# Project Proposal Document: Water Waste Management for New York

Team: Jean Pino

Nidhi Patel

Proposal Date:11/2/19  
Project Due: 11/7/19

**Statement of work:** Successfully Extract, Transform and Load Water waste management data for the state of New York and Load into a database. This will help to analyze the location(city/state), type of treatment and usage of waterbodies, Major drainage basin state pollution discharge elimination system, contact details of employees, year built and updated of treatment plant and sludge treatment.

**Data sources:** Water Waste management data (Treatments Used) and Facility information from Kaggle.com and Data.gov

**Proposed ETL:**

**Extracted** the waste management data from the mentioned sources.

**Transformed** the data using a python notebook (converted and cleaned the files using Pandas data frame). Created tables for different facilities and different locations from water waste treatment csv file. From municipal waste water treatment plant csv file tables were created for detailed recycling of waste water and treatment information.

**Load** the data into a Postgress database by creating a connection from the python notebook. Join both tables using a common primary key. Perform some basic aggregation to analyze number of facilities per city/county, type of facilities by waste treatment and flow of wastewater.

**Proposed Final Schema:**

PostgreSQL

Tables: Facilities\_info= Containing facilities info, such as address, facility type, city, zip code

Location = city,lat,long,zip code

Water\_waste\_info = Containing facilities info such as treatment type, recycling type.

**Task done by:**

Nidhi - Collected the data from the mentioned sources.

Jean - Uploaded, cleaned and turned data into data frames and created the mentioned tables.

Jean and Nidhi - load the transformed data into the database.

**Approved by:** Satish Anthony

**Final Project Report (11/6/19 or 11/7/19)**

At about 8 PM, your team will submit a Final Report that describes the following:

* **E**xtract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).
* **T**ransform: what data cleaning or transformation was required.
* **L**oad: the final database, tables/collections, and why this was chosen.

Please upload the report to Github and submit a link to Bootcampspot.

Present 3-4 minutes on the project discussing some pain points and how did you resolve them. Only one student from each team should present. 1 min for any Q&A to the class.