Flight Delay Challenge!

Nazila Fazeli, Johannes Kutsch, Markus Lohmeyer, Sebastian Hagemann
30.01.2026

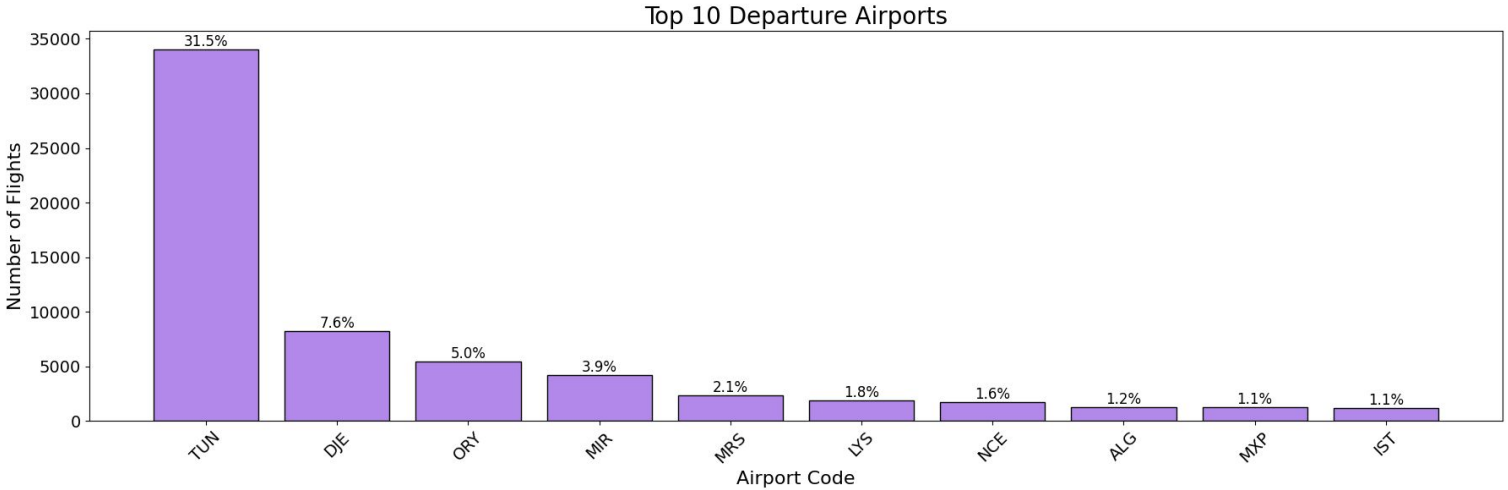
Categorical Challenges

- Multiple Categorical Features (Airports, Aircrafts, Flight ID's…) with (100+ Categories)

