Identifying optimum neighborhood location for a bike service shop in Melbourne, Australia

Mansour Bagherzadeh

June 18, 2019

Introduction

1. **Background**
For the past two decade large cities around the world are providing more bicycles paths and lanes. Biking as a mode of transportation and recreation has become more popular. City of Melbourne in Australia is progressive in making changes to accommodate bikers more and more every year. According to city of Melbourne website In March 2017, bicycles made up 16 per cent of all vehicle movements into the city in the morning peak period (between 7am and 10am). In March 2008 the figure was nine per cent". Consequently number of bikers has been increasing steadily. Like a car that needs a regular service, bicycles require regular maintenance and service. As the number of bikers increases there is a need for more bicycles services.

2. **Problem**

   A small business entrepreneur has approached our consulting firm and wants us to find the optimum neighborhood in Melbourne to set up a bicycles shop that maximizes return on investment.

Data

Our data collection involves four major sources.

1. **Wikipedia:** Where we scrape location and neighborhood data. This data will enable us to determine the number and names of neighborhoods and also provides data to proceed with mapping the neighborhoods using folium.

   https://en.wikipedia.org/wiki/Local_government_areas_of_Victoria#Greater_Melbourne

2. **Google search**: Where we locate and get data on existing bike service shops in Melbourne area. We need to determine how many bike services shops already exist in Melbourne neighborhoods. So we need to do a search and find all the names and get their address. Once we get the address we can use geocoders to get location data and superimpose that data with neighborhood map

   https://www.google.com/maps/search/bicycle+repair+shop+in+melbourne+vic/@-37.823515,144.9436148,9z/data=!3m1!4b1

3. **Australian Bureau of Statistics (ABS):** We will use ABS to gather population, median age, income for each neighborhood.
   https://quickstats.censusdata.abs.gov.au/census_services/getproduct/census/2016/quickstat/SSC21495?opendocument

4. **Foursquare:** We will use Foursquare to determine venues for each neighborhood. This will be a major feature for our model.

Data gathering, cleaning

First dataset was obtained from Wikipedia,

	0	1	2	3	4	5	6	7	8	9	10	11
0	Local government area	Council seat	Region	Year est.	Land area	Population(2001 census)	Population(2016 census)	Councillors(2012)	Notes	Map	NaN	NaN
1	km²	sq mi	Density(2016)	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	City of Melbourne	Melbourne	Inner Melbourne	12 August 1842	37	14	3486	60,745[1]	135,959[2]	11	Capital city of Victoria.	NaN
3	City of Port Phillip	St Kilda	Inner Melbourne	22 June 1994	21	8	5101	78,227[3]	100,863[4]	7	Amalgamation of the cities of St Kilda, South ...	NaN
4	City of Stonnington	Malvern	Inner Melbourne	22 June 1994	26	10	4152	86,090[5]	103,832[6]	9	Amalgamation of the cities of Malvern and Prahran	NaN

This data was processed. Column names were defined and columns that were not needed were dropped.

|	Local government area	Council	Region	Neighborhood
0	City of Melbourne	Melbourne	Inner Melbourne	Melbourne,vic,australia
1	City of Port Phillip	St Kilda	Inner Melbourne	St Kilda,vic,australia
2	City of Stonnington	Malvern	Inner Melbourne	Malvern,vic,australia
3	City of Yarra	Richmond	Inner Melbourne	Richmond,vic,australia
4	City of Banyule	Ivanhoe	Metropolitan Melbourne	Ivanhoe,vic,australia

Next step was to get the Latitude and Longitude for each neighborhood. This was done using Geolocator.

	Neighborhood	Latitude	Longitude
0	Altona,vic,australia	-37.883608	144.758616
1	Broadmeadows,vic,australia	-37.673105	144.940159
2	Camberwell,vic,australia	-37.838462	145.074077
3	Caulfield North,vic,australia	-37.870828	145.021801
4	Cheltenham,vic,australia	-37.963418	145.061567

Once we had the coordinates we gathered and created a csv file from Australian Bureau of Statistics.

