Ideal Locations for New Restaurant Business in Berlin

Sancak Özdemir

Data Science Certificate Program-Capstone Project

Content

- Introduction
- Data gathering and preliminary analyses
- Methodology
  - Explore the Neighborhoods of Berlin
  - Analyze and Cluster Neighborhoods
  - Analyze and Make Prediction Models
- Results and Discussion
- Conclusion

Introduction

- importance of location in restaurant business
- Why Berlin
  - Berlin is the capital and largest city of Germany by both area and population.
  - Its economy is based on high-tech firms and the service sector,
  - A continental hub for air and rail traffic
  - Hosting many universities, museums, movie theaters, and diverse historical and cultural places.

Introduction (Cont'd)

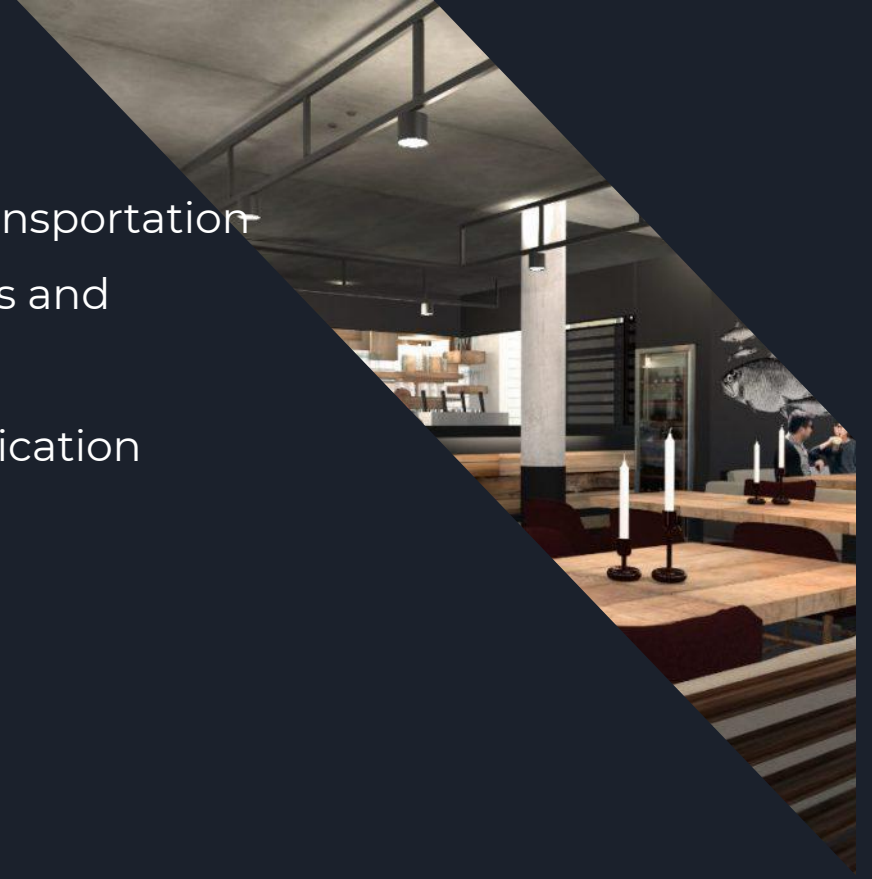
Target:

- Classifying neighborhoods
  - Based on restaurant venues
    - K-means clustering model
- Making prediction models
  - Estimating the ratings
    - Multiple linear regression model
    - Ridge regression model

Main data source: Foursquare API

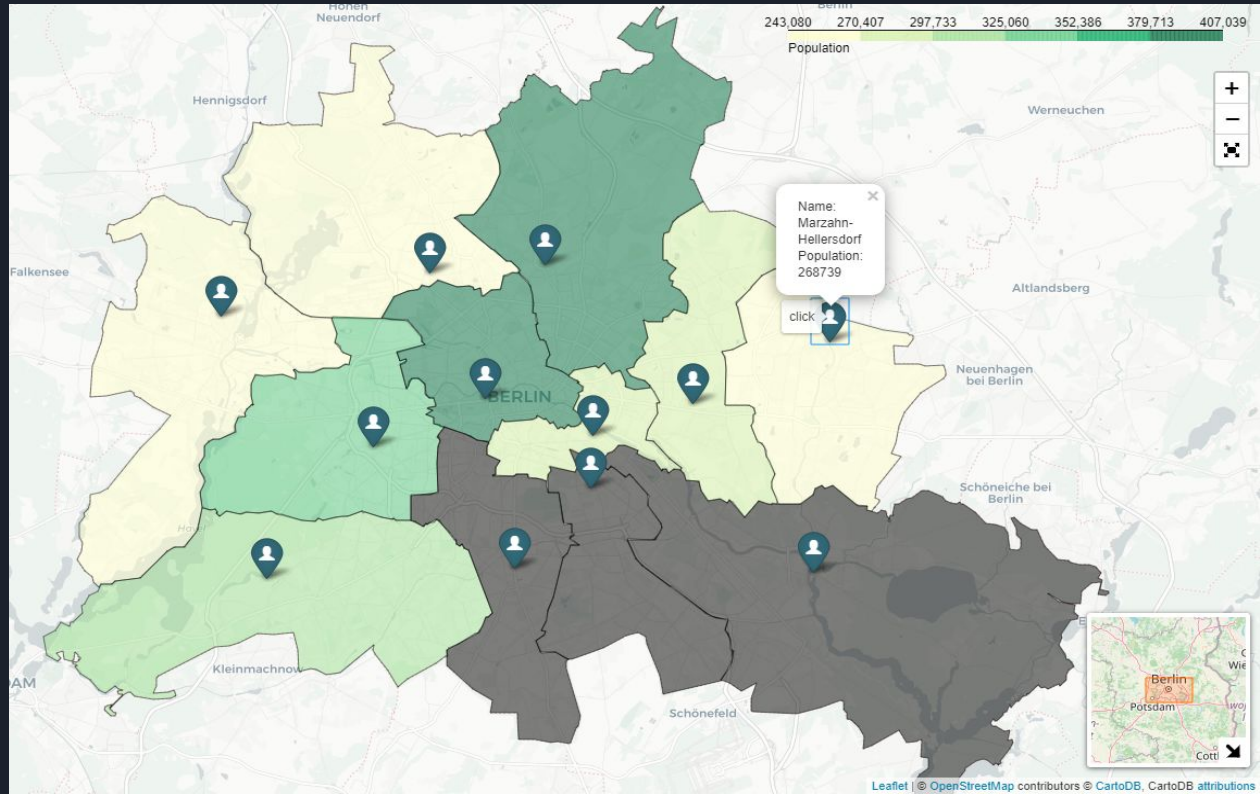
Data: What we need

- Neighborhoods
- Locations of stadiums, theaters, transportation hubs, airports, malls, parking places and universities
- General population and age classification
  - 15-35 (Fast-food)
  - 25-45 (Bar-Bistro)
  - 30-50 (Casual dining)
  - 35-65 (Fine dining)
  - 65+
- Male/female distribution
- Ethnicity (or domestic and foreign population)
- Cuisine types and/or restaurant styles, etc



