

Coursera Capstone IBM Applied Data Science Capstone

The better sector to invest in Brasilia, Brazil.



By: Simone Pereira

2021

1. Introdução

Verificar se é possível definir os gostos dos brasilienses com base em sua renda e seus locais mais visitados.

Para muitos investidores que desejam abrir um novo negócio muitas dúvidas surgem em decidir em qual negócio investir seu capital e em qual lugar. A localização central e consumidores do lugar oferece um ótimo ponto de partida. Por isso tomaremos o local como uma variável importante para essa análise assim como a renda dos consumidores de Brasília.

O resultado da análise demonstra que com base no perfil escolhido pelo investidor o melhor local para abrir um negócio é dentro de um shopping Center em uma cidade de classe média-alta.

Problema de Negócio: O objetivo deste projeto é analisar e selecionar os melhores locais na cidade de Brasília, Brasil para abrir um novo negócio. Usando metodologia de ciência de dados e machine learning: clustering. Este projeto visa fornecer soluções para atender ao negócio.

Pergunta:

Qual o melhor negócio em Brasília baseado no gosto dos consumidores medido pela marcação de lugares mais frequentados pelo Foursquare e pela renda?

2. Materiais e Métodos

Dados utilizados:

- Dados de todas as regiões administrativas de Brasília e a respectiva renda per capita;
- Foursquare para encontrar o locais mais visitados para cada cidade e
- As coordenadas de cada cidade e cada local.

Os dados das regiões administrativas de Brasília com as respectivas rendas per capita foram extraídos da Wikipedia, endereço: ([https://pt.wikipedia.org/wiki/Lista_de_regi%C3%B5es_administrativas_do_Distrito_Federal_p
or_renda_per_capita](https://pt.wikipedia.org/wiki/Lista_de_regi%C3%B5es_administrativas_do_Distrito_Federal_por_renda_per_capita)). Foi utilizado a biblioteca BeautifulSoup para a extração.

Metodologia de análise:

Inicialmente, todas as 31 regiões administrativas (RAs) de Brasília foram usadas para a análise. Não sendo necessário excluir nenhum para as primeiras análises. Entretanto ao fazer a extração dos locais mais visitados a RA Jardim Botânico retornou vazia e não foi utilizada ao agrupar os clusters com os locais mais visitados. Os dados relativos a renda per capita foram transformados em float para servir como variável do X.

Método de clusterização: k-means.

3. Análise de Dados – Desenvolvimento

O parâmetro para o Kmeans foi o dataframe composto das seguintes variáveis:

- Renda,
- Latitude e longitude.

O número de k foi determinado através do coeficiente da silhoueta mais próximo de 1.

4. Discussão dos Resultados e Conclusão

Os resultados do k-means clusterização mostrou que podemos agrupar os perfis de clientes com base em sua renda e sua localização geográfica. Foram encontrados 3 clusters, que posteriormente, foram agrupados aos locais mais visitados em cada grupo de cluster, mostrando assim os locais mais visitados por cada grupo.

- Cluster 0 : Os locais mais visitado pela classe D:

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Bakery)	9.0	7.0	14.0	7.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Longitude, Bakery)	9.0	7.0	14.0	7.0
(Venue Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0

- Cluster 1: Os locais mais visitado pela classe A:

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Longitude, Shopping Plaza)	1.0	8.0	0.0	0.0
(Venue Latitude, Shopping Plaza)	1.0	8.0	0.0	0.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Shopping Plaza)	1.0	8.0	0.0	0.0

- Cluster 2: Os locais mais visitado pela classe B.

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0

Cluster 3: Os locais mais visitado pela classe C.

Cluster Labels	0	1	2	3
PCI	11828.8	47074.2	16643.4	15620.1
(City Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(City Longitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue Latitude, Brazilian Restaurant)	9.0	8.0	17.0	13.0
(Venue, Brazilian Restaurant)	9.0	8.0	17.0	13.0

A análise mostra que se o investidor que deseja investir em alguma cidade de Brasília deveria investir em Restaurantes que é o local mais marcado no FourSquare pelos brasilienses. E caso queira consumidores de alta renda poderá escolher como local um Shopping Center que assim obterá maior número de clientes de alta renda. Dado que a frequência de Shoppings Centers é igual à frequência de marcações em restaurantes de acordo com o FourSquare. Visto que todas as classes frequentam shopping ele terá como clientes de todas as classes.

5. Referências

Curso IBM Data Science Professional Certificate - [https://www.coursera.org/IBM Data Science Professional Certificate](https://www.coursera.org/IBM%20Data%20Science%20Professional%20Certificate).

6. Links Github

Notebook:

https://github.com/siperolli/IBM-Data-Science-Capstone-/blob/main/Final_Capstone_bsb.ipynb