

## Suggested solution

### Section A

#### Question 1

To: CRO of EssanPay Co

From: An Accountant

Date: XXXX

Subject: Fraud management in EssanPay Co

#### Introduction

This report reviews the performance of the fraud prediction model during January 2026. It then evaluates the reasons raised by the head of the data science function for the model's underperformance. The report concludes with recommendations to mitigate the governance issues for future fraud predictive models.

#### (i)

The head of the data science function has produced the following confusion matrix to which has been added the identification of the true/false negatives/positives:

	<b>Predicted Legitimate</b>	<b>Predicted Fraudulent</b>
<b>Actual Legitimate</b>	1,216 (True Negatives)	131 (False Positives)
<b>Actual Fraudulent</b>	23 (False Negatives)	7 (True Positives)

From the confusion matrix these are the transaction outcomes from January 2026

- Total Transactions Processed: 1,377
- Confirmed Fraudulent Transactions: 30
- Fraud Prevalence: 2.2%

The model scores high in accuracy, but accuracy is simply answering what percentage of all predictions were correct. Fraud is rare, so a model which identifies zero fraud would still have high accuracy. A better evaluation metric would be recall, because recall measures how many actual fraud cases were detected by the model. Precision is also important, as it answers how many of the transactions flagged as potentially fraudulent

were actually fraudulent. Based on this, the best evaluation metric would combine recall and precision, F1 Score.

The following performance metrics can help to evaluate the performance of the model:

### **Accuracy**

$(\text{True Positives} + \text{True Negatives}) / \text{Total}$

$= (7 + 1,216) / 1,377$

**= 88.8%**

While superficially high, this metric is misleading. Given the 2.2% fraud rate, a naïve model classifying all transactions as legitimate would achieve accuracy in excess of 97%.

### **Precision**

$\text{True Positives} / (\text{True Positives} + \text{False Positives})$

$= 7 / (7 + 131)$

**= 5.1%**

Over 95% of flagged transactions were legitimate, resulting in unnecessary customer disruption and significant operational burden.

### **Recall (Sensitivity)**

$\text{True Positives} / (\text{True Positives} + \text{False Negatives})$

$= 7 / 30$

**= 23.3%**

Fewer than one in four fraudulent transactions were detected, directly explaining the elevated fraud losses, chargebacks, and regulatory exposure observed during the quarter.

### **F1 Score**

$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$

$= 2 \times (5.1\% \times 23.3\%) / (5.1\% + 23.3\%)$

**= 8.3%**

F1 Score is the harmonic mean of Precision and Recall. The F1 score of approximately 8% confirms very weak overall performance on the fraud class, reflecting both low recall (most frauds missed) and extremely low precision (most alerts incorrect).

### **False Positives (incorrect fraud flags)**

Approximately 9.7% of legitimate transactions were incorrectly flagged as fraudulent (131 transactions flagged out of 1,347 legitimate transactions). This explains the observed increase in customer disruption, unnecessary transaction declines, and operational burden within the fraud investigation team.

Tutorial note: other valid metrics e.g. balanced accuracy instead of F1 score would be awarded credit where appropriate.

## **(ii) Potential reasons for underperformance relative to expectations**

### **Class imbalance**

Class imbalance occurs when one class in a classification problem significantly outweighs the other class. A model trained on an imbalanced dataset perform poorly on the minority class.

Fraud represents approximately 2% of the transactions in January 2026, creating a rare-event classification problem. The ‘Accuracy Paradox’ means that overall accuracy is dominated by legitimate transactions. An accuracy rate of c.98% could be achieved simply by predicting 2% of transactions as fraudulent, even if none of the predicted cases were fraudulent. Accuracy therefore provides little insight into the effectiveness of fraud detection.

The reported training accuracy of 97.8% is therefore misleading. A random classifier predicting all transactions as legitimate would be expected to achieve comparable accuracy. Live performance falling below this benchmark indicates that performance expectations were not calibrated using appropriate metrics (such as F1 score, precision and recall) during development.

