

Finding suitable neighborhoods when moving from São Paulo to Camberra and vice-versa

Capstone Project by José
Augusto Meirelles

Introduction

- The objective is compare neighborhoods from São Paulo/Brasil with Canberra/Australia to provide information to people who wants to move from the first one to the second one and vice-versa.
- This way neighborhoods of both cities will be grouped by it's similarities and main venues characteristics.

Data

- Source of data:
 - Boroughs and Neighborhoods
 - São Paulo was obtained from https://www.prefeitura.sp.gov.br/cidade/secretarias/subprefeituras/subprefeituras/dados_demograficos/index.php
 - Canberra was obtained from https://en.wikipedia.org/wiki/List_of_Canberra_suburbs
 - Geo Location of the Neighborhoods was obtained from Nominatim from geopy
 - Trending Venue data from foursquare api

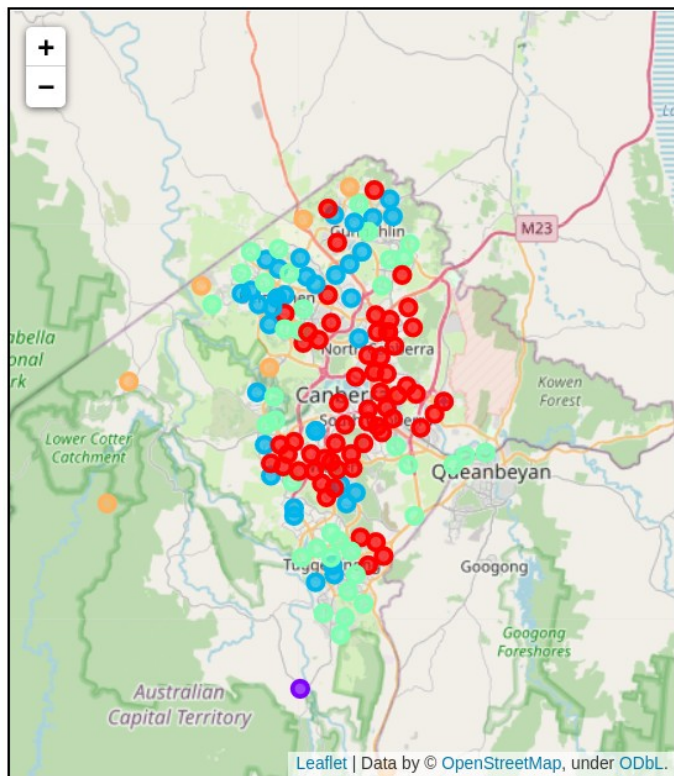
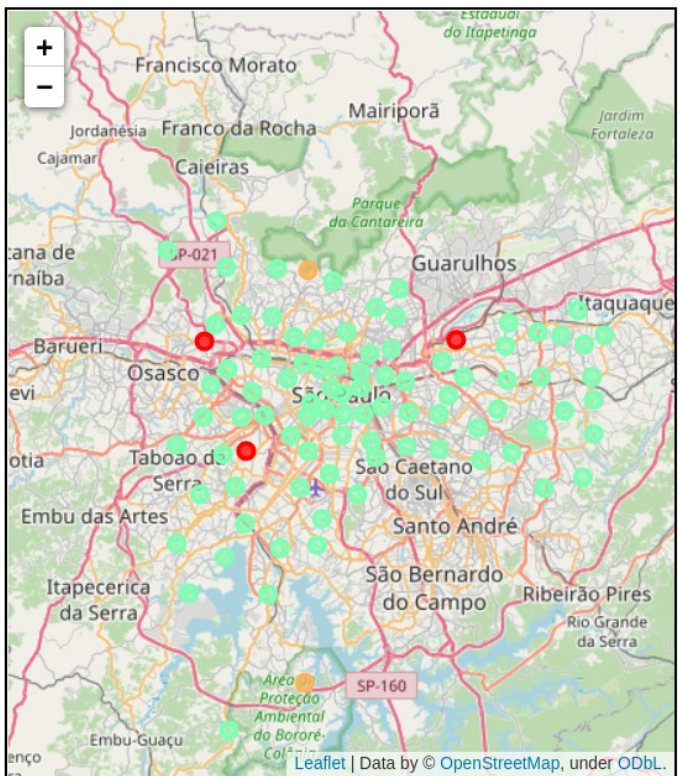
Methodology

