

Visualização Avançada de Dados

Miguel Galvão - miguelepgalvao@gmail.com
Davi Nascimento - davicamposd@hotmail.com

Março 2022

1 Introdução

O objetivo deste projeto é demonstrar como o dono de um negócio pode, através de técnicas de visualização de dados, retirar conclusões a partir das informações recolhidas para poder tomar decisões que possam ajudar a melhorar o serviço, reduzir despesas e até expandir o negócio. Para isso, é usado um dataset relativo a um serviço de aluguer de bicicletas na área da baía de São Francisco (*Bay Area*), nos Estados Unidos da América. Neste serviço, existem variadas estações espalhados pela zona onde é possível levantar ou devolver uma bicicleta. Assim, é possível retirar informações como a utilização de cada ponto de recolha, a disponibilidade de bicicletas, os percursos mais comuns, entre outras.

2 Trabalho Relacionado

Como inspiração para algumas das técnicas propostas deixo a referência ao material de apoio da disciplina ("T3 - design principles and marks and channels", slide 87) em que é exibido uma mapa com ligações entre várias cidades europeias e também ao gráfico exibido pelo motor de busca Google relativo à movimentação de clientes num certo estabelecimento, ao longo do dia (Figure 1).



Figure 1: Exemplos de Visualização de Dados

3 Plano de Trabalho

De modo a conseguir retirar as conclusões referidas na introdução, planeio desenvolver as seguintes tarefas:

- Um mapa da *Bay Area* de São Francisco com as várias estações de bicicletas identificadas por pontos. O tamanho do respetivo ponto estará diretamente relacionado com o tráfego relativo a essa estação. Isto permite avaliar a utilização do serviço por área e, portanto, permite retirar conclusões sobre estações que possam ser removidas ou até sobre novas estações a serem instaladas.
- A unir os vários pontos haverá linhas cuja grossura corresponde ao volume de viagens entre as respetivas estações, que permite observar quais as rotas com maior tráfego e ajuda também na análise da distribuição do serviço pela cidade.
- Ao fazer *hover* numa linha que una duas estações é apresentado o número de viagens entre as duas, para cada sentido. Com esta informação, é possível descobrir se é necessário fazer transferência de bicicletas de um ponto para outro caso, por exemplo, seja muito mais frequente o tráfego num sentido apenas.
- Ao clicar num dos pontos, a vista muda para um gráfico correspondente à disponibilidade, em média, do número de bicicletas da respetiva estação ao longo da semana, em intervalos de horas definidos pelo utilizador. Além disso, o utilizador tem também a escolha de mudar a vista para a distribuição de horas ao longo de um dia da semana apenas, podendo escolher qual. Aqui, é possível concluir se o número de bicicletas naquele ponto é adequada, reduzida ou exagerada. Além disto, é também possível avaliar qual o melhor horário para fazer manutenção ou substituição de bicicletas.
- Haverá a opção de restringir os dados em análise a partir de uma data de início e uma data de fim definidas pelo utilizador.

4 Dados

O dataset a utilizar contém dados relativos ao intervalo de tempo entre agosto de 2013 e agosto de 2015 e está disponível neste [link](#). Aqui podemos observar que a informação necessária para realizar a análise está distribuída por vários ficheiros:

- O ficheiro *stations.csv* contém informações relativas às várias estações, nomeadamente o nome, a cidade, localização e capacidade de bicicletas.
- O ficheiro *status.csv* revela, para cada estação, quantas bicicletas (e quantos espaços livres) existem, num determinado momento. Estes estão separados em intervalos de aproximadamente 1 minuto.

- O ficheiro *trip.csv* contém informações acerca das viagens realizadas, tendo uma referência para o instante e a estação de partida, assim como o instante e a estação de destino.
- Existe também um ficheiro com informações sobre a meteorologia que não será utilizado neste projeto.

Além destes ficheiros em formato csv está também disponível uma base de dados no formato SQLite com toda a informação centralizada que poderá ser usada para fazer as filtrações necessárias para extrair a informação.