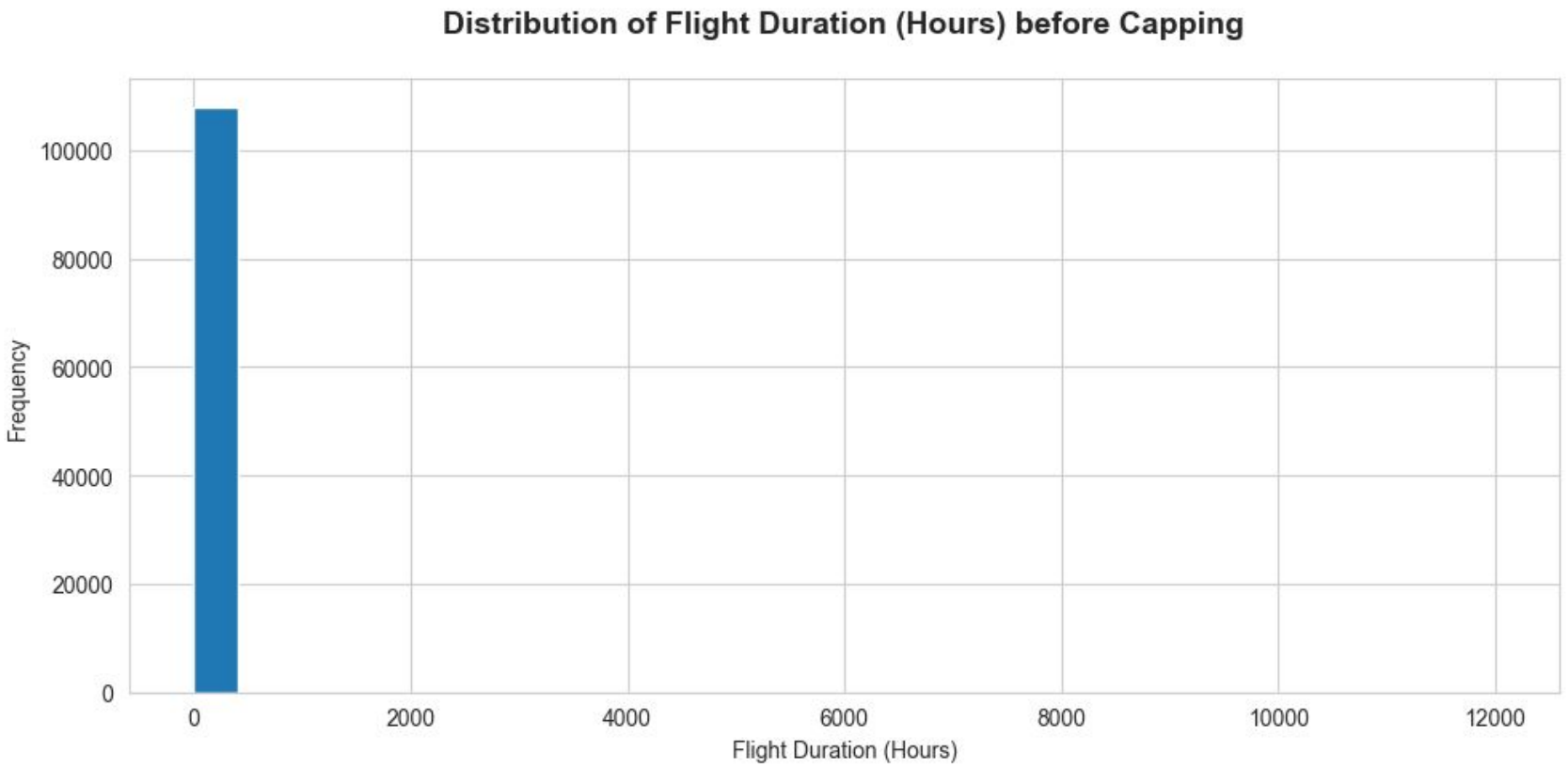
	n_unique
ARRSTN	127
DEPSTN	125
AC	68
dep_tz	53
arr_tz	53
dep_country	52
arr_country	51
STATUS	5

Categorical Challenges

