Data Science Internship Challenge

In this challenge you are tasked to complete Exploratory Data Analysis (EDA) of microservice tracing data.

Background:

- **Microservice**: Microservices are an architectural and organizational approach to software development where software is composed of small independent services that communicate over well-defined APIs. These services are owned by small, self-contained teams. More information 1, 2, 3.

- **Observability:** Microservice observability is the ability to collect and present the data to gain visibility of various services. This is achieved by 3 pillars of observability logs, metrics and traces. More information 1, 2, 3.

- **Traces:** Traces track the end-to-end behavior of a request as it moves through a distributed or microservice system. The data collected in distributed tracing brings higher visibility to requests that use multiple internal microservices. Traces provide insight into how a request behaves at specific points in an application. More information 1, 2, 3.

Data:

- How to get the data:
    - Download the csv file containing trace data
- Explanation of the data:
    - First line contains header and rest of the lines are values
    - Columns details:
        - Timestamp: Time of occurrence of the event
        - traceID: Unique ID of an execution path through the system
        - spanID: Unique ID associated with execution of a logical unit
        - parentSpanID: spanID of parent span that called a given span
        - serviceName: Name of microservice associated with the span
        - Name: Method/function/endpoint name associated with the span
        - durationNano: Time taken in nanoseconds to execute the span

Requirements:

1. Create a Jupyter notebook for exploratory data analysis (EDA) of data provided.
    a. Make sure to add ample comments to convey what you are doing and why.
    b. Try to see if there are valuable insights for `duration` of same service or method name
    c. One trace can have multiple spans and some spans can have parent spans
2. Once you complete EDA, briefly answer:
    a. What algorithm, machine learning or AI approaches would you take to find anomalies in the duration of a span? And why do you think that approach is a good approach?
3. You have 1 week to complete the challenge.
4. Once completed, upload your EDA notebook to Github and share the link along with an answer to the above question (point #2).
    a. Reply back to the email from **neel@opsfree.com**