* Project Subject Area: In this project I am planning to do the analysis on Battery Electric Vehicle (BEV) and Plug-in Hybrid Electric Vehicle (PHEV) and also try to find which geo location and which age group people are more likely to by these cars.
* Data Sources:
  + Flat File:
    - Here is the comma separated file for the electric vehicle data
    - <https://catalog.data.gov/dataset/electric-vehicle-population-data/resource/fa51be35-691f-45d2-9f3e-535877965e69?inner_span=True>
  + API:
    - Random user data for the vehicles
    - <https://randomuser.me/api/>
  + Website:
    - 2022 Third Quarter cost of living index by state
    - <https://meric.mo.gov/data/cost-living-data-series>
* Relationships

The 3 data sources I am planning to connect are not related to EV cars, but I will change it accordingly, the first data is completely about EV. Below are the list of attributes for that data source.

VIN (1-10) object  
County object  
City object  
State object  
Postal Code int64  
Model Year int64  
Make object  
Model object  
Electric Vehicle Type object  
Clean Alternative Fuel Vehicle (CAFV) Eligibility object  
Electric Range int64  
Base MSRP int64  
Legislative District float64  
DOL Vehicle ID int64  
Vehicle Location object  
Electric Utility object  
2020 Census Tract int64

It has DOL vehicle ID, which I will use to generate dummy users for these vehicles using the API. So that I can use the DOL Vehicle ID to join the user back to the data.

State wise cost of living index has the state abbreviation, which is related to the State in the EV source data where the car is registered for.

In this project I am trying to do analysis on EV cars. The data used in this project have two types of vehicles Battery Electric Vehicle and Plug-in Hybrid Electric Vehicle. I have used three sources for the data, from one source I am getting the car details, the users of the vehicles I am creating with random user API, that is an API that generates data for random users. The third source I am using for cost of living for different states. The car information source is Washington State Department of Licensing (DOL), where different cars are registered for different states. It will be more clear when we will do the state level analysis.

What I am trying to find out, is who are the users and what age group they belong to who purchases the EV cars, also do the states that have excessive cost of living only purchases the EV cars or its uniformly distributed among the states.

Assumptions : The user data is randomly generated data, so there is no relationship between the cars and the users, for this analysis purposes we are assuming that the users are the actual users for the EV cars.

Challenges : The user API has all types of information for the users, I have to be selecting while choosing the attributes that may have high impact on the EV car purchase. Selecting proper attributes for the analysis seems to be a little challenging. Another challenge could be creating new attributes based on the existing attributes, to identify the actual users of EV cars.