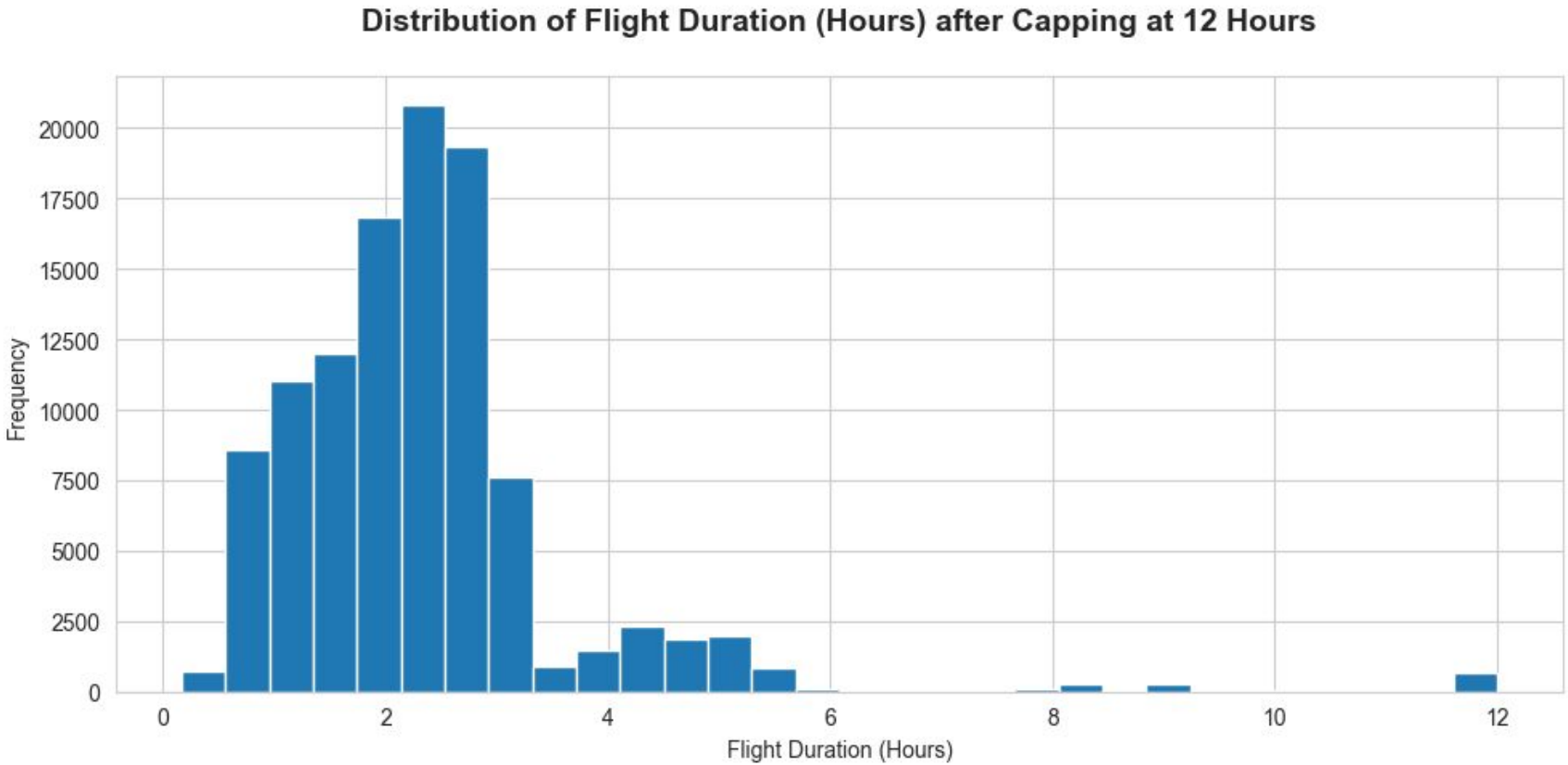
- Few features are the majority of the dataset
- We have features that are very similar (DEPSTN, dep_country)



