PYDATA GLOBAL 2021

KNOW YOUR DATA FIRST:

AN INTRODUCTION TO EXPLORATORY DATA ANALYSIS

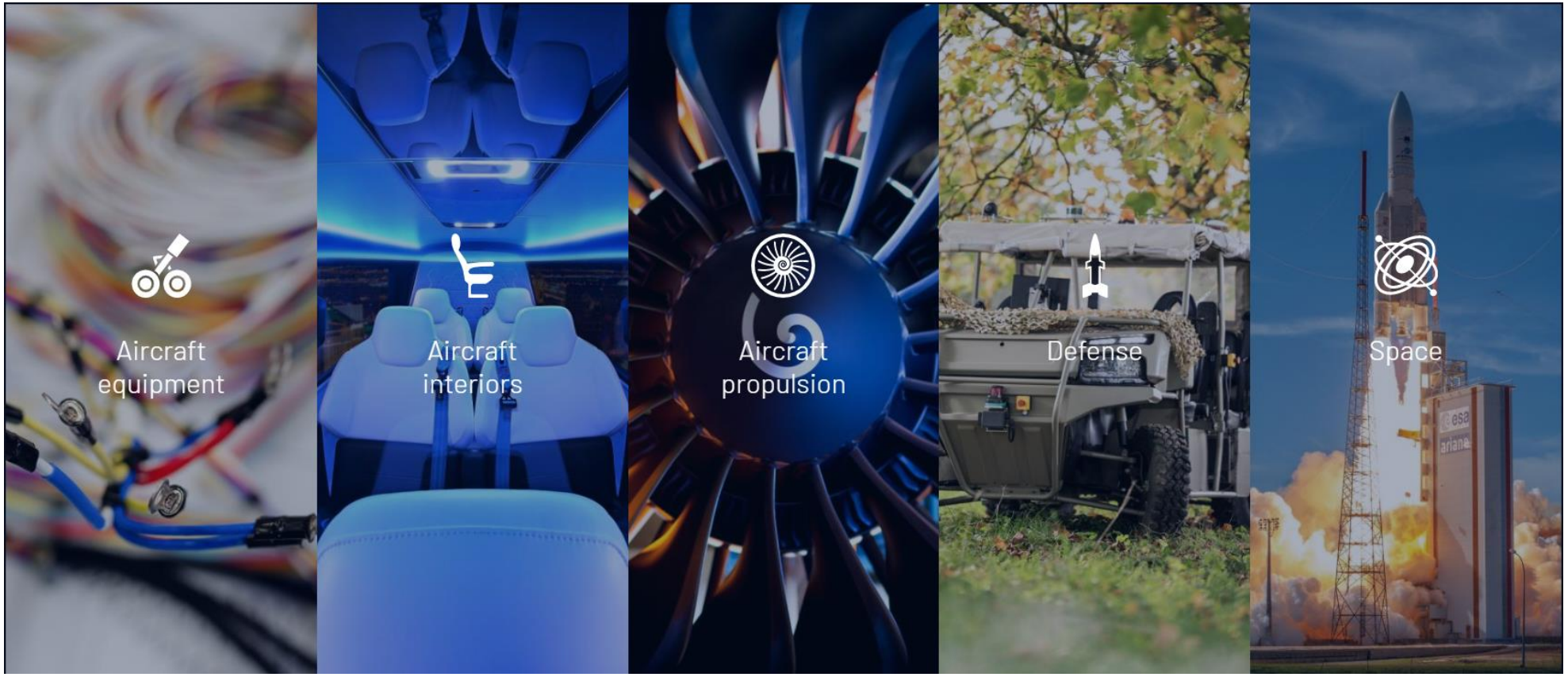
Sin-seok SEO, Safran Tech, Safran SA

OUTLINE

1. Introduction (this presentation)
   - ✓ Safran and Me
   - ✓ EDA
   - ✓ Prerequisites
2. Data Loading and Preprocessing
   - ✓ Essential check
   - ✓ *Sidetable*
3. Statistical Visualizations
   - ✓ *Matplotlib*
   - ✓ *Pandas*
   - ✓ *Seaborn*
4. (Easy Enough) Interactive Visualizations
   - ✓ *Ipywidgets*
   - ✓ *Plot.ly* and *Plot.ly express*
   - ✓ *Bokeh*
   - ✓ *Altair*
5. Automatic EDA Report
   - ✓ *Dtale*
   - ✓ *Pandas-profiling*
   - ✓ *Sweetviz*
   - ✓ *Autoviz*
6. Wrap-up and Some Tips

ABOUT SAFRAN GROUP

More than 76000 employees in 350 locations across 31 countries



SAFRAN'S AIRCRAFT ENGINES

