

Flight Delay Feature Selection - Justification via EDA

Selected Features

Categorical Features:

- Airline
- Origin
- Dest
- DepTimeBlk
- ArrTimeBlk
- DayOfWeek
- Month

Numerical Features:

- CRSDepTime
- DepDelayMinutes
- Distance

Categorical Feature Analysis

Airline:

- Delay proportions vary significantly across different airlines.
- Operational differences contribute to this variation.

Origin & Dest:

- Some airports have higher delay rates.

- Affected by congestion, weather, and runway capacity.

DepTimeBlk & ArrTimeBlk:

- Midday and evening blocks show more delays.
- Peak times lead to air traffic congestion.

DayOfWeek:

- Weekends, especially Sunday, show higher delays.
- Possibly due to travel demand or staffing.

Month:

- Summer months (June-July) show higher delays.
- Weather and demand contribute.

Numerical Feature Analysis

CRSDepTime:

- Slight time shift between delayed/non-delayed.
- Time of day impacts traffic and delays.

DepDelayMinutes:

- Strong indicator of delay.
- High values correlate with delay class.

Distance:

- Slightly higher delays on long flights.
- More variables involved in longer routes.

Conclusion

The selected features capture airline operations, airport congestion, and temporal patterns.

They are statistically significant and visually interpretable for predictive modeling.