VISUALIZATION TOOL FOR TAXI-PESSENGER DEMAND PREDICTION

Shaheer Asghar Awan, Robin Grau, MH Nikbakhtmalvajerdi, Mehdi Alijani Renani

Motivation

Cities have a lot of taxi traffic with different demands through the areas. Although the amount of taxis needed changes all the time, there are patterns to detect. Some of this patterns are easy to find with an algorithm but hard to find for humans but some patterns are easier to find for humans. That's why we combined this two methods and predict the taxi demand at the given taxi stands for the next time interval and also visualizes it so humans can also detect patterns.

Data transformation

The first step is to transform the data to a more useful form. We used a data set of taxis in Porto[4] for our project. In the transformation we deleted some useless attributes from the data and counted the amount of taxis that started per taxi stand per month as shown in Figure 1.

	A	B	C	D	E	F	G	H
1	TRIP_ID	CALL_TYPE	ORIGIN_CALL	STAND	TAXI_ID	TIMESTAMP	LONGITUDE	LATITUDE
2	1.37263730362E+018	B	0	7	20000596	1372637303	-8.639847	41.159826
3	1.37263736962E+018	B	0	53	20000381	1372637369	-8.61399	41.141313
4	1.37263761062E+018	B	0	13	20000497	1372637610	-8.585145	41.164857
5	1.37263848162E+018	B	0	28	20000403	1372638481	-8.584263	41.163156

Figure 1: Data before transformation

	A	B	C	D	E	F
1	Stand ID	Stand name	Latitude	Longitude	passengers this month	year - month
2	53	Ribeira	41.1412082	-8.614012261	6809	2013-07
3	8	Azevedo	41.1515921	-8.568195397	1664	2013-07
4	34	Infante	41.14051688	-8.615952191	3267	2013-07
5	18	Carregal	41.14832091	-8.619603476	4290	2013-07

Figure 2: Data after transformation

Visualization

For the visualization we created a html page with JavaScript and the following libraries:

- D3.js [1]
- Leaflet [2]
- Mapbox [3]

