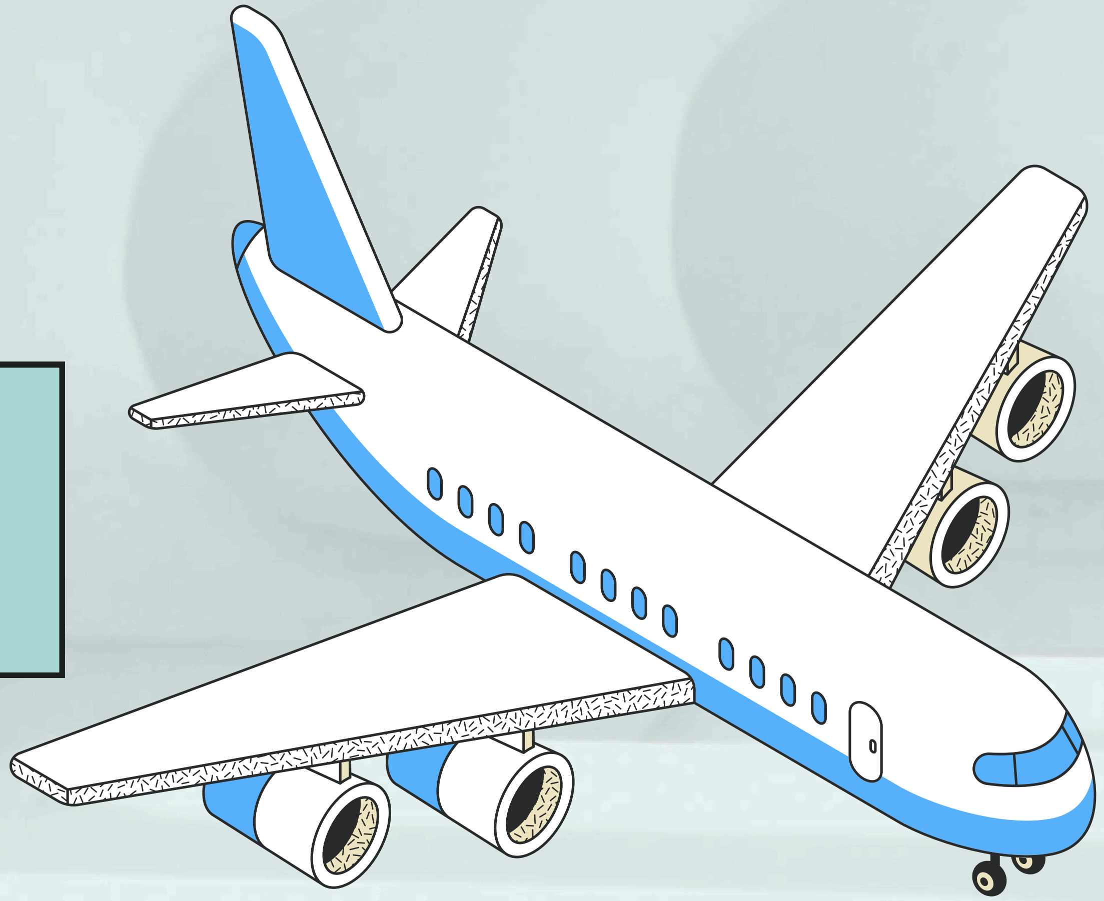
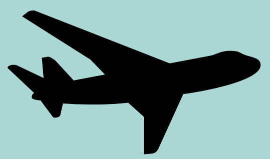
AIRLINE DELAYS AND CANCELLATION ANALYSIS

PRESENTED BY
AMINA SADIQ AND NIKHILA BOOSAM



Introduction

- The surge in Air Travel has resulted in a notable uptick in flight delays, intensifying airport congestion and imposing financial burdens on the airline sector.
- These delays not only inconvenience passengers but also impose significant economic burdens, with estimates suggesting annual costs reaching billions of dollars.
- Our study aims to forecast flight delays to improve both passenger experience and operational efficiency within the aviation industry.
- The data for 2018 airline delay and cancellations was sourced from Kaggle.



Helps optimize schedules, reduce operational costs, and improve customer satisfaction.

Provides insights into which airlines and airports are more reliable.

Can aid in better regulations and infrastructure improvements.

RESEARCH QUESTIONS

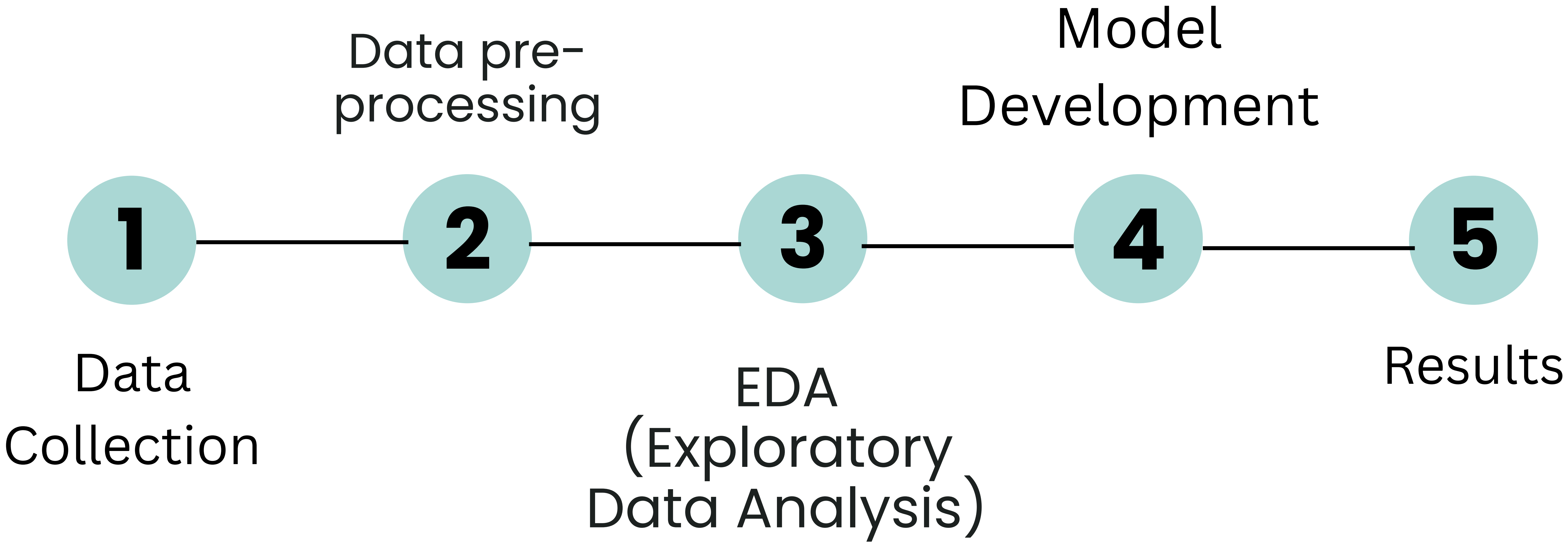
1.What are the main causes of flight delays and cancellations?

2.How do flight delays vary by airline, airport, and time of day?

3.What scheduling strategies could airlines adopt to reduce delays?

4.Can we develop a predictive model to anticipate flight delays and cancellations?

