Airfare Markets Under Pressure

Modeling Fare Prediction and Market Structure in U.S. Domestic Air Travel

By:Abhigyan Dey, Derek Li, Michelle Fang

DATA SOURCE

U.S. DOT Domestic Airline Consumer Airfare Report

2021–2025 Q2

KEY QUESTIONS

Can route-level fares be predicted from distance, demand, competition, and hub characteristics?

How do hub dominance and low-cost carrier (LCC) penetration shape fare levels?

Two Core Questions Drive This Analysis

Modeling & Prediction

Can route-level fares be predicted using distance, demand, competition, and endpoint hub characteristics?

Which features contribute most to explaining fare variation across markets?

Market Structure & Pricing

How do fares differ between routes that touch highly dominant hub cities versus more competitive markets?

Do routes with greater low-cost carrier (LCC) penetration exhibit systematically lower fares?

DATASET SCOPE

14,004 route-quarter observations across U.S. domestic markets (2021–2025 Q2), covering 21 variables including distance, passenger volume, dominant carrier market share, LCC share, and city-level hub characteristics.

Average U.S. Domestic Fares Range from ~$89 to ~$452

MEAN FARE	STANDARD DEVIATION	FARE PER MILE	DATASET SCOPE
$237.70	$63.93	$0.29	2021–2025
Median: $227.64	High Variation	Median: $0.24	~1,000 routes/quarter

Most Expensive Routes (Mean)

ROUTE	FARE
New York City → San Francisco	$420
Los Angeles → New York City	$412
Detroit → San Francisco	$408
Atlanta → Salt Lake City	$405

Cheapest Routes (Mean)

ROUTE	FARE
Knoxville → Sanford, FL	$92
Phoenix → Provo, UT	$89
Atlantic City → Orlando	$101
Bellingham → Las Vegas	$110

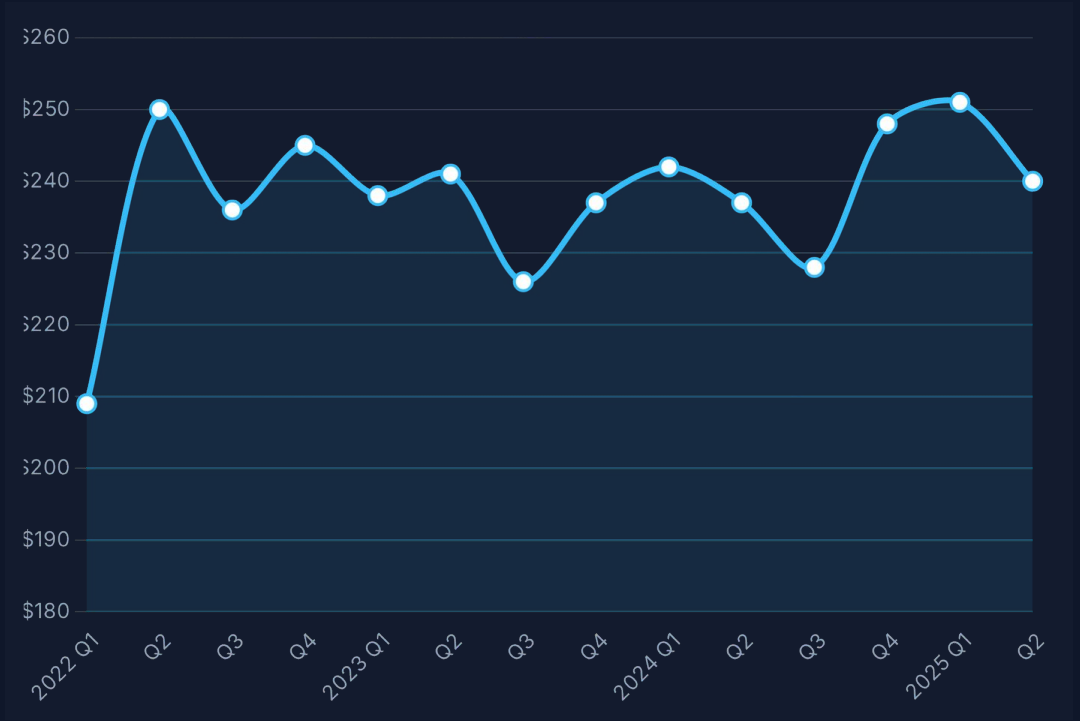
Q2 Peaks and Post-Pandemic Fare Inflation Are Visible in the Data

Seasonal Peaks
Q2 (April–June) consistently shows peak fares, driven by summer travel demand.

