**ETL Project**

**The libraries used:**

import pandas as pd

from sqlalchemy import create\_engine

import datetime as dt

import json

import requests

import time

import urllib

from api\_keys import weather\_api\_key

from citipy import citipy

import matplotlib.pyplot as plt

**Databases used:**

CSV, JSON, SQL (Postgrase)

**Steps followed:**

1. Load and read data from CSV file
2. Clean the data
3. Reformat the date and time
4. Plot chart for the best used stations
5. Plot chart with the comparison between two types of bike users: Subscribers (Use Annual Pass) and Customers (Use daily or 3 day pass)
6. Get data only for the 10:38 am, 21st February 2019
7. Load and read JSON file
8. Clean and retrieve necessary data from JSON file
9. Use Weather API to check the current Weather of the cities the where the bike stations are located (NYC and periphery)
10. Wrote queries to create DB and tables in SQL
11. Use codes to load the citibike\_2019 and citibike\_2020 data frames in postgres
12. Join the dataframes generated from CSV and JSON
13. Query and check if the databases are loaded properly

**Documents:**

1. Jupyter Notebook used for extracting, transforming and loading Data
2. SQL file for tables/dataframes
3. Project Summery Document
4. Pictures