Visualizing Airfare Trends ✈️

Ishan Patel, Sristi Karamchandani, Juntae Park, Royce Arockiasamy, Ved Sanjanwala, Luc Gau

Motivation:

Airfare pricing has traditionally been driven by short-term factors like booking timelines and seat demand. However, these dynamic pricing models often fail to account for broader, long-term influences such as economic conditions and global disruptions. As a result, price trends appear unpredictable and reactive, limiting the ability of both consumers and industry stakeholders to make informed decisions.

The airline industry is a critical driver of global connectivity and economic development. Understanding the forces behind airfare fluctuations is essential for travelers planning ahead, airlines optimizing revenue strategies, and policymakers ensuring market transparency. By identifying and visualizing how external factors influence airfare trends over time, our project offers a foundation for smarter forecasting and better decision-making for airfare pricing

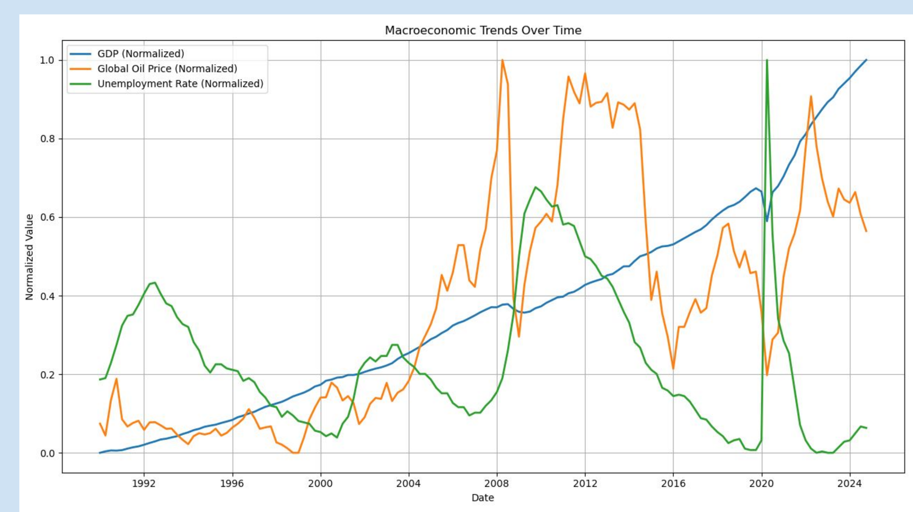
Data:

