# Data Science Guild: Hackathon

# AIRCRAFT FUEL EFFICIENCY ANALYSIS & PREDICTION

## 12-14 Jun, 2017

# Important Information

## ➢ Timeline

- **Launch: date & time** **:** 12-Jun-2017 10:00 AM – IST
- **Launch: Q&A Call** **:** 12-Jun-2017 12:00 PM – IST
- **End: date & time** **:** 14-Jun-2017 08:00 PM – (Local Time)
- **Submission Deadline** **:** 14-Jun-2017 08:00 PM – (Local Time)
- **EDA Presentations** **:** 15-Jun-2017 IST Morning (Time – TBD)
- **Winner Announcement** **:** 19-Jun-2017

## ➢ Rules & Regulation:

- Private code/analysis sharing is **not** allowed
- Public code/analysis sharing over Yammer with all participants is allowed
- Data sharing outside GE network is strictly prohibited
- Data sharing outside Hackathon participants prohibited
- Any revision/augmentation of rules will be published on Yammer and such revision(s) will be effective immediately once published
- In case of any conflict, Organizing committee's decision will be final and can't be disputed later
- A request for support, feedback or any communication between Participants & Organizers must be done through Yammer. Any other form of communication, for the purpose & duration of Hackathon, will not be entertained
- Participants are encouraged to have a healthy discussion over Yammer to promote collective learning
- Any direction/tip to help others is acceptable as long as they are shared publicly with all participants

## ➢ Starter Scripts

- To assist you in "getting started":
    - Starter Script in R & Python is provided for both – Part-A & Part-B
    - You can leverage these scripts to build your final solution on top of them

# Problem Statement

Reducing fuel consumption is extremely important for aviation industry as fuel constitutes ~ 30% of the operating cost of airlines. Reducing fuel intake can also have a significant positive impact on the environment. Hence, developing cost saving strategies especially on fuel is of prime importance to airlines. Driving fuel efficiency involves developing strategies that touch upon various aspects of airplanes – broadly some of which are highlighted below:

- Aspects related to Aircraft's actions on the ground - e.g. include reducing taxiing times to reduce engine running times which translate in to reduced fuel intake.
- Aspects related to route planning – e.g. taking shorter routes when inflight to destination taking in to consideration any altitude restrictions that exist.
- Aspects related to aircraft design – e.g. improving aerodynamics, redesigning aircraft components to conserve fuel or reducing the weight on board like installation of lighter seats.

As a Data Scientist, you are tasked with using a data driven approach in understanding the drivers of fuel consumption in aircrafts using Flight Data Recorder (FDR) archives. The challenge is divided into two parts – Part-A & Part-B. Both parts need to be solved using the same dataset.

- **Part-A:** EDA - Exploratory Data Analysis - Conduct an exhaustive EDA and derive actionable insights
- **Part-B:** Predictive Modelling - Build a predictive model

## ***Scoring***

- Total Score for a team will be calculated using following formula.
- Actionable insights derived through EDA and/or predictive model has higher weight than only a standalone predictive model

**Total Score = *0.7*Part-A Score + 0.3*Part-B Score***

## ***Data***

Flight Data Recordings (FDR) data of multiple flight instances of an aircrafts is made available. The master data represents readings of numerous sensors on aircrafts measuring detailed aircraft dynamics, system performance, and other engineering parameters. As different sensors have different sampling rates (*number of data points recorded per second*), data has been processed and normalized to provide minimum, maximum and average readings per second. The 'PH' column can be used to decipher phases of flight. The PH enumerated codes are:

- 0 = Unknown
- 1 = Preflight
- 2 = Taxi
- 3 = Takeoff
- 4 = Climb
- 5 = Cruise
- 6 = Approach
- 7 = Rollout

***For Part-A & Part-B***

- Each training file is a separate instance of flight
- Training Data: Available on Yammer *(Files section)*
- Test Data: Will be available on Yammer *(Files section)* on **14-Jun-17 Wednesday 10:00 AM**
- 'ACID' is not to be used as independent variable for prediction.
- You are expected to use Training data to build models. Any validation set you may need needs to be created from the training set. Separate validation set will not be provided.
- You may build multiple models & identify the best model for final predictions. Final predictions needs to be generated on Test data and prediction file must be submitted for evaluation.
- The variable to be predicted is **“FF - Fuel Flow”**

