Independent assessment of the Spanish road system

Analysing the characteristics of the existing roads and identifying an optimal warehouse network.

acalleja@outlook.es

Independent assessment of the Spanish road system:

Analysing the characteristics of the existing roads and identifying an optimal warehouse network.

INTRODUCTION:

IMPROVE AND SAVE IN THE LAST MILE DELIVERY SYSTEM WITH AN OPTIMIZED LOCATION OF WAREHOUSES.

Distribution networks are becoming more and more efficient worldwide. Its development allows international commerce to flourish and allow for goods produced on one side of the world to be sold in another.

Freight rail networks and container ships are very efficient ways of transporting goods through very long distances. However, once these goods arrive to high capacity freight stations of ports, they have to be delivered to the final delivery point, which may be a person's house or a store in a town away from the main cities, i.e. away from the main ports.

This final part of the supply chain is usually referred as the "Last Mile Problem", since this last step often accounts for a very high percentage of the total cost of moving the goods, despite its relatively small distance. Different studies suggest that this could be around 40% of the total transportation costs.

This last mile delivery is most often accomplished by means of road transport, which suggests the importance for assessing the road system of a country and being able to evaluate the optimal routes in order to organize the logistics of a company.

In this project, we will analyse the Spanish road system, to find how well or bad are all the different villages and towns away from the main cities connected to the main road system. We will use all the data gathered to identify the best locations for the warehouses of a company that aims to give delivery service to all the villages, towns and cities of Spain.

This would allow for a company to optimize its delivery system and data-wise tackle the "Last mile problem", saving by optimizing the last leg of the supply chain.

DATA GATHERING:

INFORMATION ABOUT THE SPANISH GEOGRAPHY INHABITED PLACES AND THEIR CONNECTIVITY TO THE ROAD SYSTEM AND ITS SERVICES

We will get data from 3 main sources: Databases, APIs and Webpages.

DATABASES:

NATIONAL STATISTICS INSTITUTE OF SPAIN (INE):

This public entity that gathers information about the population and demography of Spain provides a database from which we will be able to download information that includes:

A list of every single city, town and village in the country with:

- Its location (latitude and longitude)
- Altitude over the sea level
- Population

We will download such information as a .xls file that we import to our project.

The following image displays an example of the data obtained from this database:

	Comunidad	Provincia	Población	Latitud	Longitud	Altitud	Habitantes	Hombres	Mujeres
0	Andalucía	Almería	Abla	37.14114	-2.780104	871.16840	1504	783	721
1	Andalucía	Almería	Abrucena	37.13305	-2.797098	976.93870	1341	682	659
2	Andalucía	Almería	Adra	36.74807	-3.022522	10.97898	24373	12338	12035

APIS

Foursquare API:

We will use this API to get information about the different services to be found in the different roads, such as service and filling stations. we will have to take into account this information when looking for optimal delivery routes.

It follows an example of the data that we get from the API when we ask for a gas station within 5km of a Spanish village with coordinates: (37.14114, -2.780104).

It is the text extracted from a JSON file:


```
{'meta': {'code': 200, 'requestId': '5e88d8f102a172002869bedc'},
 'response': {'venues': [{'id': '4ea001d1be7b667c5fc2a865',
    'name': 'GASOLEOS ABLA, S.L.',
    'location': {'address': 'N324 km 268,100',
     'lat': 37.13475,
     'lng': -2.767694,
     'labeledLatLngs': [{'label': 'display',
       'lat': 37.13475,
       'lng': -2.767694}],
     'distance': 1311,
     'postalCode': '04510',
     'cc': 'ES',
     'city': 'Abla',
     'state': 'Andalucía',
     'country': 'España',
     'formattedAddress': ['N324 km 268,100',
      '04510 ABLA Andalucía',
      'España']}}
```


Trueway Matrix API:

Given the location of two points, this API provides information about the distance and the time needed to perform a road trip between the two of them. we opted for this API instead of the more known google maps because it is free.