Flight Data

tbl	Year	quarter	citymarketid_1	citymarketid_2	city1	city2	airportid_1	airportid_2	airport_1	airport_2	nsmiles	passengers	fare	carrier_lg	large_ms	fare_lg	carrier_low	lf_ms	fare_low	tbl1apk
Table1a	2021	3	30135	33195	Allentown/Bethlehem/Easton, PA	Tampa, FL (Metropolitan Area)	10135	14112	ABE	PIE	970	180	81.43	G4	1.0000	81.43	G4	1.0000	81.43	202131013514112ABEPIE
Table1a	2021	3	30135	33195	Allentown/Bethlehem/Easton, PA	Tampa, FL (Metropolitan Area)	10135	15304	ABE	TPA	970	19	208.93	DL	0.4659	219.98	UA	0.1193	154.11	202131013515304ABETPA
Table1a	2021	3	30140	30194	Albuquerque, NM	Dallas/Fort Worth, TX	10140	11259	ABQ	DAL	580	204	184.56	WN	0.9968	184.44	WN	0.9968	184.44	202131014011259ABQDAL
Table1a	2021	3	30140	30194	Albuquerque, NM	Dallas/Fort Worth, TX	10140	11298	ABQ	DFW	580	264	182.64	AA	0.9774	183.09	AA	0.9774	183.09	202131014011298ABQDFW
Table1a	2021	3	30140	30466	Albuquerque, NM	Phoenix, AZ	10140	14107	ABQ	PHX	328	398	177.11	WN	0.6061	184.49	AA	0.3939	165.77	202131014014107ABQPHX
Table1a	2021	3	30140	30721	Albuquerque, NM	Boston, MA (Metropolitan Area)	10140	10721	ABQ	BOS	1974	153	324.97	AA	0.4263	323.73	WN	0.1609	298.20	202131014010721ABQBOS
Table1a	2021	3	30140	30721	Albuquerque, NM	Boston, MA (Metropolitan Area)	10140	13296	ABQ	MHT	1974	16	315.90	WN	0.7285	270.42	WN	0.7285	270.42	202131014013296ABQMHT
Table1a	2021	3	30140	30721	Albuquerque, NM	Boston, MA (Metropolitan Area)	10140	14307	ABQ	PVD	1974	22	329.22	WN	0.5415	271.60	WN	0.5415	271.60	202131014014307ABQPVD
Table1a	2021	3	30140	30852	Albuquerque, NM	Washington, DC (Metropolitan Area)	10140	10821	ABQ	BWI	1670	159	255.89	WN	0.7212	244.89	WN	0.7212	244.89	202131014010821ABQBWI
Table1a	2021	3	30140	30852	Albuquerque, NM	Washington, DC (Metropolitan Area)	10140	11278	ABQ	DCA	1670	151	291.16	AA	0.4404	296.88	WN	0.3197	247.20	202131014011278ABQDCA
Table1a	2021	3	30140	30852	Albuquerque, NM	Washington, DC (Metropolitan Area)	10140	12264	ABQ	IAD	1670	59	343.58	UA	0.5646	382.06	WN	0.1402	266.61	202131014012264ABQIAD
Table1a	2021	3	30140	30977	Albuquerque, NM	Chicago, IL	10140	13232	ABQ	MDW	1121	99	231.66	WN	0.9934	230.85	WN	0.9934	230.85	202131014013232ABQMDW
Table1a	2021	3	30140	30977	Albuquerque, NM	Chicago, IL	10140	13930	ABQ	ORD	1121	206	241.25	AA	0.6869	229.41	AA	0.6869	229.41	202131014013930ABQORD
Table1a	2021	3	30140	31453	Albuquerque, NM	Houston, TX	10140	12191	ABQ	HOU	759	166	203.93	WN	0.9712	203.44	WN	0.9712	203.44	202131014012191ABQHOU
Table1a	2021	3	30140	31453	Albuquerque, NM	Houston, TX	10140	12266	ABQ	IAH	759	148	232.85	UA	0.8512	235.69	WN	0.0447	213.54	202131014012266ABQIAH
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	11618	ABQ	EWR	1861	77	342.13	UA	0.3845	382.33	AA	0.3324	296.40	202131014011618ABQEWR
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	12197	ABQ	HPN	1861	3	356.92	DL	0.6154	358.24	AA	0.3462	350.22	202131014012197ABQHPN
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	12391	ABQ	ISP	1861	4	240.10	WN	0.8049	220.94	WN	0.8049	220.94	202131014012391ABQISP
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	12478	ABQ	JFK	1861	177	251.78	B6	0.8338	234.07	B6	0.8338	234.07	202131014012478ABQJFK
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	12953	ABQ	LGA	1861	125	295.22	AA	0.5013	302.37	WN	0.2145	229.04	202131014012953ABQLGA
Table1a	2021	3	30140	31703	Albuquerque, NM	New York City, NY (Metropolitan Area)	10140	15070	ABQ	SWF	1861	0	427.50	AA	1.0000	427.50	AA	1.0000	427.50	202131014015070ABQSWF
Table1a	2021	3	30140	32457	Albuquerque, NM	San Francisco, CA (Metropolitan Area)	10140	13796	ABQ	OAK	896	216	194.36	WN	0.9638	193.20	WN	0.9638	193.20	202131014013796ABQOAK
Table1a	2021	3	30140	32457	Albuquerque, NM	San Francisco, CA (Metropolitan Area)	10140	14771	ABQ	SFO	896	188	246.07	UA	0.6493	255.54	WN	0.1346	221.69	202131014014771ABQSFO
Table1a	2021	3	30140	32457	Albuquerque, NM	San Francisco, CA (Metropolitan Area)	10140	14831	ABQ	SJC	896	61	209.22	WN	0.6566	200.07	WN	0.6566	200.07	202131014014831ABQSJC
Table1a	2021	3	30140	32575	Albuquerque, NM	Los Angeles, CA (Metropolitan Area)	10140	10800	ABQ	BUR	677	39	205.45	WN	0.6592	205.43	AA	0.2430	196.01	202131014010800ABQBUR
Table1a	2021	3	30140	32575	Albuquerque, NM	Los Angeles, CA (Metropolitan Area)	10140	12892	ABQ	LAX	677	572	172.93	DL	0.3790	181.73	WN	0.2572	158.14	202131014012892ABQLAX
Table1a	2021	3	30140	32575	Albuquerque, NM	Los Angeles, CA (Metropolitan Area)	10140	12954	ABQ	LGB	677	12	201.28	WN	0.5636	187.75	DL	0.1182	180.77	202131014012954ABQLGB

