



Lloyds Banking Group

Data Analyst Incubation

**Sprint 3 : Time Series Forecast,
Fraud flagging, deployed to GCP**





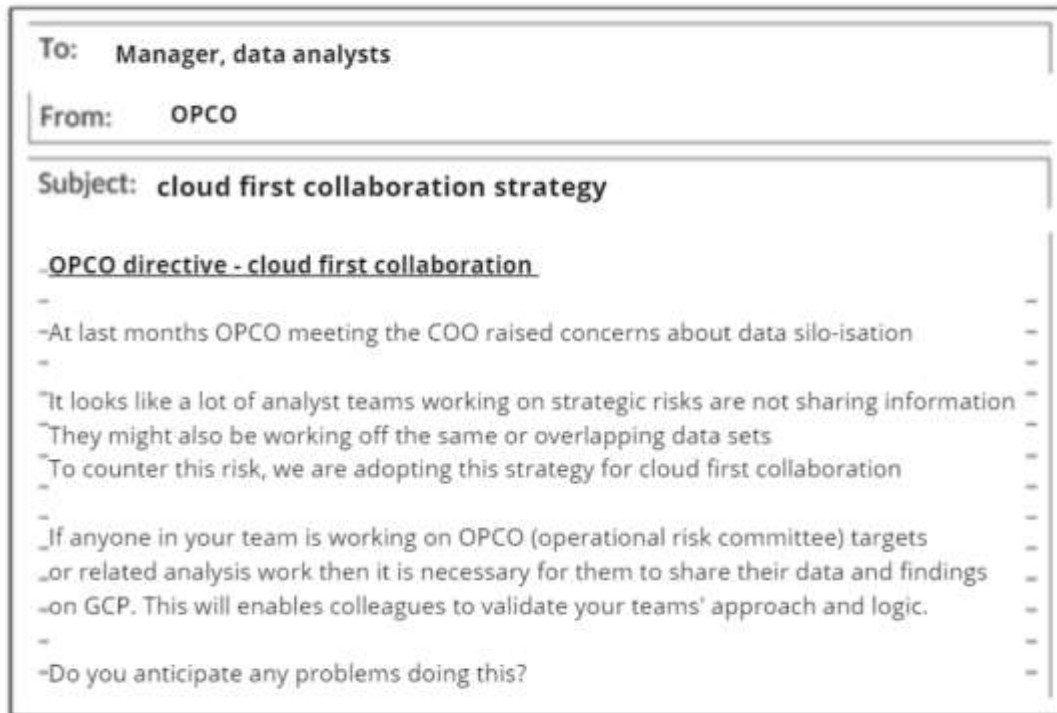
Contents

Business Requirements	3
Assignment detail (deliverables).....	4
Data and local time series forecast model deployed to GCP.....	4
Output of time series forecast produced and saved in GCP.....	4
Iteration of time series forecast model inside GCP.....	5
Further assignment (if you have time)	5
Migrate data and model for transaction fraud analysis.....	5



Business Requirements

In the middle of your work with the Markets and Finance team, you receive an email from your team lead which requires you to re-consider the approach you are taking.



It appears that the Bank is adopting what they call 'cloud-first collaboration' as an approach to data and analysis sharing. This makes it your responsibility for ensuring that all data subsets and extracts used in strategic analyses are made available for colleagues to access via the Google Cloud Platform.

Your team lead has recommended that you first migrate the historical transaction data you used to support your time series forecast model to GCP. In addition, you should promote your python descriptive analysis and time series forecast model to the cloud so that this too can be reviewed by relevant colleagues before any of your insights are communicated to the Markets and Finance team.

If you have enough time, your team lead indicated you should do the same with the fraud analysis model you have also worked on – get the data and model up on GCP.



Assignment detail (deliverables)

Data and local time series forecast model deployed to GCP

1. Data set used for time series analysis (extracted and aggregation from SQL database) migrated to appropriate storage service (Big Query, Cloud SQL or Cloud Storage) on Google Cloud Platform. Consideration should be given to data security in the Cloud at the point of migration. If the raw data set is too large for GCP, consider migrating an aggregated version of the transactional data (by appropriate time period).
2. Guide for colleagues to access the data set you have migrated to GCP including providing ambiguous column definitions, documenting filters that you have applied to select the data from the original source and details of any calculations or aggregation logic that are built into the selected data set.
3. Time series forecast model connected to the above data set, hosted on GCP using appropriate services. Consideration should be given to the capabilities of the GCP ML services :
 - BQML natively supports time series model creation (Arima Plus)
 - Vertex AI supports the creation of Arima and Prophet models
 - Vertex AI supports the import of existing models for improvement

Output of time series forecast produced and saved in GCP

1. Forecast data points (predictions into the future) created by the initial model produced and saved as a separate data source in GCP (using the appropriate storage service) to enable future validation by colleagues



2. Exploratory Data Analysis performed on the data produced by the first forecast model, inside the Google Cloud Platform. This can be managed in BigQuery, Colab or Workbench.

Iteration of time series forecast model inside GCP

1. Proof and documentation showing that the time series model can be updated or retrained as appropriate, and the forecast recalculated, when new transactional data becomes available (this can be automated, or manually iterated at least once).
2. Forecast data points (predictions into the future) created by the second version of the model produced and saved as a separate data source in GCP (using the appropriate storage service) to enable future validation by colleagues. This should be clearly labelled in GCP to differentiate it from the forecast data produced by the initial model.
3. Exploratory Data Analysis performed on the data used for the second forecast model and the forecast data points created by the second forecast model, inside the Google Cloud Platform. This can be delivered in BigQuery, Colab or Workbench.

Further assignment (if you have time)

Migrate data and model for transaction fraud analysis

1. Following on from the memo in the business requirements section, if you have time during this Sprint, you should replicate the approach you have taken to the time series forecast model and migrate the fraudulent transactions analysis & model you worked on earlier in the sprint. Applying the same process to two different scenarios will be a good test of the technology and ease of use, enabling you to offer practical feedback to the Bank on their strategy of 'cloud first collaboration'.

