

Project 1 Data Card

Title

Authors	Last updated	Created
SL56393	01/30/2026	01/29/2026
DR39873	01/30/2026	01/29/2026
KBP856	01/30/2026	01/29/2026

Contents

- 1. [Card ID](#)
 - 2. [Dataset at a glance](#)
 - 3. [Usage information](#)
 - 4. [Source and collection process](#)
 - 5. [Exploration](#)
-

Card ID

Primary Dataset	<i>Airline Fight Routes in The US [1993-2024]</i>
Description	<i>This dataset is an overview of domestic airline routes in the US. It includes information regarding passenger travel patterns, market trends, and airline pricing strategies.</i>
Supplemental Dataset 1	<i>Geospatial Data Pack for Visualization</i>
Description	<i>This dataset contains:</i> Base Map Boundaries, Geographic Reference Points, Chloropleth Mapping Data, Geophysical and Atmospheric Data <i>Last updated 10/2025</i> Link for quick viewing: Geospatial Data Pack for Visualization

Add more rows to the above table for each additional supplemental dataset.

Datasets at a glance

Primary dataset

The primary dataset contains *31 years of* flight routes, quarterly earnings, and passengers on each route for each corresponding year ... in the period from 1993 to 2024.

Fields (if applicable)

Field name	Data Type	Description
tbl	str	Table identifier
Year	int	Short description, value range, and example values
Quarter	Year of record	Quarter in the year
citymarketid_1	int	US DOT identifier grouping airports in the same city
citymarketid_2	int	US DOT identifier grouping airports in the same city
city1	str	Departure city
city2	str	Arrival city
airportid_1	int	US DOT airport code for the departure airport
airportid_2	int	US DOT airport code for the arrival city
airport_1	str	Departure airport code
airport_2	str	Arrival airport code
nsmiles	int	Distance of the route in nautical miles
passengers	int	Number of daily passengers
fare	float	Average fair
carrier_lg	str	Largest market share carrier (IATA code)
large_ms	float	Market share proportion for the largest market share carrier
fare_lg	float	Fare of the largest market share carrier
carrier_low	float	Carrier with the lowest fare
lf_ms	float	Lowest carrier market share proportion
fare_low	float	Lowest fair
tbl1apk	str	Unique ID

Supplemental dataset 1

The supplemental dataset contains 3376 *iata values (airport codes)* with the corresponding coordinates, names of the airports, and cities they reside in, and the airport percentage in each state. Last updated 10/2025

Fields (if applicable)

Field name	Data Type	Description
iata	alphanumeric	Airport codes
name	String	Airport name
city	string	The city the airport is in
state	string	The state the airport is in
country	string	The country the airport is in
latitude	float	The latitude of the airport
longitude	float	The longitude of the airport

Additional notes

Provide any additional notes on the data here (e.g. changes in data collection and labeling practices).

Usage information

Use(s)

Primary Dataset:

1. Identify popular routes, and understand seasonal variations in passenger traffic.
2. Analyze airline competition on specific routes and assess pricing strategies.
3. Airlines can use this data to evaluate existing routes and identify potential new routes with high passenger demand.
4. Businesses can use this data to understand travel patterns relevant to their industry and make informed decisions.

Supplemental Dataset 1:

1. Allows us to map airports onto the US map using their latitude and longitude
2. Tracks the percentages of the total amounts of airports that reside in each state

3. Analyze airport density by state and country to understand regional differences in aviation infrastructure.

Licensing

License type for each data source and the license's corresponding requirements for sharing, transforming, logging changes, citing, and restricting usage of the data.

Primary Dataset:

- Provided under the Open Database License (ODbL) v1.0
- It is a "copyleft" or reciprocal license specifically designed for databases
- Sharing: You are free to copy, distribute, and use the database. However, if you redistribute it, you must do so under the same ODbL license.
- Transforming: You can modify, transform, and build upon the database. If you publicly use an adapted version, you must offer that adapted version under the ODbL.
- Logging changes: You must make it clear that the database has been modified and that it is a derivative of the original.
- Citing/Attribution: You may not use the trademarks or names of this original licensor in a way that suggests they endorse your work.
- Usage Restrictions: You cannot hold the original contributors liable for any damages arising from your use of the data. There are no restrictions on using this data for commercial purposes.

