

CS5122 Descriptive and Predictive Analytics

Project

Due – April 22 before 11:55 PM (hard deadline)

Task

You are required to perform a comprehensive, descriptive and predictive analytics of a dataset related to Sri Lanka. A team consists of 2 members; therefore, find a class mate of your choice. No team can have more than 2 students without prior approval from instructor.

Steps

1. Find a suitable dataset with following characteristics:
 - Dataset should have at least a few 1,000s of entries and size of data should exceed over 0.5 MB. Data set should be related to Sri Lanka.
 - Dataset should be multi-attribute (having multiple columns) or multi-variate time series
 - Each student group should pick a separate dataset
 - You may look for sources such as <http://www.agridept.gov.lk/index.php/en/markert-information>, <http://data.gov.lk/>, <http://www.statistics.gov.lk>, <http://data.un.org/> for a dataset
2. Get instructor approval for your dataset. Instructor may suggest some changes depending on the feasibility and time constraints.
3. Briefly describe the dataset
4. Formulate 10 questions you may want to answer based on the dataset
5. Using suitable descriptive and predictive techniques that we discussed in the class, as well as other suitable techniques critically analyze your dataset
6. Answer the questions you formulated under Step 4. Provide relevant outputs, tables, and graphs
7. Provide any further findings specially ability to reduce dimensions, cluster, and predict values
8. Critically evaluate your finding and then discuss the takeaways from the dataset, e.g., business or societal impact

Rubric

#	Item	No of Pages	Marks
1	Description about the dataset	0.5	2
2	10 questions	0.5 - 1	2
3	Answers to 10 questions	5 - 10	10
4	Further findings/analysis	3 - 5	4
5	Critical evaluation of finding	1 - 2	4
6	Tools and techniques used	-	4
7	Self-reflection Give a synopsis on what you learned from the assessment, your time and effort, level of support you got from your team member, and feedback on what aspects of the assignment need to improve and how. This should be uploaded to Moodle as a separate document by each student.	0.5	2
8	Quality of report	-	2
	Total	10.5 – 19	30

What to Submit

- A report consisting of description about the dataset, 10 questions, answers, further findings, Critical evaluation, etc., as a PDF.
- Dataset as a separate file
- R commands as a separate file
- Name the file as **DPA_Project<index no1>_<index no2>.pdf** and **DPA_SelfReflection<index no>.pdf** Replace <index no> with your index number.