It follows an example of the data that we get from the API when we ask for the connection between the Spanish village with coordinates (37.14114, -2.780104) and another one with coordinates (39.14114, -1.780104).

The response is obtained as plain text:


```
{
"distances":[[376128]]
"durations":[[14534]]
}
```

WEBPAGES:

WIKIPEDIA WEBPAGE:

We will scrape Wikipedia webpage in order to get further statistical information about each Spanish region, including:

- GDP per capita in each region

We will use this data to look for correlation of some macroeconomic indices to the connectivity of the roads, which will help us in better understanding the structure of the road system.

The following image displays an example of the data obtained from this webpage:

	Population	Latitude	Longitude	GDP per capita
Comunidad Autonoma				
Andalucía	8409000	37.367	-5.983	18470

METHODOLOGY

Given the sources of data data described in the previous section, we analysed it to reach our project objective.

The structure of the data analysis process can be structured in three main steps.

1.- Connectivity between capital cities

The Spanish geography is organized in regions, called "Comunidades Autónomas". Each of these regions has a capital city which centralises a lot of its resources.

We analysed how all those capital cities are connected via road transport. For that, we used the Trueway Matrix API to evaluate the time and distance travel for every pair of capital cities.

We used that to obtain the following features to assess the connectivity between the cities:

- **Average travel speed** for a trip between every two capital cities.
- **Road distance / Straight distance ratio**: this is a measure of how direct are the roads between every two capital cities.

In addition to the pairwise trip evaluation, with which we obtained the quality of the connection between cities, we also evaluated the average for all the trips from each capital city in order to look for correlations in the variables. To do so, evaluated the **Pearson correlation.**

The discussion of the findings can be consulted in the following section.

In order to transform this analysis into a clear input for our warehouse location model, we combined the average travel speed and the distance ratio into one single variable called **mark**.

To evaluate mark, we normalized the travel speed and the inverse of the distance ratio dividing both of them by their maximum value and multiplied them, getting an order for the connectedness of each capital city with each other’s.

OUTPUT OF THE ANALYSIS → COMPARISON OF THE QUALITY OF THE ROADS CONNECTING EACH REGIONS

MARK: SINGLE NUMBER TO COMPARE HOW WELL CONNECTED ARE THE CAPITAL CITIES ACCORDING TO THE AVERAGE SPEED AND DISTANCE RATIO OF THEIR TRAVELS.

FEATURES CORRELATED WITH THE DEVELOPMENT OF ROADS IN EACH REGION

2.- Connectivity of towns and cities

We repeated a similar analysis, but in this case, we checked the connectivity of towns and cities within a region to the capital city of each region.

We calculated the **average travel speed** and **road distance / straight distance** order to find how are the last mile connections, since we want to design a warehouse system that can reach every single place most efficiently.

The calculations were performed for the 7955 towns and villages in the Iberian peninsula.

Similarly to the previous case, we evaluated the Pearson Correlation for each of the coefficients.

In this case, since we obtained a lot of geographical information about the roads, we displayed it on a map together with the existing roads maps to visually analyse it.

By taking the average of all the trips within the region, we could also evaluate the mark for the transportation within the Comunidad Autónoma.

OUTPUT OF THE ANALYSIS → VISUAL DATA DISPLAYING THE GEOGRAPHICAL INFORMATION OF THE ROADS

MARK: SINGLE NUMBER TO COMPARE HOW WELL CONNECTED ARE THE CAPITAL CITIES ACCORDING TO THE AVERAGE SPEED AND DISTANCE RATIO OF THEIR TRAVELS.

3.- Final output

Given the results of:

- How well connected is every village/town with the main city of its region.
- How well connected is the capital city of a region with all the other capital cities.
- The population of each village/town.

For a given number of warehouses, we evaluate the optimum locations of the warehouses such that the higher amount of people can be reached in the minimum time.

