NEW YORK CITY FLIGHT DELAYS

Study Completed by Ridwan Alam

Summer Flight Departure Delays from New York City

01

TARGET AIRPORTS

La Guardia (LGA), John F. Kennedy (JFK), and Newark (EWR)

02

DATA SOURCE

June, July, and August 2019 Data from Bureau of Transportation Statistics

03

TARGET AUDIENCE

Frequent Business & Leisure Travelers

WHY IS MY FLIGHT DELAYED?

01 AIR CARRIER

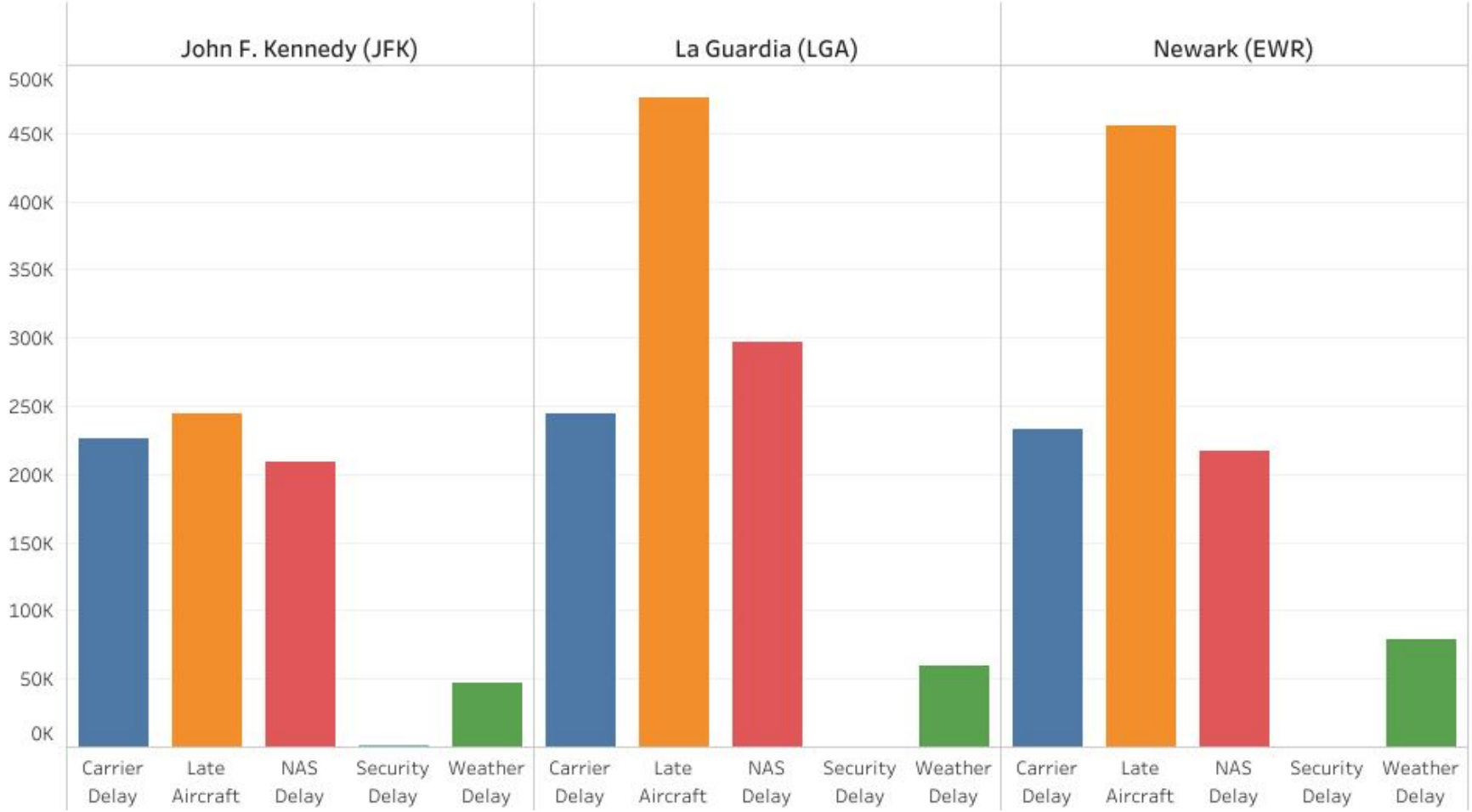
02 EXTREME WEATHER

03 NATIONAL AVIATION SYSTEM (NAS)

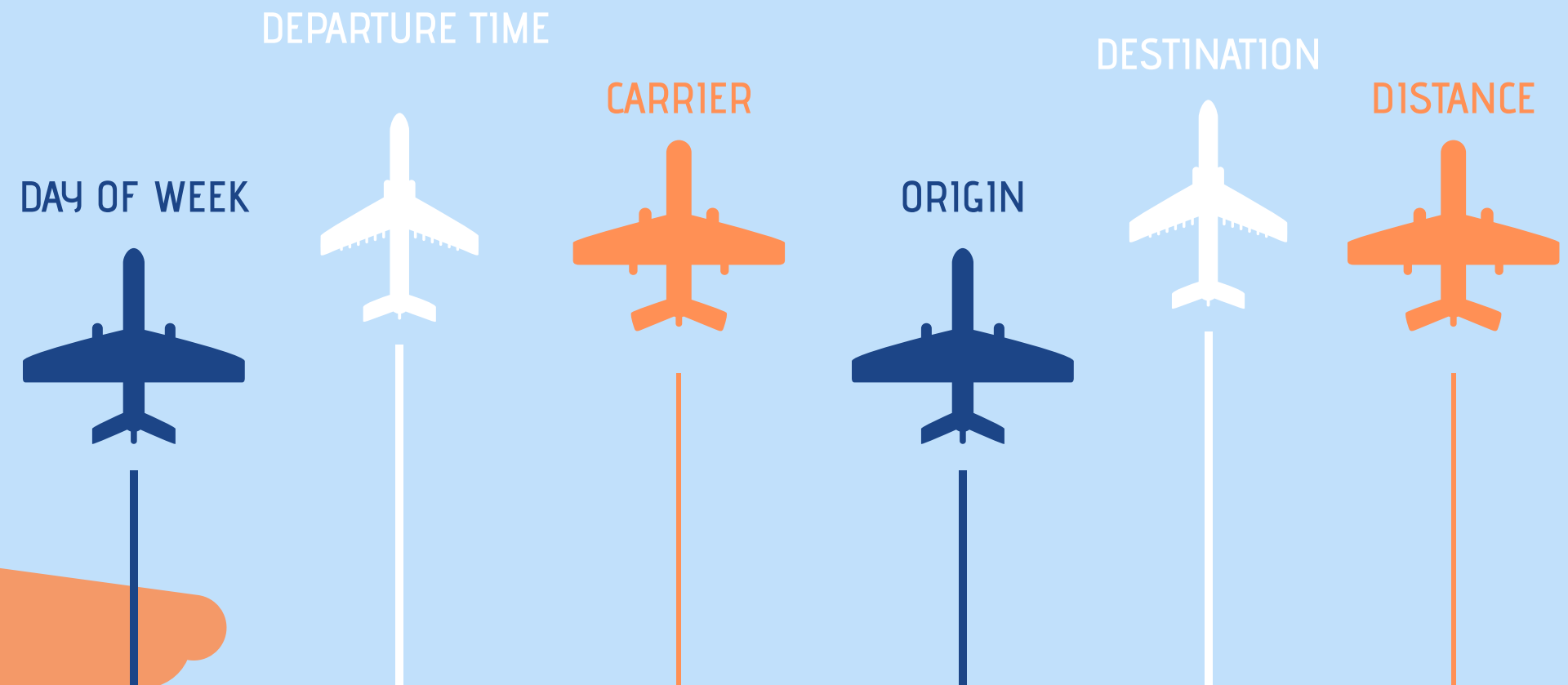
04 LATE ARRIVING AIRCRAFT

05 SECURITY DELAYS

