



Girlfriend's new apartment

Capstone Project – IBM Data Science course



São Paulo Real Estate




- The pandemic has caused several changes to our society.
- One of the effects in the city of São Paulo was a sharp increase in the demand for rental properties
- During the third trimester of 2020, there was an 18% increase in the number of searches in a popular real state website compared to the same period in 2019.
- The number of officialized real state transactions was 336,968, a 37% increase compared to 10 years prior.
- Our goal is to aid our client in her search for a new apartment.

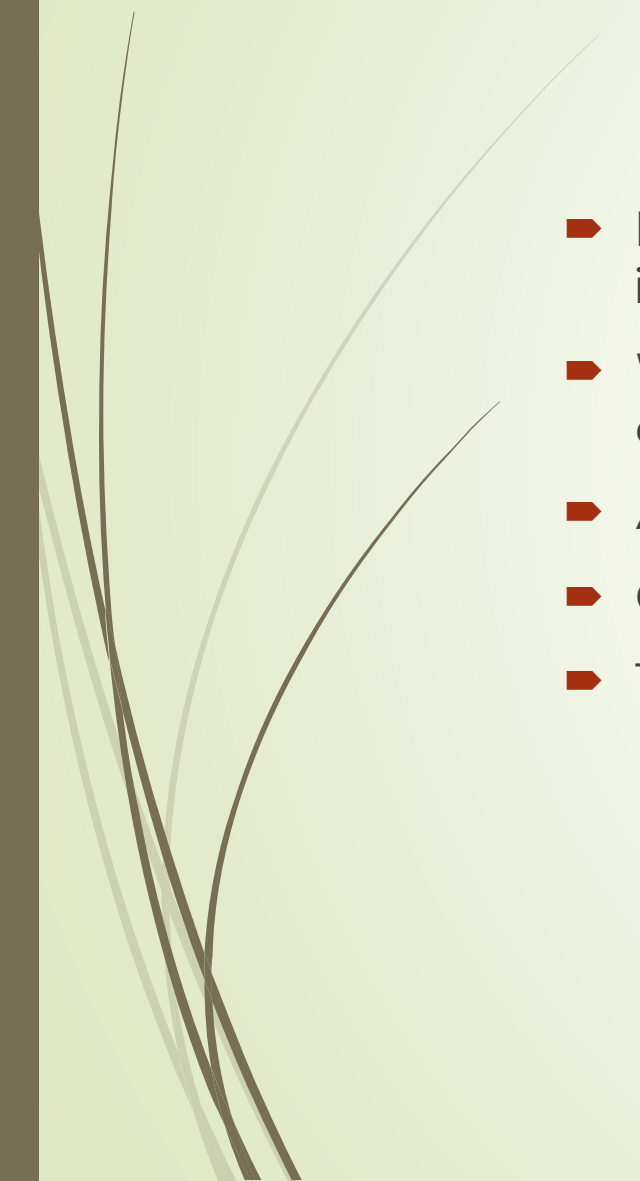



Data sources – Quinto Andar

- ▶ Quinto Andar is one of the most famous real state websites in São Paulo.
- ▶ We filter for apartments that comply with the followin requirements given by the client:
 - ▶ Total rental price (Rent + condominium fee + taxes) larger than 1000,00 BRL and lower than 2800,00 BRL.
 - ▶ Total number of bedrooms larger than 2.
 - ▶ Number of parking spots larger than 1.
 - ▶ Close to train or metro stations.
 - ▶ Accepts pets.
 - ▶ Currently available.