Naïve Bayes does not inherently address class imbalance. Without explicit class weighting, threshold optimisation, or cost-sensitive learning, predictions naturally favour the majority class, contributing to low recall and thus poor predictive power.

If majority-class under-sampling were applied to the training data to address the class imbalance, important predictive information may have been lost from the training data.

### **Data quality issues**

A misspelling of `merchant_riskscore` in the live production data relative to `merchant_risk_score` in the training data may have prevented this feature from being recognised by the predictive model. Consequently, the deployed model may be operating with fewer predictors than the trained model, reducing its predictive power.

There are missing values in the live data. While Naïve Bayes can theoretically omit missing predictors under its independence assumption, it is possible that not imputing the missing data could have led to poor model performance in production.

If missing data existed in the training data but was not appropriately analysed or imputed, conditional probability estimates may be biased, leading to poor live performance. Conversely, even if imputation of missing training data was appropriate, missing data in

the production environment may have been handled inconsistently or may not have been handled at all, altering the effective feature distributions in the live production data.

The training data and live data include categorical variables. Inconsistent categorical encoding between the training and production environments may have distorted probability estimates, contributing to poor live performance.

### **Overfitting**

Overfitting occurs when a model memorises the training set so closely that the model fails to make correct predictions on new data.

The only metric reported in respect of the model's training performance was accuracy. There is no evidence of independent test evaluation or time-based validation. The model may therefore have learned historical patterns which failed to generalise to the unseen live data.

If class imbalance in the data were addressed in training, the methods used to address imbalance may themselves introduce overfitting, inflating the apparent performance of the model in training. This is because oversampling prior to data splitting can cause leakage of training data into the testing data, duplication of fraud cases in the training data can lead to pattern memorisation rather than generalisable learning and synthetic oversampling of training data can generate unrealistic feature combinations.

Naïve Bayes assumes conditional independence among predictors. However, fraud risk often depends on interactions and non-linear relationships. The model may not have captured complex behavioural patterns effectively. A tree-based model may have been performed better as they naturally capture interactions and non-linear effects.

The model was trained on 5,000 transactions, representing six months of data. The volume of training data will not have been sufficient to learn stable fraud patterns. In reality it should have been trained on at least two years of data. With only a limited number of fraud observations, parameter estimates are likely to be highly sensitive to noise within the training sample. This increases variance and raises the risk that the model has captured noise in the training data rather than generalisable relationships, contributing to overfitting and poor live performance.

### **Model drift**

Model drift is the decline in a machine learning model's predictive performance over time. It may have occurred due to feature drift, concept drift or a combination of both.

Feature drift occurs when the distribution of input variables changes over time, meaning the model is applied to data which differ from those used during training.

The six-month training period is not specified and may therefore represent behaviour which is no longer reflective of current transaction patterns. If the training data related to an earlier period, changes in customer behaviour, fraud strategies or operational conditions may mean the relationships learned in training were outdated by January 2026. Even a well-trained model may therefore perform poorly when applied to out-of-date data.

The January 2026 dataset includes an additional product category field which was not included in the training feature set, indicating that the operational data structure and business activity have evolved since model development. The introduction of new products or transaction types changes the feature space and may alter the distribution of related predictors such as transaction value, merchant behaviour, or customer purchasing patterns.

The number of transactions in January 2026 was higher than the average monthly transactions in the training data. Even where predictors remain unchanged, rapid growth in transaction volumes and increasing digital usage may have shifted the distribution of the live data relative to the six-month training data.

If January behaviour was not represented in training data, seasonal effects such as post-holiday spending patterns or changes in fraud targeting may have changed the distribution of the data further.

Concept drift occurs when the relationship between predictors and fraud risk changes, meaning patterns previously associated with fraud no longer have the same predictive meaning.

Changes in authentication methods, customer behaviour or fraudster strategies may alter which transaction patterns indicate fraud. Behavioural signals which were previously associated with fraudulent activity may have become common among legitimate users, while new fraud patterns may emerge which were not present during training. As a result, conditional probability estimates learned by the model become outdated, directly reducing discriminatory power.

