# 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.

**Q1- Is there a correlation between the number of EV charging stations in a particular area, the number of EV cars registered in that area, and the time period in which they were registered? And if so, how can we use this information to optimize our expansion strategy and better serve our customers over time?**

Parts Unlimited **Business Development** team is interested in using this information to plan and prioritize their expansion strategy for EV parts and charging station installations. The marketing team can use this information to tailor their marketing campaigns and promotions to specific regions.

**Q2:What is the current range of EV parts manufacturers and their product offerings that Parts Unlimited is working with, and how can this information be leveraged to optimize their product mix and pricing strategy for increased revenue in the growing EV market? Additionally, how has this range of manufacturers and their product offerings evolved over time, and what trends can be identified for future business planning?**

**Purchasing team** is interested in this question , they are responsible for sourcing and procuring products for Parts Unlimited, including EV parts from different manufacturers. By understanding the current range of EV parts manufacturers and their product offerings, the Procurement department can make informed decisions about which manufacturers to work with, what products to stock, and how to optimize their product mix and pricing strategy for increased revenue in the growing EV market

**Q3-How does the popularity of different EV models and plug types vary by geographic region, and how can Parts Unlimited use this information to target their marketing and sales efforts?**

**The marketing and sales department** in Parts Unlimited can use the information to target their efforts more effectively. For example, if a certain geographic region shows a higher preference for a particular EV model or plug type, the marketing and sales team can focus their promotional activities and campaigns in that region to increase sales. They can also use this information to tailor their messaging and product offerings to better meet the needs and preferences of customers in each region.

**Q4-What is the relationship between the location and price of existing EV charging stations over time, and how can this information be used to determine the feasibility of adding new charging stations in the vicinity of Parts Unlimited's stores in partnership with companies like Tesla? Additionally, how can this information be leveraged to increase revenue and customer convenience?**

Parts Unlimited is considering providing EV charging stations as a new service. Therefore the Research and Development team can determine if there is a relationship between price and location. This information can provide insights into the cost and demand for EV charging stations in different locations and inform the decision on where to install new stations.

**Q5-What are the most popular types of EV charging plugs for different model years, and how have these trends changed over time? How can Parts Unlimited leverage this information to ensure that they stock the appropriate parts for EV charging stations and stay ahead of industry trends?**

The Supply Chain department at Parts Unlimited would need this information to know the most popular types of EV charging plugs for different model years, as well as how these trends have changed over time, in order to ensure that they stock the appropriate parts. By analyzing this information, they can optimize their inventory management, pricing strategy, and product mix to meet the changing demand and stay ahead of industry trends.

**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 a 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.

1. **Product Info- EV Vest**

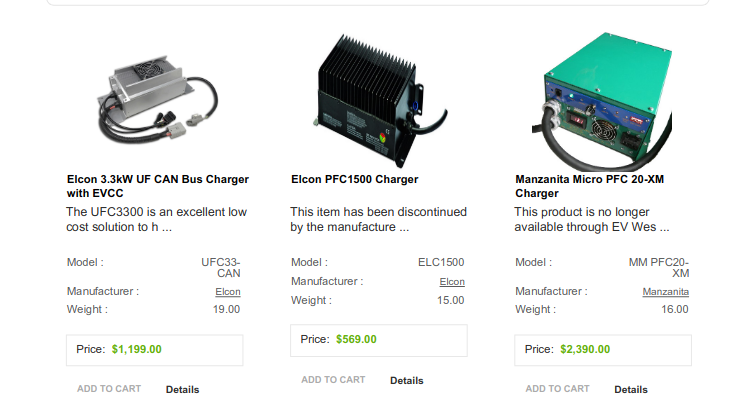
**Num Columns:**5

**Num Row:**50

**URL:**[**https://www.evwest.com/catalog/index.php?cPath=4&osCsid=55f0661764899b07540a29eff0700422**](https://www.evwest.com/catalog/index.php?cPath=4&osCsid=55f0661764899b07540a29eff0700422)

The data set provides information about EV parts, manufacturer , weight and price information.

The data is not in the pdf version. I need to create a python script to extract and format the information.