Dealing with Numerical Challenges

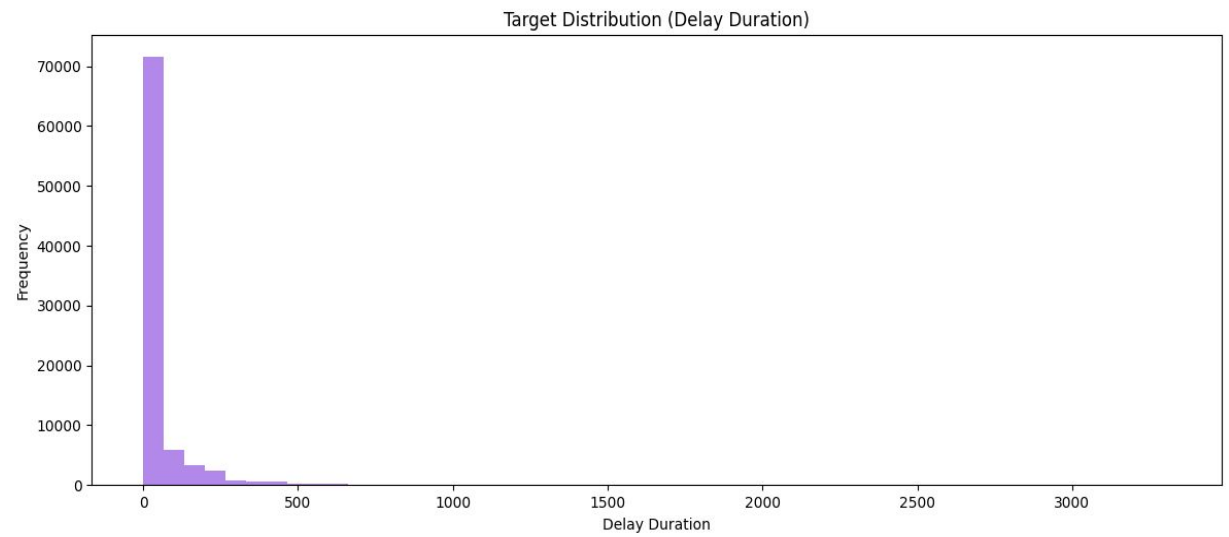


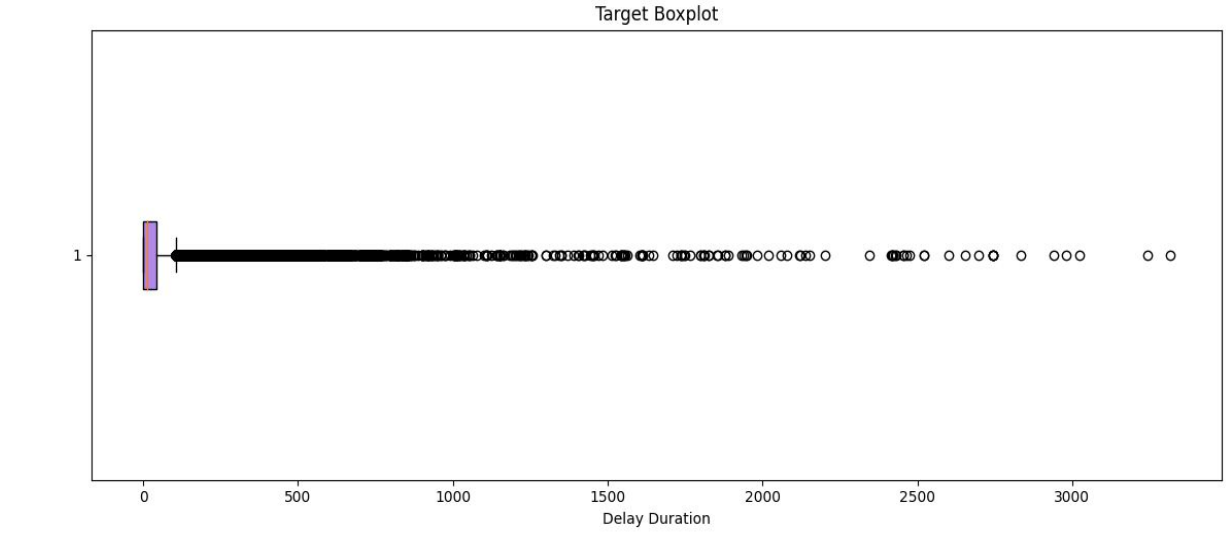
Dealing with Numerical Challenges



Dealing with Categorical Challenges

- Clustering
  - Let an algorithm reduce the cardinality of a feature
- Top-K
  - Use only the n most frequent categories from a feature
  - Make all other categories a cumulative category “other”
- Frequency encoding
  - Use frequency of categories within feature
- Target encoding
  - Use the target value instead of category
- Count encoding
  - Use counts of the category instead





Challenge with Target Data

- The target variable with high right skewness
- Outliers over 3000 minutes (more then 2 days)
- And a lot of Zeros