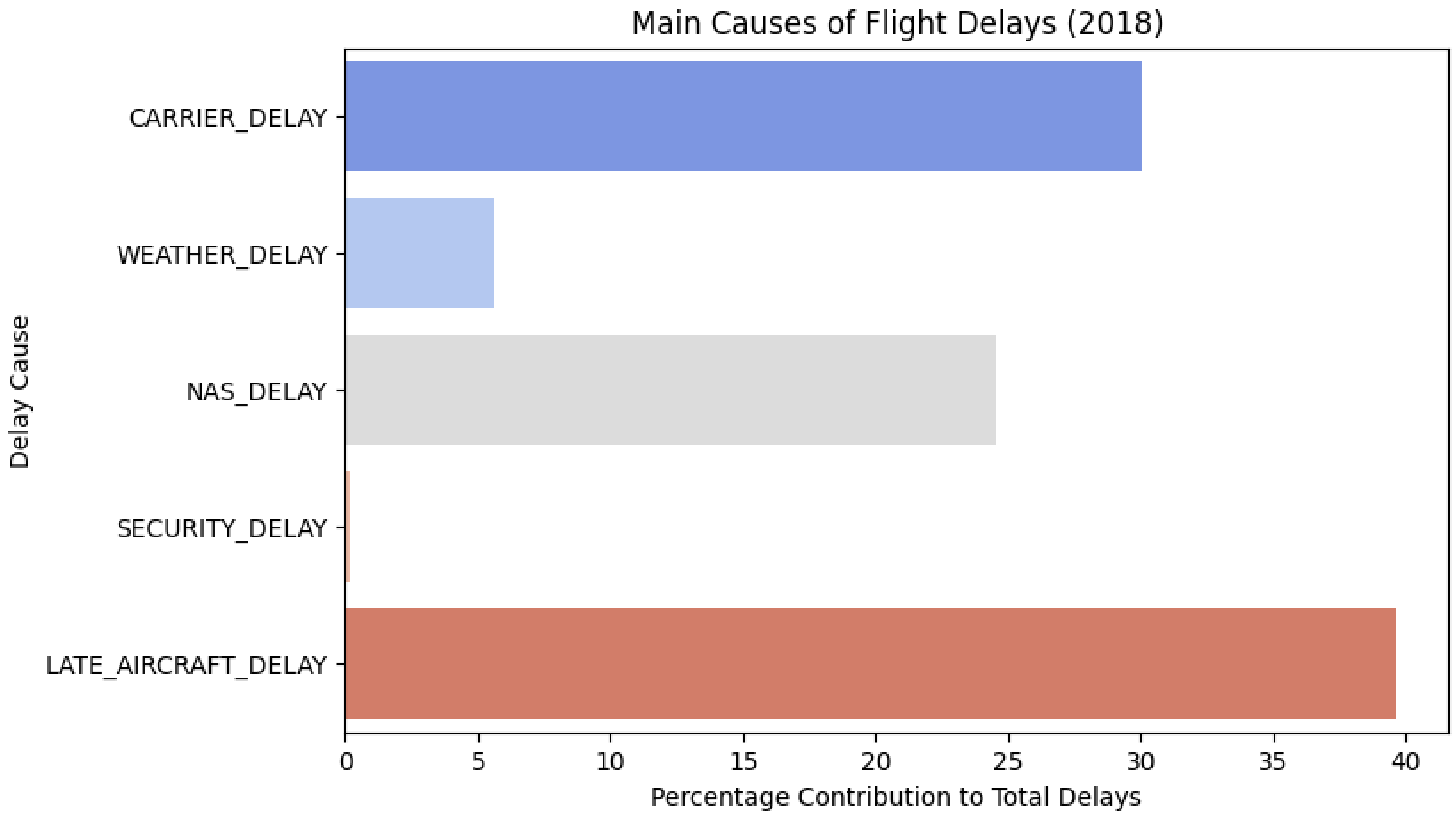
5.How do airline-specific operational strategies influence delay management?


Data pre-processing
Model Development
1
2
3
4
5
Data Collection
EDA (Exploratory Data Analysis)
Results