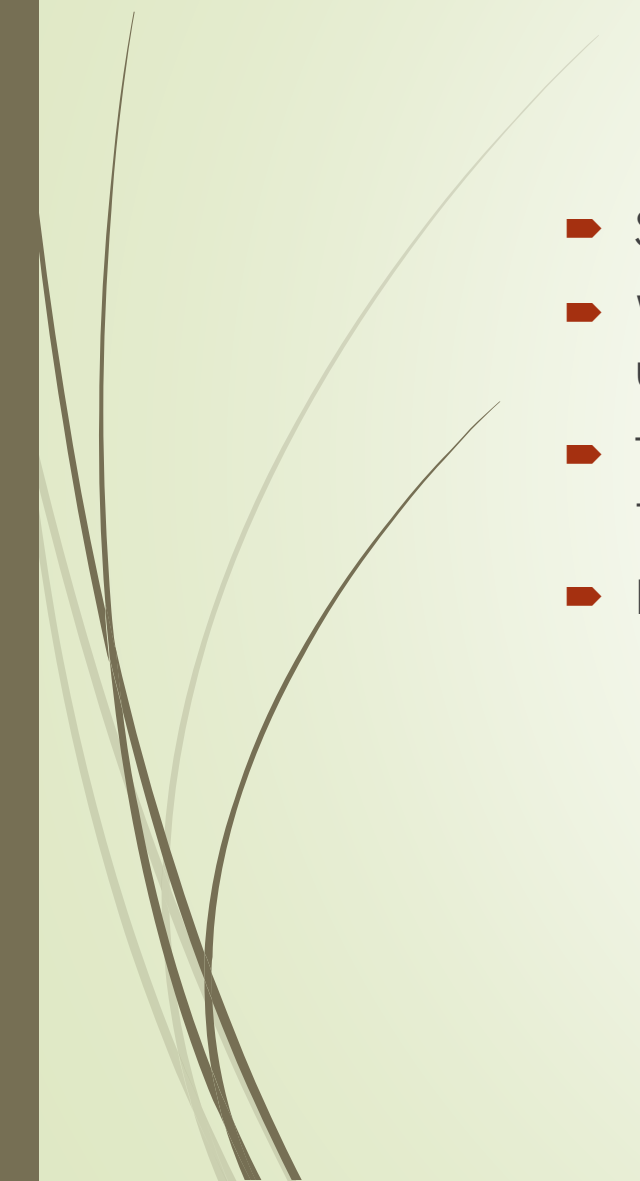


Data sources – Foursquare

- Foursquare is an app, which is purpose is to clients discover and share information about businesses and attractions around designated locations.
 - We obtained data on the venues surrounding each of the remaining apartments after filtering.
 - Area of 800 m around each apartment.
 - Only the venues categories are stored.
 - This data is used for clustering appartements.
- 



Data sources - Google Maps

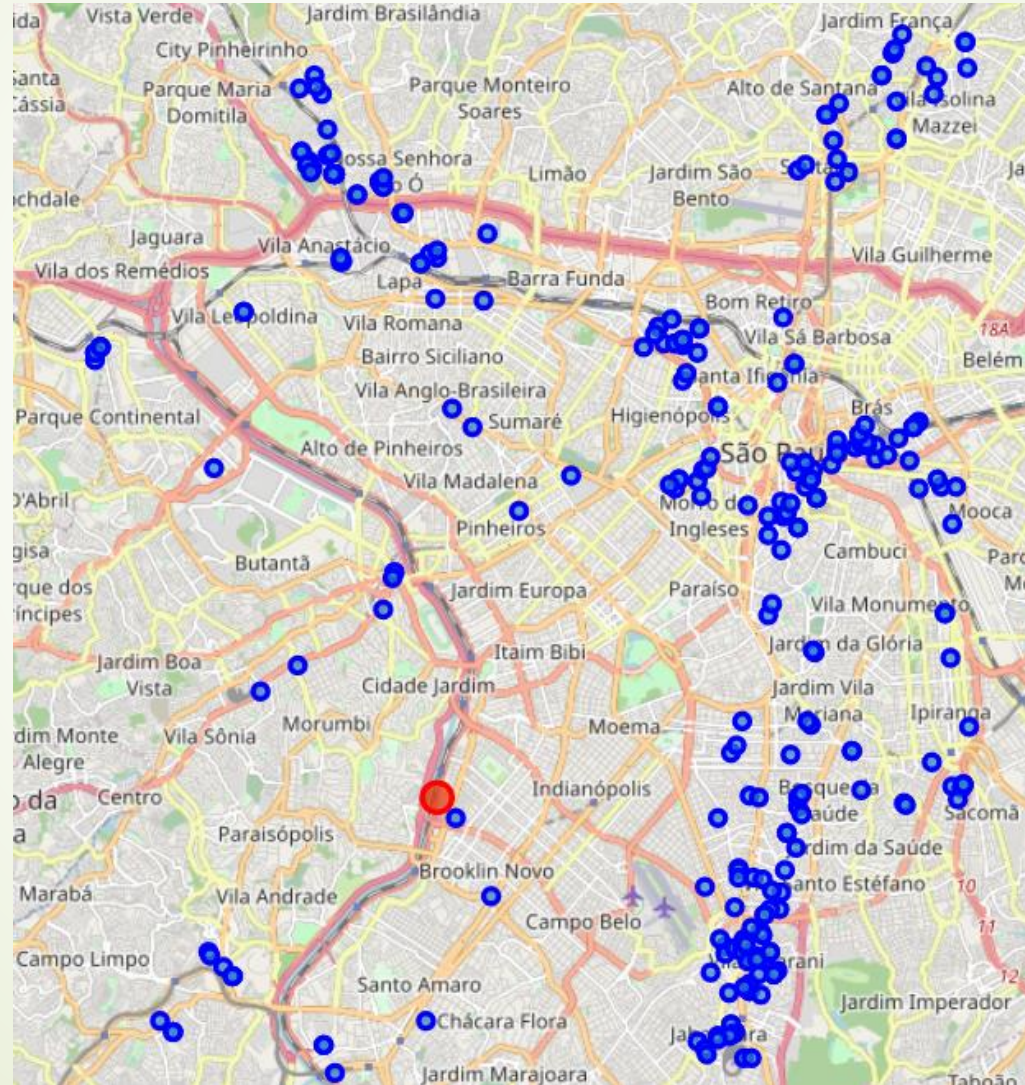
- Similar to other large cities in the world, São Paulo has problems with traffic.
 - We use Google Maps "Distance Matrix" API to calculate the transit time using public transportation from each apartment.
 - The time is calculating considering an arrival time of 9 am the next Monday to the time the code is run.
 - For this presentation July 26, 2021.
- 