	Neighborhood	population	median age	yealy income
0	Altona,vic,australia	12689	42	74880
1	Broadmeadows,vic,australia	13305	31	47476
2	Camberwell,vic,australia	21400	41	115024
3	Caulfield North,vic,australia	20967	34	99944
4	Cheltenham,vic,australia	22291	40	78104

We added the information from both sets of data and created a Data frame. This was used to create a map using folium. Below is the map of all neighborhoods.
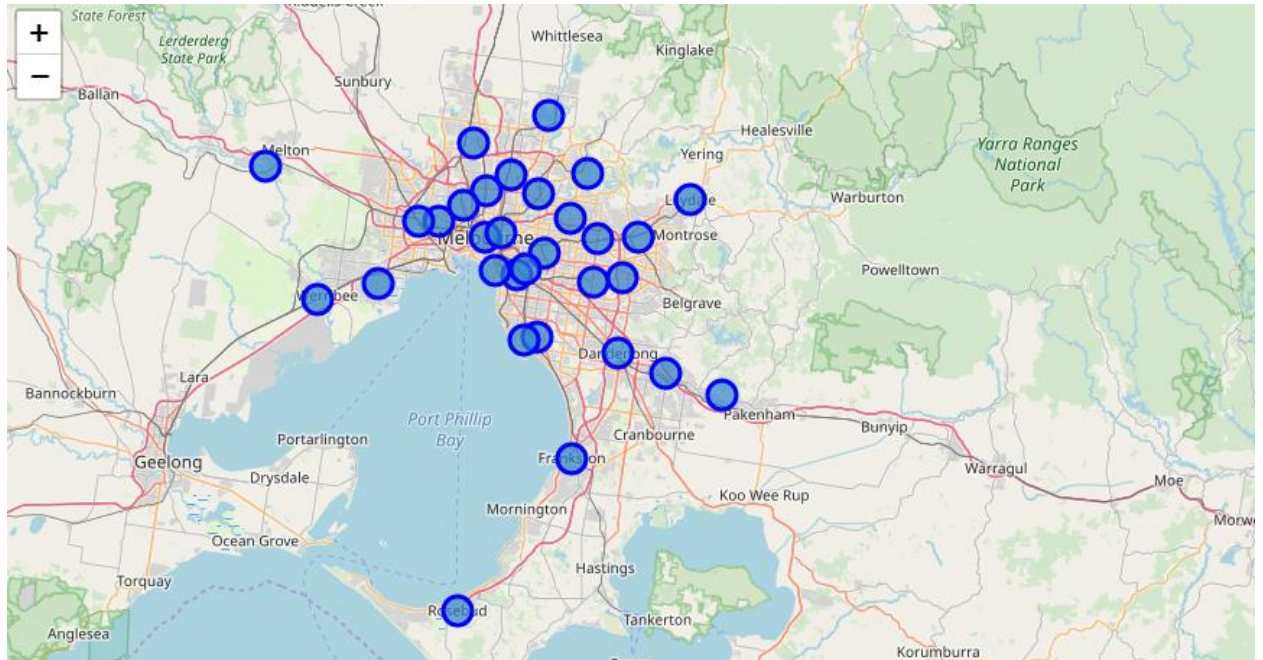
Next step in data gathering was to obtain current locations of existing bike service shops in and around Melbourne. We used Google to do the search and created a csv file with names and addresses of these shops and used Geolocator the get their coordinates.

	Name	Address	Latitude	Longitude
0	SC Cyclery	39 Bridge Road, Richmond, Melbourne VIC 3121, ...	-37.817557	144.990743
1	The Freedom Machine - Port Melbourne	265/267 Bay St, Port Melbourne VIC 3207, Austr...	-37.837649	144.943206
2	Back2Bikes Ltd	525 Williamstown Rd, Port Melbourne VIC 3207, ...	-37.836467	144.917105
3	Good Cycles	1/750 Collins St, Docklands VIC 3008, Australia	-37.820903	144.948068
4	Cycles Galleria	385 Bourke St, Melbourne VIC 3000, Australia	-37.814664	144.962520

Final piece of data was to gather the venues for each neighborhood within 1000 meters. This was done using Foursquare API calls.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
Altona,vic,australia	5	5	5	5	5	5
Broadmeadows,vic,australia	11	11	11	11	11	11
Camberwell,vic,australia	14	14	14	14	14	14
Caulfield North,vic,australia	20	20	20	20	20	20
Cheltenham,vic,australia	30	30	30	30	30	30

Feature set

Our final feature set was put together. Below is our featured data frame. There are four features per neighborhood (population, median age, yearly income, Venue)

	Neighborhood	Latitude	Longitude	population	median age	yealy income	Venue
0	Altona,vic,australia	-37.883608	144.758616	12689	42	74880	5
1	Broadmeadows,vic,australia	-37.673105	144.940159	13305	31	47476	11
2	Camberwell,vic,australia	-37.838462	145.074077	21400	41	115024	14
3	Caulfield North,vic,australia	-37.870828	145.021801	20967	34	99944	20
4	Cheltenham,vic,australia	-37.963418	145.061567	22291	40	78104	30

Modeling

Our choice of modeling in this scenario was Sklearn KMeans . Clustering of unlabeled data can be performed with this module from sklearn. According to their document The KMeans algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criterion known as the *inertia* or within-cluster sum-of-squares. This algorithm requires the number of clusters to be specified. It scales well to large number of samples and has been used across a large range of application areas in many different fields.