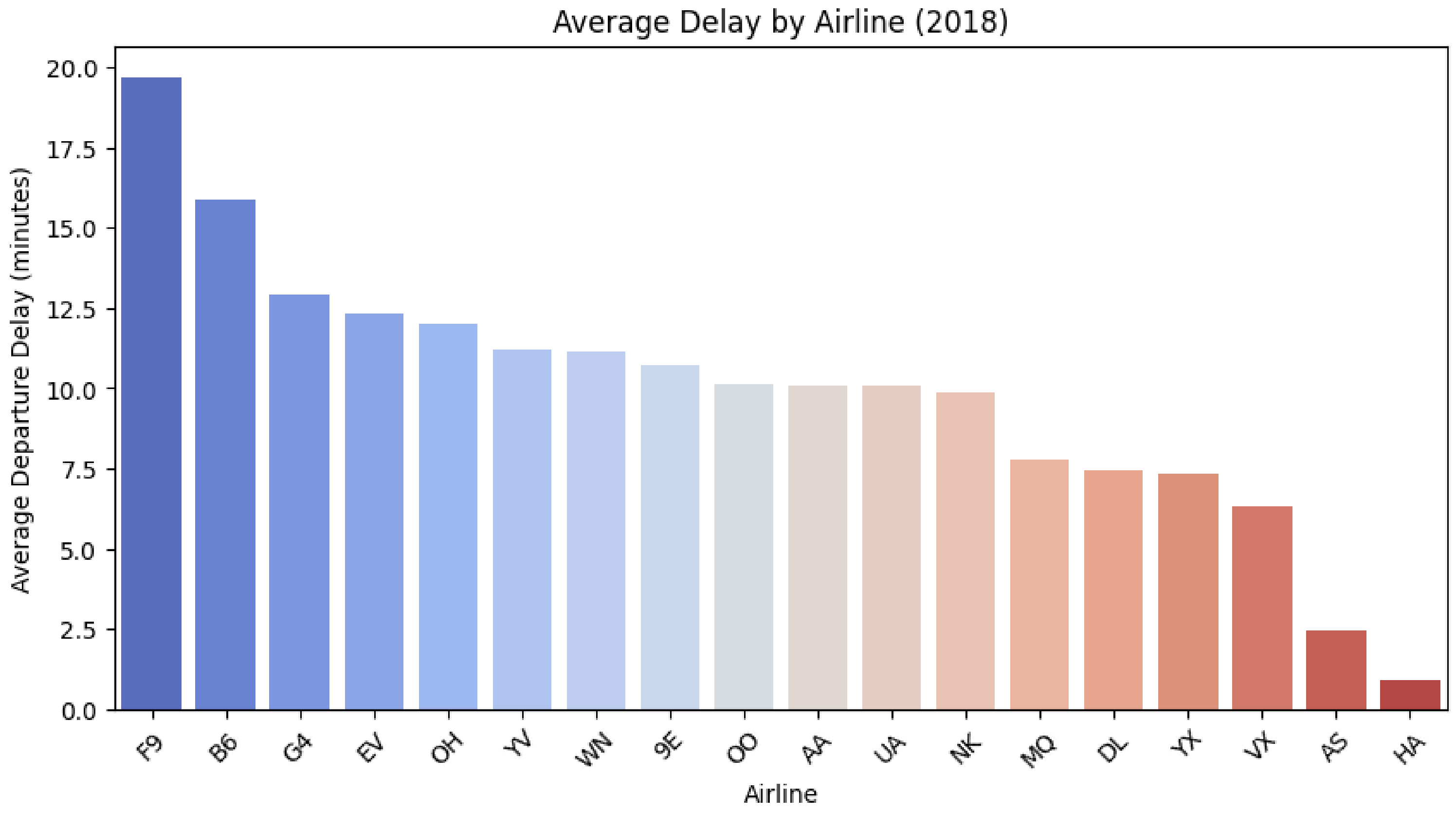
Dataset Overview

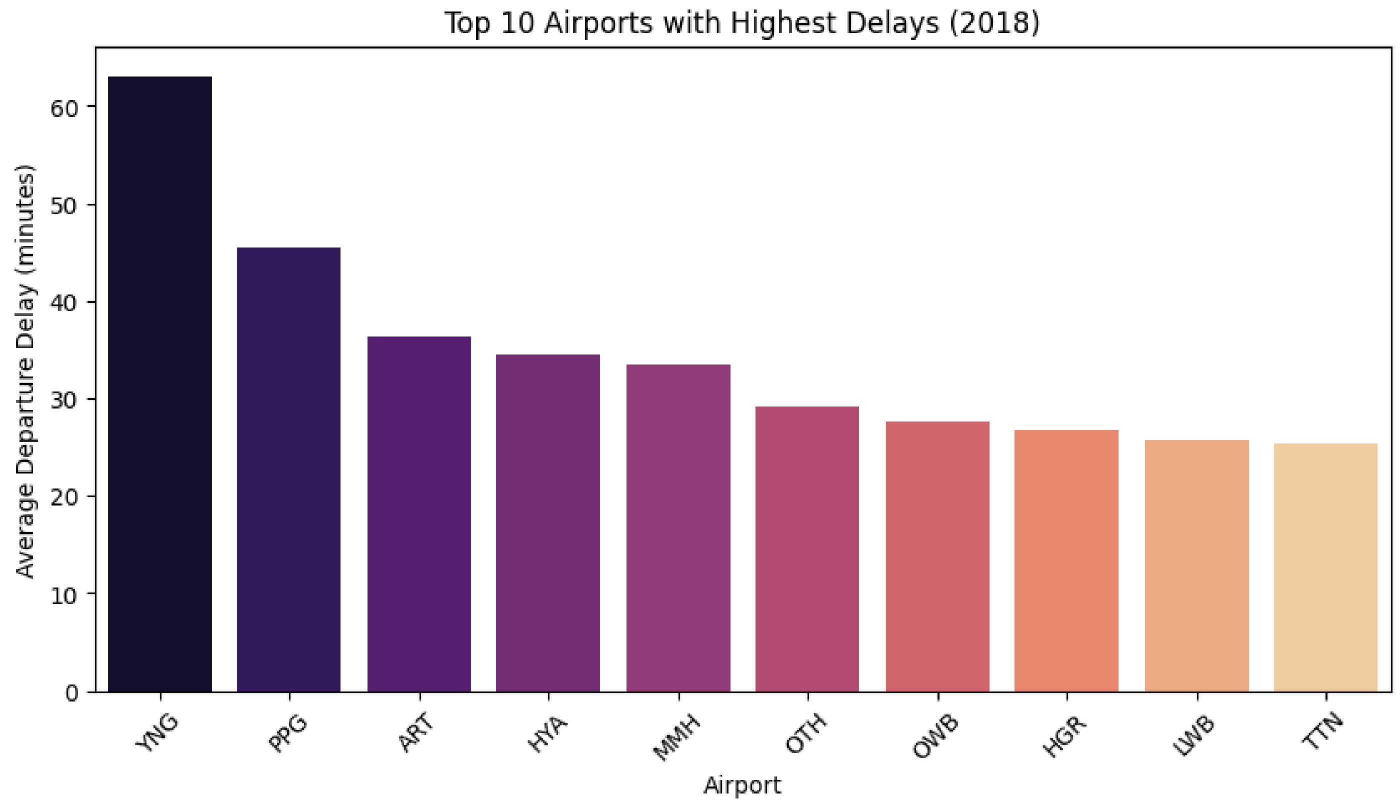
Categories	Feature Name	Sample Values	Feature Description
DATE & TIME	FL_DATE	2018-01-01	The Date of the Flight
	CRS_DEP_TIME	2147, 1050, 700	Schedule Departure Time (HHMM)
	DEP_TIME	2147, 1050, 700	Actual Departure Time (HHMM)
	CRS_ARR_TIME	2250, 1404, 757	Scheduled Arrival time (HHMM)
	ARR_TIME	2245., 1403., 813.	Actual Arrival time (HHMM)
FLIGHT DETAILS	OP_CARRIER	'NK', 'MQ', 'OO', 'EV', 'HA'	The Name of the Carrier
	OP_CARRIER_FL_NUM	195, 197, 198	Flight Number of the Carrier
	ORIGIN	'MCO', 'LGA', 'FLL', 'IAH'	Origin Airport
	DEST	'FLL', 'MCO', 'LAS', 'ORD'	Destination airport
	DISTANCE	177., 1076., 1222.	Distance between airports (miles)
TIME METRICS	TAXI_OUT	15., 20., 19., 8.	Taxi Out Time, in Minutes; The time elapsed between departure from the origin airport gate and wheels off.
	WHEELS_OFF	2158., 1124., 731.	Wheels Off Time (local time) in HHMM
	WHEELS_ON	2158., 1124., 731.	Wheels On Time (local time) in HHMM
	TAXI_IN	7., 9., 10., 4., 5.	Wheels down and arrival at the destination airport gate, in minutes
	CRS_ELAPSED_TIME	63., 194., 57., 196.	Estimated Elapsed Time of Flight, in Minutes
	ACTUAL_ELAPSED_TIME	63., 194., 57., 196.	Elapsed Time of Flight, in Minutes
	AIR_TIME	40., 150., 32., 164.	Flight time in Minutes
DELAY INFORMATION	DEP_DELAY	-4., 14., 12.	Difference in minutes between scheduled and actual departure time. Early departures show negative numbers.
	ARR_DELAY	-5.0, -1.0, 16.0	Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers.
	CANCELLED	0., 1.	Cancelled Flight Indicator (1=Yes); was the flight cancelled?
	CANCELLATION_CODE	'A', 'B', 'C', 'D'	Reason for cancellation (A = carrier, B = weather, C = NAS, D = security
	DIVERTED	0., 1.	Diverted Flight Indicator (1 = Yes)
	CARRIER_DELAY	1., 15., 127., 174.	Carrier Delay, in Minutes
	WEATHER_DELAY	31., 17., 24., 61.	Weather Delay, in Minutes
	NAS_DELAY	16., 18., 25., 19.	National Air System Delay, in Minutes

