

Where to Open an Italian Restaurant in Boston?



DENİZ AYTEMİZ



- In this project; I examined the areas of Boston city for a person considering opening an Italian restaurant.

- Data on number of Italian restaurants around the area, the popularity of the area, safety of the area, the housing prices of the area and the restaurant rankings around the area will give a description of the areas.

- These areas are determined by distinct postal codes.

Data Acquisition and Cleaning



- Foursquare data for number of Italian restaurants in a postal code area, id of venues, latitude and longitude, ratings of Italian Restaurants and number of tips.
- House prices in Boston and Crime Incidents in Boston datasets from www.boston.gov.tr

	PID	CM_ID	GIS_ID	ST_NUM	ST_NAME	ST_NAME_SUF	UNIT_NUM	ZIPCODE	PTYPE	LU	...	U_BTH_STYLE2	U_BTH_STYLE3	U_KITCH_TYPE	U_KITCH_STYLE	U_HEAT_TYP	U_AC	U_FPLACE	U_INT_FIN	U_INT_CND	U_VIEW
0	502550008	502550000.0	502550000	87	BEACON	ST	2-F	2108.0	102	CD	...	M	NaN	O	M	W	N	1.0	N	G	A
1	502550010	502550000.0	502550000	87	BEACON	ST	2-R	2108.0	102	CD	...	NaN	NaN	O	S	W	N	1.0	N	A	A
2	502550012	502550000.0	502550000	87	BEACON	ST	3-F	2108.0	102	CD	...	S	NaN	O	M	W	N	1.0	N	A	G
3	502550014	502550000.0	502550000	87	BEACON	ST	3-R	2108.0	102	CD	...	NaN	NaN	O	N	W	N	1.0	S	A	G
4	502550016	502550000.0	502550000	87	BEACON	ST	4	2108.0	102	CD	...	M	M	O	L	W	C	2.0	N	E	G

INCIDENT_NUMBER	OFFENSE_CODE	OFFENSE_CODE_GROUP	OFFENSE_DESCRIPTION	DISTRICT	REPORTING_AREA	SHOOTING	OCCURRED_ON_DATE	YEAR	MONTH	DAY_OF_WEEK	HOURL	UCR_PART	STREET	Lat	Long	Location	
0	TESTTEST2	423	NaN	ASSAULT - AGGRAVATED	External		0	2019-10-18 00:00:00	2019	10	Wednesday	0	NaN	RIVERVIEW DR	NaN	NaN	(0.00000000, 0.00000000)
1	S97333701	3301	NaN	VERBAL DISPUTE	C6	915	0	2020-07-18 14:34:00	2020	7	Saturday	14	NaN	MARY BOYLE WAY	42.330813	-71.051368	(42.33081300, -71.05136800)
2	S47513131	2647	NaN	THREATS TO DO BODILY HARM	E18	530	0	2020-06-24 10:15:00	2020	6	Wednesday	10	NaN	READVILLE ST	42.239491	-71.135954	(42.23949100, -71.13595400)
3	I92102201	3301	NaN	VERBAL DISPUTE	E13	583	0	2019-12-20 03:08:00	2019	12	Friday	3	NaN	DAY ST	42.325122	-71.107779	(42.32512200, -71.10777900)
4	I92097173	3115	NaN	INVESTIGATE PERSON	C11	355	0	2019-10-23 00:00:00	2019	10	Wednesday	0	NaN	GIBSON ST	42.297555	-71.059709	(42.29755500, -71.05970900)

