Modeling Market Share Between Competing Adjacent Airports

Jessica Christa Wira Hadipoernomo

BA 612 – Data Analytics for Aviation Business
Instructor: Dr. Ahmed Abdelghany

April 25, 2025

Abstract

This study analyzed the factors influencing passenger market share between competing adjacent airports within shared origin or destination regions. Using a dataset of over 44,000 airport pair records from 2022–2023, we examine how variables such as fare ratio, flight frequency, airport capacity, and geographic distance impact market share outcomes. The dataset was segmented into four competitive scenarios, including cases where larger airports lead or lose market share to their smaller competitors. To evaluate these relationships, we implemented a combination of regression models (Ordinary Least Squares, Ridge, Lasso) and ensemble learning methods (Random Forest, XGBoost). The analysis reveals that the fare ratio, departure count difference, and capacity difference are the most consistent predictors in all scenarios. Moreover, larger airports do not always dominate —convenience and pricing often influence passenger decisions. The findings support data-driven decisions in route planning, pricing, and infrastructure, offering valuable insights into airport competition in multi-airport regions.

1 Introduction

In regions where multiple airports serve overlapping markets, airport competition becomes a significant factor influencing passenger behavior and industry strategy. This phenomenon is particularly evident in metropolitan areas or regional clusters with two or more airports within proximity that offer service to the same or similar origin-destination (O&D) markets. While it is commonly assumed that larger airports dominate due to their scale, connectivity, and resources, emerging evidence suggests that smaller airports may outperform their larger counterparts under certain conditions. This research investigates the determinants of passenger market share between pairs of adjacent airports to better understand the conditions under which smaller airports can attract a competitive share of travelers.

This topic is of growing importance to both academic researchers and industry stakeholders. For airport authorities, understanding what drives passengers to choose one airport over another enables more targeted investments in infrastructure, pricing strategies, and service improvements. For airlines, airport choice impacts route profitability, hub operations, and network planning. From a policy standpoint, better insights into airport competition support more effective regional planning and transport integration. Despite extensive literature on airport choice and competition, there is a lack of research that directly compares adjacent airports within the same market, especially through the lens of how airport characteristics interact to influence passenger distribution.

The primary objective of this study is to analyze the factors that drive passenger market share between competing airports serving the same origin or destination. Specifically, we aim to examine whether larger airports consistently maintain higher market share and to identify conditions in which smaller airports outperform larger ones. The research questions guiding this study are: (1) What are the most important factors influencing market share in adjacent airport pairs? (2) Does having greater size or capacity necessarily correlate with higher market share? (3) How do operational characteristics such as fare, flight frequency, and capacity interact with airport accessibility and market type (domestic vs. international) to shape outcomes?

To answer these questions, we use a dataset containing over 44,000 records of airport pairs with overlapping markets across 20 global regions, covering the period from 2022 to 2023. Each record compares two airports within the same market cluster based on attributes such as fare, capacity, departure frequency, and geographical distance. The dataset was cleaned, structured, and processed to ensure that one airport in each pair represents the "larger" airport based on size metrics. Feature engineering included the creation of variables such as fare ratio, departure count difference, and origin/destination size differences, as well as dummy variables for months, regions, and international market classification.

Using this enriched dataset, we conducted multiple regression analyses, including Ordinary Least Squares (OLS), Ridge, and Lasso regressions, to identify key predictors of market share. These results were then compared with ensemble methods like Random Forest and XGBoost to evaluate model performance and gain additional insights. Our findings indicate that factors such as fare ratio, departure frequency, and accessibility are strong predictors of market share. Notably, the results challenge the assumption that larger airports always dominate, showing that smaller airports can gain a competitive edge when they offer more favorable pricing or access.

By bridging existing research gaps and applying a combination of statistical and machine learning approaches, this study contributes to a more nuanced understanding of airport competition. The findings offer actionable insights for airport managers, airlines, and policymakers seeking to optimize operations and improve passenger satisfaction in multi-airport environments.

2 Literature Review

Understanding the dynamics of airport competition and the factors influencing passenger preferences is critical in the context of multi-airport regions. Numerous studies have emphasized that passengers choose between adjacent or competing airports based on a combination of operational performance, accessibility, and service quality. These preferences are not uniform and may vary significantly depending on the type of traveler, regional infrastructure, and market conditions.

Operational characteristics such as airfare levels, flight frequency, and schedule reliability consistently emerge as core determinants of airport preference [1, 2]. Airports that offer more frequent services or lower prices tend to capture greater market share, particularly among price-sensitive or time-constrained travelers. Additionally, passengers often consider the overall convenience and reliability of airport operations when making decisions, which directly ties into the airport's ability to attract and retain demand in competitive environments.

Beyond operations, accessibility to the airport plays a substantial role in influencing passenger choices. Studies have shown that shorter travel distances and stronger transportation links, including the availability of public transit or express road access, make certain airports more attractive to travelers [3]. Even when two airports serve the same city or region, those with better integration into the local transportation network typically enjoy a competitive edge. This highlights the importance of regional planning and investment in supporting airport competitiveness.

Another critical dimension influencing airport competition is service quality. Research has demonstrated that the overall passenger experience—including terminal amenities, ease of navigation, and the comfort of waiting areas—can significantly affect traveler satisfaction and loyalty [4, 5]. High service quality not only enhances the perception of an airport but also encourages repeat usage, further reinforcing market share advantages. In multi-airport regions, where passengers often have multiple options, airports that invest in superior passenger experiences are better positioned to retain and grow their user base.

The literature also explores how infrastructure and airport size affect competitiveness. Airports with greater capacity and more efficient operations generally perform better in multi-airport systems [6]. Adequate infrastructure enables these airports to offer more services, reduce congestion, and maintain on-time performance, all of which contribute to improved passenger satisfaction. However, this advantage can be diminished if the airport lacks accessibility. Some studies suggest that even larger airports may lose market share if smaller alternatives are more conveniently located or offer comparable services [7, 8, 9]. In this context, strategic infrastructure investments—including those targeting sustainability—are gaining importance. Airports that adopt net-zero energy building practices and prioritize environmental consciousness are increasingly appealing to both passengers and airline partners [10].

To understand and quantify airport competition, various analytical methods have been employed in the literature. The Hotelling model has been used to study spatial competition among airports and to simulate how market share shifts with changes in flight frequency and location [11]. Logistic regression models and discrete choice models, such as multinomial or nested logit, are widely used to estimate the likelihood of passengers choosing one airport over another based on characteristics like access time, flight availability, or pricing [12, 13]. These models capture the probabilistic nature of airport selection and help identify the marginal effects of key variables.

Simulation-based approaches have also been developed to assess the impact of intermodal connectivity on airport competitiveness. For example, integrating high-speed rail into existing airport networks has been shown to reshape market boundaries and shift passenger flows [14]. Meanwhile, other studies explore passenger preferences for specific features, such as additional legroom or upgraded services, to assess the elasticity of demand [15].