1. **EV Car Dataset:**

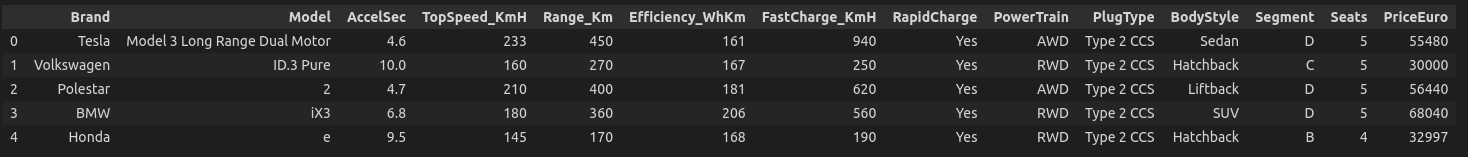
**Num Columns:**14

**Num Row:**103

**URL**[**https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset?select=ElectricCarData\_Norm.csv**](https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset?select=ElectricCarData_Norm.csv)

Provides information about EV car and models and charging plug types.

This dataset is clean, currently I do not see any issues.



1. **EV CarCharging**

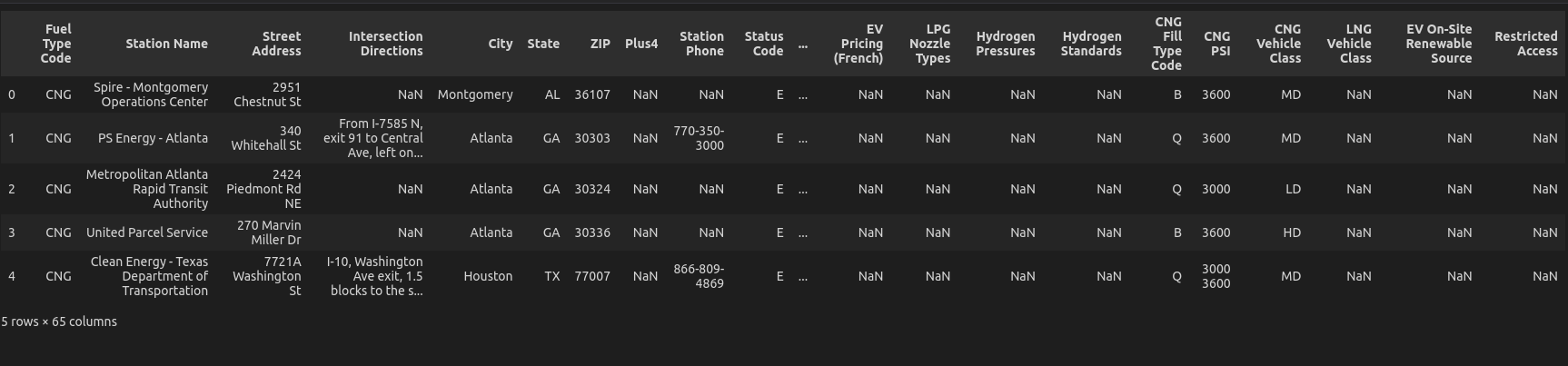
**Num Columns:**65

**Num Row:**56800

**URL:**[**https://catalog.data.gov/dataset/alternative-fueling-station-locations-422f2/resource/341957d8-daf6-4a38-ab1d-8ec1bc21cfb9**](https://catalog.data.gov/dataset/alternative-fueling-station-locations-422f2/resource/341957d8-daf6-4a38-ab1d-8ec1bc21cfb9)

Provides information about the current list of charging stations and their locations.

There are a lot of missing values that need to be cleaned and lat and long location info needs to be identical with the EV Car Population dataset and price information in a text format we need to format it so it is a number.



