SEGMENTATION OF SHOPS

by
Median Household Income
in Metro Vancouver

Background

- Objective: Find segmentation of shops and effect of income levels
- Location: Metro Vancouver
- Reasons for this location:
 - Highly diverse urban area
 - Wide range of income level
 - Large collection of shops

Data Source

- 2016 Census Profile from Statistics Canada
- Geolocation Services API
- Foursquare API

Methodology

- Python Jupyter Notebook
 - Pandas
 - Requests
 - Sklearn
 - Etc.

Data Process Pipeline

Read CSV and import
Metro Vancouver
FSAs

Query the Geolocation
Service and get
coordinates

Read CSV and import
median household
income from 2016
Canada Census

Visualize the results

in Folium

Run the sklearn's

KMeans algorirhtm

Join the income and

FSA coordinates

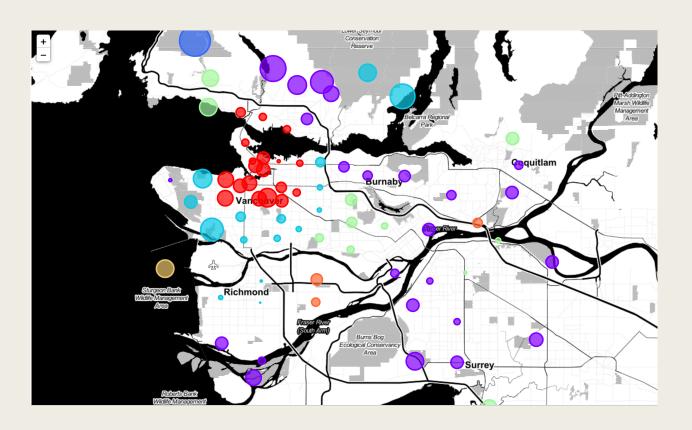
datasets

Normalize the income

and prepare for

KMeans Clustering

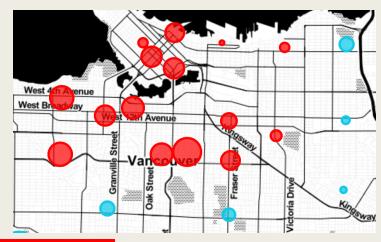
Results



Color: Clusters based on common restaurants

Size: Median Household Income

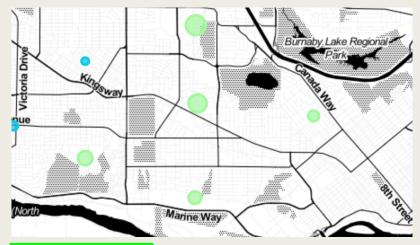
Example



Red Cluster:

Most Common Shops:

- Grocery Store
- Apparel
- Pharmacy
- Liquor Store



Green Cluster:

Most Common Shops:

- Big Box Store
- Supermarket
- Discount Store

Income: \$95407 Income: \$91936

Discussion

- Higher income -> More expensive shops
- More factors may affect the shop categories, such as:
 - Culture
 - Race
 - Household size
 - Etc.

THANK YOU

Bingbing Zhang