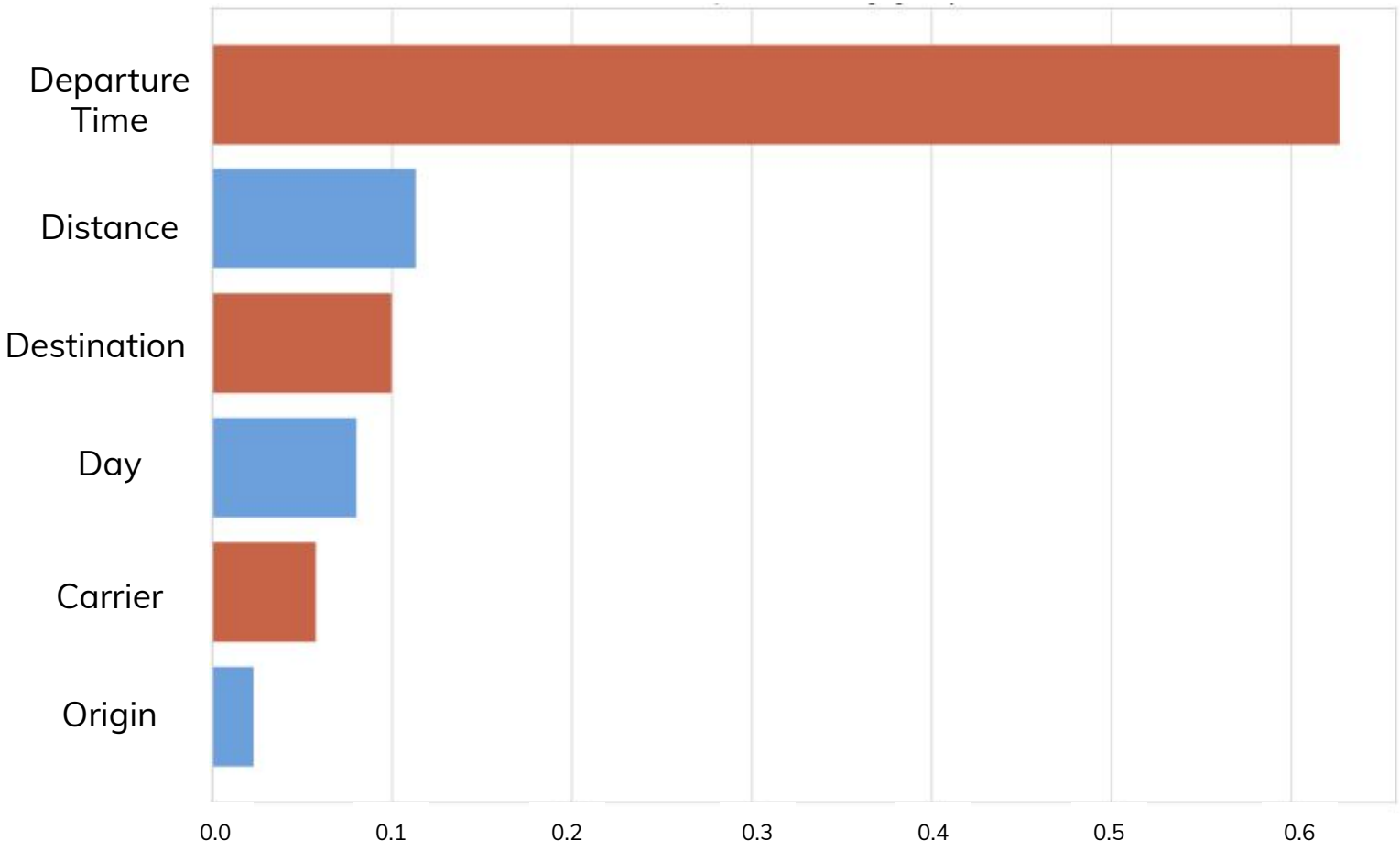
TOTAL DEPARTURE DELAY IN MINUTES



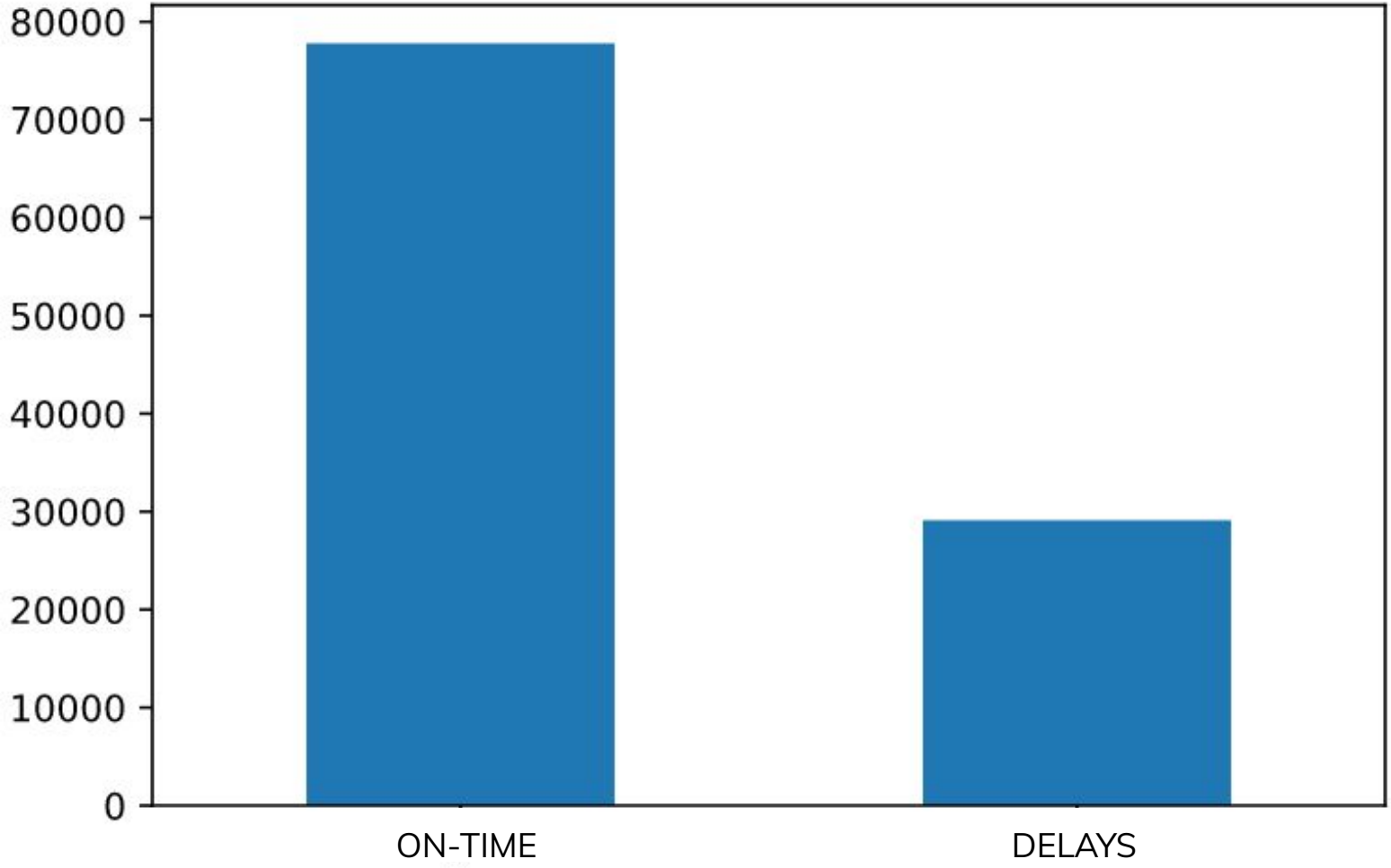
FEATURE SELECTION



FEATURE IMPORTANCE



CLASS IMBALANCE



TIME TO BALANCE!