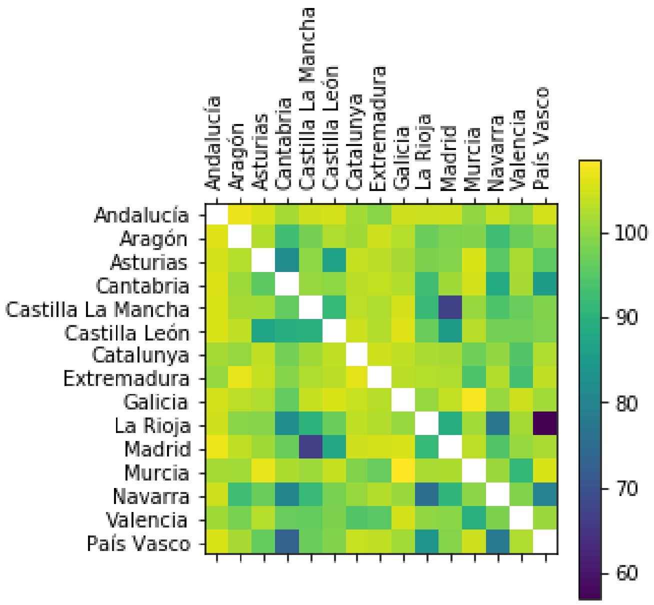
In particular, we order the regions according to the priority given to set a warehouse there.

RESULTS

1.- Connectivity between capital cities

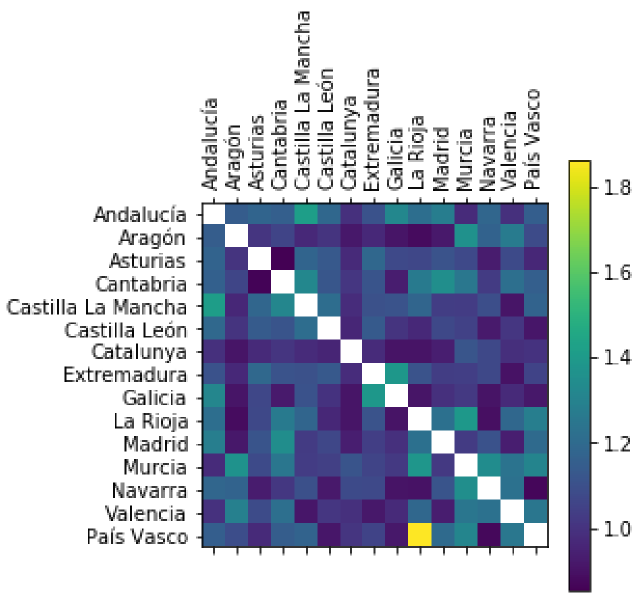
The results for the connectivity are displayed in the following confusion matrices.

AVERAGE SPEED BETWEEN REGIONS (KM/H)



The highest average speed is found for Andalucía, followed by Galicia and Extremadura. The slowests are Navarra, La Rioja and Cantabria.

