Relocation Support

Capstone Presentation
IBM Applied Data Science Capstone on Coursera

Rebeka
Bhattacharyya
July 2024

Relocation opportunities

- Globalised world facilitates people moving around to take good opportunities

- Relocation by itself also brings opportunities for service providers

- Supporting people selecting possible districts or neighbourhood in their target location brings great value

Project Goal

- Create a Districts Rank for selected target location using relocator requirements.

Input data

- Relocation family profile
  - 2 adults / 2 kids / 1 dog as pet
- Family priority list
  - Primary school
  - Outdoor park
  - Supermarket
  - Pharmacy
  - Metro station

Input data

- Housing wishes
  - Apartment 2 or 3 bedrooms
  - 80 m^2 approx.
  - 1 garage spot
- Rental budget
  - BRL 2000.00 per month (BRL 25.00/m^2)
  - Tolerance range +-5%: 23.75 - 26.25 BRL/m^2

Input data

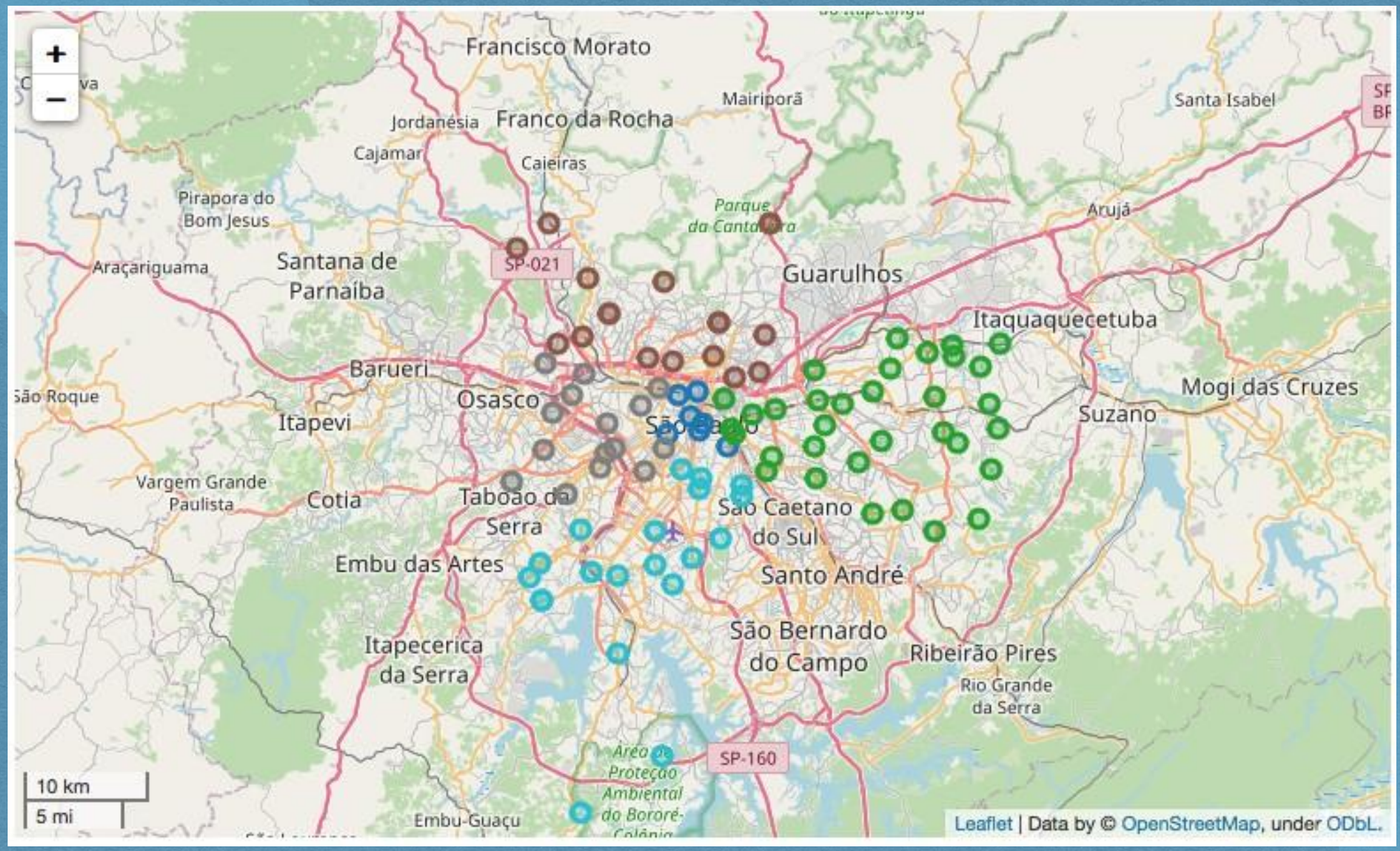
- Target location - São Paulo, Brazil
  - Over 12 million inhabitants in capital city
  - Over 22 million inhabitants in metropolitan area
  - Main financial and business location in South America
  - It responds for 11% of Brazilian GDP

