Transport and Mobility analytics

Christos Kargidis, Msc in Big Data

UNIVERSITY *of* STIRLING

Problem & Aims

Traak Systems obtains a large dataset from various transport sources, both real-time and historical. The data relates to various travel modes, e.g. car and bus. The goal is by applying data science techniques to this data, to baseline normal operation and identify anomalies in real time that should be investigated further. Also to investigate the potential for fusing different data sets and looking at new outcomes and insights.
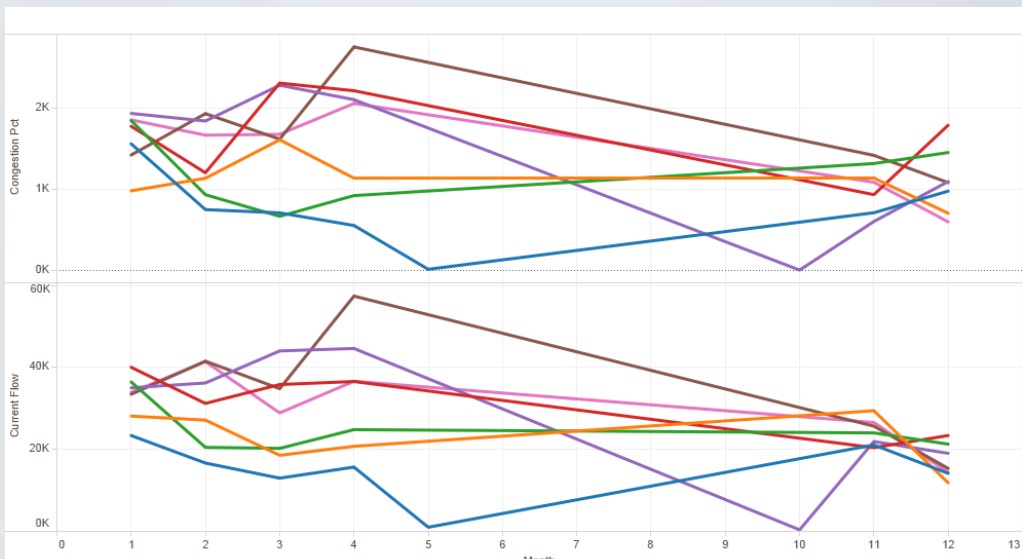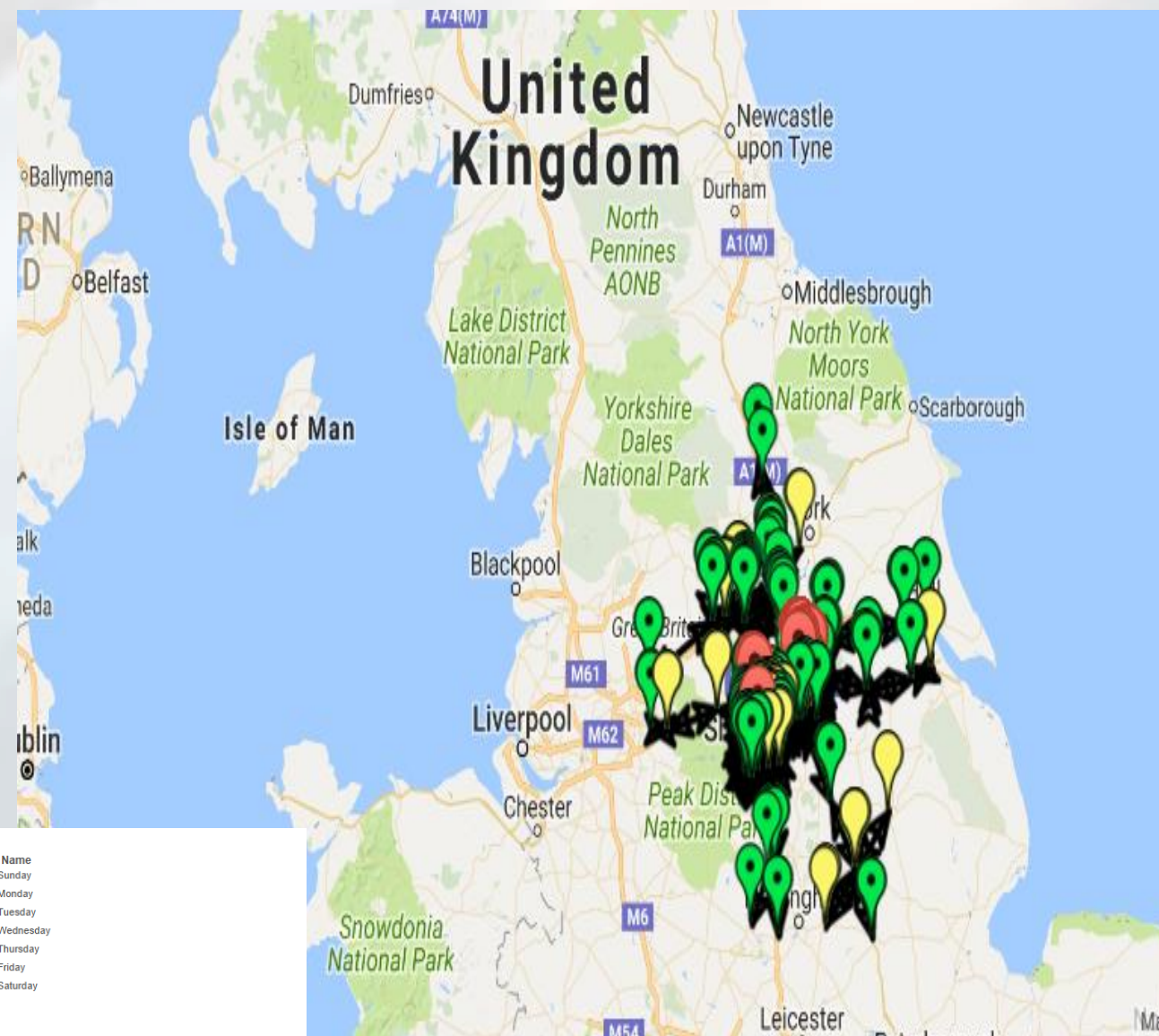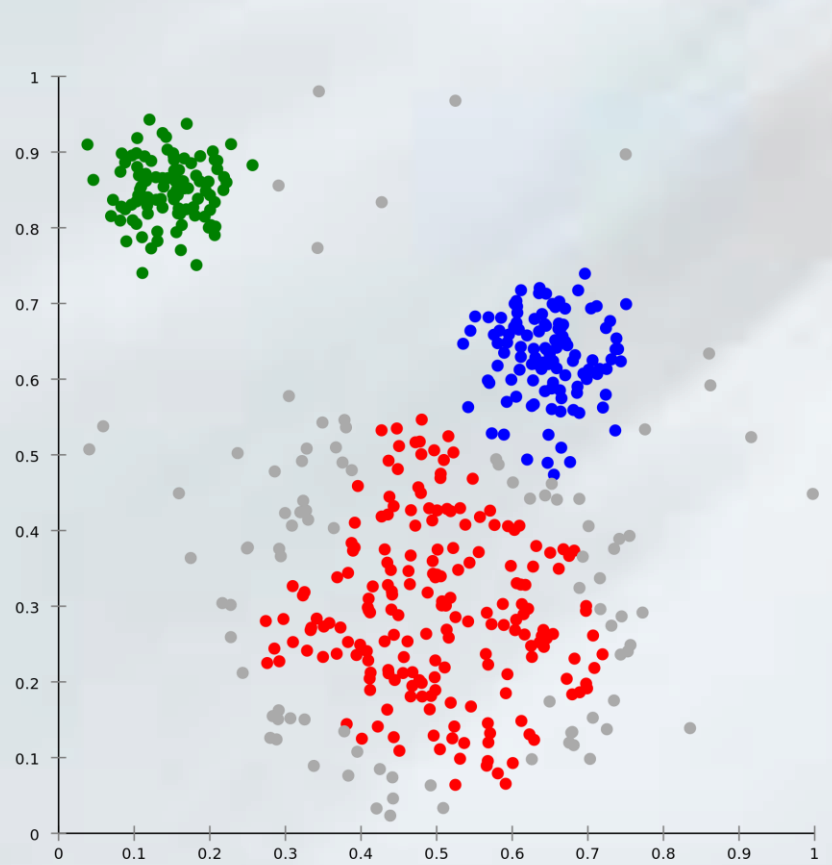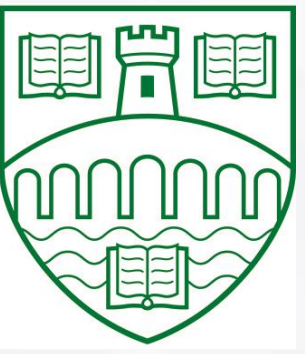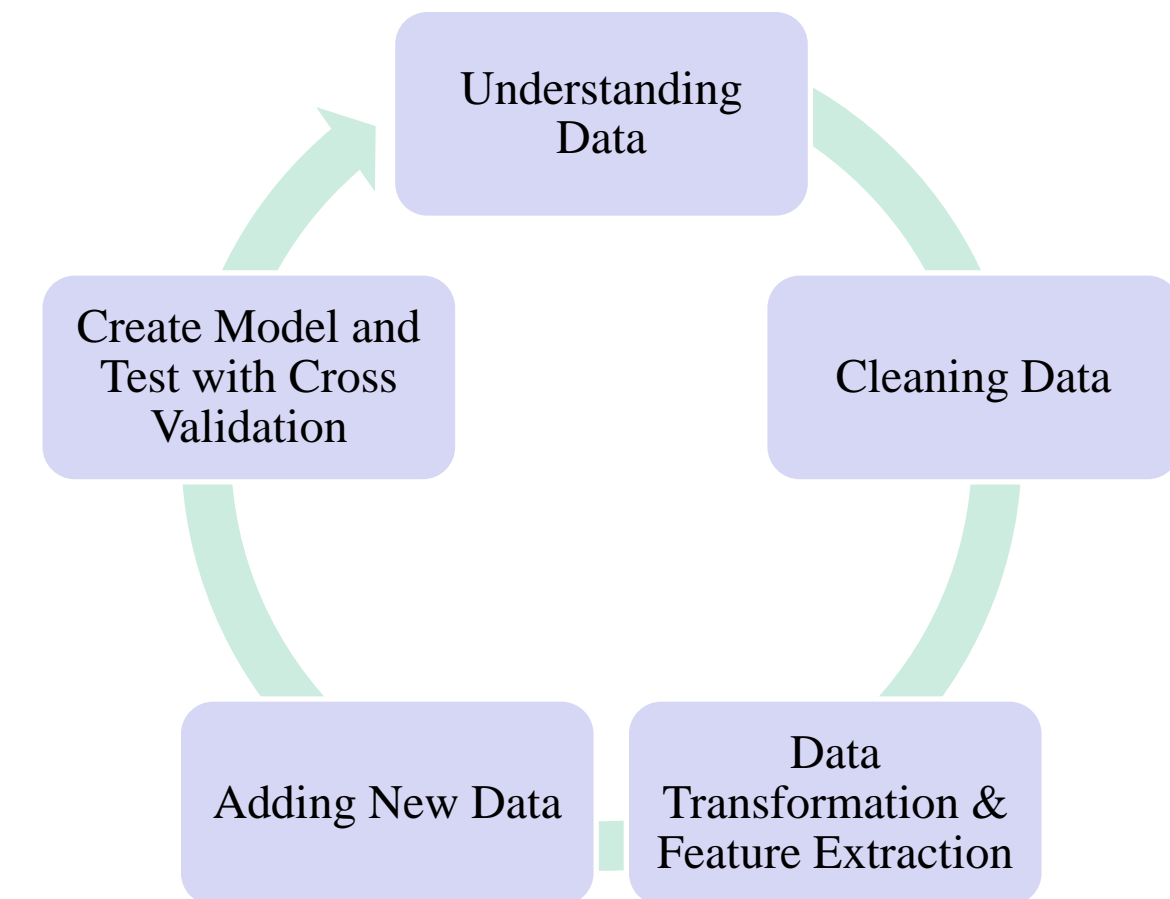
Approach

- **Understanding the Data**: Study the data set and understand the data it represents
- **Cleaning the Data**: spot and eliminate null values and outliers
- **Data Transformation & Feature Extraction**: Using the existing dataset, create new features, encode or scale existing ones, transform categorical data
- **Adding New Data**: Add other data sets and combining with the existing one (e.g. geospatial data, adding Holiday timestamps etc.)
- **Create Model**: Experiment with different algorithms (Linear Regression, DBSCAN clustering etc.) to cluster the data depending the location and spot patterns, anomalies and correlations between the features of the datasets.