**"A Chef of her Own Kitchen"​: Determining the Best Place for a Pizza Place in the City of Rome by Using Data Science!@Zafer Uzun**

**Introduction**

* One of the important points for establishing a new business is the location of the workplace.For example, a crowded population can provide the best chance experiences such as social gatherings, entertainment, performances, festivals, and tourists to earn money. And all the things provided can help your business. So we have to chose the right place.

**Businees Problem**

* The objective of this capstone project is to analyze and select the best locations in the city of Rome, Italy to open a new **pizza place**. Using data science methodology and machine learning techniques like clustering, this project aims to provide solutions to answer the business question: In the city of Rome , Italy, if a property developer is looking to open a new pizza place, where would you recommend that they open it?

to be countinue :)

**Data**

* In this section we will use Wikipedia page "[https://en.wikipedia.org/wiki/Category:Subdivisions\_of\_Rome"](https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome%22), contains a list of neighborhoods in Rome from Italy
* And we will use the Latitude and longitude coordinates of those neighborhoods to explore the city.
* We will use Python Geocoder package to to examine the coordinates of the neighborhoods in the city.
* Venue data, particularly data related to restaurants.
* To get the venue data for those neighborhoods we will use Foursquare API ( <https://foursquare.com/> com/)

**Methodology**

**At this stage, to focus on the research problem we use a holistic analysis approach.**

* The first step is preparing the data:
  + to get the list of neighborhoods in the city of Rome, available on the Wikipedia page "[https://en.wikipedia.org/wiki/Category:Subdivisions\_of\_Rome"](https://en.wikipedia.org/wiki/Category:Subdivisions_of_Rome%22).
  + to web scraping using Python requests and "beautifulsoup" packages to extract the list of neighborhoods.
  + To use the Geocoder package that will allow us to convert address into geographical coordinates in the form of latitude and longitude.
* The second step is checking the data:
  + to make sure we are working on the right form of the data, we will use one of the geographical visualization library "Folium package" and visualization library Pandas DataFrame.
* The third step is analyzing the data:
  + Firstly, we will convert the data into a pandas DataFrame and then visualize the neighborhoods in a map using the Folium package.
  + Next, we will use Foursquare API to get the top 100 venues that are within a radius of 2000 meters. To use Foursquare API we will use our Foursquare ID and Foursquare secret key.
  + Then, we will focus on the mean of the frequency of occurrence of each venue category to analyze each neighborhood by grouping the rows by neighborhood.
  + To solve our Bussiness Problem, we will filter the “Pizza Place” as venue category on the neighborhoods.
  + Lastly, to cluster the neighborhoods into 3 clusters based on their frequency of occurrence for “Pizza Place”, we will cluster on the data by using k-means clustering.
  + By making a selection among the clustered neighborhood groups we have acquired, we will provide the process of making the optimum decision by using data science support.

**Results**

* We categorize the neighborhoods of Roma into 3 clusters based on the frequency of the occurence for "Pizza Place (Restaurants)".
  + Cluster 0: Neigbordhoods with moderate number of Pizza Place.
  + Cluster 1: Neigbordhoods with low number of Pizza Place.
  + Cluster 2: Neigbordhoods with high number of Pizza Place.

**Discussion**

* The highest number of Pizza Place in cluster 2 and moderate number of Pizza Placein cluster 0.
* On the other hand, cluster 1 has very low number to no Pizza Place in the neighborhoods.
* As a recommendation:
  + **The neighborhoods in cluster 1** are the most preferred locations to open a new Pizza Place.

**Conclusion**

* In this project we use the power of the data science, to make a desicion and chose the right place for a Pizza Place.
* To answer the business question :
  + The neighbourhoods in cluster 1 are the most preferred locations to open a Pizza Place.
  + The findings of this project will help the relevant stakeholders to capitalize on the opportunities in high potential locations while avoiding overcrowded areas, are high competition, in their decisions to set up an entrepreneurial business.