

COURSERA IBM DATA SCIENCE CAPSTONE ASSIGNMENT

WHICH NEIGHBORHOODS OF TORONTO LACK SHOPPING MALLS

By Viatcheslav Liachenko

SECTION 2 – Data collection and preprocessing

For the final capstone project assignment, we continue to build upon the data obtained during the third week of the studies.

The necessity of data for the Toronto's neighborhoods was solved by utilization of Wikipedia's source of Toronto, CA neighborhoods, which was scrapped into the notebook's Pandas dataframe using BeautifulSoup library. The geographical coordinates were imported from the available Geospatil_data file.

Data preprocessing included removing of "\n" tags and "not assigned" boroughs in some postal codes. Then, the neighborhoods and coordinates were merged into a single dataframe.

The following step would be to use Foursquare to find and filter the venues to include "Shopping malls" only.