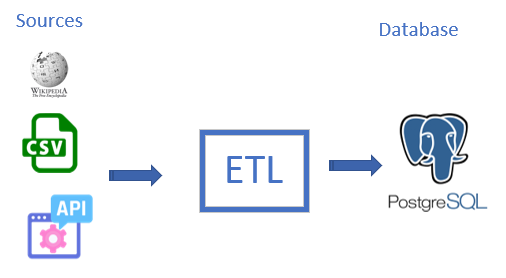
Project name: ETL for different sources and files

Project Proposal

This project will analyze crimes in cities in America that have almost the same Toronto’s population. In this case, we chose data set from Toronto, Vancouver and Chicago.

Workflow



**Extract**

Data set

Wikipedia:

<https://en.wikipedia.org/wiki/Demographics_of_Toronto#cite_note-2011censuspop-18>

CSV:

<https://data.cityofchicago.org/Public-Safety/Crimes-2001-to-present/ijzp-q8t2/data>

API return JSON:

http://manoelburgos.azurewebsites.net/api/Crimes

**Transform**

HTML



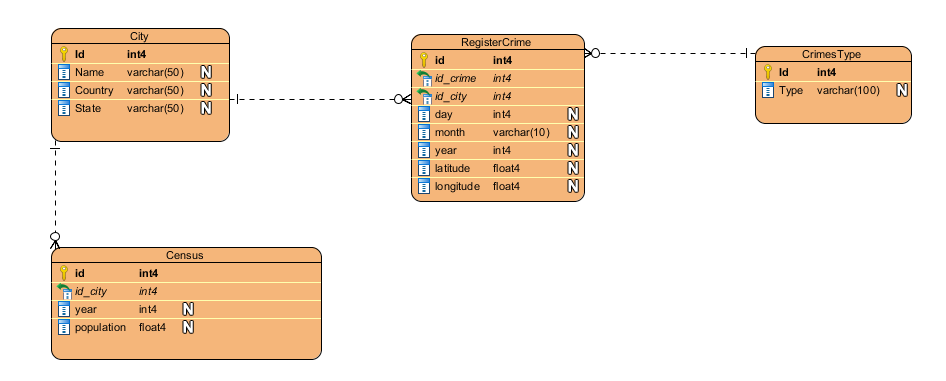
JSON



CSV



**Load**



**Results**

Sample quey:

select c.name,

c.country,

rg.year,

cr.type,

count(cr.type)

from city c

inner join registercrime rg

on c.id = rg.id\_city

inner join crimestype cr

on rg.id\_crime = cr.id

group by c.name,

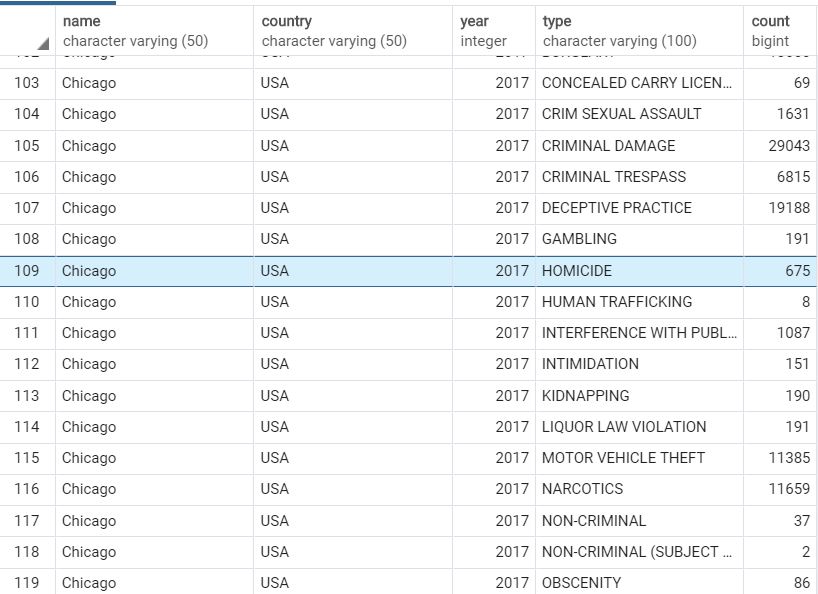
c.country,

rg.year,

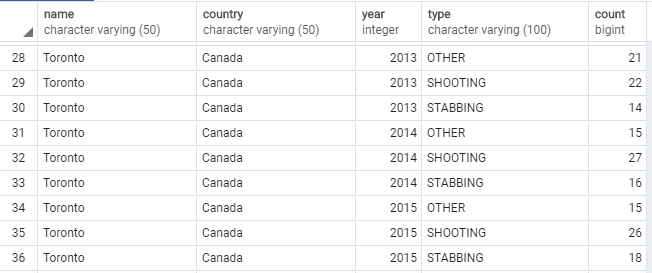
cr.type

order by rg.year

Chicago:



Toronto:



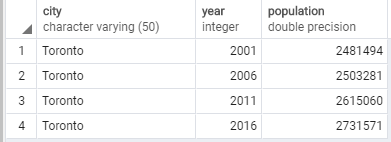
select (select name

from city where id = c.id\_city) as City,

c.year,

c.population

from census c;



Features:

PostgreSQL database

ERD designed in Visual Paradigm

Python with libraries:

Pandas,

Sqlalchemy

Json,

Requests

By:

Manoel,

Mir,

Banafshe