Acquired data

- São Paulo districts information
  - 96 districts in 5 regions
  - Data scraped from city São Paulo official website
  - http://www.capital.sp.gov.br/
- Rental prices
  - Mean rental prices per district in BRL/m^2
  - Data scraped from Blog SP Imóvel
  - http://www.spimovel.com.br/

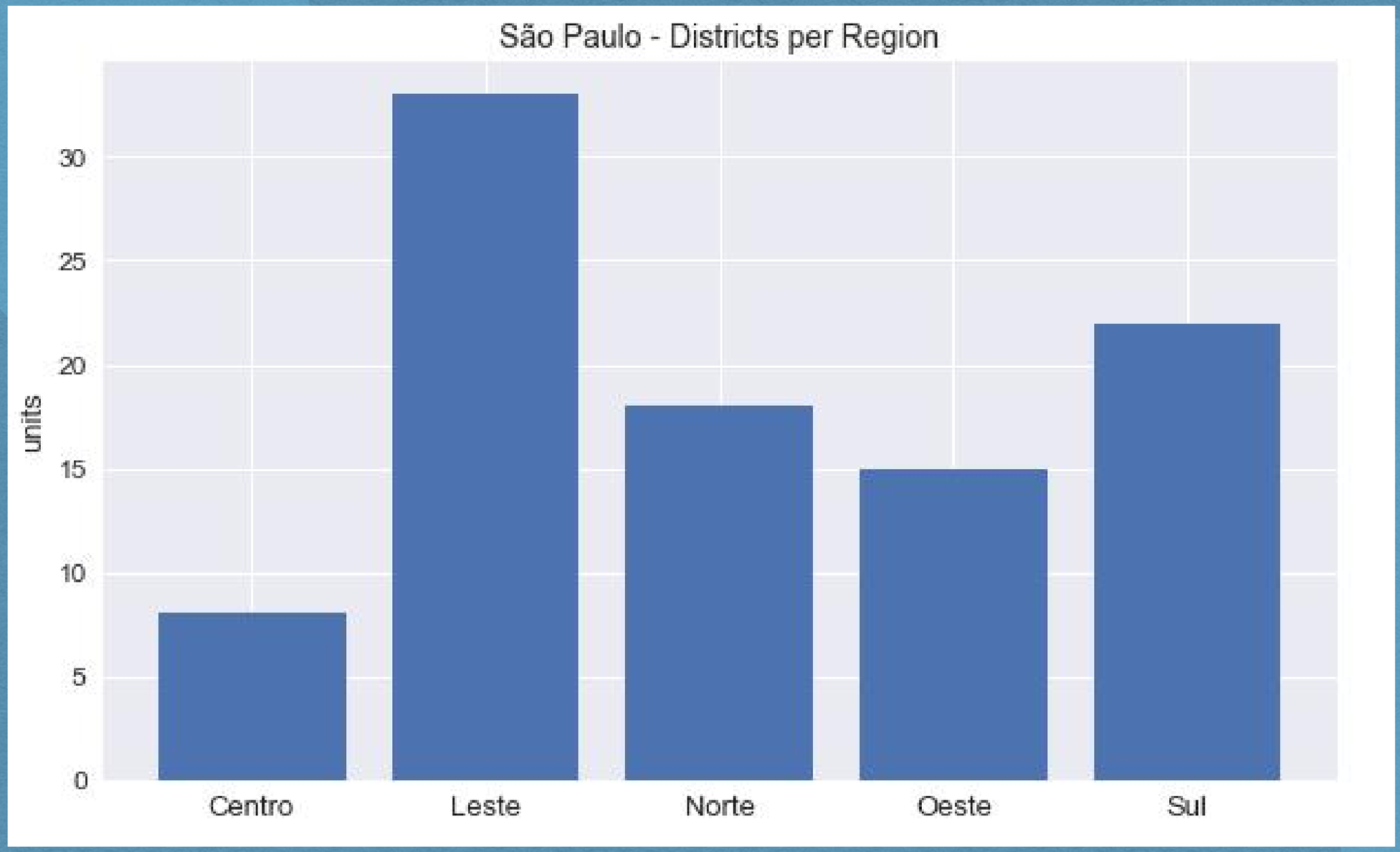
Acquired data

- Districts geolocation
  - Coordinates per district
    - Data acquired with *Geopy* package
- Venues information
  - Venues categories per district
    - Data acquired with *API Foursquare*



