**Group Projet 2 ETL**

We extracted NYPD “arrest data by bias motivation for 2017 1st quarter to 2019 2nd quarter” with A Socrata API.:

Data was from a NYPD website <https://www1.nyc.gov/site/nypd/stats/reports-analysis/hate-crimes.page>

We limited the result to get 700000 rows in a descending order, in order to get all of 2018 and 2017 data.

Data was divided into groups: (Racial/Ethnicity/Ancestry Bias, Religious Bias, Sexual Orientation Bias, Gender Identity Bias)

NYPD doesn’t divide bias motivation so we did it and added the corresponding col’s. We looked closely at hate crime by religious bias motivation - Census data doesn’t provide info on religion so we counted the places of worship (synagogues and mosques) in areas where hate crimes were committed to estimates religious population “using google places”. We created 2 new tables that connect places of worship to precinct areas using precinct id as a foreign key.

We connected the number of hate crime with the minority population in neighborhoods using census data (2013) to get population by zip code.

We also extracted each precinct zip code from its full address.

Each csv files was transformed and copied to required elements.

We created a DataFrame. Created a filtered df from specific col’s, Renamed col’s for SQL, Change the date format from object into datetime64

Renamed col headers on all csv to "ID" as primary keys for SQL table join

Cleaned the data by dropping duplicates and setting the index

Create database connections

Loaded df’s into database

Joined tables in ProgreSQL