

Project H6: KAGGLE-OpenSky Aircraft States Snapshot

Team members: Kristo Hark

Task 1:

Repository link: <https://github.com/Kristo-H/opensky>

Task 2

- Identifying your business goals
 - **Background:** Air-traffic monitoring is important for aviation research and understanding global flight patterns. OpenSky Network provides free data from worldwide sensors and this project uses a single snapshot of global air traffic. The dataset can be used by air-traffic analysts and students to analyze flight behaviors, detect anomalies, and explore predictive modeling. Although the dataset is static, it can reveal traffic density, flight clustering and basic prediction of barometric altitude. The project focuses on practical outcomes useful for analysis and education.
 - **Business goals:** Clean the dataset and handle missing or inconsistent values. Visualize global flight positions, altitudes, velocities. Identify clusters of flights to detect patterns and anomalies. Predict barometric altitude using flight features.
 - **Business success criteria:** Clear and interpretable visualizations. Meaningful clusters which reflect flight behaviors. Predictive model with acceptable accuracy (MAE <15%)
- Assessing your situation
 - **Inventory of resources:** dataset (877.35kB), Python tools (pandas, scikit-learn, visualization libraries), laptop, Kaggle notebooks, analyst
 - **Requirements, assumptions, and constraints:** access to dataset and tools. Some fields may be missing.
 - **Risks and contingencies:** Missing values - can drop values. Unclear clusters - try multiple algorithms. Low prediction accuracy - do feature selection/tuning
 - **Terminology:**
 - icao24 - aircraft identifier, used to track and identify it
 - callsign - radio ID used to identify an aircraft during flight, used by pilots
 - origin_country - the country where the aircraft is registered
 - longitude - the east/west coordinate of the aircraft's position, measured in degrees, used together with latitude to determine exact location
 - latitude - the north/south coordinate of the aircraft's position, measured in degrees
 - baro_altitude - aircraft's altitude calculated from barometric air pressure

- **on_ground** - shows if the aircraft is on ground, true when on ground, false when airborne
 - **vertical_rate** - the rate at which an airplane ascends or descends, shows how quickly it is climbing or descending during flight
 - **velocity** - the aircraft's speed over the ground
 - **clustering** - data analysis technique that groups aircraft based on similar attributes to identify traffic patterns and anomalous flights
- **Costs and benefits:** costs are minimal(time, computing). Benefits are understanding flight patterns, detecting anomalies and predictive modeling.
- Defining your data-mining goals
 - **Data-mining goals:** Clean and prepare the data. Visualize positions, altitudes, velocities. Cluster aircraft based on behavior. Train decision tree/random forest to predict barometric altitude.
 - **Data-mining success criteria:** Visualizations clearly communicate patterns. Clusters are interpretable. The predictive model meets the accuracy threshold. Output is useful for intended users.