriate and complete	Appropriate, complete and advanced
Data Aggregation and Group Operations	Overly simplistic or incomplete	Appropriate and complete	Appropriate, complete and advanced
Analysis and visualization	Choice of analysis is overly simplistic or incomplete. Inappropriate choice of plots; poorly labelled plots; plots missing	Analysis appropriate. Plots convey information but lack context for interpretation	Analysis appropriate, complete, advanced, and informative. Plots convey information correctly with adequate and appropriate reference information
Machine Learning	Overly simplistic or incomplete	Appropriate and complete	Appropriate, complete and advanced
Conclusion	Conclusions are missing, incorrect, or not based on analysis	Conclusions relevant, but partially correct or partially complete	Relevant conclusions explicitly tied to analysis and to context

Overall Writing	Explanation is illogical, incorrect, or incoherent	Explanation is correct, complete, and convincing	Explanation is correct, complete, convincing, and elegant
PART B – ipynb PYTHON CODE			
Code Review	Code is messy and poorly organized; unused or irrelevant code distracts when reading code. Variables and functions names do not helpful to understand code.	Code is reasonably well organized. There is little unused or irrelevant code, or this code has been moved out of the main project files. Variable and function names generally meaningful and helpful for understanding.	Code very well organized. No irrelevant or distracting code. Variable and function names have clear relationship to their purpose in the code. Code is easy to read and understand.
PART C – VIDEO PRESENTATION			
Verbal Presentation	Illogical, incorrect, or incoherent.	Partially correct but incomplete or unconvincing	Correct, complete, and convincing
Visual Presentation	Cluttered, disjoint, or illegible	Readable and clear	Appealing, informative, and crisp
Overall Presentation	Verbal and visual presentation unrelated	Verbal and visual presentation related	Verbal and visual presentation clearly related