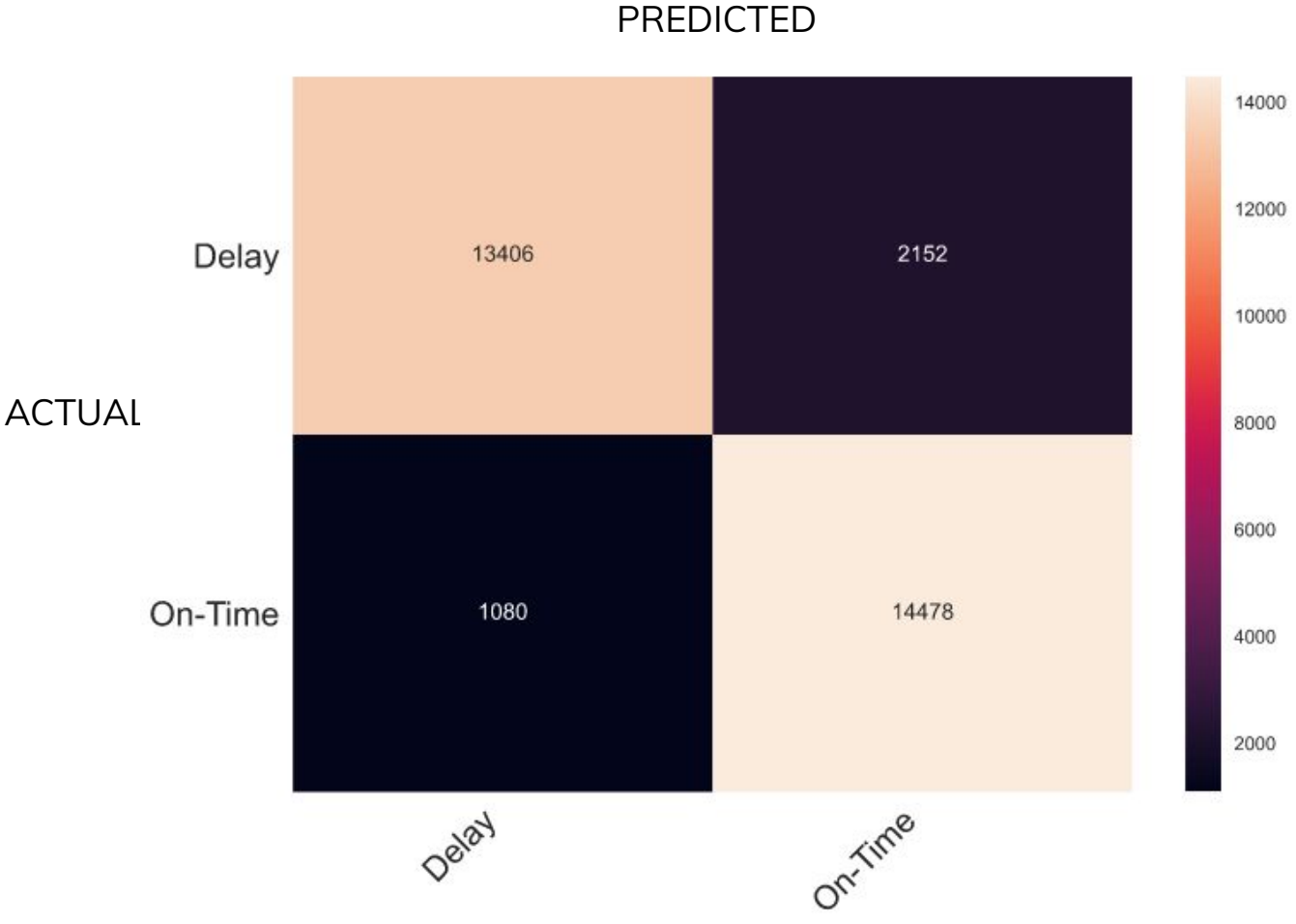
01 RANDOM OVERSAMPLING

02 SMOTE

03 ADASYN

04 RANDOM UNDERSAMPLING

SMOTE


PREDICTED
ACTUAL
Delay
On-Time
13406
2152
1080
14478
Delay
On-Time
14000
12000
10000
8000
6000
4000
2000

