# Finding the Best Area to Live for Birdwatchers in Singapore

## Nugroho Budianggoro

## January 21, 2021

## 1. INTRODUCTION

Humans have a innate instinct to connect with nature and other living beings (1). One such connection is through watching birds in their habitat. Indeed, birdwatching is a popular hobby around the world (2). Some people, including the author, can have stronger need to connect with nature more than others. This need can influence their buying decisions, such as choosing a place of residence. In connection with the birdwatching activity, factors that maximize the birdwatching experience in an area, such as bird abundance and diversity, can affect whether a birdwatching enthusiast would decide to live in that area or not.

Singapore is a country in South East Asia. The country is a well developed island city state with numerous venues or facilities such as restaurants, medical centers, malls, schools, parks, and transit systems all over the country built by/for the people. It was chosen as the subject area of this study because the bird observation data from there is extensive, the area is not too large but also not too small, it is a city that the author has visited several times, and the Foursquare API data from the area (that this study used) seemed updated enough.

The goal of this study is to help birdwatching enthusiasts choose an area to live in Singapore with maximum birdwatching experience while also considering availability of important venues using machine learning.

## 2. DATA

There are two datasets that were used in this study. The first is Singapore bird observation data obtained from The Global Biodiversity Information Facility (GBIF) (3). This dataset contains 423211 rows of Singapore bird observations from year 1800 to 2020. The second dataset is the venues data from the Foursquare API.

## 3. METHODOLOGY

### Data acquisition

A query in GBIF was made for all available records data for birds of Singapore. The resulting dataset consists of 423211 rows and 249 columns. Data preprocessing The acquired dataset was then cleaned from unused columns and missing data. The process was conducted with Microsoft Excel. The resulting dataset contained 401607 rows and 4 columns. The 4 columns were: latitude, longitude, species name, species id.

### Sectors determination

To convert the data points into a more manageable form, a clustering was performed using latitude and longitude features of the data. The result was 40 clusters of data, which we called sectors, from sector 0 to sector 39. The quantity of 40 was chosen because the resulting clusters seemed to have the right sizes, specifically for exploration through walking or taking a short trip on bus. For each sectors, a center was determined.

### Venues exploration

For each of the 40 sector centers, 3 API calls were made to the Foursquare API. Each API calls requested venues of certain categories within 1 km radius of the centers. The first API call requested venues data of the category 'Food' that contained restaurants, cafes, bars, etc. This first API call requested 10 results. The second and third API requested venue data of categories that were considered important and desirable features for a person to choose to live in the area. Such categories were: residence area, transit system, medical center, convenience store, mall, park, school, university, nature reserve. The second API call requested 50 results, while the third call requested 40. So for each of the 40 sector centers, the maximum API call result was 100 venues within 1 km of the centers.

### More processing

The top 10 most popular venue types were calculated for each sectors. The species data were aggregated to obtain the number of uniques species for each sectors. To prepare for cluster analysis, the venue types data were transformed into integer data with one-hot encoding using the mean.

### Clustering

Clustering analysis were implemented on the prepared data using k-means clustering algorithm. To determine the optimum number of k, Yellowbrick library were used. The elbow method visualization were produced quickly with Yellowbrick. The resulting cluster data with the optimum k were then merged with the species numbers data for each sectors and the sectors data that contain the top 5 most popular venue types for each sectors.

### Mapping

The final clustering result were then mapped using Folium library.

## 4. RESULTS

The Singapore bird observation data used in this study consists of 401067 rows and are plotted as follows:
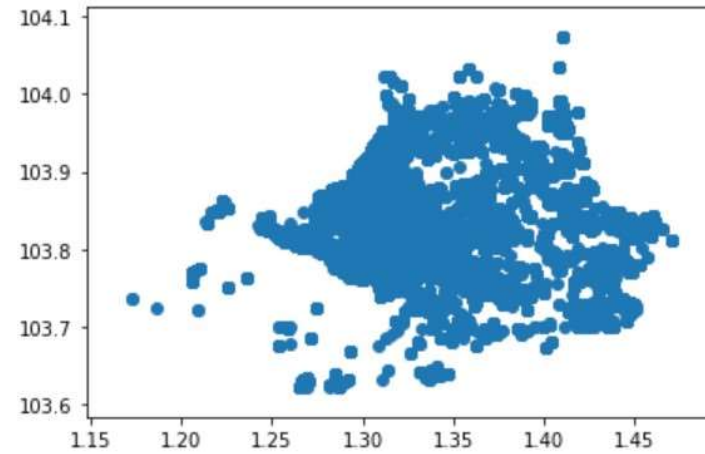


Figure 1. Initial mapping of Singapore bird observation data.