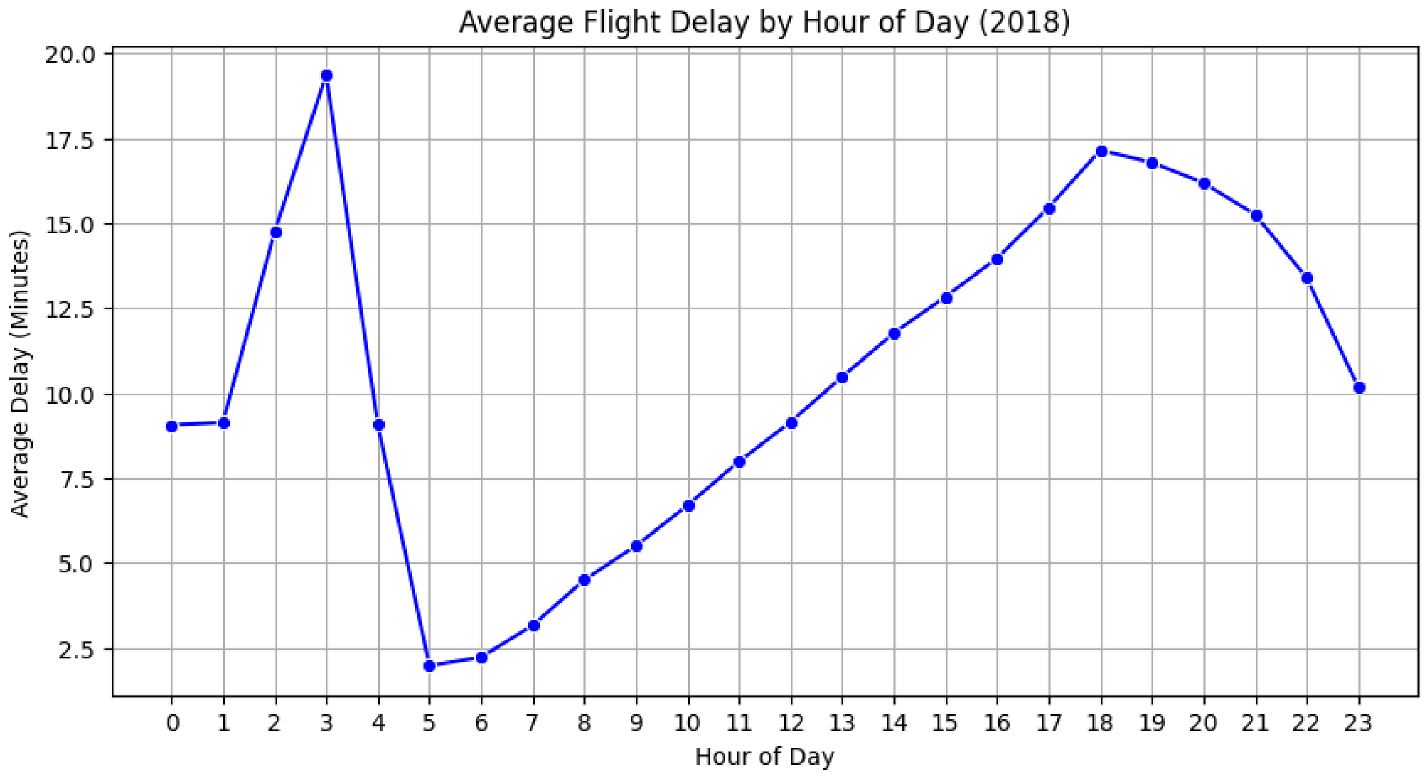
Main causes for flight delays and cancellations



Flight delays by airline, airport, and time of day







Scheduling strategies that airlines could adopt to reduce delays

Fleet and Aircraft Utilization Strategy Standardized fleets = fewer maintenance delays Overused aircraft = higher risk of cascading delays	**Crew Scheduling and Reserve Crew Availability** Sufficient reserve crew = faster recovery from staff shortages Poor crew planning = last-minute flight cancellations
Boarding and Turnaround Efficiency Faster boarding & deplaning = reduced departure delays Slow turnaround = cascading delays throughout the day	**Hub-and-Spoke vs. Point-to-Point** Point-to-Point = fewer connection-related delays Hub-and-Spoke = delays from congestion & missed connections

Predictive model to anticipate flight delays and cancellations

Logistic Regression

Logistic regression is a statistical method used to predict the probability of a binary outcome

K-Nearest Neighbour

K-Nearest Neighbors (KNN) predicts an outcome by comparing a data point to its closest neighbors based on similar characteristics

Random Forest

Random Forest is a algorithm that builds multiple decision trees and combines their predictions to make a more accurate and robust decision