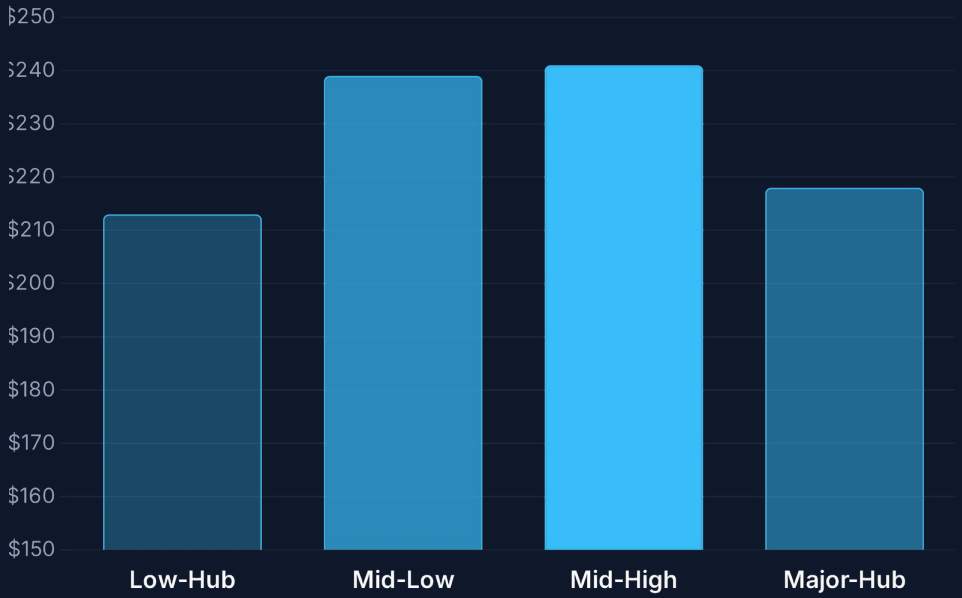
Post-Peak Normalization
Q3 (July–September) shows a notable dip relative to Q2 across all years.

Inflationary Trend
2025 Q1 recorded the highest mean fare ($251), continuing an upward trend since 2022.

Year	Q1	Q2	Q3	Q4
2022	$209	$250	$236	$245
2023	$238	$241	$226	$237
2024	$242	$237	$228	$248
2025	$251	$240	—	—



Major-Hub Routes Are Not Necessarily the Most Expensive



HUB TIER	MEDIAN FARE	MEAN FARE
Low-Hub	**$213**	$222
Mid-Low	**$239**	$248
Mid-High	**$241**	$251
Major-Hub	**$218**	$229

KEY INSIGHT

Mid-tier hub routes carry the highest median fares. Major-hub cities attract more competition (including LCCs), which suppresses fares despite high demand.

High LCC Penetration Reduces Median Fares by ~$14 Across All Hub Tiers

MEDIAN FARE BY LCC TIER

LCC TIER	MEDIAN FARE
Low-LCC	$232
Mid-Low	$233
Mid-High	$228
High-LCC	**$217**

KEY FINDING

LCC presence acts as a consistent downward pressure on fares regardless of hub tier.

JOINT EFFECT: HUB TIER × LCC TIER

	LOW-LCC	MID-LOW	MID-HIGH	HIGH-LCC
Low-Hub	$221	$215	$211	**$198**
Mid-Low	$231	$248	$239	$234
Mid-High	$243	$245	$244	$229
Major-Hub	$245	$219	$218	**$211**

$34 Reduction on Major-Hub routes when LCC penetration is high ($245 → $211).

The Dominant Carrier Charges a Median Premium of $47.67 Over the Lowest-Fare Carrier

MEDIAN FARE SPREAD

$47.67

Difference between dominant carrier fare and lowest available fare.