K-means clustering algorithm was implemented with the above Singapore bird observation data with 40 centroids. The result is plotted as follows:
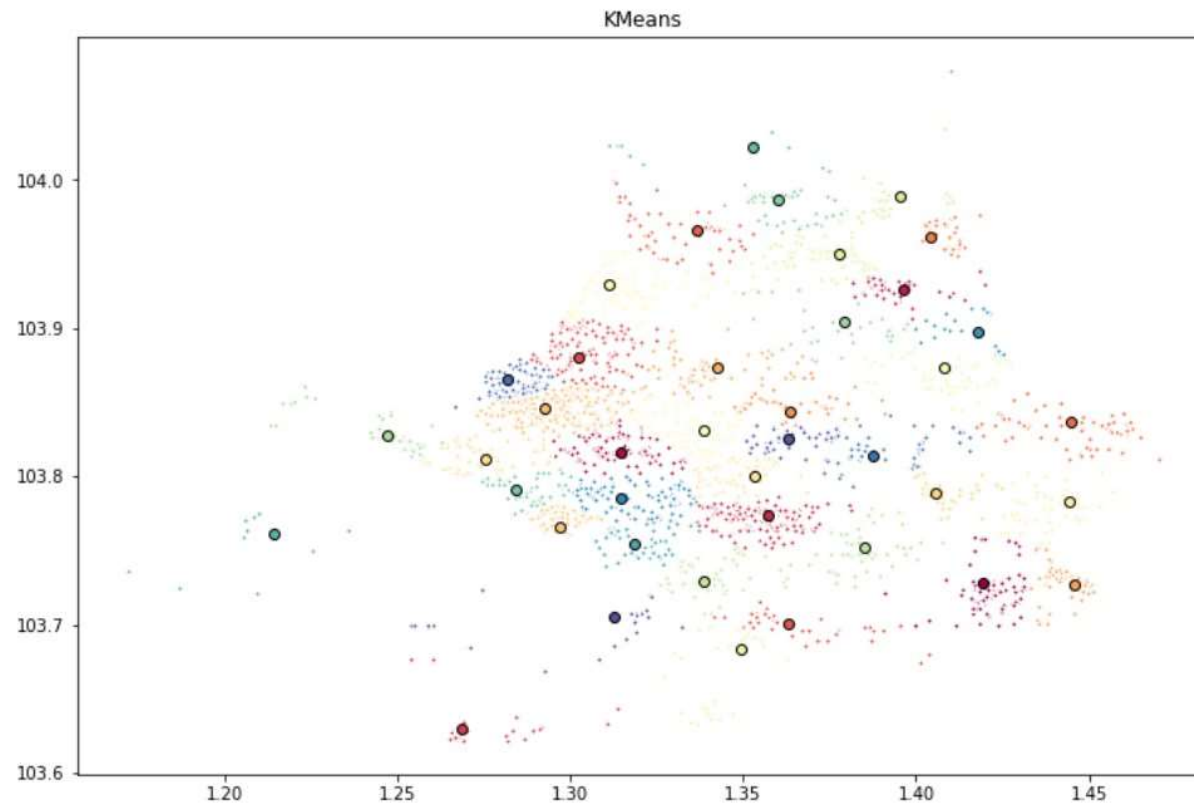


Figure 2. Clustering of Singapore bird observation data (k = 40)

