March 7th 2020

Corey Davidson

Patrick Bourke

Sandy Lucev

**Data Analytics Boot Camp - University of Denver**

**Project 2 - E.T.L.**

**Project Requirements Overview:**

Several fire departments in the north Denver area are in the process of changing the way fire and emergency medical services units are dispatched. The basic premise of the project is to send the closest unit to true emergencies. This seems to be a simple concept but once you take jurisdictional boundaries and various 911 call centers into the mix the simplicity vanishes. Without delving too deeply into the details of the overall project one of the necessary steps is to aggregate response data from all agencies. This will enable us to perform various detailed analysis. This project is a realtime real world application of E.T.L.

Fire Project leaders are in the process of extracting, cleaning, and reorganizing the data from six fire departments. The project is reviewing five years of call data including call types, address, dates and geolocations. There are two main challenges in merging the datasets. One, each of the six fire departments use their own Records Management System (RMS). Two, each fire department has changed RMS vendors within the past five years.

**Step 1:**

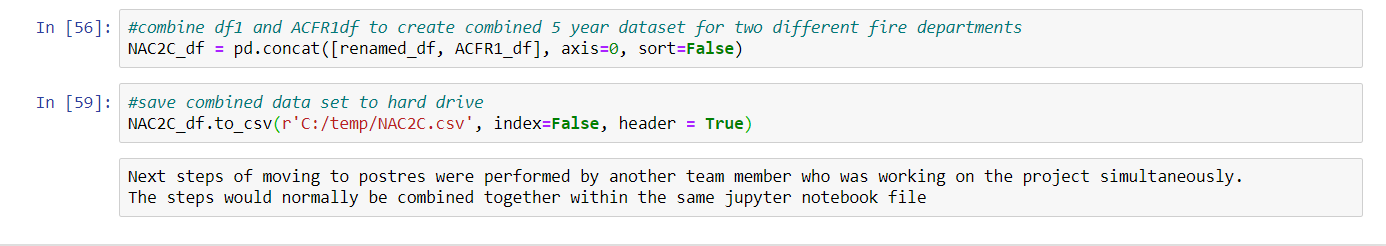
There were a total of three datasets used for the BootCamp ETL project. One fire department has already aggregated all of their data which will serve as the template for the data this team extracted, cleaned and organized. The template data set was provided in csv format. Data set number two is from Adams County Fire Rescue (ACFR). The current RMS used has a customizable report writer feature that allows the user to write and execute queries and reports using a graphical user interface. That report was then exported as a CSV file and covered response data from March of 2017 through the end of 2019. The third set, FH\_data, of data was pulled from a SQL server using SQL server management studio. The three data sets were ultimately merged together to create a single dataset.

The FH\_data set presented the biggest challenge. The SQL query selected the needed fields, concatenated a few fields, used RTRIM to trim blanks and also used four different joins. The design of the database itself created a significant challenge. The table identified as Lkp\_Inci contains 3219 rows with 128 different categories. Each category ranged from 4-58 items. The total incident count for this small date range was just over 15K. At one point the joins into the Lkp\_inci table created a result of over 2.3 million rows. Ultimately we required three different joins into the Lkp\_inci table to produce the desired results.

The results of the query were exported from the SQL management studio into a csv file. See the FHSQLQuery.txt file for the actual SQL query.

