Rockwall Analytics
DATA & ANALYTICS ADVISORY

To: Aviation

From: Rockwall Analytics

Subject: Data Analytics

Date: December 19, 2020

Combine and visualize data to drive meaningful insights using enterprise-grade technology to assist Aviation's team make decisions based off real time data

Data Engineering

Using data engineering to combine full-flight engine data, part manufacturing data, airport location data to determine the distance travelled for each airplane

Background Story

In this business case Rockwall Analytics will combine and visualize data to drive meaningful insights using enterprise-grade technology. We are provided with work samples for both data engineering and data visualization to assist the Aviation team to make decisions based off real time data

Aviation rises to the challenge to invent the future of flight, lift people up, and bring them home safely

Abstract

At Aviation, we use the concept of a Data Lake. A Data Lake is a single database instance that contains data from all around Aviation. Everything from financial, delivery (of parts and engines), supplier, engine data, customer data, and so on

The advantage of using a *Data Lake* allows the developer to make a single connection string to the *Data Lake*, and as long as the developer has permissions to see the data – *they are able to immediately start creating data-driven insights in a centralized repository of data*

To simplify this experience, Aviation has put all tables into a single schema (database inside the data lake)

There are 8 datasets required to accomplish this task

```
The 'Y': Remaining Useful Life (RUL) calculations for ESN's.  This is the 'Y'
that we want to predict so that we can prevent unscheduled maintenance and/or
identify manufacturing issues that are causing RUL to decrease more rapidly than
expected.

Flight Data (4 datasets): Engine measurement data from 4 airline operators – Air
India (AIC), Frontier (FRON), Pegasus (PGS), and AirAsia (AXM).  This data that
can be used to determine the health of the engine.  Each airline has provided to
GE their flight data in 4 separate tables.

Location Data: Airport ICAO codes and their respective locations (latitude and
longitude).  This could be valuable in understanding where the airline operators
are flying when their data is collected.

Manufacturing Data: Manufacturing data for 3 (completely imaginary) parts:
44321P02 (blade), 54321P01 (shroud), and 65421P11 (disk).  The data contains key
characteristic measurements (and their max and mins) for these parts at certain
operations within their production.  Every part contains a unique serial number.

MBOM: Manufacturing Bill of Material (BOM) that matches every serial numbered
part to the respective engine via engine serial number (ESN).  Within this data
set you will find the engines used by the 4 airframe operators and the three
types of life limiting parts within the previous dataset.
```

OUT TASK

Rockwall has been asked by Aviation leadership to create a single dataset that combines all 8 datasets into a single table

Lastly, Aviation can then use this table to either create a prediction of RUL or visualize the data for easy user consumption

To assist us with this task, Aviation has designed an instance using the Dataiku tool. Here are our instructions to get started:

1. Assess the database tables provided in the source file
2. Log into the Dataiku instance:
    - Check email for an account name and password for access to the Dataiku platform - *Rockwall* will need this to complete the tasks
    - **Due to high demand in this program, please note our Dataiku account will be removed after 30 days of inactivity**
3. Use the Dataiku instance to complete the task
    - Aviation included some useful materials to help us get started in the links below
        - How to use Dataiku
        - How to select data from multiple tables by using the PostgreSQL INNER JOIN clause

ADDITIONAL RESOURCE

These videos showcase fundamental actions that one can take inside of the Dataiku application