

Visualização Avançada de Dados - Meta 1

Tiago Fernandes
2017242428

2020/2021

1 Definition of target users and a problem that the visualization should solve – conceptualization

As visualizações a realizar serão para ser apresentadas à Câmara Municipal do Porto, com vista a mostrar dados sobre o impacto da pandemia COVID-19 durante 1 ano (desde o dia 9 de março de 2020 até 9 de março de 2021).

2 Developing of analytical questions for a data analysis and/or communication

- Preço médio por zona em cada mês, conseguindo até reconhecer alguns padrões como quando Portugal esteve num período de confinamento por exemplo
- Variação dos preços de cada casa individualmente, podendo até perceber com base na localização da casa se houve algum evento ou atividade que possa ter feito variar esse valor
- Variação mensal e por zona dos preços das casas, conseguindo perceber o desvio padrão e o custo que os clientes das várias habitações de Airbnb pagam enquanto visitam a região do Porto (se forem turistas)

3 Definition of task abstractions, as well as specification of the requirements

- Nas representações geográficas, de forma a minimizar os cliques desnecessárias, o mapa pode estar focado, por default, na região abrangida, contudo, é também possível fazer zoom in e zoom out no mapa
- Nas várias visualizações que vão ser criadas, a possibilidade de criar, ajustar e remover filtros (exs.: filtragem por preços, filtragem por meses, ...)

4 Definition of data and acquisition of the necessary data, its parsing and filtering

Os dados relativos aos *Airbnb* da região do Porto foram obtidos através do site fornecido pelos docentes (todos os ficheiros *calendar.csv.gz* e *listings.csv.gz* desde 27 de fevereiro de 2020 até 13 de fevereiro de 2021 e ainda os ficheiros *neighbourhoods.csv* e *neighbourhoods.geojson* de 13 de fevereiro de 2021). Nos ficheiros de *calendar* e *listings* estão informações relativas aos apartamentos como preços, localização, ... Já nos ficheiros *neighbourhoods* estão dados geográficos como limites das zonas quais freguesias pertencem a cada zona.

Estes ficheiros requerem algum tratamento como eliminar colunas desnecessárias, corrigir os dados de strings para float (preços) e ainda corrigir o formato das datas.

No fim, juntar os dados de ambas as tabelas tudo numa só tabela para facilitar o processamento dos dados como está no exemplo a seguir:

	listing_id	date	available	price	adjusted_price	minimum_nights	maximum_nights	neighbourhood_cleansed	neighbourhood_group_cleansed
0	510404	2020-03-09	t	25.0	25.0	4.0	1125.0	Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, V...	PORTO
1	510404	2020-03-09	t	25.0	25.0	4.0	1125.0	Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, V...	PORTO
2	510404	2020-03-09	t	25.0	25.0	4.0	1125.0	Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, V...	PORTO
3	510404	2020-03-09	t	25.0	25.0	4.0	1125.0	Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, V...	PORTO
4	510404	2020-03-09	t	25.0	25.0	4.0	1125.0	Cedofeita, Ildefonso, Sé, Miragaia, Nicolau, V...	PORTO

Figure 1: Tabela com a junção dos dados dos ficheiros *calendar* e *listings*

5 Identification of an appropriate solution for each of the views of the visualization application

- Saber o preço médio por zona - representação geográfica
- Flutuação do preço médio de uma casa em específico - representação temporal
- Estatísticas mensais (média e desvio padrão)

6 Implementation of one of the chosen views in Plotly (no Dash required)

Nota: A label que aqui aparece é temporária, será mais tarde alterada.

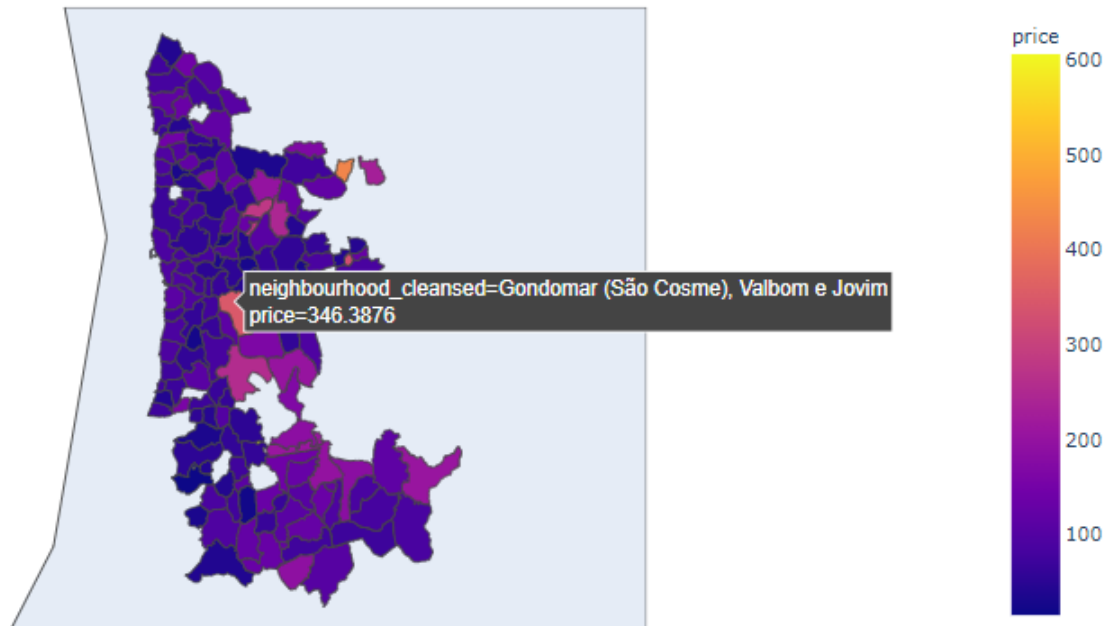


Figure 2: Exemplo da representação geográfica