Data Table Schema

airlines

Mapping of airline codes to names.

*61 rows & 2 columns.* Size: ~0.1MB. Source: the Internet, generally.

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| airline\_id | STRING | 2-letter airline identifier |
| airline\_name | STRING | Name of the airline |

airports

Important details (name, state, identifier, latitude, longitude, etc.) on various US airports.

*322 rows & 6 columns.* Size: ~0.1MB. Source: [US Department of Transportation](https://www.transportation.gov/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| airport\_id | STRING | 3-letter airport identifier |
| airport\_name | STRING | Name of the airport |
| city | STRING | Airport’s city |
| state | STRING | Airport’s state |
| latitude | FLOAT | Airport’s latitude |
| longitude | FLOAT | Airport’s longitude |

events\_US

Public events from around the US throughout 2017.

*1,151 rows & 4 columns.* Size: ~0.1MB. Source: [Shore Fire Media](https://shorefire.com/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| date | STRING | Date. Format DD/MM/YYYY. |
| city | STRING | Name of city event takes place in |
| state | STRING | Name of state event takes place in |
| event\_name | STRING | Name of event |

fares

Airline fare distributions for each quarter-route-airline combination in 2017 with a bucket size of $10.

*80,823 rows & 255 columns.* Size: ~44MB. Source: [US Department of Transportation](https://www.transportation.gov/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| quarter | INTEGER | Financial quarter of the datum |
| airline\_id | STRING | Airline identifier |
| origin\_airport | STRING | Unique 3-character identifier of origin airport |
| destination\_airport | STRING | Unique 3-character identifier of destination airport |
| distance | INTEGER | Distance between airports, in miles |
| fare\_bucket\_x (249 total) | INTEGER | Number of passengers whose fares were between x and x+10 dollars. The first column excludes values below $10. Last column includes all values above $2500 |

flight\_traffic

Information about delays for US domestic flights in 2017.

*~600,000 rows & 24 columns.* Size: ~130MB zipped, ~550MB unzipped. Source: [Bureau of Transportation Statistics](https://www.bts.gov/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| year | INTEGER | Year |
| month | INTEGER | Month |
| day | INTEGER | Day of month |
| airline\_id | STRING | Airline identifier |
| origin\_airport | STRING | Origin airport code |
| destination\_airport | STRING | Destination airport code |
| scheduled\_departure | INTEGER | Scheduled departure time. Format HHMM |
| actual\_departure | INTEGER | Actual departure time. Format HHMM |
| taxi\_out | INTEGER | Taxi-out time, in minutes |
| wheels\_off | INTEGER | Wheels off time. Format HHMM |
| wheels\_on | INTEGER | Wheels on time. Format HHMM |
| taxi\_in | INTEGER | Taxi-in time, in minutes |
| scheduled\_arrival | INTEGER | Scheduled arrival time. Format HHMM |
| actual\_arrival | INTEGER | Actual arrival time. Format HHMM |
| cancelled | BINARY | 1 if flight was cancelled, 0 otherwise |
| diverted | BINARY | 1 if flight was diverted, 0 otherwise |
| scheduled\_elapsed | INTEGER | Scheduled elapsed time of flight, in minutes |
| actual\_elapsed | INTEGER | Actual elapsed time of flight, in minutes |
| distance | INTEGER | Distance between airports, in miles |
| airline\_delay | INTEGER | Delay due to the airline, in minutes |
| weather\_delay | INTEGER | Delay due to weather, in minutes |
| air\_system\_delay | INTEGER | Delay due to the National Air System, in minutes |
| security\_delay | INTEGER | Delay due to security, in minutes |
| aircraft\_delay | INTEGER | Delay due to late aircraft, in minutes |

stock\_prices

Daily closing stock prices of various US airlines from late-2016 to early-2018.

*380 rows & 10 columns.* Size: ~0.1MB. Source: [Alpha Vantage](https://www.alphavantage.co/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| timestamp | STRING | Date. Format MM/DD/YY. |
| airline\_stock\_price (9 total) | FLOAT | Stock closing price |

weather

Weather data (temperature, wind, precipitation, cloud cover, etc.) collected at various US airports every 6 hours through 2017.

*~375,000 rows & 12 columns.* Size: ~40MB. Source: [National Centers for Environmental Information](https://www.ncei.noaa.gov/).

|  |  |  |
| --- | --- | --- |
| Field | Type | Description |
| airport\_id | STRING | 3-letter airport identifier |
| airport\_name | STRING | Airport’s name |
| latitude | FLOAT | Airport’s latitude |
| longitude | FLOAT | Airport’s longitude |
| elevation | FLOAT | Airport’s elevation, in meters |
| datetime | STRING | Date and time of the measurement. Format YYYY-MM-DDTHH:MM:SS |
| temperature | FLOAT | Air temperature surrounding the airport, in degrees Celsius |
| visibility | INTEGER | Visibility, in meters |
| wind\_direction | INTEGER | Horizontal wind direction, in degrees from North clockwise |
| wind\_speed | FLOAT | Horizontal wind speed, in meters per second |
| snow\_depth | INTEGER | Depth of snow accumulation, in centimeters |
| cloud\_status | INTEGER | Code (0 – 9) representing the cloud cover status |

***cloud\_status*** *key:*

0: None

1: ACSL (Altocumulus Standing Lenticular)

2: ACCAS (Altocumulus Castelanus)

3: TCU (Towering Cumulus)

4: MDT CU (Moderate Cumulus)

5: CB/CB MAM DISTANT (Cumulonimbus or Cumulonimbus Mammatus visible)

6: CB/CBMAM (Cumulonimbus or Cumulonimbus Mammatus within 20 nautical miles)

7: Unknown

9: Missing