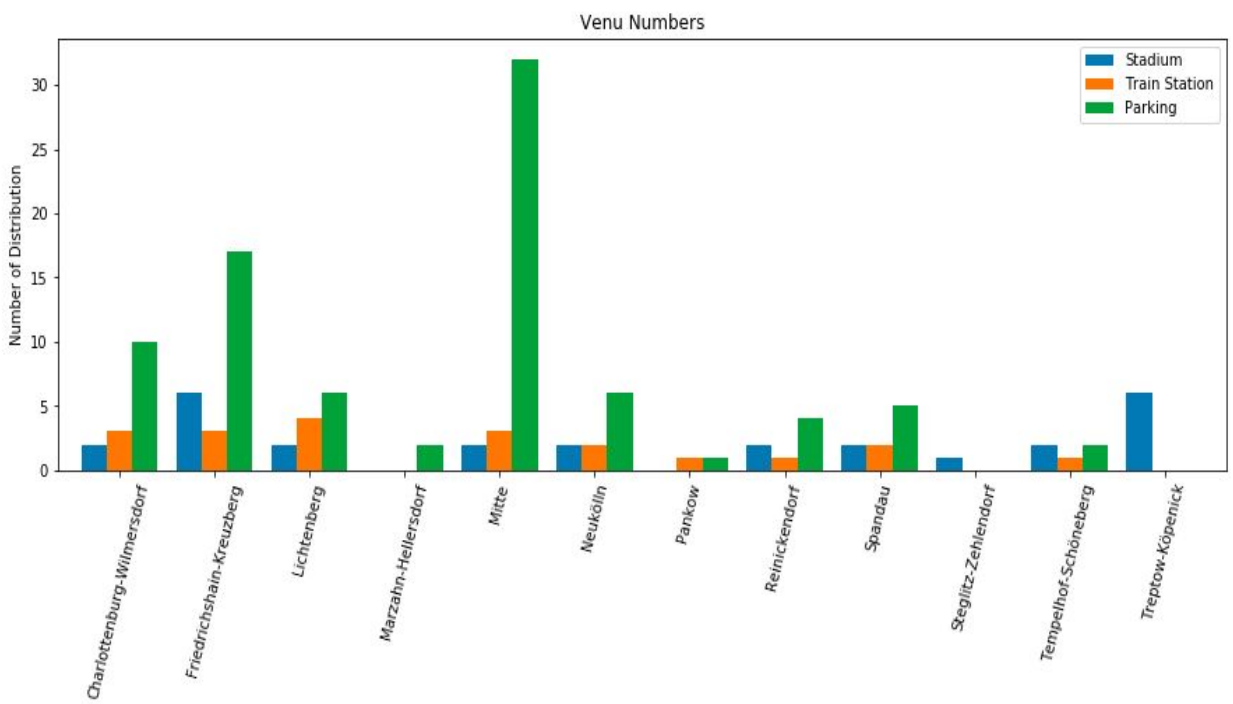
Neighborhoods of Berlin

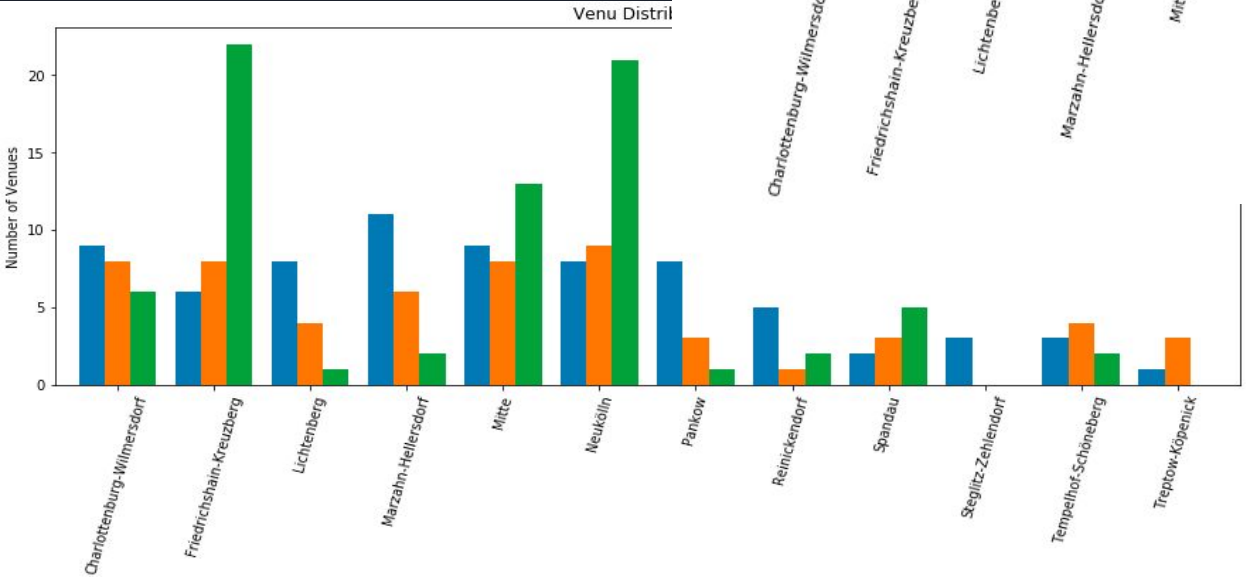
Neighborhood	Population
Charlottenburg-Wilmersdorf	319.628
Friedrichshain-Kreuzberg	268.225
Lichtenberg	259.881
Marzahn-Hellersdorf	248.264
Mitte	332.919
Neukölln	310.283
Pankow	366.441
Reinickendorf	240.454
Spandau	223.962
Steglitz-Zehlendorf	293.989
Tempelhof-Schöneberg	335.060
Treptow-Köpenick	241.335



Number of;

