PROJECT – Groupings and Data Proposal (10 points) Due: April 05, 2024 Friday

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

**Address the following questions about the term project. Only one proposal 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 proposal.**

**(See the end of Links document on BlackBoard for potential Big Data public data resources.)**

**Project:** General information (a short summary – detailed version will be provided later):

* Must be a programming project,
* For the final project you will turn in Source Code, link – I have to be able to run, instructions on how to do that –

and resulting output files,

* In class presentation,
* A project report, including background and motivation, project description, algorithms used, system and implementation design, tools used, project results and conclusions

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.

5. List at least 4 types of “useful” and “significant” queries that you may want to ask of the data

(an initial query design, will be adding more queries later).

6. List each member of the group. A group may consist of 1 to 2 individuals.

Name CmuLoginID

**What to Submit:**

Submit your answers to the questions above as a Word document or PDF file through Blackboard.