The centroids were then assigned as sectors. The sectors are plotted in map with Folium as follows:
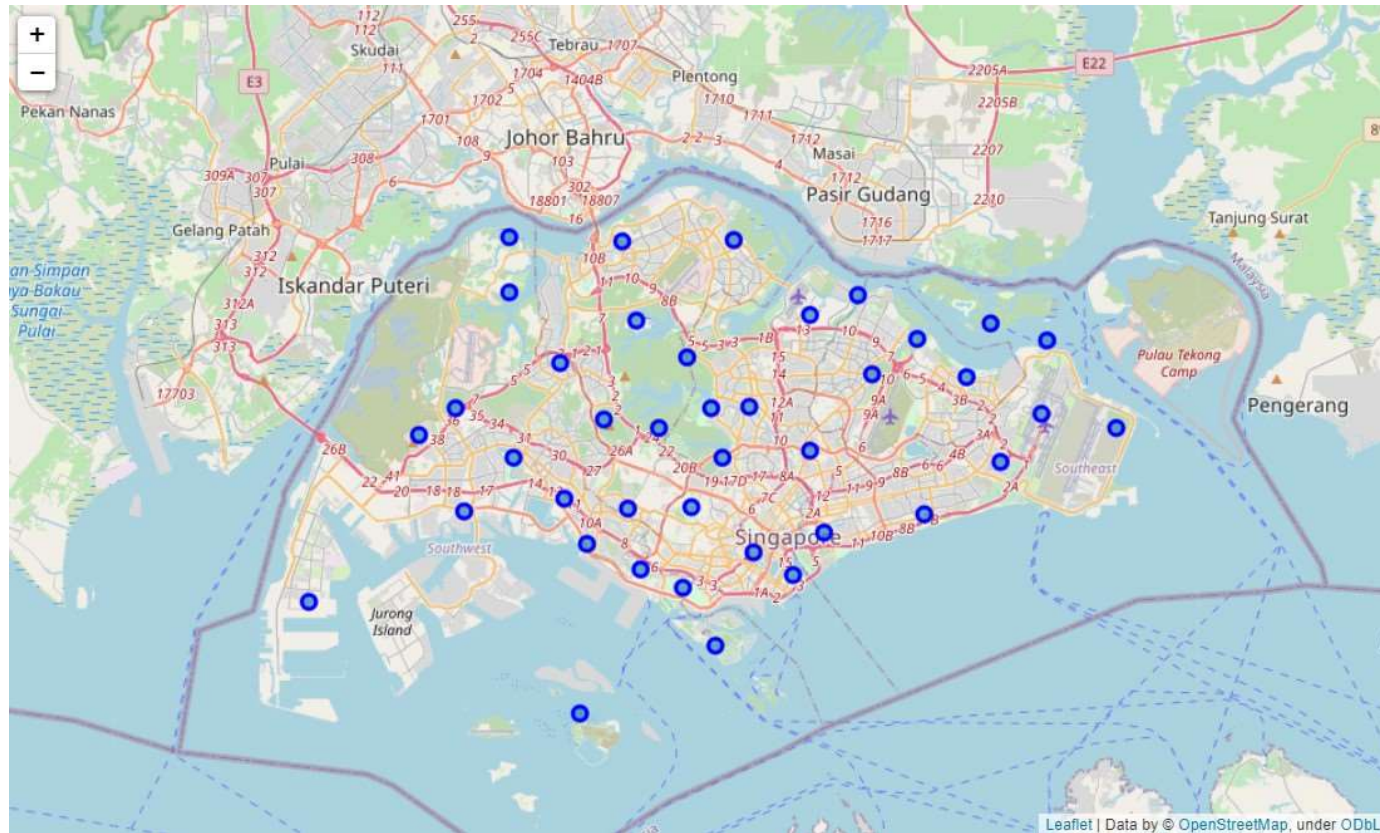


Figure 3. Centers of bird observation data points (sectors) in the form of a map

After collecting the venues and species number data for each sector, we again ran k-means clustering algorithm with the sectors data. Before that, 3 sectors were ommitted because there were no venues nearby.

The elbow method suggested 13 as the optimum number of k. However, as the curve seems continuous, it is expected that some of the resulting clusters would show high overlap with another.



Figure 4. The elbow to find optimum k for sectors clustering.