Productivity and efficiency analyses further enhance the understanding of airport competitiveness. Techniques like Data Envelopment Analysis (DEA) and the Malmquist index have been used to measure how well airports utilize their resources relative to market performance. Findings suggest a strong link between airport size, market share, and operational efficiency [16], reinforcing the value of data-driven evaluation tools in strategic planning.

Despite the breadth of research on airport competition, several gaps remain. Many studies tend to analyze airports individually or in isolation rather than focusing on direct competition between airport pairs. Moreover, behavioral factors such as passenger loyalty programs, ground access time variability, or demographic influences are often excluded due to data limitations. Limited attention has also been given to how these factors operate simultaneously in competitive multi-airport systems, where the dynamics can be more complex. While discrete choice models offer valuable insights, they may not fully capture the temporal fluctuations or multivariate interactions that drive market share differences over time.

This study addresses these gaps by analyzing adjacent airport pairs serving the same origin or destination region using real-world data that incorporates market share, fare, capacity, and distance measures. Unlike earlier research, it explicitly compares larger versus smaller airports within pairs, enabling a nuanced understanding of competitive positioning. By applying both traditional regression models and advanced machine learning techniques, the study contributes to the methodological literature as well, offering insights into how model complexity can improve predictive accuracy and feature interpretability.

3 Data Collection & Preprocessing

The dataset used in this study was provided by Dr. Ahmed Abdelghany. It contains monthly observations (averaged) of competing airports within shared origin or destination clusters. The original dataset includes combinations of origin-destination (O&D) records with up to 30 or more competing airports in a single cluster. For the purpose of this study, only records with exactly two competing airports were selected, allowing for pairwise comparisons and clearer interpretation of competitive dynamics. To ensure consistency, each airport pair was reordered such that the larger airport—based on average daily passenger volume—was listed first.

The filtered dataset was further segmented into two main categories: records in which the competing airports shared the same origin and those that shared the same destination. To ensure meaningful comparisons, any observations where the larger airport's market share was 100% (i.e., cases with no observable competition) were removed. Additionally, the data was further split into subgroups where the larger airport has either a larger or smaller market share. The sample distribution of each scenario is presented in Table 1.

Table 1: Market Segmentation Scenarios and Dataset Sizes

ID	Scenario	Size of subset
1	Same Destination (All)	21,672 (48.65%)
2	Same Destination – Larger Airport with Larger Share	15,969 (35.85%)
3	Same Destination – Larger Airport with Smaller Share	5,703 (12.80%)
4	Same Origin (All)	21,672 (48.65%)
5	Same Origin – Larger Airport with Larger Share	15,952 (35.81%)
6	Same Origin – Larger Airport with Smaller Share	5,720 (12.84%)

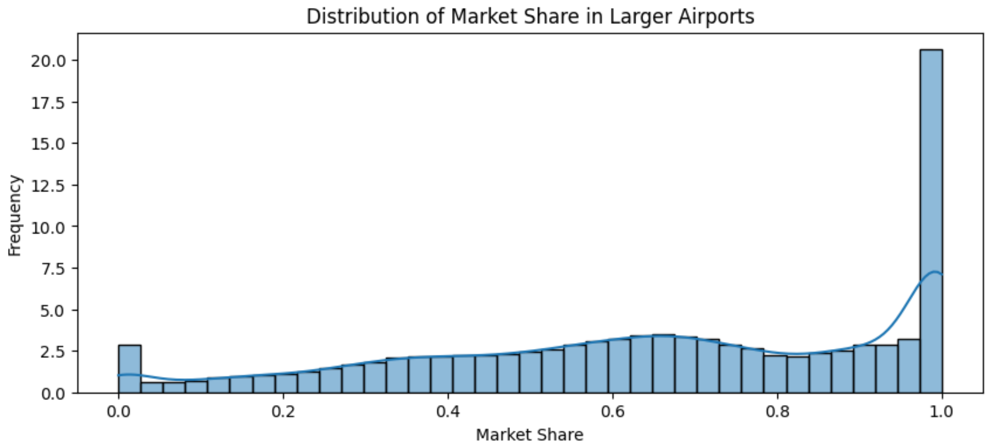

Figure 1: Distribution of Market Share in Larger Airports for All Records

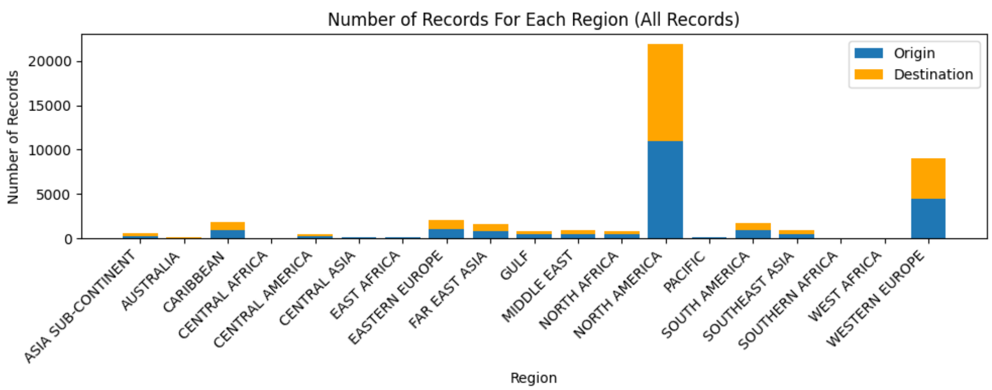

Figure 2: Number of Records by Region (Origin and Destination)

Figure 1 shows the distribution of market shares for the larger airports, highlighting the spread of competitiveness across markets.

To understand the spatial distribution of the records, we plotted the frequency of records by region. As shown in Figure 2, North America dominates the dataset due to the high density of competing airports in the U.S. air travel network. Other well-represented regions include Western Europe and Southeast Asia.

To encode the competitive relationship between each airport pair, several key variables were constructed as differences or ratios between the attributes of the two competing airports. These transformations ensure that the predictors directly reflect the relative performance or characteristics of these competing airports. For example, `fare_ratio` is calculated as the average fare from Airport 1 divided by that from Airport 2, providing a normalized indicator of fare competitiveness. Similarly, `depcount_diff`, `capacity_diff`, `origin_size_diff`, and `destination_size_diff` capture differences in departures, available seats, and airport sizes between the two competitors.

In addition to the engineered features, categorical variables were incorporated to capture seasonality and geographic variation in airport competition. A series of dummy variables was created for each month (`month_January` to `month_December`) to control for potential seasonal effects such as holidays, school breaks, or weather-related travel patterns. Similarly, the regions associated with each airport—both origin and destination—were encoded as one-hot vectors using region-specific dummy variables (e.g., `origin_region_SOUTHEAST ASIA`, `destination_region_NORTH AMERICA`). These regional indicators help account for differences in demand that can vary significantly between geographic areas.