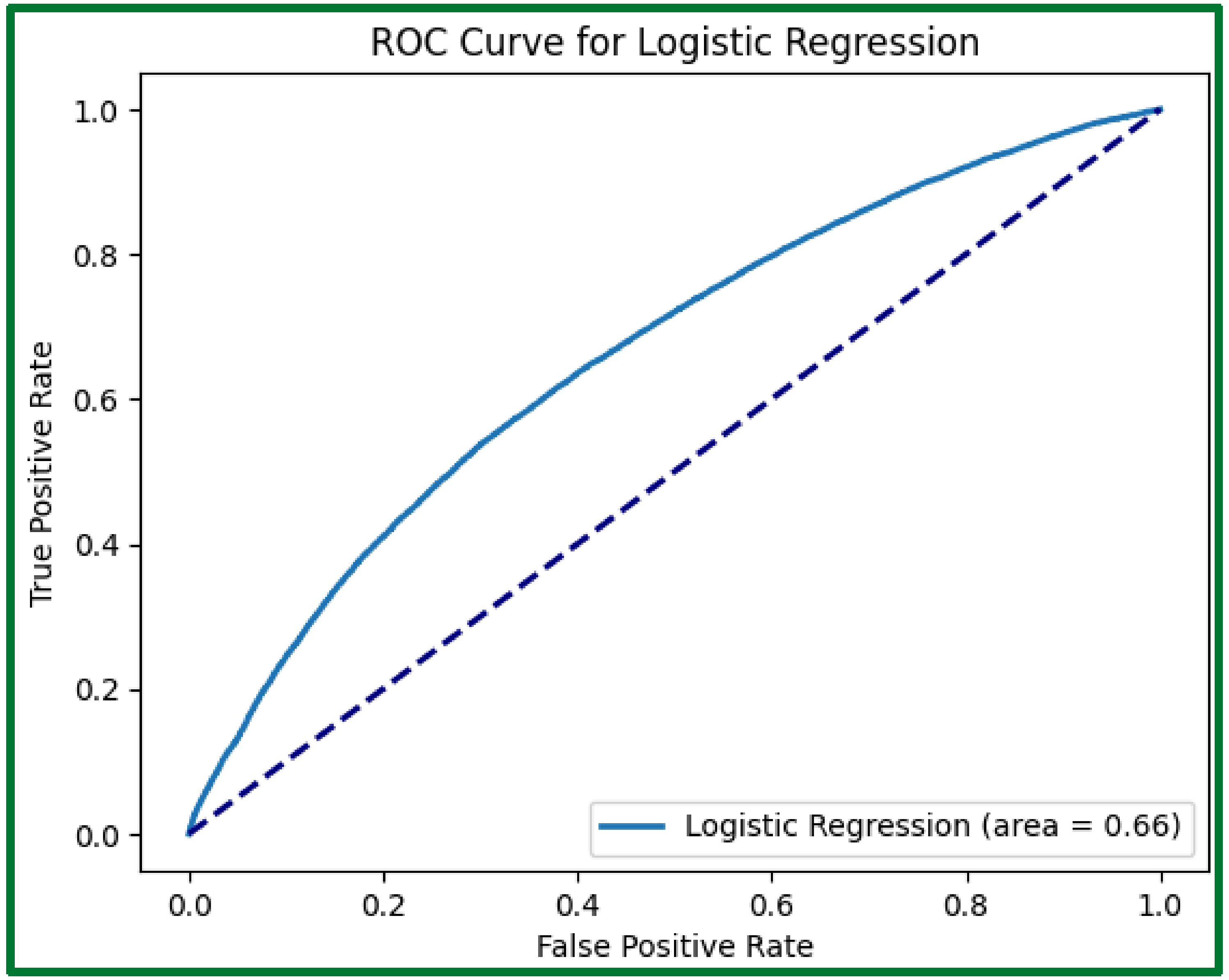
XG Boost

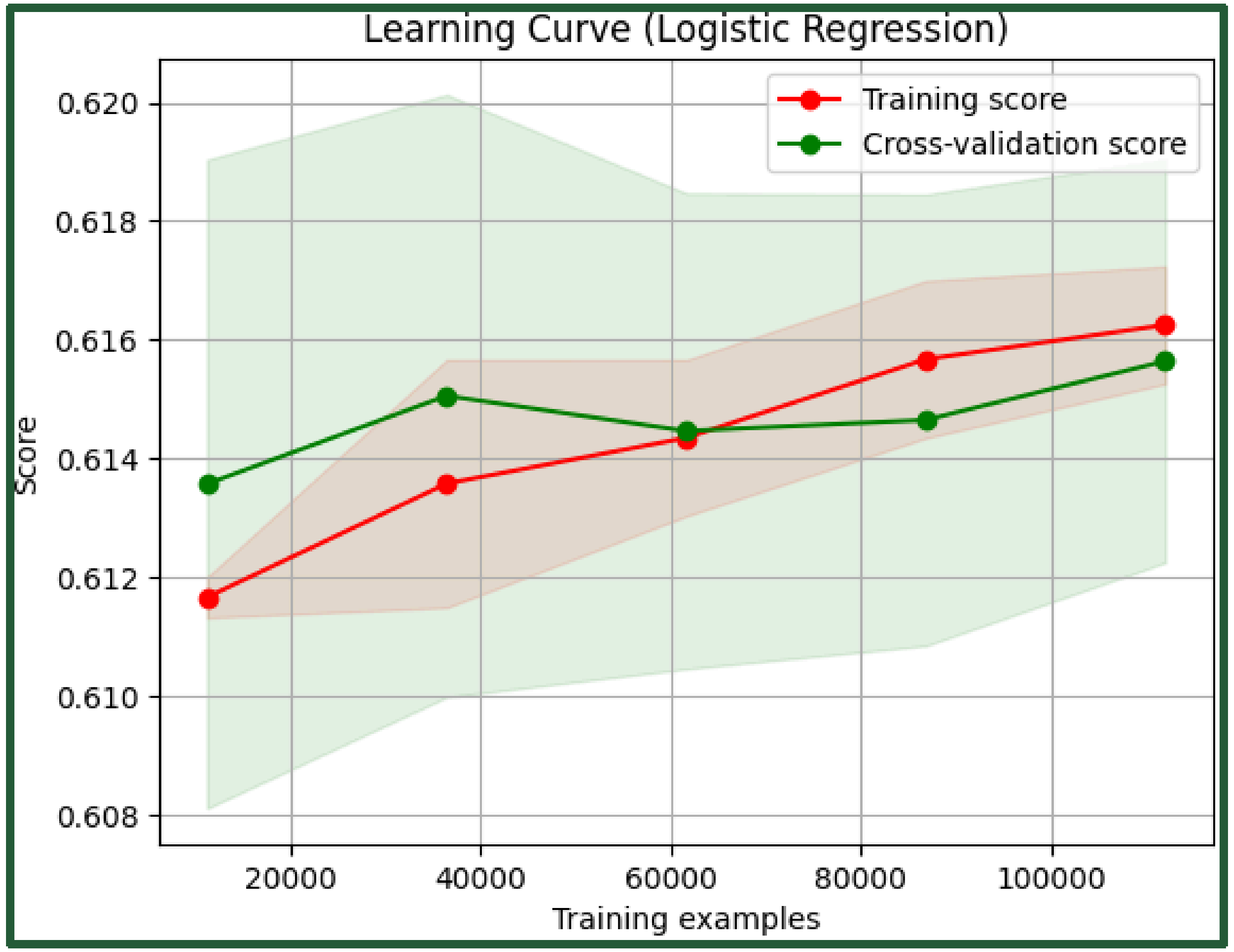
XGBoost is a fast and accurate algorithm that combines decision trees using gradient boosting to improve predictions.

Logistic Regression

- Area Under ROC Curve : 66%
- Accuracy: 62%

The model is a good fit.

