Predicting Flight Delays: How Can Airports Increase Revenue from Predictive Analytics?

THE UNITED STATES DOMESTIC AVIATION SECTOR

The United States domestic aviation industry is one of the most robust and economically significant transportation sectors in the world. Over 800 million domestic passengers pass through U.S. airports every year (Statista, 2024). There are an average of 45,000 daily domestic flights in the U.S., representing nearly half of the 106,000 daily flights worldwide (FAA, n.d.-a; OAG, n.d.).

This volume of domestic air traffic generates hundreds of billions in economic activity and supports millions of jobs across the nation. In numbers, commercial aviation drives 5% of U.S. GDP (Airlines for America, 2025). The domestic aviation industry revenue has grown at a CAGR of 20.6% over the past five years, and it is estimated to reach $204.5 billion in 2025 (Samorajski, 2025).

The U.S. aviation system operates under a complex regulatory framework. The Federal Aviation Administration (FAA), established in 1958 and operating under the Department of Transportation, is responsible for managing the American skies. They are in charge of directing air traffic, airport safety, and regulating flight inspection standards (FAA, n.d.-b; FAA, n.d.-c).

Unlike many international counterparts, US airports typically operate under the oversight of local governmental authorities. Most large commercial airports are owned by counties, cities, or regional airport authorities. While the FAA provides regulatory oversight and safety standards, day-to-day management decisions rest with local airport authorities.

MAIN AIRLINES IN THE UNITED STATES

The airline industry in the United States is highly segmented. There are Ultra-Low-Cost Carriers (ULCCs), Low-Cost Carriers (LCCs), Legacy Carriers, Regional Carriers, and "start-up" airlines (see Figure 1).
Legacy carriers, also known as network carriers, include American Airlines, Delta Air Lines, and United Airlines. These airlines operate an extensive network of domestic and international flights, with multiple service classes, diverse aircraft fleets, robust frequent flyer programs, and belong to global airline alliances- Oneworld, SkyTeam, and Star Alliance, respectively—

Figure 1. Segmentation of the domestic commercial aviation industry in the United States.

Low-Cost Carriers are led by Southwest Airlines, the nation's largest domestic airline by passengers carried and by flights (see Exhibit A). JetBlue and Alaska Airlines also fall into this category. These usually feature lower operating costs than legacy carriers and often focus on domestic and short-haul international flights.

Ultra-Low-Cost Carriers such as Spirit Airlines, Frontier Airlines, and Allegiant Air
operate with the industry's lowest cost structures. These airlines feature fare models where passengers pay only for basic transportation service, with all amenities available for additional fees, such as seat selection, baggage, and onboard snacks and refreshments.Their planes usually have smaller leg room to allow for high-density configurations.

Regional airlines operate smaller aircraft. Companies like SkyWest, Republic Airways, and Mesa Airlines fly under brands like American Eagle, Delta Connection, and United Express to connect smaller communities to major hubs.

The "start-up" carriers include newer entrants like Avelo Airlines and Breeze
Airways. These airlines identify routes between secondary markets, often with little direct competition.

REVENUE IN U.S. AIRPORTS

Airports have two main revenue streams: aeronautical and non-aeronautical.
Aeronautical revenue includes landing fees, terminal rentals, and passenger facility charges. Non-aeronautical revenue comes from concessions, parking, rental car operations, and real estate development. See Appendix I for your assigned airport's financials.

To uniquely identify each airport globally, there are two identifiers to code airports.
The ICAO identifier, which is a four-letter code, and the IATA identifier, a three-letter code.Most passengers are familiar with the IATA identifier, since this the code displayed on baggage tabs and passengers ticketing (see Exhibit B).

Airports strategically Leverage their IATA codes-simple three-letter identifiers likeLAX, JFK, or SFO-as powerful branding tools to enhance recognition and attract more passengers. These codes are prominently displayed on signage, marketing materials, and even merchandise, making them instantly recognizable and easy to remember. By embedding IATA codes into wayfinding signs, promotional campaigns, and digital content, airports create a strong local identity and global presence. This branding encourages travelers to associate the code with efficiency, convenience, or destination appeal, influencing airline route decisions and passenger choice, and ultimately driving more traffic through their facilities.