Methodology

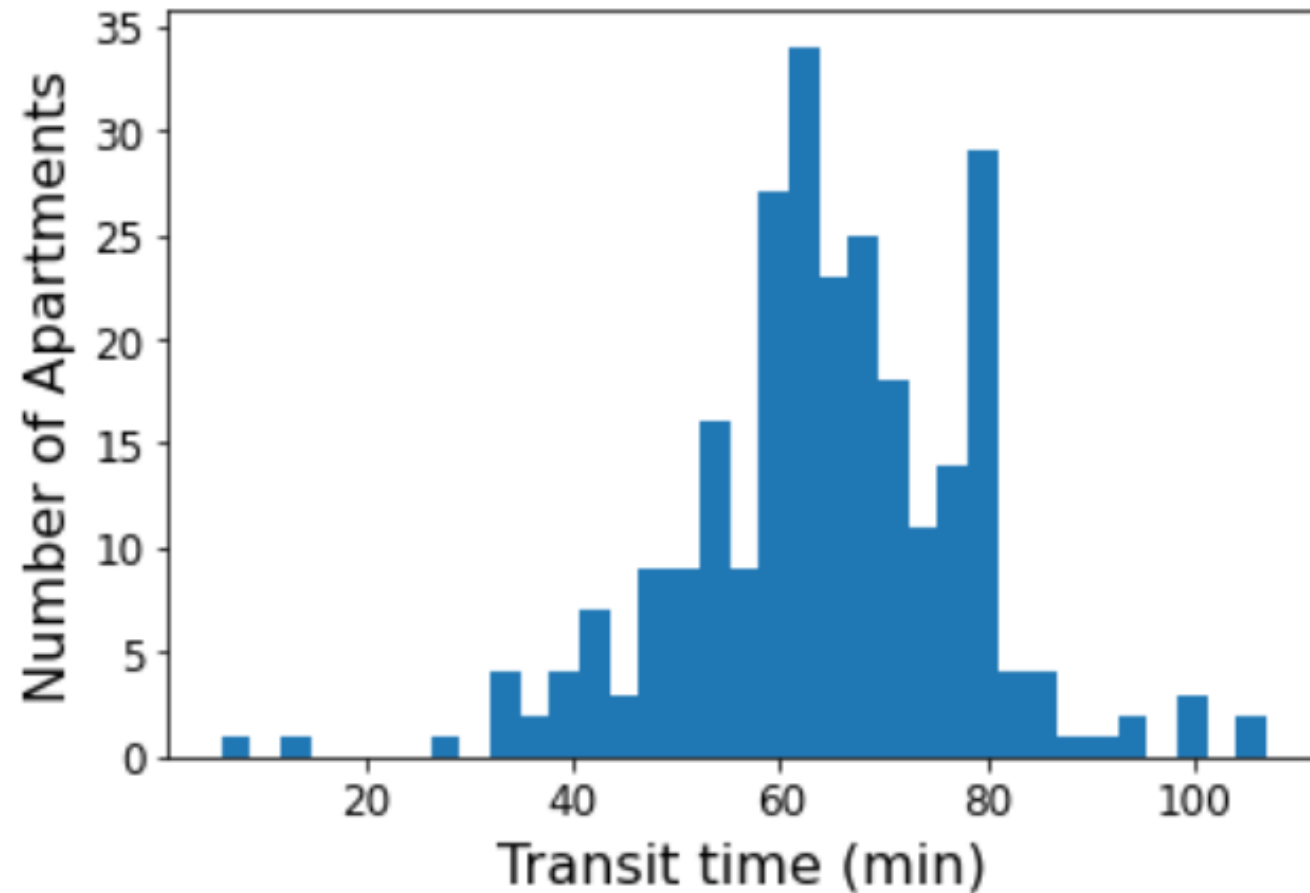
- First step – Filtering
 - Done during data acquisition.
- Clustering:
 - K-Means of the apartments given the frequency of categories of venues surrounding each apartment.
 - Test locations are used to choose the number of clusters and the chosen cluster.
- Transit time filtering
 - Filter the apartments within 60 minutes commute time.

