# Overview of the Assignment:

It is time to start design this is just a draft

**Part 1**: Are you working on your own or with a partner? If with a partner, provide their name. If on your own, just state that this is the case.

I am working alone on this project.

**Part 2**: Determine the project scope

* In a short paragraph, describe the topic you wish to explore – an update if any

The topic I wish to explore for my data warehouse project is the management of Parts Unlimited's EV parts business. This topic was inspired by "The Unicorn Project," which describes the challenges and opportunities of digital transformation in a large organization. Specifically, I plan to focus on storing and analyzing data related to charging stations, EV product price lists, EV car information, and EV customer data, such as their geographic location, the model of the car they own, and location data. Parts Unlimited already sells EV parts, and my goal is to improve the organization's data management, reporting, and analysis capabilities related to this business.

* Update the five business questions that your data warehouse will answer.

**Part 3:** Data Sources

* Provide two data sources you will be using, for each data source list the number or columns and rows that are in each data source. Provide a header and first 5 rows from each source in separate file.
* What is the URL or location of the data?
* What information does this data provide that will help answer one or more of the above questions?
* Do you see any issues in the data that will require transformation.

**Part 4**: Dimensions - Review the data and the business questions from part 2.

* What fields (attributes) are in the data that will be used for the dimensions.
* Determine the dimension tables. There should be at least two non-date dimensions and one date dimension for each fact table.
* At least one (non-date) dimension in your design should have a hierarchy.
* What are the attributes that will be tracked via slowly changing dimensions?
* What attributes within the dimensions will need transformation before they are loaded into the dimension, for example it could be to build consistency or any other issues? This is where for example you might build case statements in your code to handle various scenarios. Two to three examples showing some sample data and what you think the transformation will be during your ETL would be helpful here.

| Table Name | Table Attributes | SCD Type  (info) | Transform  Needed  Attributes | SCD Tracked  Attributes | Q- No |
| --- | --- | --- | --- | --- | --- |
| ev-car-population | PK DOL Vehicle-Id  **FK location-id**  FK **EV-Charg-Stat-Rec-Date**  Make  Model  Model-Year  Electric-Vehicle-Type  **Status-Flag**  **Status-DeAct-TimeStamp** | SCD type 2, if a car no longer exist we can check the status and check deactivation date, we can track the record date  The location of the car might change. So we can track this information by FK to the location table. | **Loc-ID:** Not exsist  this will be latitude and longitude concatenation  **EV-Charg-Stat-Rec-Date, Status-Flag**  **Status-DeAct-Timestamp:**  Not exist,  The data creation date mentioned in the source but not included in the csv file | **FK location-id**  **EV-Charg-Stat-Rec-Date**  **Status-Flag**  **Status-DeAct-TimeStamp** | 1 |
| Location Dimension Table  (SCD type 3) | PK- location-ID  City  State  ZIP  Latitude  Longitude  Latitude-prev  Longitude-prev | The location of the car might change. So we can track this information by current and previous. We add surrogate key the track the changes | **Loc-ID:**  this will be latitude and longitude concatenation  **Latitude-prev Longitude-prev**not exist I will need to add this info when I load | Latitude-prev  Longitude-prev | 1 |
| EV Charging Station  (SCD type 2) | PK Id  FK Location-ID  **FK EV-Charg-Stat-Rec-Date**  Station Name  **Status**  **Expire\_timeStamp** | We keep track of of the change over time with with the record creation date  , status and record expiration date | EV-Charg-Stat-Rec-Date and  Status  Expire\_timeStamp. This columns will be added | **FK EV-Charg-Stat-Rec-Date**  **Status**  **Expire\_timeStamp** | 1 |

**Part 5**: Facts – Review the data and the business questions from step 1.

* What measurements are in the data that will be used for the fact tables?
* What measures will you be calculating (i.e. using an aggregate function, or some other transformation – recall as an example some of the aggregation you did in assignment 1A)

**Part 6**: Design – Create a Draw.io, Visio or Lucidchart diagram of your constellation data warehouse design.



Project scoping is graded based on the following:

1 – On track, 0-Off track, .5 – partially on track