Airline Market Data Dictionary

U.S. Airline Industry Dataset (2005–2019)

September 2025

1   Data Construction Process

This dataset was constructed by processing quarterly DB1B Market **10% sample** files from the U.S. Department of Transportation's Bureau of Transportation Statistics and merging with auxiliary data on airport characteristics, demographics, and market structure. The DB1B 10% sample is the standard public-use file provided by BTS.

1.1   Data Sources

- **Primary**: U.S. DOT Bureau of Transportation Statistics, Airline Origin and Destination Survey (DB1B Market)

- **Website**: `https://www.transtats.bts.gov/DatabaseInfo.asp?DB_ID=125`

- **Auxiliary**: Census population data, airport characteristics, vacation destinations

1.2   Processing Steps

1. **Market Definition**: Origin-destination city pairs based on city market IDs

2. **Time Aggregation**: Quarterly data aggregated from individual trip records

3. **Geographic Scope**: Limited to Continental U.S. (CONUS) airports

4. **Market Size Filter**: Markets with $\geq$20 passengers/day on average ($\geq$182.5 quarterly passengers in 10% sample)

5. **Price Filter**: Real fares between $25-$2,500 (2019 dollars)

6. **Competition Filter**: Excludes monopoly markets (rival_carriers > 0)

1.3   Dataset Overview

- **File**: `airline_data_main.dta`

- **Period**: 2005 Q1 - 2019 Q4 (60 quarters)

- **Level**: Product-market-time (carrier-route-quarter)

- **Observations**: 1,677,867 product-market-quarter combinations

- **Markets**: 8,059 unique origin-destination city pairs

- **Carriers**: 134 unique airline codes

2 Variable Definitions

2.1 Identifiers and Time

Variable	Type	Description
year	integer	Year (2005-2019)
quarter	integer	Quarter (1-4)
time	string	Time identifier in format YYYY-Q
origin	integer	Origin city market ID (BTS designation)
destination	integer	Destination city market ID (BTS designation)
carrier	string	Two-letter airline code (IATA/ICAO standard)
market_code	string	Origin-destination market identifier
market_name	string	Human-readable market name (e.g., "NYC-LAX")

2.2 Demand and Price Variables

Variable	Type	Description
total_passengers	integer	Total quarterly passengers for this carrier on this route
average_fare	float	Average ticket price in 2019 real dollars. Constructed as passenger-weighted average of individual fares
share_nonstop	float	Share of carrier's passengers on nonstop flights (0-1)
average_distance	float	Average flight distance in miles, including connections
average_nonstop_miles	float	Average nonstop flight distance in miles
average_extra_miles	float	Additional miles due to connections. Calculated as average_distance - average_nonstop_miles

2.3 City Demographics and Airport Characteristics

Variable	Type	Description
origin_pop	float	Origin metropolitan area population. Merged from Census/BEA data by airport-year
dest_pop	float	Destination metropolitan area population. Merged from Census/BEA data by airport-year
pop_o_d_geo_mean	float	Geometric mean of endpoint populations: $\sqrt{\text{origin\_pop} \times \text{dest\_pop}}$. Primary market size measure

Variable	Type	Description
pop_sum	float	Sum of endpoint populations: origin_pop + dest_pop
mean_pop	float	Average of endpoint populations: (origin_pop + dest_pop)/2
max_pop	float	Maximum of endpoint populations: max(origin_pop, dest_pop)
origin_hub	binary	Origin airport is a major hub. Based on T_MASTER_CORD classification
destination_hub	binary	Destination airport is a major hub
origin_slot_controlled	binary	Origin airport has slot restrictions (capacity constraints)
destination_slot_controlled	binary	Destination airport has slot restrictions
destination_vacation	binary	Destination city is classified as a vacation/leisure market

2.4 Carrier Types and Classification

Variable	Type	Description
major	binary	Major carrier indicator. Large network carriers with extensive route networks
legacy	binary	Legacy carrier indicator. Traditional full-service carriers (American, Delta, United, etc.)
lcc	binary	Low-cost carrier indicator. Discount/low-cost carriers (Southwest, JetBlue, etc.)
fringe	binary	Fringe carrier indicator. Calculated as 1 - major

2.5 Competition and Network Structure

Variable	Type	Description
fringe_carriers	integer	Number of fringe carriers serving this market in this quarter
rival_carriers	integer	Number of rival carriers in market (total carriers minus focal carrier)
presence	float	Carrier's passenger share at origin airport. Calculated as carrier_passengers_at_origin / total_passengers_at_origin
total_presence	float	Sum of all carriers' presence measures in this market
num_destinations	integer	Number of destinations served by carrier from this origin airport in this quarter
num_markets	integer	Number of distinct origin-destination markets served by carrier in this quarter

Variable	Type	Description
total_num_destinations	integer	Total destinations served by all carriers from this origin
total_num_markets	integer	Total markets served by all carriers
average_distance_rival	float	Average distance flown by rival carriers in this market. Calculated as leave-one-out mean: $\frac{\sum \text{rival\_distance}-\text{own\_distance}}{N-1}$
average_presence_rival	float	Average presence of rival carriers at origin airport. Leave-one-out mean of presence measures
average_num_destinations_rival	float	Average network size of rival carriers (destinations). Leave-one-out mean
average_num_markets_rival	float	Average network size of rival carriers (markets). Leave-one-out mean

3 Data Quality and Usage Notes

3.1 Missing Data

- Population data: ~1.25% missing (primarily small airports/MSAs)

- Core variables: 99%+ coverage for passengers, fares, distance

- Complete temporal coverage: All quarters 2005–2019

4 File Dependencies

- **Source Data**: DB1B Market quarterly files (2005Q1-2019Q4)

- **Auxiliary Data**: Airport characteristics, demographics, vacation destinations

- **Crosswalk**: crosswalk_airlines_code.csv maps carrier codes to airline names

- **Processing Code**: processing_data.do contains full data construction methodology

5 Citation

When using this dataset, please cite:

- U.S. Department of Transportation, Bureau of Transportation Statistics, Airline Origin and Destination Survey (DB1B)

- Data processing and construction methodology from this class

Last updated: September 2025
For questions about variable construction or methodology, contact professors.