Impact Evaluation

Executive Summary

- **Business Context & Problem**
  This portfolio documents a series of data science projects undertaken within Virgin Media O2 to address challenges in construction-based broadband installations, including long-running work-in-progress (WIP), limited visibility of internal failure points, manual financial tracking, and a largely spreadsheet-driven data ecosystem.
- **End-to-End Analytical Approach**
  An end-to-end approach was adopted, combining an assessment of existing data infrastructure, the engineering of clean and joined datasets, analytical modelling, and dashboard-based communication to support operational and managerial decision-making across construction and in-house installation workflows.
- **Key Analytical & Technical Outputs**
  Outputs included a logistic regression model predicting whether construction jobs would exceed 35 days using WIP features (AUC = 0.84), a Power BI dashboard analysing In-House NBR rejections by time and geography, engineered finance datasets linking purchase orders to installation costs, and a critical evaluation of current tools and data maturity.
- **Impact & Business Value**
  Collectively, this work improved visibility of operational risk in open construction jobs, reduced manual effort in monitoring NBR rejections and finance consumption, and demonstrated how predictive analytics and well-designed dashboards can support earlier intervention, better partner management, and more efficient use of internal resources.
- **Future Development & Iteration**
  The projects highlight clear opportunities for progression, including automating data ingestion, migrating from Excel-based workflows to SQL or cloud platforms, embedding predictive models into operational reporting, and improving governance as data volumes and reliance on analytics continue to grow.

Data Infrastructure and tools

The data infrastructure supporting this project reflects a moderate level of data maturity within Virgin Media O2, where analytics capabilities are established but remain heavily dependent on manual processes and spreadsheet-based workflows. Core operational and financial data is distributed across multiple internal systems, with limited support for direct querying or automated data access. As a result, data extraction and preparation often rely on scheduled downloads and file-based storage within secure Microsoft Teams environments.

Power BI is the primary analytics and reporting tool used across the organisation and was therefore selected as the end-to-end solution for visualisation and stakeholder-facing outputs. Its integration with Excel files stored in Teams enables relatively quick development

and broad accessibility for non-technical users. However, this architecture places Excel as an intermediary layer between source systems and reporting, which introduces constraints around scalability, version control, and data refresh automation.

Operational construction data is sourced via WIPMaster, a VBA-based tool that extracts work order information from the ICOMS system into Excel format. While this approach provides access to essential work-in-progress data, it limits flexibility and requires manual intervention. Similarly, finance data relating to installation costs and purchase orders is extracted from the CPA finance system, which restricts direct interaction and necessitates manual downloads. These constraints influenced tool selection, reinforcing the use of Excel and Power BI as pragmatic choices aligned with existing access controls and user capability.

Despite these limitations, the current toolset is sufficient to support descriptive analytics, dashboarding, and initial predictive modelling. However, the analysis highlights clear opportunities for improvement, including the introduction of Python for advanced analytics, automated data pipelines to reduce manual handling, and the use of SQL or cloud-based data storage to improve scalability, governance, and reliability. These enhancements would support the organisation’s transition toward more mature, data-driven decision-making.

Data Engineering

The data engineering processes for this project focused on transforming operational, administrative, and financial data from multiple internal sources into structured datasets suitable for analysis and reporting. Due to system access limitations, data ingestion was largely manual, with extracts downloaded from source systems and stored within secure Microsoft Teams locations. While not fully automated, this approach reflects current organisational constraints and ensures compliance with internal data governance requirements.

Work-in-progress (WIP) data relating to construction jobs was sourced via WIPMaster, a VBA-based tool used to extract work order information from ICOMS. This dataset included job duration, partner, region, product type, work type, and work location. To support analytical modelling, records were split into completed WIP, where job duration was known, and open WIP, where completion outcomes were unknown. Data cleaning steps included removing incomplete records, standardising categorical variables, and validating date fields used to derive job duration.

