Coursera
Capstone

Othmane Laoufir

Assignment

Clearly define a problem or an idea of your choice, where you would need to leverage the Foursquare location data to solve or execute.

Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

Business Problem

The project owner wants to offer housing rental to international students in Montreal.

International students pay well, and have to prepare their arrival from abroad.

The project owner expect his clients to seek for tenements:

- At a fair distance from their university

- Offering a proper level of indoor comfort

- Close enough to appealing venues and useful services for their daily life

Where does the project owner need to invest in the city to satisfy his clients and optimize his return on investment?

Which neighborhood is the most suitable?

The Data

Geographical information

- The scope of the study limits the research to Montreal, QC.
- Top Montreal universities are easy to find
- The Foursquare API enables geocoding

Lifestyle and comfort information

- Services, facilities and other venues must be qualified and categorized
- We will admit a range of 500m around each venue to collect information

Financial aspects

- Gathering information about property valuation would be more than useful
- Unfortunately most of the available sources are proprietary e.g paid ones.
- Valuation information will be eventually appended to the project if possible

Methodology

In a nutshell:

1. Get the top Montreal universities location
2. Collect rental ads nearby Montreal
3. Collect venues information nearby rental locations
4. Compute the average distance from universities
5. Aggregate venues by category
6. Score city spots by neighborhood

Data Collection :

- Addresses registry:
  - Web search for top universities
  - Web scraping from zumper.com
- Geocoding from addresses:
  - Universities
  - Rental ads
- Foursquare Venues exploration:
  - From geocoded addresses

Tools and libraries:

– Google Search

– BeautifulSoup

– Nokia HERE API

– Foursquare API

Resulting Datasets

Rental offers in Montreal

- – Neighborhood
- – Address
- – Geocode
- – Average distance from universities
- – Price

Venues in Montreal

- – Neighborhood
- – Address
- – Geocode
- – Category

About Venues Dataset

306 initial categories merged into 8:

– **Casual Food:** Whatever nonessential food

– **Entertainment:** Nightclubs, Bars, Shows an so on.

– **Essentials:** Supermarkets, Banks, and daily services

– **Leisure:** Parks, Cultural places, and recreational activities

– **Retail:** Shops, Boutiques, and so on

– **Sport:** Sport infrastructures, clubs and facilities

– **Transportation:** Public transportation services

– **Other:** Everything that does not fit any of the above

Data Analysis

- Learning Method: Unsupervised
- Algorithm: K-Means Clustering
- Justification:

Supervised learning would have output poor results regarding to variable dependence. Rental offers do not target only students. Thus addressing the business problem with a biased control variable would have ended to biased results.

Datasets usage:

- Venues in Montreal is used to perform neighborhoods clustering.

- Rental Offers in Montreal is used to control the output results and deem whether unsupervised classification is plausible and reliable or not.

Results Section

Data computation with 4 clusters:

- Has successfully excluded outliers in clusters 1 and 4:

  - Cluster 1 is far from universities and offers very few nearby activities

  - Cluster 4 is far from universities and offers only irrelevant activities ("Other" venues category)

14

Data computation with 4 clusters:

- Has successfully identified clusters 2 and 3 as attractive
  - Cluster 2 offers balanced activities and is still at an acceptable distance from universities
  - Cluster 3 is the closest to universities with a quite high proportion of "Essentials" and "Leisure" activities nearby

Data computation with 4 clusters:

- Did not input:
  - Average distance from universities
  - Prices
  - Number of rental offers
- And still has matched the 3 above features

Discussion

- Model Limitations:
  - More than 90% of the rental offers dataset falls into the cluster 3
  - In-cluster distance among cluster 3 is quite high
  - 2 by 2 correlations are weak within the cluster
- Corrective:
  - Perform a supervised learning approach restricted to cluster 3

Conclusion

- Historical neighborhoods in Montreal offer a wide range of activities. Top Montreal universities are also located overthere.
- Thus it is the most attractive area to launch a rental business addressed to international students
- Peripheral areas must be avoided since it rather targets families than individual households
- Ville-Marie is the most attractive and the most expansive neighborhood in town.
- Mercier-Hochelaga, though less attractive turns out to be at a fair distance from universities, with an acceptable level of facilities available. It is also the cheapest district in the third cluster which makes it a good choice.