Coursera

IBM Applied Data Science Capstone

Final Project

Find Good Places for Selling Pizza in Bandung, Indonesia

By:

Joshua Abel O.

1 Introduction

Pizza is one of the most popular food in the whole world. A lot of people love to eat pizza, no exception to the city of Bandung. Bandung is known as the culinary city of Indonesia. People from around Indonesia often come to Bandung to enjoy the food.

The objection of this project is to choose which place would be most beneficial for someone who wants to sell pizza in Bandung City. The selection of the place is based on the number of pizza sellers around the area. The best place to sell pizza is the place where no pizza sellers around. This project will able to handle the question: "Where would be most beneficial if someone wants to sell pizza?".

2 Data

The first data used in this project is the district data from Wikipedia (https://id.wikipedia.org/wiki/Daftar_kecamatan_dan_kelurahan_di_Kota_Bandung). The table consists of a larger district (*Kecamatan* in the Indonesian Language) and a smaller district (*Kelurahan* in the Indonesian Language). Bandung City has 30 *Kecamatan* and 152 *Kelurahan.*

The second data is data of longitude and latitude of each *Kelurahan* Head Office from official Bandung City Governor (http://data.bandung.go.id/dataset/0ac32316-8450-4064-b7f2-48049439ff5e/resource/6a568b69-5e41-4ea0-80ca-0487341fe9f3/download/koordinat-dan-ketinggian-kantor-kelurahan-di-kota-bandung-2014.csv)

The last data is data of pizza sellers around each *Kelurahan* head office. This data is collected from foursquare.

3 Methodology

The first data was scrapped by using beautiful soup algorithm from the wikipedia webpage. The data used from this webpage is *Kelurahan* and *Kecamatan* data of Bandung City and another data was removed from dataframe.

```
# remove the unused columns
df.drop('Kode Kemendagri', axis=1, inplace=True)
df.drop('Jumlah Kelurahan', axis=1, inplace=True)
df.drop('Status', axis=1, inplace=True)
df.head()
```

	Kecamatan	Kelurahan
0	Andir	Campaka\nCiroyom\nDunguscariang\nGaruda\nKebon...
1	Astana Anyar	Cibadak\nKaranganyar\nKarasak\nNyengseret\nPan...
2	Antapani	Antapani Kidul\nAntapani Kulon\nAntapani Tenga...
3	Arcamanik	Cisaranten Bina Harapan\nCisaranten Endah\nCis...
4	Babakan Ciparay	Babakan\nBabakanciparay\nCirangrang\nMargahayu...

Figure 1 Dataframe collected from Wikipedia Webpage

From **Figure 1**, multiple *Kelurahan* is located in one cell. The dataframe was processed so each *Kelurahan* separated in each row.

```
# split the Kelurahan data into one row for each kelurahan
new_df = df.Kelurahan.apply(pd.Series)
new_df = new_df.merge(df, left_index=True, right_index=True)
new_df.drop(["Kelurahan"], axis = 1, inplace=True)
new_df = new_df.melt(id_vars = ['Kecamatan'], value_name =
"Kelurahan")
new_df.drop("variable", axis = 1, inplace=True)
new_df.dropna(inplace=True)
new_df.sort_values(by=['Kecamatan'], inplace=True)
new_df.reset_index(drop=True, inplace=True)
new_df.head()
```

	Kecamatan	Kelurahan
0	Andir	Campaka
1	Andir	Ciroyom
2	Andir	Garuda
3	Andir	Kebonjeruk
4	Andir	Maleber

Figure 2 Dataframe after processed.

After that, the data of each *Kelurahan* merged with longitude and latitude data from second dataset.

```
df_merged = pd.merge(new_df, coordinates, how='left',
on='Kelurahan')
df_merged.dropna(inplace=True)
df_merged.drop("Kecamatan_y", axis=1, inplace=True)
df_merged.columns = ['Kecamatan','Kelurahan','Latitude',
'Longitude','Height']
df_merged.drop("Height", axis=1, inplace=True)
df_merged.head()
```

	Kecamatan	Kelurahan	Latitude	Longitude
0	Andir	Campaka	-6.89787	107.56314
1	Andir	Ciroyom	-6.91295	107.58617
2	Andir	Garuda	-6.91596	107.57656
3	Andir	Kebonjeruk	-6.91908	107.60107
4	Andir	Maleber	-6.90734	107.57344