Finance data was extracted from the CPA system and required significant manual cleaning due to inconsistent formatting, missing values, and variations in how costs and purchase orders were recorded. Purchase order (PO) budget data was provided separately and integrated using partner and regional identifiers. This enabled the tracking of cost consumption against allocated budgets and supported analysis of financial exposure across construction partners and regions.

For the In-House NBR Rejections dashboard, a large administrative dataset containing form submissions was filtered to isolate in-house installation failures. Records rejected by the admin team were identified and aggregated by time, geography, and franchise. Additional reference datasets, including franchise and regional mappings, were used to enrich the data and enable geographic slicing within Power BI.

Overall, the data engineering approach prioritised accuracy, traceability, and alignment with business logic, while acknowledging the risks introduced by manual handling. The process demonstrates how structured, analysis-ready datasets can be created within constrained environments and highlights the potential efficiency gains achievable through future automation and pipeline development.

Data Visualisation & Dashboards

Data visualisation played a central role in communicating operational insights to stakeholders within Virgin Media O2, particularly for monitoring in-house installation failures under Project Maverick. A Power BI report was developed to analyse In-House NBR (No Build Required) rejection forms, which are completed when internal technicians are unable to complete an installation and the job cannot proceed as planned. Prior to this work, insight into these failures relied on manual review of administrative data, limiting visibility and timeliness.

The dashboard was designed to provide a clear, multi-dimensional view of NBR rejections over time and geography. Line charts were used to display weekly trends, enabling the identification of spikes or sustained increases in rejection volumes. A matrix visual showing days of the month against months, supported by conditional formatting, allowed for rapid detection of recurring temporal patterns. Bar charts were used to analyse rejection volumes by failure reason, supporting root-cause investigation.

Geographic and organisational context was incorporated through slicers and linked visuals based on region, regional manager, and franchise. Bubble maps and matrix tables were used to compare rejection volumes across franchises, enabling targeted operational review. Interactivity between visuals allows users to drill down from high-level trends to specific regions or franchises, improving usability and decision support.

Overall, the dashboard significantly reduced the effort required to monitor in-house NBR rejections, improved transparency, and provided a scalable reporting solution as in-house installation volumes continue to increase.

Data Analytics

The data analytics component of this portfolio focused on predicting delays in construction-based broadband installations, a key operational challenge within **Virgin Media O2**. Construction jobs that exceed 35 days in duration represent increased cost, operational risk, and potential customer dissatisfaction. To address this, a binary classification approach was adopted to estimate whether a job would complete within or exceed this threshold.

A logistic regression model was selected due to its interpretability and suitability for binary outcomes. The model was trained using completed work-in-progress (WIP) data, where final job durations were known. Predictor variables were derived from operational datasets and included construction partner, partner region, product type, type of work, and work location (road or path). These features reflect factors known to influence construction complexity and duration, enabling the model to align closely with operational understanding.

Model performance was evaluated using the Area Under the Receiver Operating Characteristic Curve (AUC), achieving a score of 0.84. This indicates strong discriminative capability and suggests that the model captures meaningful relationships within the data. The model was subsequently applied to open WIP jobs, where completion outcomes were unknown, to identify jobs at higher risk of exceeding the 35-day threshold. This enables proactive intervention, such as prioritisation, escalation, or partner engagement, rather than relying solely on retrospective reporting.

The analytical outcomes demonstrate the feasibility and value of predictive analytics within the organisation's current data constraints. By moving beyond descriptive reporting, the model provides forward-looking insight that can support more effective operational decision-making and resource allocation. While the model was not deployed into production systems, it serves as a proof of concept for embedding predictive outputs into operational dashboards or workflow tools in future iterations.

Ethical and governance considerations were assessed throughout the analysis. The dataset did not include personal customer information, reducing privacy risk and supporting compliance with data protection requirements. However, care must be taken to avoid reinforcing bias against specific partners or regions if model outputs are used without contextual understanding. Transparency in model assumptions and limitations is therefore essential. Logistic regression was chosen in part to support interpretability, allowing stakeholders to understand how predictors influence outcomes and to ensure responsible use of analytics in operational decision-making.