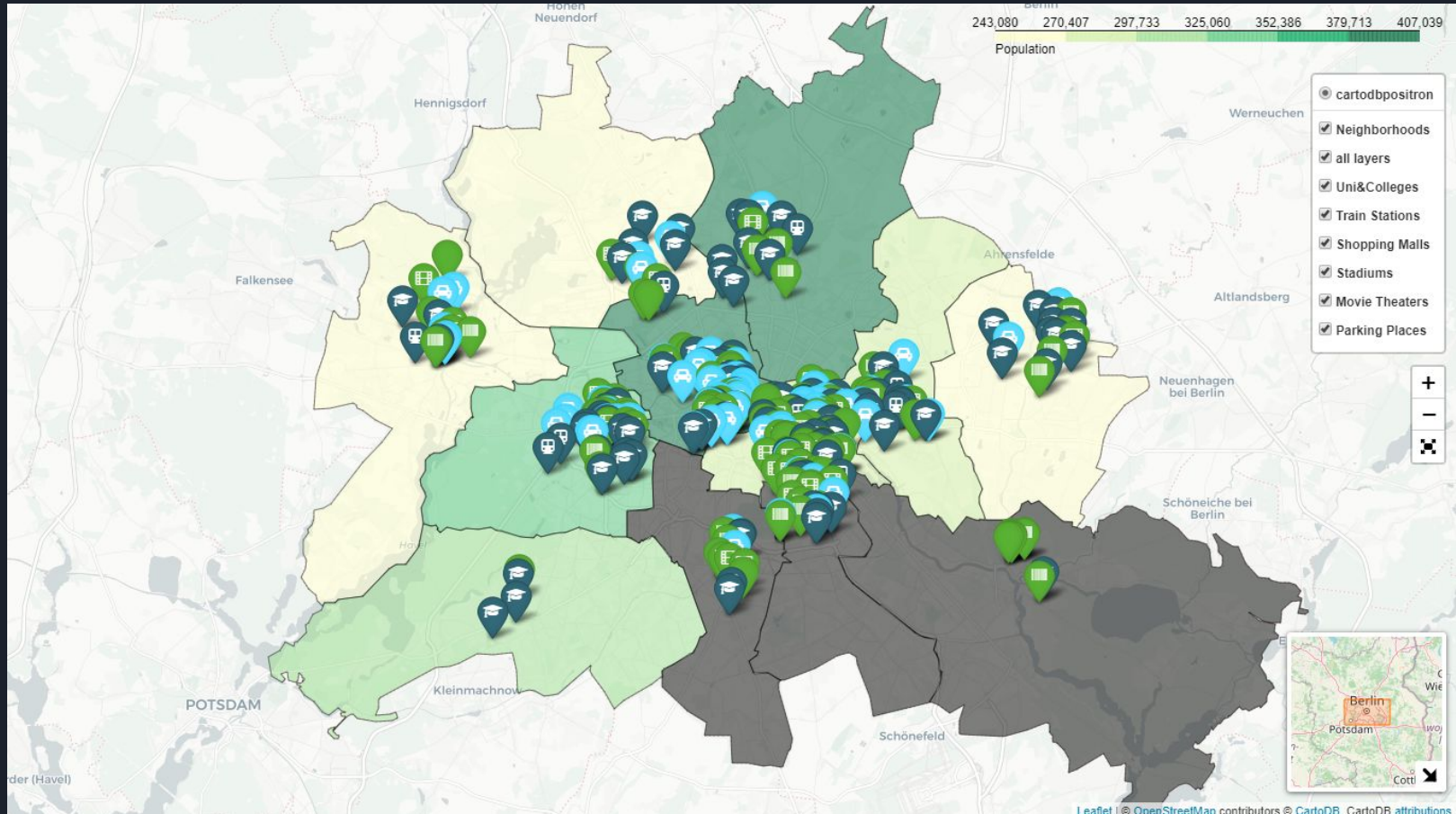
- ❑ Stadiums
- ❑ Train stations,
- ❑ Parking places





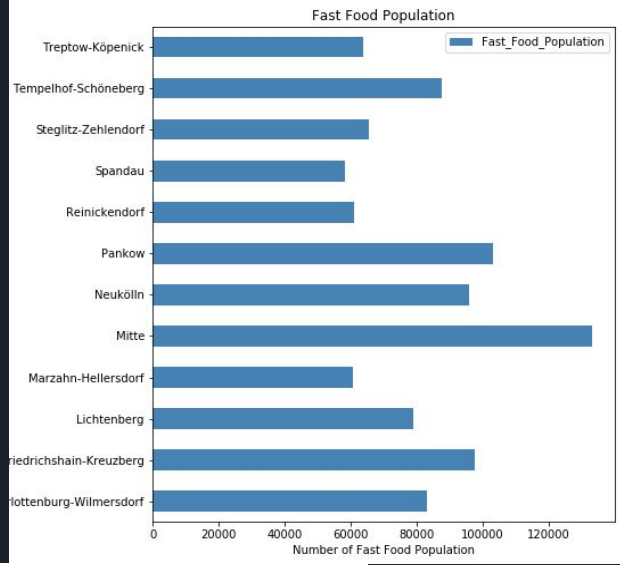
- ❑ Universities&Colleges,
- ❑ Shopping Malls,
- ❑ Movie Theaters

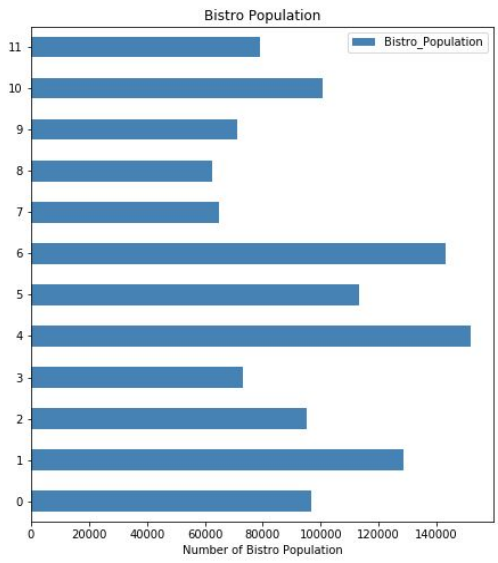
People gathering places shown on the interactive map

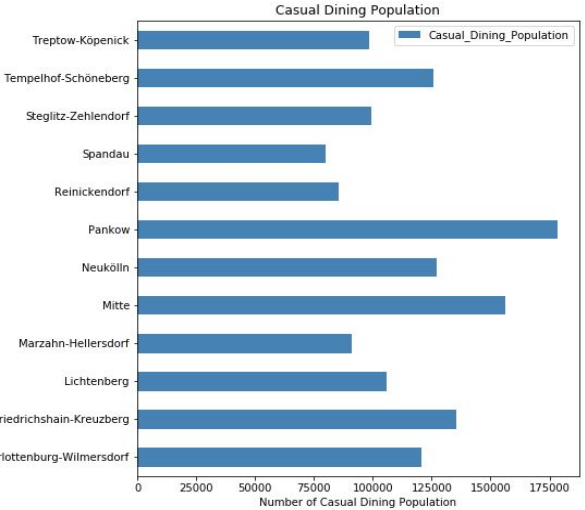


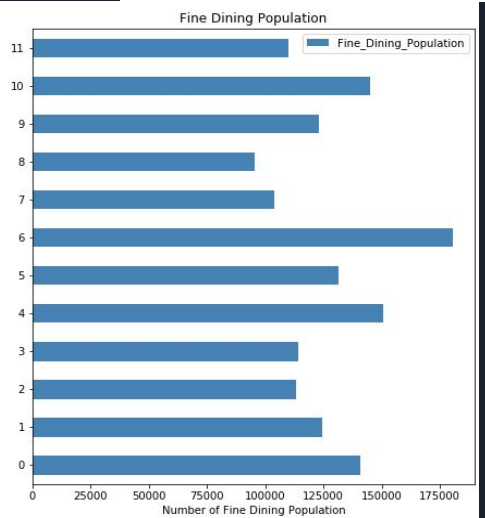
Number of;

