This project focuses on building and evaluating a Logistic Regression model in BigQuery ML to predict flight diversions using historical flight performance data. The goal is to develop a deployable predictive tool that can forecast rare but costly operational disruptions directly within BigQuery. The workflow began with importing the Carrier On-Time Performance dataset from Kaggle into Colab, authenticating with Google Cloud, and setting up the BigQuery environment using the google.cloud library. The dataset was configured in everybody's personal projects to ensure consistency across the workspace. Screenshots from the development process show the successful connection and initialization of the dataset, confirming that the environment was properly prepared for model training and evaluation.

The baseline logistic regression model used key flight parameters such as departure delay, carrier, and origin-destination airports. The results reflected the difficulty of modeling rare events. Although the model achieved a high accuracy of 0.998115, both precision and recall were 0.0, indicating that it failed to identify any diverted flights. This outcome was expected due to the highly imbalanced data, where diversions represent a very small proportion of total flights. The ROC AUC of 0.820762 suggested that the model captured some level of class distinction, but this was not sufficient for reliable classification. The low log loss value of 0.013002 showed the model was confident in its predictions, but it was consistently predicting the majority class.

A second, improved model was successfully developed using AUTO_CLASS_WEIGHTS=TRUE to implement cost-sensitive learning, which corrected the baseline model's failure to predict diversions. This "engineered" model achieved a significantly better ROC AUC of 0.837579, and, crucially, a Recall of 0.947 demonstrating a high success rate in identifying true diversion events, which is the key operational goal.

Future work will involve threshold tuning and operational calibration to determine an acceptable false negative rate for diversion prediction. In practical terms, the business must balance the cost of false positives, which could lead to unnecessary resource allocation, against the cost of false negatives, which represent missed diversions and potential disruptions. The separate Regression model for predicting arrival delay demonstrated an MAE of approximately 9 minutes and an R^2 of 0.4972. This MAE is the average error that must be absorbed by the operations schedule for staffing and gate assignments, signaling room for improvement in prediction of accuracy. Overall, this project highlights the process of integrating data management, feature engineering, and evaluation within BigQuery to improve operational decision-making in flight management.

Our classification model predicts whether a flight will be Diverted or not, which is an asymmetric risk problem as the costs of the two types of errors are completely different. In the matrix, there are false negatives and false positives, and both of these errors can incur unwanted costs to the airline company. False positives are essentially false alarms. These are manageable costs, such as alerted staff, alternate gates prepared, and other preemptive tasks. However, the false negative, or missed diversion, is completely catastrophic. Under the assumption that this situation has no contingency plan, the company will waste resources at the original destination, face regulatory fines, and, most impactfully, massive compensatory costs for passengers. With the extreme cost difference, the model should be tuned to minimize the false negative rate, which comes with prioritizing recall over precision. The final model is the weighted model, which uses "AUTO_CLASS_WEIGHTS=TRUE". This tells the model that the cost of misclassifying a rare positive is much higher than misclassifying a negative. This method resulted in a high recall of 0.947 as well as an ROC AUC of 0.838. Deploying a model tuned this way almost eliminated the risk of false negatives, making the model safe and effective for real time decision making.

The most significant performance gain came not from the use of traditional feature engineering but engineering AUTO_CLASS_WEIGHTS. This form of data engineering created a cost-sensitive learning algorithm. The model quantified risk mathematically by embedding the high cost of a false negative into the model's objective. Forcing the model to prioritize maximizing recall (about 95%) ensures almost no missed diversions. Future ideas for predictive improvement can include three parts: external data integration, advanced modeling, and dedicated infrastructure. For external data integration, the model will incorporate real-time weather information at the destination and origin. This could come from Air Traffic Control, National Weather Service, and other governmental entities that provide valuable information sources. Our model could also use more advanced modeling algorithms such as XGBoost, as those models are better equipped to handle the sparsity and extreme class imbalance. Our baseline model performs decently, but having an algorithm that captures even more patterns in the data will be infinitely more useful to the airline company. Last, the model can have a dedicated infrastructure. The airline can fully deploy the model to a hosting service (i.e. Vertex AI) which allows for real-time predictions and establishment of automated, continuous pipelines. Flight patterns change continuously and suddenly; hence it is important for a pipeline to update as soon as the data can be updated.