COMMUTE OR RELOCATE

COMPARING TWO CITIES RESIDENTIAL POTENTIAL

BACKGROUND

- Problem definition:
  - Pretoria and Johannesburg are close to one another, but have many preconceived ideas around the other city
  - The result is a reluctance to move from the one to the other causing massive commutes for people working in one and residing in the other
- Intended audience
  - Commuters between Pretoria and Johannesburg
  - Intercity job seekers
- Project aim
  - Give commuters an alternative to simply staying taking traffic head on, by giving them an indication to areas that confirm to their current residence environment

DATA ACQUISTION

- There were four sources used:
  - Municipal regions data (from most recent census)
  - Residential property sales data (average number of beds and price)
  - Surrounding facilities

here

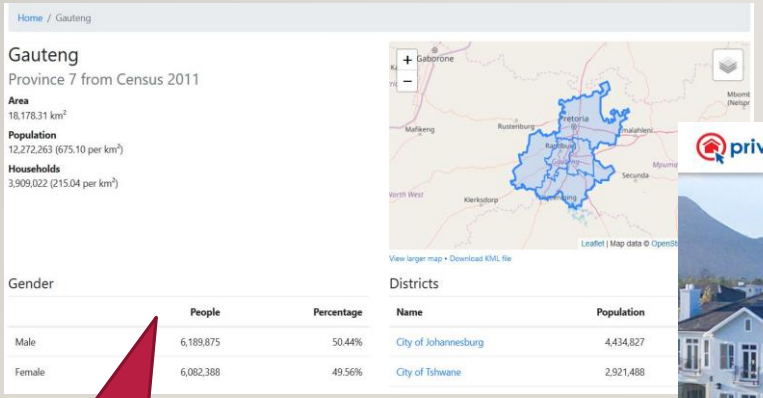
Geocoder API

FOURSQUARE

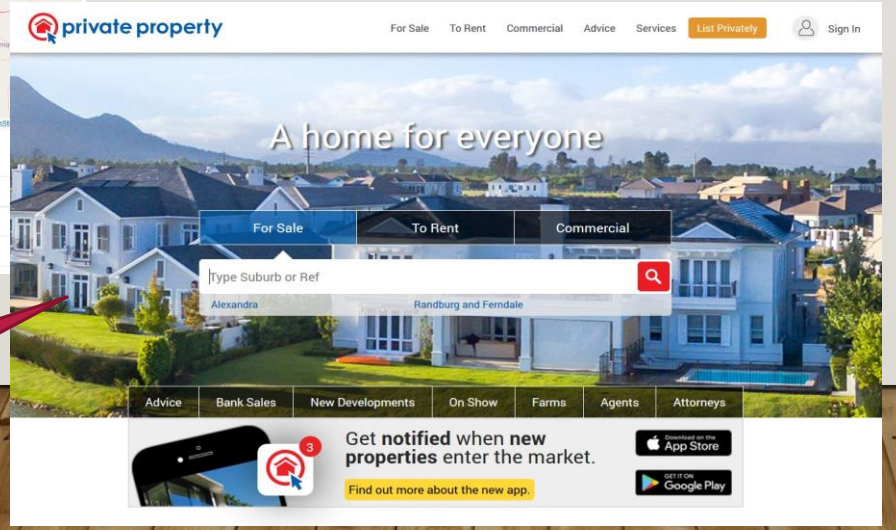
Places API

Power location-based experiences in your app or website.

Geocoding and venue info API's



Most recent census data



Public property listings

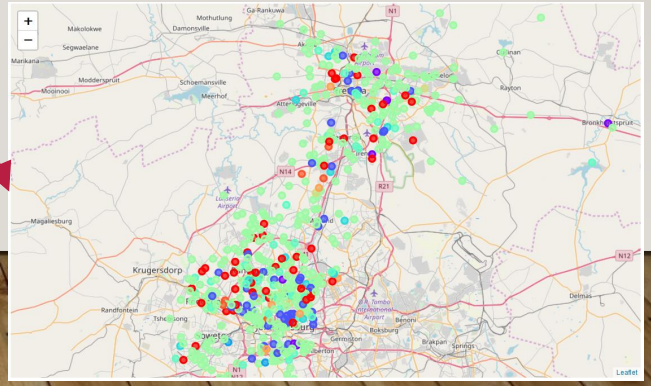
APPROACH

	region	city
84	Rietfontein	2
86	Roodepoort	2

	city	region	suburb	beds	price
0	City of Tshwane	Akasia	Amandasig	1.50	382500.00
1	City of Tshwane	Akasia	Chantelle	1.81	561600.00
2	City of Tshwane	Akasia	Clarina	1.00	332454.00
3	City of Tshwane	Akasia	Heatherdale AH	5.66	4309000.00
4	City of Tshwane	Akasia	Akasia	3.00	1031727.00