- ❏ Fast food population
- ❏ Bistro population





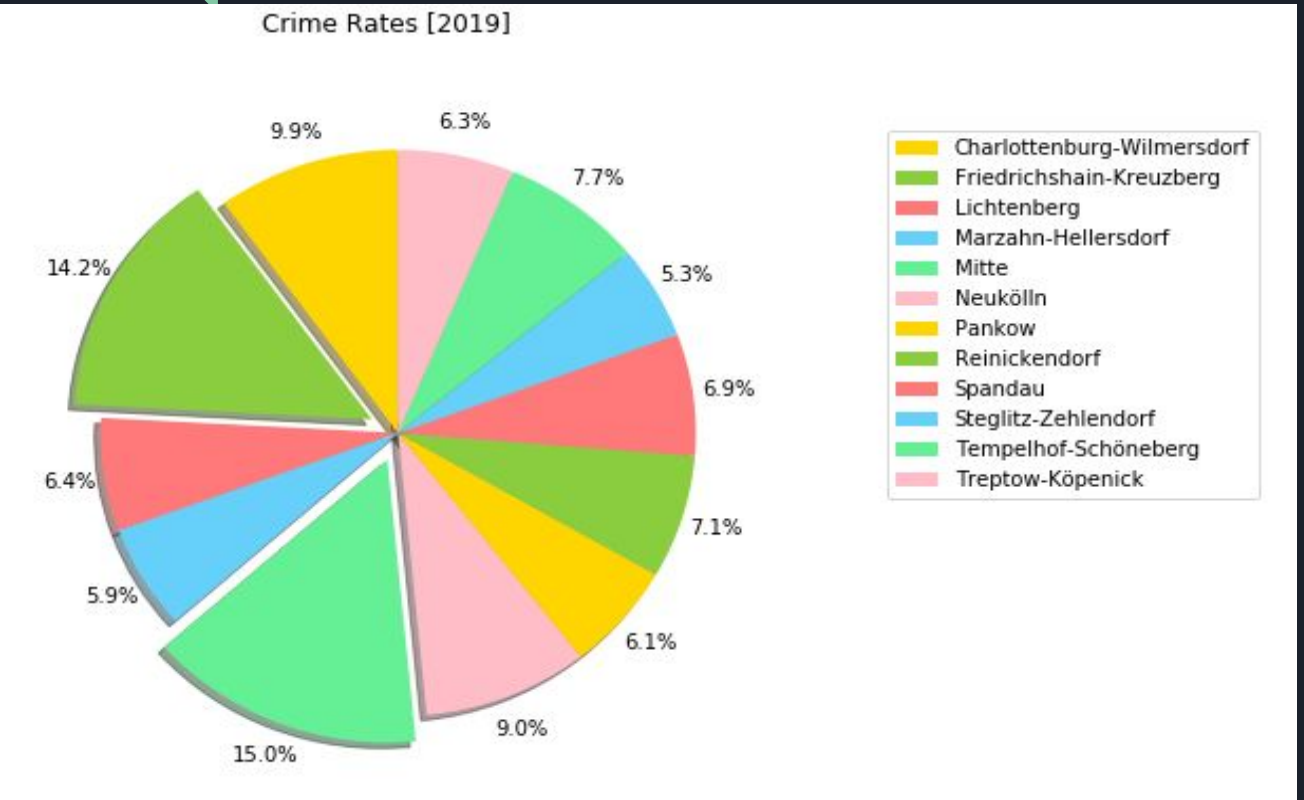




Number of;

- ❏ Casual dining population
- ❏ Fine dining population

Crime rates of Neighborhoods of Berlin



Restaurants related information

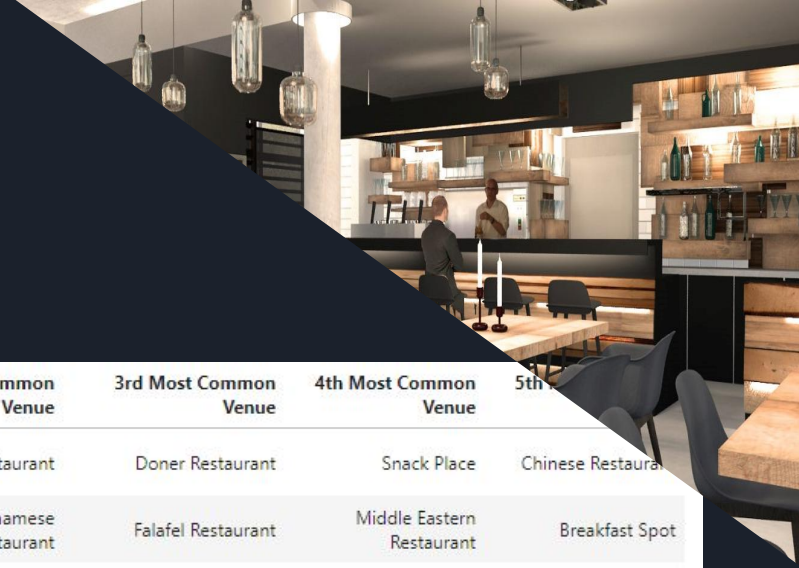
Neighborhood	546 non-null object
Neighborhood Latitude	546 non-null float64
Neighborhood Longitude	546 non-null float64
Venue Id	546 non-null object
Venue	546 non-null object
Venue Latitude	546 non-null float64
Venue Longitude	546 non-null float64
Venue Distance	546 non-null int64
Venue Category	546 non-null object
Price	546 non-null int64
Rating	546 non-null float64
Likes	546 non-null int64
Borough	546 non-null object
Population	546 non-null float64
Latitude	546 non-null float64
Longitude	546 non-null float64
Uni&College	546 non-null int64
Shopping Mall	546 non-null float64
Movie Theater	546 non-null float64
Stadium	546 non-null float64
Train Station	546 non-null float64
Parking	546 non-null float64
Crime_numbers	546 non-null int64
Male_Population	546 non-null int64
Female_Population	546 non-null int64
Foreign_Population	546 non-null int64
German_Population	546 non-null int64
FasFood_Population	546 non-null int64
Bistro_Population	546 non-null int64
Casual_Dining_Population	546 non-null int64
Fine_Dining_Population	546 non-null int64

