**IBM COURSE**

**APPLIED DATA SCIENCE CAPSTONE**

Peer-graded Assignment: Capstone Project - The Battle of Neighborhoods (Week 5)

SUMMARY

1. INTRODUCTION
2. DATA DESCRIPTION
3. METHODOLOGY
4. RESULTS
5. DISCUSSION
6. CONCLUSION
7. **INTRODUCTION**

*Question: Introduction where you discuss the business problem and who would be interested in this project.*

The problem that want to be addressed with such work is to measure the attractiveness of the City of Melbourn, Australia in terms of facilities (parks, cafes,…), in particular in the Northern part of the city.

Such work could be in the interest of those who are looking for an apartment in such area, to understand whether it satisfy their needs.

1. **DATA DESCRIPTION**

*Question: Data where you describe the data that will be used to solve the problem and the source of the data.*

The data that are going to be used are mainly the postcodes and the median house sales price in the state of Victoria (VIC), Australia. Such data are already well structured and clean, this is the reason why I have choosen to perform this work on such area.

Links:

<https://www.matthewproctor.com/full_australian_postcodes_vic>

<https://www.data.vic.gov.au>

1. **METHODOLOGY**

*Question: Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.*

To measure the attractiveness of the Northern part of Melbourne have been analysed different attribute of the Suburb: number of cafes, parks, avg price, distance.

In order to do this the following step have been done:

1. Serach of the abovementioned data from which start the analysis
2. Ensure that data where enough and consistent to allow the analysis
3. Clean the data and structure (fliter and sort) in order to allow to be used for the analysis on the basis of the abovementioned attribute (cafes and parks)
4. Scrape the data in order to analyse the location. This has been done thanks to **Foursquare API**
5. Group the data. This has been done thanks to **k-means**
6. Analysis of the identified groups in terms of house price and distance.
7. Identify the best group
8. **RESULTS**

*Question: Results section where you discuss the results.*

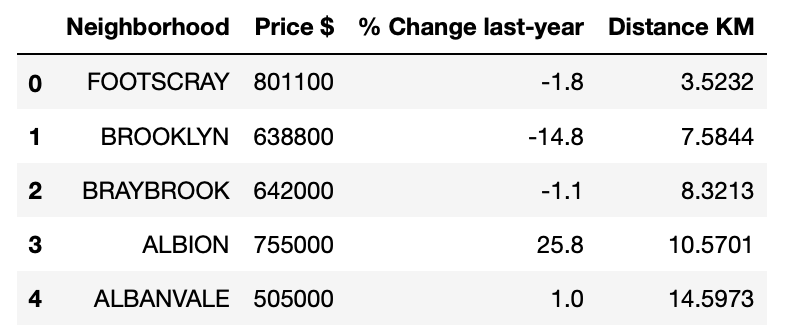
The main output of the analysis have been the following:

1. Identification of suburbs whith the attributes

Immagine che contiene mappa

Descrizione generata automaticamente

1. Identified the suburbs analysed the price and the distance



1. DISCUSSION

*Question: Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.*

The most difficult part of the work has been identify the data on which perform the analysis. Unfortunately, it is difficult to find such structured data for smaller cities. So, until when all city will have a good database on which base the analysis, such works could be performed on a really limited set of cases. Overall, the analysis performed has showed up good results, maybe setting more attribute could have been possible to come up with a even shorter set of final suburbs.

1. CONCLUSIONS

*Question: Conclusion section where you conclude the report.*

Starting from a subset of already identified attributes and needs has been possibile to identify a set of suburbs which mostrly fit such data. Probably identifying more attribute could bringh more precise results.