Projeto Coursera Capstone

Com o objetivo estudar possíveis locais para a abertura de um negócio ainda pouco explorado no Brasil, este projeto traz o conceito do *Dry Bar*, com a proposta de identificar locais vocacionados que receberam esse tipo de estabelecimento, visando uma futura extrapolação para outras cidades.

O conceito do *Dry Bar* surgiu na Califórnia, em 2010 e se baseia na oferta de serviços rápidos de salão de beleza, a preços que prometem ser mais acessíveis do que em espaços tradicionais. Pelas facilidades, agilidade e preços convidativos, já se tornaram presença comum em cidades como Nova York e Los Angeles.

Para o desenvolvimento do projeto, a escolha foi o mapeamento de uma região da Cidade de Nova York, buscando entender elementos de geolocalização e suas relevâncias, para a instalação do *Dry Bar.* Considera-se que esse seja um serviço dependente do fluxo de pessoas do sexo feminino, com pouca disponibilidade de tempo e perfil que valoriza a boa apresentação no ambiente de trabalho. Esses são fatores determinantes, para a possibilidade de sucesso deste modelo de negócio.

**Objetivo do Projeto:**

Encontrar *Dry Bars* na região de observação, analisando suas proximidades a partir do estudo de clusters.

Uma forma encontrada para complementar este estudo, foi analisar a avalições dadas pelos clientes, utilizando a nota sete como fator de corte, considerando aí que, além da qualidade de serviço esteja implícita a facilidade de acesso, elemento relacionado a localização. Assim, o objetivo é tentar compreender a relação de localização destes agrupamentos e buscar similaridades com a região central da Cidade do Rio de Janeiro.

O ponto de partida foi o mapeamento da região de Midtown,na Ilha da Manhattan, que além de ser um dos principais pontos turísticos da Cidade de Nova York, abriga inúmeras empresas de grande porte, onde, por essas características, apresenta concentração do perfil do público para este tipo de negócio; mulheres com muitas atividades e pouco tempo para salões de beleza tradicionais.

Baseados nestes critérios, será utilizado o Foursquare, como fonte de dados destas áreas. O algoritmo de aprendizado não supervisionado K-means fornecerá esclarecimento de vantagens de cada área, para que a melhor localização possível seja definida, para futuras indicações aos interessados em implementar este negócio no Centro Rio de Janeiro.

Com base na definição do problema, as fontes de dados foram:

- Lista com os dados de localização do distrito de Manhattan

Para isso foi utilizado o arquivo Json do Curso do Coursera IBM Certificação, que fornece dados de localização dos 5 distritos de Nova YORK com os seus 306 bairros. Como projeto NÃO tem finalidade de explorar todos estes lugares, apenas um, então foi feito um filtro selecionando o distrito de Manhattan, e salvo em arquivo csv que se encontra no GitHub. (manhattan.csv)

Além disso no GitHub tem um notebook só com esta parte do processo: Extrair, filtrar e salvar estes bairros.

* -Link: <https://developer.foursquare.com/docs>
* Descrição: Foursquare API, um provedor de dados de localização, será usado para fazer chamadas RESTful API para recuperar dados sobre locais em diferentes bairros. Este é o link para a [Hierarquia da Categoria de Locais do Foursquare](https://developer.foursquare.com/docs/resources/categories) . Os locais recuperados de todos os bairros são categorizados amplamente em "Artes e entretenimento", "Faculdade e universidade", "Evento", "Comida", "Vida noturna", "Atividades ao ar livre e recreação" etc. Um extrato de uma chamada de API é do seguinte modo:
* 'categories': [{'id': '4bf58dd8d48988d110941735',
* 'name': 'Italian Restaurant',
* 'pluralName': 'Italian Restaurants',
* 'shortName': 'Italian',
* 'icon': {'prefix': 'https://ss3.4sqi.net/img/categories\_v2/food/italian\_',
* 'suffix': '.png'},
* 'primary': True}],
* 'verified': False,
* 'stats': {'tipCount': 17},
* 'url': 'http://eccorestaurantny.com',

'price': {'tier': 4, 'message': 'Very Expensive', 'currency'

Como parte do processo, próxima etapa é buscar avalições dadas pelos clientes, como notas e likes, considerando aí, de forma implícita a qualidade de serviço e a facilidade de acesso, com a intenção de justificar a hipótese de que estabelecimentos mais frequentados e curtidos estão em pontos estratégicos.

Com Foursquare e posteriormente com o algoritmo K-means, poderá ser conhecido estes pontos estratégicos.

Esta etapa requer acesso ao Fousquare mais específico, pois informações como rating e likes é preciso usar o ID do estabelecimento, então para cada nota ou quantidade de likes se faz um “request” o que diminui sua cota diária de 50 acessos ou request.