- Data gathering
 - First we collected the data of neighborhood from the both cities, find all geolocations using geopy
 - After that we merge all neighborhoods in a single dataframe with a distinctive city index
 - After the first foursquare exploration I discovered that's Camberra have significantly less venue data than São Paulo (767 vs 2658) when is used a radius of 500 meters for both, so I tripled the radius just for Camberra.
- Data preparation
 - Transpose venue category information to columns
 - Summarize categories by neighborhood using mean for normalization
 - Sort top 10 venue categories by neighborhood
- Clustering
 - Clustering neighborhoods altogether using K-Means considering top 10 venue categories
- Comparison
 - Plot clustered maps side by side to visualize similar neighborhoods
 - Compare the characteristics of the nth more common venues categories

Analysis

- The first discovery that's Camberra have significantly less venue data than São Paulo (767 vs 2658) when I used a 500 meter radius, what make me think that Camberra could be a much more residential city than São Paulo. To have a similar amount of venue data I had to triple the radius for Camberra's exploration (2802 vs 2658).
- Even using a larger radius for Camberra it has fewer venue categories than São Paulo (215 vs 290) which implies that São Paulo has a variety of venues 34,88% bigger even exploring a radius 3 times smaller.

All Clusters in Maps



São Paulo and Canberra have different predominant clusters

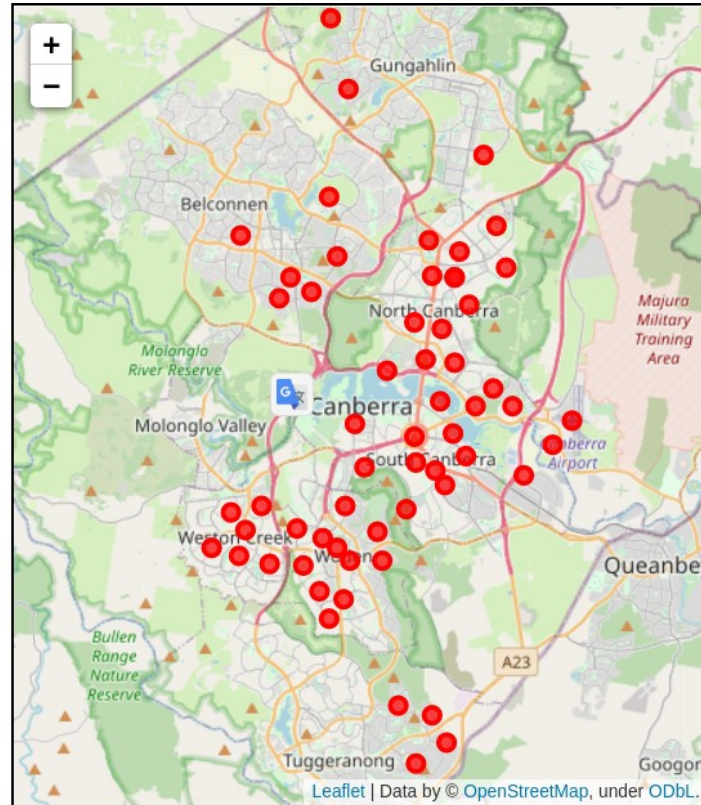
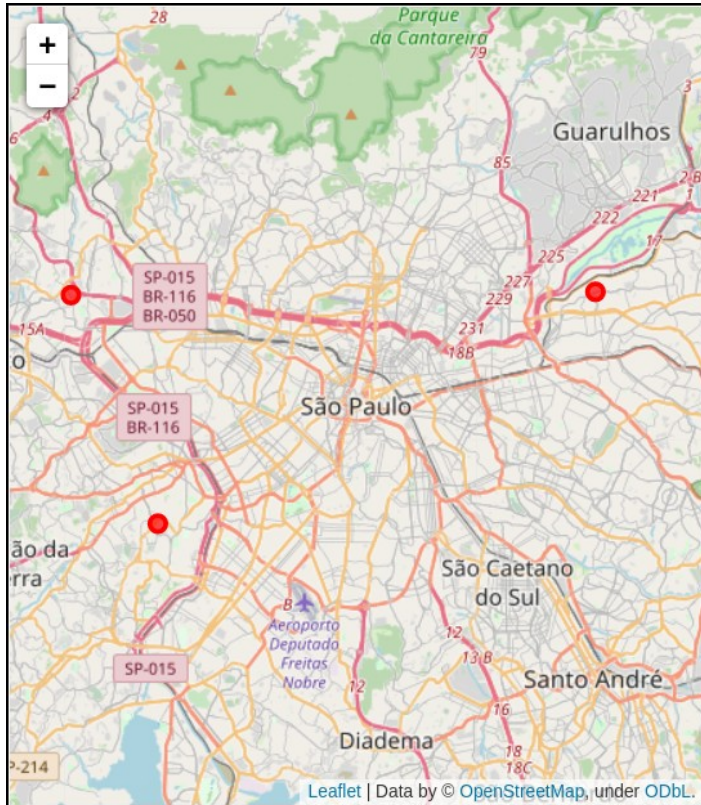
Also as can be seen Canberra had its neighborhoods more distributed in different clusters, but São Paulo have the majority clustered all together

