RedOptima Placement Work Sample

RedOptima uses a variety of data sources to calculate objective contextual information related to the physical conditions surrounding a vehicle at a place and time. A deep understanding of the road network is important because it brings detailed information such as what options a driver has when making a journey, how close a point is to a junction, and how likely it is that a given driver is travelling at high speeds given traffic.

A fundamental part of deriving insight from network-style data is handling messy data. In geospatial domains this is often volunteered by organisations and is not always provided in a user-friendly or analyst-friendly manner. Handled correctly, such information can be deeply insightful.

The UK provides open road traffic data, specifically measurements of Average Annual Daily Flow, the number of vehicles that will travel along a road length on an average day of the year. The data contains columns listing AADF for vehicle types and text identifiers for the two ends of each road junction (StartJunction, EndJunction). Other columns detail, for example, the length of the road in question, and co-ordinates of the junction in Easting and Northing (see https://en.wikipedia.org/wiki/Easting_and_northing).

Your task is to:
1. Download an AADF dataset for any region ([https://www.dft.gov.uk/traffic-counts/download.php], top right column on the table under "Download traffic datasets")
2. Construct a network representation of the road network using the python packages NetworkX or NetworKit
3. Perform an analysis which identifies:
    a. which of the network metrics which can be calculated using network representation packages, if any, correlate with road network AADF, and how strongly they do so
    b. whether there are significant differences between different road categories

Consideration should be given to the performance and efficiency of the code and how simple your solution is to scale up and automate, and to enable it to process much larger datasets (as a hint/example, if you use jupyter notebooks for analysis we would strongly encourage providing a scalable executable version of your solution code outside of your analysis notebook). Consideration should also be given to handling data quality and issues with data integrity e.g. missing data.

You should attach all the code you have written and a brief 2-3 page summary of:
- your method
- your findings
- how you would extend your solution and perform further improvements given more time
- how you have designed your analysis to be scalable on larger datasets, identifying relevant limits and workarounds

You have **10 days** to complete and return your submission in a compressed file to Navin.Seeburn@redoptima.ai. Enquiries should also be sent to that email address. This work sample is confidential, and the instructions, data and solution shouldn't be shared with anyone other than yourself.