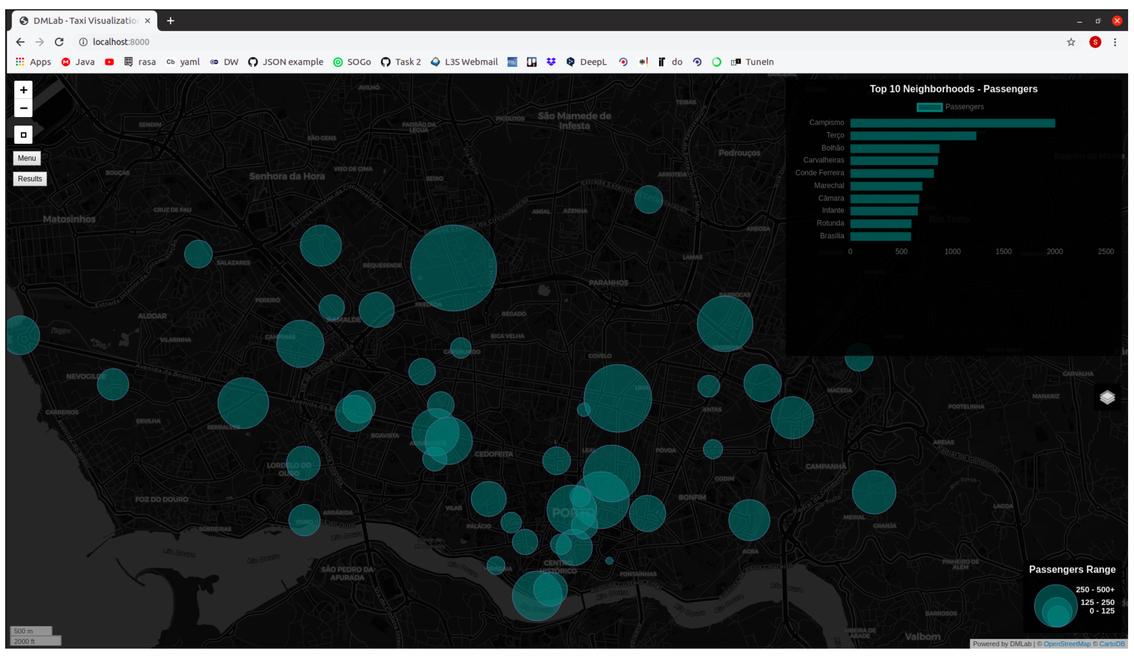

Figure 3: Application

Each blue circle in Figure 3 represents one taxi stand and if the mouse hovers about the one of the circle the exact amount of taxis started from that taxi stand at the selected time will be shown. In the top right are the 10 most used neighborhoods at the current selected time. Below the bar chart is a little icon to change the kind of background map for a pleasures user experience. In the bottom left is the legend for the circle sizes. In the top left is a slider to change and select the shown time. Left to the slider is the legend for the distance in the map. It is possible to zoom in and out with the "+" and "-" in the top left corner.

Prediction

We didn't concentrated on the prediction so our prediction algorithm is really simple. To predict the demand of taxis at each taxi stand for the next month we just took the average from the taxi stand at it's neighbors at the past. Neighbors of a taxi stand are defined as taxi stands with a haversine distance below 500 meters. Haversine distance is an algorithm created by José de Mendoza y Ríos in 1801 to calculate the distance between two points on a sphere given their longitudes and latitudes.

```
9,57

10,14,19,27,60

11

12,13,30,32,54

13,12,21,32,54

14,10,23,27,57,60
```

Figure 4: Neighbors of taxi stands

In Figure 4 you can see a few taxi stands and their neighbors. The first number is the ID of the taxi stand and the following numbers represent the IDs of the neighbors. Since the neighbors are based on distance the amount of neighbors differs on the taxi stand and can also be zero, like the taxi stand with the ID 11.

Find important places

To find important places we clustered the neighborhoods of taxi stands and added the amount of taxis that started from the neighborhoods. The neighborhoods with the most taxis are called the most important neighborhoods. For this places we looked to find out why that locations are important. To do that we took the median coordinates and reversed searched them with the help of the geopy library for python. The code only returns the area of the location and not if there are public places like a university, church or public park. To check if an public place is at one of the important places we still would have to search for public places manually. We didn't see a gain in doing this since we still had to do the search for public places manually and therefore stroke it out of the project.

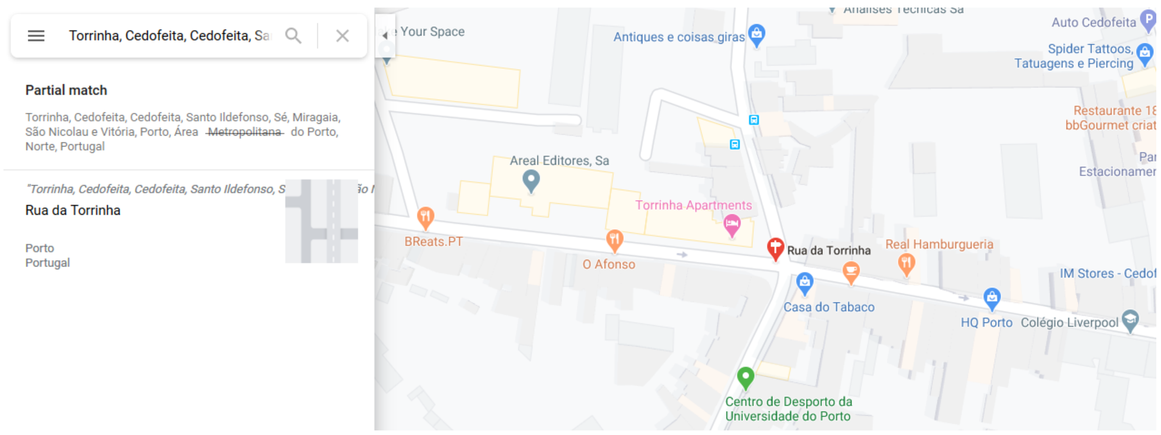

Figure 5: Reverse search for a important location

In Figure 5 you can see an example of the reverse search of an important location. You can see that the "Colégio Liverpool", the "Centro De Desporto Universidade Do Porto - Boa Hora" and the huge street "Rua da Torrinha" are near the center of the important neighborhood.

References

[1] D3.js.

[2] Leaflet, 2010.

[3] Mapbox, 2010.

[4] Taxi service trajectory prediction challenge, ecml pkdd 2015, 2015.