ETL Project  
August 2019

Team Members: Hiral Patel, Olufunke Olaleye, Waleed Hanna

* **Project Background and Description**

Annual salary information including gross pay and overtime pay for all active, permanent employees of Montgomery County, MD paid in calendar year 2014. This information will be published annually, by the end of January each year.

* **Extract:**

Data was loaded into a Jupyter Notebook for analysis and transformation.

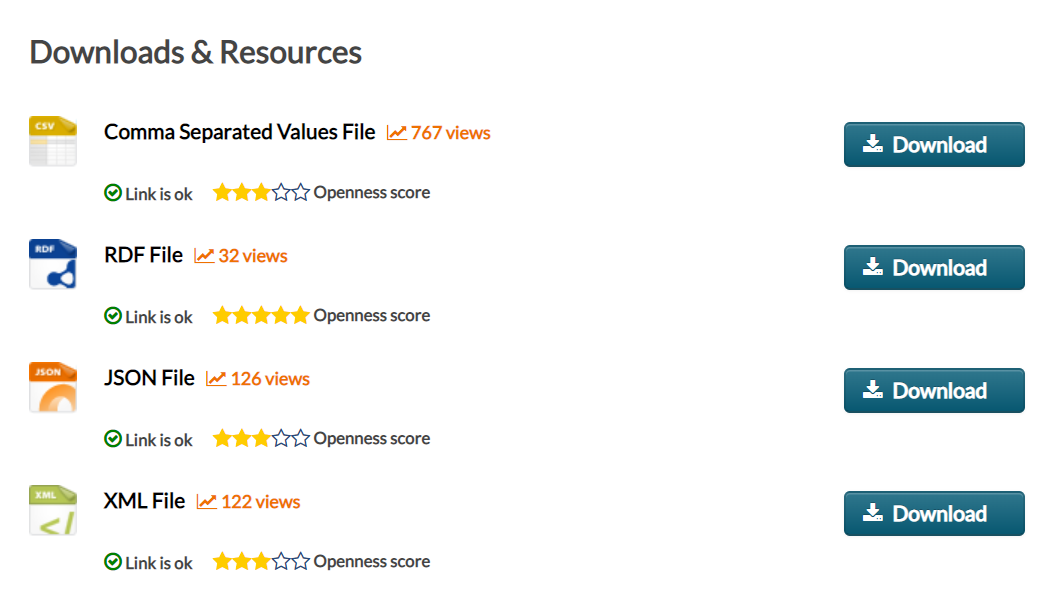
# Data Source: Data.Gov

Data about Montgomery County of Maryland police depoartment.

# Employee Salaries – 2014 & Employee Salaries – 2015

<https://catalog.data.gov/dataset/employee-salaries-2014>

**Format:** CSV, json, xml



* **Transform.**

Before transform the data into Postgres, we did it cleaning, joining, filtering, merging, renaming, aggregating and grouping. This process was as follows

1. Import all the dependences and libraries needed for the process such as Pandas, Json, PPrint, Sqlchemy, numpy.
2. Made sure we can read the data files
3. Created a new data frame and drop all the columns that are not needed for the analysis process.
4. Cleaned the data and remove all the duplicates and Null values.
5. Store CSV data into a new data frame that contains selected columns.

### Store JSON data into a variable.

1. Created a new data frame and drop all the columns that are not needed for the analysis process.
2. Cleaned the data and remove all the duplicates and Null values.

### Store JSON data into a data frame.

### Created an engine.

### Connect to local database in PgAdmin4.

### Use pandas to load csv converted data frame into database.

1. Confirmed the data has been loaded to local database by querying the table in Jupyter.

**Conclusion**

We have done the type of transformation needed for this data (cleaning, joining, filtering, aggregating). The type of final production database to load the data into (relational or non-relational). The final tables or collections that will be used in the production database.