No venue data (non-clustered nbs)

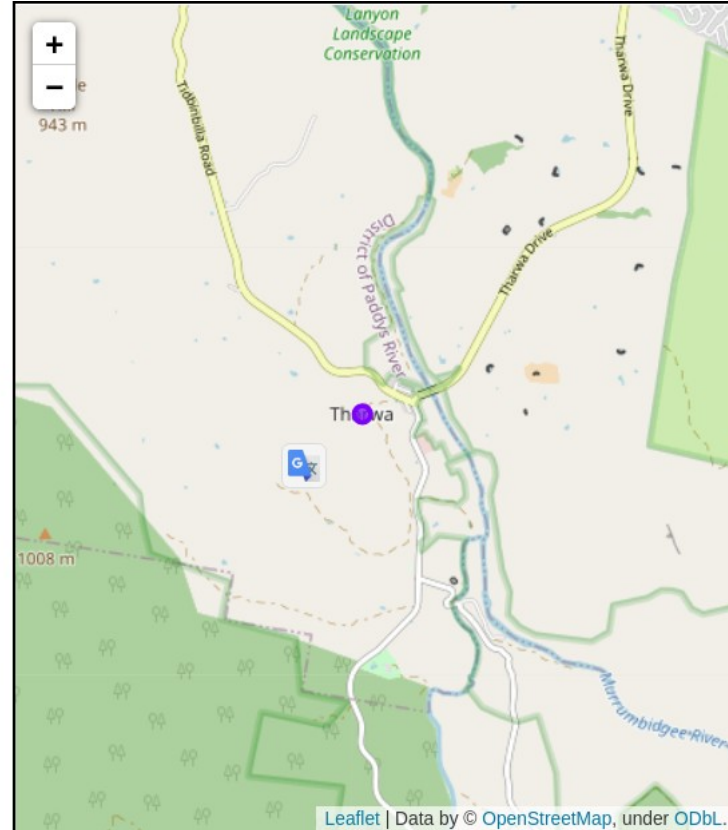
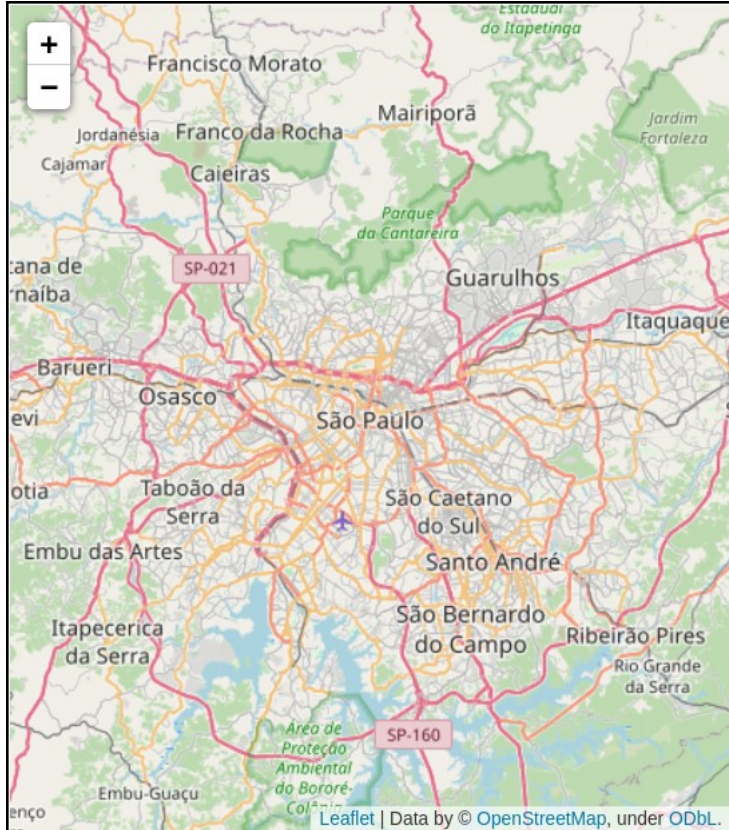