Figure 3 Dataframe after combined by second dataset

Finally, for each longitude and latitude data of each *Kelurahan,* was used to find nearby pizza seller venues from FourSquare. We got data of venue's name, longitude, latitude, and category. The dataframe processed to remove duplicate data in case if one Pizza seller collected from multiple *Kelurahan.*

```
# Drop duplicate value in case if one venue collected from
multiple Kelurahan
final_df = bandung_venues.sort_values('Venue Latitude')
.drop_duplicates(subset='Venue Latitude', keep='first')
.reset_index(drop=True)
print(final_df.shape)
final_df.head()
```

```
(108, 4)
```

	Venue	Venue Latitude	Venue Longitude	Venue Category
0	J-qeeys pizza	-6.945992	107.660345	Pizza Place
1	Pizza Hut	-6.945940	107.641876	Pizza Place
2	Martabak Asia	-6.943416	107.667167	Pizza Place
3	Warung Nasi Suryalaya	-6.943411	107.623078	Pizza Place
4	Pizza Hut Buah Batu	-6.943380	107.674318	Pizza Place

Figure 4 Final dataframe

The final dataframe was used to produce a map of pizza seller. The map is used to produce a heat map based on the number of pizza seller around. Based on the intensity of heat map, can be concluded which area will be the best suitable place to sell pizza.