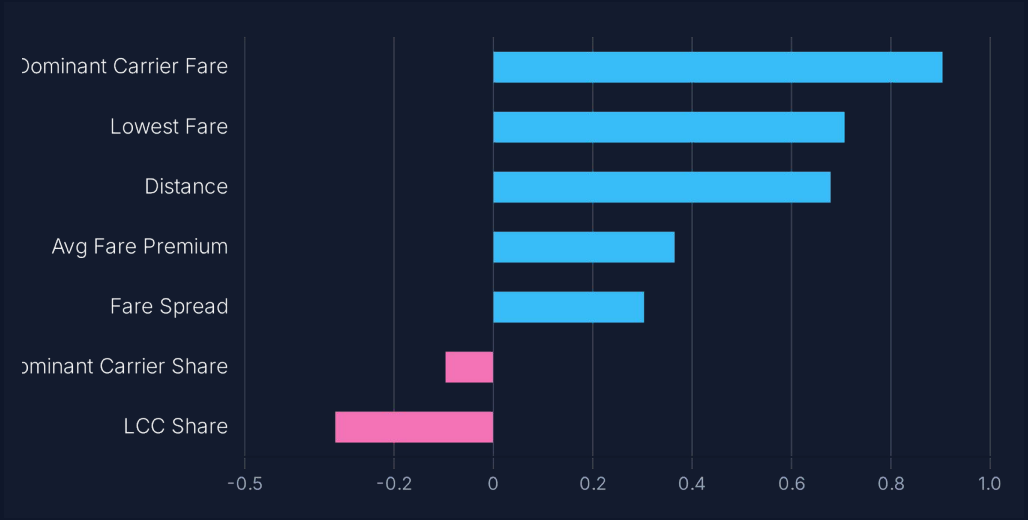
ZERO SPREAD ROUTES

31.2%

On nearly **one-third** of routes, the dominant carrier is the lowest-fare carrier.

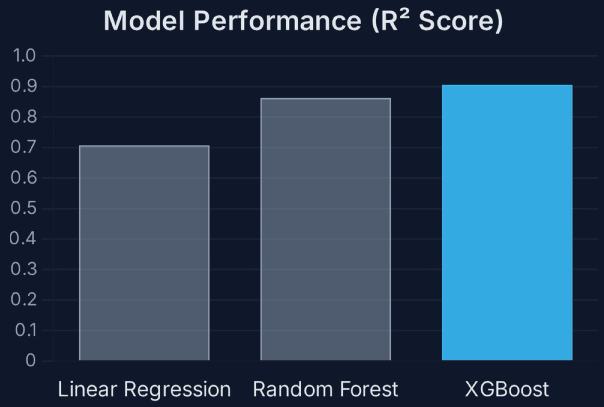
Key Insight: Dominant carrier fare level strongly predicts route average (r = 0.903).

Correlation with Average Fare (r)



XGBoost on Structural Features Achieves $R^2 = 0.906$

MODEL	MAE	RMSE	R^2 SCORE
Linear Regression	$27.13	$34.88	0.708
Random Forest	$17.91	$23.89	0.863
XGBoost	**$14.86**	**$19.81**	**0.906**