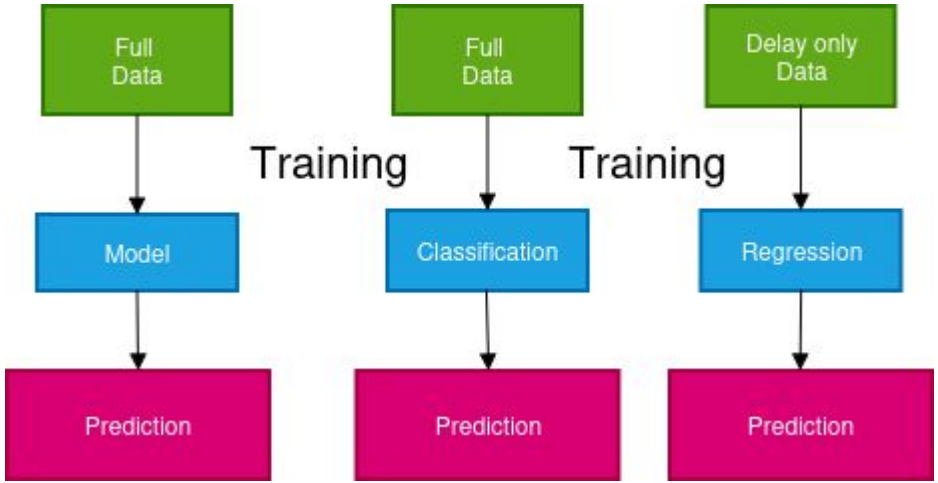
Target Preprocessing Pitfalls

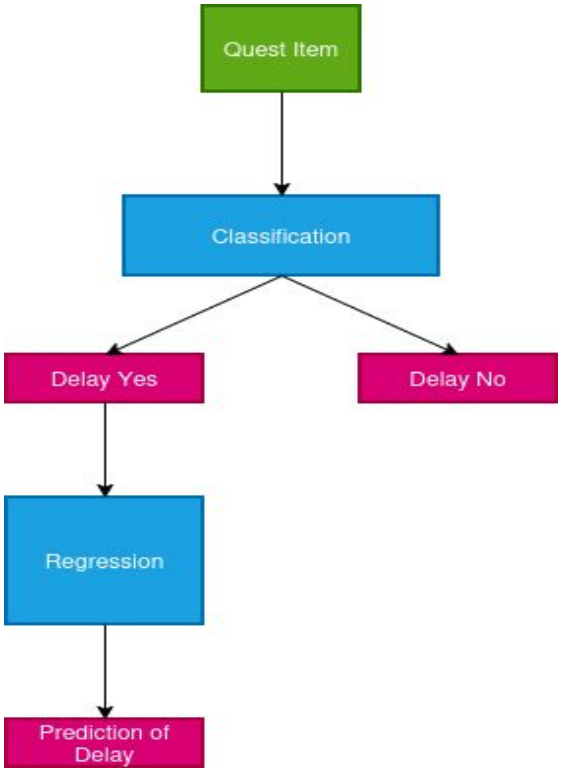
Idea: Capping the target value (put outliers within IQR multiples) will make the data more predictable
=> RMSE decreased significantly

Did the predictions became better?
=> No!

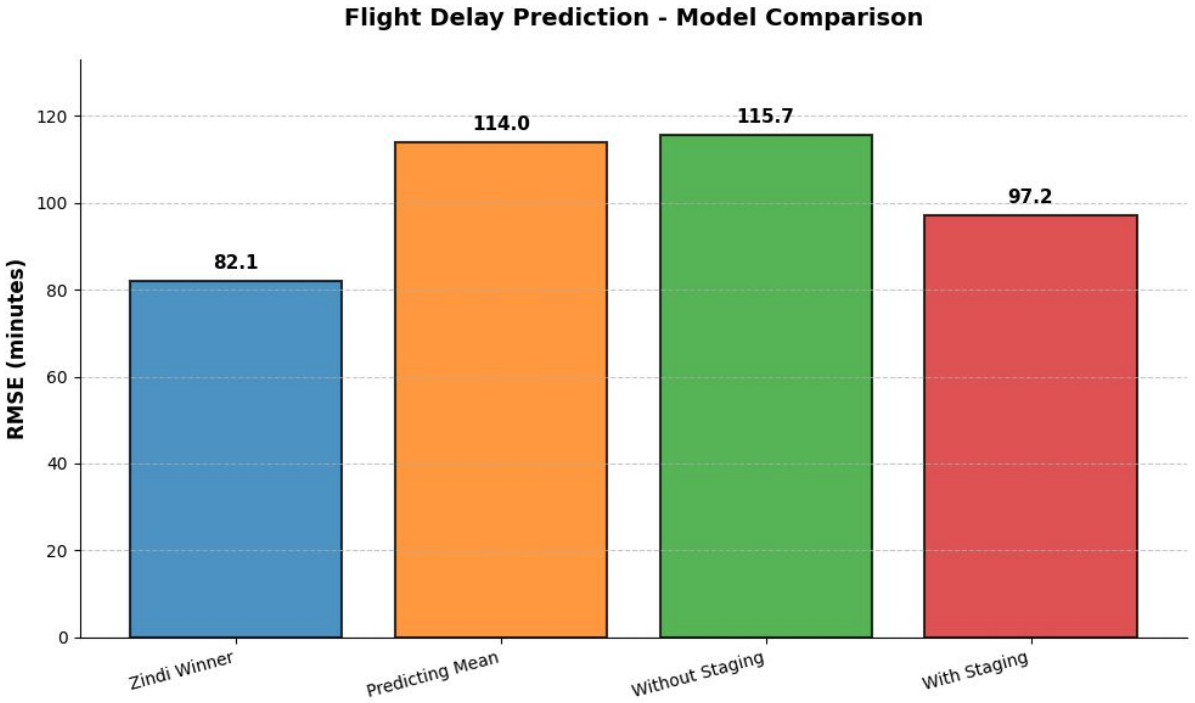
=> Changing the scale of target data changes the magnitude of error calculation

Layered Modeling





Performance Comparison



Staging improves model significantly!

Take Home Messages

- High Cardinality can be a problem
  - Binning, Clustering, Replacing (Frequency, Count, Target value)
- Regression is problematic when target contains many zeros
  - => Staged model: Classification + Regression
- Garbage in Garbage out
  - Don’t underestimate the importance of data preparation