Restaurant Categories

```
array(['Seafood Restaurant', 'German Restaurant', 'Thai Restaurant',
       'Italian Restaurant', 'Doner Restaurant', 'Gourmet Shop',
       'Snack Place', 'Breakfast Spot', 'Fast Food Restaurant',
       'Japanese Restaurant', 'Chinese Restaurant', 'Steakhouse',
       'Burger Joint', 'Sushi Restaurant', 'French Restaurant',
       'Greek Restaurant', 'Food Court', 'Kofte Place', 'Pizza Place',
       'Falafel Restaurant', 'Middle Eastern Restaurant',
       'Vietnamese Restaurant', 'Gastropub', 'Korean Restaurant',
       'Sandwich Place', 'Lebanese Restaurant', 'Ramen Restaurant',
       'Mexican Restaurant', 'Deli / Bodega', 'Pakistani Restaurant',
       'Restaurant', 'Asian Restaurant', 'Turkish Restaurant',
       'BBQ Joint', 'Indian Restaurant', 'Diner', 'Persian Restaurant',
       'Vegetarian / Vegan Restaurant', 'Bistro', 'Bavarian Restaurant',
       'Currywurst Joint', 'Brasserie', 'Cocktail Bar',
       'Trattoria/Osteria', 'Kebab Restaurant', 'Argentinian Restaurant',
       'Mediterranean Restaurant', 'Brewery', 'Food & Drink Shop',
       'North Indian Restaurant', 'Taverna', 'Food Stand', 'Taco Place',
       'Burrito Place', 'Eastern European Restaurant', 'Hot Dog Joint'],
      dtype=object)
```

