**ETL Project Technical Report**

We did our project on Chicago crime data that analyzed 2018 / 2019 Chicago crimes and the police stations that they corresponded too. We started by extracting our data via the Chicago data portal, transforming the data in pandas, and loading dataframes into SQL database.

**Extract**

The data that we used was in the .csv format from the Chicago data portal. We looked for relevant data that we found interesting.

**Transformed**

Our data set that we got from the City of Chicago was still incomplete for our project needs. Our plan was to transform the data using the Pandas library in Python. The first thing that we did was concatenate the 2019 and 2018 crime data. After doing this, we wanted to make sure that there were no fields that had no numbers in it. To solve for this, we did a drop.na on the data frame. After that, we made sure to create a new data frame that only had the columns that we desired – ‘ID', 'Case Number', 'Date', 'IUCR','Description','Community Area'.

In order to join our two data frames, we had to clean up the IUCR data set to include only ‘DISTRICT’, ‘DISTRICT NAME’,’ADDRESS’, ‘WEBSITE’ columns.

**Load**

All the data was uploaded into a SQL database using PgAdmin – PostgreSQL. We decided to load in two separate tables, one with the crimes and one with police stations so users can load in the data that they want to analyze. We did this in Jupiter Notebooks using the create\_engine function in SQLALCHEMY.

**Next Steps**

We hope that we can continue to grow upon this data set to further analyze Chicago crime data. With this, the city can start making better and more analytical decisions so that can live in a safer city