Overview of the Assignment: Let's focus on the ETL Process.

First iteration of ETL: Take a part of your design and implement ETL. The scope at this point is limited and should include just two of your dimensions and one fact table. Using Python, SQL document and walk through the initial process of loading your data into staging (SQL staging or a data frame) doing some transformation and loading into your dimensions and a fact table.

Part 7: Provide code and screenshots of loading your data into staging/data frame

Part 8: Provide code and screenshots of transforming the data. Perhaps you are adjusting for consistency of data or calculating aggregates.

Part 9: Provide code and screenshots of loading the data into the two dimensions and the fact. At this time, you do not need to worry about maintenance of slowly changing dimensions, the focus is on the initial data load. If you are loading into SCD2 or SCD3, make sure to show the SCD maintenance attributes populated.

For part 7, 8 9: This week I worked on extracting data from EV West pdf documents. Extract : First I have extract all the pdfs and saved as a csv files and later I combined them as a big csv Transform : I have transformed data so that it has all the attributes. Load : I have created two dimention table and one fact table for this dataset.

I am going to attach these steps in two parts.

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 in this project

Part 2: Determine the project scope

- **In a short paragraph, describe the topic you wish to explore --an update if any**
- **Update the five business questions that your data warehouse will answer.**

I haved added the project objective part

Project Description 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 products, customer purchases, and geographic location. 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.

Project Objective The primary objective of this project is to create a concept data warehouse project using a slowly changing dimension approach. I will identify the business requirements by formulating specific business questions that the data warehouse will help answer. Based on these requirements, I will develop a consolidated ERD schema for the data warehouse.

I will focus on storing and analyzing data related to EV charging stations, product information, customer purchases, and geographic location. To accomplish this, I will use a slowly changing dimension approach to ensure that data is properly tracked over time. Specifically, I will use SCD Type 2 to track changes to the customer's location, SCD Type 3 to track changes to the EV product information, and SCD Type 1 to track changes to the charging station information.

Finally, I will use Tableau and Postgres SQL to create reports that will help the business make data-driven decisions based on the data stored in the data warehouse. By using a consolidated ERD schema and slowly changing dimensions, I will ensure that the data warehouse accurately reflects the business needs and can be used to provide valuable insights to the organization.

Business Questions:

Business Question 1
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.

Business Question 2

What is the currento 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

Business Question Q3

How does the popularity of different EV models and electric vehichle 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.

Business Question 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.

Business Question Q5
What are the top five manufacturers by percentage of products sold by Parts Unlimited, and where are these companies located?

The Supply Chain department at Parts Unlimited can utilize this information to determine the percentage of products supplied by each manufacturer, as well as their respective locations. This can enable the department to better manage logistics and distribution of the products, while also ensuring compliance with any legal requirements related to supply chain operation

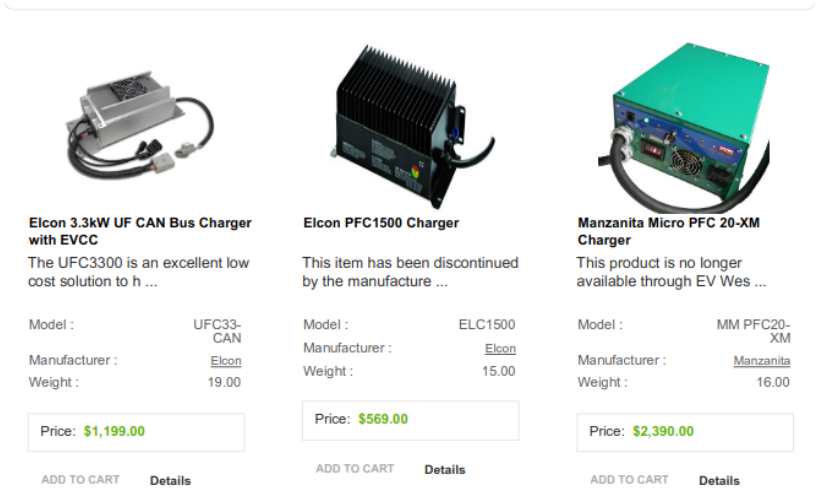
Part 3: Data Sources

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.