Results – Total apartments Map

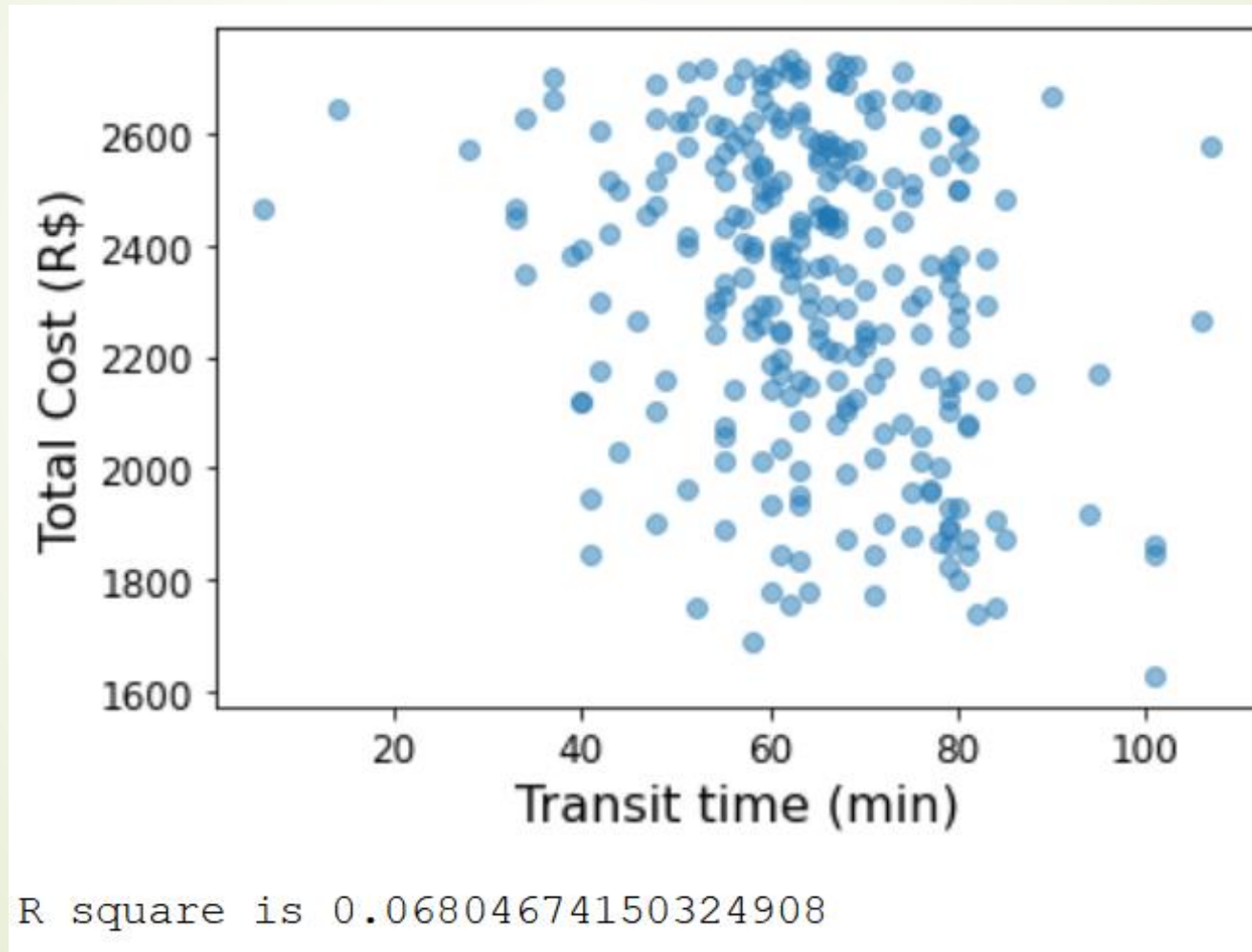


Results – Transit time histogram

The transit time median is 64.5 minutes.