Marco Data

year	date	quarter	GDP	Global Oil Price	Unemployment Rate
1990	1990-01-01	1	5872.701	19.7191798418972	5.3
1990	1990-04-01	2	5960.028	16.3834368530021	5.33333333333330
1990	1990-07-01	3	6015.116	26.3570783926219	5.7
1990	1990-10-01	4	6004.733	32.3667607754564	6.13333333333330
1991	1991-01-01	1	6035.178	20.9463957902001	6.6000000000000000
1991	1991-04-01	2	6126.862	18.9095092226614	6.83333333333330
1991	1991-07-01	3	6205.937	19.9218253968254	6.86666666666670
1991	1991-10-01	4	6264.54	20.5237511763599	7.1000000000000000
1992	1992-01-01	1	6363.102	17.9752042160738	7.36666666666670
1992	1992-04-01	2	6470.763	20.0618326118326	7.6000000000000000
1992	1992-07-01	3	6566.641	20.1358601543384	7.63333333333330
1992	1992-10-01	4	6680.803	19.2600542694021	7.36666666666670
1993	1993-01-01	1	6729.459	18.2650838509317	7.13333333333330



U.S. Airline Fare Data (Kaggle):
- ➢ Years Covered: 1993–2024
- ➢ Size: 63 MB, 2 million records
  Frequency: Quarterly
- ➢ Key Features:
  - ○ Origin/destination cities and airport codes
  - ○ Distance, number of passengers
  - ○ Average fare and carrier-specific fare info
- ➢ Purpose: Core dataset used to model fare trends across domestic routes

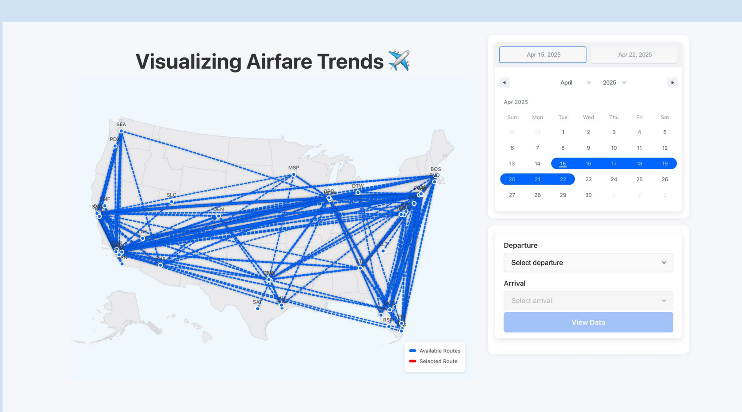
U.S. GDP Data (FRED):
- ➢ Years Covered: 1947–2024
- ➢ Frequency: Quarterly
- ➢ Units: Billions of U.S. Dollars (Seasonally Adjusted Annual Rate)
- ➢ Purpose: Captures macroeconomic growth, used to understand long-term pricing trends