Model drift may also have arisen if the overall fraud prevalence has changed over time. Naïve Bayes explicitly incorporates class priors into posterior probability calculations. Failure to update these assumptions can bias predictions even where feature distributions remain stable.

### **(iii) Recommendations of actions to improve governance of future fraud prediction models**

As can be seen from the case of FernLink Co, there can be significant commercial consequences arising from issues of model governance, especially in relation to its

development, deployment and monitoring. Strengthening model governance would help ensure fraud detection systems remain reliable, transparent and responsive to evolving fraud risks. The following improvements are recommended.

### **Introduce formal model validation before deployment**

The current model appears to have been deployed primarily based on a 97.8% training accuracy, which is misleading given the very low fraud prevalence. Future models should undergo independent validation before model training. This can be achieved by splitting the historical data to train, validate and test subsets before training begins. This ensures the model can be validated on unseen data and avoids model overfitting. Validation should specifically assess whether the model can identify fraudulent transactions rather than simply classifying legitimate transactions correctly, using appropriate evaluation metrics, such as precision, recall, and F1 score.

### **Benchmark alternative algorithms**

The current model uses a Naïve Bayes algorithm, which assumes independence between predictors. Fraud detection often depends on interactions between variables such as transaction amount, merchant type and customer behaviour. Governance procedures should therefore require systematic benchmarking of alternative models such as Random forests, Gradient boosting models, Logistic regression with interaction terms. Tree-based models in particular may capture complex behavioural patterns which Naïve Bayes cannot detect, potentially improving fraud detection performance.

### **Strengthen feature engineering and feature governance**

The current model does not capture important behavioural indicators of fraud. Governance processes should require structured feature engineering and documentation to ensure relevant predictive signals are included. For example, additional features could include deviation of transaction value from a customer's historical average, number of transactions within a short time window, or merchant risk indicators or geographic anomalies. Improved feature design would help future models identify suspicious transaction patterns more effectively.

### **Address class imbalance through controlled training procedures**

Fraud represents approximately 2% of transactions, creating a severe class imbalance problem. Without corrective measures, models tend to favour the majority class. Governance frameworks should require explicit handling of class imbalance during model development, such as oversampling fraud cases, undersampling legitimate transactions, or using a model algorithm which penalises missed fraud more heavily. These techniques would help ensure that a future model focuses on detecting fraud rather than maximising overall accuracy.

### **Implement continuous performance monitoring**

Model performance should be monitored continuously using operational metrics such as recall, precision, or F1 Score. The current results show that 76% of fraud cases were

missed and over 95% of alerts were false, indicating serious performance deterioration. Governance procedures should require regular performance reporting to the CRO and trigger investigation when performance falls below predefined thresholds.

### **Introduce model drift monitoring and regular retraining**

The model was trained using approximately 5,000 historical transactions over six months, which may not reflect current transaction patterns. Governance controls should therefore include monitoring for changes in transaction behaviour or fraud patterns. For example, scheduled retraining of the model using more recent transaction data, and periodic recalibration of model parameters and fraud thresholds. This would help ensure future models remain aligned with evolving fraud strategies.

### **Implement data governance and data pipeline controls**

Errors in production data can significantly degrade model performance. For example, inconsistent variable names or missing features could prevent the model from using key predictors. Automated data validation checks should therefore be introduced to detect missing or incomplete fields, changes in feature definitions, and inconsistencies between training and production datasets. These controls would reduce the risk that the model operates on incomplete or inconsistent information.

### **Documentation**

Comprehensive documentation should be maintained for the fraud detection model, covering the training data used, feature definitions, preprocessing steps, model assumptions, performance metrics and limitations. In this case, reliance on a single training accuracy metric and potential inconsistencies between training and production data highlight weaknesses in transparency and traceability. Proper documentation would allow risk managers and internal auditors to understand how the model operates, verify that appropriate validation has been performed, and identify whether changes in data, features or assumptions could explain any underperformance.