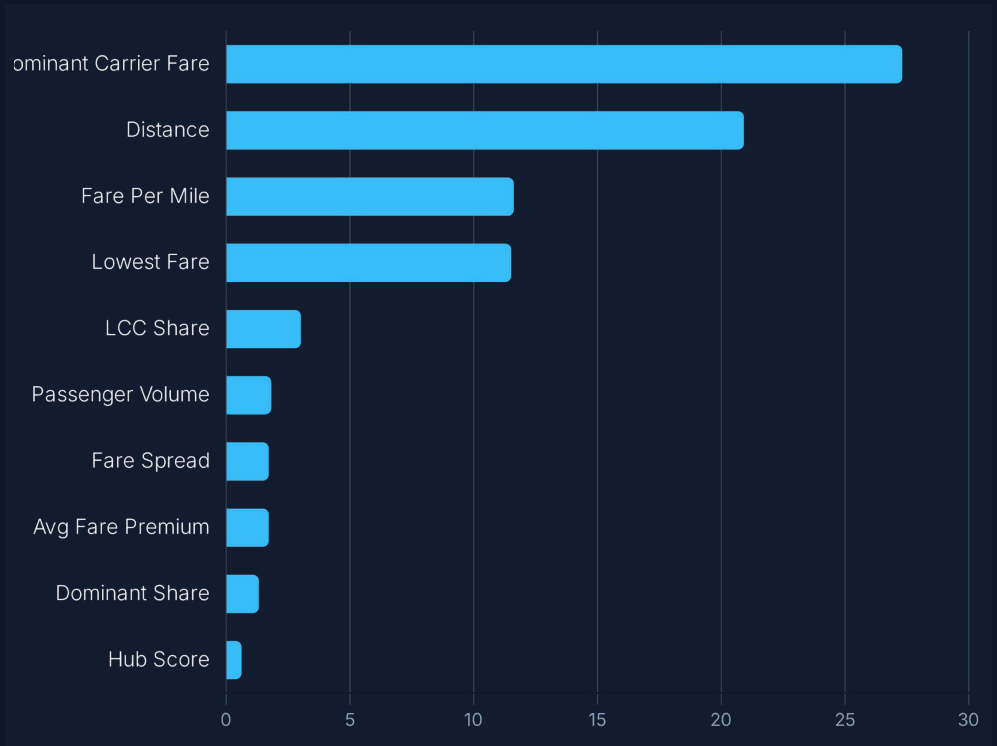
KEY INSIGHT

Using only structural market features (distance, demand, competition), XGBoost predicts fares to within $14.86 on average.

Adding carrier-level pricing data improves performance to near-perfect prediction ($R^2 = 0.991$, MAE = $4.34).

Distance, Dominant Carrier Fare, and LCC Share Are the Top Predictors

SHAP Feature Importance (Mean |Impact|)



Structural-Only Model Drivers

Distance	34.0%
Avg Fare Premium	18.8%
City Premium Index	5.3%
Passenger Volume	4.5%
LCC Share	4.1%

KEY TAKEAWAY

Among purely structural features, distance is the dominant driver (34% of importance), followed by the average fare premium at the city endpoints.

Evidence-Based Insights for Consumers, Industry, and Policymakers

CONSUMERS & TRAVELERS

Routes with high LCC penetration offer fares **~$17–$34 lower** on average. Booking on routes touching major-hub cities with active LCC competition (e.g., Atlanta, Chicago) can yield meaningful savings. Seasonal patterns show Q3 fares dip relative to Q2 peaks.

TRAVEL PLATFORMS

The high predictability of fares (**$R^2 = 0.906$**) enables robust fare-scoring tools. Distance, hub score, and LCC penetration can power affordability indicators and route comparison features that improve pricing transparency and user trust.

AIRLINES

Hub dominance does not guarantee premium pricing—LCC entry on major-hub routes suppresses fares by up to $34. The median **$47.67 fare spread** signals competitive vulnerability on routes with growing LCC presence.

POLICYMAKERS

Mid-tier hub routes show the highest median fares with less LCC competition—these markets warrant monitoring. The **31.2% of routes** where the dominant carrier is also the cheapest suggests complex competitive dynamics beyond simple concentration metrics.

Market Structure Is Highly Predictive of Fares; LCC Presence Is the Key Competitive Lever

01 **FARE PREDICTION IS HIGHLY FEASIBLE**

XGBoost achieves **$R^2 = 0.906$** using only structural features (distance, demand, competition), with an MAE of $14.86 per route.

02 **DISTANCE IS THE DOMINANT DRIVER**

Distance accounts for **34%** of structural feature importance. Fare per mile averages $0.29 but varies significantly by route type.

03 **LCC PENETRATION LOWERS FARES**

High-LCC routes average **$17 less** than Low-LCC routes. The effect is strongest on Major-Hub routes with a **$34 reduction**.

04 **HUB DOMINANCE IS NON-LINEAR**

Mid-tier hub routes carry the highest fares. Major-Hub routes attract enough competition to moderate pricing despite high demand.

05 **DOMINANT CARRIER PREMIUM**

Median fare spread is **$47.67**, but 31.2% of routes show zero spread—the dominant carrier is often the cheapest option.

06 **TEMPORAL PATTERNS MATTER**

Q2 peaks and a 2025 Q1 high of **$251** suggest continued post-pandemic fare inflation, with Q3 offering relative value.

Dataset and Analytical Approach

DATA SOURCE

- **U.S. DOT Domestic Airline Consumer Airfare Report**

Period: 2021 – 2025 Q2

- 14,004 route-quarter observations
- 21raw variables (distance, passengers, fares, shares)

DATA SPLIT

Train: 11,177 observations

Test: 2,795 observations

FEATURE ENGINEERING

DERIVED METRICS

- **Hub Score:** Composite index of city-level hub intensity
- **Competition Intensity:** From dominant carrier & LCC share
- **Fare Premium & Spread:** Dominant vs avg and lowest fares

MODELING

ALGORITHMS

Linear Regression | Random Forest | XGBoost

EVALUATION METRICS

MAE | RMSE | R^2 Score | 5-fold CV

STRATEGY

- Compare Structural-Only vs Full Feature sets