- The approach followed the process below:
  1. Scraping municipal regional data (duplications needed to be resolved) [creating initial base]
  2. Scraping public property sales listings (information on two different logical levels needed to be resolved, and missing data needed to be catered for) [shrinking initial base]
  3. Merging 1 and 2 into stable base
  4. Adding geocode data for use in step 5
  5. Enriching base data with surrounding venue info
  6. Pre-processing for use in clustering
     1. Normalising sales data
     2. Creating dummy data for categorical variables
  7. Performing clustering

1. Municipal regional data
2. Property sales data
3. Creating a stable base
4. Adding geocodes
5. Surrounding venue data
6. Cluster pre-process data
7. Clustered data set

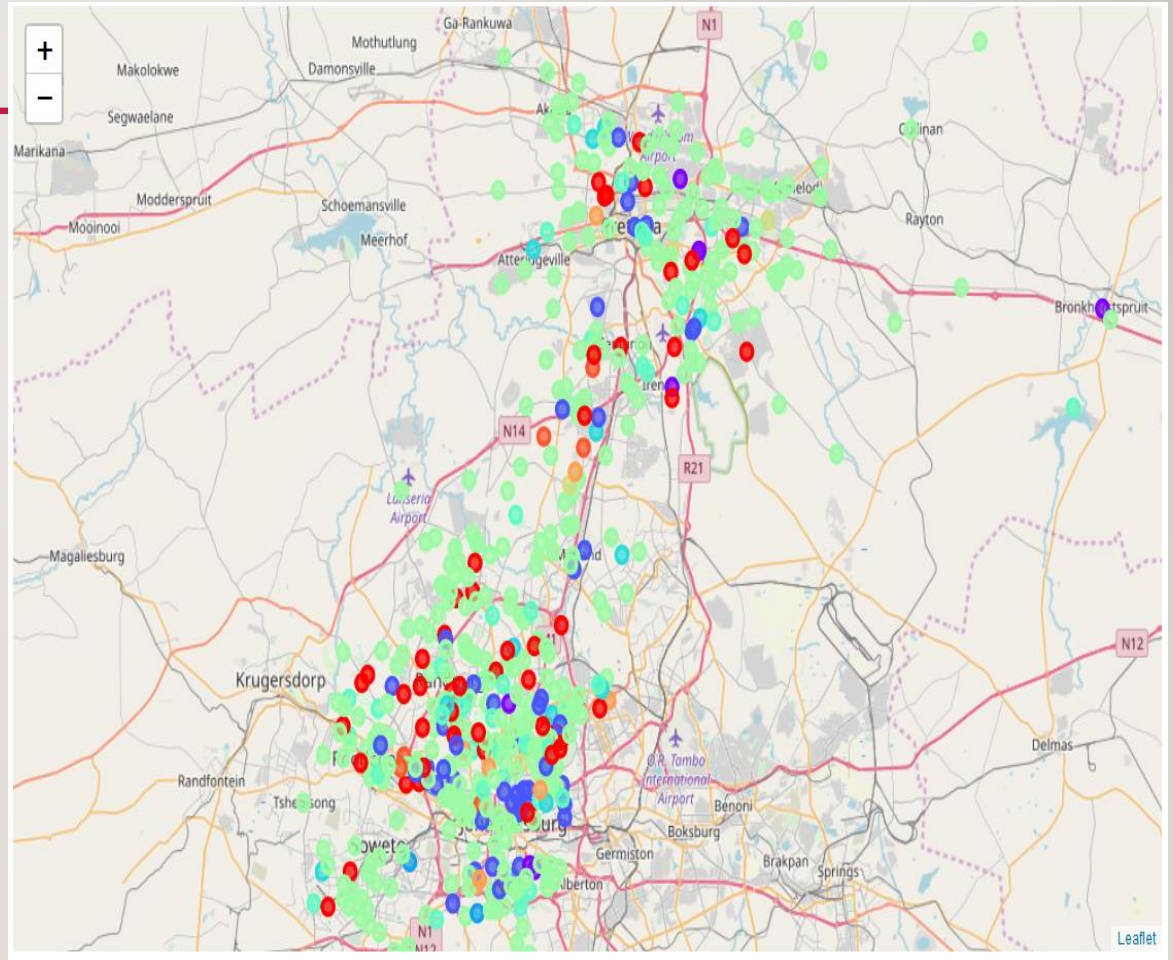


Technologies used:

- ***Web scraping*** – Selenium, BeautifulSoup, Requests
- ***Data processing*** – pandas, numpy, sklearn
- ***Data visualization*** – matplotlib, folium

FINDINGS

- There were three major findings from the analysis
  1. Strong similarity over the entire geographical area
  2. Johannesburg does seem to be a bit more metropolitan, in the sense that it carries more diverse clusters over a smaller geographical area
  3. Progressive communities have already moved to the centre between the two cities



CONCLUSION

- Considering to commute or relocate comes down to a choice of preference:
    1. Stable and homogenous – Pretoria
    2. Outgoing and divers – Johannesburg

- PS. In the code pack provided, there is a tool to give feedback to which cluster your area belongs