### **Governance structures**

EssanPay Co should establish clear governance structures for the oversight of predictive models used in fraud detection. This could include a formal model risk management framework with defined responsibilities for model development, independent validation and ongoing monitoring. For example, the data science function would remain responsible for model development, while the risk management team could provide independent challenge and approval before deployment. Regular governance reviews would ensure that model performance, fraud losses and operational impacts are monitored at senior management level, allowing the organisation to respond quickly if fraud detection performance deteriorates.

## Section B

### Question 2

#### (a)

The SQL data types proposed by the junior analyst are poorly aligned with Watlee Co's operational and financial reporting requirements, particularly given the need to allocate overheads accurately and aggregate results across products, plants, and months.

A major weakness is the use of FLOAT for financial values, including material costs, labour costs and overhead costs. Floating-point types store approximate values and can introduce rounding errors, which may compound when unit costs are multiplied by production volumes and aggregated across multiple dimensions. In a costing and overhead allocation context, such inaccuracies undermine the reliability of reported results. For example, an overhead rate calculated using FLOAT may store slight approximations, leading to inconsistent allocated costs and reconciliation issues when results are aggregated.” Exact numeric types such as DECIMAL or NUMERIC with appropriate precision and scale are therefore more suitable.

The use of FLOAT for units produced is also inappropriate. Production volumes represent discrete count data, yet floating-point storage allows fractional values which have no operational meaning. This creates a risk of unintended non-integer values emerging during calculations or data transformations, particularly when aggregating production quantities across plants or periods. Such inconsistencies may weaken confidence in operational metrics and complicate reconciliation between production and financial systems. Integer data types such as INT or BIGINT should instead be used, as they enforce whole-number values, improve data validation, and more accurately reflect the underlying business reality of unit-based production.

Storing the reporting period as VARCHAR(7) presents further issues. Character data types do not enforce valid date values, meaning incorrectly formatted or invalid months could be stored without error. This reduces data quality controls and limits the effectiveness of time-based filtering, sorting, and indexing operations. Queries requiring chronological ordering or period comparisons may become more complex and computationally inefficient, particularly as data volumes grow. Given that Watlee's reporting operates on a monthly cycle, a DATE data type (for example, storing the first day of each reporting month) would be more appropriate. This enables reliable temporal validation, improves query performance, and supports consistent time-series aggregation and analysis.

The use of unbounded character types such as TEXT for identifiers (e.g. product SKU) is also inefficient. While flexible, unbounded text fields consume more storage and reduce indexing efficiency, which can slow joins across reporting tables and downstream

reporting datasets. In an environment where data from multiple operational systems must be integrated regularly, inefficient joins may materially affect reporting performance and scalability. Furthermore, unrestricted text fields provide limited control over data quality, increasing the risk of inconsistent identifier formats. Using bounded character types such as VARCHAR(n) aligned with expected identifier length improves storage efficiency, strengthens validation, and enhances join performance while still accommodating alphanumeric product codes.

Overall, the proposed data type choices increase the risk of inaccurate cost calculations, inefficient joins, reduced data quality and unreliable aggregation in Watlee Co's management reporting. Adopting appropriate numeric, temporal, and identifier data types would significantly improve accuracy, performance, and scalability of the reporting framework.

## **(b)**

### **CFO – ability to quickly assess performance**

The CFO stated that the dashboard should quickly indicate whether financial performance is improving or deteriorating.

The dashboard meets this need in some respects. Trend charts and consistent monthly layouts enable rapid identification of performance movements. Adverse variances highlighted in red effectively draw attention to underperformance. Logical grouping of revenue and cost information supports intuitive navigation for users. Interactive year selection improves usability during executive review.

However, it does have some limitations. The title “Executive Summary” is generic and does not clearly communicate the dashboard's purpose or decision focus. There are also some financial presentation inconsistencies which reduce clarity, for example budget labelled ‘Budget’ in tables but ‘BudMonth’ in charts, the Revenue and Cost values include dollar symbols while Diff does not and the repetition of dollar symbols in each row creates unnecessary visual clutter.

