ETL Project

By:

Manoel,

Mir,

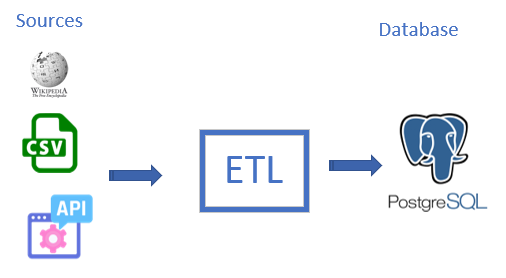
Banafshe

Project Proposal

Track crimes

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

Load

Results

Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

Transform: what data cleaning or transformation was required.

Load: the final database, tables/collections, and why this was chosen.

Queries

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

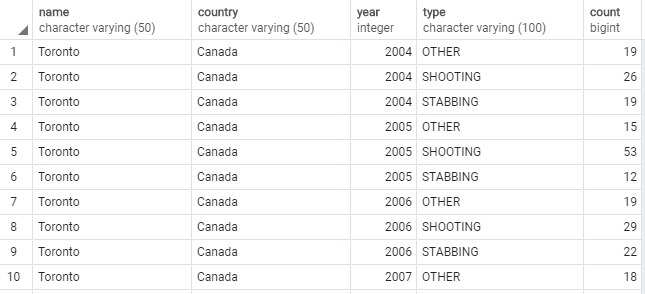
group by c.name,

c.country,

rg.year,

cr.type

order by rg.year



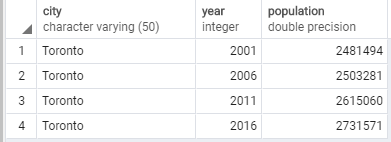
select (select name

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

c.year,

c.population

from census c;



select rg."year",

count(rg.year) as "Number of crimes" ,

population

from registercrime rg

left join census cen

on rg.id\_city = cen.id\_city

and rg.year = cen.year

group by rg.year,

cen.population

order by rg.year

