**ETL Project Report**

**Group Members: Rick, Petra and Emrah**

**Introduction:**

Our group is using The Big Mac index as an economic indicator along with Electric Vehicle sales to analyze if there is a correlation between the economic state of the countries and the Electric Vehicle usage.

The Big Mac Index is a way of measuring the purchasing power parity between two currencies and provides a test of the extent to which market exchange rates result in goods costing the same in different countries. It "seeks to make exchange-rate theory a bit more digestible"

**Main Work Flow:**

* Extract Dataframe from HTML
* Cleaning up unnecessary data
* Filtering out data
* Importing data to MySQL Database
* Plot

**1.Extraction**

We used two datasets from Wikipedia Electric car use by country, and the Big Mac index and the sources for our dataset are as follows,

Our original data source is from

* HTML / Electric Car use by country

<https://en.wikipedia.org/wiki/Electric_car_use_by_country>

* HTML/ The Big Mac Index.

<https://www.worldatlas.com/articles/the-world-s-most-expensive-countries-measured-by-the-big-mac-index.html>

**2.Transforming**

Our first steps in cleaning up the datasets involved figuring out which variables were not relevant. For the EV df, we dropped irrelevant columns and renamed the State Code column to State.

For the Electric Vehicle (EV), first we imported all necessary dependencies, scrapped EV data from the provided url and create Dataframe through panda. We cleaned all unnecessary columns and rows, also renamed columns as well as the values by deleting unnecessary data using splitting method.

Same with Big Mac (BM), used pandas to read the HTML and set it to a Dataframes and revised the columns so they are shorter. After cleaning the datasets, we merged EV and BM tables on countries.

**3. Loading**

We created a database and respective tables to match the columns from the final Panda’s Data Frame using MYSQL and then connected to the database using SQLAlchemy library by creating connection and engine then we load our tables into databases. We also checked if the values are loaded properly by pd.read\_sql\_query to show final df

And we’ve scatter plots for,

“EV Percent vs BM\_Dolloars”, “BM vs EV per Thousand people”, “BM vs Percent of EV on the road”

We did not observe a solid correlation between Big Mac index and Electric Vehicle Usage.