- 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.

Dataset Name	Columns	Rows	URL
1-Product Info - EV West	5	50	EV West

The data set provides information about EV parts, manufacturer, weight and price information. The data is not available in the PDF version and requires a Python script to extract and format the information.



Dataset Name	Columns	Rows	URL
2-Alternative Fueling Station Locations	65	56,800	Data.gov

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. Price information is available in a text format and needs to be formatted so it is a number.

	EV station Dataset Column Names		EV Car Dataset Column Names		EV station Dataset Column Names		EV station Dataset Column Names
Index		Index		Index		Index	
1	Fuel Type Code	21 | EV Ot	EV Other Info	41 |	Access Days Time (French)	61	Intersection Directions (Russian)
2	Station Name	22 | EV Ne	EV Network	42 |	BD Blends (French)	62	Access Days Time (Russian)
3	Street Address	23 | EV Ne	EV Network Web	43 |	Groups With Access Code (French) | 63 |	63	BD Blends (Russian)

Index	EV station Dataset Column Names	Index	EV Car Dataset Column Names	Index	EV station Dataset Column Names	Index	EV station Dataset Column Names
4	Intersection Directions	24	Geocode | Status Geoco	44	Hydrogen Is | Retail	64	Groups With | Access Code (Russian)
5	City	25	Latitude | Latit	45	Hydrogen Status Link (French)	65	Hydrogen | Status Link (Russian) H
6	State	26	Longitude	46	NG Vehicle | Class (French) NG V		
7	ZIP	27	Date Last | Confirmed Date	47	LPG Primary (French)		
8	Plus4	28	ID | ID	48	E85 Blender | Pump (French)		
9	Station Phone	29	Updated At | Updat	49	EV Connector Types (French)		
10	Status Code	30	Owner Type | Code Owner	50	Country (French)		
11	Expected Date	31	Federal | Agency ID Feder	51	Intersection Directions (Spanish)		
12	Groups With Access Code	32	Federal | Agency Feder Name	52	Access Days Time (Spanish)		
13	Access Days Time	33	Open Date | Open	53	BD Blends | (Spanish)		
14	Cards Accepted	34	Hydrogen | Status Link Hydro	54	Groups With Access Code (Spanish)		
15	BD Blends	35	NG Vehicle | Class NG Ve	55	Hydrogen Status Link (Spanish)		

EV station Dataset Column Index	Names	EV Car Dataset Column Index	Names	EV station Dataset Column Index	Names	EV station Dataset Column Index	Names
16	NG Fill Type Code	36	LPG Primary | LPG P	56	NG Vehicle Class (Spanish)		
17	NG PSI	37	E85 Blender Pump | E85 B	57	LPG Primary (Spanish)		
18	EV Level1 EVSE Num	38	EV Connector Types | EV Co	58	E85 Blender Pump | (Spanish)		
19	EV Level2 EVSE Num	39	Country | Count	59	EV Connector Types (Spanish)		
20	EV DC Fast Count	40	Intersection Directions (French) | 60 | Inter	60	Country (Spanish)		

```
import warnings
warnings.simplefilter(action='ignore', category=Warning)
```

```
import  pandas as pd
pd.set_option("display.max_columns",10 )
```

```
import  pandas as pd
ev_station = pd.read_csv('data/alt_fuel_stations.csv')
ev_station.head(5)
```

```
  Fuel Type Code                                       Station Name  \
0            CNG                  Spire - Montgomery Operations Center
1            CNG                                    PS Energy - Atlanta
2            CNG         Metropolitan Atlanta Rapid Transit Authority
3            CNG                                  United Parcel Service
4            CNG  Clean Energy - Texas Department of Transportation

         Street Address                                Intersection Directions  \
0      2951 Chestnut St                                                    NaN
1      340 Whitehall St  From I-7585 N, exit 91 to Central Ave, left on...
2   2424 Piedmont Rd NE                                                    NaN
```

```
3  270 Marvin Miller Dr                                                  NaN
4   7721A Washington St  I-10, Washington Ave exit, 1.5 blocks to the s...

         City  ...    CNG PSI CNG Vehicle Class  LNG Vehicle Class  \
0  Montgomery  ...       3600                MD                NaN
1     Atlanta  ...       3600                MD                NaN
2     Atlanta  ...       3000                LD                NaN
3     Atlanta  ...       3600                HD                NaN
4     Houston  ...  3000 3600                MD                NaN

  EV On-Site Renewable Source Restricted Access
0                         NaN               NaN
1                         NaN               NaN
2                         NaN               NaN
3                         NaN               NaN
4                         NaN               NaN

[5 rows x 65 columns]
```

