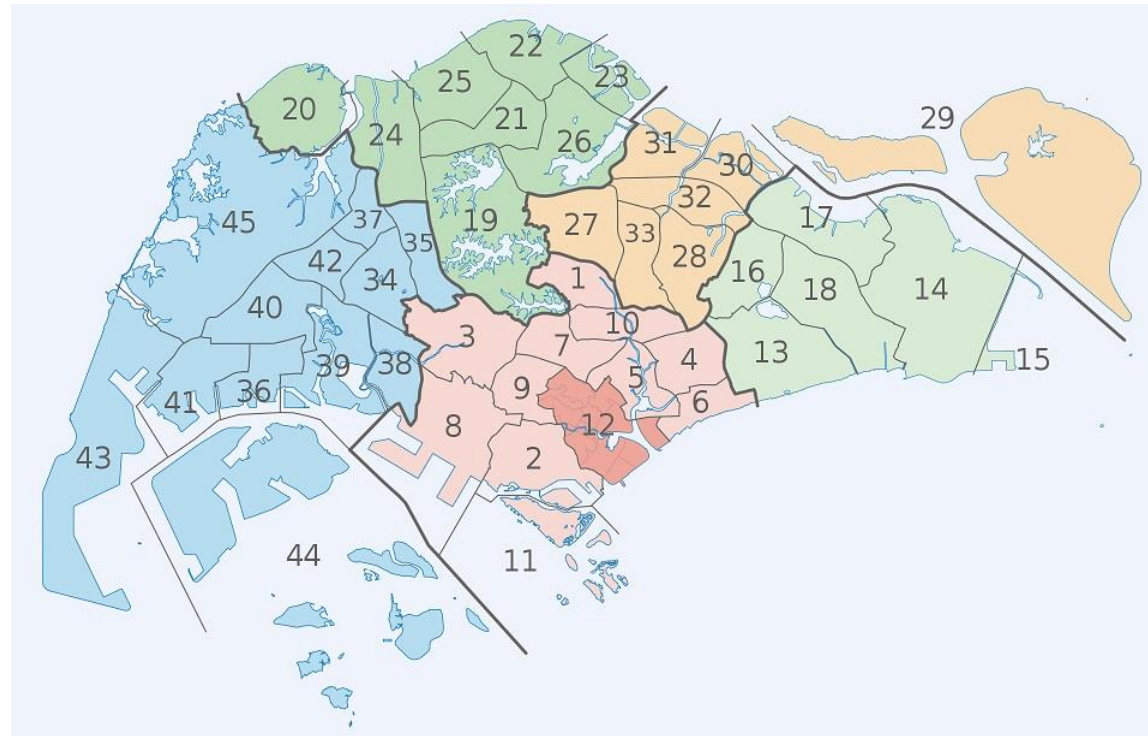




***Which Planning Area to
Live for Young
Immigrant Couple
in Singapore***

Young Immigrants need to choose a district to live

- Singapore keeps attracting a large number of young immigrants every year. There are different districts with different characteristics, new immigrants need to consider where to buy or rent a flat, especially for young couples.



Data Acquisition and Cleaning

- OneMap is the authoritative national map of Singapore with the most detailed and timely updated information developed by the Singapore Land Authority. Use OneMap API to get the multi-polygon boundaries of all planning areas of Singapore. Then, use OneMap API to get the population statistical data of each planning area.
<https://docs.onemap.sg/#planning-area>
<https://docs.onemap.sg/#population-query>
- Use Foursquare API to get the venue information within each planning area.
<https://developer.foursquare.com/docs/api/venues/search>
- Change GeoJSON format returned by OneMap API so that Planning Area Polygons are compatible with Foursquare API or Folium Map

Methodology

- Use Foursquare API to explore the venues inside each planning areas. I adopted the method of searching by polygon area. Then group by planning area and get the mean number of each type of venue.
- For population statistical data of OneMap API, pick average income, average age, tenancy ratio, etc.