4 Results

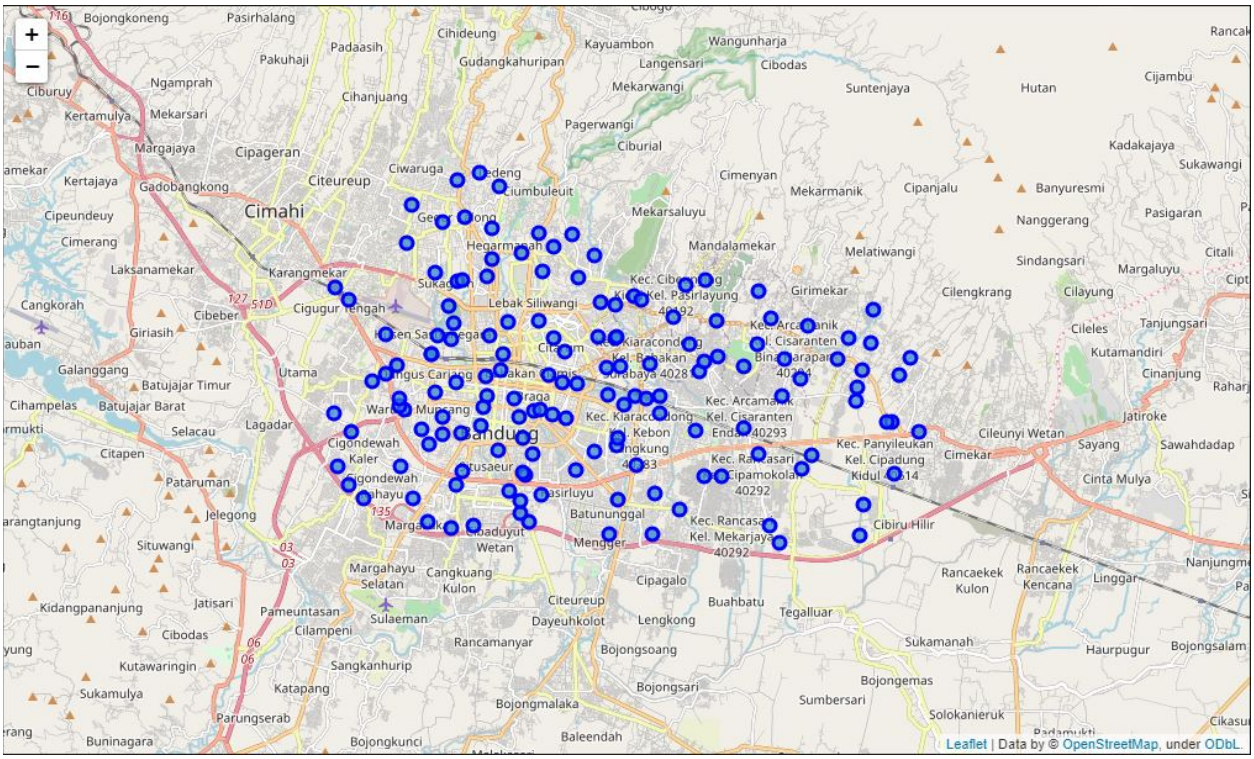

Figure 5 Bandung City Map of *Kelurahan*

Figure 5 show mapping for each *Kelurahan* head office. From this figure, we can see the distribution of *Kelurahan* in Bandung City is not really balance from the area. Some districts may have larger area and some districts may have smaller area. That's why it is needed to process the data in case if there is duplicated pizza seller from multiple *Kelurahan.*

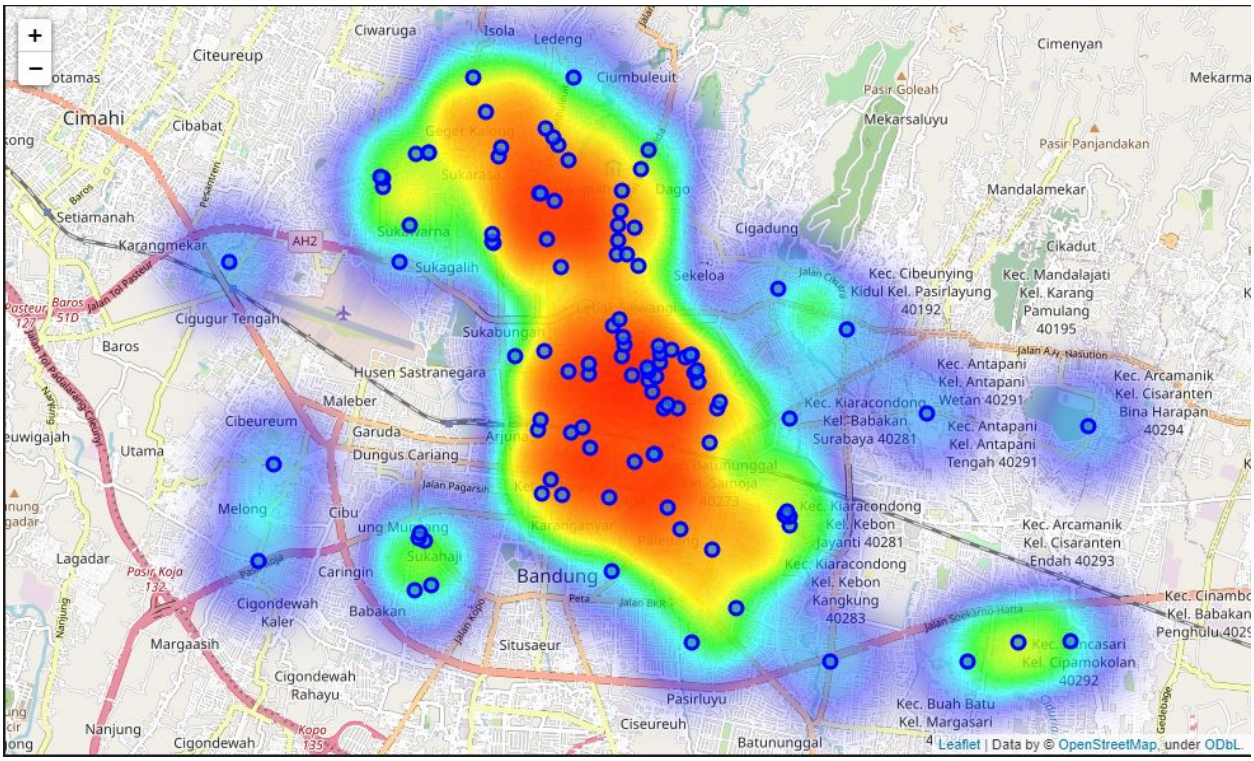

Figure 6 Heat Map of Bandung City for Pizza Seller Distribution

	Venue Category
Pizza Place	75
Café	7
Coffee Shop	4
Bakery	4
Steakhouse	3
American Restaurant	2
Eastern European Restaurant	2
Bar	2
French Restaurant	1
Lounge	1
Cocktail Bar	1
College Science Building	1
Burger Joint	1
Italian Restaurant	1
Japanese Restaurant	1
Food Court	1
Diner	1

Figure 7 Number of Pizza Sellers of Each Category

From **Figure 7**, we can see that most of Pizza Sellers is Pizza Place.

