VANCOUVER

IBM Data Science Capstone Project
- The Battle of Neighborhoods

About Vancouver, Canada

Key facts

Vancouver is a coastal seaport city in western Canada, located in the Lower Mainland region of British Columbia. As the most populous city in the province, the 2016 census recorded 631,486 people in the city, up from 603,502 in 2011. The Greater Vancouver area had a population of 2,463,431 in 2016, making it the third-largest metropolitan area in Canada. Vancouver has the highest population density in Canada, with over 5,400 people per square kilometre, which makes it the fifth-most densely populated city with over 250,000 residents in North America, behind New York City, Guadalajara...

Local time: Pacific Standard Time

Population: 631,486 (2016)

Area: 44.39 sq miles (115 km^2)

Nearby airport: Vancouver International Airport, Abbotsford International Airport

Colleges and universities: University of British Columbia, Langara College, Regent College, Emily Carr University of Art and Design, Vanc...

Host of: 2010 Winter Olympics

Mayor: Kennedy Stewart

CONTENTS

INTRODUCTION

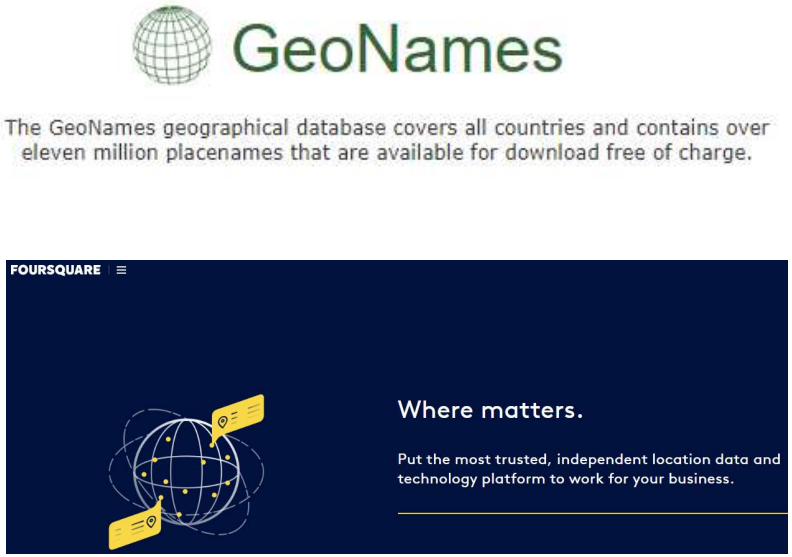
A new student is moving to Vancouver, Canada to study at University of British Columbia. To help her decide where to live, s/he would like to know which neighborhood has which characteristics and how each neighborhood differs from each other.

Other potential interested parties may include real estate agents, housing developers, etc.

DATA

In courtesy to GeoNames (https://www.geonames.org), a downloadable csv file is available and contains all Canadian postal codes and corresponding geospatial data.

We also utilize the Foursquare, which is the location data provider we practiced using in this course, API to explore the neighborhoods and segment them.



METHODOLOGY AND RESULTS

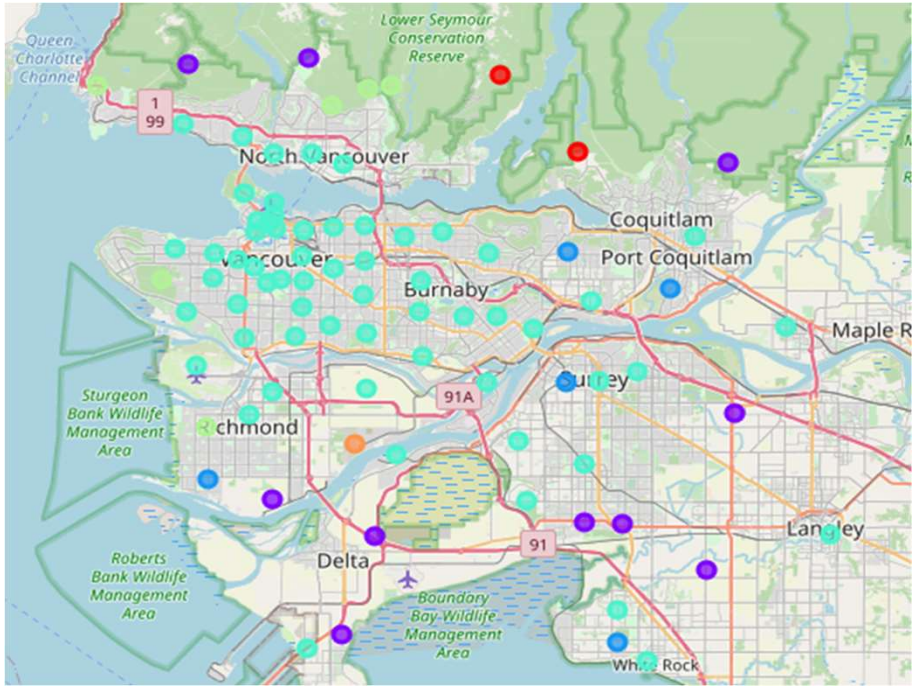
Top 20 venues within 1000-meter radius of each neighborhood in Greater Vancouver.

1,223 venues in 219 unique categories across our target 88 neighborhoods.

Ranging from ATM, various types of restaurants, public facilities, to yoga studios, etc.

Top 10 venue categories in each neighborhood.

k-means clustering to cluster the neighborhoods into 6 clusters.



CONCLUSION

Neighborhood clustering based on Foursquare venues data effectively display each neighborhood's venues or lifestyle characteristics.

For potential client interest, we could enhance neighborhood segmentation by including other neighborhood-specific socio-economic data. E.g. income level, education level, crime rate, housing prices, etc.