Through CFM International (the 50/50 joint company between Safran Aircraft Engines and GE) we produce the LEAP® turbofan, successor to the best-selling CFM56®. The LEAP powers new-generation single-aisle commercial jets: the Airbus A320neo, Boeing 737 MAX and COMAC C919. We're also a leading military aircraft engine manufacturer, supplying the M88 for the Rafale fighter, and as part of a consortium making the TP400 turboprop engine for the Airbus A400M transport aircraft

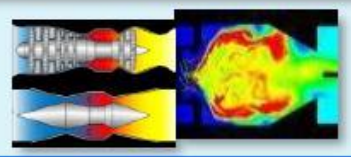
Safran Tech

Safran Research Center at Paris-Saclay

About 500 persons including 80 experts

6 Research Departments
Up to TRL 3-5*

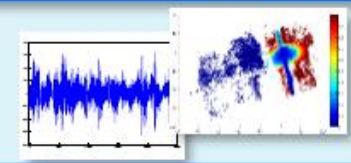
- **Energy & Propulsion**



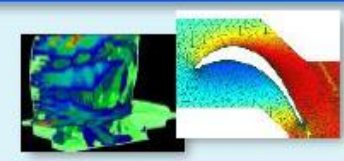
- **Electrical & Electronical Systems**

- **Materials & Processes**

- **Signal and Information Technologies**



- **Modelling & Simulation**



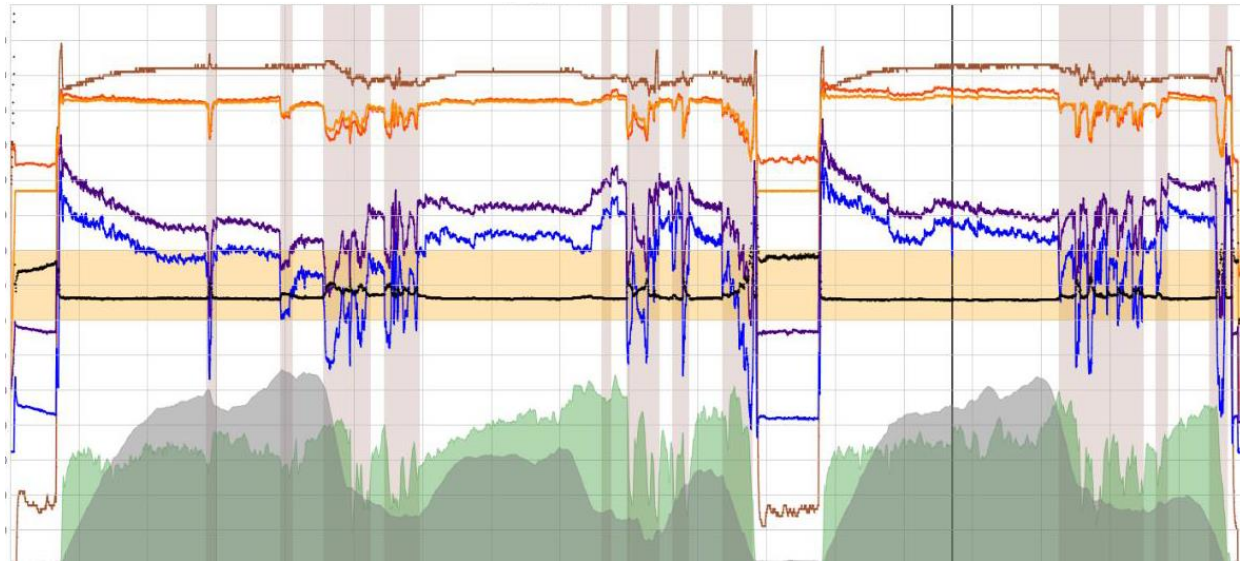
- **Sensors Technologies & applications**