**Step 2: Transform the data in Python and put data into dataframes**

****

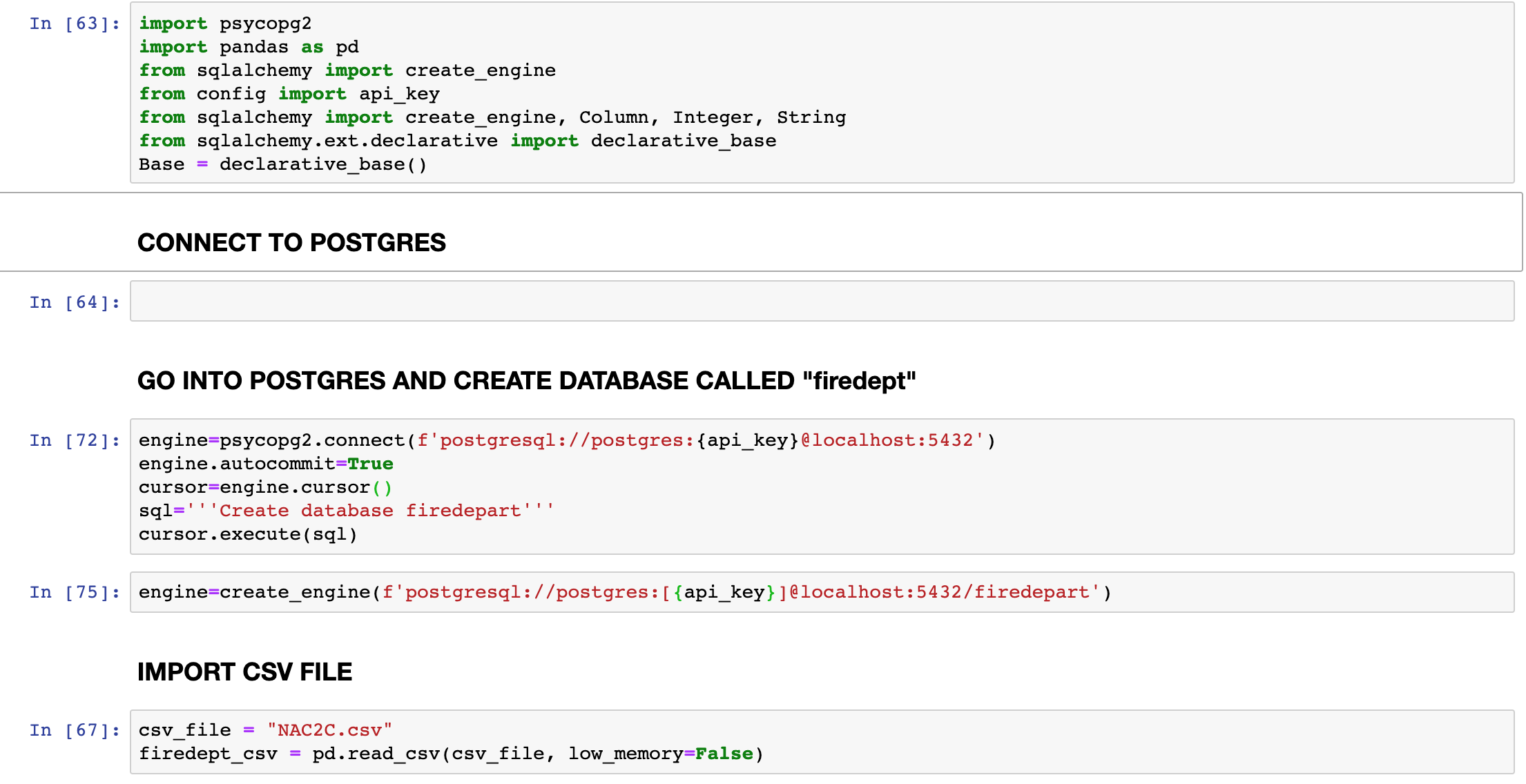
****

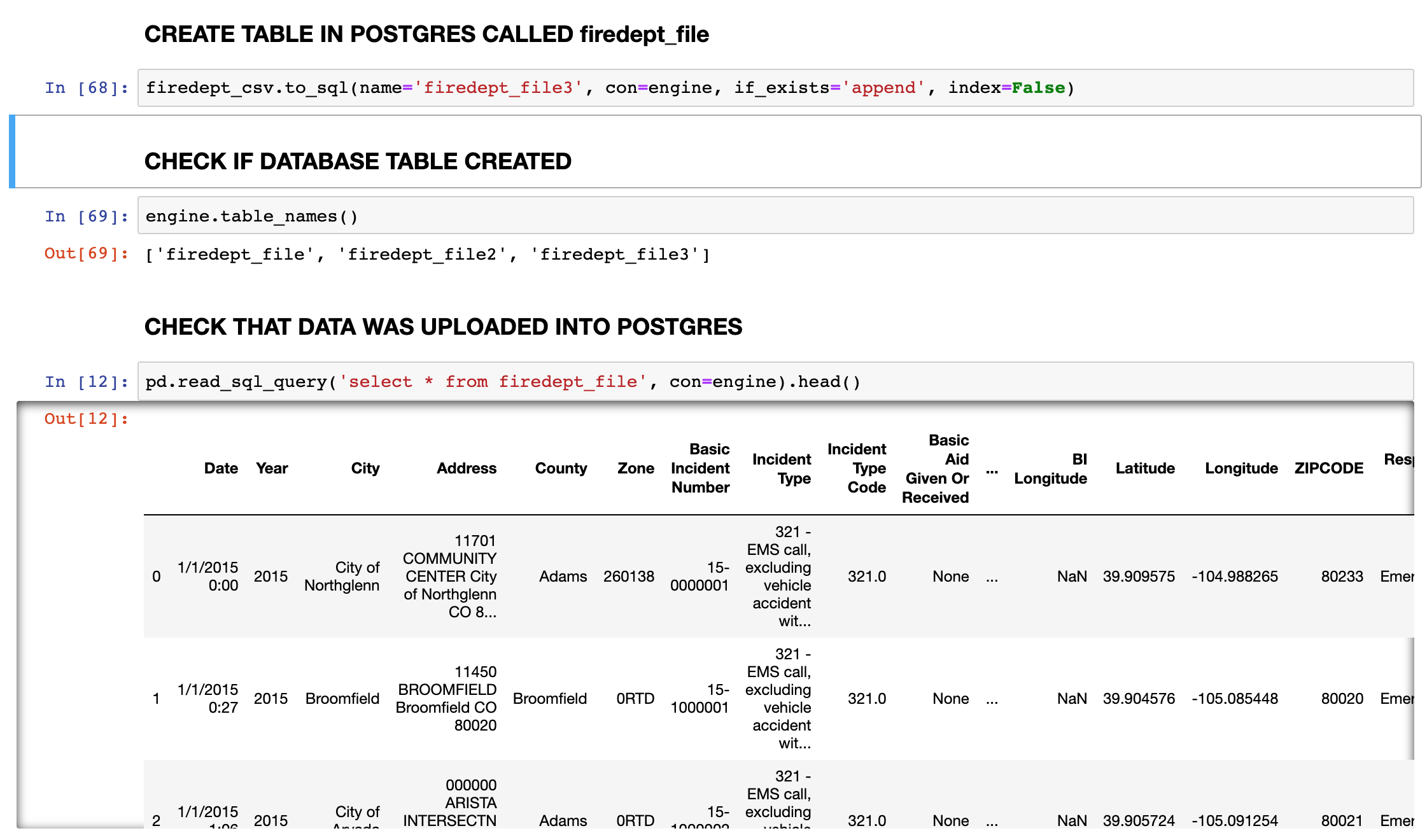
**Step 3: Import cleaned CSV data through Python into Postgres**