## ***Data Dictionary***

Available on Yammer *(Files section)*

# PART-A: Exploratory Data Analysis (EDA)

## *Objective*

- Analyze why some flying instances are more fuel efficient than others. What makes an airplane more fuel efficient during the difference phases of a flight?
- How can FDR data be used to understand drivers of fuel flow (*consumption*) and derive best practices that make flights fuel efficient under different conditions?
- Visualize given data and identify insights on what makes flights fuel efficient during different phases?
- Come up with **Top-3 Actionable Insights (with supporting approach/analysis, visualization, recommendation),** that you believe, will result in maximum improvement in the fuel efficiency

## *What Is Expected?*

1. **Explore the data**. You are free to use any tools for exploring, representing and visualizing the data to identify evidence-based insights.
2. **Visualize the data**. Data provides you the scope to draw comparisons of different flying instances of same aircraft. You may compare similar travel times in order to make sense of the data and derive useful insights. You may also need to transform the data for better visualization.
3. **Analyze and derive insights**. You may analyze different phases of the flight independently. As the drivers of fuel flow may tend to be different in different phases of the flight, it is important to understand how fuel flow patterns behave in each phase. We expect you to identify actionable insights from the analysis for different phases of the flight.

## *What Is Not Expected?*

- You are not expected to design a dashboard for ongoing use. This is about exploring and visualizing the data in innovative ways to identify key insights.
- You are not expected to simply come up with observations or information based on the data.
- It's critical that you find the underlying drivers of fuel consumption. For example, you may notice the fuel flow rate decreasing with altitude. This is expected as the drop in density lowers engine thrust requirement resulting in lower fuel flow. What would be more interesting to understand is - at what ground speed does the fuel flow rate minimize during descent and how does this compare between two different flights? Such insights can help plan a flight trajectory while minimizing fuel.

## ***Team Deliverables:***

- **Top-3 Insights (with supporting visualization) & recommended action,** that you believe, will result in maximum improvement in the fuel efficiency
- Supporting approach/analysis document, markup files, any form or reproducible research document will be "good-to-have" & will be an added advantage for the participating Teams

## ***Evaluation Criteria***

- **EDA will be assessed using following criteria:**
    - Approach Maturity
    - Analysis Quality
    - Actionability Of Insight
    - Creativity
    - Presentation/Communication

## ***General Question***

- **What is an actionable insight?**
    - An insight that is derived from your analysis that can be used in practice to improve fuel efficiency of airplanes.

# PART-B: Predictive Modelling

## *Objective*

- Predict fuel flow rate of airplanes during different phases of a flight. You can model fuel flow in different phases of the flight independently.

## *What Is Expected?*

You are required to model fuel flow rate to understand:

- What makes an airplane perform at higher levels of fuel efficiency during the difference phases of a flight (Taxi, Takeoff, Climb, Cruise, Approach, and Rollout)?
- How can the FDR data be used to understand drivers of fuel flow (consumption) and derive best practices that make flights fuel efficient under different conditions?

## *What Is Not Expected?*

- You are not expected to develop a model to be used only for prediction purposes. <u>This challenge is about utilizing the predictive model in innovative ways to identify key insights.</u>
- It's critical that you find the underlying drivers of fuel consumption. For example, you may notice that the fuel flow rate decreasing with altitude. This is expected as the drop in density lowers engine thrust requirement resulting in lower fuel flow. What would be more interesting to understand is the influence of ground speed on fuel flow rate during descent and how ground speed can be controlled to minimize fuel flow rate without dropping the speed which can delay the airplanes. Such insights can help plan a flight trajectory while minimizing fuel.

## *Team Deliverables*

- Mandatory
    - Prediction File
    - Code
- Good To Have
    - Supporting Documentation with Approach, Variable importance, Performance comparison of more than one model etc.

## *Submission Process*

- **Only one prediction file** (.csv format) needs to be submitted per team. Sample format of submission file is attached in Yammer *(Files section).*
- Code for more than one modelling technique can be submitted, however, the model used to generate the final prediction file needs to be clearly differentiated from other code.

- All submission must be sent to – ***Preet Kanwal Singh via Email @ preetkanwal.singh@ge.com***
- In case of more than one submission, only **the first** submission will be considered for evaluation – *be REALLY sure before making a submission!*
- All submissions made after contest end date & time will **not** be evaluated.
- Submission file must have prediction for each observation in Test dataset. Partial/incomplete prediction file will not be evaluated.

## ***Evaluation Criteria***

- Model performance will be evaluated based on **RMSE (Root Mean Square Error) metric.**
- Lower the RMSE, better the model.

*Good Luck & Have Fun…..*