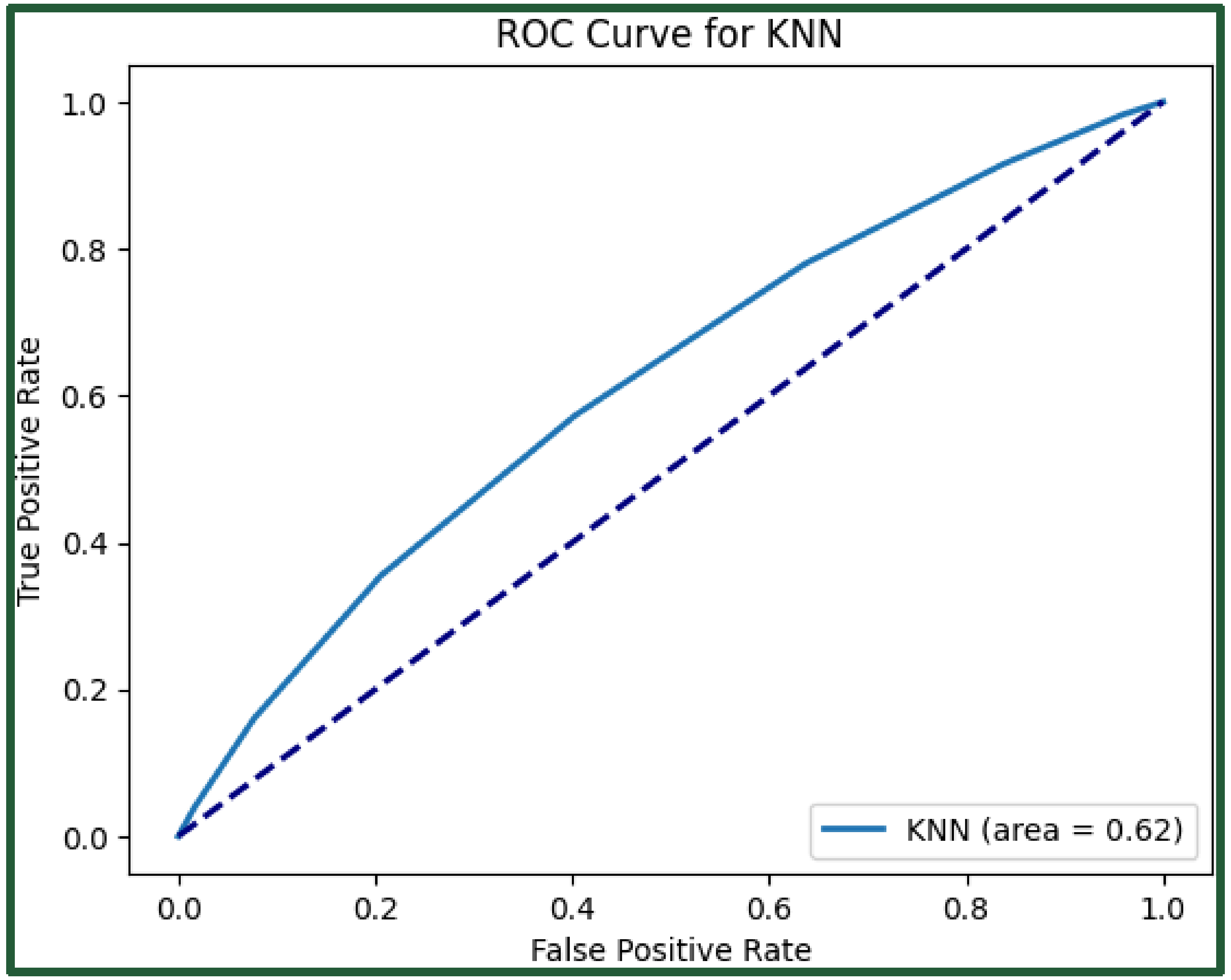


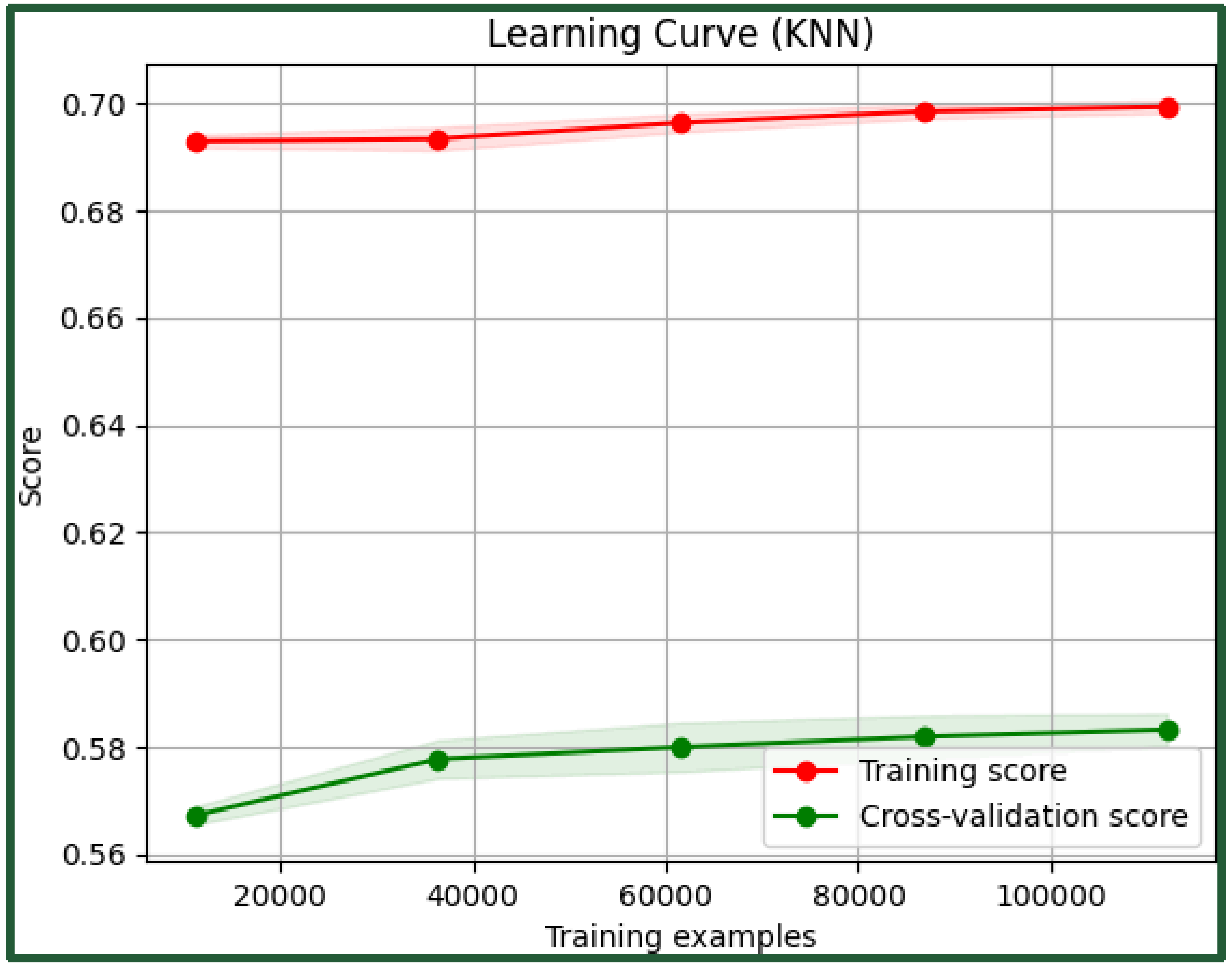


K-Nearest Neighbour

- Area Under ROC Curve : 62%
- Accuracy: 59%

The model is overfitted.