São Paulo Map

Districts coloured by region



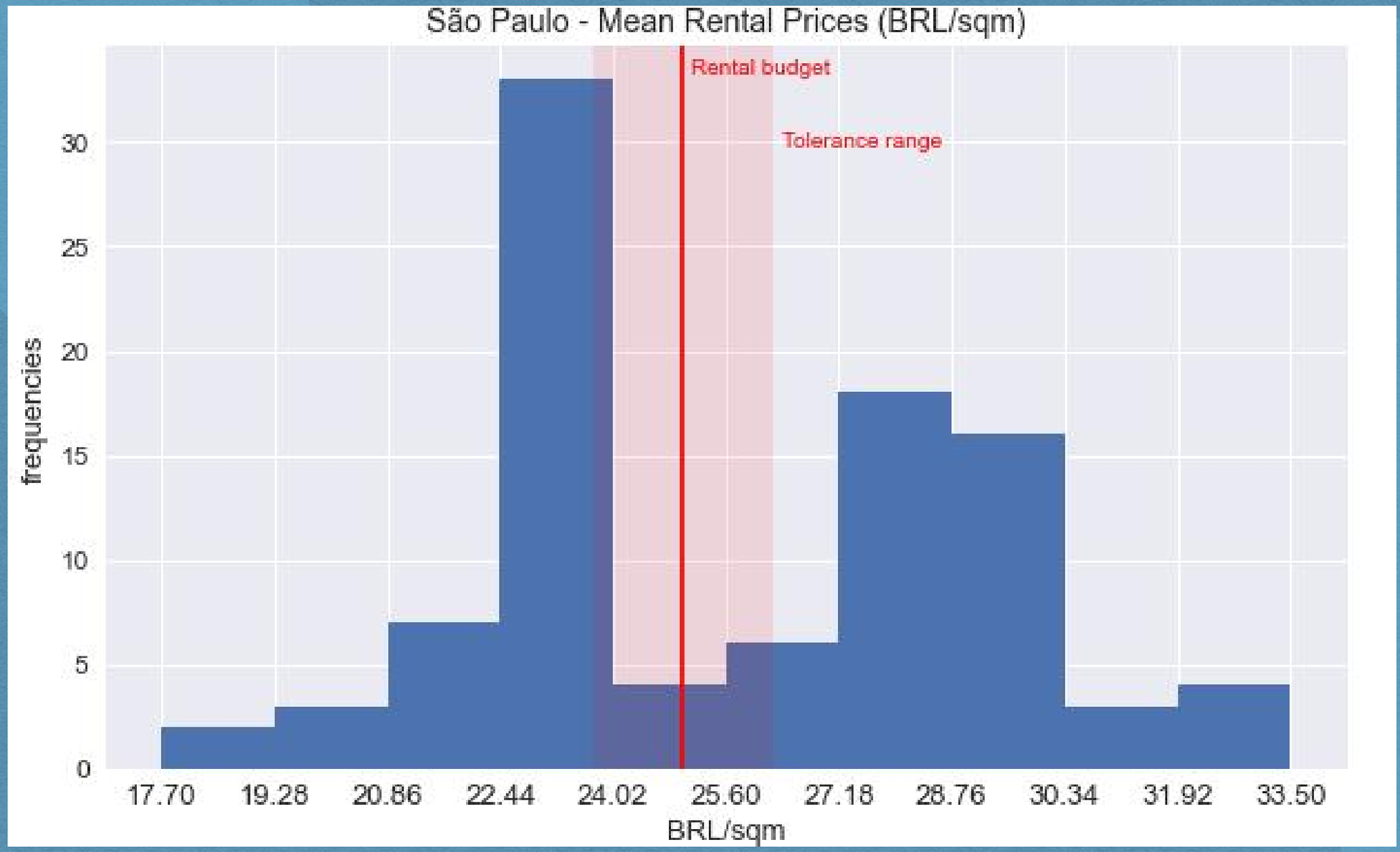
São Paulo Regions

Districts counting per region

Districts and Rental Prices

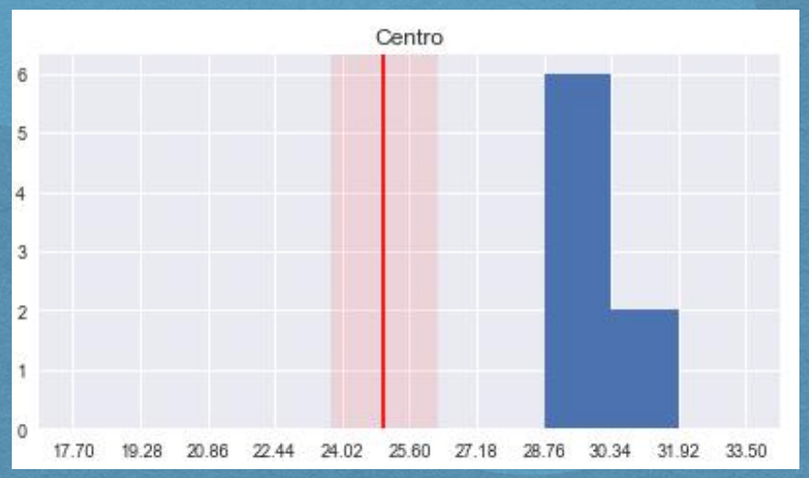
- Rental prices (BRL/$m^{2)}$
  - Min = 17,70
  - Max = 33,50
  - Q1 = 23.05
  - Q3 = 28,47

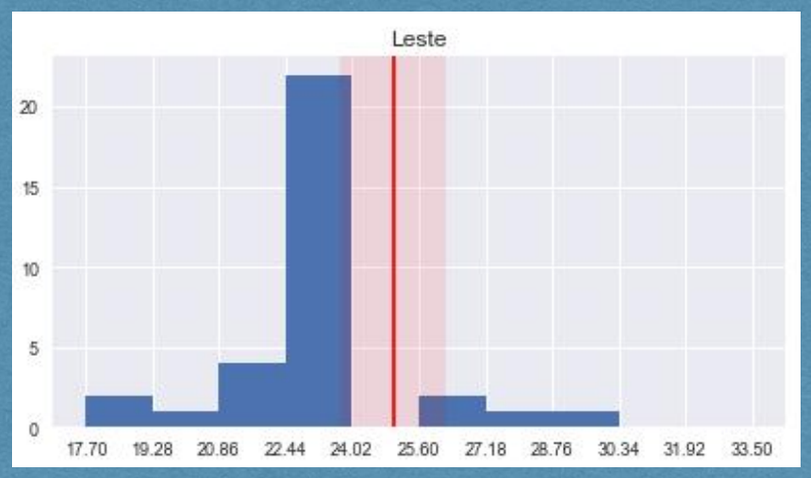
	region	borough	district	area_sqkm	mean_price_sqm
count	96	96	96	96.000000	96.000000
unique	5	32	96	NaN	NaN
top	Leste	Sé	Ponte Rasa	NaN	NaN
freq	33	8	1	NaN	NaN
mean	NaN	NaN	NaN	15.912396	25.785521
std	NaN	NaN	NaN	26.968078	3.597908
min	NaN	NaN	NaN	2.190000	17.700000
25%	NaN	NaN	NaN	7.395000	23.050000
50%	NaN	NaN	NaN	9.780000	25.150000
75%	NaN	NaN	NaN	13.560000	28.467500
max	NaN	NaN	NaN	208.190000	33.500000