FLIGHT DELAYS

Flights are considered delayed when they arrive 15 or more minutes after the scheduled time (BTS, n.d.-a). Flight delays are an inconvenience for passengers and increase the operating costs of an airline. When a flight arrives late, it impacts the entire transportation system. Passengers lose time, they may miss a connection flight, and sometimes have to pay out pocket expenses to cover for hotels or meals while waiting for the next available flight. Airlines, on the other hand, need to also deal with the logistical challenges of rebooking passengers, rescheduling crew, and dealing with additional fuel consumption.

These additional challenges come with a serious price tag for airlines. Crew members have strict work hour limits, and when delays push them past their scheduled shifts, airlines either pay overtime or find replacements, which is expensive. Fuel costs skyrocket when planes are forced to wait in taxiways, circle in holding patterns, or take longer routes because of congestion. When delays become extreme, airlines have to compensate passengers with meal vouchers, hotel stays, or rebooking.

The US Department of Transportation records minutes of delay for five different causes of flight arrival delays: carrier, weather, NAS, security, and late arrival (FAA, n.d.-d).

- **Carrier delay:** A delay that is within the control of the airline. Some examples include aircraft cleaning, aircraft damage, awaiting the arrival of connecting passengers or crew, fueling, maintenance, oversales, slow boarding or sitting, or removal of unruly passengers.
- **Weather delay:** Caused by extreme or hazardous weather conditions that are forecasted or manifest themselves on point of departure, enroute, or on point of arrival.
- **National Air System (NAS) delay:** a delay that is within the control of the NAS.It can be due to non-extreme weather conditions, airport operations, heavy traffic volume, or air traffic control.
- **Security delay:** Caused by evacuation of a terminal or concourse, re-boarding of aircraft because of security breach, and inoperative screening equipment, among others.
- **Late arrival delay:** Happens when the aircraft is arriving late because of the late arrival of the same plane at the previous airport. This effect is usually referred to as delay propagation.

AIRPORTS, HUBS, RUNWAYS, AND DELAYS

Figures 2 and 3 display the busiest U.S. airports for domestic flights, by total passengers and by total flights, respectively. Your assigned airport is one of the highlighted in red.

A hub, in aviation, refers to an airport that serves as a central connecting point for flights, where passengers can transfer between multiple domestic and international destinations (CTI, n.d.). For example, since DFW is American Airlines main hub, if someone wants to flight from Nashville (BNA) to Tokyo (HND) with American Airlines, a very likely route would be BNA-DFW, a layover in Dallas, and then DFW-HND. You can see which airlines have a hub in your assigned airport in Appendix II.

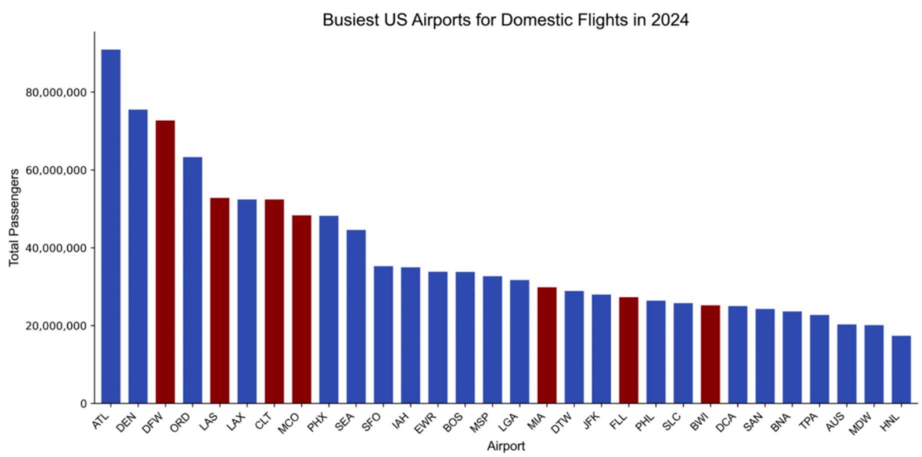

Figure 2. Busiest US airports for domestic flights, by total passengers.

