Analysis on Formula1 Race

→ We collected raw data from the API and stored it in containers on the Azure.

→ We have joined the key information required for reporting to create a new table.

→ We processed this data using Databricks notebook to ingest into ingested raw layer.

→ We have ingested data which has requirements like - the data should be in columnar format and the schema must have been applied to the ingested data and the status must be analyzed via SQL.

→ We have created a project structure and was able to categorize different kinds of data in the project.

→ We have setup notebooks to mount ADLS storages in Databricks.

→ We need to transform and analyze the data.

→ We must use Azure data factory to schedule and monitor the data with different requirements like scheduling should be done on a particular day of a week and ability to re-run the failed pipelines and ability to setup alerts on failures.

→ We are trying to add a feature like – the ability to roll back to a previous version.

→ We are yet to finish the programming part of the project.

→ We still need to work on the visualization of the results.