But before running the data through our model we used standard scaler to scale our feature. This will avoid domination by certain features with larger values in our case population or income.

Since our total number of neighborhoods was 30 it made sense to choose 4 clusters.

```
pop_nei.groupby('Labels').mean()
```

Labels	Latitude	Longitude	population	median age	yealy income	Venue
0	-37.800733	145.046198	16828.666667	38.833333	106903.333333	23.500000
1	-37.883238	145.066264	26187.416667	35.583333	72158.666667	28.416667
2	-37.870181	145.085983	13277.000000	37.875000	74061.000000	8.375000
3	-37.840531	144.893311	57602.400000	36.600000	82326.400000	11.200000

By examining the mean for each Label, Cluster one has an average of 28.4 Venues, youngest median age. Below we show cluster one with its Neighborhoods in Y-axis and (population, median age, yearly income, Venue) in x-axis.
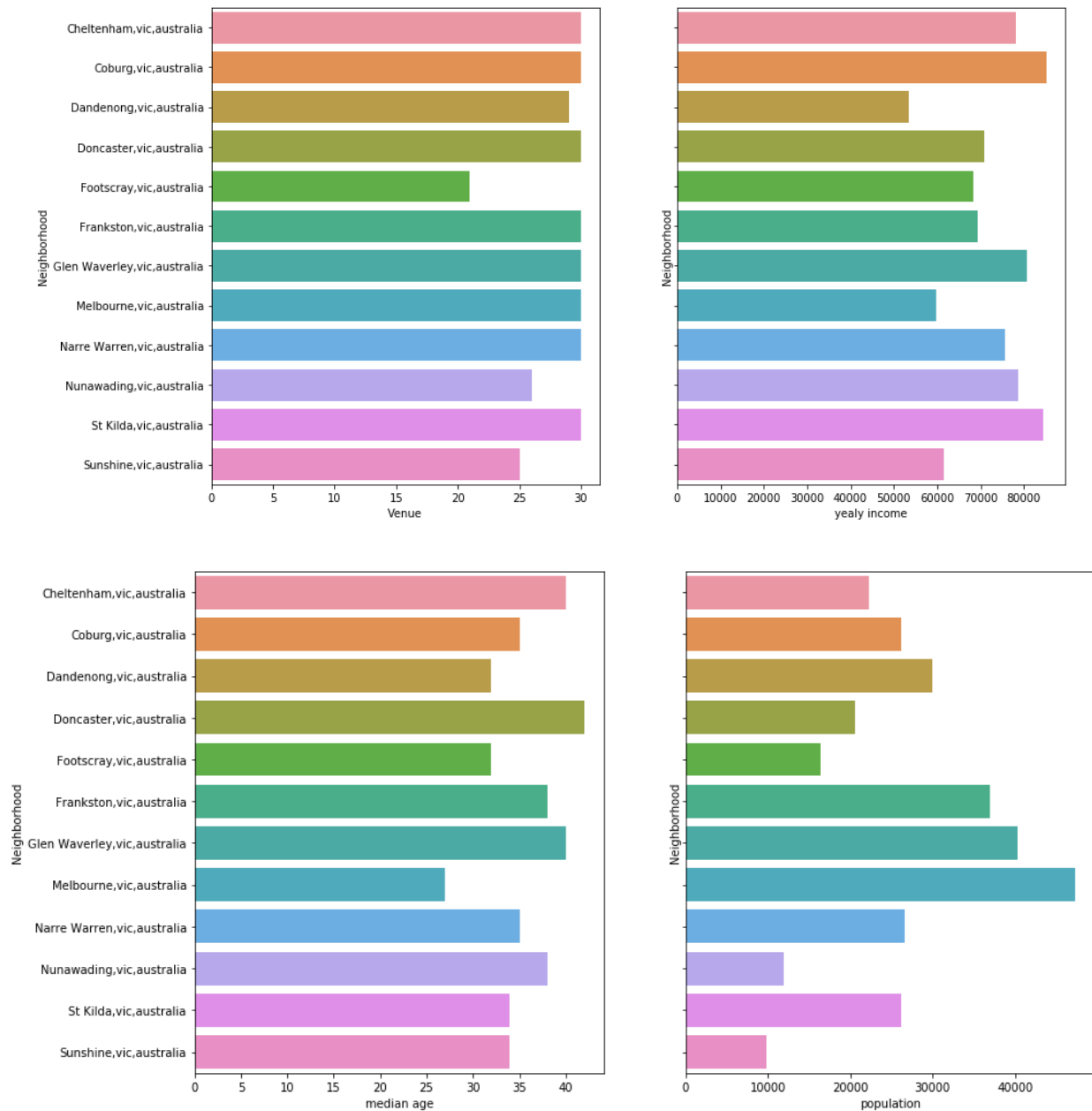
Comparing locations