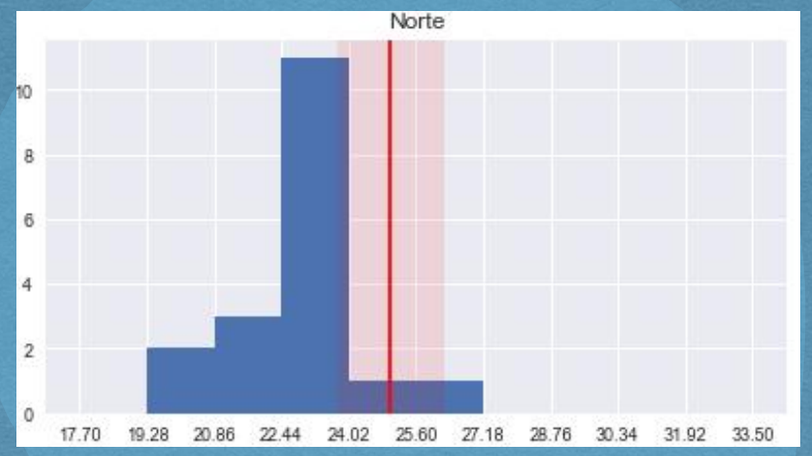


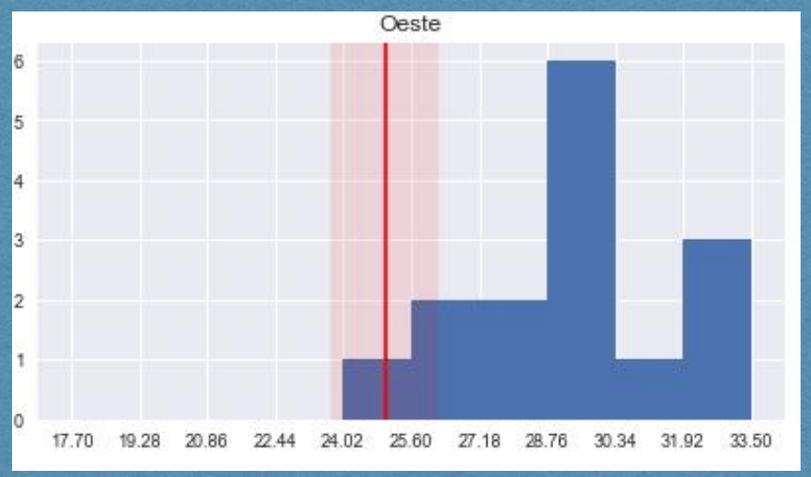
Mean Rental Price

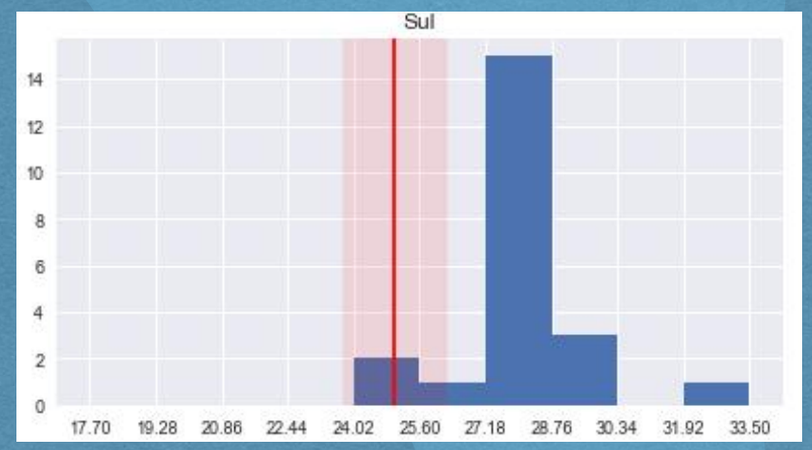
Histogram with rental budget











Mean Rental Price

Region histograms with rental budget

Clustering Algorithm

- ***DBSCAN*** - Density-Based Spatial Clustering of Applications with Noise
  - Clusters with different densities
  - Clusters with different sizes
  - Clusters with different shapes
  - No need of setting number of clusters
  - Outliers and noise identification

Clustering features

- Venues categories counting
- Mean rental price

```
# print results
df_clusters
```

	district	elementary_school	metro_station	park	pharmacy	supermarket	mean_price_sqm
0	Bela Vista	0.0	0.0	0.0	0.0	0.0	30.60
1	Bom Retiro	2.0	5.0	4.0	19.0	5.0	29.10
2	Cambuci	3.0	0.0	1.0	21.0	3.0	29.93
3	Consolação	0.0	4.0	1.0	28.0	12.0	29.93
4	Liberdade	0.0	5.0	1.0	29.0	10.0	30.70
...	...	...	...	...	...	...	...
91	Campo Grande	3.0	4.0	5.0	25.0	8.0	28.39
92	Santo Amaro	2.0	4.0	5.0	28.0	8.0	26.80
93	Moema	0.0	4.0	3.0	28.0	11.0	32.60
94	Saúde	0.0	7.0	2.0	31.0	8.0	28.20
95	Vila Mariana	1.0	5.0	2.0	26.0	12.0	29.90