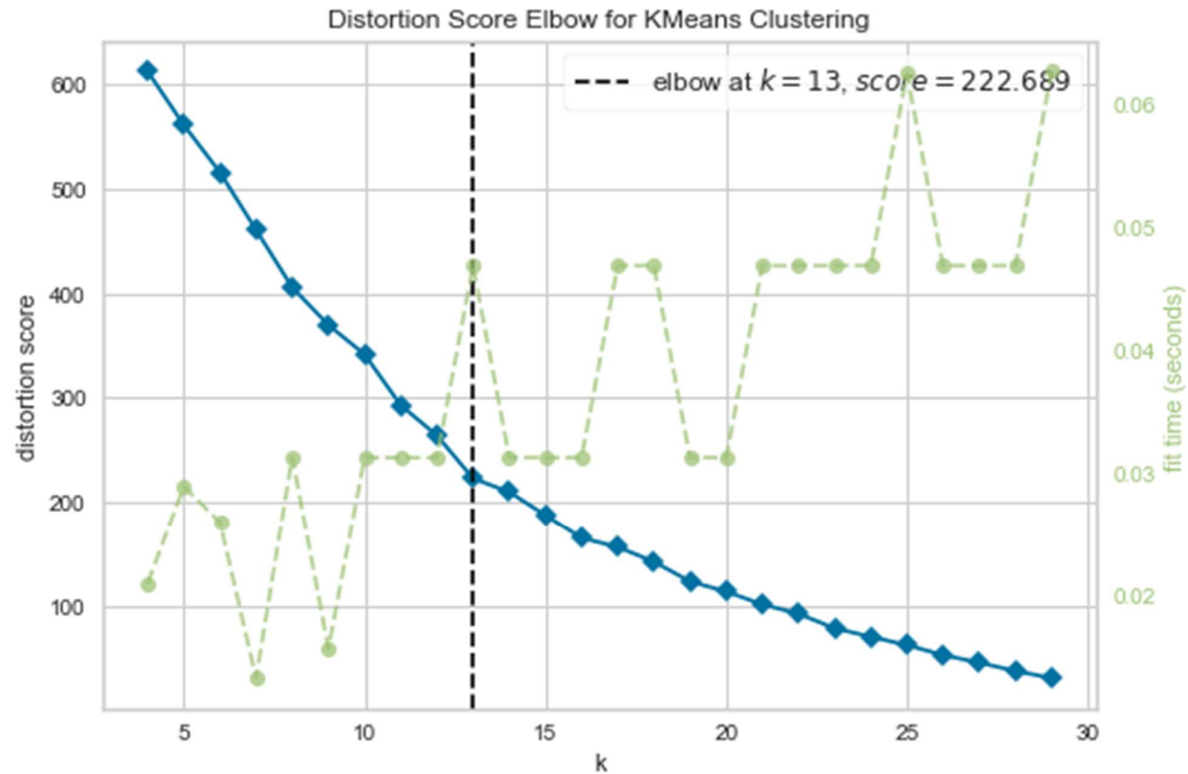K-means algorithm were them implemented again with the sectors, producing the clusters that were mapped as follows:
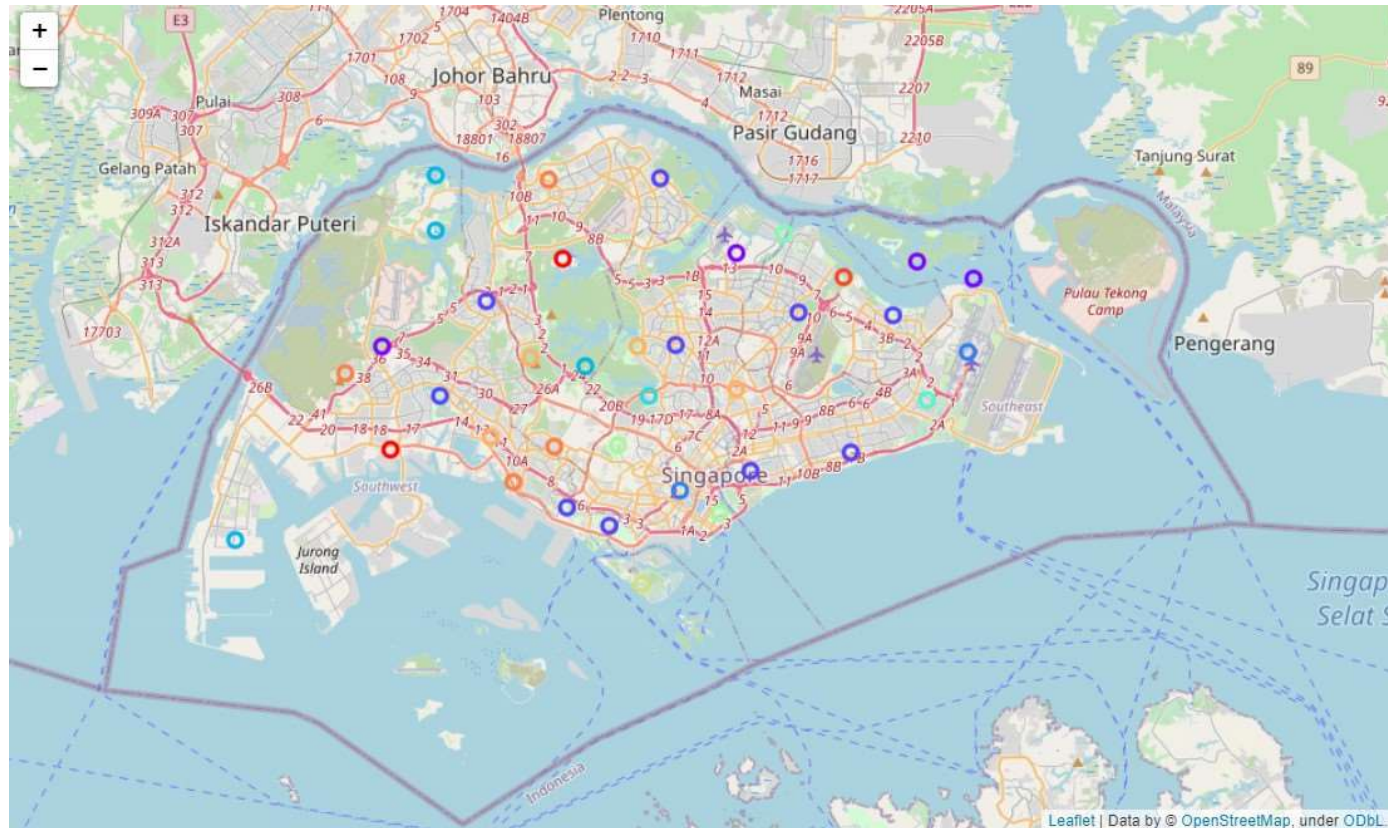


Figure 5. Map of centers of clusters of these sectors that have been implemented with k-means algorithm.