Classifying the Neighborhoods

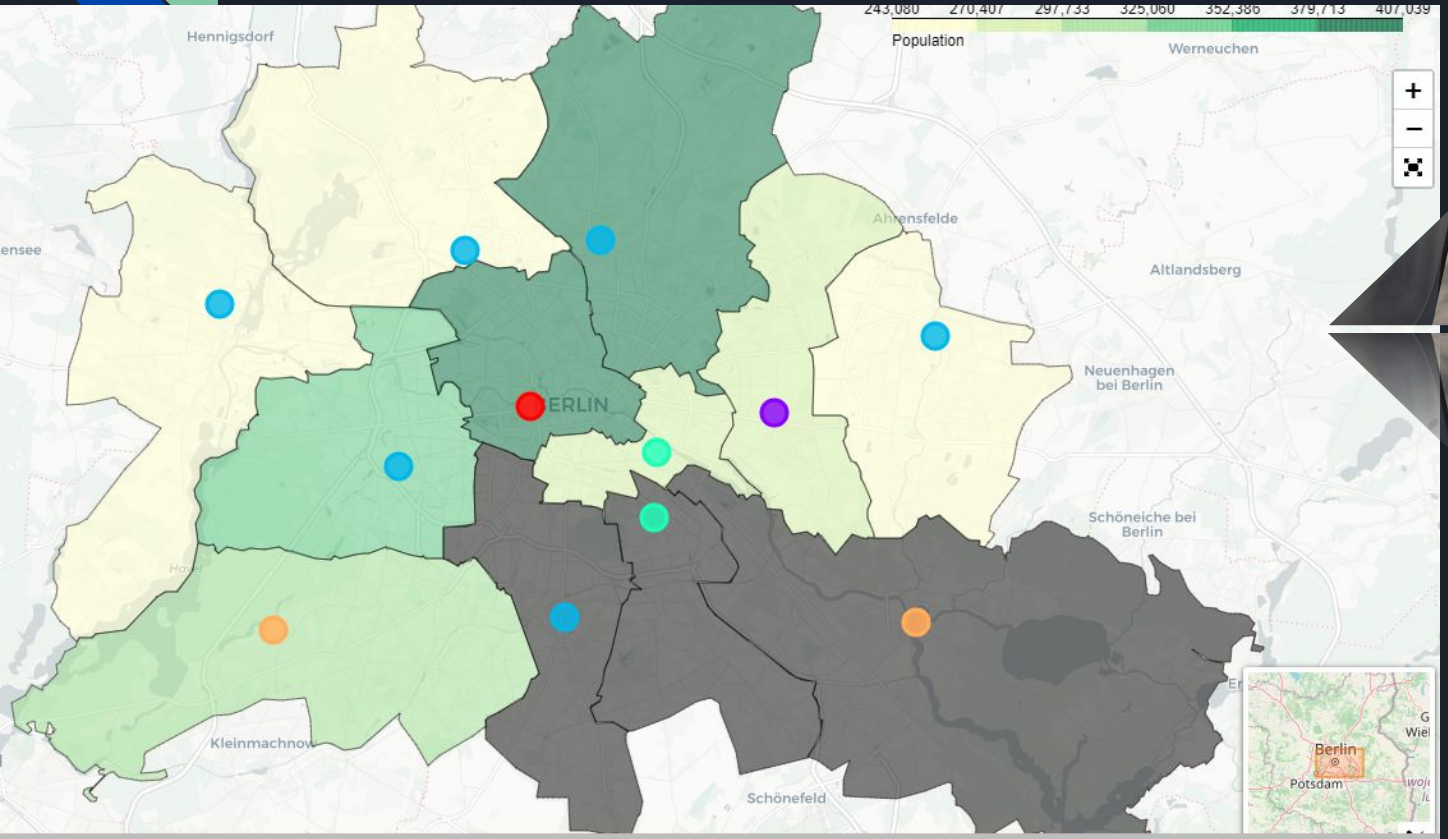
K-means clustering model

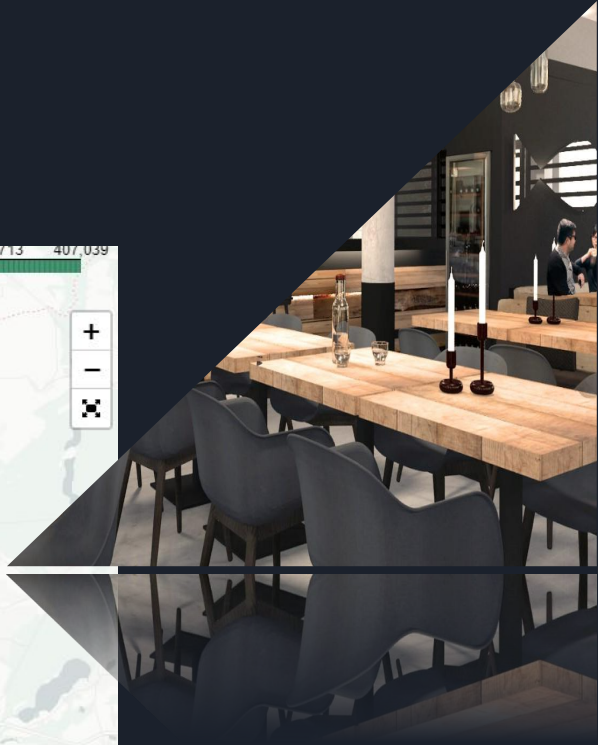


	Borough	Neighborhood	Population	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th
0	Berlin	Charlottenburg-Wilmersdorf	341327	52.497058	13.296490	2	Japanese Restaurant	Sushi Restaurant	Doner Restaurant	Snack Place	Chinese Restaura
1	Berlin	Friedrichshain-Kreuzberg	289120	52.501500	13.435120	3	Pizza Place	Vietnamese Restaurant	Falafel Restaurant	Middle Eastern Restaurant	Breakfast Spot
2	Berlin	Lichtenberg	290493	52.514581	13.498392	1	Vietnamese Restaurant	Pizza Place	Asian Restaurant	Currywurst Joint	Gastropub
3	Berlin	Marzahn-Hellersdorf	268739	52.539720	13.584280	2	Doner Restaurant	Snack Place	Italian Restaurant	Restaurant	Asian Restaurant
4	Berlin	Mitte	383457	52.516740	13.366790	0	Fast Food Restaurant	Burger Joint	Vegetarian / Vegan Restaurant	German Restaurant	Sandwich Place
5	Berlin	Neukölln	330786	52.480200	13.433640	3	Pizza Place	Breakfast Spot	Korean Restaurant	Middle Eastern Restaurant	Bistro
6	Berlin	Pankow	407039	52.571050	13.404970	2	Breakfast Spot	Fast Food Restaurant	Trattoria/Osteria	Italian Restaurant	Burger Joint
7	Berlin	Reinickendorf	264826	52.567550	13.331650	2	Fast Food Restaurant	Doner Restaurant	Trattoria/Osteria	Italian Restaurant	Restaurant
8	Berlin	Spandau	243080	52.550090	13.200356	2	Turkish Restaurant	Italian Restaurant	Fast Food Restaurant	Doner Restaurant	German Restaurant
9	Berlin	Steglitz-Zehlendorf	308077	52.443640	13.229080	4	Italian Restaurant	German Restaurant	Greek Restaurant	Asian Restaurant	Trattoria/Osteria
10	Berlin	Tempelhof-Schöneberg	351429	52.447630	13.385350	2	Restaurant	Indian Restaurant	Taverna	German Restaurant	Gastropub
11	Berlin	Treptow-Köpenick	269775	52.445817	13.574580	4	German Restaurant	Fast Food Restaurant	Italian Restaurant	Greek Restaurant	Asian Restaurant

Classifying the Neighborhoods

K-means clustering model



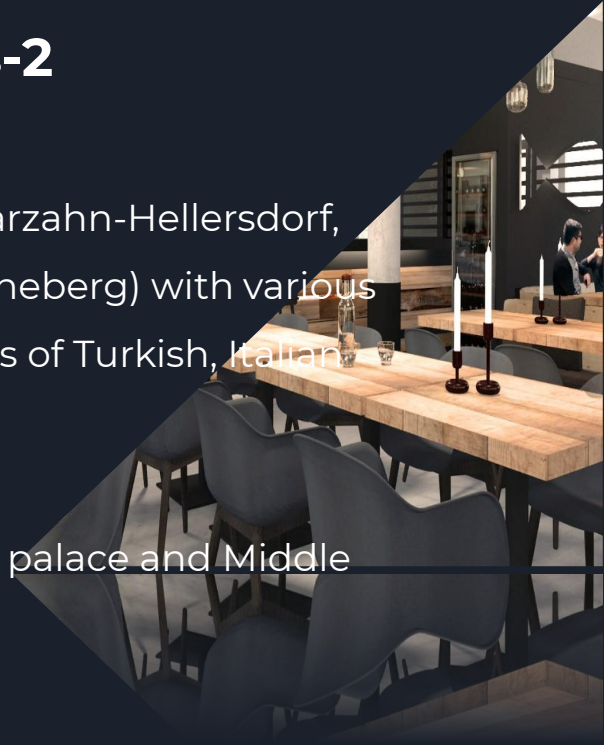


K-means clustering model-Results-1

- **Cluster-1**
  - Only Mitte region is classified in this cluster
    - Most common venues in the top two are fast-food restaurants.
    - No Asian restaurants, Asian restaurant business can be promising.
- **Cluster-2**
  - Just the Lichtenberg region is classified in this cluster
    - The most common restaurant venue categories in this neighborhood are Vietnamese Restaurant, Pizza Places and Asia restaurants.
    - Saturation of Asian restaurants (Vietnamese Restaurant)

K-means clustering model-Results-2

- **Cluster-3**.
  - 5 neighborhoods (Charlottenburg-Wilmersdorf, Marzahn-Hellersdorf, Pankow, Reinickendorf, Spandau, Tempelhof-Schöneberg) with various types of cuisines, especially with the commonalities of Turkish, Italian and Fast Food restaurants.
- **Cluster-4**
  - Friedrichshain-Kreuzberg and Neukölln have Pizza palace and Middle east Restaurants in common.
- **Cluster-5**
  - Finally, we see that Italian, German and Greek restaurants are in common
- In addition to this clustering study, the population, number of these venues, number and location of touristic regions must be taken into account.

