Group 4 Frustration Report.

In following the week 2 instructions our group encountered many difficulties when selecting the three initial indicators. Most of this occurred when understanding what the indicators represent. Some of them can be very vague or too concise and as a result it was hard to find three that would work well together. Also the options were too much. Just for the databases mentioned in week 2 assignment where the indicators had to be decided from us, there were always thousands to choose from. For example World Development Indicators had over 1400 indicators alone. Some would sound too similar, resulting in some confusion. WIthout a data dictionary it was really hard to understand what each indicator stood for.

From our understanding of the structure of the DB, there was not much issue. It was pretty much understanding what is in the tables and what each indicator meant. In the case of the tables the indicators were pretty evident to be attributes of the tables and the years to be a column as well or its own table if needed. There was a known issue of finding which databases should be connected in the ER diagram, due to some not matching one another like Enterprise Survey and Climate Change Knowledge portal.

Once all the Indicators have been decided for our databases, we worked on creating the 60 questions. This one took the longest for our group to figure out, because based on the three different database tables we selected and the many indicators, we had to come up with a lot of questions and some that made sense. Some issues that our group experienced was finding analytical questions and another issue is using the indicators to come up with questions. Coming up with these questions requires us to have a good understanding of the indicators, which if not understood can cause poor questions.

With all these challenges, how were they overcome? For the Indicators and DB, we only could just use our best assumptions and go with that. Choosing the DB, we just chose easier ones that related to one another and with the indicators, chose three that sounded interesting. Regarding the indicators selecting the 15 indicators for each table, we used ones that sound interesting but can also relate to one another. For example in the World Development Indicators, there were indicators regarding electricity, so some of those were chosen as indicators to make the table seem more cohesive.

With the research questions, that one took our group the longest to figure out how to get through. Meeting the requirements for the project along with coming up with 60 of them was a difficult task. A way we decided to go around it is by using the indicators as a way to help. The way the indicators are structured, many questions can be made based on each of them. Outside help was used, using tools like researching what the topics

meant, using AI to help clarify information and assist in getting ideas for questions.. After submission we later understood that this may not have been the best approach for this as will be shown in the self reflection part of the report.

Now that the ERD, indicators and questions are created what reflections and lessons do we have? For the ERD/database selection there should have been the inclusion of the actual databases, as that would make the ERD look more complete with the world development indicators. Only using the indicator as a table and not stating the table/indicator looks to have been the wrong decision. This should have been in combination or expanded when choosing the 30 Indicators for each database.

When coming up with questions, the reflections we have was what type of questions we have made. From the class and review of the questions, these questions seemed more like that they were more a basic bar graph could answer. If this were to be done again, I believe a more emphasis on questions that can use analytical models should have been thought off. This could involve looking at individual analytical problems/models and see what questions we could come up with that could take advantage of analytics and not just a basic chart.

This assignment in combination with the lecture has shaped our group's thinking on handling bigger datasets. It shows how little information on the dataset can be given and how we will need to approach how. In this context if this project were to be redone then we would need to better look into what the indicators were in context of the DB and see how analytics will help the entire database as a whole (the DB lined out in the ERD) that way we can come up with better questions.