Planning Area	Ice Cream Shop	Indian Restaurant	Indonesian Restaurant	Indoor Play Area	Intersection	Island	Italian Restaurant	Japanese Curry Restaurant	Japanese Restaurant	Education	HDB	Tenancy	Married	Age_Per	Income_Per
ANG MO KIO	0.020000	0.03	0.00	0.0	0.0	0.000000	0.010000	0.0	0.030000	0.119718	0.237560	0.123397	0.585885	1.060903	0.921914
BEDOK	0.040000	0.04	0.00	0.0	0.0	0.000000	0.020000	0.0	0.010000	0.159574	0.228261	0.094978	0.598147	1.032859	0.956226
BISHAN	0.060000	0.01	0.00	0.0	0.0	0.000000	0.010000	0.0	0.020000	0.213793	0.353791	0.083333	0.599459	1.020555	1.027919
BOON LAY	0.000000	0.00	0.00	0.0	0.0	0.000000	0.000000	0.0	0.000000	NaN	NaN	NaN	NaN	NaN	NaN
BUKIT BATOK	0.020000	0.02	0.00	0.0	0.0	0.000000	0.030000	0.0	0.000000	0.176707	0.341686	0.057078	0.590517	0.978717	0.985199
BUKIT MERAH	0.027523	0.00	0.00	0.0	0.0	0.009174	0.009174	0.0	0.045872	0.113744	0.260000	0.194545	0.572112	1.079188	0.915892
BUKIT PANJANG	0.010000	0.03	0.01	0.0	0.0	0.000000	0.020000	0.0	0.010000	0.107143	0.378641	0.046117	0.646958	0.933948	1.019983
BUKIT TIMAH	0.040000	0.05	0.00	0.0	0.0	0.000000	0.060000	0.0	0.040000	0.189394	0.038136	0.110638	0.681034	0.999559	1.106425
CENTRAL WATER CATCHMENT	0.020000	0.03	0.00	0.0	0.0	0.000000	0.010000	0.0	0.010000	NaN	NaN	NaN	NaN	NaN	NaN
CHANGI	0.010000	0.03	0.00	0.0	0.0	0.000000	0.000000	0.0	0.010000	NaN	NaN	NaN	NaN	0.872823	NaN

Methodology (continued)

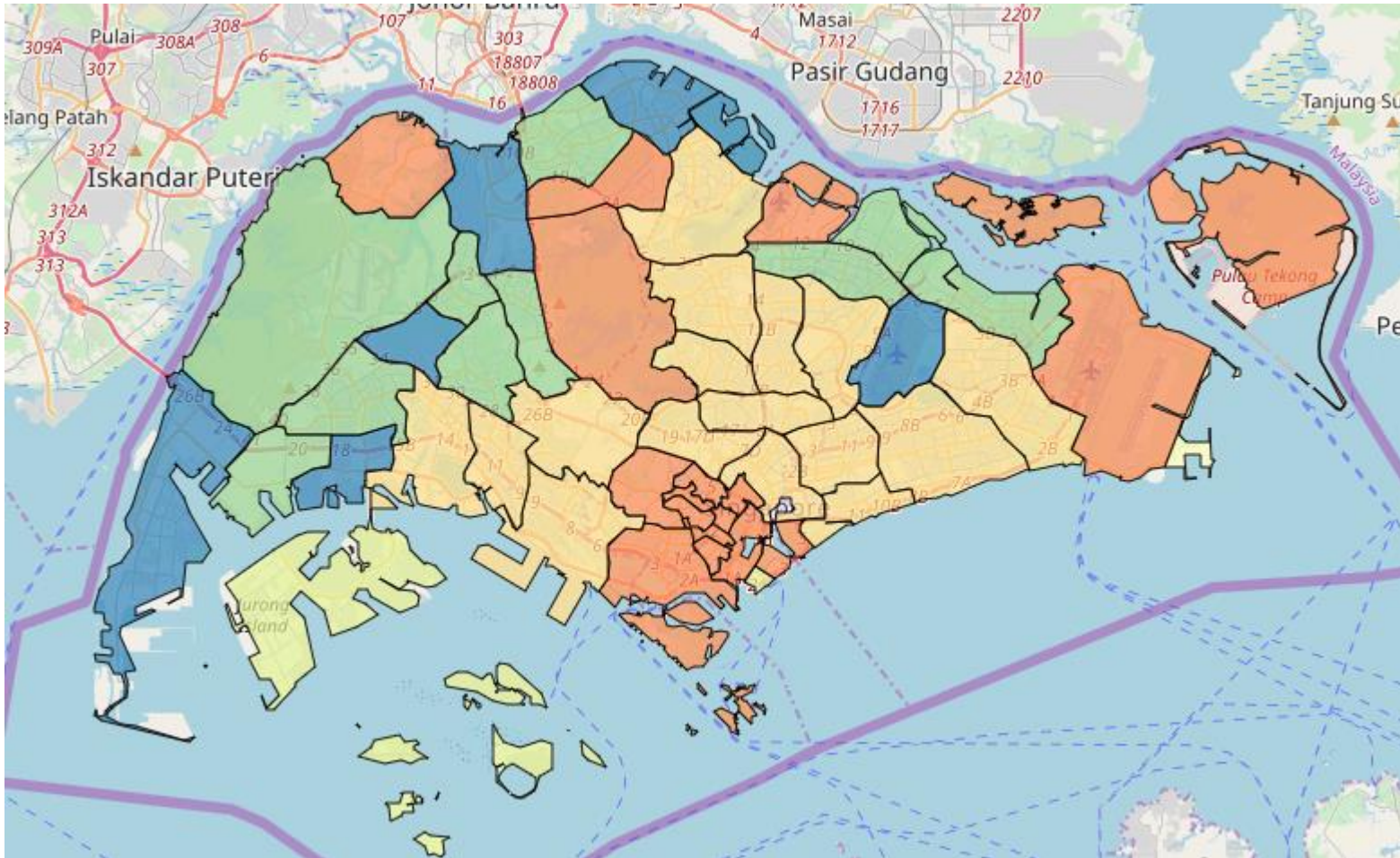
- There some common venue categories in planning areas.
Use unsupervised learning K-means algorithm to cluster the planning areas.
- Run K-Means to cluster the planning areas into 6 clusters.

```
kclusters = 6
# drop the population data because there are too much NaN
singapore_grouped_clustering = singapore_grouped.drop(['P']
kmeans = KMeans(n_clusters=kclusters, random_state=0).fit(
kmeans.labels_
array([[1, 1, 1, 5, 4, 0, 4, 1, 0, 0, 2, 4, 1, 0, 1, 1, 1,
0, 1, 0, 0, 0, 1, 0, 0, 4, 5, 4, 4, 1, 0, 0, 0, 5,
3, 5, 1, 0, 5, 1, 5, 3, 4, 4, 1], dtype=int32))
```