96 rows × 7 columns

Clustering data preparation

- District name removed
- Venues categories counting converted to
  - 0 when counting = 0
  - 1 when counting > 0
- Mean rental price converted to rental groups
  - 0 when below tolerance range (below 23.75)
  - 1 when into tolerance range (between 23.75 and 26.25)
  - 2 when above tolerance range (above 26.25)

Clustering features ready

```
# print results
df_clusters
```

	elementary_school	metro_station	park	pharmacy	supermarket	rental_group
0	0.0	0.0	0.0	0.0	0.0	2
1	1.0	1.0	1.0	1.0	1.0	2
2	1.0	0.0	1.0	1.0	1.0	2
3	0.0	1.0	1.0	1.0	1.0	2
4	0.0	1.0	1.0	1.0	1.0	2
...	...	...	...	...	...	...
91	1.0	1.0	1.0	1.0	1.0	2
92	1.0	1.0	1.0	1.0	1.0	2
93	0.0	1.0	1.0	1.0	1.0	2
94	0.0	1.0	1.0	1.0	1.0	2
95	1.0	1.0	1.0	1.0	1.0	2

96 rows × 6 columns

cluster_label	elementary_school	metro_station	park	pharmacy	supermarket	rental_group
-1	9	9	9	9	9	9
0	19	19	19	19	19	19
1	13	13	13	13	13	13
2	7	7	7	7	7	7
3	4	4	4	4	4	4
4	19	19	19	19	19	19
5	3	3	3	3	3	3
6	11	11	11	11	11	11
7	4	4	4	4	4	4
8	7	7	7	7	7	7

Clustering Results

Clusters size

cluster_label	elementary_school	metro_station	park	pharmacy	supermarket	rental_group
-1	0.333333	0.222222	0.333333	0.555556	0.444444	0.888889
0	1.000000	1.000000	1.000000	1.000000	1.000000	2.000000
1	1.000000	0.000000	1.000000	1.000000	1.000000	2.000000
2	0.000000	1.000000	1.000000	1.000000	1.000000	2.000000
3	1.000000	0.000000	0.000000	1.000000	1.000000	0.000000
4	1.000000	0.000000	1.000000	1.000000	1.000000	0.000000
5	1.000000	0.000000	1.000000	1.000000	1.000000	1.000000
6	1.000000	1.000000	1.000000	1.000000	1.000000	0.000000
7	1.000000	1.000000	0.000000	1.000000	1.000000	0.000000
8	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Clustering Results

Clusters characteristics by mean value

Districts Rank (cluster analysis)

- Outliers
  - Group -1
- Priority List not satisfied
  - Cluster 1, missing Metro Station
  - Cluster 2, missing Elementary School
  - Cluster 3, missing Metro Station and Park
  - Cluster 4, missing Metro Station
  - Cluster 5, missing Metro Station
  - Cluster 7, missing Park