The table for the resulting clustering of sectors is as follows:

| center_lat | center_lon | sect | cluster | species_numb | 1st_most_common_venu | 2nd_most_common_venu | 3rd_most_common_venu | 4th_most_common_venu |
|---|---|---|---|---|---|---|---|---|
| 1.312898966588966 | 103.70548717948719 | 39 | 0 | 125 | Zoo | Food | Bus Transit | Park |
| 1.405399240041928 | 103.7889142033543 | 15 | 0 | 158 | Zoo | Food | Park | Rail Transit |
| 1.407938871958712 | 103.8736994101745 | 21 | 1 | 162 | Food | Bus Transit | College | Coast |
| 1.395489743658931 | 103.98894603345926 | 24 | 1 | 190 | Food | Bus Transit | Coast | College |
| 1.3630427683544304 | 103.70101684367089 | 6 | 1 | 135 | Food | Bus Transit | Park | Mall |
| 1.4039747035040253 | 103.9613842678897 | 9 | 1 | 260 | Food | Convenience Store | Trail | Coast |
| 1.377627694399132 | 103.94970738323966 | 23 | 2 | 223 | Convenience Store | Bus Transit | Residence | Food |
| 1.3636938904716682 | 103.84406028806585 | 10 | 2 | 154 | Residence | Convenience Store | Bus Transit | Food |
| 1.4446758263723165 | 103.83675119331743 | 8 | 2 | 176 | Residence | Bus Transit | Food | Convenience Store |
| 1.3387780796617692 | 103.72914752002669 | 25 | 2 | 197 | Bus Transit | Residence | Convenience Store | Food |
| 1.3022430326620453 | 103.88042003723135 | 5 | 2 | 155 | Bus Transit | Residence | Food | Convenience Store |
| 1.385055628095452 | 103.75240882485367 | 26 | 2 | 133 | Residence | Convenience Store | Food | Bus Transit |
| 1.379236919903747 | 103.90385867995874 | 28 | 2 | 138 | Residence | Bus Transit | Convenience Store | Food |
| 1.2843734533616802 | 103.79099706749933 | 30 | 2 | 150 | Bus Transit | Residence | Food | Park |
| 1.27560908506566 | 103.81188501730887 | 16 | 2 | 191 | Residence | Bus Transit | Convenience Store | Food |
| 1.3114496562116378 | 103.92938563455742 | 19 | 2 | 155 | Bus Transit | Residence | Convenience Store | Food |
| 1.2928250926759022 | 103.84609161418031 | 13 | 3 | 177 | Bus Transit | Mall | Convenience Store | Food |
| 1.3599356215213358 | 103.98629128014842 | 29 | 3 | 107 | Convenience Store | Food | Bus Transit | Mall |
| 1.4194708469714172 | 103.72748910063811 | 0 | 4 | 239 | Food | Nature Reserve | Zoo | Mall |
| 1.353501416331836 | 103.80012787166156 | 17 | 4 | 234 | Food | Trail | Nature Reserve | Zoo |
| 1.4459940663168214 | 103.7271607470437 | 11 | 4 | 266 | Food | Nature Reserve | Zoo | Park |
| 1.268860751947126 | 103.62964894784047 | 4 | 4 | 206 | Food | Zoo | Trail | Clinic |
| 1.338740013534076 | 103.83112455497718 | 20 | 5 | 203 | Hospital | Bus Transit | Clinic | Food |
| 1.3365111335724404 | 103.96630152157375 | 7 | 6 | 195 | Bus Transit | Food | Convenience Store | Park |
| 1.418069013853441 | 103.8970163507109 | 34 | 7 | 192 | Food | Coast | Residence | Rail Transit |