Overall performance cannot be assessed immediately because profitability indicators are absent; revenue and cost information are presented separately rather than as financial outcomes.

### **Head of costing – level and clarity of costing information and drivers of change**

The head of costing noted that performance is normally analysed by cost type. As an Executive Summary, aggregated cost information is appropriate and avoids overwhelming executives with operational detail. The dashboard provides no indication that more detailed costing analysis is available, limiting investigation of cost drivers.

The head of costing also highlighted the need to understand why performance changes occur.

The dashboard does show the Revenue Trend by Plant chart which enables comparison of plant performance patterns, however the revenue trends lack cost or margin context, limiting diagnostic insight. In addition, the Unit cost visuals highlight operational cost differences between plants but one plant appears missing from the trend visual, reducing comparability across operations.

### **Commercial manager – completeness of financial picture**

The commercial manager questioned whether the dashboard presents a complete financial picture.

Both revenue and cost information are included in the dashboard, supporting high-level monitoring and trends over time are clearly visible.

However, the dashboard focuses on activity measures rather than outcomes, making overall performance difficult to judge. For example, revenue trends are presented without profitability or contribution context and the Unit Cost by Product visual provides limited decision usefulness. Products are displayed in descending cost order but without revenue, margin or budget context, so management cannot determine whether higher costs represent inefficiency or product mix differences. Watlee Co manufactures six products, yet only four are shown, and it is unclear whether these represent the highest-cost products or a selected subset. This ambiguity supports the concern that the dashboard does not present a complete financial picture.

### **Data analyst – accessibility and colour usage**

The data analyst raised accessibility considerations.

The dashboard does use consistent colours to represent plants across multiple visuals, aiding recognition. It has a clean layout and consistent font to improve readability.

The colour meaning is, however, inconsistent as orange represents both adverse variances and Plant D and blue represents both the actual revenue and cost of sales and Plant A. Using colours simultaneously for categorisation and performance signalling may confuse users. The costing department should check that this does not create accessibility risks for colour-vision deficiencies.

White space has been used to separate the content effectively, but the charts do not strongly emphasise key performance signals, which limits visual hierarchy.

## **Operations manager – risk of misleading visuals and lack of commentary**

The operations manager expressed concern that charts may lead to incorrect conclusions.

A positive about the dashboard is that several visuals appropriately use zero-based axes, supporting proportional comparisons. There are, however, some issues. The Unit Cost by Plant chart uses a y-axis which does not start at zero, exaggerating differences between plants. The Unit Cost by Plant chart includes unnecessary bar labels, redundant y-axis labelling and gridlines which add visual noise without improving insight. Revenue and Cost charts positioned side-by-side use different axis scales, limiting meaningful comparison. The Revenue Trend by Plant chart shows December twice, with different values, creating ambiguity and reducing confidence in dashboard reliability.

The operations manager also noted the absence of explanatory commentary.

The visuals allow trends to be observed quickly without extensive explanation, however without highlighted insights on key areas, managers may interpret results differently, reducing consistency in decision-making.

### Question 3

(a)

#### Stage 4: Model Evaluation

The evaluation of the model is suggested to be performed by the model developer without peer review, however there should be structured validation and sign-off involving key areas such as finance and the risk management department.

Accuracy metrics are most commonly used as can be used for comparison but reliance on accuracy alone may not reflect business usefulness for identifying emerging risks. There should be metrics appropriate to model purpose and aligned with decision usefulness rather than accuracy alone. The metrics should be based on the data used and the business problem being addressed. This ensures the model provides actionable insight rather than merely accurate predictions.

Comparing predicted results with historic financial results is an essential element of model development so this suggestion by the junior data scientist is correct. This would allow MoyWorth Co to ascertain that the model is predicting values which are similar to the historical data.

Adjustment of the model during evaluation blurs the boundary between model tuning (Stage 3) and evaluation. The adjustment should occur at Stage 3 Model Engineering which is likely to have been addressed by the senior data scientist in the part of the plan drafted by them.

