|  |
| --- |
| Electric Vehicle - ETL Project |
| Project Proposal  Extract location information for:   * Location specific information; and * Registration information   Of Electric Vehicles in Washington State (WA) USA.  Transform and clean data frames within pandas and prepare clean CSV files available for load directly into PostgreSQL |
| Team Members:  Srivasthan Krishna  Kevin Kirkpatrick  Ali Asghir  Sonam Bhandari |
| Tools Used:   * Jupyter Notebook   + Pandas   + SQL Alchemy * PostgreSQL |
| Extract  Imported the csv files from <https://www.atlasevhub.com/public/dmv/wa_ev_registrations_public.csv>  <https://catalog.data.gov/dataset/electric-vehicle-population-data> |
| Transform  From Location extract we retained   * VIN (1-10) * County * City * ZIP Code * Model Year * Make * Model * Electric Range   Graphical user interface, text, application  Description automatically generated  From Registration extract we retained   * Vehicle ID * ZIP Code * VIN Prefix * Vehicle Name * TechnologyGraphical user interface, text, application    Description automatically generated |
| Load  PostgreSQL   1. Created EVP database in PostgreSQL with separate tables for ‘Locations’ and ‘Registrations’   Graphical user interface, text, application  Description automatically generated   1. Imported data into tables using PostgreSQL import   Graphical user interface, text, application  Description automatically generated |
| Insights  An important aspect of this project was (prior to building the ETL phase and database) determining the required fields and tables, there relevance for analysis and future use.  Given the commonality of VIN details it made sense to use PostgreSQL so that the two tables could be related to each other using the VIN information.  This then guided what cleansing and modification of data was required in order to have an efficient and relevant ETL process for the current but also future datasets. |