| center_lat | center_lon | sector | cluster | species_numbe | 1st_most_common_venue | 2nd_most_common_venue | 3rd_most_common_venue | 4th_most_common_venue |
|---|---|---|---|---|---|---|---|---|
| 1.2816921983418426 | 103.86516586734692 | 37 | 8 | 217 | Park | Food | Bus Transit | Convenience Store |
| 1.3148320519795194 | 103.81620061595058 | 1 | 8 | 208 | Park | College | Food | Hospital |
| 1.247302823559809 | 103.82756574528545 | 27 | 9 | 191 | Theme Park | Food | Park | Bus Transit |
| 1.357278108317763 | 103.7733352932907 | 3 | 10 | 239 | Bus Transit | Trail | Food | Residence |
| 1.3423697788332056 | 103.87340413978876 | 12 | 10 | 260 | Residence | Bus Transit | Convenience Store | Food |
| 1.3187535820327367 | 103.75431351351351 | 33 | 10 | 136 | Bus Transit | Residence | Food | Convenience Store |
| 1.3631777770681717 | 103.82532088638052 | 38 | 10 | 203 | Bus Transit | Residence | Food | Trail |
| 1.3495148028600605 | 103.68361385086823 | 22 | 11 | 181 | College | Bus Transit | Food | Residence |
| 1.3143983949152558 | 103.78521400635593 | 35 | 11 | 144 | College | Residence | Bus Transit | Food |
| 1.2969024463749794 | 103.76532389053862 | 14 | 11 | 151 | Bus Transit | College | Residence | Food |
| 1.4440661024925858 | 103.78252518738016 | 18 | 11 | 144 | College | Residence | Bus Transit | Food |
| 1.396323177317086 | 103.92617867065614 | 2 | 12 | 230 | Food | Coast | Convenience Store | Elementary School |

Figure 6. Table of the clustering result and the most popular venue types for each sectors.

# 5. DISCUSSION

The clustering of Singapore area based on birds observation data and nearby venues produced 13 clusters. After considering the features of each clusters, the clusters can be named based on their characteristics as follows.

Cluster 0 : Zoo

Cluster 1 : Food

Cluster 2 : Residence 1

Cluster 3 : Mall

Cluster 4 : Foo & Nature Reserve

Cluster 5 : Hospital

Cluster 6 : Bus Transit

Cluster 7 : Coast 1

Cluster 8 : Park

Cluster 9 : Theme Park

Cluster 10 : Residence 2

Cluster 11 : College

Cluster 12 : Coast 2

| center_lat | center_lon | sector | cluster | cluster_name | species_numb | 1st_most_common_ven | 2nd_most_common_ven | 3rd_most_common_ven | 4th_most_common |
|---|---|---|---|---|---|---|---|---|---|
| 1.405399240041928 | 103.7889142033543 | 15 | 0 | Zoo | 158 | Zoo | Food | Park | Rail Transit |
| 1.312898966588966 | 103.7054871794871 | 39 | 0 | Zoo | 125 | Zoo | Food | Bus Transit | Park |
| 1.403974703504025 | 103.9613842678897 | 9 | 1 | Food | 260 | Food | Convenience Store | Trail | Coast |
| 1.395489743658931 | 103.9889460334592 | 24 | 1 | Food | 190 | Food | Bus Transit | Coast | College |
| 1.407938871958712 | 103.8736994101745 | 21 | 1 | Food | 162 | Food | Bus Transit | College | Coast |
| 1.363042768354430 | 103.7010168436708 | 6 | 1 | Food | 135 | Food | Bus Transit | Park | Mall |
| 1.377627694399132 | 103.9497073832396 | 23 | 2 | Residence 1 | 223 | Convenience Store | Bus Transit | Residence | Food |
| 1.338778079661769 | 103.7291475200266 | 25 | 2 | Residence 1 | 197 | Bus Transit | Residence | Convenience Store | Food |
| 1.27560908506566 | 103.8118850173088 | 16 | 2 | Residence 1 | 191 | Residence | Bus Transit | Convenience Store | Food |
| 1.444675826372316 | 103.8367511933174 | 8 | 2 | Residence 1 | 176 | Residence | Bus Transit | Food | Convenience Store |
| 1.302243032662045 | 103.8804200372313 | 5 | 2 | Residence 1 | 155 | Bus Transit | Residence | Food | Convenience Store |
| 1.311449656211637 | 103.9293856345574 | 19 | 2 | Residence 1 | 155 | Bus Transit | Residence | Convenience Store | Food |
| 1.363693890471668 | 103.8440602880658 | 10 | 2 | Residence 1 | 154 | Residence | Convenience Store | Bus Transit | Food |
| 1.284373453361680 | 103.7909970674993 | 30 | 2 | Residence 1 | 150 | Bus Transit | Residence | Food | Park |
| 1.379236919903747 | 103.9038586799587 | 28 | 2 | Residence 1 | 138 | Residence | Bus Transit | Convenience Store | Food |
| 1.385055628095452 | 103.7524088248536 | 26 | 2 | Residence 1 | 133 | Residence | Convenience Store | Food | Bus Transit |
| 1.292825092675902 | 103.8460916141803 | 13 | 3 | Mall | 177 | Bus Transit | Mall | Convenience Store | Food |
| 1.359935621521335 | 103.9862912801484 | 29 | 3 | Mall | 107 | Convenience Store | Food | Bus Transit | Mall |
| 1.445994066316821 | 103.7271607470437 | 11 | 4 | Food & Nature Reserve | 266 | Food | Nature Reserve | Zoo | Park |
| 1.419470846971417 | 103.7274891006381 | 0 | 4 | Food & Nature Reserve | 239 | Food | Nature Reserve | Zoo | Mall |
| 1.353501416331836 | 103.8001278716615 | 17 | 4 | Food & Nature Reserve | 234 | Food | Trail | Nature Reserve | Zoo |
| 1.268860751947126 | 103.6296489478404 | 4 | 4 | Food & Nature Reserve | 206 | Food | Zoo | Trail | Clinic |
| 1.338740013534076 | 103.8311245549771 | 20 | 5 | Hospital | 203 | Hospital | Bus Transit | Clinic | Food |
| 1.336511133572440 | 103.9663015215737 | 7 | 6 | Bus Transit 1 | 195 | Bus Transit | Food | Convenience Store | Park |
| 1.418069013853441 | 103.8970163507109 | 34 | 7 | Coast 1 | 192 | Food | Coast | Residence | Rail Transit |
| 1.281692198341842 | 103.8651658673469 | 37 | 8 | Park | 217 | Park | Food | Bus Transit | Convenience Store |
| 1.314832051979519 | 103.8162006159505 | 1 | 8 | Park | 208 | Park | College | Food | Hospital |