The final variables used for the models are shown in Table 2. A total of more than 50 predictor (in-

Table 2: List of Variables Used in the Models

Variables	Description
origin_distance	Distance between origin airports in (Nautical miles)
destination_distance	Distance between destination airports (Nautical miles)
fare_ratio	Ratio of average fares between larger and smaller airports
depcount_diff	Difference in average daily flight departures
capacity_diff	Difference in average daily seat capacity of operating airlines
origin_size_diff	Difference in origin airport size (average total daily passenger volume)
destination_size_diff	Difference in destination airport size (average total daily passenger volume)
international_market	Dummy variable for international market
1_percentage_Passengers	Market share of larger airport for the route **(Dependent Variable)**
month_Month	Dummy variable for Month
origin_region_{REGION}	Dummy variable for origin airport region
destination_region_{REGION}	Dummy variable for destination airport region

dependent variables) were used in the final regression models, which included operational, geographic, and categorical information.

4 Methodology

The primary objective of this study is to identify the factors that influence airport-level passenger market share in overlapping origin-destination (O&D) markets. Because the dependent variable—`1_percentage_Passengers`—is a continuous variable representing the proportion of passengers choosing one airport over another, this research uses a regression-based modeling approach.

Five supervised learning algorithms were selected to support both interpretability and predictive performance. These include:

- Ordinary Least Squares (OLS) Regression
- Lasso Regression
- Ridge Regression
- Random Forest Regressor
- XGBoost Regressor

This combination of models was chosen to allow both interpretability (OLS, Lasso) and predictive flexibility (Random Forest, XGBoost). Time series models were not implemented as the data was aggregated monthly and did not show individual airport behavior over time.

Each scenario-specific dataset, as well as the full dataset, was split into a training set (70%) and a testing set (30%) using random sampling. More than 50 predictor variables were used in the model. For OLS and Ridge regression, all features were retained. For Lasso regression, automatic feature selection was performed using L1 regularization, resulting in the removal of less-influential variables. For Random Forest and XGBoost, all features were included.

All models were evaluated using the following metrics:

- **R-squared (R^2)**: Measures the proportion of variance in market share explained by the model.
- **Root Mean Squared Error (RMSE)**: Assesses average prediction error, with stronger penalties for large deviations.

These metrics enable comparison across models and help determine the trade-off between explanatory power and predictive accuracy. The inclusion of both linear and nonlinear models provides a comprehensive view of the underlying patterns in airport competition dynamics.

5 Results

This section presents the outcomes of each model applied to the full dataset and the four market scenarios defined earlier. Model performance is evaluated using R^2 and RMSE, while the top contributing variables are analyzed based on regression coefficients (for linear models) and feature importance scores (for ensemble methods).

5.1 Ordinary Least Squares (OLS)

Table 3 shows the coefficients of the OLS model across all six datasets. Key variables such as `fare_ratio`, `depcount_diff`, and regional dummies (e.g., `origin_region_SOUTHEAST ASIA`) appear consistently across scenarios. While OLS provides interpretable estimates, it performs moderately overall, with R^2 values ranging from 0.22 to 0.33. The highest performance was observed in the scenarios 1 and 5.

5.2 Ridge Regression

Ridge regression (Table 4) demonstrates similar predictive patterns to OLS, with slightly improved model stability and reduced variance in coefficient estimates. Regularization reduces the impact of multicollinearity among region and month dummies while retaining explanatory power. Performance was particularly stable across scenarios 4, 5, and 6, reaching $R^2 = 0.30$.

5.3 Lasso Regression

Table 5 illustrates Lasso's selective nature in retaining only the most influential variables. In "Same Destination – Smaller Share," Lasso delivered its best performance with an R^2 of 0.30, suggesting that price sensitivity and targeted regional effects dominate in niche markets.

5.4 Random Forest Regression

As shown in Table 6, Random Forest outperformed all linear models in terms of predictive accuracy, achieving R^2 values above 0.84 in multiple scenarios. It successfully captured nonlinear interactions between features such as `fare_ratio`, `capacity_diff`, and regional dummies. The model performed best in the "Same Destination (All)" and "Same Origin (Larger Share)" scenarios, demonstrating its effectiveness in explaining dominant market behaviors.

5.5 XGBoost Regression

XGBoost (Table 7) also demonstrated strong performance, particularly in markets where the larger airport had a smaller share. This may indicate that boosted trees capture subtle pricing and regional effects that linear models miss. Important variables in this model included `fare_ratio`, `origin_distance`, and various regional dummies, with R^2 values reaching up to 0.70.

5.6 Key Predictors

Across all models, the most influential predictors of passenger market share included:

- **Fare Ratio:** Passengers favor airports offering more competitive fares.
- **Departure Count:** Higher flight frequency attracts more passengers.
- **Capacity Difference:** Larger seat availability correlates with greater share.
- **Distance:** Proximity increases passenger preference for an airport.
- **International Indicator:** Larger airports tend to dominate in international markets.

These findings were consistent across both linear and nonlinear models.