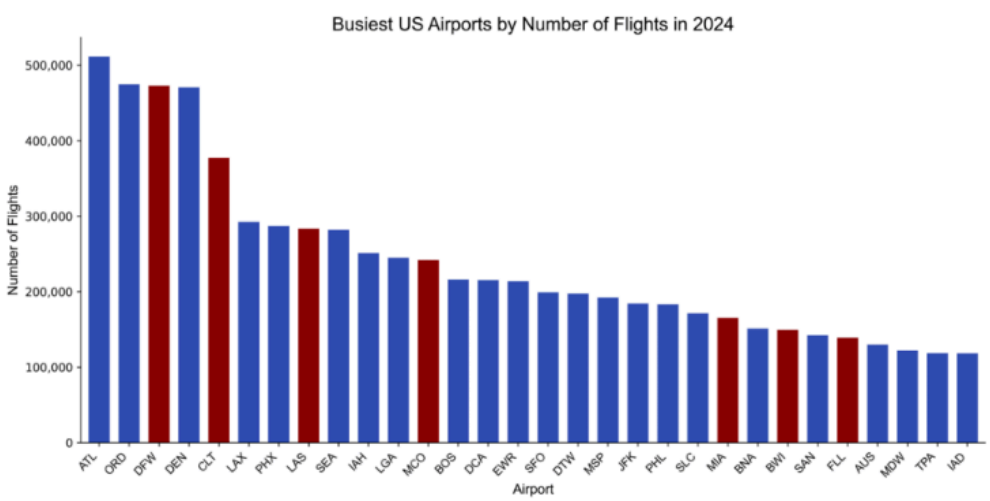

Figure 3. Busiest US airports for domestic flights, by number of flights.

In aviation, heading refers to where the aircraft is pointing relative to the magnetic north, measured in degrees from 0 and 360. A heading of 0 or 360 means the airport is facing north, 90 is east, 180 is south, and 270 is west.

Runways are named between 01 and 36, which is the runway's heading rounded to the nearest 10 degrees. For example, a runway aligned at 87° would be named Runway 09, while one at 274° would be Runway 27. Since aircraft can take off and land from either direction, each runway has two names, one for each approach. At large airports with parallel runways, they are labeled with "L" (left), "C" (center), and "R" (right) to distinguish them (Pilot Institute, 2025). You can see your assigned airport's map, with their runways and terminals, in Appendix III.

Flight delays can be analyzed from two different perspectives. Departure delays occur when an aircraft fails to leave the origin airport within a specified time of its schedule departure, typically attributed to factors controlled by the departing airport. Arrival delays happen when the plane arrives more than 15 minutes late. However, arrival delays are usually not the arrival airport's responsibility. For example, if a plane departs very delayed from an airport, it will arrive late. So, it will count as "arrival delay" even though the delay was originated somewhere else.

Figure 4 displays how the top 30 airports in the United States are in terms of departure delays and arrival delays. Your assigned airport has a lot of traffic but that does not have good on-time performance. However, other airports with even more traffic, like ATL, are able to maintain way better on-time performances.

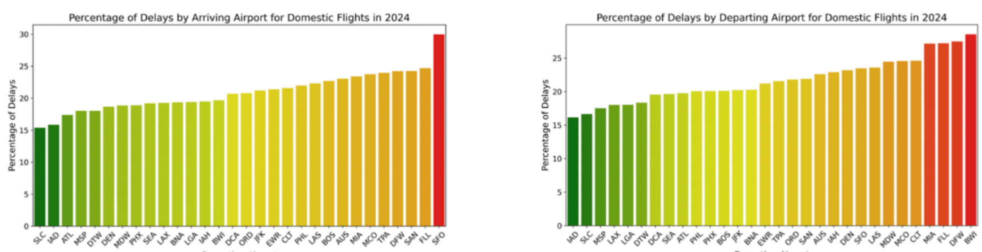

Figure 4. Percentage of delays by arriving and departing airport, respectively, of the top 30 American airports, for domestic flights.

Controlling these delays in the airport is not easy, since it's usually not the airport's fautt. The only type of delay that can be "attributable" to the airport's efficiency management is the NAS delay, and noffentirely. And this accounts for a very small portion of the minutes delayed, as it can be seen in Exhibit D.

PREDICTING DELAYS USING MACHINE LEARNING

Airports have a vested interest in keeping passengers moving and happy while passing through their facilities. Often airports have no control on flight delays, but they may be able to improve the passenger experience and, hence, increase revenue when airports are able to predict when and which flights can be delayed.

You are a business analyst of the Analytics Innovation Team at a major airport. Your team was formed six months ago as part of the airport's Strategic Plan 2025 initiative to leverage advanced analytics for operational and financial improvements.

You team collected airport data for 2023 and 2024, including scheduled times, actual times of departure and arrival, airline operating the flight, flight number, origin and destination, weather, and information about the plane flying the route.