Cluster Labels	Planning Area	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue
1	ANG MO KIO	Food Court	Chinese Restaurant	Coffee Shop	Park	Fast Food Restaurant	Indian Restaurant	Japanese Restaurant
1	BEDOK	Chinese Restaurant	Coffee Shop	Food Court	Dessert Shop	Asian Restaurant	Beach	Seafood Restaurant
1	BISHAN	Chinese Restaurant	Coffee Shop	Food Court	Ice Cream Shop	Café	Thai Restaurant	Park
5	BOON LAY	Exhibit	Boat or Ferry	Bus Station	Bus Stop	Fishing Spot	Scenic Lookout	Coffee Shop
4	BUKIT BATOK	Food Court	Coffee Shop	Fast Food Restaurant	Chinese Restaurant	Supermarket	Café	Gas Station

Methodology (continued)

- Visualize the cluster result by Folium Choropleth map



Result and Conclusion

- Get the top venues and statistical data for all planning areas of each Cluster.
After comparing the 6 clusters, we can conclude that Cluster 5, the Emerging Residential Area are suitable for young immigrant couples.

Cluster 5: Emerging Residential Area, full of Fast Food Restaurants

```
neighborhoods_venues_sorted.loc[neighborhoods_venues_sorted['Cluster Labels'] == 4, neighborhoods_venues_sorted.columns[[1
```

Planning Area	Age_Per	Income_Per	Education	HDB	Tenancy	Married	PublicTranWork	PublicTranSchool	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue
BUKIT BATOK	0.978717	0.985199	0.176707	0.341686	0.057078	0.590517	0.608828	0.745283	Food Court	Coffee Shop	Fast Food Restaurant
BUKIT PANJANG	0.933948	1.019983	0.107143	0.378641	0.046117	0.646958	0.484520	0.664894	Chinese Restaurant	Coffee Shop	Fast Food Restaurant
CHOA CHU KANG	0.924337	1.012545	0.136612	0.425963	0.030612	0.616552	0.624549	0.798561	Coffee Shop	Fast Food Restaurant	Park
JURONG WEST	0.928378	0.999297	0.101449	0.381127	0.062731	0.640598	0.585470	0.795000	Fast Food Restaurant	Food Court	Coffee Shop
PASIR RIS	0.931741	1.024287	0.166113	0.301235	0.049505	0.614853	0.566176	0.744186	Park	Coffee Shop	Fast Food Restaurant
PIONEER	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	Fast Food Restaurant	Food Court
PUNGGOL	0.819803	1.166680	0.065574	0.442771	0.039157	0.771753	0.561587	0.611111	Fast Food Restaurant	Supermarket	Chinese Restaurant
SENGKANG	0.876428	1.067404	0.073434	0.434109	0.035659	0.708817	0.614141	0.667785	Fast Food Restaurant	Food Court	Coffee Shop
WESTERN WATER CATCHMENT	0.929173	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Coffee Shop	Fast Food Restaurant	Gym
WOODLANDS	0.901022	0.970771	0.073864	0.424113	0.062589	0.631959	0.652837	0.810160	Food Court	Fast Food Restaurant	Park

Discussion

- Areas need improvement in the future:
 - The usage of Foursquare in Singapore is not high enough, although higher than most of the other countries. Therefore, the statistical result is not too meaningful.
 - The official statistical data of Singapore is available in many fields, but there is few public data based on planning area, such as the statistical data used in this project, the number is given by interval, so I just take the mean value of the interval, which is not accurate enough.



Thank you