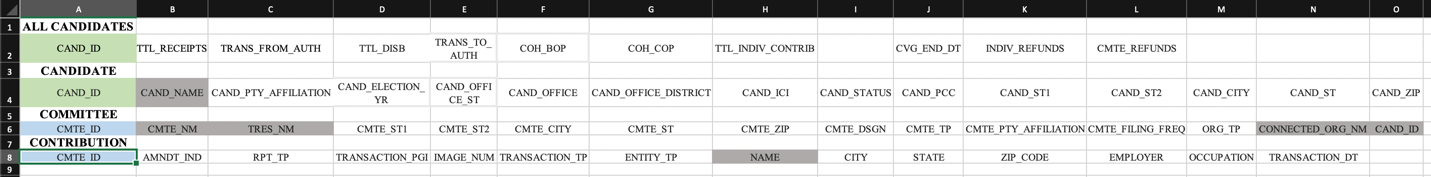
1. Develop and draw an ER-diagram that describes the listed data files. Show on the diagram the following

* Entities: Each entity should have a unique name (10 Marks)
* For each entity, List the fields representing the entity. The fields can be found using the Metadata file link column (Column c). Only list fields indicated in column D.  (15 Marks)
* For each field clearly indicate if the field is a key (Underline it) (10 Marks)
* Indicate if the field is atomic, multivalued or composite fields. (10 Marks)
* Provide on the diagram the relationship names and cardinalities (one:one, one:many, many:many) (15 Marks)

Diagram

Description automatically generated

2. Draw a relational DB schema and provide relationships among all the entities (20 Marks)



3. What is your opinion on the data files provided in column B? Would you consider them to be tidy? Provide any suggestion on improving the design of the data files. Examples of not tidy data could be:

(1) columns headers are values,

(2) columns need separation,

(3) whether variables are stored in both rows and columns, or

(4) whether the tables have just one kind of observation -- 20 Marks

When I firstly downloaded the file that located on column B, it was the .txt file which I always had hard time to understand and read the data. As soon as I understand and read it, I started to find any common sign that divide the values on txt file, which was “| ” in this txt. file. If I were in the creator of this data, I would like to use “,” rather than | because it was hard to understand it as divider.

It need to be clean which we need to tidy the data. Since each txt file contains a many value with many attributes, it should have the name of variable. For example, I took screenshot from txt file for the example purpose.

Text

Description automatically generated

If it contains names of column or indicate name of attribute on here, it would be much easy to make it clean and organize the data. Also, easy to understand the data.

On top of that, we didn’t use all attributes on our work.. Most of txt files, we used all variables however, we did not used all variables at “candidate-committee linkage” which we need to remove or make new data frame that only contain specific variables.

Also, It might only happen on my system, which was a few link were only hyperlinked and others weren’t. some URL that does not contain hyperlink, I had to copy and paste the URL on Internet to download the file.

Lastly, I could observe some blank were existed on txt file which we need to clean it. I would like to remove and combine variable where blank are existed. To remove blank value and combine with others if we can, I would like to separate the variable into multiple since we have a lot of attributes on data.