1. **EV Population Data**

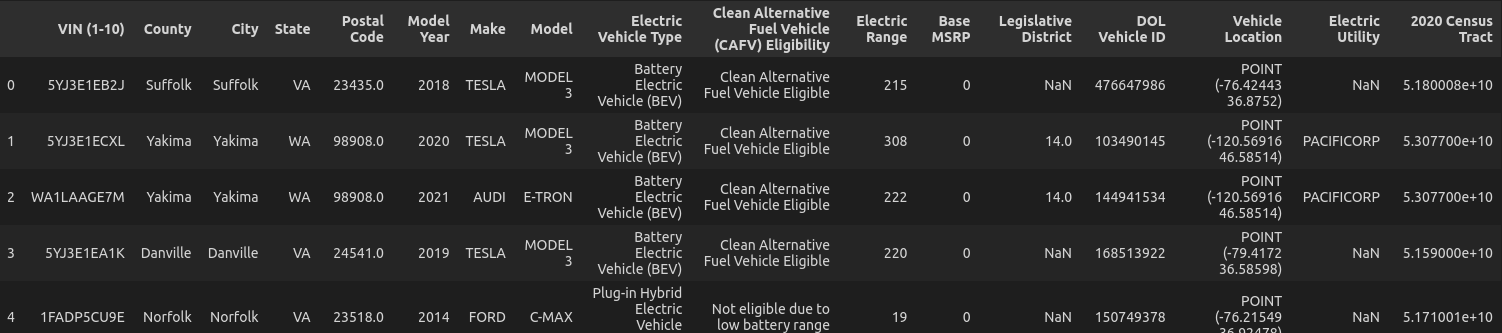
**Num Columns:**17

**Num Row:** 121978

**URL:**[**https://catalog.data.gov/dataset/electric-vehicle-population-data**](https://catalog.data.gov/dataset/electric-vehicle-population-data)

This dataset shows EV cars that are currently registered through Washington State Department of Licensing (DOL).Provides information about EV cars, cars type and registered locations.

Location fields have to match with the format of the EV Charging dataset. There are some missing fields. Currently I do not see any other issues related to this dataset.



**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.

**1-Table Name :** ev-car-population

**Table Attributes:**

* PK DOL Vehicle-Id
* SK Car-Pop-Id
* FK location-id
* FK EV-Charg-Stat-Rec-Date
* Make
* Model
* Model-Year
* Electric-Vehicle-Type
* Status-Flag
* Status-DeAct-TimeStamp

**SCD TypeInfo:** SCD type 2. if a car no longer exists we can check the status and check deactivation date, we can track the record date. and sk help us track the history

**SCD Tracked Attributes :**

* SK CAR-POP-ID
* Status-Flag
* Status-DeAct-TimeStamp

**Transform Needed Attributes:**

* Loc-ID:Location Id Not exist this will be latitude and longitude concatenation
* EV-Charge-State-Rec-Date, The data creation date mentioned in the source but not included in the csv file, I will need to insert this data
* Status-Flag: I will need to insert this date based on other columns info
* Status-DeAct-Timestamp: I will need to insert this date based on other columns info

—-----

**2-Table Name :** EV Charging Station

**Table Attributes:**

* PK Id
* SK Charging\_Stat\_ID
* fk location-ID
* Location-ID
* EV-Charg-Stat-Rec-Date
* Station Name
* Updated At
* Date Last Confirmed
* Updated At

**SCD TypeInfo:** SCD type 2. if a charging station has been updated. Updated at field allow us to track we also created sk to track history.

**SCD Tracked Attributes :**

* SK Charging\_Stat\_ID
* Updated-At

T**ransform Needed Attributes:**

* We need to create surrogate key

—-----

**3-Table Name :** Location Dimension Table

Table Attributes:

* PK location-ID
* City
* State
* ZIP
* Latitude
* Longitude
* Latitude-prev
* Longitude-prev

**SCD TypeInfo:** SCD type 3. We have prev attributes for lat and long

**SCD Tracked Attributes :**

* SK Charging\_Stat\_ID
* Updated-At

**Transform Needed Attributes:**

* location columns needs some transformation

—---

**4-Table Name :** EV Charging Station

**Table Attributes:**

* PK ManU-ID
* sk man-track-id
* FK Vendor\_since
* manufacturer-name
* Updated\_at

**SCD Tracked Attributes :**

* vendor since
* Updated-At
* status flag

**SCD TypeInfo:** SCD type 2.

**SCD Tracked Attributes :**

* SK Charging\_Stat\_ID
* Updated-At

**Transform Needed Attributes:**

* we are missing vendor since and updated at we will fill those fields as current date

**5- Product :** EV Charging Station

**Table Attributes:**

* PK produc-ID
* Product Category
* product-name
* product-name-prev
* Product-category-prev
* updated-timestamp

**SCD Tracked Attributes :**

* product-name-prev
* Product-category-prev
* updated-timestamp

**Transform Needed Attributes:**

* we need to bring all the values from the pdf and make sure that they are in the correct format.