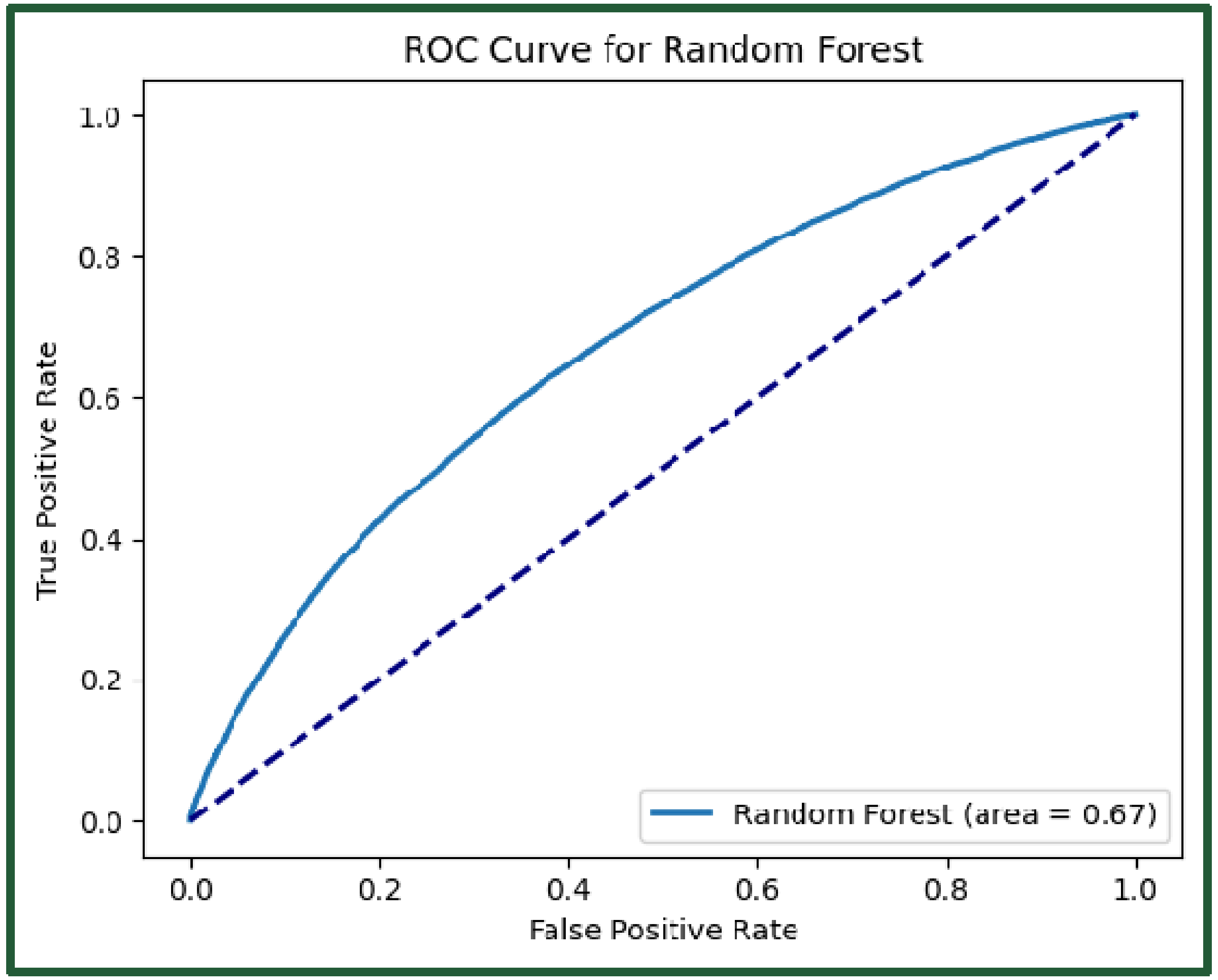


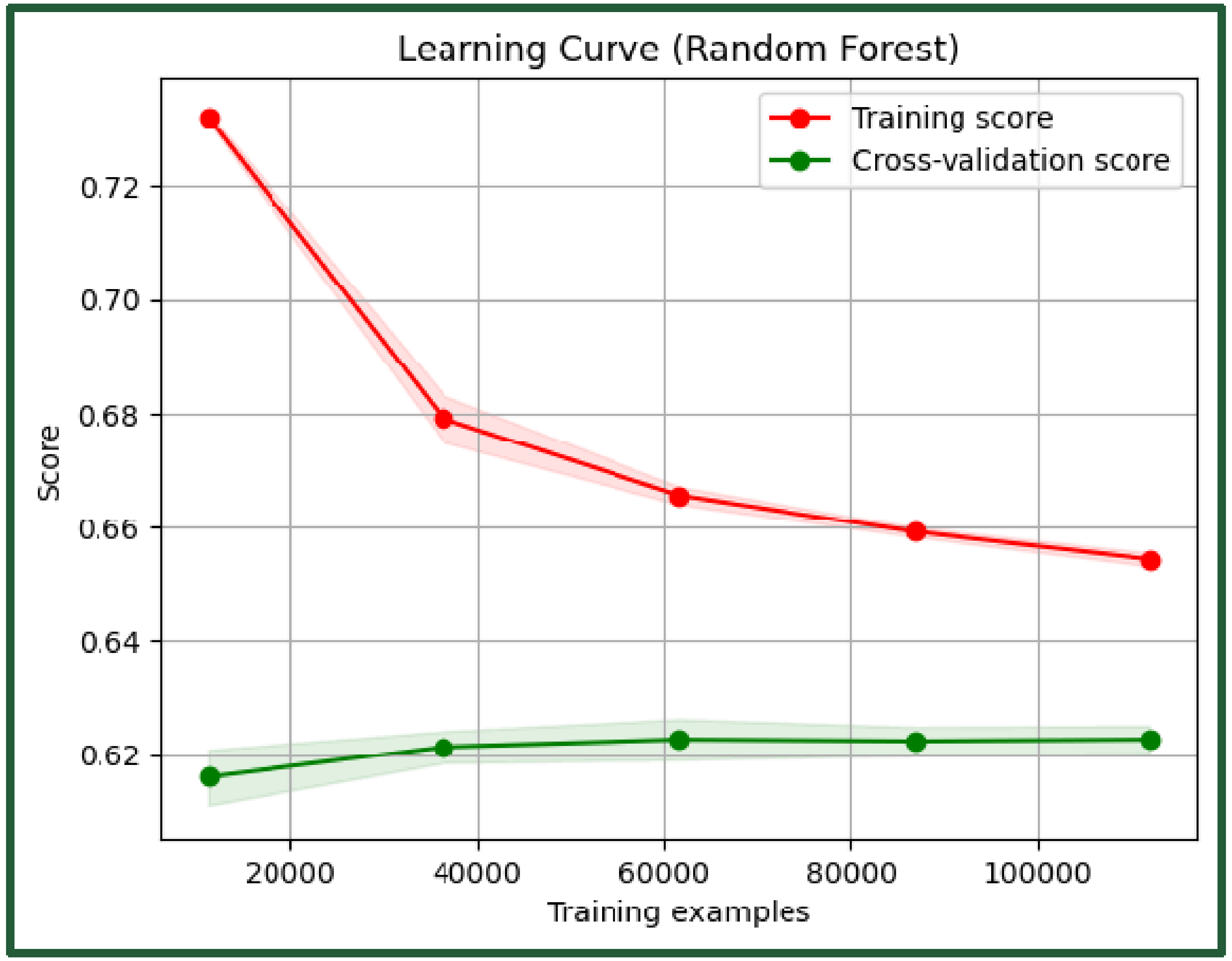


Random Forest

- Area Under ROC Curve : 67%
- Accuracy: 62%

The model is a good fit.