After the data has been cleaned through Python, we need to upload into Postgres.

We needed to import psycopg2 and connect to Postgres through Python. A config file was created to store the password for Postgres server.

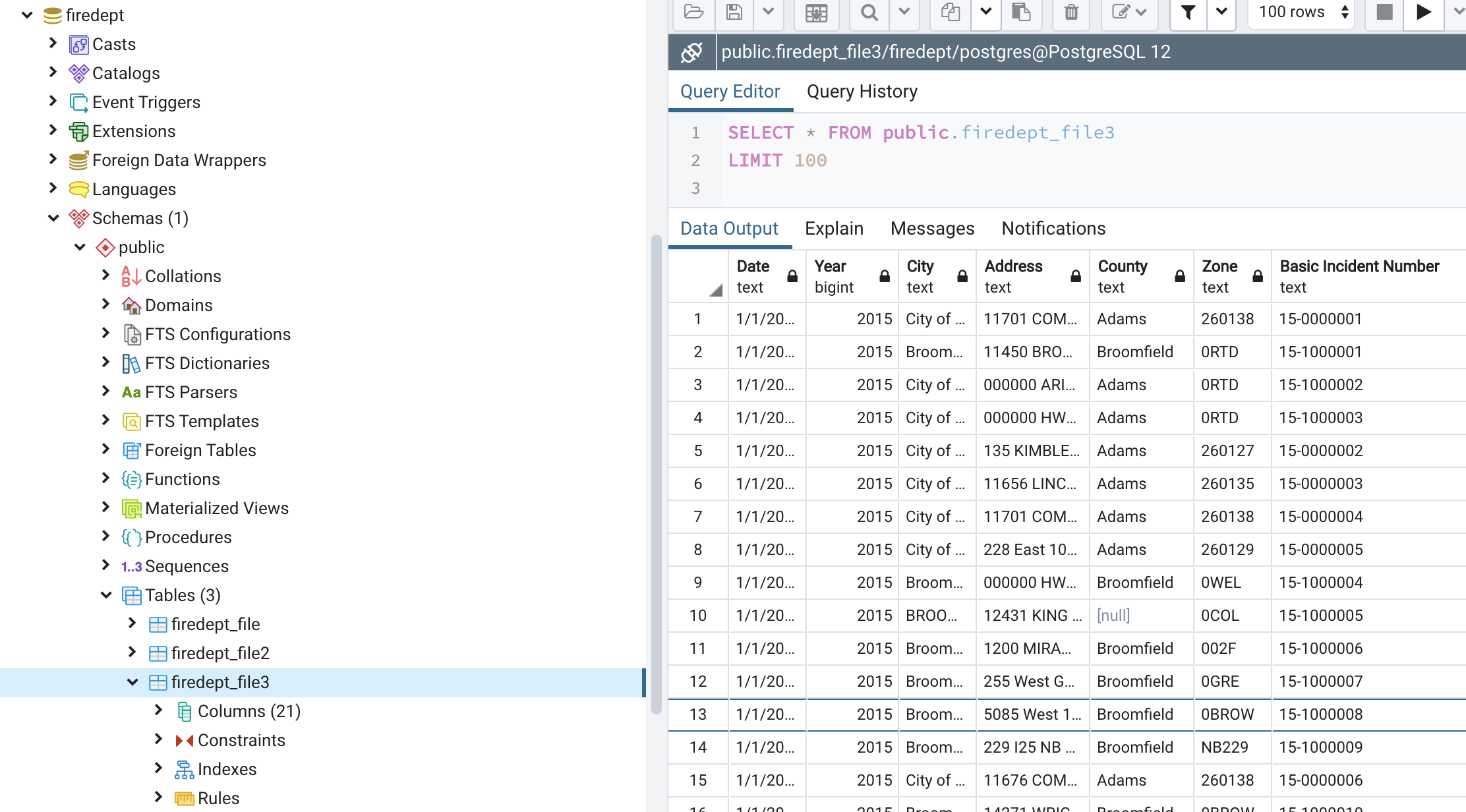
The steps below go through the import process:





**Step 4:**

We double checked Postges to ensure the database exists and looked at “First 100 rows” to ensure all the data uploaded.



**# ### CONNECT TO POSTGRES**