**5-Table Name** : EV Car

(we can implement a hierarchy in this table like car company -brand-model)

**Table Attributes:**

* PK Id
* Brand
* MODEL
* PLUG-TYPE
* MODEL YEAR

**SCD Tracked Attributes :**

* vendor since
* Updated-At
* status flag

**SCD TypeInfo:** SCD type 1.

**SCD Tracked Attributes :** None

**Transform Needed Attributes:** Currently I do not see any adjustment

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

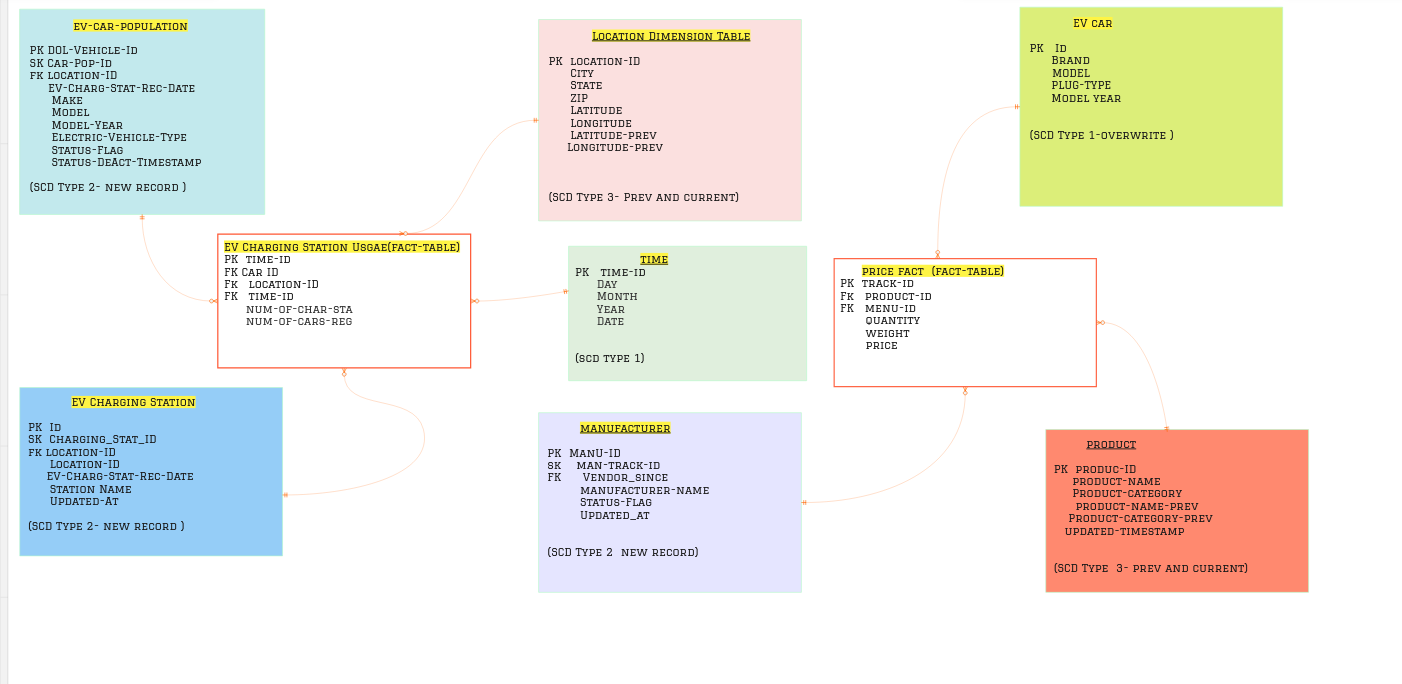
* What measurements are in the data that will be used for the fact tables?
* What measures will you be calculating (i.e. using an aggregate function, or some other transformation – recall as an example some of the aggregation you did in assignment 1A)

EV Charging Station Usage: Calculating number of cars and stations to measure, it can also calculate average price per location etc

price (fact-table):Calculating product quantity, weight, and price , percentage of product quantity and average weight and price

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

I ran out of time this week because I had to search for new project ideas and adjust my business questions. I will iterate through the whole process again.





Project scoping is graded based on the following:

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