The proposed approach to Stage 4 (Model Evaluation) is limited and focuses primarily on prediction accuracy and whether results “appear reasonable.” For a predictive decision-support model, evaluation should confirm that outputs provide reliable early-warning insight for users rather than simply achieving statistical performance.

The junior data scientist has not included evaluation success criteria. There should be predefined acceptance criteria linked to early detection of adverse trends.

#### Stage 5 — Model Deployment

The junior data scientist’s plan assumes dashboards can be published and immediately adopted by finance teams but there is no indication of how this will happen. There should be clarity on who will own the model and its outputs and whether the teams external to the data analysis team will have access to the model itself or just the dashboard. It should also be clear about where the model is housed e.g. is it an internal server, the cloud etc and what the implications of that is. Similarly, what are the data pipelines which link the model to the dashboard.

Scheduling automatic data refreshes should be expected, however the plan does not explicitly state when and how often these will be performed and by whom. Similarly, running the model prior to reporting deadlines to generate updated outputs needs more clarity in terms of who will undertake this task and the timeframes for it.

The junior data scientist refers to finance team incorporating insights into reporting processes; however, this is the explicit purpose of the dashboard and its self-explanatory. The plan does not address how insights will be embedded into management monitoring or decision-making processes. There should also be a description of how model outputs should be interpreted and how it be integrated into management monitoring workflows in advance of each reporting cycle.

### **Stage 6 – Monitoring and Maintenance**

The monitoring arrangements in the junior data scientist’s plan are weak and largely informal. Periodic review and informal user feedback are not sufficient for managing the risks associated with a predictive model as predictive models can degrade over time due to data changes and/or evolving business conditions.

The key weaknesses of the plan include undefined monitoring criteria, absence of performance thresholds, escalation procedure and structured review responsibilities. Monitoring is therefore considered to be a reactive rather than proactive activity in the junior data scientist’s plan.

Recommended improvements would be defining model performance and data-quality metrics, scheduling review cycles aligned with reporting periods and monitoring of model drift and data changes. This would allow a structured approach to maintaining the model rather than an ad hoc approach.

In addition, there should be a formal escalation and remediation process where users identify issues with the model. This may not always be about problems with the model but additional information which is deemed useful and can be included in the dashboard.

### **Other points**

It is not clear as the Stages 1-3 of the plan are not available whether the senior data scientist raised the matters of documentation and model governance, however both are important for user training and audit purposes.

A more rigorous investigation of the exact information being maintained by each department is likely to reveal more data in common and it is also very likely that several departments maintaining the same data may reduce data integrity, as there might be confusion about which data set (among many versions) is reliable and up-to-date.

**(b)**

### **Finance Leadership Team (FLT)**

The FLT are primary users of model insights and should therefore be closely engaged throughout the project.

During the early stages of the project, the project manager should consult finance teams to clarify business objectives and define what constitutes useful and actionable insight.

Finance teams should also be engaged to validate the data used to train the models. Finance can provide context on accounting definitions, adjustments, and unusual historical events which may affect data interpretation. This helps ensure the model learns meaningful business patterns.

Regular demonstrations should be conducted for FLT during development to ensure model outputs are understandable and aligned with management decision-making needs.

During implementation, FLT should receive clear communication and training explaining how outputs should be interpreted and their limitations. Expectations regarding how insights support, rather than replace, existing reporting processes should be agreed.

Following implementation, structured feedback mechanisms should be established so finance users can report concerns and highlight changes in business conditions affecting model usefulness.

### **Risk Management Department (RMD)**

RMD should be engaged as an independent governance stakeholder throughout the lifecycle to ensure that innovation occurs within the organisation's risk framework.

The project manager should engage RMD early to align the project with the organisation's machine learning model risk management framework, including expectations relating to documentation, transparency and appropriate use.

During development and prior to implementation, RMD should provide structured challenge of assumptions, review evaluation approaches and confirm that accountability and control arrangements are defined.