Additionally, you have data access to the number of passengers that went in and out of a plane at the airport since 2000. With such data you can forecast the number of passengers for the next year. See data dictionary in Exhibit E.

How do delays affect customer experience, retail spending, and airport efficiency?
Can identifying high-risk flights help optimize gate assignments or even create premium services that generate additional revenue? More importantly, how can predictive analytics be leveraged as a strategic advantage rather than just an operational tool?

EXHIBIT A

There are two main ways to measure the size of an airline. First, the number of passengers carried in a year; second, the number of flights per airline in a year. For 2024, these two ways are displayed in Figures A1 and A2.

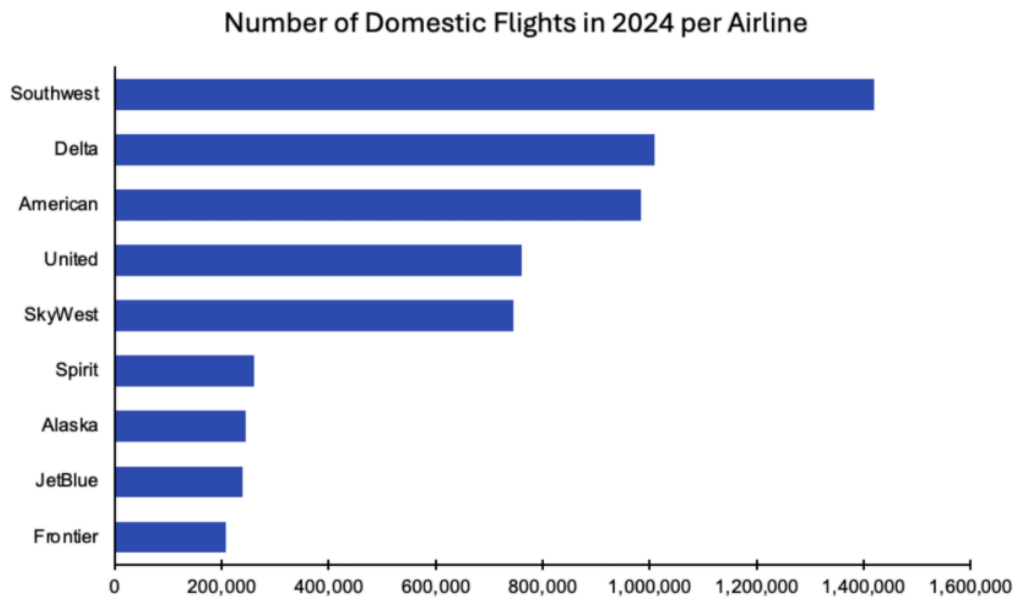

Figure A1. Number of US domestic flights, during 2024, per airline.