Table 3: OLS Regression Coefficients Across Scenarios

	Feature	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
0	const	0.667083	0.782198	0.337222	0.705305	0.777303	0.358002
1	capacity_diff	0.000053	0.000010	0.000031	0.000051	0.000006	0.000022
2	depcount_diff	0.023148	0.012642	0.020432	0.024176	0.013815	0.020281
3	destination_distance	NaN	NaN	NaN	-0.000605	-0.000786	-0.001319
4	destination_region_AUSTRALIA	-0.063115	-0.071464	0.056802	-0.089127	-0.066982	0.049613
5	destination_region_CARIBBEAN	-0.050411	0.056426	0.063264	-0.228257	-0.127797	-0.161499
6	destination_region_CENTRAL AFRICA	-0.361974	NaN	-0.178844	-0.169333	-0.262650	0.031898
7	destination_region_CENTRAL AMERICA	-0.113909	-0.034634	0.029194	0.050328	-0.007379	-0.071572
8	destination_region_CENTRAL ASIA	0.298326	0.161247	NaN	NaN	NaN	NaN
9	destination_region_EAST AFRICA	-0.165832	-0.193122	-0.018253	0.032808	0.009591	0.132818
10	destination_region_EASTERN EUROPE	-0.019690	0.071976	-0.014839	-0.262803	-0.168830	-0.088700
11	destination_region_FAR EAST ASIA	0.078542	0.088422	0.080824	-0.204292	-0.097558	-0.175518
12	destination_region_GULF	0.092449	0.044357	0.151160	-0.053502	-0.056303	-0.064341
13	destination_region_MIDDLE EAST	0.024435	0.029925	0.070111	NaN	NaN	NaN
14	destination_region_NORTH AFRICA	0.044116	0.070368	0.144479	-0.064371	-0.002282	-0.166532
15	destination_region_NORTH AMERICA	-0.040231	0.027676	0.025652	-0.126959	-0.107228	-0.065950
16	destination_region_PACIFIC	0.037816	-0.007669	0.000000	0.027213	-0.010929	0.000000
17	destination_region_SOUTH AMERICA	-0.003915	-0.004669	0.080290	-0.119768	-0.013329	-0.061912
18	destination_region_SOUTHEAST ASIA	0.004736	0.019655	0.226457	-0.221830	-0.145258	-0.201773
19	destination_region_SOUTHERN AFRICA	-0.220914	-0.167337	0.121419	0.123876	0.096286	-0.017539
20	destination_region_WEST AFRICA	-0.534267	-0.154856	-0.176709	NaN	NaN	NaN
21	destination_region_WESTERN EUROPE	-0.014511	0.055918	0.077264	-0.145574	-0.100510	-0.118511
22	destination_size_diff	NaN	NaN	NaN	0.000001	0.000001	0.000001
23	fare_ratio	-0.026888	0.004115	-0.013368	-0.026233	-0.000101	-0.016608
24	international_market	0.020568	0.034207	-0.030923	0.019533	0.033150	-0.046732
25	month_April	0.003694	0.002213	0.005664	0.012361	-0.000426	0.000645
26	month_August	0.002494	-0.009727	0.020717	0.006687	-0.013418	0.024840
27	month_December	0.015704	-0.006563	0.016702	0.016223	-0.010642	0.012494
28	month_February	0.012514	0.002631	-0.002847	0.014916	-0.000785	-0.007971
29	month_July	-0.010735	-0.009282	0.012567	-0.015435	-0.014422	0.002661
30	month_June	0.008890	-0.011844	0.024366	0.012230	-0.014702	0.021363
31	month_March	0.022616	0.002453	0.018796	0.025214	-0.002238	0.019835
32	month_May	-0.003554	0.001693	0.002964	0.004562	-0.008343	-0.003133
33	month_November	-0.000064	-0.014316	-0.000450	0.014433	-0.012162	0.004583
34	month_October	0.008092	-0.006795	0.010681	0.008189	-0.010061	-0.007474
35	month_September	0.004463	-0.013670	0.023965	0.013675	-0.013850	0.009952
36	origin_distance	-0.000428	-0.000839	-0.001237	NaN	NaN	NaN
37	origin_region_AUSTRALIA	-0.063115	-0.071464	0.056802	-0.089127	-0.066982	0.049613
38	origin_region_CARIBBEAN	-0.182590	-0.160642	-0.122386	-0.050716	0.019292	0.088589
39	origin_region_CENTRAL AFRICA	-0.142065	-0.312591	0.032997	-0.380406	NaN	-0.164056
40	origin_region_CENTRAL AMERICA	0.070540	-0.029094	-0.068111	-0.144541	-0.067841	0.030030
41	origin_region_CENTRAL ASIA	NaN	NaN	NaN	0.282755	0.159955	NaN
42	origin_region_EAST AFRICA	-0.016259	-0.078903	0.170551	-0.240119	-0.262325	0.031785
43	origin_region_EASTERN EUROPE	-0.240815	-0.209257	-0.096005	-0.040445	0.029363	0.001090
44	origin_region_FAR EAST ASIA	-0.146502	-0.137448	-0.000506	0.087065	0.063983	0.155661
45	origin_region_GULF	-0.042926	-0.117222	-0.061509	0.050296	0.040981	0.176142
46	origin_region_MIDDLE EAST	NaN	NaN	NaN	0.007066	-0.025617	0.087814
47	origin_region_NORTH AFRICA	-0.183631	-0.033342	-0.187588	0.020416	0.017535	0.154854
48	origin_region_NORTH AMERICA	-0.095275	-0.155246	-0.053791	-0.051650	-0.013201	0.032333
49	origin_region_PACIFIC	0.037816	-0.007669	-0.000000	0.027213	-0.010929	0.000000
50	origin_region_SOUTH AMERICA	-0.078565	-0.052388	-0.070963	-0.020156	-0.039056	0.063261
51	origin_region_SOUTHEAST ASIA	-0.201498	-0.210495	-0.168322	-0.001321	-0.022235	0.270192
52	origin_region_SOUTHERN AFRICA	0.061545	0.028188	-0.067385	-0.460471	-0.342448	0.083495
53	origin_region_WESTERN EUROPE	-0.130260	-0.145216	-0.118482	-0.032730	0.020891	0.079090
54	origin_size_diff	0.000001	0.000001	0.000001	NaN	NaN	NaN
	R-squared (Testing Set)	0.3287	0.2219	0.2888	0.3016	0.2203	0.2265
	RMSE (Testing Set)	0.2180	0.1450	0.1259	0.2197	0.1479	0.1321