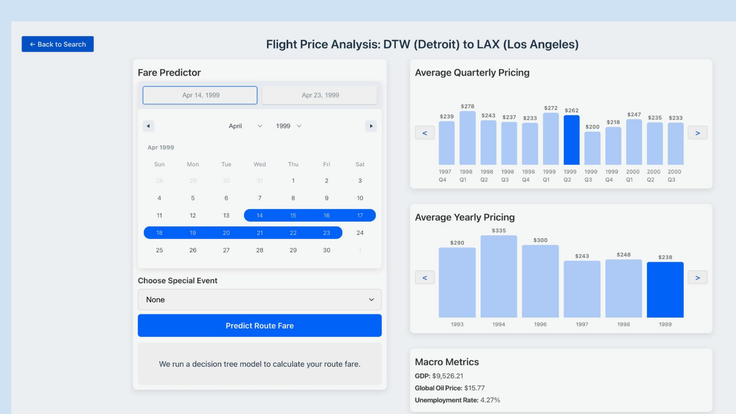
Brent Crude Oil Prices (FRED):
- ➢ Years Covered: 1990–2025
- ➢ Frequency: Quarterly
- ➢ Units: U.S. Dollars per Barrel
- ➢ Purpose: Included as a key cost driver of airline operations; captures fuel-related fare fluctuations

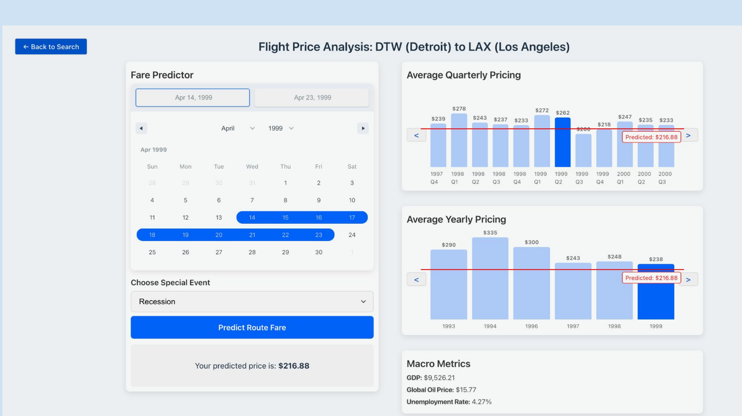
U.S. Unemployment Rate (FRED):
- ➢ Years Covered: 1948–2025
- ➢ Frequency: Monthly
- ➢ Units: Percentage (Seasonally Adjusted)
- ➢ Purpose: Indicator of consumer demand and travel behavior; aggregated to quarterly for model input

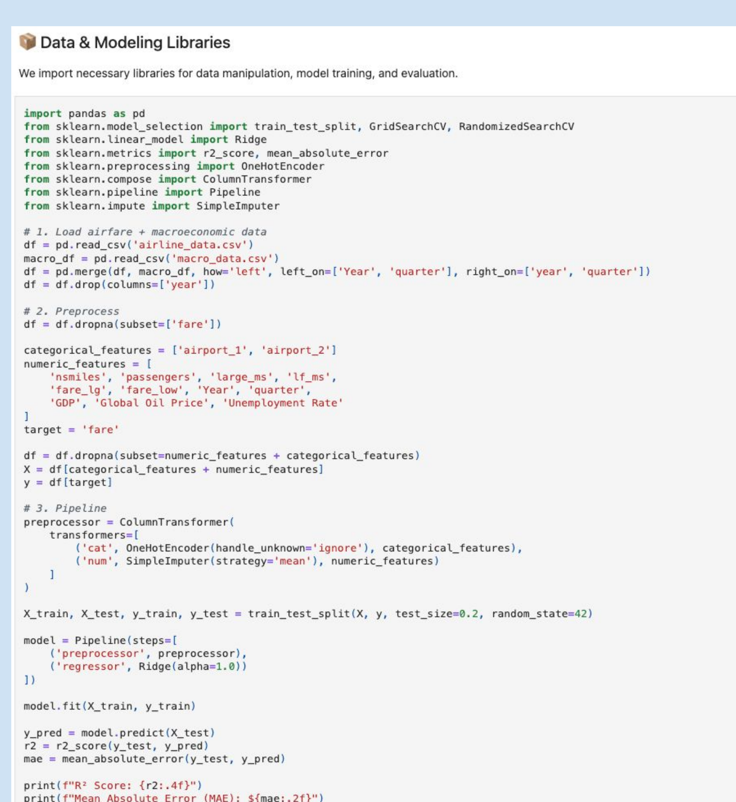
Data Collection and Cleaning:
- ➢ Downloaded 20+ years of U.S. domestic airfare data from Kaggle
- ➢ Pulled macroeconomic indicators (GDP, oil prices, unemployment) from FRED
- ➢ Added event-based markers (e.g., 9/11, COVID-19, 2008 recession) for contextual modeling
- ➢ Parsed GDP and oil prices directly from web tables (quarterly)
- ➢ Scraped unemployment data (monthly) and aggregated to quarterly for consistency
- ➢ Removed rows with NaNs, duplicates, and outliers (e.g., zero fares, extreme values)
- ➢ Merged macroeconomic indicators into airfare data by quarter and year
- ➢ Ensured alignment between route-level fare data and national-level economic trends