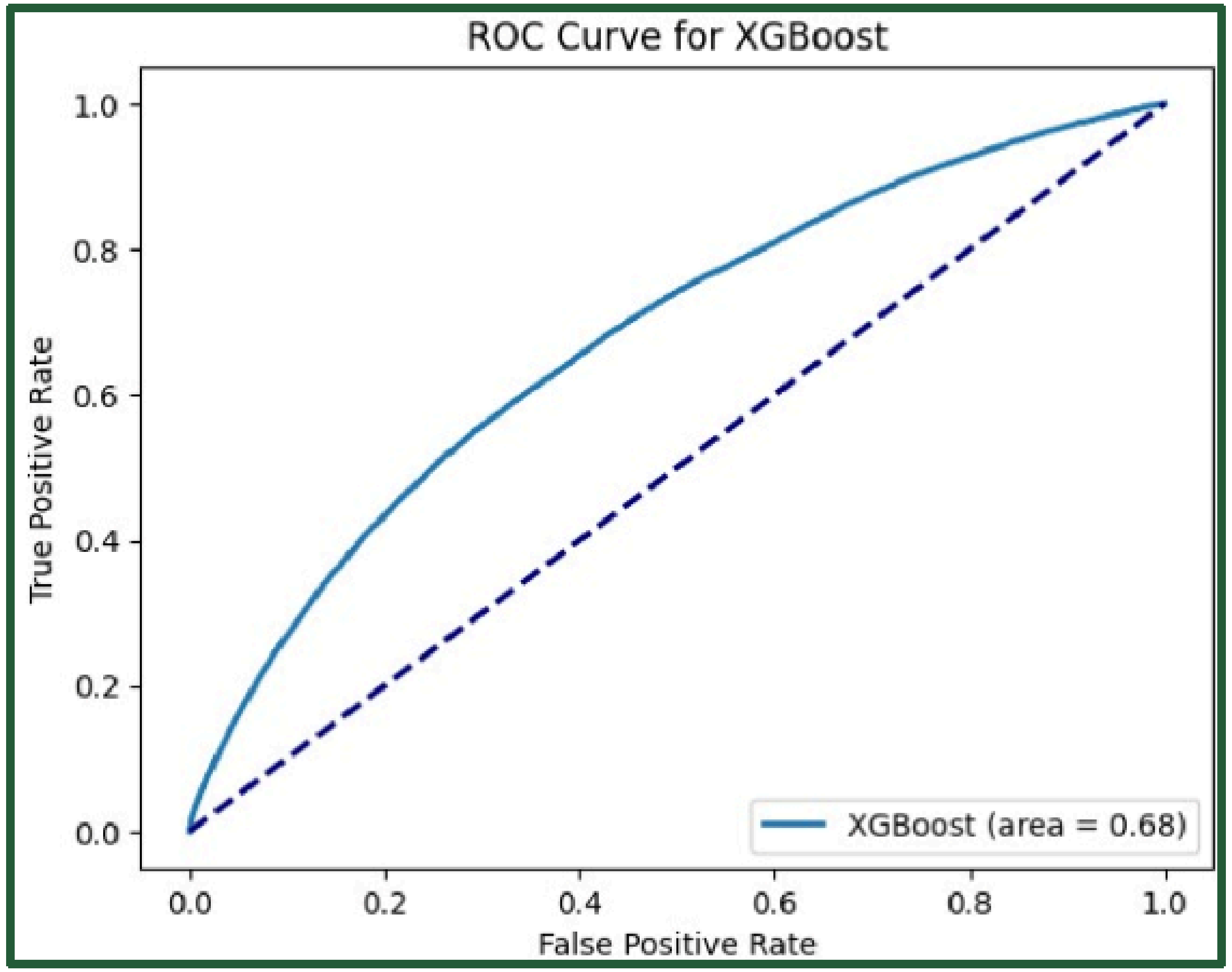


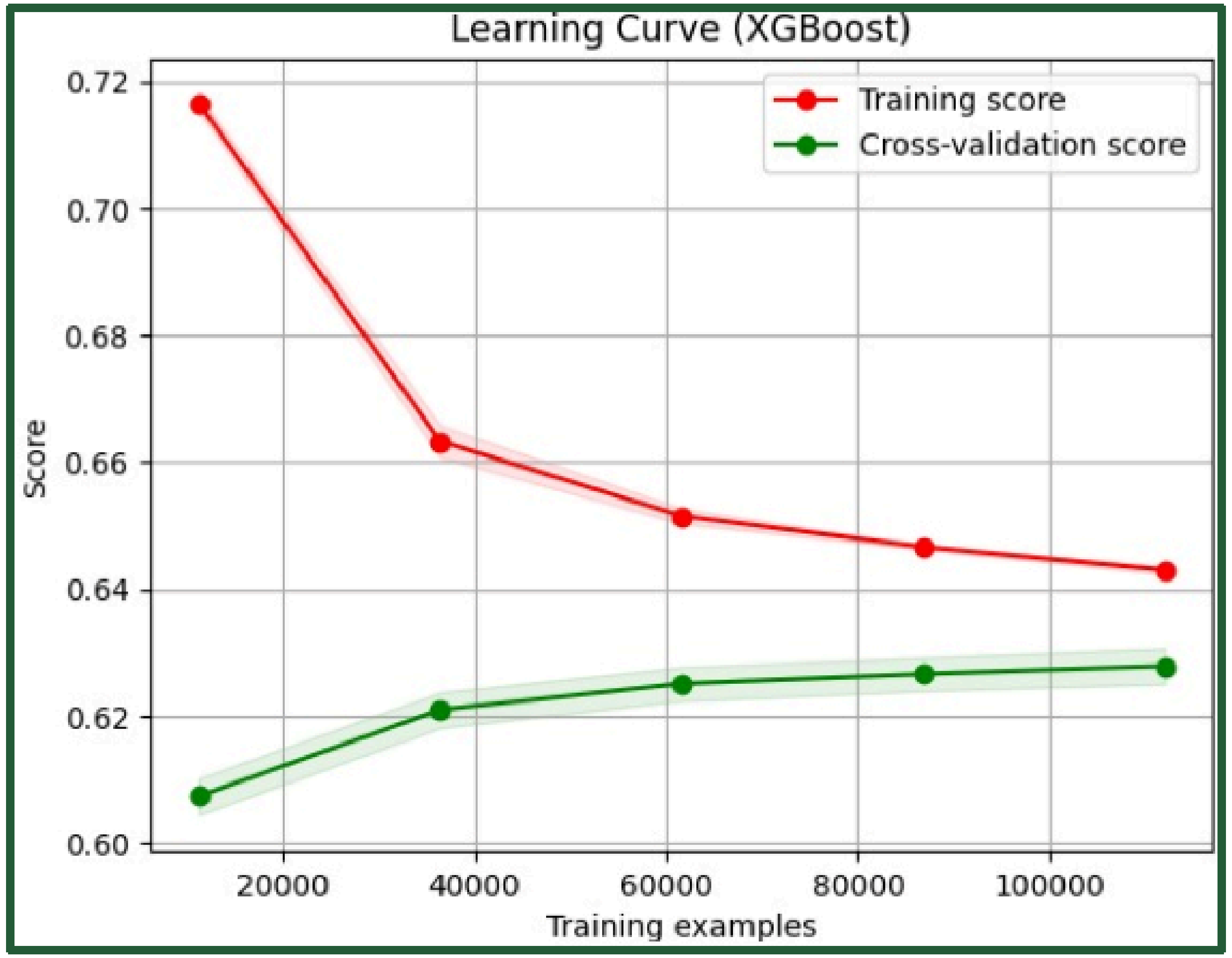


XG Boost

- Area Under ROC Curve : 68%
- Accuracy: 63%

The model is a good fit.





Analysis and Interpretation

XGBoost model achieves 68% accuracy, ensuring effective flight delay management for improved operational efficiency and passenger satisfaction

Strategic Planning: Optimizing Operations, Routes, and Resources for Enhanced Efficiency and Customer Satisfaction.

Implement targeted strategies for routes prone to delays and invest in preventive maintenance programs to enhance operational efficiency and passenger satisfaction

Using machine learning for predicting airline delays improves operational efficiency, punctuality, and customer satisfaction in the dynamic airline industry, highlighting the importance of continuous monitoring.

Thank you