# Requirements Document: Comparative Analysis of Flight Data between Somalia CAA and NavPass

## Introduction

We are undertaking a project to compare the flight data received from the Somalia Civil Aviation Authority (CAA) for July, August, and September with the data collected internally at NavPass. The goal is to identify discrepancies between the two datasets, understand any gaps in our data collection, and enhance our overall data accuracy and coverage.

## Objectives

* **Identify Missing Data**: Determine which flights are present in the Somalia CAA data but missing from NavPass data.
* **Discover Additional Flights**: Find flights recorded by NavPass that are not present in the Somalia CAA data.
* **Analyze Discrepancies**: Understand the common characteristics of the missing or additional data to identify patterns or reasons for discrepancies.
* **Enhance Data Collection**: Utilize findings to improve NavPass’s data collection methods and processes.

## Scope of Work

The Data Science team is expected to:

1. **Data Comparison**:
   * Perform a comprehensive comparison between the Somalia CAA flight data and NavPass’s internal data for the specified months.
   * Use unique identifiers to match flights across both datasets.
2. **Gap Identification**:
   * List all flights present in the Somalia CAA data but absent in NavPass data.
   * Identify flights recorded by NavPass that do not appear in the Somalia CAA data.
3. **Characteristic Analysis**:
   * Analyze the common traits of missing or additional flights, such as:
     + Flight times (e.g., time of day, dates)
     + Airlines or operators
     + Aircraft types or models
     + Flight paths or routes
     + Departure and arrival locations
4. **External Data Utilization**:
   * Leverage external data sources to supplement analysis as needed, including:
     + FlightAware
     + RadarBox
     + Web searches
     + ChatGPT with web search capabilities
   * Cross-reference discrepancies to validate findings.
5. **Data Requests**:
   * Identify any additional data required from internal teams (e.g., detailed aircraft data) and communicate requests promptly.

## Deliverables

The following deliverables are expected from the Data Science team, ideally this is iterative delivery.

1. **Initial Findings Report**:
   * A summary of initial discrepancies found between the two datasets.
   * Preliminary insights into missing or additional flights.
2. **Detailed Analysis Report**:
   * An in-depth analysis of the gaps identified.
   * Visual representations (charts, graphs) illustrating key patterns and trends.
   * A breakdown of common characteristics among discrepant data.
3. **Recommendations Document**:
   * Actionable suggestions on how to address the identified gaps.
   * Strategies to improve data collection and processing at NavPass.
   * Recommendations for ongoing data validation methods using external sources.
4. **Supplementary Data Files**:
   * Lists of flights missing from each dataset.
   * Annotated datasets with notes on discrepancies and validations.

## Purpose and Importance

Understanding the discrepancies between our data and that of the Somalia CAA is crucial for several reasons:

* **Data Accuracy**: Ensures that NavPass maintains a high standard of data accuracy and reliability.
* **Operational Efficiency**: Identifies potential blind spots in our data collection processes that could affect operational decisions.
* **Strategic Insights**: Provides valuable insights that can inform future strategies for data acquisition and management.
* **Defensibility:** This allows us to build confidence with the CAA that we are capturing all flights.

## Expectations

* **Collaboration**: The Data Science team should work closely the core team as needed.
* **Communication**: Regular updates on progress are expected to keep stakeholders informed.
* **Quality Assurance**: All findings should be thoroughly validated and documented.
* **Timeliness**: Deliverables should be provided in a timely manner, with iterative progress allowing for feedback and adjustments.