SELECTED MODELS

KNNeighbors

Logistic Regression

Gaussian Naive Bayes

Decision Tree Classifier

Random Forest Classifier

Gradient Boosted Classifier

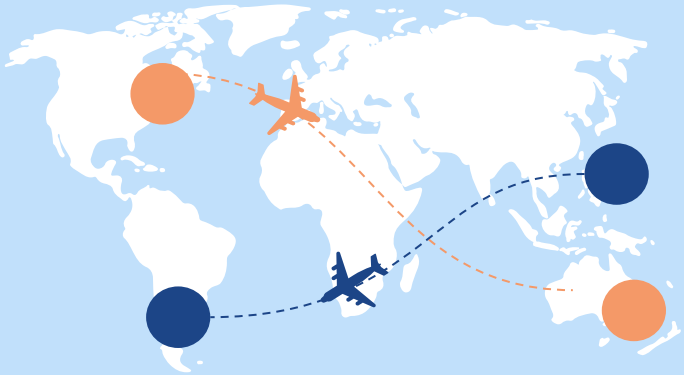
FOCUSED MODELS

Random Forest Classifier

Gradient Boosted Classifier

GRADIENT BOOSTED CLASSIFIER + SMOTE

	SCORES
ACCURACY	90%
PRECISION	90%
RECALL	90%
F1	90%



GRADIENT BOOSTED CLASSIFIER ROC-AUC CURVE

Score = 0.95

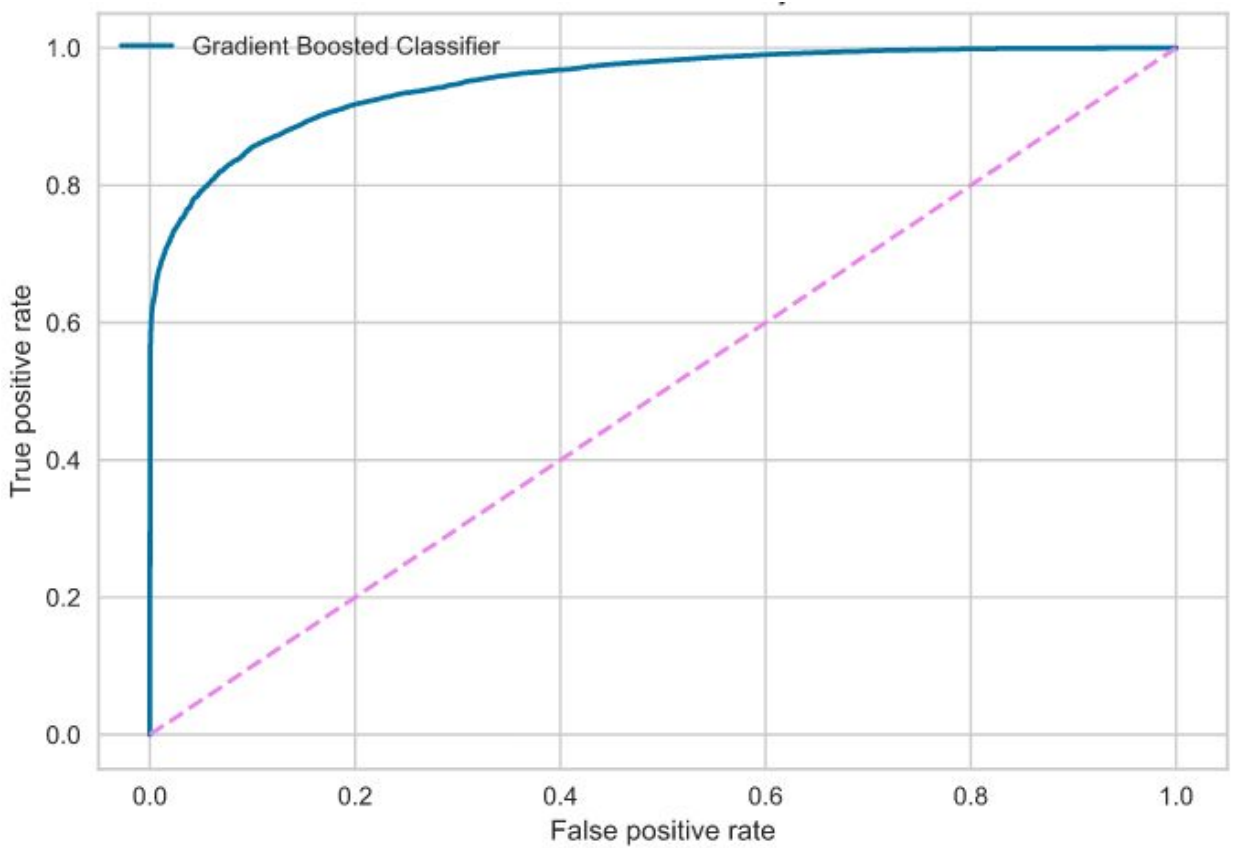
