

# Project Report: Estimating Take-Off Weight (TOW) of Airplanes for Eurocontrol & OpenSky Network PRC Data Challenge 2024

**Authors:** Andrey Belkin & Anton Bogachev

**Project Team:** Team Bold Emu

---

## 1. Introduction

The objective of this project is to develop a robust model for estimating the take-off weight (TOW) of airplanes based on a dataset provided by Eurocontrol and the OpenSky Network. The project encompasses the integration of an initial dataset with trajectory-derived variables, enhancing predictive accuracy by factoring in critical real-time conditions and economic influences on aviation performance.

Our approach to achieving this is through a multivariate analysis and variable extraction strategy, targeting specific influential components such as takeoff processes, jet streams, and meteorological conditions. By combining statistical analysis and comprehensive dataset structuring, this report summarizes our findings, methodologies, and proposed improvements.

---

## 2. Initial Data Analysis

We began by analyzing the initial dataset (`challenge_set.csv`), which served as a foundation for estimating TOW. This preliminary analysis allowed us to establish baselines and identify which existing variables required enhancement or supplementation.

Through R-based statistical modeling, we identified that the initial variables provided insufficient context for accurately modeling TOW. This necessitated the addition of trajectory-derived and economically driven variables, leading us to incorporate external datasets and custom scripting in Python 3.10 using `pandas` and `pyarrow`.

---

## 3. Variable Extraction and Model Improvement

Our model focuses on three main components to achieve improved accuracy in TOW estimations:

- **Takeoff Process Variables:** By examining the initial altitude and ground speed changes post-take-off, we were able to extract variables that reflect take-off dynamics in real time. These parameters were integrated into the model to capture variations directly impacting fuel consumption and, by extension, TOW.
- **Jet Stream Influence:** We introduced variables to measure jet stream intensity and direction at take-off locations, including wind speed thresholds that correlate with increased fuel consumption. These jet stream indicators provide insights into the impact

of crosswinds and tailwinds on TOW, aligning model predictions more closely with practical observations.

- **Weather Conditions:** Atmospheric variables such as humidity, wind, and temperature were critical additions, though they posed integration challenges due to data fluctuations. These metrics were modelled to simulate conditions accurately at various take-off points, creating a more adaptable and responsive TOW model.
- 

#### 4. Incorporating Economic Indicators

Acknowledging the economic influences on TOW, we included the MSCI World ETF index and oil prices as proxies for broader market conditions. The integration of these indices provided an economic context that aligned with our model's predictive performance, reflecting demand cycles and cost structures influencing airplane loading practices.

This integration yielded insights on how economic factors affect TOW, establishing a direct correlation between these variables and TOW adjustments. The fact that these variables improved our model hints to a close relationship between global economic factors and the TOW.

Due to database constraints and high cost of specific economic data our approach included a simplified version of this approach. Precise data on global changes in specific factors could improve the model even further.

---

#### 5. Future Improvements

Given the complexity of TOW estimation, we recommend exploring additional enhancements, such as:

- **Broader Economic Models:** Incorporate GDP per capita, trade volume indicators and price indices to reflect seasonal and region-specific loading trends.
  - **Fuel Tankering Practices:** Model adjustments for fuel tankering under varying fuel costs, simulating load adjustments based on the price differential across regions.
  - **CO2 Emission Pricing:** Integrate emission cost data to simulate weight adjustments airlines may make in response to environmental pricing structures.
- 

#### 6. Conclusion

Our work on the PRC Data Challenge has underscored the importance of integrating real-time, trajectory-derived data and economic indicators to estimate TOW with greater accuracy. The project's multi-dimensional approach, leveraging both atmospheric conditions and economic context, enhances the predictive robustness of TOW modeling. Further development along the lines of economic factors and operational practices will enhance our model's adaptability to evolving aviation demands.

