Election Data Analysis

# Data Sources

For the purpose of the ETL project data was sourced from following 3 sources:

* [www.Kaggle.com](http://www.Kaggle.com) 🡪 SQLite database containing the facts and attributes related to the population demographics across various counties in states of US.
* <https://Data.world> 🡪 Data related to presidential election for 2016 was obtained in a csv format: <https://data.world/kidpixo/election-usa-2016-results>
* [www.pewforum.com](http://www.pewforum.com) 🡪 Data related to voter distribution amongst Democrats, Republicans and Independents across states of United States: <http://www.pewforum.org/religious-landscape-study/compare/party-affiliation/by/state/>

# Data Cleanup

Data was cleaned up in some of the following aspects:

* Removing information that wouldn’t add value to the analysis
* Removing incomplete data set information
* Removing blank/NaN field values
* Removing redundant/duplicate columns
* Renaming columns to more relevant names
* Signature content was removed

# Data Analysis

Data was analyzed in following ways:

* Identifying the ways of combining the data to have relevant information – looking for keys
* Joining the data to prevent loss of data.
* Joins were applied to combine the data based on key values
* Columns were reordered to make data intuitive
* For working with sqlite, SQLAlchmey was utilized along with DB Browser tool to view data in a GUI.
* All manipulations were performed in Jupyter Notebook after extracting the datasets.

# Output

After analysis and cleaning of data, a non-relational file election\_data.csv was obtained which can be manipulated in numerous ways to present different aspects of the data.