Final Dataframe after cleaning and merging with other dataframes

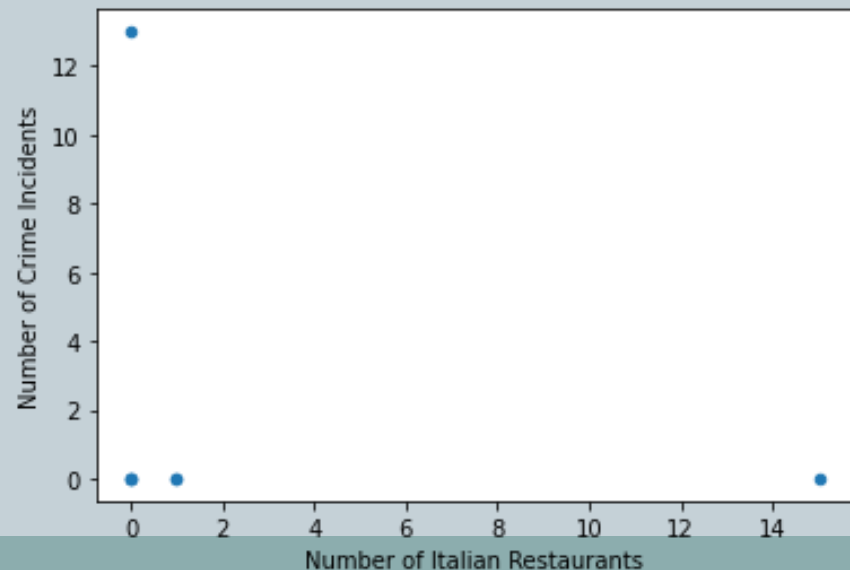
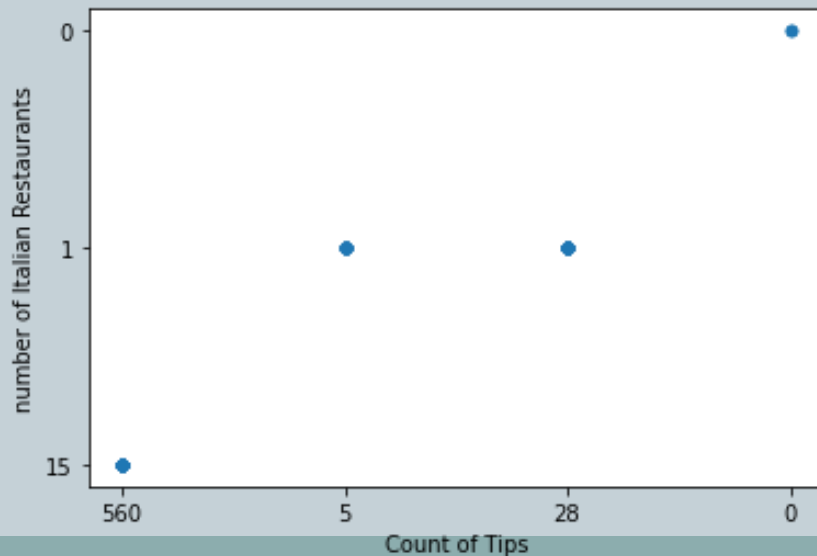


	postalCode	Average Ranking of Italian Restaurant_x	Count of Tips_x	Average of House Prices	Number of Crime Incidents	Number of Italian Restaurants
0	02113	8.2	560	981968.24	0	15
1	02113	8.2	560	981968.24	0	15
2	02113	8.2	560	981968.24	0	15
3	02113	8.2	560	981968.24	0	15
4	02113	8.2	560	981968.24	0	15
...
90	02109	8.6	28	981968.24	0	1
91	02114	0.0	0	981968.24	13	0
92	02114	0.0	0	981968.24	13	0
93	02110	0.0	0	541200.00	0	0
94	02110	0.0	0	541200.00	0	0

Feature Set



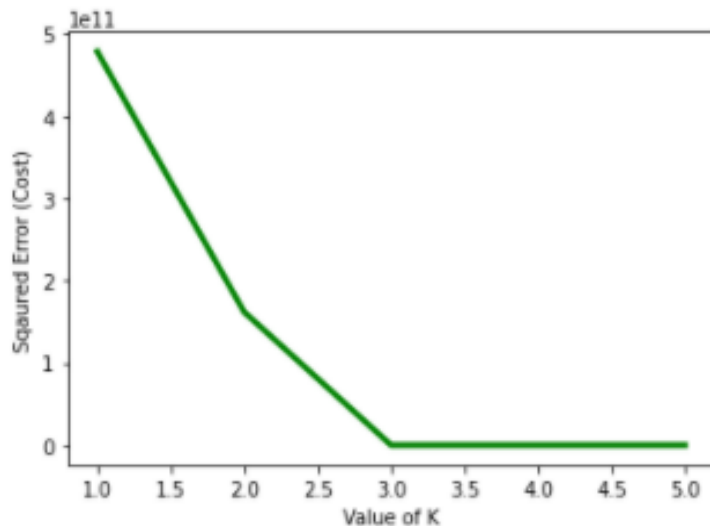
- Feature set is number of Italian Restaurants, Number of Crime Incidents, Average House Price, Total Number of Tips, Average ratings of Italian Restaurants.
- Correlation among Features:
- Positive correlation between count of tips and number of Italian restaurants and no correlation between crime rates and number of Italian restaurants.



Model



- K-means for modelling is used, since the aim is to cluster the postal code areas into groups.
- Selecting the optimal k value:



Elbow point is $k=3$, hence that is the optimal value.

Model Result



- What this labels indicate can be interpreted from the below table;

	Average Ranking of Italian Restaurant	Count of Tips	Average of House Prices	Number of Crime Incidents	Number of Italian Restaurants
Labels					
0	0.00	0.0	871776.180	3.25	0.0
1	8.45	16.5	1249200.785	0.00	1.0
2	8.20	560.0	981968.240	0.00	15.0

- Label 0 can be interpreted as, low ranking, low popularity, cheaper house prices, high rates of crime with very few Italian Restaurants.
- Label 1 indicates places with, very high rankings, average popularity, expensive housing prices, minimum crime incidents and few Italian Restaurants.
- Label 2 indicates popular areas with high rankings, average housing prices, low rates of crime and many Italian Restaurants.

The Map after the labels are assigned to the postal codes