Table 4: Ridge Regression Coefficients Across Scenarios

	Feature	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
0	capacity_diff	0.000053	0.000010	0.000031	0.000051	0.000006	0.000025
1	depcount_diff	0.023154	0.012656	0.020326	0.024176	0.013815	0.019929
2	destination_distance	NaN	NaN	NaN	-0.000604	-0.000785	-0.001328
3	destination_region_AUSTRALIA	-0.056985	-0.066451	0.054346	-0.088867	-0.065978	0.043912
4	destination_region_CARIBBEAN	-0.048857	0.052146	0.021347	-0.227833	-0.125598	-0.087896
5	destination_region_CENTRAL AFRICA	-0.334692	NaN	-0.168900	-0.169146	-0.258987	0.044501
6	destination_region_CENTRAL AMERICA	-0.111608	-0.038455	-0.012538	0.050659	-0.005531	-0.001942
7	destination_region_CENTRAL ASIA	0.287260	0.153262	NaN	NaN	NaN	NaN
8	destination_region_EAST AFRICA	-0.146158	-0.188292	0.010209	0.031816	0.007103	0.083686
9	destination_region_EASTERN EUROPE	-0.017574	0.068449	-0.053161	-0.262434	-0.166922	-0.018497
10	destination_region_FAR EAST ASIA	0.078642	0.084201	0.030655	-0.203848	-0.095534	-0.042253
11	destination_region_GULF	0.094532	0.042023	0.110177	-0.053192	-0.054759	-0.005974
12	destination_region_MIDDLE EAST	0.026176	0.026240	0.029471	NaN	NaN	NaN
13	destination_region_NORTH AFRICA	0.046140	0.066014	0.106193	-0.063991	-0.000450	-0.084192
14	destination_region_NORTH AMERICA	-0.038607	0.023727	-0.016127	-0.126586	-0.105336	0.004668
15	destination_region_PACIFIC	0.043051	-0.003486	0.000000	0.027431	-0.010112	0.000000
16	destination_region_SOUTH AMERICA	-0.002344	-0.008737	0.034486	-0.119376	-0.011476	0.007052
17	destination_region_SOUTHEAST ASIA	0.006036	0.017386	0.171467	-0.221424	-0.143445	-0.070812
18	destination_region_SOUTHERN AFRICA	-0.182751	-0.147345	0.079061	0.121979	0.090197	0.003927
19	destination_region_WEST AFRICA	-0.509733	-0.108274	-0.202736	NaN	NaN	NaN
20	destination_region_WESTERN EUROPE	-0.012699	0.051992	0.037199	-0.145210	-0.098631	-0.048331
21	destination_size_diff	NaN	NaN	NaN	0.000001	0.000001	0.000001
22	fare_ratio	-0.026817	0.004264	-0.013203	-0.026231	-0.000082	-0.016528
23	international_market	0.020467	0.034156	-0.030836	0.019534	0.033134	-0.045980
24	month_April	0.003648	0.002196	0.005484	0.012362	-0.000397	-0.000064
25	month_August	0.002414	-0.009715	0.020546	0.006685	-0.013384	0.024172
26	month_December	0.015701	-0.006581	0.016910	0.016221	-0.010624	0.012496
27	month_February	0.012454	0.002631	-0.002935	0.014914	-0.000770	-0.008793
28	month_July	-0.010812	-0.009228	0.012642	-0.015435	-0.014385	0.002533
29	month_June	0.008807	-0.011868	0.024078	0.012228	-0.014664	0.020514
30	month_March	0.022561	0.002439	0.018321	0.025213	-0.002216	0.019008
31	month_May	-0.003563	0.001720	0.002711	0.004564	-0.008310	-0.003648
32	month_November	-0.000153	-0.014388	-0.000293	0.014434	-0.012136	0.004271
33	month_October	0.008070	-0.006818	0.010479	0.008189	-0.010027	-0.007983
34	month_September	0.004446	-0.013736	0.023860	0.013674	-0.013814	0.009780
35	origin_distance	-0.000428	-0.000837	-0.001243	NaN	NaN	NaN
36	origin_region_AUSTRALIA	-0.056985	-0.066451	0.054346	-0.088867	-0.065978	0.043912
37	origin_region_CARIBBEAN	-0.172000	-0.146230	-0.079977	-0.050645	0.018975	0.012252
38	origin_region_CENTRAL AFRICA	-0.136792	-0.294853	0.034218	-0.379625	NaN	-0.152864
39	origin_region_CENTRAL AMERICA	0.078256	-0.016558	-0.018440	-0.144440	-0.068029	-0.043811
40	origin_region_CENTRAL ASIA	NaN	NaN	NaN	0.282342	0.158579	NaN
41	origin_region_EAST AFRICA	-0.023913	-0.072271	0.124501	-0.238732	-0.258292	0.040716
42	origin_region_EASTERN EUROPE	-0.231180	-0.196851	-0.056728	-0.040340	0.029211	-0.070817
43	origin_region_FAR EAST ASIA	-0.135224	-0.124723	0.048428	0.087083	0.063603	0.020840
44	origin_region_GULF	-0.033681	-0.105874	-0.034464	0.050410	0.040836	0.095004
45	origin_region_MIDDLE EAST	NaN	NaN	NaN	0.007159	-0.025777	0.006257
46	origin_region_NORTH AFRICA	-0.172784	-0.021277	-0.143268	0.020513	0.017201	0.082710
47	origin_region_NORTH AMERICA	-0.085729	-0.142885	-0.012511	-0.051571	-0.013472	-0.042573
48	origin_region_PACIFIC	0.043051	-0.003486	0.000000	0.027431	-0.010112	0.000000
49	origin_region_SOUTH AMERICA	-0.068919	-0.039900	-0.025594	-0.020091	-0.039291	-0.010252
50	origin_region_SOUTHEAST ASIA	-0.190777	-0.199270	-0.114039	-0.001240	-0.022273	0.133483
51	origin_region_SOUTHERN AFRICA	0.041704	0.026354	-0.035231	-0.457530	-0.325369	0.046897
52	origin_region_WESTERN EUROPE	-0.120979	-0.132917	-0.079344	-0.032643	0.020640	0.003759
53	origin_size_diff	0.000001	0.000000	0.000001	NaN	NaN	NaN
	R-squared (Testing Set)	0.3284	0.2222	0.2885	0.3016	0.2203	0.2249
	RMSE (Testing Set)	0.2181	0.1450	0.1259	0.2197	0.1479	0.1323

Table 5: Lasso Regression Coefficients Across Scenarios

	Feature	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
0	capacity_diff	0.000067	0.000011	0.000020	0.000067	0.000012	0.000029
1	depcount_diff	0.021090	0.012277	0.019465	0.021817	0.012545	0.018392
2	destination_distance	NaN	NaN	NaN	-0.000556	-0.000693	-0.001469
3	destination_region_AUSTRALIA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
4	destination_region_CARIBBEAN	-0.013587	0.000000	-0.000000	-0.000000	-0.000000	-0.013672
5	destination_region_CENTRAL AFRICA	-0.000000	NaN	-0.000000	-0.000000	-0.000000	-0.000000
6	destination_region_CENTRAL AMERICA	-0.000000	-0.000000	-0.000000	0.000000	0.000000	-0.000000
7	destination_region_CENTRAL ASIA	0.000000	0.000000	NaN	NaN	NaN	NaN
8	destination_region_EAST AFRICA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
9	destination_region_EASTERN EUROPE	-0.000000	0.000000	-0.072419	-0.101101	-0.040232	0.000000
10	destination_region_FAR EAST ASIA	0.029315	0.014183	0.000000	0.000000	0.000000	-0.000000
11	destination_region_GULF	0.028169	0.000000	0.000000	0.034010	0.000000	0.000000
12	destination_region_MIDDLE EAST	0.000000	-0.000000	-0.000000	NaN	NaN	NaN
13	destination_region_NORTH AFRICA	0.000000	0.000000	0.023603	0.000000	0.000000	-0.000000
14	destination_region_NORTH AMERICA	-0.016370	-0.019469	-0.009595	-0.000000	-0.016096	0.000000
15	destination_region_PACIFIC	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	destination_region_SOUTH AMERICA	0.018383	0.000000	0.000000	0.000000	0.000000	0.000000
17	destination_region_SOUTHEAST ASIA	-0.000000	-0.000000	0.027025	-0.012882	-0.021745	0.000000
18	destination_region_SOUTHERN AFRICA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
19	destination_region_WEST AFRICA	-0.000000	-0.000000	-0.002250	NaN	NaN	NaN
20	destination_region_WESTERN EUROPE	-0.010334	0.000000	-0.000000	-0.000000	0.000000	-0.024082
21	destination_size_diff	NaN	NaN	NaN	0.000001	0.000001	0.000001
22	fare_ratio	-0.024215	0.001989	-0.011815	-0.024227	0.000000	-0.015449
23	international_market	0.006162	0.019069	-0.024634	0.007258	0.019196	-0.032661
24	month_April	-0.000000	0.000000	-0.000000	0.000000	0.000000	-0.000000
25	month_August	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.006555
26	month_December	0.000273	-0.000000	0.000000	0.000000	-0.000000	0.000000
27	month_February	0.000000	0.000000	-0.001815	0.000000	0.000000	-0.002235
28	month_July	-0.005186	-0.000000	0.000000	-0.014448	-0.000000	-0.000000
29	month_June	0.000000	-0.000000	0.000000	0.000000	-0.000000	0.001480
30	month_March	0.005554	0.000000	0.000000	0.003110	0.000000	0.000625
31	month_May	-0.000000	0.000000	-0.000000	-0.000000	-0.000000	-0.000000
32	month_November	-0.000000	-0.000000	-0.000000	0.000000	-0.000000	0.000000
33	month_October	0.000000	-0.000000	0.000000	-0.000000	-0.000000	-0.001662
34	month_September	-0.000000	-0.000000	0.001103	0.000000	-0.000000	0.000000
35	origin_distance	-0.000373	-0.000754	-0.001395	NaN	NaN	NaN
36	origin_region_AUSTRALIA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
37	origin_region_CARIBBEAN	-0.000000	-0.000000	-0.000000	-0.007088	0.000000	0.000000
38	origin_region_CENTRAL AFRICA	-0.000000	-0.000000	-0.000000	-0.000000	NaN	-0.000000
39	origin_region_CENTRAL AMERICA	0.000000	0.000000	0.000000	-0.009261	-0.000000	-0.000000
40	origin_region_CENTRAL ASIA	NaN	NaN	NaN	0.000000	0.000000	NaN
41	origin_region_EAST AFRICA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
42	origin_region_EASTERN EUROPE	-0.095122	-0.036569	-0.000000	-0.000000	0.000000	-0.062204
43	origin_region_FAR EAST ASIA	0.000000	0.000000	0.000000	0.029262	0.023409	-0.000000
44	origin_region_GULF	0.032955	0.000000	0.000000	0.019045	0.000000	0.000000
45	origin_region_MIDDLE EAST	NaN	NaN	NaN	0.000000	-0.000000	-0.000000
46	origin_region_NORTH AFRICA	-0.000000	0.000000	-0.000000	0.000000	-0.000000	0.025538
47	origin_region_NORTH AMERICA	0.000000	-0.023555	0.000000	-0.021065	-0.026463	-0.015313
48	origin_region_PACIFIC	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
49	origin_region_SOUTH AMERICA	0.000000	0.000000	0.000000	0.002491	0.000000	0.000000
50	origin_region_SOUTHEAST ASIA	-0.024793	-0.046592	0.000000	-0.000000	-0.000000	0.031852
51	origin_region_SOUTHERN AFRICA	-0.000000	-0.000000	0.000000	-0.000000	-0.000000	0.000000
52	origin_region_WESTERN EUROPE	-0.004004	0.000000	-0.032278	-0.007823	0.000422	-0.004392
53	origin_size_diff	0.000001	0.000000	0.000001	NaN	NaN	NaN
	R-squared (Testing Set)	0.3045	0.2004	0.2316	0.2854	0.2113	0.2113
	RMSE (Testing Set)	0.2219	0.1470	0.1308	0.2223	0.1334	0.1334