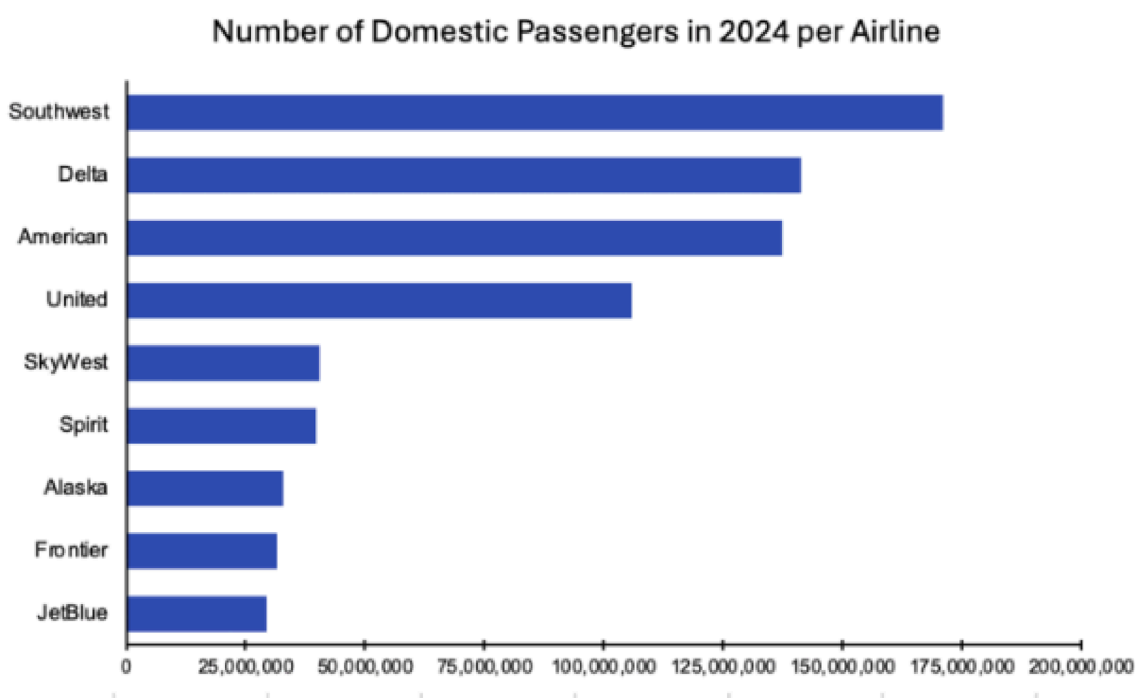

Figure A2. Number of US domestic passengers carried by the main airlines.

EXHIBIT B

Figure B1. Baggage tag with the airports of Hamburg (HAM), Istanbul (IST), and Tocumen, Panama City (PTY) (Travel Dealz, 2023).

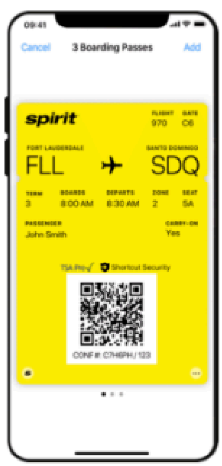

Figure B2. Boarding pass for a Spirit Airline flight between Fort Lauderdale International Airport (FLL) and Las Américas International Airport, Santo Domingo (SDQ) (Spirit, n.d.).

EXHIBIT C

Aircraft used in commercial aviation are generally categorized into narrow-body and wide-body planes. Narrow-body planes have a single aisle and are used for domestic or regional flights. In contrast, wide-body aircraft have two aisles and are designed for longer routes. Sometimes, wide-body cover domestic flights, for high-demand routes, because of their higher capacity.

The main aircraft manufacturers are Boeing (US), Airbus (Europe), Embraer (Brazil), and Bombardier (Canada). Airbus and Boeing dominate the short- and long-haul markets, while Embraer and Bombardier specialize in smaller regional jets.

The Canadair Regional Jets (CRJ) family, from Bombardier, has the CRJ-100, 200, 440, 550, 700, 705, 900, and 1000. They are all regional jets with capacities between 50 and 100 passengers.

The Embraer Regional Jets (ERJ) family includes the ERJ-135, 140, and 145, all regional jets with capacities below 50 passengers. Embraer also manufactures the E-Jet family, with the models E170, E175, E190, and E195, with capacities between 50 and 100 passengers and a longer range than the ERJ family.

Out of the regional jets, the two major players in the aircraft manufacturing industry are Boeing and Airbus. For short-haul flights, the main airplanes are the B737 family and the A320 family, along with the next-generation models, the "MAX" family and the "neo" family for the B737s and the A320s, respectively. The B757 family is a narrow-body, medium-haul, higher capacity plane, very common for routes that go from coast to coast. Additionally, some B717 are still flying, which is the regional plane of Boeing.

The wide-body planes are also used for domestic flights, on certain occasions.Boeing has the B767, B777, and B787 families, while Airbus has the A330 and A350 family.

Also, Bombardier originally developed a short-haul, narrow-body family of planes, the C-series. But this was later acquired by Airbus and it was rebranded to what we know today as the A220 family (Hartley, 2024).

Model	Range (km)	Capacity	Period
B717-200	2,620	100-120	1999-2006
B737-700	5,575	126-149	1997-2020
B737-800	5,436	162-189	1998-2020
B737-900	5,463	178-220	2001-2019
B737-MAX8	6,480	162-210	2017-pr
B737-MAX9	6,110	178-220	2018-pr
B757-200	7,572	200-228	1983-2005
B757-300	5,287	243-280	1999-2005
B767-300	10,002	218-350	1986-2014
B767-400	9,688	245-375	2000-2009
B777-200	10,730	276-364	1995-2021
B777-300	14,685	290-304	1998-2024
B787-8	13,530	210-248	2011-pr
B787-9	14,010	250-296	2014-pr
B787-10	11,730	300-336	2018-pr