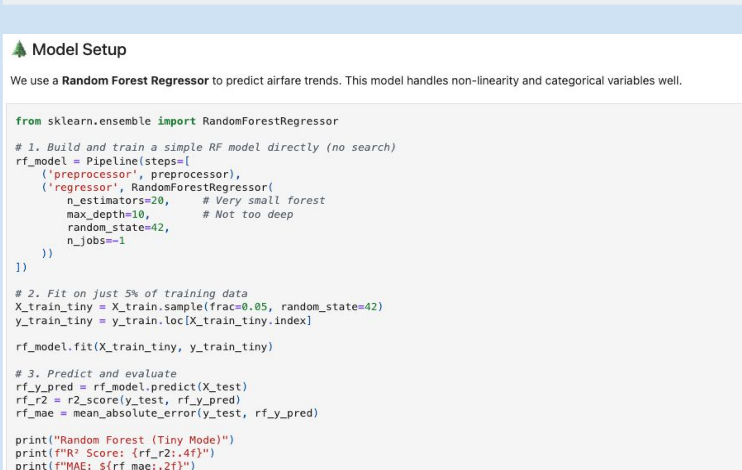
Algorithm and Interactive Visualization:

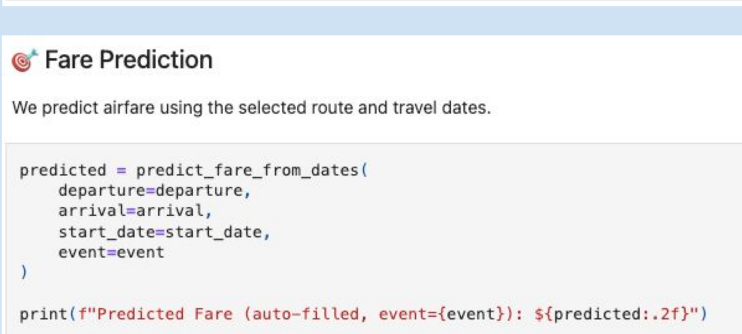


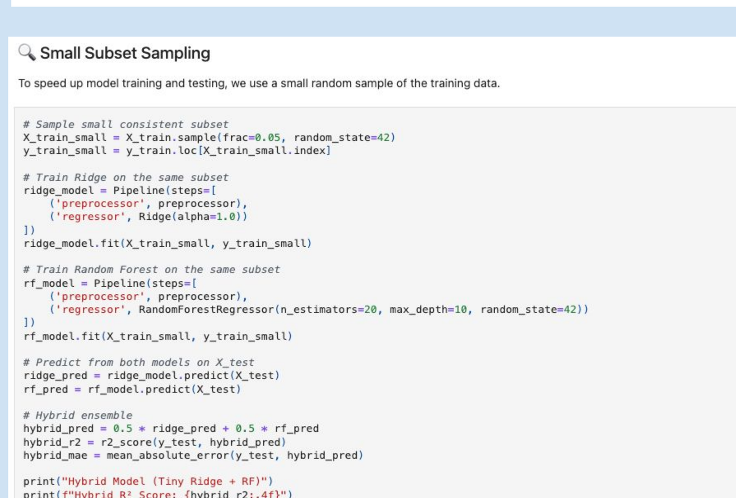












Our Approach:

