Introduction

Moscow is the capital and most populous city of Russia, with 13.2 million residents within the city limits and 17 million within the urban area. Moscow is one of Russia's federal cities. Moscow is a major political, economic, cultural, and scientific center of Russia and Eastern Europe, as well as the largest city (both by population and by area) entirely on the European continent. By broader definitions Moscow is among the world's largest cities, being the 14th largest metro area, the 18th largest agglomeration, the 14th largest urban area, and the 11th largest by population within city limits worldwide. According to Forbes 2013, Moscow has been ranked as the ninth most expensive city in the world by Mercer and has one of the world's largest urban economies, being ranked as an alpha global city according to the Globalization and World Cities Research Network, and is also one of the fastest growing tourist destinations in the world according to the MasterCard Global Destination Cities Index.

The city is served by a transit network, which includes four international airports, nine railway terminals, numerous trams, a monorail system and one of the deepest underground rapid transit systems in the world, the Moscow Metro, the fourth-largest in the world and largest outside Asia in terms of passenger numbers, and the busiest in Europe. It is recognized as one of the city's landmarks due to the rich architecture of its 200 stations. [1]

In this project, we will study in detail the area classification using Foursquare data and machine learning segmentation and clustering. The aim of this project is to segment areas Moscow metro station surrounding areas based on the most common places captured from Foursquare. Using segmentation and

clustering, we hope we can determine classification of area located inside the city whether it is residential, tourism places, or others.

Data

The data used for this project will be acquired from wikipedia https://en.wikipedia.org/wiki/List_of_Moscow_Metro_stations . The dataset consist of station names and coordinates. In addition, the Foursquare API search will be used to extract venue information.

Details about local venues and locality will provide insight into the qualities of a metro station surrounding area. In addition to Foursquare, various python packages will be used to create maps and machine learning models to gather further insights and provide efficient recommendations and results. This packages include:

- Pandas   -   Library for Data Analysis
- NumPy   –   Library to handle data in a vectorized manner
- Geopy   –   To retrieve Location Data
- Requests   –   Library to handle http requests
- BeautifulSoup   – Library for pulling data out of HTML
- Matplotlib   –   Python Plotting Module
- Sklearn   –   Python machine learning Library
- Folium   –   Map rendering Library