It's possible to observe that these neighborhoods are located mainly in forest and preservation areas

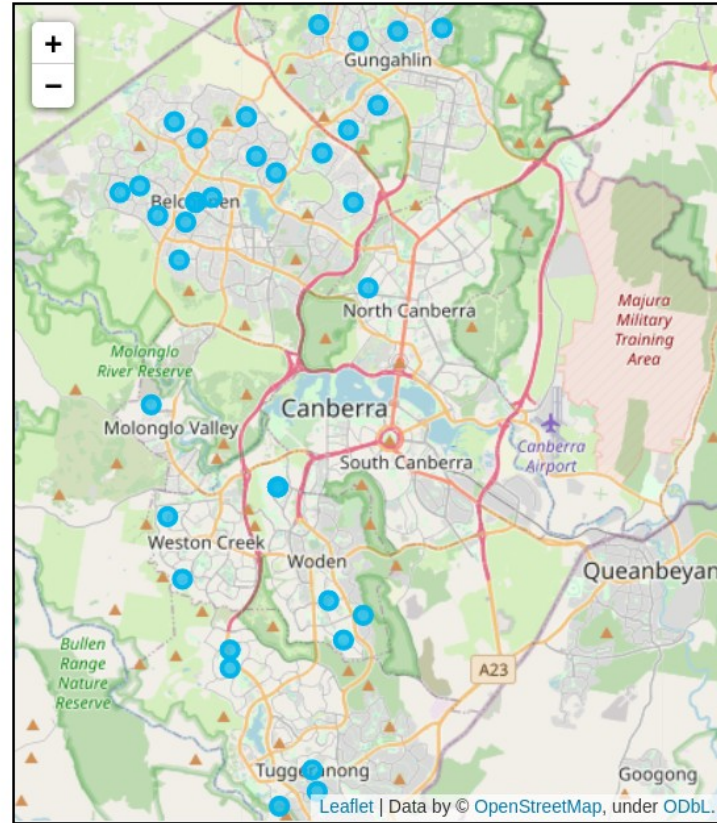
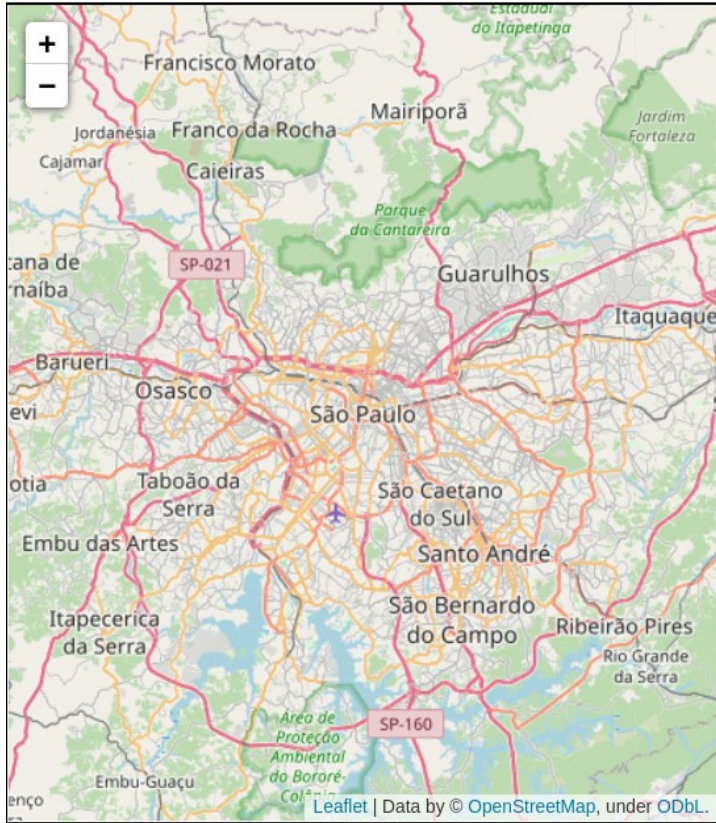
Cluster 0



