

Flight Prices



This dataset is a CSV file where each row is a purchasable ticket found on Expedia between 2022-04-16 and 2022-10-05, to/from the following airports: ATL, DFW, DEN, ORD, LAX, CLT, MIA, JFK, EWR, SFO, DTW, BOS, PHL, LGA, IAD, OAK.

Question:

What is the lowest price from the top airport?

Content:

- **legId:** An identifier for the flight.
- **searchDate:** The date (YYYY-MM-DD) on which this entry was taken from Expedia.
- **flightDate:** The date (YYYY-MM-DD) of the flight.
- **startingAirport:** Three-character IATA airport code for the initial location.
- **destinationAirport:** Three-character IATA airport code for the arrival location.
- **fareBasisCode:** The fare basis code.
- **travelDuration:** The travel duration in hours and minutes.
- **elapsedDays:** The number of elapsed days (usually 0).
- **isBasicEconomy:** Boolean for whether the ticket is for basic economy.
- **isRefundable:** Boolean for whether the ticket is refundable.
- **isNonStop:** Boolean for whether the flight is non-stop.
- **baseFare:** The price of the ticket (in USD).
- **totalFare:** The price of the ticket (in USD) including taxes and other fees.
- **seatsRemaining:** Integer for the number of seats remaining.
- **totalTravelDistance:** The total travel distance in miles. This data is sometimes missing.
- **segmentsDepartureTimeEpochSeconds:** String containing the departure time (Unix time) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsDepartureTimeRaw:** String containing the departure time (ISO 8601 format: YYYY-MM-DDThh:mm:ss.000±[hh]:00) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsArrivalTimeEpochSeconds:** String containing the arrival time (Unix time) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsArrivalTimeRaw:** String containing the arrival time (ISO 8601 format: YYYY-MM-DDThh:mm:ss.000±[hh]:00) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsArrivalAirportCode:** String containing the IATA airport code for the arrival location for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsDepartureAirportCode:** String containing the IATA airport code for the departure location for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsAirlineName:** String containing the name of the airline that services each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsAirlineCode:** String containing the two-letter airline code that services each leg of the trip. The entries for each of the legs are separated by '|'|.

- **segmentsEquipmentDescription:** String containing the type of airplane used for each leg of the trip (e.g. "Airbus A321" or "Boeing 737-800"). The entries for each of the legs are separated by '|'|.
- **segmentsDurationInSeconds:** String containing the duration of the flight (in seconds) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsDistance:** String containing the distance traveled (in miles) for each leg of the trip. The entries for each of the legs are separated by '|'|.
- **segmentsCabinCode:** String containing the cabin for each leg of the trip (e.g. "coach"). The entries for each of the legs are separated by '|'|.

Problem statement

Our purpose is to use the data To find the lowest price from the top airport, we first identify the airport with the highest number of departures, then look for the minimum ticket price among all flights departing from that airport."

Data Descrip3on

The data that will be used in this project has been downloaded from Kaggle

<https://www.kaggle.com/datasets/dilwong/flightprices/data>

Algorithm

We are going to build different classification models, then we will choose the best model based on the classification metrics that we are going to use. Some feature engineering will be done to enhance the performance of the models.

Tools

- Technologies: Python, Jupyter notebook.
- Libraries: NumPy, Pandas, Sklearn, Matplotlib, Seaborn, Tableau.