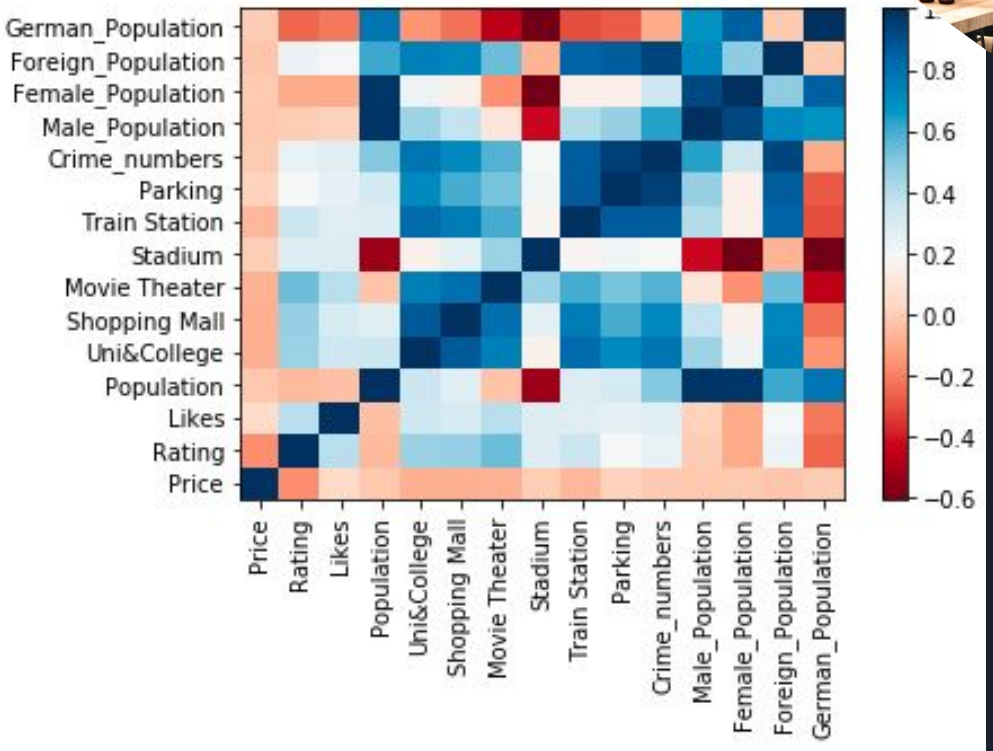


Multiple regression model
Prediction of Rating

Determination of independent and dependent variables

```
dfnew.corr()['Rating'].sort_values()
```

```
German_Population     -0.249195
Price                 -0.159745
Female_Population     -0.093123
Population            -0.046426
Male_Population       -0.000610
Parking                0.203134
Foreign_Population     0.241458
Crime_numbers          0.250090
Stadium                0.287544
Train Station          0.350554
Likes                  0.392928
Uni&College            0.456468
Shopping Mall          0.471028
Movie Theater          0.545049
Rating                 1.000000
Name: Rating, dtype: float64
```



Multiple regression model-Results-1

- **Multiple Regression Model Coefficients:**

  [9.71291557e-04 2.92238414e+09 1.70648096e-01 -1.65198471e-01
  3.02673696e-02 4.61008151e-01 7.13941708e-02 5.72347906e-02
  -1.51110784e-04 -2.92238414e+09 -2.92238414e+09]

- **The R-square:** 0.4126715675897107
- **Residual sum of squares:** 0.86 and
- **Variance score:** 0.22
- **The mean square error of 'Rating' and predicted value:**
  0.7654593187366441

Multiple regression model-Results-2

(With normalized data)

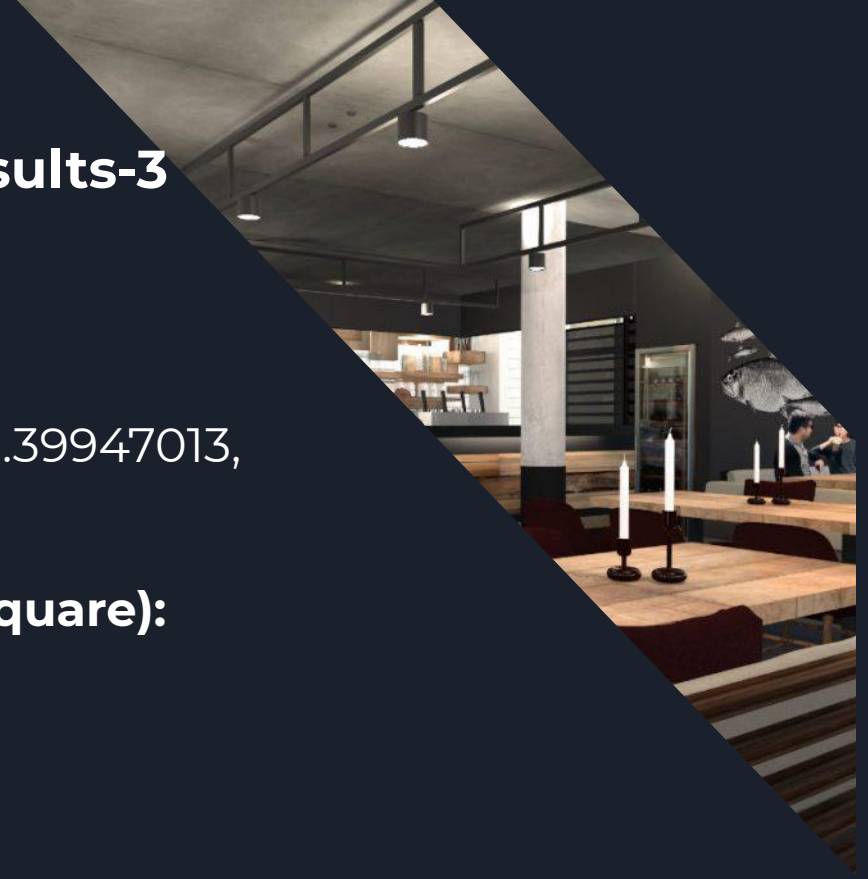
- **Multiple Regression Model Coefficients:**

  [0.25418166 0.70721605 0.8650806 -0.49001694
  0.23463565 0.88998317   0.09311164 0.5791981
   -2.83848416 -0.24770434 1.4420277]

- **The R-square:** 0.435636983137335 56
- **Residual sum of squares:** 0.86 and
- **Residual sum of squares:** 0.77 and
- **Variance score:** 0.30
- **The mean square error of Rating and predicted value:**
  0.7654593187366455

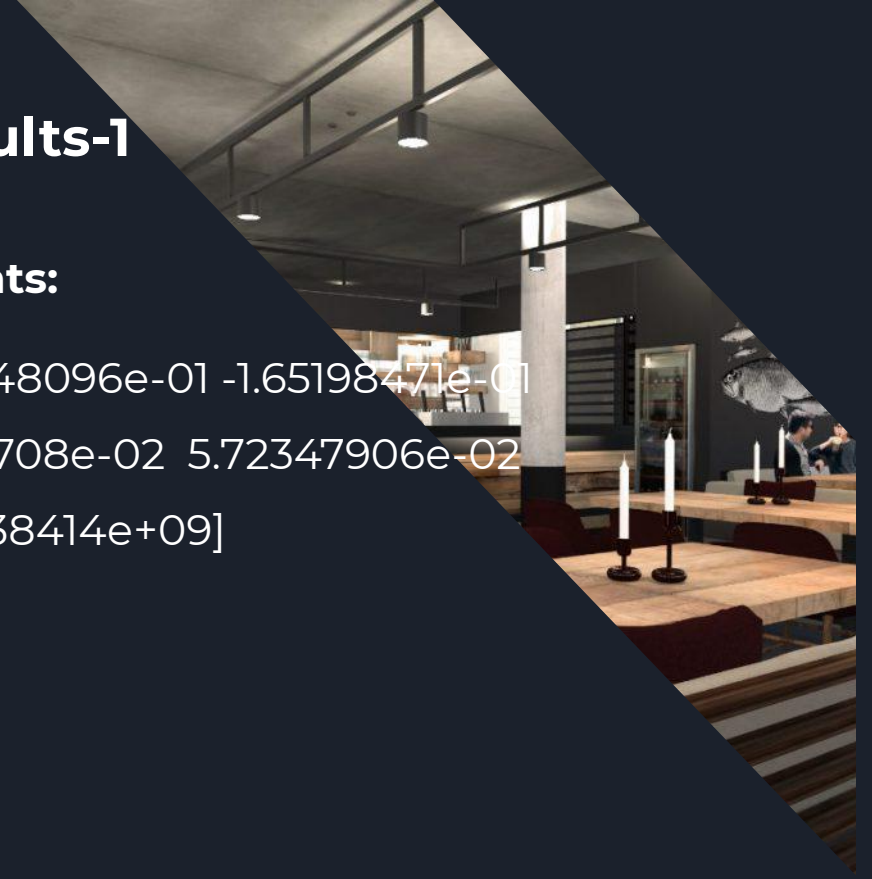
Multiple regression model-Results-3

(Cross-validation scores)