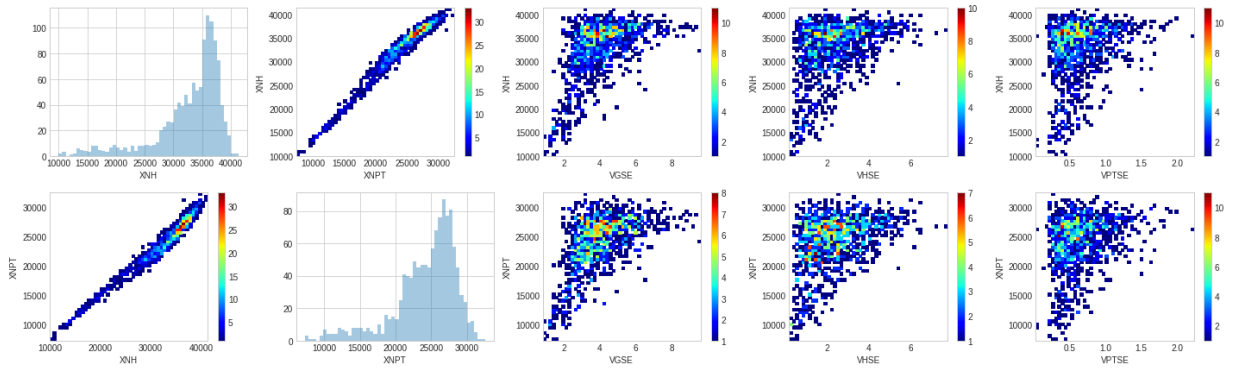
4 Technological Platforms
Up to TRL 6*

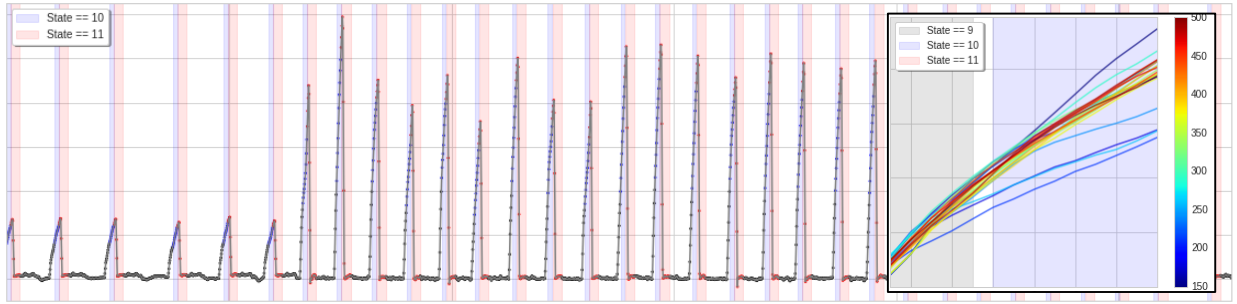
- **Safran Composites**
- **Safran Advanced Turbine Airfoils** (Experimental Foundry)
- **Safran Additive Manufacturing**
- ***Safran Ceramics***

