# README

# Energy efficiency project API

## About

This repository contains code for accessing and analysing energy efficiency project data from the US government's API. The API offers a comprehensive look at the energy efficiency projects implemented in the state of New York, which depicts the energy savings in kWh and MMbtu units. It also represents the estimated annual savings in dollars.

To use this code, you will need to have access to the API provided by the US government. Once you have access to the API, you can use the code in this repository to access and analyse the data. The code is written in Python and is designed to be easily modifiable to suit your needs.

To get started, you will need to clone this repository to your local machine. Once you have done that, you can open the code in your favourite Python editor and begin exploring the pollution data. The code is well-commented and should be easy to follow.

API Used:

## https://data.ny.gov/api/views/4a2x-yp8g/rows.json?accessType=DOWNLOAD

## Pre-requisites

Visual Studio- 1.77.3

Python version- 3.11

MySql server in openstack- 8.0.32

PostgresSql server in openstack- 11.19

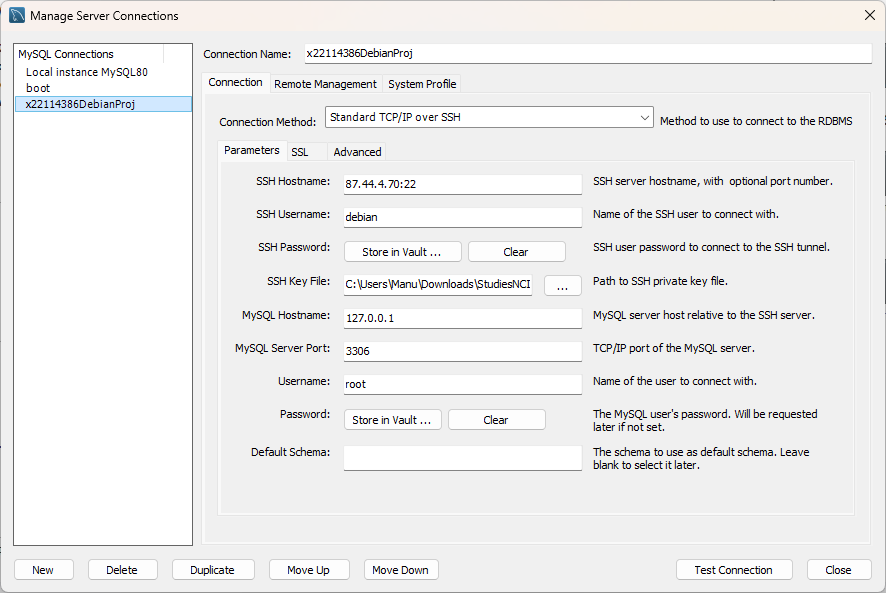
PEM file in the link – new.pem

PPK file in the link – new.ppk

Debian server in openstack- 10.0

Debian Instance- [x22114386Project](https://openstack.cloudenci.ie/project/instances/b22b1dab-19d0-494f-b5c0-9616dfb7ea1b/) - 87.44.4.70

* Mysql server details and database details:



* Postgres Server and database details:

user = "dap",

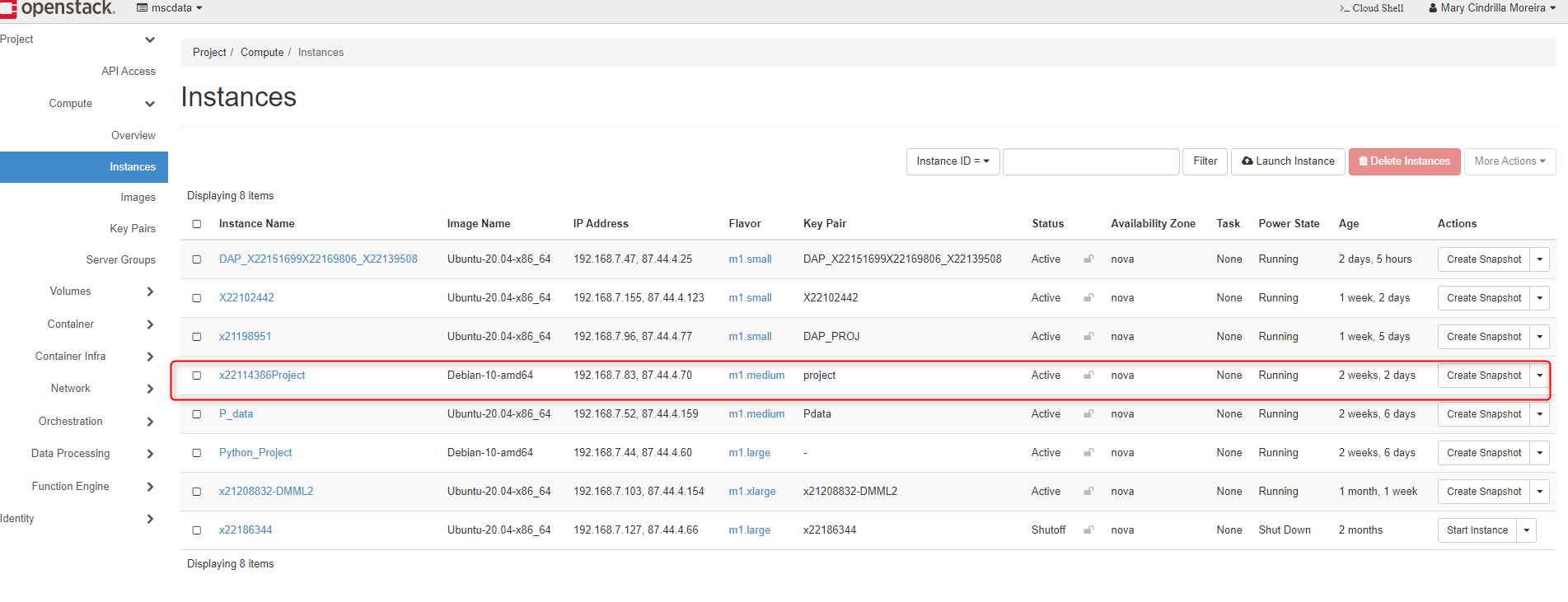
password = "dap",

host = "87.44.4.70",

port = "5432",

database = "postgres"

* Openstack cloud



More insights on the DB connections are attached in file “MySQL Server DATA BASE CONNECTION DETAILS.DOCX” in the file structure.

## Sequence of execution of Program

1. **1\_apiDataFetch. ipynb**

This is the first file to execute this file consist of

a. Triggering the open API.

b. Extracting the Json Output to a CSV file in the local path. The output file is RawData\_EnergyEfficiency.csv

1. **2\_DB\_MySQL. ipynb**

The code is related to Data base connection.

a. Connection string to create a connection pipeline to Database.

b. Creation of Database Schema

c. Dump the entire EnergyEfficiencyProject to the table.

d. Normalization of the data

e. Insertion of data with respect to table.

1. **x22105522\_DAP.ipynb**

This code gives the connectivity to PostgreSQL pipeline in case we are unable to create a pipeline in dagster just an intermediate file for reference.

a. Connection string to create a connection pipeline to Database.

b. Creation of Database Schema

c. Dump the entire EnergyEfficiencyProject to the table.

d. Normalization of the data

e. Insertion of data with respect to table.

1. **3\_DataCleaning.ipynb**

The CSV file is taken, and cleaning analysis of the file has been done.

Post this, all the 4 dataset is merged, analysed on a common ground.

## Other files in the folder structure:

1. RawData\_EnergyEfficiency.csv – Raw structured data
2. CleanedEnergyEfficiency.csv- CSV data
3. ShortDescriptionOnDataSets.txt- about the dataset
4. MySQL Server DATA BASE CONNECTION DETAILS- for the data base connectivity using ssh tunnel.
5. new.pem- private key for ssh tunnelling
6. pvtKey.ppk- for putty connection
7. 6\_MergedDatasets.ipynb – file for merged datasets and visualisation
8. x22105522\_DAP.ipynb – File which contain all the code.