Dataset Name	Columns	Rows	URL
3- EV Population Data	17	121,978	Data.gov

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

Index | Elect	Electric_Vehicle_Population |
1 | IN (1	IN (1-10) |
2 | Count	County |
3 |	City |
4 | State	State |
5 | Posta	Postal Code |
6 | Model	Model Year |
7 | Make	Make |
8 | Model	Model |
9 |	ctric Vehicle Type |
10	Clean Alternative Fuel Vehicle (CAFV) Eligibility |
11 | Elect	Electric Range |
12 | Base	Base MSRP |
13	Legislative District |
14 | DOL V	DOL Vehicle ID |

Index | Elect	Electric_Vehicle_Population |
15 | Vehic	Vehicle Location |
16	Electric Utility |
17	2020 Census Tract |

```
ev_pop = pd.read_csv('data/Electric_Vehicle_Population_Data.csv')
ev_pop.head(5)
```

```
   VIN (1-10)    County      City State  Postal Code  ...  \
0  5YJ3E1EB2J   Suffolk   Suffolk    VA      23435.0  ...
1  5YJ3E1ECXL    Yakima    Yakima    WA      98908.0  ...
2  WA1LAAGE7M    Yakima    Yakima    WA      98908.0  ...
3  5YJ3E1EA1K  Danville  Danville    VA      24541.0  ...
4  1FADP5CU9E   Norfolk   Norfolk    VA      23518.0  ...

   Legislative District DOL Vehicle ID             Vehicle Location  \
0                   NaN      476647986    POINT (-76.42443 36.8752)
1                  14.0      103490145  POINT (-120.56916 46.58514)
2                  14.0      144941534  POINT (-120.56916 46.58514)
3                   NaN      168513922    POINT (-79.4172 36.58598)
4                   NaN      150749378   POINT (-76.21549 36.92478)

  Electric Utility 2020 Census Tract
0              NaN      5.180008e+10
1       PACIFICORP      5.307700e+10
2       PACIFICORP      5.307700e+10
3              NaN      5.159000e+10
4              NaN      5.171001e+10

[5 rows x 17 columns]
```

Dataset Name	Columns	Rows	URL
4- Manufacture Location	14	103	Kaggle

Provides information about EV car and models and charging plug types. This dataset is clean, currently I do not see any issues.

Index	EV Car Dataset Column Names	Index	EV Car Dataset Column Names
1	Brand	9	PowerTrain
2	Model	10	PlugType
3	AccelSec	11	BodyStyle
4	TopSpeed_KmH	12	Segment
5	Range_Km	13	Seats

Index	EV Car Dataset Column Names	Index	EV Car Dataset Column Names
6	Efficiency_WhKm	14	PriceEuro
7	FastCharge_KmH		
8	RapidCharge		

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 :1-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 |