Model	Range	Capacity	Period
A220-100	6,390	100-135	2018-pr
A220-300	6,297	120-160	2018-pr
A319-100	6,950	110-156	1996-2021
A320-200	6,200	140-180	1988-2020
A321-200	5,950	170-220	1994-2021
A320neo	6,300	150-194	2016-pr
A321neo	7,400	180-244	2017-pr
A330-200	15,094	220-406	1998-2019
A330-300	11,750	250-440	1994-2020
A330-900	13,334	260-465	2018-pr
A350-900	15,742	300-440	2014-pr
A350-1000	16,482	350-480	2018-pr

Table C1. Airbus and Boeing commercial aircraft capacity, range, and delivery period. There are "ER" versions for some models, this is, extended range. The indicated ranges are just an idea. All data extracted from Boeing (n.d.) and Airbus (n.d.).

In the datasets, multiple planes appear. In tables C1 and C2, basic information is displayed about these planes' typical range, normal capacity, and delivery period.

Model	Range (km)	Capacity	Period
CRJ100	2,417	50	1992-2006
CRJ200	2,275	50	1994-2006
CRJ440	2,275	44	1994-2006
CRJ550	1,850	50	2001-2020
CRJ700	2,650	66-78	2001-2020
CRJ705	3,140	75	2001-2020
CRJ900	2,500	81-90	2003-2021
CRJ1000	3,050	97-104	2010-2020

Model	Range	Capacity	Period
ERJ135	3,100	37	1999-2005
ERJ140	3,000	44	2001-2003
ERJ145	2,000	50	1996-2011
E170	3,100	70-78	2004-pr
E175	2,800	78-86	2005-pr
E190	3,200	94-106	2005-pr
E195	2,650	106-118	2006-pr

Table C2. Bombardier and Embraer commercial aircraft capacity, range, and delivery period. There are "ER" versions for some models, this is, extended range. The indicated ranges are just an idea. All data extracted from Embraer (n.d.) and Airlines Inform (n.d.).