Table 6: Random Forest Feature Importance Across Scenarios

	Feature	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
0	capacity_diff	0.587593	0.326076	0.127044	0.601662	0.336620	0.144945
1	depcount_diff	0.092102	0.115626	0.328613	0.082472	0.107998	0.327675
2	destination_distance	NaN	NaN	NaN	0.064722	0.132303	0.108062
3	destination_region_AUSTRALIA	0.000056	0.000050	0.000259	0.000054	0.000054	0.001429
4	destination_region_CARIBBEAN	0.003991	0.006871	0.005452	0.000857	0.000039	0.004763
5	destination_region_CENTRAL AFRICA	0.000022	NaN	0.000138	0.000203	0.000054	0.000385
6	destination_region_CENTRAL AMERICA	0.001472	0.003627	0.000930	0.000370	0.000202	0.000006
7	destination_region_CENTRAL ASIA	0.000795	0.001234	NaN	NaN	NaN	NaN
8	destination_region_EAST AFRICA	0.000487	0.001402	0.000165	0.000179	0.000270	0.000013
9	destination_region_EASTERN EUROPE	0.009247	0.011130	0.028595	0.013165	0.008901	0.003636
10	destination_region_FAR EAST ASIA	0.001848	0.004790	0.004328	0.000615	0.001139	0.002037
11	destination_region_GULF	0.001128	0.001317	0.001276	0.000432	0.000664	0.000247
12	destination_region_MIDDLE EAST	0.003169	0.009644	0.003543	NaN	NaN	NaN
13	destination_region_NORTH AFRICA	0.001244	0.003036	0.001318	0.000437	0.003893	0.000114
14	destination_region_NORTH AMERICA	0.007853	0.024698	0.007739	0.004222	0.006852	0.007187
15	destination_region_PACIFIC	0.000045	0.000051	0.000000	0.000068	0.000037	0.000000
16	destination_region_SOUTH AMERICA	0.002191	0.004497	0.003986	0.000707	0.001308	0.003112
17	destination_region_SOUTHEAST ASIA	0.002199	0.005611	0.006332	0.000814	0.002033	0.000881
18	destination_region_SOUTHERN AFRICA	0.000083	0.000038	0.000287	0.000073	0.000014	0.000190
19	destination_region_WEST AFRICA	0.003080	0.000007	0.013376	NaN	NaN	NaN
20	destination_region_WESTERN EUROPE	0.009870	0.009926	0.009608	0.004617	0.006109	0.008399
21	destination_size_diff	NaN	NaN	NaN	0.074860	0.117844	0.115381
22	fare_ratio	0.065303	0.111443	0.124070	0.064400	0.113168	0.135998
23	international_market	0.010319	0.021631	0.011630	0.009586	0.025869	0.011163
24	month_April	0.002498	0.003923	0.005135	0.002518	0.004352	0.006137
25	month_August	0.002458	0.004582	0.006654	0.002486	0.004144	0.005826
26	month_December	0.002144	0.005059	0.002734	0.002198	0.004347	0.003978
27	month_February	0.002225	0.004634	0.004481	0.002475	0.004871	0.005431
28	month_July	0.002652	0.004100	0.008163	0.002895	0.004355	0.005974
29	month_June	0.002483	0.003724	0.005889	0.002131	0.003843	0.005433
30	month_March	0.002769	0.004615	0.005173	0.002615	0.004111	0.004923
31	month_May	0.002462	0.004389	0.004619	0.002316	0.003866	0.004495
32	month_November	0.002308	0.003585	0.004993	0.002277	0.004033	0.004461
33	month_October	0.002738	0.004295	0.005988	0.002722	0.004250	0.006487
34	month_September	0.002531	0.004183	0.006033	0.002228	0.004150	0.004680
35	origin_distance	0.065287	0.129561	0.087147	NaN	NaN	NaN
36	origin_region_AUSTRALIA	0.000085	0.000036	0.000010	0.000064	0.000039	0.001254
37	origin_region_CARIBBEAN	0.000465	0.000962	0.002364	0.003562	0.007049	0.005165
38	origin_region_CENTRAL AFRICA	0.000082	0.000174	0.000434	0.000072	NaN	0.000382
39	origin_region_CENTRAL AMERICA	0.000067	0.000156	0.000001	0.001925	0.003152	0.001467
40	origin_region_CENTRAL ASIA	NaN	NaN	NaN	0.000894	0.001790	NaN
41	origin_region_EAST AFRICA	0.000275	0.000380	0.000016	0.000639	0.001632	0.000205
42	origin_region_EASTERN EUROPE	0.010621	0.004917	0.002837	0.008080	0.012268	0.016757
43	origin_region_FAR EAST ASIA	0.000811	0.001029	0.003129	0.003614	0.006224	0.002185
44	origin_region_GULF	0.000342	0.000836	0.000199	0.001642	0.001995	0.004550
45	origin_region_MIDDLE EAST	NaN	NaN	NaN	0.003049	0.014067	0.002931
46	origin_region_NORTH AFRICA	0.000621	0.001950	0.000187	0.001646	0.003217	0.003949
47	origin_region_NORTH AMERICA	0.003440	0.006169	0.008114	0.009033	0.020782	0.007608
48	origin_region_PACIFIC	0.000053	0.000096	0.000000	0.000094	0.000107	0.000000
49	origin_region_SOUTH AMERICA	0.001078	0.001758	0.003450	0.002025	0.003649	0.002264
50	origin_region_SOUTHEAST ASIA	0.000459	0.001896	0.000457	0.002635	0.003872	0.007565
51	origin_region_SOUTHERN AFRICA	0.000086	0.000081	0.000703	0.000127	0.000013	0.000128
52	origin_region_WESTERN EUROPE	0.006158	0.006579	0.011927	0.009526	0.008448	0.010144
53	origin_size_diff	0.078707	0.133625	0.140473	NaN	NaN	NaN
	R-squared (Testing Set)	0.8478	0.7215	0.6659	0.8459	0.7566	0.6614
	RMSE (Testing Set)	0.1038	0.0868	0.0863	0.1032	0.0826	0.0874