Results – Transit time vs Total cost scatter plot



Results – Chosen Candidates

Website_Id	Lat	Lng	Bedrooms	Parking_Slots	Total_Value	Total_Area	Floor	Neighborhood	Status	Traffic_time(min)
101409	-23.608980	-46.630734	2	1	2717.0	57.0	6	Bosque da Saúde	publicado	57
119491	-23.524734	-46.697685	2	1	2690.0	90.0	8	Vila Romana	publicado	56
257592	-23.586649	-46.723164	2	1	2448.0	60.0	10	Butantã	publicado	33
291088	-23.601331	-46.643220	2	1	2575.0	72.0	13	Vila Clementino	publicado	51
335416	-23.607796	-46.619220	2	1	2448.0	70.0	7	Vila Gumerindo	publicado	57
518304	-23.608234	-46.630049	2	1	2454.0	47.0	3	Bosque da Saúde	publicado	56
615232	-23.622575	-46.641706	2	1	2538.0	65.0	8	Saúde	publicado	59
625436	-23.571916	-46.705593	2	1	2628.0	68.0	7	Butantã	publicado	34
627613	-23.600038	-46.642174	2	1	2689.0	70.0	1	Vila Clementino	publicado	48
632196	-23.526911	-46.733039	2	1	2653.0	75.0	18	Vila Leopoldina	publicado	52
632693	-23.603011	-46.606042	2	1	2400.0	67.0	7	Vila Nair	publicado	58
635542	-23.596033	-46.629236	2	1	2416.0	60.0	4	Vila Mariana	publicado	51
643807	-23.596178	-46.641170	2	1	2609.0	60.0	3	Vila Mariana	publicado	55
662260	-23.612393	-46.645429	2	1	2710.0	72.0	3	Planalto Paulista	publicado	51
671551	-23.591110	-46.729930	2	1	2262.0	60.0	14	Vila Sônia	publicado	46
83787	-23.625629	-46.687598	2	1	2646.0	85.0	3	Jardim das Acácias	publicado	14

Results – Chosen Candidates top 5

	Website_Id	1	2	3	4	5	labels
0	101409	Fruit & Vegetable Store	Gym / Fitness Center	Pizza Place	Burger Joint	Bar	3
2	119491	Bakery	Gym / Fitness Center	Pharmacy	Restaurant	BBQ Joint	3
28	257592	Pharmacy	Pizza Place	Gym / Fitness Center	Coffee Shop	Brazilian Restaurant	3
35	291088	Gym / Fitness Center	Brazilian Restaurant	Pet Store	Burger Joint	Dessert Shop	3
45	335416	Bakery	Gym	Gym / Fitness Center	Pharmacy	Dessert Shop	3
83	518304	Gym / Fitness Center	Japanese Restaurant	Fruit & Vegetable Store	Burger Joint	Dance Studio	3
131	615232	Bakery	Pizza Place	Brazilian Restaurant	Pharmacy	Gym / Fitness Center	3
146	625436	Bar	Café	Brazilian Restaurant	Chocolate Shop	Ice Cream Shop	3
152	627613	Bar	Coffee Shop	Brazilian Restaurant	Café	Pet Store	3
160	632196	Dessert Shop	Brazilian Restaurant	Japanese Restaurant	Gym / Fitness Center	Burger Joint	3
161	632693	Pizza Place	Gym / Fitness Center	Optical Shop	Bar	Pet Store	3
165	635542	Burger Joint	Pet Store	Pizza Place	Pharmacy	BBQ Joint	3
179	643807	Bar	Coffee Shop	Brazilian Restaurant	Dessert Shop	Pizza Place	3
223	662260	Cosmetics Shop	Bakery	Pizza Place	Dessert Shop	Pharmacy	3
245	671551	Pharmacy	Pizza Place	Bakery	Japanese Restaurant	Grocery Store	3
257	83787	Gym	Salon / Barbershop	Mineiro Restaurant	Supermarket	Italian Restaurant	3
264	GF0	Dessert Shop	Cocktail Bar	Restaurant	Hostel	Spa	3
265	GF1	Cosmetics Shop	Dessert Shop	Gym / Fitness Center	Pizza Place	Bar	3



Conclusion



- From over 4200 apartments the algorithm resulted in 16 candidates
- All of the requirements were met
- The candidates in the same cluster clearly share the same characteristics.
- Transit times of 60 minutes
- The final decision on the optimal apartment will be made by the client's preferences.