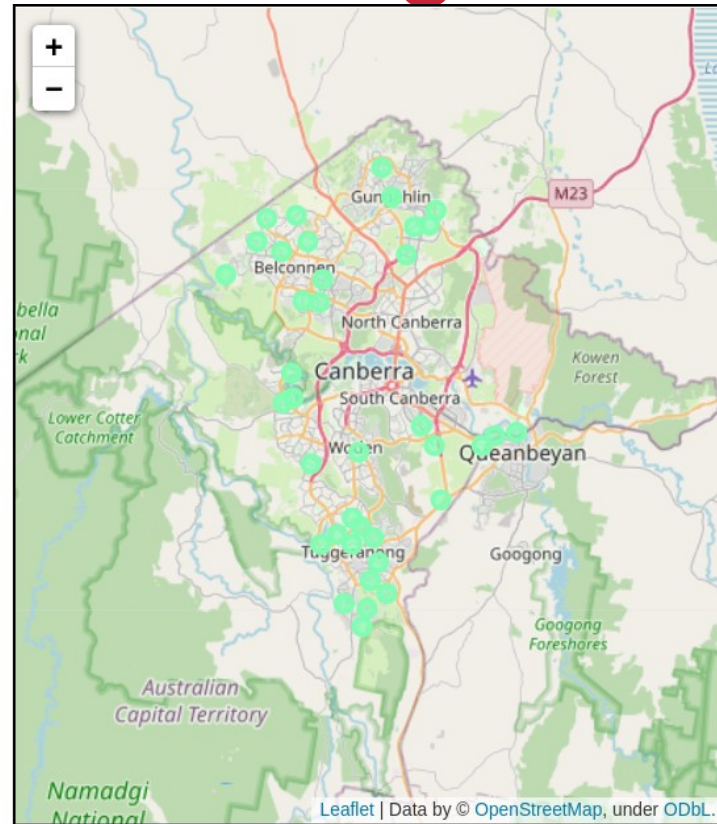
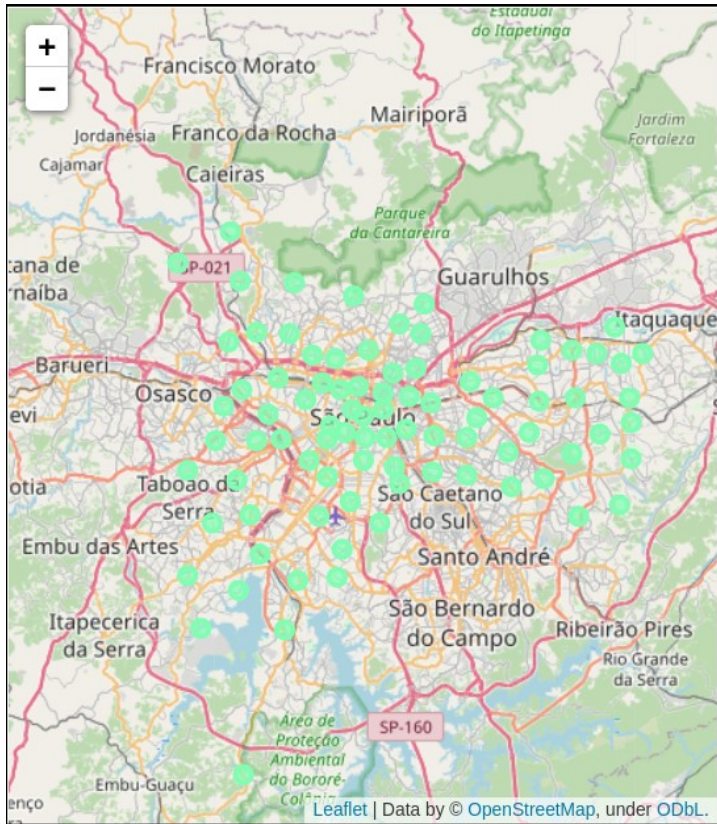
Cluster 1 – no match on SP