| center_lat | center_lon | sector | cluster | cluster_name | species_numb | 1st_most_common_venu | 2nd_most_common_venu | 3rd_most_common_venu | 4th_most_common_ |
|---|---|---|---|---|---|---|---|---|---|
| 1.247302823559809 | 103.8275657452854 | 27 | 9 | Theme Park | 191 | Theme Park | Food | Park | Bus Transit |
| 1.342369778833205 | 103.8734041397887 | 12 | 10 | Residence 2 | 260 | Residence | Bus Transit | Convenience Store | Food |
| 1.357278108317763 | 103.7733352932907 | 3 | 10 | Residence 2 | 239 | Bus Transit | Trail | Food | Residence |
| 1.363177777068171 | 103.8253208863805 | 38 | 10 | Residence 2 | 203 | Bus Transit | Residence | Food | Trail |
| 1.318753582032736 | 103.7543135135135 | 33 | 10 | Residence 2 | 136 | Bus Transit | Residence | Food | Convenience Store |
| 1.349514802860060 | 103.6836138508682 | 22 | 11 | College | 181 | College | Bus Transit | Food | Residence |
| 1.296902446374979 | 103.7653238905386 | 14 | 11 | College | 151 | Bus Transit | College | Residence | Food |
| 1.314398394915255 | 103.7852140063559 | 35 | 11 | College | 144 | College | Residence | Bus Transit | Food |
| 1.444066102492585 | 103.7825251873801 | 18 | 11 | College | 144 | College | Residence | Bus Transit | Food |
| 1.396323177317086 | 103.9261786706561 | 2 | 12 | Coast 2 | 230 | Food | Coast | Convenience Store | Elementary School |

Figure 7. Table of the cluster names and the most popular venue types for each sectors.

The cluster that contains the most sectors is Residence 1 (10 sectors). Clusters Residence 2, College, Food & Nature Reserve, and Food have the second most sectors with 4 sectors each.

There are multiple approach to using the resulting information to decide which area to choose. One of the ways is to simply sort the species number of all sectors and then select which of the top sectors have the most favorable venues. In this discussion part, we will decide which area to choose by first choosing the clusters we would like to live in, sort the bird species numbers and then select a few sectors from those for more specific consideration.

First, we would like to choose the Residence 1 and Residence 2 clusters. These clusters are dominated by residential areas. The assumption is that the more extensive the residence are are, the more the house price would decrease because of competition. So we choose those Residence clusters because there may be more alternatives in terms of price range and the overall price may be less expensive than the houses in the less denser area.