DISTANCE RATIO BETWEEN REGIONS



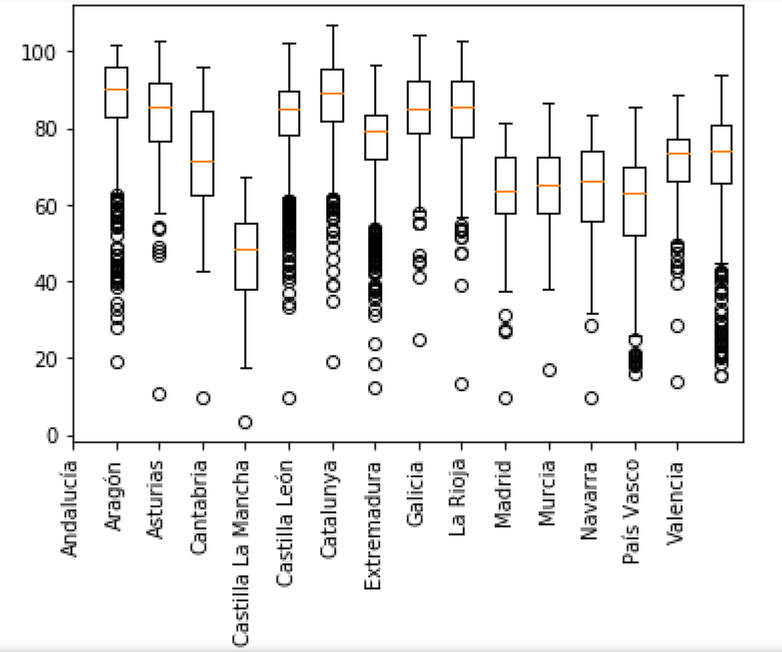
The results are summarized in the following dataframe, which displays both the raw average speed (km/h) and distance ratio, as well as the normalized values and the final mark used to order them.

	Comunidad Autonoma	Distance ratio	avg speed	Normalised inv ratio	Normalised avg speed	Mark
0	Galicia	1.027165	103.471848	0.989737	0.995182	0.984968
1	Catalunya	1.016623	101.887530	1.000000	0.979944	0.979944
2	Aragón	1.051073	100.705957	0.967224	0.968580	0.936834
3	Asturias	1.061217	99.619665	0.957978	0.958132	0.917870
4	Extremadura	1.089043	102.033133	0.933501	0.981345	0.916086
5	Castilla León	1.051858	98.353837	0.966502	0.945958	0.914270
6	Valencia	1.089710	98.657462	0.932930	0.948878	0.885236
7	Madrid	1.089939	97.104272	0.932734	0.933939	0.871117
8	Andalucía	1.167775	103.972775	0.870564	1.000000	0.870564
9	Navarra	1.062272	94.328492	0.957027	0.907242	0.868255
10	Murcia	1.162912	101.304282	0.874204	0.974335	0.851768
11	Castilla La Mancha	1.124226	96.554913	0.904287	0.928656	0.839771
12	País Vasco	1.118955	96.004446	0.908547	0.923361	0.838917
13	Cantabria	1.132714	95.596480	0.897511	0.919438	0.825205
14	La Rioja	1.129471	94.620098	0.900088	0.910047	0.819122

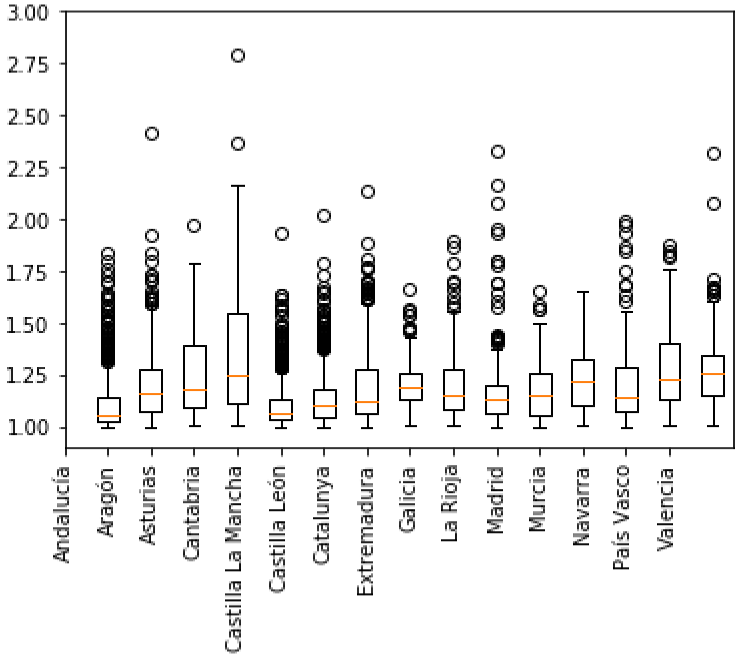
2.- Connectivity within regions

The results for the connectivity are displayed in the following box plots, where each point is one town/village:

AVERAGE SPEED TO CAPITAL CITY (KM/H)



DISTANCE RATIO BETWEEN REGIONS



The results are summarized in the following dataframe, which displays both the raw average speed (km/h) and distance ratio, as well as the normalized values and the final mark used to order them.