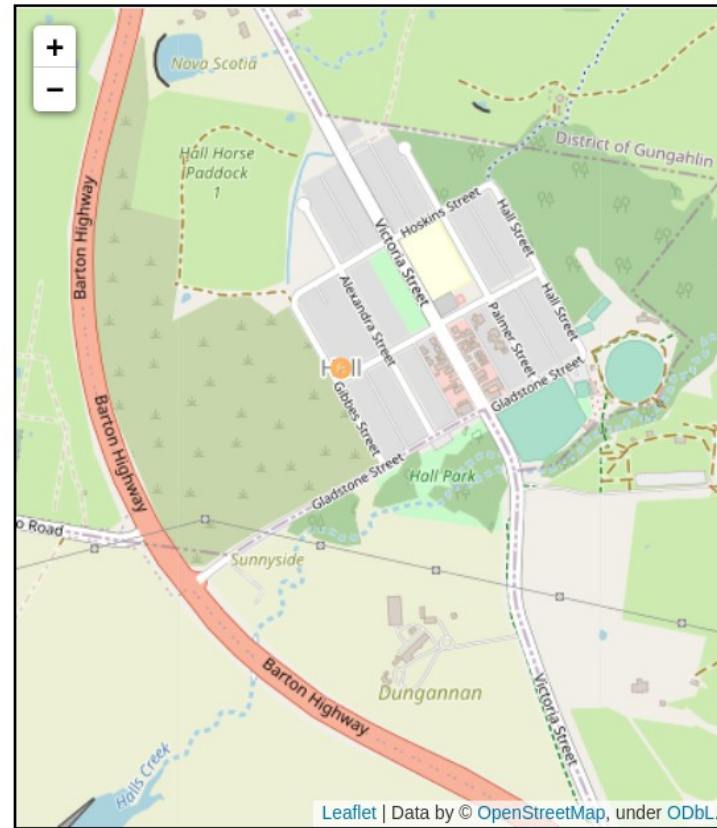
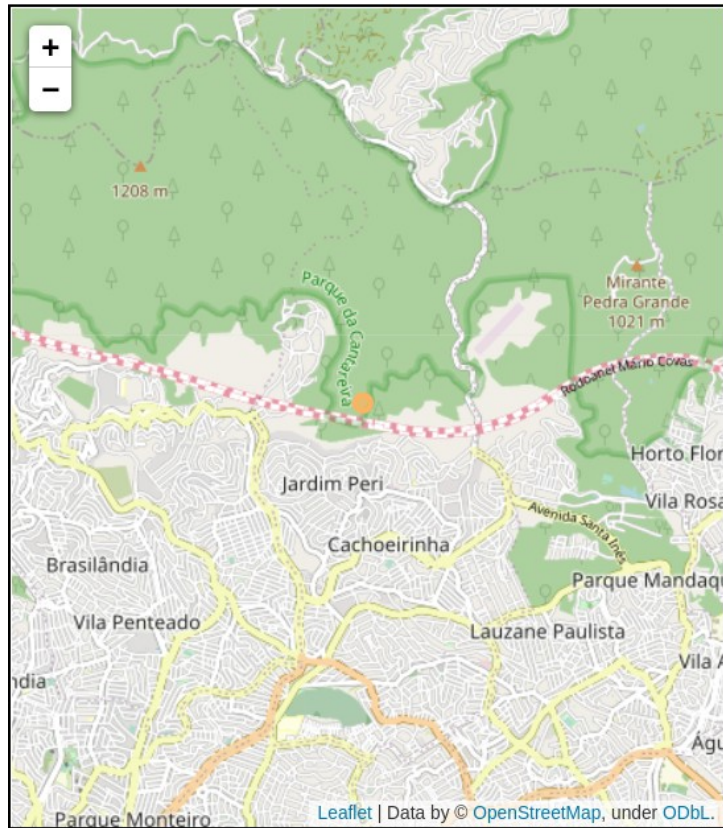
Cluster 2 – no match on SP



Cluster 3 – best matching in both



Cluster 4



Results and Discussion

- São Paulo and Camberra are very different, although both are large Metropolis, some clusters in Camberra has no equivalence in São Paulo. However all neighborhoods in São Paulo have correspondence in Camberra.
- Clusters who has large amount of elements in Camberra have few elements in São Paulo and vice-versa. Only one cluster has a significant amount of elements in both cities.
- The exchangeable neighborhoods follows:

Similar neighborhoods

- Cluster 0
 - SP neighborhoods:
 - "Morumbi" "Tremembé" "Jaguara" "Cangaíba"
 - CB neighborhoods:
 - "Aranda" "Bruce" "Cook" "Lawson" "Jamison Centre" "Page" "Acton" "Ainslie" "Barton" "Braddon" "Campbell" "Duntroon" "Capital Hill" "City" "Deakin" "Dickson" "Dickson Centre" "Downer" "Forrest" "Fyshwick" "Griffith" "Manuka" "Hackett" "Kingston" "Lyneham" "North Lyneham" "Parkes" "Red Hill" "Reid" "Russell" "Turner" "Watson" "Yarralumla" "Jacka" "Kenny" "Kinlyside" "Nicholls" "Canberra Airport" "Pialligo" "Chisholm" "Fadden" "Gilmore" "Macarthur" "Holder" "Rivett" "Stirling" "Waramanga" "Weston" "Weston Creek Centre" "Chifley" "Garran" "Hughes" "Lyons" "Southlands Centre" "O'Malley" "Pearce" "Phillip" "Woden Town Centre" "Swinger Hill" "Torrens"

Similar neighborhoods

- Cluster 3

- SP neighborhoods:

- "Aricanduva" "Carrão" "Vila Formosa" "Butantã" "Raposo Tavares" "Rio Pequeno" "Vila Sônia" "Campo Limpo" "Capão Redondo" "Vila Andrade" "Cidade Dutra" "Socorro" "Casa Verde" "Limão" "Cidade Ademar" "Pedreira" "Cidade Tiradentes" "Ermelino Matarazzo" "Ponte Rasa" "Brasilândia" "Freguesia do Ó" "Lajeado" "Guaianases" "Cursino" "Ipiranga" "Sacomã" "Itaim Paulista" "Vila Curuçá" "Cidade Líder" "Itaquera" "José Bonifácio" "Parque do Carmo" "Jabaquara" "Jaçanã" "Barra Funda" "Jaguare" "Lapa" "Perdizes" "Vila Leopoldina" "Jardim Ângela" "Jardim São Luís" "Água Rasa" "Belém" "Brás" "Mooca" "Pari" "Tatuapé" "Parelheiros" "Artur Alvim" "Penha" "Vila Matilde" "Anhanguera" "Perus" "Alto de Pinheiros" "Itaim Bibi" "Jardim Paulista" "Pinheiros" "Jaraguá" "Pirituba" "São Domingos" "Mandaqui" "Santana" "Tucuruvi" "Campo Belo" "Campo Grande" "Santo Amaro" "Iguatemi" "São Rafael" "São Mateus" "São Miguel" "Jardim Helena" "Vila Jacuí" "Sapopemba" "Bela Vista" "Bom Retiro" "Cambuci" "Consolação" "Liberdade" "República" "Santa Cecília" "Sé" "Vila Guilherme" "Vila Maria" "Vila Medeiros" "Moema" "Saúde" "Vila Mariana" "São Lucas" "Vila Prudente"

- CB neighborhoods:

- "Emu Ridge" "Dunlop" "Fraser" "Latham" "Macgregor" "Macquarie" "Melba" "Strathnairn" "Weetangera" "Narrabundah" "Franklin" "Gungahlin" "Gungahlin Town Centre" "Harrison" "Mitchell" "Moncrieff" "Throsby" "Beard" "Hume" "Oaks Estate" "Symonston" "Coombs" "Molonglo" "Sulman" "Wright" "Banks" "Calwell" "Conder" "Gordon" "Gowrie" "Greenway" "Monash" "Oxley" "Richardson" "Theodore" "Wanniassa" "Erindale Centre" "Fisher" "Harman"

Similar neighborhoods

- Cluster 4
 - SP neighborhoods:
 - "Cachoeirinha"
 - CB neighborhoods:
 - "Hall"

No matching neighborhoods

- Cluster 1
 - CB neighborhoods:
 - "Tharwa"
- Cluster 2
 - CB neighborhoods:
 - "Belconnen" "Belconnen Town Centre" "Charnwood" "Evatt" "Florey" "Flynn" "Giralang" "Hawker" "Higgins" "Holt" "Kippax Centre" "Kaleen" "McKellar" "Scullin" "Spence" "O'Connor" "Amaroo" "Bonner" "Casey" "Crace" "Forde" "Ngunnawal" "Palmerston" "Denman Prospect" "Bonython" "Tuggeranong Town Centre" "Isabella Plains" "Kambah" "Kambah Village Centre" "Chapman" "Duffy" "Curtin" "Curtin Centre" "Farrer" "Isaacs" "Mawson"

Conclusion

- Most of the people who wants to leave São Paulo will find a suitable neighborhood in Camberra, however the contrary situation don't have the same fitting.