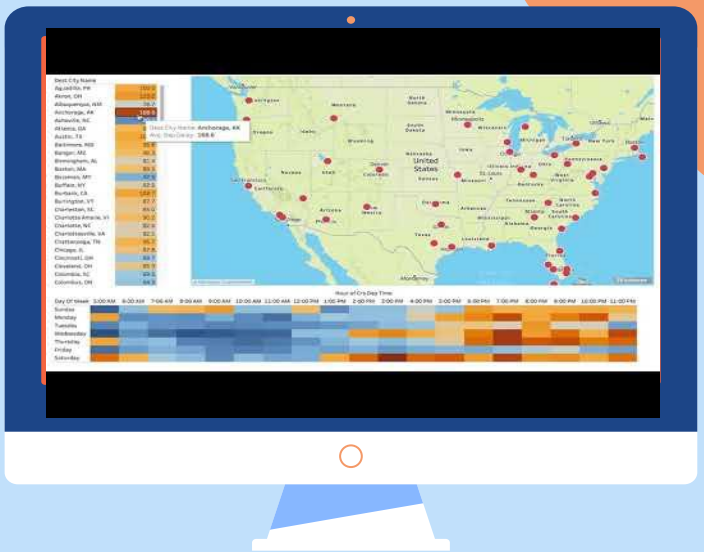
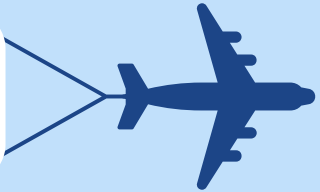


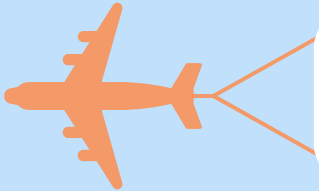
TABLEAU DASHBOARD DEMONSTRATION



CONCLUSION

Highest Impact:
Departure Time & Distance





FUTURE ITERATIONS

Focus on Multiclass (Weather Delay, Carrier Delay, etc.)

THANKS

Do you have any questions?

Tableau Dashboard
GitHub Repo

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik
Please keep this slide for attribution

slidesgo