EXHIBIT D

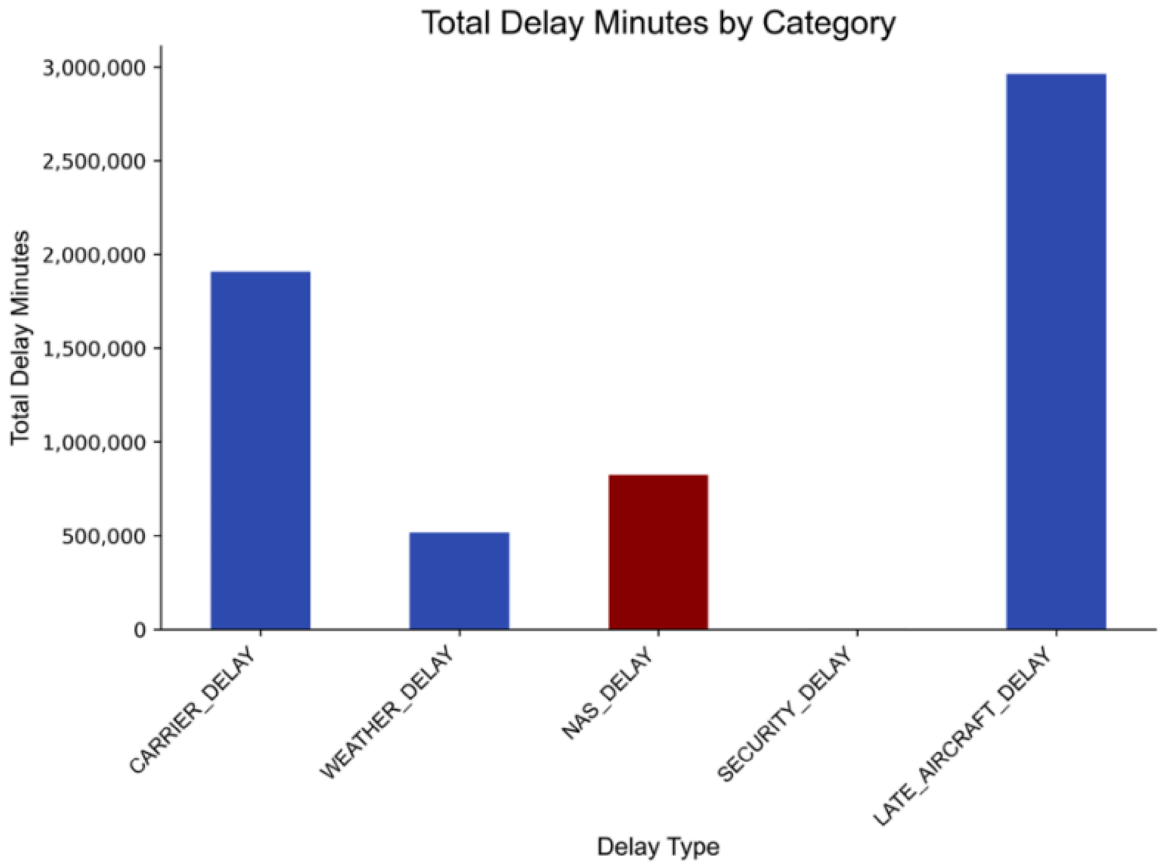

Figure D1. Total delay minutes by category, in 2024 domestic flights.

EXHIBIT E

You will be given access to five files. The first one will be called CDE _data, osv, where CDE is the IATA code of the assigned airport. The structure of the dataset, as well as an explanation of its columns, is detailed in Table E1. This data was adapted from BTS (n.d.-b) and National Weather Service (n.d.).

YEAR	Year of the flight
QUARTER	1: Jan 1 to Mar 31:2: Apr 1 to Jun 30; 3: Jul 1 to Sep 30; 4: Oct 1 to Dec 31.
MONTH	Numeric code of the month (e.g. January is 1).
DAY_OF_MONTH	Day of the month.
DAY_OF_WEEK	Numeric code of the day of the week (a-g., Monday is 1).
FL_DATE	Flight date in *YYYY-MM-DD" format.
MKT_UNIQUE_ CARRIER	JATA code of the marketing airline, i.e., the airline that sells the ticket.
OP_UNIQUE_ CARRIER	IATA code of the operating airline, t.e., the airline that operates the flight (provides the aircraft and crew).
TAIL_NUM	Tail number of the aircraft that operated the flight (the license plate equivalent of a plane).
OP_CARRIER_FE_NUM	Operating carrier flight number (only the numbers).
ORIGIN_ AIRPORT_ID	An identification number assigned by US DOT to identify a unique airport.
ORIGIN AIRPORT SEQ_ID	An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time.
ORIGIN_CITY_MARKET_ID	City Market ID is an identification number assigned by US DOT to identify a city market.
ORIGIN	IATA code of the origin airport.
ORIGIN_CITY_NAME	Origin city name.
ORIGIN_STATE_ABR	Origin state two-letter abbreviation (for example, for DFW it would be TX).

DEST_AIRPORT_ID	An identification number assigned by US DOT to identity a unique airport.
DEST_AIRPORT_SEQ_ID	An identification number assigned by US DOT to identity a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time.
DEST CITY MARKET	City Market ID is an identification number assigned by US DOT to identify a city market.
DEST	JATA code of the destination airport.
DEST_CITY_NAME	Destination city name.
DEST_STATE_ABR	Destination state two-letter abbreviation (for example, for DFW it would be TX)
CRS_DEP_TIME	Scheduled departure time (local time).
DEP_TIME	Actual departure time (local time).
DEP_DELAY	Difference in minutes between scheduled and actual departure time. Early departures show negative numbers.
DEP_DELAY_NEW	Difference in minutes between scheduled and actual departure time. Early departures set to 0.
DEP_ DEL15	Departure delay indicator, 15 minutes or more (1=Yes).
TAXI_OUT	Taxi-out time, in minutes. Duration between gate out time and take off (wheels off) time.
TAXI_IN	Taxi-in time, in minutes. Duration between landing (wheels on) time and gate in time
CRS_ ARR_TIME	Scheduled arrival time (local time).
ARR_TIME	Actual arrival time (local time).
ARR_DELAY	Difference in minutes between scheduled and actual arrival time. Early arrivals show negative numbers.
ARR_DELAY NEW	Difference in minutes between scheduled and actual arrival time. Early arrivals set to 0.
ARR DEL15	Arrival delay indicator, 15 minutes or more (1=Yes).