- **Rcross:** [0.18368166, 0.35121787, 0.39947013, 0.34026652, 0.27882088]
- **The mean of the iterations(R-Square):** 0.3106914121403912 5
- **The standard deviation:** 0.07422391892277923

Multiple regression model-Results-1



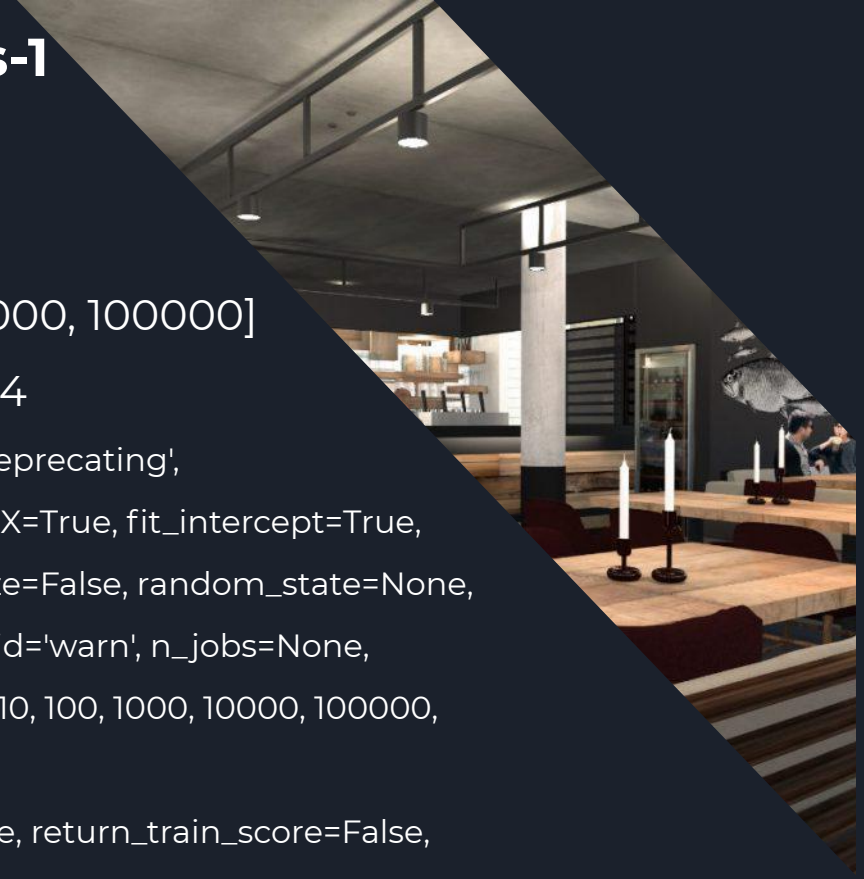
- **Multiple Regression Model Coefficients:**

  [9.71291557e-04 2.92238414e+09 1.70648096e-01 -1.65198471e-01
  3.02673696e-02 4.61008151e-01 7.13941708e-02 5.72347906e-02
  -1.51110784e-04 -2.92238414e+09 -2.92238414e+09]

- **The R-square:** 0.4126715675897107
- **Residual sum of squares:** 0.86 and
- **Variance score:** 0.22
- **The mean square error of 'Rating' and predicted value:**
  0.7654593187366441

Ridge regression model-Results-1

(With GridsearchCV)



- **Alpha parameters**

  [0.001,0.1,1, 10, 100, 1000, 10000, 100000, 100000]

- **Cross Validation Parameter (CV):** 4

```
GridSearchCV(cv=4, error_score='raise-deprecating',
             estimator=Ridge(alpha=1.0, copy_X=True, fit_intercept=True,
                             max_iter=None, normalize=False, random_state=None,
                             solver='auto', tol=0.001),  iid='warn', n_jobs=None,
             param_grid=[{'alpha': [0.001, 0.1, 1, 10, 100, 1000, 10000, 100000,
                                    100000]}],
             pre_dispatch='2*n_jobs', refit=True, return_train_score=False,
             scoring=None, verbose=0)
```

- **The R-square:** 0.3470253669057909

Discussion & Conclusion

- Comparison of models
  - multiple linear regression model
    - R-Square values: 0.45/ Cross validated score: 0.31
  - ridge regression model
    - R-Square values: 0.35 (With GridsearchCV
- These values are not so promising.
  - limited data (Almost half of the restaurant avenues data have not rating score
- can be extended by calculating the distances from people's gathering places to restaurants.
- as a target variable, revenues of restaurants can be used if available.

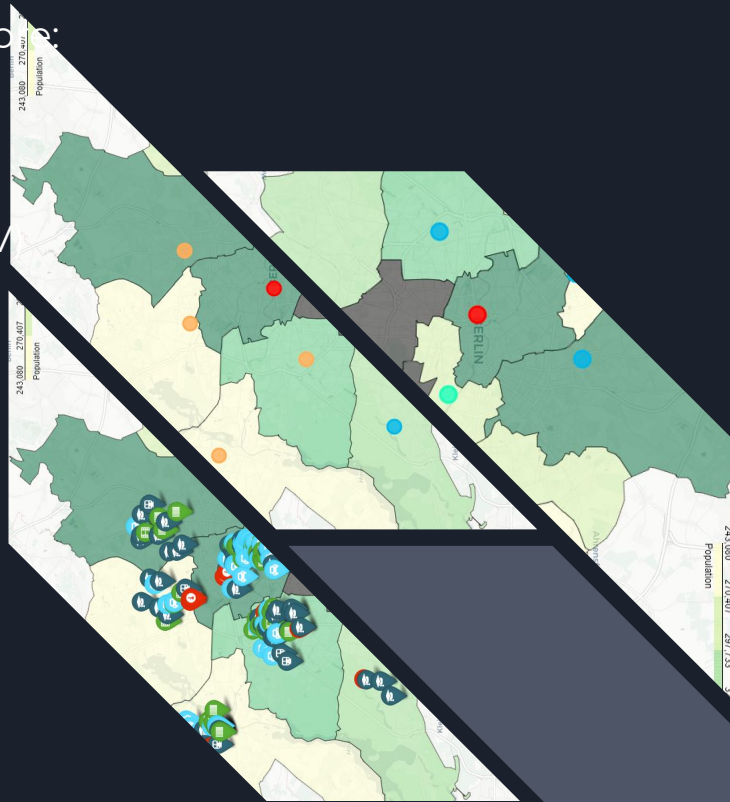


image source: https://www.juni.studio/project/interior-design-restaurant-in-berlin/