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:

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

DATA ANALYSIS

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

1.- Connectivity of 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 will analyse how all those capital cities are connected via road transport. For that, we will use the Trueway Matrix API to evaluate the time and distance travel for every pair of capital cities.

We will use 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.

We will also check the presence of **gas stations** with the FourSquare API.

We will compare the value of our new features with the GDP per capita of this area to find any possible existing **correlations**.

2.- Connectivity of towns and cities

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

We will calculate 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.

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 will use this data to evaluate the optimum locations of the warehouses such that the higher amount of people can be reached in the minimum time.