- ➢ A hybrid machine learning model paired with an interactive web-based visualization tool that predicts and explains airline fare trends across U.S. routes
- ➢ Model: Learns from 20+ years of fare data, economic indicators (GDP, oil, unemployment), and special events to predict future prices
- ➢ Visualization: Lets users explore historical pricing, select routes/dates, and simulate the effects of disruptions like recessions in real time
- ➢ Traditional fare models react to short-term demand.
- ➢ Airfare pricing is driven by more than demand, it is influenced by seasonality, economic trends, and global events
- ➢ Ours model is proactive, capturing both linear patterns (e.g., inflation) and nonlinear shocks (e.g., pandemics)
- ➢ Our visualization makes these insights accessible and actionable
- ➢ Offers a user-facing platform that ties pricing, demand, and economic signals into one decision-support system

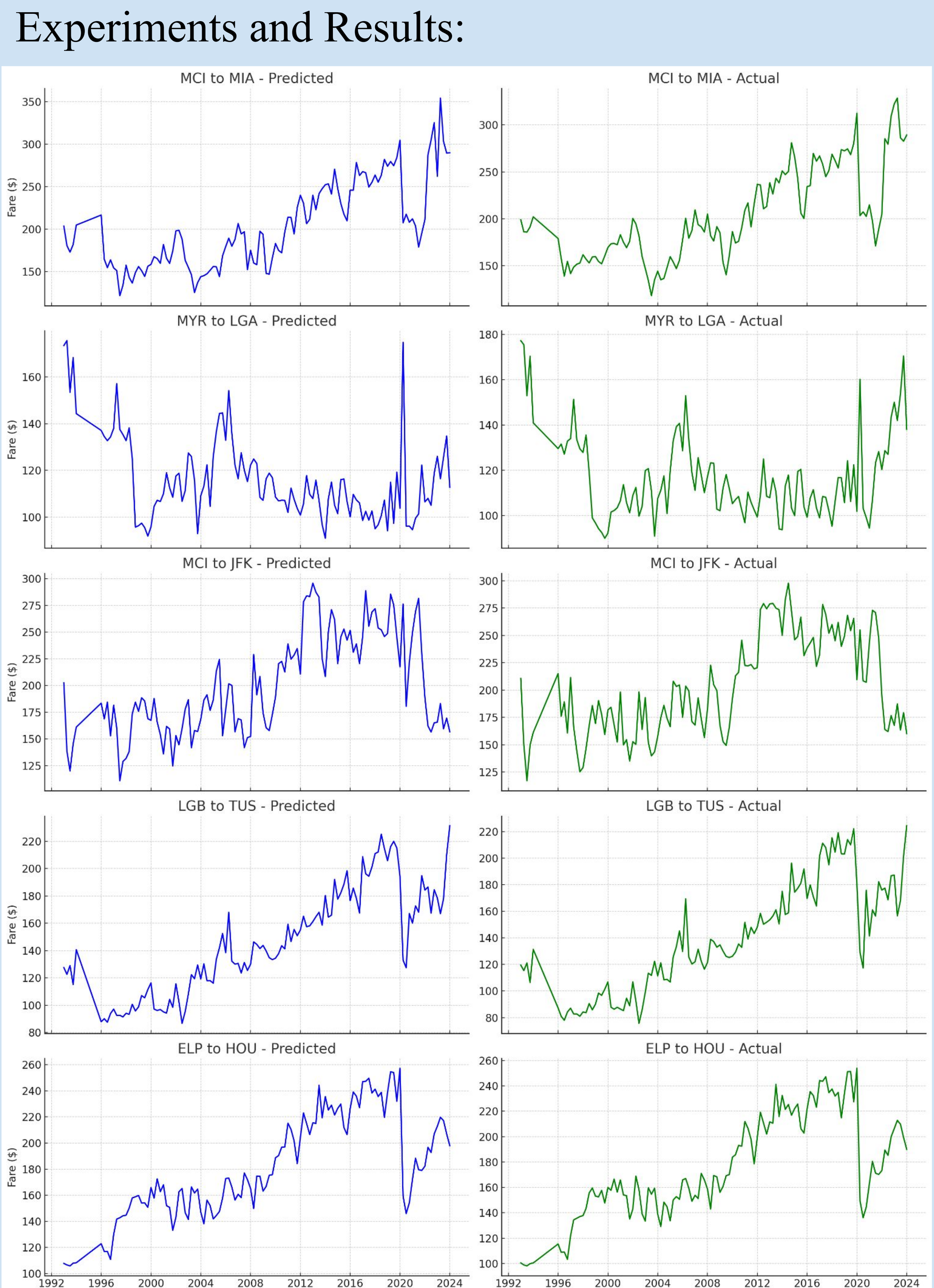
Visualization Features:

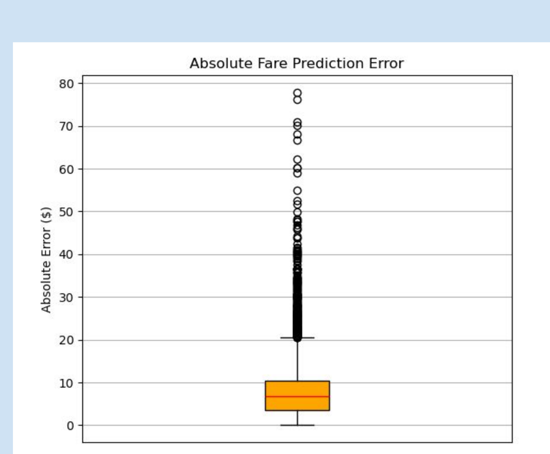
- ➢ Route Map Visualization: Interactive U.S. map displays all available routes; users can explore pricing by selecting specific origin-destination pairs.
- ➢ Custom Date Selection: Allows users to view fare predictions for custom travel dates
- ➢ Event Simulation Tool: Users can simulate the impact of external events like recessions on fare predictions allowing for scenario-based analysis
- ➢ Quarterly and Yearly Pricing Trends: Bar charts show historical fare patterns over time, helping users detect seasonality, inflationary effects, and airline pricing cycles.
- ➢ Macro Metrics Dashboard: Displays GDP, oil prices, and unemployment rates relevant to the selected time period, helping connect fares to economic context

ML Model:

- ➢ Random Forest: handles complex, nonlinear impacts, like oil prices and economic shocks
- ➢ Ridge Regression: captures smooth, long-term economic trends

Experiments and Results:





Absolute Fare Prediction Error

Statistic	Difference ($)
Mean	8.46
Standard Deviation	7.92
Minimum	0
25th Percentile	3.59
75th Percentile	10.36
Maximum	77.92

Model Performance Metrics

Metric	Value
Mean Absolute Error	8.461986
Root Mean Squared Error	11.590751
R^2 Score	0.962821
Mean Absolute Percentage Error	4.402026

Prediction vs Actual Fare Graphs:

- ➢ Predictions closely track actual fare trends, reflecting real market behavior
- ➢ Seasonal spikes and long-term shifts are accurately captured
- ➢ Route-specific dynamics (e.g., busy vs. regional routes) are well preserved

Absolute Fare Prediction Error:

- ➢ 95.6% prediction accuracy
- ➢ Model fares align closely with real prices
- ➢ Most predictions are within a few dollars, making it practical for planning and pricing
- ➢ Performs well even in volatile periods, such as recessions and pandemics
- ➢ Tight error range shows consistent and trustworthy performance
- ➢ Model explains 96% of variation in airfare prices
- ➢ Benchmarked traditional models (e.g., linear regression) show lower performance of $R^2 \approx 0.61$
- ➢ Captures both macro and micro trends
- ➢ Provides a reliable, high-confidence tool for forecasting and analysis