	Comunidad autónoma	Ratio	avg speed	norm inv Ratio	norm speed	Mark
0	Extremadura	1.179595	83.977968	1.256695	0.957027	1.202691
1	Galicia	1.178737	83.795579	1.255781	0.954949	1.199206
2	Aragón	1.175917	83.958242	1.252777	0.956802	1.198660
3	Andalucía	1.091762	86.501342	1.163121	0.985784	1.146586
4	Castilla León	1.065361	87.748785	1.134994	1.000000	1.134994
5	País Vasco	1.280485	70.134526	1.364179	0.799265	1.090341
6	Valencia	1.256281	70.584280	1.338393	0.804390	1.076590
7	Asturias	1.230082	71.766022	1.310481	0.817858	1.071787
8	Catalunya	1.157771	76.161291	1.233445	0.867947	1.070565
9	Castilla La Mancha	1.067705	82.513091	1.137492	0.940333	1.069621
10	Murcia	1.283979	63.048620	1.367901	0.718513	0.982855
11	Madrid	1.146202	64.966052	1.221119	0.740364	0.904073
12	Navarra	1.188964	59.973314	1.266676	0.683466	0.865730
13	La Rioja	1.104675	62.867519	1.176877	0.716449	0.843173
14	Cantabria	1.319999	46.566628	1.406275	0.530681	0.746284

Regarding not the average values, but the results for each village, we display them plotted in the following maps, which display the characteristics of a trip departing from each of the markers to the capital city of the region.

AVERAGE SPEED TO CAPITAL CITY (KM/H)

DISTANCE RATIO BETWEEN REGIONS

DISCUSSION

The road system in Spain has a radial shape with its centre in Madrid. Other capillary roads leave from this radial road connecting the locations that are not on a main radial.

When assessing the preference in order to locate a warehouse we have to take this into account, as there are two main clusters of locations:

- those well connected to Madrid via the main roads
- those which are les well connected to this main system

The analysis of the connectivity between capital cities allowed us to identify those regions. We could see from the correlation of the variables that this fact is no linearly correlated to the GDP or the population of the regions, but instead to the orography of them: Spain is a mountainous country and therefore some connections are less favoured due to geography.

This is especially true for a group of regions in the north, comprised by: La Rioja, Cantabria, País Vasco and Navarra.

Regarding the communications within each region, we note that the communications around the main cities tend to be slower due to the traffic and the fact that the speed tends to be more regulated, while more desertic areas display higher average velocities as expected.

Again, another relevant factor is the geography of the areas, since regions with mountains tend to have less connectivity. This greatly affects Cantabria, la Rioja and Navarra, as can be seen in the plots.

Regarding the presence of gas stations, we found that it is not a very relevant feature since they are very present throughout all the road system in Spain.

Given the results, the recommended order in which to locate the new warehouses would be:

1. **Madrid**
2. **País Vasco**
3. **Cataluña**
4. **Andalucía**
5. **Murcia**
6. **Galicia**
7. **Castilla la Mancha**
8. **Castilla León**

CONCLUSION

In this project we analysed the Spanish road system and developed a rating system to asses the location for a set of warehouses (Madrid, País Vasco, Cataluña, Andalucía, Murcia, Galicia, Castilla la Mancha, Castilla León).

We concluded that the main features affecting the quality of the roads are not as related as to the GDP and wealth of each region, but to the geography (presence of mountains, etc.).

For a future study, it would be interesting to combine our findings with further information, such as the ownership (public or private) of the roads analysed.

This study was done as part of the Coursera Capston project.