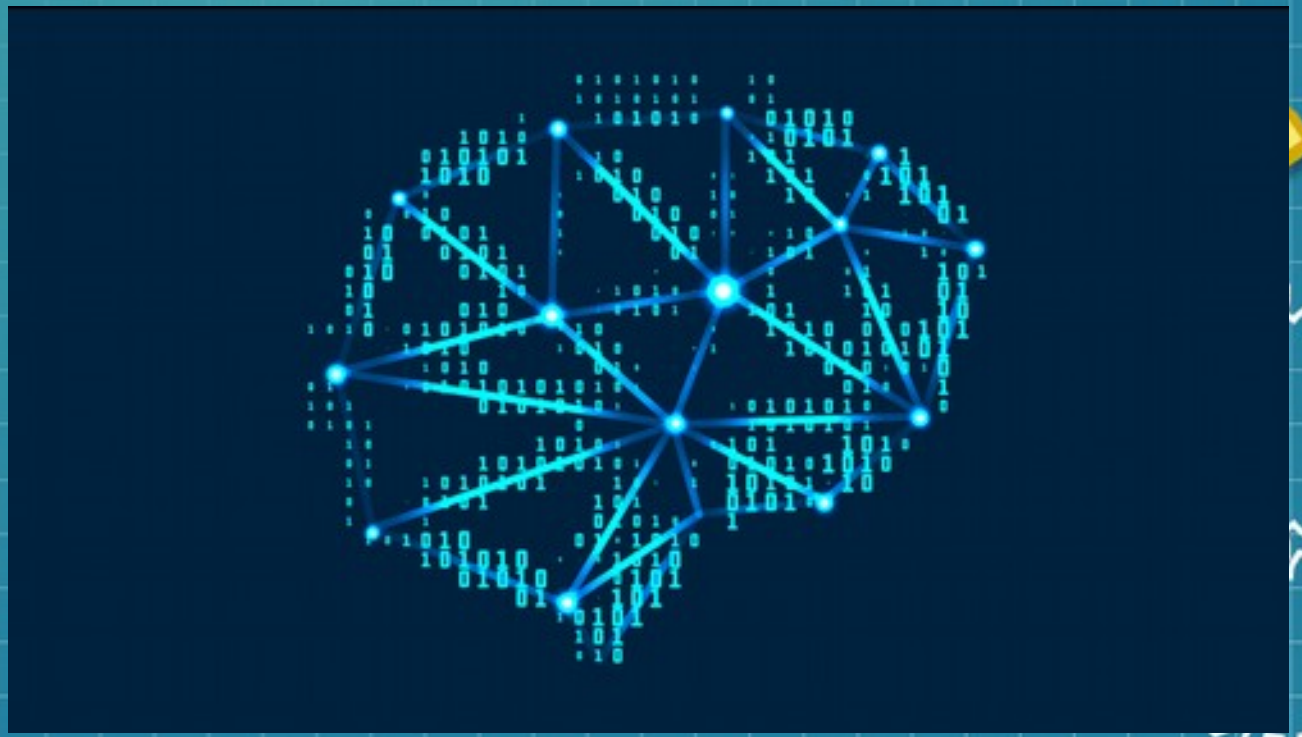
Data Science Capstone final Assignment

Author: M. Giaconi



Background

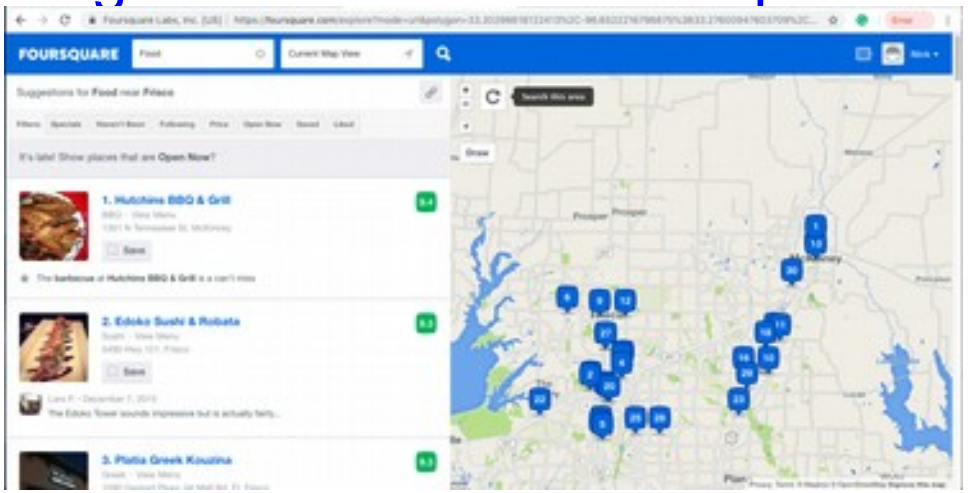
- The final assignment has been based on the various techniques learnt and acquired during the 9 modules of the IBM data science professional certification, offered by Coursera

Application

- The project work is aimed to an entrepreneur (or a group of them) that would like to make an investment related to Italian cuisine in the city of Toronto.

Data

- Scrapped from Wikipedia page related to Toronto postal codes (https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M)
- Used Foursquare API to get information related to the public activities located in the surroundings of every neighborhood
- Used the Lat and Long of every neighborhood which were provided by Coursera through a CSV file

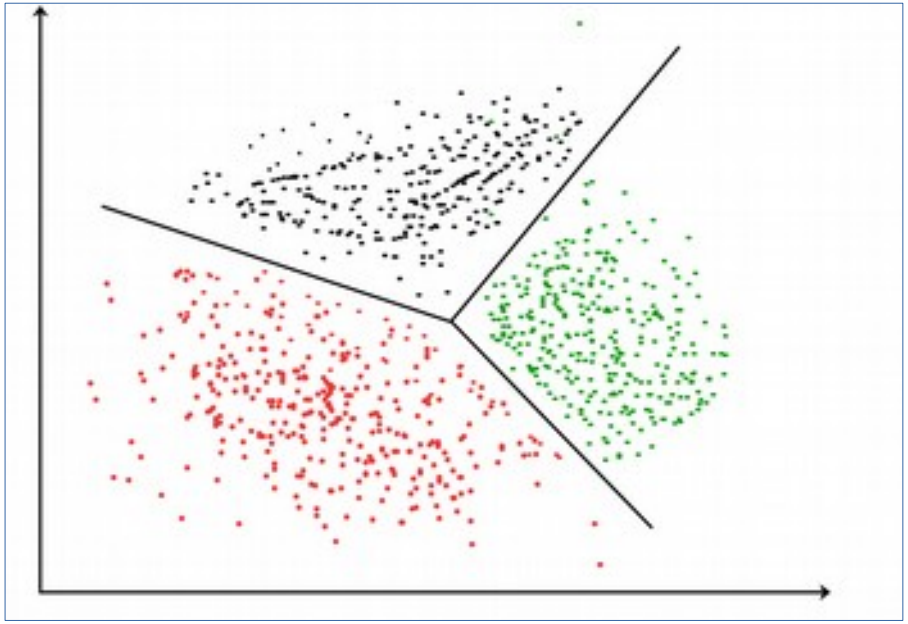


Methods

- Used Python language as main vehicle to build data analysis algorithms
- The code has been built on Jupyter Notebooks due to their high flexibility
- Libraries include: Folium, Pandas, Numpy, Matplotlib, Kmeans

Clusters

- Kmeans clustering have been used to process the various data
- Indication related to the presence of Italian Restaurants and Pizza places have been used to fit the model
- 3 Clusters have been built



Results

- Results are shown in the picture below.

  The clusters have been colour coded
- The final graph clearly shows the separation between the areas nearer t

Conclusion and further steps

- The analysis highlights the difference between neighbourhood in terms of business potentiality.
- It would be recommended to integrate the analysis with some data about land usage and costs to make it more reliable and explicative of the possible advantageous areas to build on.