Now we map the cluster one results with existing bike shops. The green dots in the map show where the existing bike shops are located. And the blue circles are cluster one.
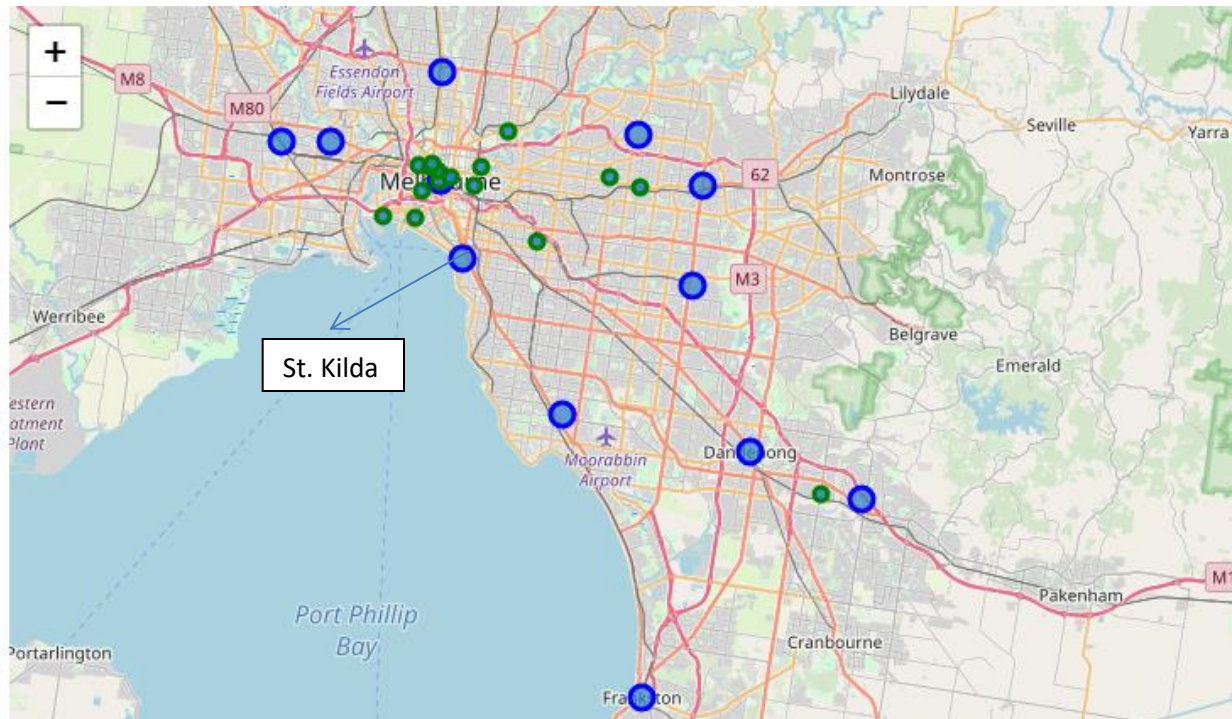
Conclusion

Based on the above map and results of modeling, St. Kilda is the optimum location for our client's bike service shop. In addition to the model result we note that it's a beautiful coastal neighborhood.

```
Neighborhood     St Kilda,vic,australia
Latitude                       -37.8638
Longitude                       144.982
population                        26124
median age                           34
yealy income                      84500
Venue                                30
Labels                                1
```