Relocation Support by Sidclay da Silva

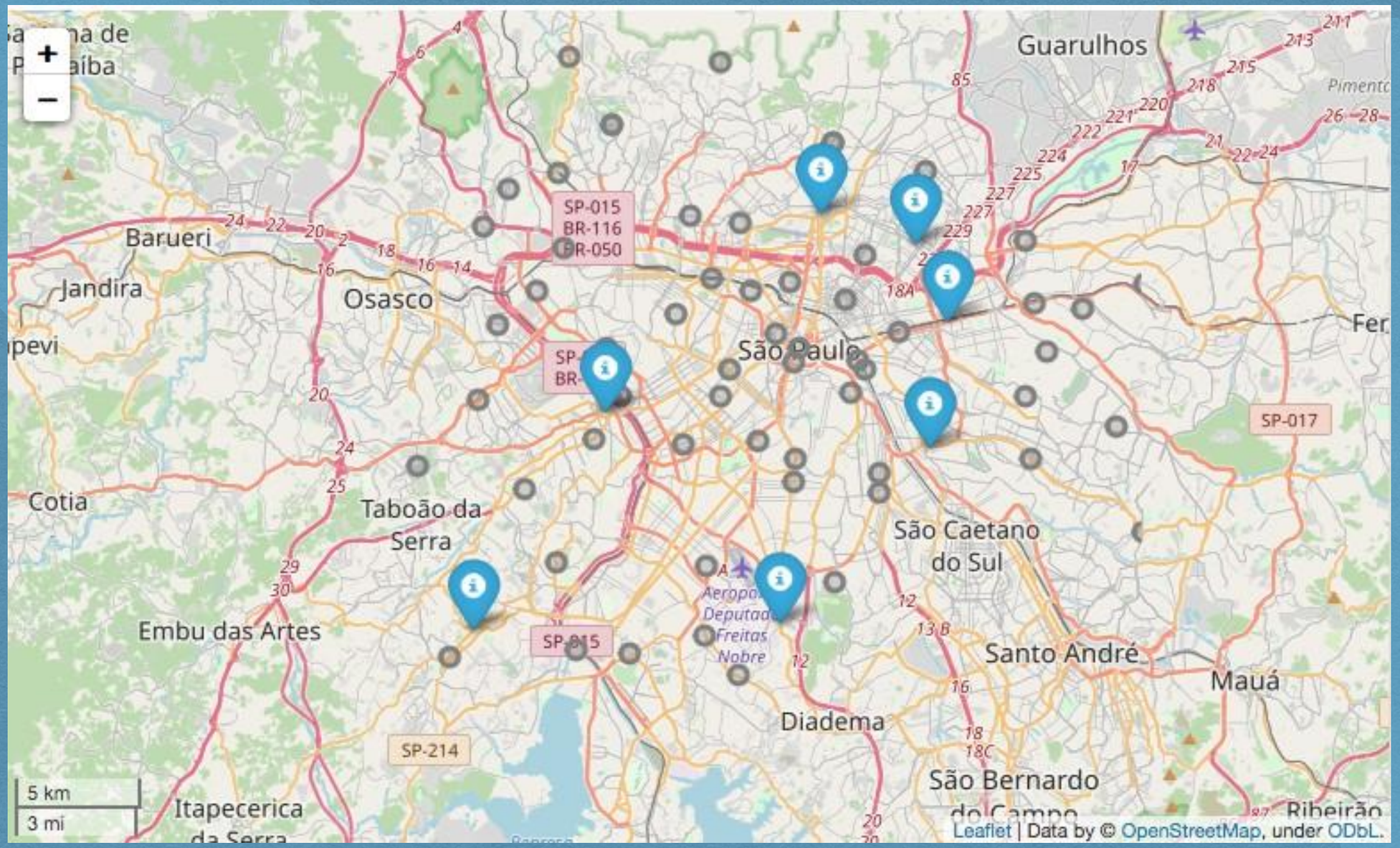
Districts Rank (cluster analysis)

- **Priority List satisfied**
  - Cluster 0, Rental Group 2 (above rental price tolerance range)
  - Cluster 6, Rental Group 0 (below rental price tolerance range)
  - **Cluster 8, Rental Group 1 (into rental price tolerance range)**

	district	elementary_school	metro_station	park	pharmacy	supermarket	cluster_label	mean_price_sqm
0	Tatuapé	1.0	6.0	5.0	26.0	9.0	8	25.7
1	Vila Prudente	5.0	3.0	7.0	23.0	8.0	8	25.9
2	Santana	2.0	5.0	2.0	31.0	8.0	8	23.8
3	Vila Maria	8.0	1.0	4.0	26.0	5.0	8	24.8
4	Butantã	1.0	4.0	7.0	28.0	7.0	8	25.0
5	Campo Limpo	7.0	5.0	3.0	24.0	10.0	8	24.5
6	Jabaquara	5.0	3.0	1.0	24.0	12.0	8	25.3

Districts Details

Cluster 8 - Relocator requirements satisfied



São Paulo Map

Selected districts in Cluster 8

Conclusion and Future directions

- The solution successfully created Districts Rank to satisfy Relocator Requirements

- It has market value, Relocation Service Providers could make use of it to improve and speed up locations selection

- It could be enhanced to a configurable and flexible solution, including additional parameters for searching features

- It could be enhanced also to select rental ads. after creating the rank

- The solution has a great potencial for future developments.