	Venue
Pizza Hut	9
PHD (Pizza Hut Delivery)	5
Domino's Pizza	3
Ngopi Doeloe	2
PizzaHut	2
Magic Pizza	2
Bober Cafe	1
Fakultas Kedokteran	1
Tizi's Restaurant & Bar	1
Pizza Hermes	1

Figure 8 Top 10 Pizza Sellers franchise in Bandung City

From **Figure 8,** we can see that Pizza Hut is placed in 9 area of Bandung City.

5 Discussion

From the result from section 4, Figure 6, we can see that the distribution of pizza seller is focused in the middle of city. Meanwhile only some pizza seller at the east and west side of Bandung city. Based on that, the best place to sell pizza will be at east or west side of Bandung city, as you can see in Figure 9 below.

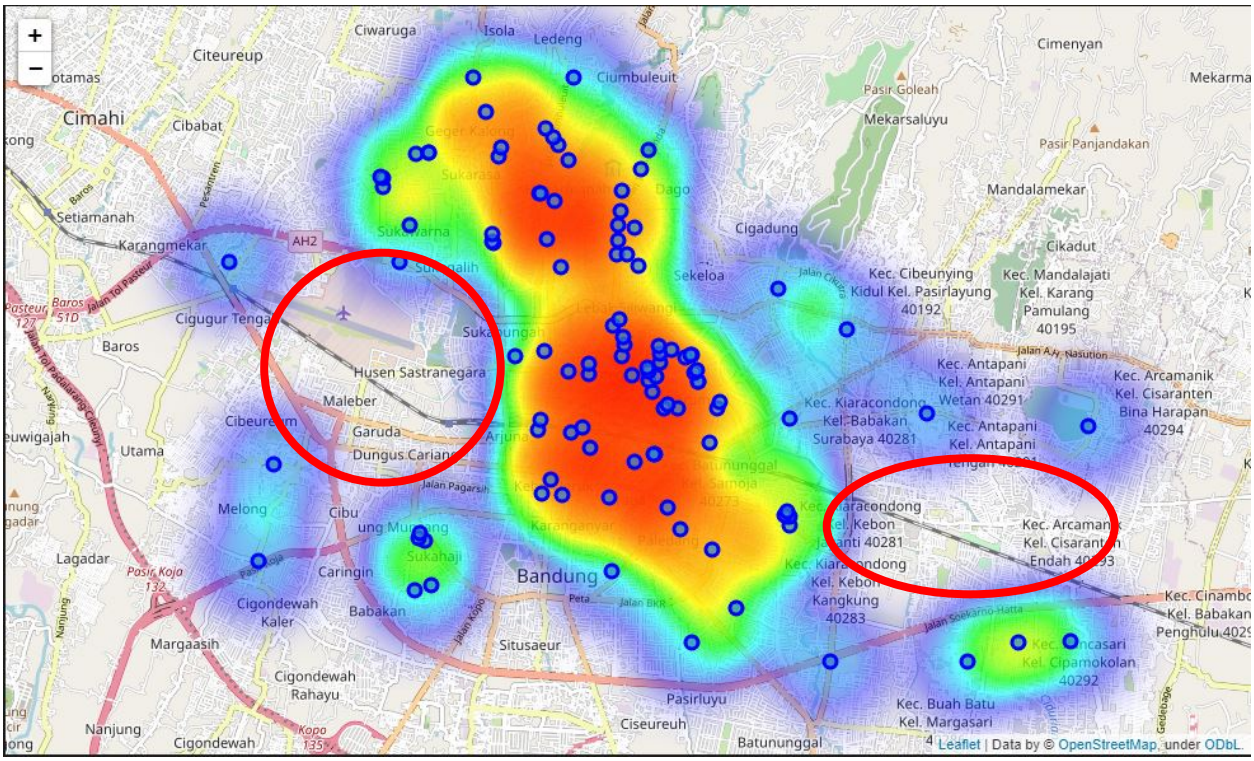

Figure 9 Best area to sell pizza

6 Conclusion

The most beneficial area to sell pizza is shown in the red circle in Figure 9, located in the east and west side of the Bandung City.