Supplemental Dataset 1:

- **CC-BY-4.0** (Creative Commons Attribution 4.0): Applies to most datasets including base maps (us-10m.json, world-110m.json), London features, earthquakes.json, windvectors.csv, volcano.json, annual-precip.json, income.json, obesity.json, species.csv, population_engineers_hurricanes.csv, and zipcodes.csv. Requires attribution.
- **Public Domain**: Applies to airports.csv, unemployment.tsv, and flights-airport.csv. No restrictions.

Privacy considerations

Indicate any anonymization of data or redaction of PII, as well as other privacy-relevant information.

Primary Dataset:

1. Consists of aggregated, public-domain aviation metric
2. The data is structured around market identifiers, airport IDs, and carrier codes. It does not contain personal identifying information
3. Passenger counts and fare information are provided as averages and totals for specific routes and quarters, ensuring that individual transactions cannot be traced back to specific persons.
4. While the dataset includes Geocoded_City1 and Geocoded_City2 coordinates, these refer to the origin and destination cities/airports, not specific residential locations.
5. Privacy protections for corporate entities are not applicable here, as the dataset explicitly identifies airline carriers by their standard codes
6. The data is derived from official records, which are subject to government standards for public data release.

Format

The primary dataset is in a CSV file.

The first supplemental dataset is in a JSON file and CSV file.

Source and collection process

Origin

Source tables used to create the primary dataset (include links) and sources tables used to create the supplemental datasets (include links).

Primary Data Source: [Consumer Airfare Report: Table 1a - All U.S. Airport Pair Markets](#)

Original datasets author/publisher

Name of source datasets author/publisher.

Primary Dataset:

1. Publisher: U.S. Department of Transportation (DOT), Bureau of Transportation Statistics (BTS)
2. Office: Office of Aviation Analysis

Original datasets point of contact

Name and contact information of source datasets point of contact.

Primary Dataset:

1. Office: Office of Aviation Analysis
2. Email: randall.keizer@dot.gov (listed contact for the Consumer Airfare Report)
3. General BTS Contact: answers@dot.gov or 202-366-2347

Original metadata

Online link to original metadata file for your source datasets.

Primary Dataset:

1. [Consumer Airfare Report Data Dictionary / Column Descriptions](#)
2. [Technical Table Descriptions \(Metadata\) - DOT Policy Page](#)

Collection process

Online link(s) to information on data source's collection and labeling processes.

Primary Dataset:

1. The data is derived from the DB1B (Airline Origin and Destination Survey), which is a 10% sample of airline tickets from reporting carriers.
2. Methodology Link: [Domestic Airline Consumer Airfare Report Description](#)

Collection or access date

Download date for each data source used in this dataset.

Primary Dataset:

1. Approximately January 2025 (stated as "Updated a year ago").
2. The DOT releases this data quarterly, typically with a lag of 3-6 months.

Refresh date/schedule

Last refresh date and refresh cadence, or N/A if there is only one version.

Primary Dataset: N/A

Exploration

Questions driving the exploration study

List of questions driving the data preparation, visualization, and analysis.

Primary Dataset:

1. Since the data spans 30 years (1993–2024), how should fare, fare_lg, and fare_low be adjusted for inflation to make year-over-year comparisons valid?
2. Have citymarketid or airportid codes changed over time due to airport reclassifications or closures? Do we need to map old codes to new ones?
3. How can we visualize the density of flight routes on a US map? Can we use line thickness to represent passengers volume and color to represent average fare?
4. What does the time-series plot of "Average Fare vs. Year" look like? Can we overlay major events (9/11, 2008 Recession, COVID-19) to see their immediate impact on air travel?
5. What is the correlation between nsmiles (distance) and fare? Does a scatter plot reveal distinct clusters for short-haul vs. long-haul pricing models?

Section

Add more sections if needed.