Table 7: XGBoost Feature Importance Across Scenarios

	Feature	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5	Scenario 6
0	capacity_diff	0.477270	0.230067	0.085400	0.502174	0.208438	0.070312
1	depcount_diff	0.048832	0.051136	0.098882	0.035742	0.040812	0.179832
2	destination_distance	NaN	NaN	NaN	0.025667	0.055458	0.036900
3	destination_region_AUSTRALIA	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
4	destination_region_CARIBBEAN	0.000000	0.000000	0.000000	0.012376	0.000000	0.018183
5	destination_region_CENTRAL AFRICA	0.004368	NaN	0.010246	0.000000	0.000000	0.010015
6	destination_region_CENTRAL AMERICA	0.005850	0.030666	0.007161	0.015124	0.000000	0.000000
7	destination_region_CENTRAL ASIA	0.007247	0.007334	NaN	NaN	NaN	NaN
8	destination_region_EAST AFRICA	0.009209	0.019531	0.000000	0.008582	0.000000	0.000000
9	destination_region_EASTERN EUROPE	0.028084	0.067234	0.081595	0.046846	0.040267	0.000000
10	destination_region_FAR EAST ASIA	0.024668	0.011328	0.000000	0.000000	0.023927	0.030969
11	destination_region_GULF	0.000000	0.000000	0.010410	0.000000	0.000000	0.000000
12	destination_region_MIDDLE EAST	0.030962	0.030112	0.009008	NaN	NaN	NaN
13	destination_region_NORTH AFRICA	0.000000	0.017527	0.020416	0.008140	0.020431	0.000000
14	destination_region_NORTH AMERICA	0.049173	0.068397	0.002206	0.005831	0.062290	0.087575
15	destination_region_PACIFIC	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
16	destination_region_SOUTH AMERICA	0.022887	0.029940	0.030733	0.010740	0.000000	0.038676
17	destination_region_SOUTHEAST ASIA	0.008442	0.019683	0.038211	0.018881	0.008028	0.000000
18	destination_region_SOUTHERN AFRICA	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
19	destination_region_WEST AFRICA	0.018861	0.000000	0.045326	NaN	NaN	NaN
20	destination_region_WESTERN EUROPE	0.021929	0.009228	0.020257	0.029220	0.094520	0.036419
21	destination_size_diff	NaN	NaN	NaN	0.027761	0.048047	0.053856
22	fare_ratio	0.016475	0.035668	0.020956	0.013334	0.022055	0.027537
23	international_market	0.013570	0.030645	0.026590	0.014568	0.030665	0.046184
24	month_April	0.000000	0.000000	0.002319	0.006897	0.000000	0.007760
25	month_August	0.000000	0.000000	0.002262	0.000000	0.000000	0.010699
26	month_December	0.000286	0.000000	0.000000	0.000000	0.000655	0.000000
27	month_February	0.000269	0.000421	0.015271	0.000000	0.000000	0.002426
28	month_July	0.000000	0.000000	0.011383	0.004980	0.000000	0.014720
29	month_June	0.000000	0.002517	0.004783	0.000000	0.000000	0.008502
30	month_March	0.000000	0.000000	0.000017	0.000000	0.000183	0.001720
31	month_May	0.000000	0.000000	0.001099	0.000000	0.000000	0.001147
32	month_November	0.000012	0.000597	0.000000	0.000016	0.000258	0.001324
33	month_October	0.000000	0.000000	0.000000	0.000000	0.000000	0.002897
34	month_September	0.000847	0.000443	0.004936	0.000000	0.000059	0.000000
35	origin_distance	0.025167	0.059868	0.030009	NaN	NaN	NaN
36	origin_region_AUSTRALIA	0.000000	0.000000	0.035838	0.000000	0.000000	0.027674
37	origin_region_CARIBBEAN	0.009574	0.008781	0.033659	0.008242	0.000000	0.006898
38	origin_region_CENTRAL AFRICA	0.000000	0.008789	0.000000	0.004905	NaN	0.008050
39	origin_region_CENTRAL AMERICA	0.000000	0.000000	0.000000	0.007462	0.016029	0.011317
40	origin_region_CENTRAL ASIA	NaN	NaN	NaN	0.008574	0.009463	NaN
41	origin_region_EAST AFRICA	0.009697	0.038193	0.005674	0.005335	0.022898	0.000000
42	origin_region_EASTERN EUROPE	0.050051	0.036169	0.007371	0.042020	0.052757	0.055695
43	origin_region_FAR EAST ASIA	0.000000	0.000000	0.014845	0.038723	0.006731	0.000000
44	origin_region_GULF	0.000000	0.000000	0.000000	0.000876	0.037996	0.020028
45	origin_region_MIDDLE EAST	NaN	NaN	NaN	0.013324	0.042563	0.007140
46	origin_region_NORTH AFRICA	0.000000	0.033874	0.016102	0.000000	0.000000	0.021563
47	origin_region_NORTH AMERICA	0.017723	0.043971	0.093350	0.025036	0.068023	0.029516
48	origin_region_PACIFIC	0.005678	0.000000	0.000000	0.006637	0.006246	0.000000
49	origin_region_SOUTH AMERICA	0.034800	0.000000	0.020318	0.027380	0.025424	0.018748
50	origin_region_SOUTHEAST ASIA	0.000000	0.015693	0.000000	0.005017	0.018652	0.056335
51	origin_region_SOUTHERN AFRICA	0.000000	0.000000	0.011673	0.000000	0.000000	0.008174
52	origin_region_WESTERN EUROPE	0.031052	0.039751	0.129274	0.019588	0.037123	0.041212
53	origin_size_diff	0.027016	0.052439	0.052421	NaN	NaN	NaN
	R-squared (Testing Set)	0.7031	0.5135	0.5379	0.6920	0.5412	0.5205
	RMSE (Testing Set)	0.1450	0.1147	0.1015	0.1459	0.1134	0.1041

5.7 Summary of Results

Across all models, `fare_ratio` and `depcount_diff` consistently appeared among the most important predictors, reinforcing the critical role of pricing and flight frequency in competitive airport environments. The variation in month and region dummy variable effects across scenarios reinforces the need for context-aware modeling. For instance, a variable like `destination_region_SOUTHEAST ASIA` may be highly significant in smaller-share scenarios, reflecting the emergence of secondary airports in growing regional hubs. Meanwhile, holiday months might play a greater role in same-origin models, where travel is more time-sensitive and seasonally driven.