CANCELLED	Cancelled flight indicator (1=Yes).
CANCELLATION_CODE	Specifies the reason for cancellation. “-” is unknown, “A” is carrier, “B” is weather, "C" is NAS, “D” is security
DIVERTED	Diverted flight indicator (1=Yes).
DISTANCE	Distance between airports (miles).
CARRIER_DELAY	Carrier delay, in minutes.
WEATHER DELAY	Weather delay, in minutes.
NAS DELAY	NAS delay, in minutes.
SECURITY DELAY	Security delay, in minutes.
LATE AIRCRAFT DELAY	Late aircraft delay, in minutes.
MANUFACTURER	Operating aircraft manufacturer. "Airbus", *Boeing*. *Bombardier*, or *Embraer*.
MODEL	Operating aircraft model.
AGE FLIGHT	Age of the plane at the moment of the flight.
ROUTE	Route of the flight, with the IATA codes of the airports (e.g., DFW-LAX).
MODEL_GEN	Operating aircraft general model.
AIRLINE FULL NAME	Operating carrier full name.
DAY_OF_WEEK FULL	Date of week full name.
DFW_date	The timestamp (year, month, day, and hour corresponding to the scheduled arrival or departure time at the assigned airport. If the assigned airport is the destination, this reflects the arrival time; if it is the origin, this reflects the departure time.
Vis	The recorded visibility at the assigned airport at the time of arrival (if it is the destination) or departure (if it is the origin), measured in miles (mi).
Prec	One hour precipitation at the assigned airport at the time of arrival (if it is the destination) or departure (if it is the origin), measured in inches (in).

Windspd	Wind speed at the assigned airport at the time of arrival (fit is the destination or departure (if it is the origin), measured in miles per hour (mph).
Gust	Maximum wind gusts recorded at the assigned airport at the time of arrival (if it is the destination) or departure (if it is the origini, measured in miles per hour (mph). If there were no gusts, the value is *NaN"
WindDir	Direction of the wind at the assigned airport at the time of arrival (if it is the destination) or departure (if it is the origin), measured in degrees (°).
Temp	The recorded temperature at the assigned airport at the time of arrival (if it is the destination) or departure (if it is the origin), measured in degrees Fahrenheit (PF).

TableE1.CDE_data.csvcolumn structure and description.

You will also be given access to 2023_data.csv and 2024_data.csv. These files contain all domestic commercial flights in 2023 and 2024, respectively. The structure Is the same as in Table E1, excluding the last seven columns, that are weather-related.

While these files will not have to be used for the Machine Learning models, they allow you to compare your assigned airport's performance metrics against other airports. The dataset CDE_data. cs has been obtained by filtering these two datasets and adding weather data.

Regarding the number of passengers, you will be provided with I100. csv, which contains monthly passenger counts per route and per airline for all U.S. domestic flights. This will be your primary data source for forecasting passenger numbers. It includes data from 2000 to 2024 and an explanation of its columns is detailed in Table E2.

Column Name	**Description**
PASSENGERS	Number of passengers enplaned.
FREIGHT	Freight enplaned (in pounds).
MAIL	Mail enplaned (in pounds).
DISTANCE	Distance between airports (in miles).
UNIQUE_CARRIER	IATA code of the airline, if it exists. If not, it uses a different unique carrier code.
AIRLINE_ID	An identification number assigned by US DOT to identify a unique airline (carrier).
UNIQUE_ CARRIER_NAME	Full name of the airline.
CARRIER	IATA code of the airline, if it exists. If not, it uses a different unique carrier code.
ORIGIN_AIRPORT_ ID	An identification number assigned by US DOT to identify a unique airport.
ORIGIN_AIRPORT_ SEQ_ID	An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time.
ORIGIN_CITY_MARKET_ ID	City Market ID is an identification number assigned by US DOT to identify a city market.
ORIGIN	IATA code of the origin airport

ORIGIN_CITY_NAME	Origin city name
ORIGIN_STATE_ ABR	Origin state two-letter abbreviation (for example, for DFW it would be TX).
DEST_AIRPORT_ ID	An identification number assigned by US DOT to identify a unique airport.
DEST_ AIRPORT_SER_ID	An identification number assigned by US DOT to identify a unique airport at a given point of time. Airport attributes, such as airport name or coordinates, may change over time.
DEST_CITY_MARKET_ID	City Market ID is an identification number assigned by US DOT to identify a city market.
DEST	IATA code of the destination airport.
DEST_CITY_NAME	Destination city name.
DEST_STATE_ABR	Destination state two-letter abbreviation (for example, for DFW it would be TX).

Table E2. `T100.csv` column structure and description.

Finally, CDE_Appendix. pdf will contain specific information to your assigned airport. Again, CDE is the IATA code of the assigned airport.