**Student Project Details**

**About us:**

~~Co Wheels is a sustainable transport pioneer which aims to move people from owning cars to using shared transport.~~

~~We are one of only three national car clubs operating in the UK, with market leader Zipcar, which is focused on London, and Enterprise Car Club which also uses its extensive rental office network.~~

~~Co Wheels was set up more than 13 years ago in Durham, and now has more than 600 vehicles and 30,000 members based across the UK from Orkney to the Isle of Wight, including 3 cars in Lancaster, and a third of the fleet are EVs. There are two sides to the business, B2C for public cars and B2B closed fleets for employees of bodies like councils, NHS trusts, Universities, and housing providers.~~

**Project Proposal**

**Title:** Optimising Car Sharing Profitability with a Regional Pricing Strategy

**Previous Title:** Develop a regional tool for car sharing

**Deliverable**: A usable pricing tool to help us model the effects of different pricing options on profitability

**Background:** Car share is price sensitive and seasonal as many B2C journeys are for social/discretionary use such as shopping or leisure trips. We have a significant amount of trip data on how and when our vehicles are used, plus a clear idea on our costs to provide the service both fixed and variable, but we have traditionally charged one fixed price across all locations even though demand and local economies vary enormously.

**Why is this piece of work required:** We want to look at how best to use our data on **usage and demand patterns** to better inform our pricing strategy. Ideally, we would want to develop a straightforward pricing tool where we could input all costs – including variables such as fuel/electricity – and output options for hourly and daily pricing based on locations, demand, seasonal variations or potentially time of day (e.g. cheaper overnight hire).

We would want this to give us a clearer idea of profitability in locations and the potential outcomes if we adopted different pricing options and what the impact of this might have on utilisation.

~~Most data are available in our booking system TripIQ and can be exported as CSV files or potentially direct from our SQL servers. We also have additional socio demographic data on locations in our GIS system QGIS. It may also need further desk research on competitor data to cost a sample of alternative travel options, as well as studying current user’s surveys or even forming new surveys for our customers.~~

Input:

1. Booking lengths
2. No of booking
3. Miles travels
4. Mileage

**Vital URL:**

1. [Dynamic pricing through data science](https://www.youtube.com/watch?v=cKPaIsOQslo)

**Data Science Project Steps:**

1. Problem Statement
2. Data Collection
3. Data Cleaning & Processing:
   1. Remove duplicate values
   2. Remove irrelevant observations
   3. Address missing values (e.g. Imputation techniques, drop features / observations)
   4. Data Standardisation and Normalisation
   5. Encoding Categorical Variables
   6. Handling data types
   7. Filter unwanted outliers
   8. Reformat strings
   9. Validate
4. Exploratory Data Analysis (EDA):
   1. Inspect all data features
   2. Define data properties
   3. Build confidence in data
   4. Conduct sanity checks
   5. Figure how to handle each feature

**NOTE:** Univariant analysis, Bivariate analysis, Missing values treatment, Outlier treatment, Variable transformation, and correlation analysis.

***Univariant Analysis:***

**Observation of the population and understand the sample distribution:** measure of spread, measure of central tendency, outlier detection.

**Graphical Analysis:** Histogram, Boxplots, Stem and leaf

***Multivariant Analysis:***

**Relationship between two or more variables:** Covariance and Correlations

**Graphical Analysis:** Bar plots, Scatterplots

1. Feature Engineering:

**Approaches:**

* + 1. **Feature Extraction:** Select and/or combine variables into features to reduce dimensionality reduction of dataset (e.g. PCA, Nonlinear dimensionality reduction, unsupervised clustering methods.)
    2. **Feature Selection:** Select the features which contributes most to the problem (e.g. Variance thresholding, Pearson correlation, LASSO)
    3. **Feature Construction:** The process of manually building more efficient features from raw data (e.g. Dynamic aggregation of relational attributes)
    4. **Feature Learning:** The automatic identification and use of features (e.g. Restricted Boltzmann machine, K-means clustering)

**Methods:**

* + 1. The correlation coefficient between the feature and the target variable (the feature that you are trying to predict)
    2. Co-integration between two time series (for time-series data)
    3. Predictive models have embedded feature selection methods (e.g. Random Forest, Gradient Boosting Machine)
    4. Chi-Squared test (between target and numerical variable)
    5. Recursive Feature Elimination (EXPLAIN)

1. Model Building & Evaluation
2. Model Results
3. Model Deployment & Maintenance