* Technology Readiness Level

Plateforme digital







Things I do @Safran Tech

Since 2017. 04

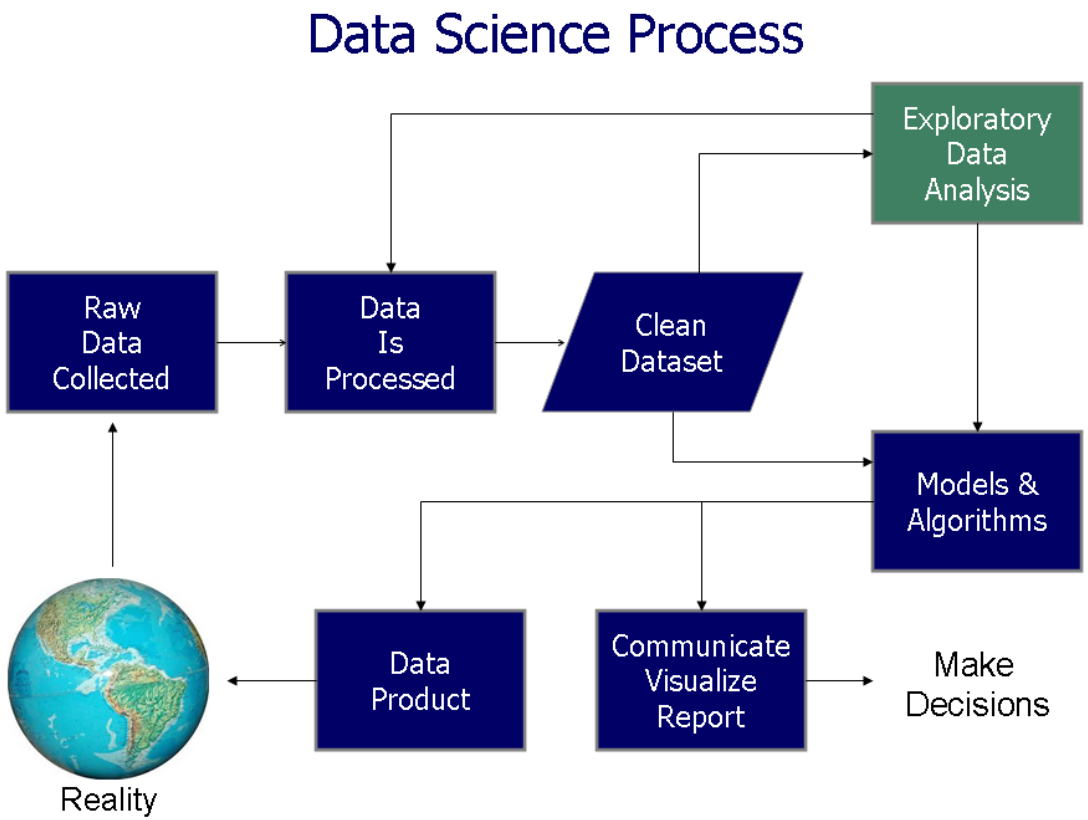
Data scientist & SW engineer

- Analyzing data obtained from airplanes and helicopters (mostly from engines)

- Applying various statistical models and machine learning algorithms to improve performances and reduce costs

- Optimizing maintenance policies

EXPLORATORY DATA ANALYSIS (EDA)



- An approach of analyzing data sets to summarize their main characteristics, often using **statistical graphics** and other **data visualization** methods

- **Objectives**
  - ✓Suggest hypotheses about the causes of observed phenomena
  - ✓Assess assumptions on which statistical inference will be based
  - ✓Support the selection of appropriate statistical tools and techniques
  - ✓Provide a basis for further data collection through surveys or experiments

