

PROJECT – (150 points total) Due: April 26, 2024 Friday

Group Assignment (You may form groups up to two individuals in one group)

Only one project needs to be submitted per group, other group members need to submit a text file indicating their group members and the CMU Login Id of the member who submitted the project.

A. Project report: 100 points (10 pts part1 + 100 pts)

(include the following in your project report)

- The name of the project, listing of group member(s) - name(s), and CMU Login Id(s),
- A brief description of the project,
- Part1 information:
 1. What is the source of data for your project? Provide a URL to a downloadable version of your dataset.
 2. How large is your dataset?
 3. Provide a description of the dataset (i.e., What is contained in the dataset?)
 4. List all “fields” of the data. If there are more than 15 fields, list the most pertinent “fields” for the types of questions you will ask of the data. Attach an extra page if more space is needed.
- Queries (at least 9 queries)
Three simple, three moderately complex, and three complex queries,
- The platform information (the Big Data environment and tools you have used for the project, including a reasoning about why you have used that environment and tools),
- EDF (exploratory data analysis) and query results,
- The analysis of results and Conclusions
- References (at least five),
- **Source code file(s)**, should be listed separately (not in the project report)
- Instructions how to run the source code.

B. The power point presentation file of the report (in a summarized way) (20 points)

C. The presentation (20 points) – about 10 mins.

Submission: (all material to both the K-drive and Blackboard)

- A. The project report due: April 26, 2024, Friday
- B and C due: April 24, 26, 2024, Wednesday & Friday

Presentation schedule: during the regular class time.

Additional time will be made available for presentations if needed.