Comparing the performance across all models, Random Forest and XGBoost models provided the highest predictive accuracy, while the linear models contributed interpretable insights on the marginal effect of each feature. These results highlight the value of using both linear and nonlinear techniques for comprehensive market share modeling.

6 Limitations and Future Research

While the results of this study offer important insights into competitive airport dynamics, several limitations must be acknowledged that may affect the validity and generalizability of the findings.

The dataset used in this analysis was limited to a 24-month period (2022–2023) and focused only on records where exactly two competing airports existed within the same O&D cluster. This pairwise design, while useful for interpretability, excluded markets with three or more competing airports. As such, the findings may not generalize to more complex airport networks or hub-and-spoke competition structures.

Moreover, the data reflects post-pandemic air travel conditions, which may not fully represent typical passenger behavior. For instance, international travel was still rebounding during this period, and regional shifts in demand may not be stable over time.

North America—particularly the United States—was heavily overrepresented in the dataset. Although this reflects the high density of commercial airports and competitive markets in the U.S., it may bias the model's learned patterns toward regions with well-developed air infrastructure and higher volumes of domestic traffic. Underrepresented regions such as Africa, Central Asia, and Oceania had relatively few records, which limits the model's ability to generalize insights to these areas.

The regression models assume that relationships between predictors and market share are stable across time and geography, which may not hold in practice. For linear models such as OLS, assumptions of linearity, homoscedasticity, and feature independence are difficult to guarantee in a real-world setting where variables may interact or be co-linear. Additionally, while ensemble models like Random Forest and XGBoost capture complex relationships, they do not offer the same interpretability as linear models. Variable importance scores can be sensitive to data noise or correlated features, which may obscure true causal relationships.

Given these limitations, several avenues for future research can be pursued to enhance the scope, depth, and generalizability of the findings. Some include:

- **Expand to Multi-Airport and Multi-Year Analysis:**
  - Include O&D markets with 3+ competing airports
  - Analyze multi-year data for temporal patterns
- **Incorporate External Factors:**
  - Add macroeconomic indicators, fuel prices, and delay performance
  - Integrate policy and infrastructure variables
- **Improve Interpretability and Causal Inference:**
  - Use SHAP or LIME for interpretable ML
  - Apply causal methods (e.g., IV regression, matching)
- **Conduct Region-Specific Modeling:**

– Build customized models for Southeast Asia, Europe, etc.
– Compare model behavior under different regulatory regimes

7 Conclusion

This study demonstrates that passenger market share between adjacent airports is influenced by a combination of operational, geographic, and pricing factors. While OLS and regularized regression models offered interpretable baselines, machine learning models—particularly Random Forest and XGBoost—delivered superior predictive performance.

The most important drivers of passenger choice include fare competitiveness, flight frequency, seat capacity, and the relative distance between competing airports. These variables consistently emerged as top predictors across different market segments and modeling frameworks.

Notably, the analysis also highlights the dynamic nature of competition: smaller airports can attract greater market share when they provide better accessibility or pricing. Conversely, larger airports typically dominate international travel due to more extensive capacity and network connections.

The insights derived from this project can support strategic decision-making in airline route planning, airport investment, and regional transportation policy. Future research may extend this work by incorporating passenger-level data, accessibility scores, and behavioral metrics such as loyalty programs or demographic preferences to better explain airport choice in competitive regions.

References

[1] B. Loo, "Passengers' airport choice within multi-airport regions (mars): some insights from a stated preference survey at hong kong international airport," *Journal of Transport Geography*, vol. 16, pp. 117–125, 2008.

[2] W. Cho, R. Windle, and M. Dresner, "The impact of low-cost carriers on airport choice in the us: a case study of the washington–baltimore region," *Transportation Research Part E Logistics and Transportation Review*, vol. 81, pp. 141–157, 2015.

[3] V. Psaraki, A. Stathopoulos, and C. Abakoumkin, "Parking capacity requirements for relocated airports: the new athens international airport," *Transportation Research Record Journal of the Transportation Research Board*, vol. 1788, pp. 19–25, 2002.

[4] W. Wattanacharoensil, M. Schuckert, and A. Graham, "An airport experience framework from a tourism perspective," *Transport Reviews*, vol. 36, pp. 318–340, 2015.

[5] L. Adiloğlu-Yalçınkaya and E. Baranskaitė, "Transformation of tourism: innovations in the world's busiest airports," 2019.

[6] B. Qian, N. Zhang, and C. Wu, "Estimate spatial spillover of airport operational efficiency in the yrd region," *Sustainability*, vol. 14, p. 1019, 2022.

[7] Y. Li, M. Teng, and Y. Feng, "Research on the relationship between beijing-tianjin-hebei airport group and regional coordinated development," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–14, 2022.

[8] H. Putro, R. Trinanda, and A. Iscahyono, "Understanding airport leakage at multi airport in west java through accessibility characteristics to the airport," *Jurnal Teknik Sipil Dan Perencanaan*, vol. 25, pp. 161–169, 2023.

[9] Y. Zhang, J. Xue, M. Li, R. Liu, and Y. Jing, "Evaluating civil aviation airport competitiveness in the yangtze river economic belt of china: a lens of spatial-temporal evolution," *Frontiers in Environmental Science*, vol. 10, 2022.

[10] M. Pandiangan, T. Kurniawan, A. Aurelia, and T. Rasya, "Enhancing airport environmental sustainability through net zero energy building principles," *Iop Conference Series Earth and Environmental Science*, vol. 1268, p. 012044, 2023.

[11] W. Wei and J. He, "Analysis of competition relationship of multiairport regional airport market based on deep learning," *Mathematical Problems in Engineering*, vol. 2022, pp. 1–10, 2022.

[12] J. Lian and J. Rønnevik, "Airport competition – regional airports losing ground to main airports," *Journal of Transport Geography*, vol. 19, pp. 85–92, 2011.

[13] S. Hess and J. Polak, "Airport, airline and access mode choice in the san francisco bay area," *Papers of the Regional Science Association*, vol. 85, pp. 543–567, 2006.

[14] A. Avenali, T. D'Alfonso, A. Nastasi, and P. Reverberi, "High-speed rail and air transport integration in hub-and-spoke networks: the role of airports," pp. 171–195, 2022.

[15] D. Lee and M. Luengo-Prado, "Are passengers willing to pay more for additional legroom?" *Journal of Air Transport Management*, vol. 10, pp. 377–383, 2004.

[16] A. Pacagnella, P. Hollaender, G. Mazzanati, and W. Bortoletto, "Infrastructure and flight consolidation efficiency of public and private brazilian international airports: a two-stage dea and malmquist index approach," *Journal of Advanced Transportation*, vol. 2020, pp. 1–15, 2020.