Prior to operational use, the project manager should obtain confirmation that appropriate controls and accountability arrangements are in place. After implementation, RMD should receive regular monitoring information and participate in periodic reviews to ensure continued appropriate use.

### **IT and Data Engineering Teams**

The IT and Data Engineering Teams should be managed as operational partners responsible for delivery and sustainability.

Early engagement should confirm technical feasibility, data pipeline design and system integration requirements.

During implementation, close coordination should ensure deployment timelines, access controls and support responsibilities are clearly defined.

Following deployment, agreed monitoring and maintenance processes should ensure system stability, data quality and timely resolution of operational issues.

## Mark scheme

### Question 1

#### (i)

Summary of figures from the confusion matrix – up to 2 marks

Justification as to why the metrics were chosen - up to 2 marks

Evaluation of the model using the metrics:

Accuracy paradox calculation and interpretation – up to 2 marks

Precision calculation and interpretation - up to 2 marks

Recall (sensitivity) calculation and interpretation - up to 2 marks

F1 score calculation and interpretation - up to 2 marks

Interpretation of false positives and their implications for the business - up to 2 marks

#### Maximum 12 marks

#### (ii)

Class imbalance – 1 mark per relevant point, up to 5 marks

Data quality issues – 1 mark per relevant point, up to 5 marks

Overfitting – 1 mark per relevant point, up to 5 marks

Model drift – 1 mark per relevant point, up to 5 marks

#### Maximum 18 marks

#### (iii)

Governance improvement recommendations – up to 2 marks each

#### Maximum 10 marks

**Professional skills marks:****Communication:**

General report format and structure (use of headings, sub-headings and an introduction)

Style, language and clarity (tone of report response, presentation of calculations, appropriate use of the tools, easy to follow and substantive amount of content)

**Analysis and evaluation**

Model performance is logically analysed with relevant metrics and clearly articulates the results of the metrics chosen

**Scepticism**

Reasons suggested by the head of the analytics department are probed and possible issues are flagged and appropriately challenged

**Commercial acumen**

Governance improvement recommendations are practical and plausible for EP Co taking into consideration their commercial context

**Maximum 10 marks****Total 50 marks**

## Question 2

### (a)

Each weakness and recommendation – up to 3 marks

Relevant areas for evaluation include:

- Use of FLOAT for financial data (material, labour, overhead costs)
- Use of FLOAT for discrete quantities (units produced)
- Use of character data types for time periods (VARCHAR(7))
- Use of unbounded character types (TEXT) for identifiers
- Inconsistent or suboptimal data type choices for identifiers (e.g. plant IDs as VARCHAR)

**Maximum 10 marks**

### (b)

1 mark per relevant point, up to 10 marks

Relevant areas for evaluation include:

- Ability to quickly assess performance
- Lack of detail and clarity and drivers of change
- Completeness of financial picture
- Accessibility and colour usage
- Risk of misleading visuals and lack of commentary

**Maximum 10 marks**

**Professional skills marks:**

### Analysis and Evaluation

Logical and structured analysis is undertaken of the SQL data types which clearly evaluates the proposed tables and offers valid improvements

### Commercial Acumen

Evaluation of the dashboard clearly raises practical and commercial considerations for its use in respect of the management team's comments

**Maximum 5 marks**

**Total 25 marks**

**Question 3****(a)**

Evaluation of each stage, including recommendations – up to 5 marks

Other points – up to 2 marks

**Maximum 12 marks****(b)**

Advice on the management of each identified stakeholder – up to 3 marks each

**Maximum 8 marks****Professional skills marks:****Analysis and evaluation**

Systematic review of the three stages of the project plan is undertaken which clearly analyses the plan in relation to the CRISP-ML framework

**Scepticism**

Suggestions made by the junior data scientist are appropriately challenged and clearly supported or rejected

**Commercial acumen**

Advice on stakeholder management offers practical solutions on how they should be managed, clearly recognising their different roles and responsibilities

**Maximum 5 marks****Total 25 marks**