Project Summary

This project focuses on the analysis and visualisation of licensing and compliance activity within the Small Public Service Vehicle (SPSV) sector in Ireland. The SPSV sector includes taxis, hackneys and limousines, and is regulated to ensure passenger safety, service quality and public confidence.

The aim of the project was to design and develop a data-driven analytics and visualisation solution that could support regulatory decision-making. To achieve this, multiple synthetic datasets were used to represent SPSV licences, inspections, complaints, enforcement actions and monthly performance indicators. The synthetic datasets were designed to follow the structure and schema of the original real-world SPSV datasets, ensuring that the analysis reflects realistic data formats, relationships and reporting requirements, while avoiding the use of restricted or sensitive data.

The project followed a structured data analytics pipeline. First, all datasets were loaded, inspected and cleaned using Python. This included handling missing values, standardising column formats, correcting data types and ensuring consistency across datasets. Exploratory Data Analysis (EDA) was then carried out to identify trends, patterns and potential risk indicators, such as increases in complaints, inspection failures and enforcement actions over time.

Next, feature engineering was performed to create aggregated and licence-level features that could support further analysis and predictive modelling. These features were used to train a machine learning model that classifies licences into higher-risk and lower-risk categories. The predictive outputs are intended to demonstrate how machine learning could be used to support proactive regulatory oversight and prioritisation.

Finally, an interactive Power BI dashboard was developed to present the results of the analysis in a clear and accessible way. The dashboard includes an executive overview of key trends, a detailed compliance and complaints analysis, and a predictive risk section based on the machine learning outputs. Filters and visualisations allow users to explore the data by county, time period and licence type.

Overall, the project demonstrates how data analytics, visualisation and machine learning techniques can be combined to support evidence-based decision-making in a regulatory transport context. While the data used is synthetic, it was intentionally structured to mirror real SPSV datasets, and the methodology and analytical approach reflect how a real-world SPSV compliance and licensing dashboard could be implemented in practice.