Source: https://en.wikipedia.org/wiki/Exploratory_data_analysis

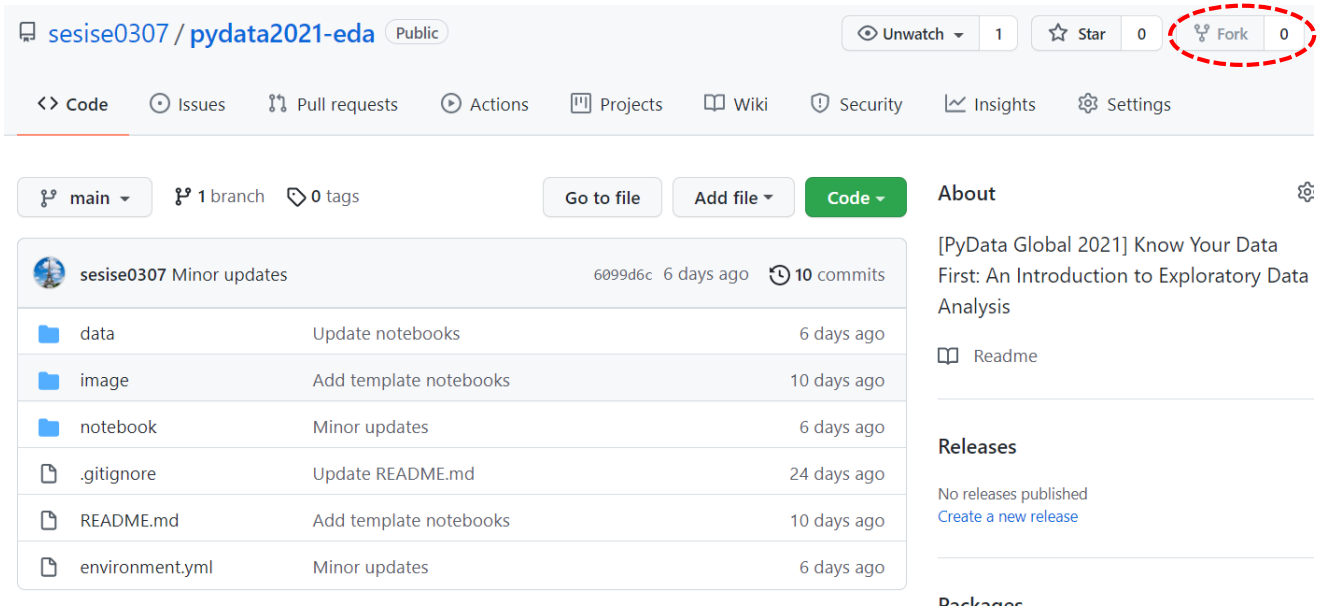
PREREQUISITE

- Some Experiences with:
  - ✓Python
  - ✓Pandas
  - ✓Matplotlib
  - ✓Jupyter Notebook (or similar)
- GitHub & Google Account

Go to: https://github.com/sesise0307/pydata2021-eda/

Fork the repo

Click!



LET'S GET YOUR HANDS DIRTY

Go to:

https://colab.research.google.com/github/{your_github_id}/pydata2021-eda/

For example:

https://colab.research.google.com/github/sesise0307/pydata2021-eda/

WRAP UP

1. Introduction (this presentation)
   - ✓ Safran and Me
   - ✓ EDA
   - ✓ Prerequisites
2. Data Loading and Preprocessing
   - ✓ Essential check
   - ✓ *Sidetable*
3. Statistical Visualizations
   - ✓ *Matplotlib*
   - ✓ *Pandas*
   - ✓ *Seaborn*
4. (Easy Enough) Interactive Visualizations
   - ✓ *Ipywidgets*
   - ✓ *Plot.ly* and *Plot.ly express*
   - ✓ *Bokeh*
   - ✓ *Altair*
5. Automatic EDA Report
   - ✓ *Dtale*
   - ✓ *Pandas-profiling*
   - ✓ *Sweetviz*
   - ✓ *Autoviz*
6. Wrap-up and Some Tips

SOME TIPS OR RECOMMENDATIONS

RTFM (Read The Fucking Manual)

Dashboarding

- Streamlit
- Dash
- Voila