The final selection of the area to choose to live in would depend on respective users. For this discussion, we will use the author's preferences. Based on the selected cluster, sorted bird species numbers, and the features of the sector, we choose sector 23 and sector 16. The considerations are:

- These sectors have relatively high bird species diversity
- The residence area of these sectors are quite dominant, so the housing prices may be lower
- These sectors have desirable venues, namely mall, convenience store, clinic, bus transit, food, park

| center_lat | center_lon | sector | cluster | cluster_name | species_number | 1st_most_common_venue | 2nd_most_common_venue | 3rd_most_common_venue | 4th_most_common_venue | 5th_most_common |
|---|---|---|---|---|---|---|---|---|---|---|
| 1.3423697788332056 | 103.8734041397887€ | 12 | 10 | Residence 2 | 260 | Residence | Bus Transit | Convenience Store | Food | High School |
| 1.357278108317763 | 103.773335293290713 | 13 | 10 | Residence 2 | 239 | Bus Transit | Trail | Food | Residence | Convenience Store |
| 1.3776276943991321 | 103.9497073832396€ | 23 | 2 | Residence 1 | 223 | Convenience Store | Bus Transit | Residence | Food | Park |
| 1.3631777770681717 | 103.8253208863805₂ | 38 | 10 | Residence 2 | 203 | Bus Transit | Residence | Food | Trail | High School |
| 1.3387780796617692 | 103.7291475200266₉ | 25 | 2 | Residence 1 | 197 | Bus Transit | Residence | Convenience Store | Food | Park |
| 1.2756090850656598 | 103.8118850173088₈ | 16 | 2 | Residence 1 | 191 | Residence | Bus Transit | Convenience Store | Food | College |
| 1.4446758263723165 | 103.8367511933174₃ | 8 | 2 | Residence 1 | 176 | Residence | Bus Transit | Food | Convenience Store | Park |
| 1.302243032662045 | 103.8804200372313₅ | 5 | 2 | Residence 1 | 155 | Bus Transit | Residence | Food | Convenience Store | College |
| 1.3114496562116378 | 103.9293856345574₂ | 19 | 2 | Residence 1 | 155 | Bus Transit | Residence | Convenience Store | Food | Clinic |
| 1.3636938904716682 | 103.8440602880658₅ | 10 | 2 | Residence 1 | 154 | Residence | Convenience Store | Bus Transit | Food | Mall |
| 1.284373453361680₂ | 103.7909970674993₂ | 30 | 2 | Residence 1 | 150 | Bus Transit | Residence | Food | Park | Convenience Store |
| 1.379236919903747 | 103.9038586799587₃ | 28 | 2 | Residence 1 | 138 | Residence | Bus Transit | Convenience Store | Food | College |
| 1.3187535820327367 | 103.7543135135135133 | 33 | 10 | Residence 2 | 136 | Bus Transit | Residence | Food | Convenience Store | Clinic |
| 1.3850556280954522 | 103.7524088248536₉ | 26 | 2 | Residence 1 | 133 | Residence | Convenience Store | Food | Bus Transit | College |

Figure 8. Table of the preferred clusters.

After further checking in Google Maps, it is known that the sector 23 (lat/lon: 1.377628/103.949707) is Pasir Ris area on the northeast coast of Singapore island near Changi Airport. Sector 16 (lat/lon: 1.275609/103.811885) is Bukit Merah area on the south part of Singapore. So, based on the analysis on this study, the choice for the author's place of residence where there are maximum experience for birdwatching and there are more desirable venues nearby are:

- Primary: Pasir Ris area
- Secondary: Bukit Merah area

Pasir Ris is also the preferred choice of residence for the author because from there he would have shorter access to the Changi airport any time he wants to come home to his hometown of Jakarta, Indonesia.

## 6. CONCLUSION

In this study we have analyzed Singapore bird observation data and Foursquare API venues data to produce insights that may be useful for birdwatchers find the best area to live in Singapore. The insights may point birdwatchers to areas where there are maximum birdwatching experience and desireable venues. The results of the study were maps of 13 clusters that consisted of 37 sectors and table that listed most common venues for each sectors. The final decision will depend on each users' needs, desireable features of an area, and other information that are not captured in the datasets. However, the analysis, maps and table may serve as useful baseline information for real-life use case of birdwatchers choosing an area of residence in Singapore.

# 7. REFERENCES

Vidovich, E. Bringing the Outdoors In: The Benefits of Biophilia. NRDC: Expert Blog. Last accessed on 21/01/2021. https://www.nrdc.org/experts/maria-mccain/bringing-outdoors-benefits-biophilia

White, J. How popular is birdwatching? Chirpbirding.com. Last accessed on 21/01/2021. https://chirpbirding.com/blog/81/how-popular-is-birdwatching/#:~:text=Millions%20of%20people%20are%20birdwatchers,is%20spent%20birdwatching%20than%20gardening.

GBIF.org (13 January 2021). GBIF Occurrence. Download: https://doi.org/10.15468/dl.43xha7