

预测航班延误：机场如何通过预测分析增加收入？

美国国内航空业

美国国内航空业是世界上最强大、最具经济意义的运输行业之一。每年通过美国机场的国内乘客超过 8 亿人次（Statista，2024 年）。美国每天平均有 45,000 个国内航班，占全球每天 106,000 个航班的近一半（FAA，n.d.-a；OAG，n.d.）。

国内航空运输量创造了数千亿美元的经济活动，并为全国提供了数百万个工作岗位。商业航空占美国国内生产总值的 5%（《美国航空》，2025 年）。在过去五年中，国内航空业收入的年复合增长率为 20.6%，预计 2025 年将达到 2,045 亿美元（Samorajski，2025 年）。

美国航空系统在复杂的监管框架下运行。联邦航空管理局（FAA）成立于 1958 年，隶属于交通部，负责管理美国的天空。他们负责指挥空中交通、机场安全和监管飞行检查标准（FAA，n.d.-b；FAA，n.d.-c）。

与许多国际机场不同，美国机场通常在地方政府部门的监督下运营。大多数大型商业机场都归县、市或地区机场管理局所有。虽然美国联邦航空局提供监管监督和安全标准，但日常管理决策仍由地方机场当局负责。

美国主要航空公司

美国航空业高度细分。其中有超低成本航空公司（ULCC）、低成本航空公司（LCC）、传统航空公司、地区性航空公司以及

"启动"航空公司（见图 1）。

传统航空公司也称为网络航空公司，包括美国航空公司、达美航空公司和美国联合航空公司。这些航空公司运营着广泛的国内和国际航班网络，拥有多种服务等级、多样化的机队、强大的常旅客计划，并分别隶属于全球航空联盟--寰宇一家、天合联盟和星空联盟--。



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

低成本航空公司以西南航空公司为首，按载客量和航班量计算，西南航空公司是全美最大的国内航空公司（见附图 A）。捷蓝航空和阿拉斯加航空也属于这一类。这些航空公司的运营成本通常低于传统航空公司，而且通常以国内和短途国际航班为主。

超低成本航空公司，如 Spirit 航空公司、Frontier 航空公司和 Allegiant 航空公司

以业内最低的成本结构运营。这些航空公司的票价模式是，乘客只需支付基本的运输服务费用，而所有的便利设施，如座位选择、行李、机上小吃和茶点等，则需额外付费。

支线航空公司运营较小的飞机。天西航空公司（SkyWest）、共和航空公司（Republic Airways）和梅萨航空公司（Mesa Airlines）等公司以美鹰航空公司（American Eagle）、达美联运航空公司（Delta Connection）和美联航快运航空公司（United Express）等品牌运营，将较小的社区与主要枢纽连接起来。

初创 "航空公司包括 Avelo Airlines 和 Breeze 等新加入的航空公司

航空公司。这些航空公司在二级市场之间确定航线，通常没有什么直接竞争。

美国机场收入

机场有两大收入来源：航空收入和非航空收入。

航空收入包括着陆费、航站楼租金和客运设施费。非航空收入来自特许经营、停车场、租车业务和房地产开发。请参阅附录 I 了解您所在机场的财务状况。

为了在全球范围内唯一标识每个机场，有两种标识符对机场进行编码。

大多数乘客都熟悉国际航空运输协会（IATA）的标识符，因为行李标签和乘客票据上显示的就是这个代码（见附图 B）。

机场战略性地利用其国际航空运输协会（IATA）代码--简单的三个字母标识符，如机场代码（LAX）、肯尼迪机场代码（JFK）或旧金山机场代码（SFO）--作为强大的品牌工具，以提高知名度并吸引更多乘客。这些代码醒目地显示在标识牌、营销材料甚至商品上，让人一目了然，易于记忆。通过将 IATA 代码嵌入路标、促销活动和数字内容中，机场可以创建强大的本地标识和全球影响力。这种品牌形象鼓励旅客将代码与效率、便利性或目的地吸引力联系起来，从而影响航空公司的航线决策和旅客选择，最终为机场带来更多客流量。

航班延误

航班比预定时间晚 15 分钟或更长时间到达，即被视为延误（BTS, n.d.-a）。航班延误会给乘客带来不便，并增加航空公司的运营成本。航班晚点会影响整个运输系统。乘客会损失时间，可能错过转机航班，有时在等待下一班航班时还需要自掏腰包支付酒店或餐饮费用。另一方面，航空公司还需要应对重新预订乘客、重新安排机组人员以及处理额外燃料消耗等后勤挑战。

这些额外的挑战给航空公司带来了严重的代价。机组人员有严格的工作时间限制，当延误导致他们超过预定的轮班时间时，航空公司要么支付加班费，要么寻找替代人员，而这都是昂贵的。当飞机被迫在滑行道上等待、在滞留模式中盘旋或因拥堵而走更长的路线时，燃油成本就会飙升。当延误达到极致时，航空公司不得不用餐券、酒店住宿或重新订票来补偿乘客。

美国运输部记录了造成航班抵达延误的五种不同原因的延误分钟数：承运人、天气、NAS、安全和晚到（FAA, n.d.-d）。

- **承运人延误**：航空公司可控制的延误。例如，飞机清洁、飞机损坏、等待转机乘客或机组人员到达、加油、维修、超售、登机或就座缓慢，或驱赶不守规矩的乘客。
- **天气延误**：在出发地、途中或到达地预报或显示的极端或危险天气条件造成的延误。
- **国家航空系统（NAS）延误**：在国家航空系统控制范围内的延误。其原因可能是非极端天气条件、机场运行、交通流量大或空中交通管制。
- **安全延误**：航站楼或大厅疏散、因安全漏洞而重新登机、安检设备失灵等造成的延误。
- **晚到延误**：由于同一架飞机在前一个机场晚到，导致飞机晚到。这种效应通常被称为延误传播。

机场、枢纽、跑道和延误

图 2 和图 3 显示了美国国内航班最繁忙的机场，分别按乘客总数和航班总数计算。您的指定机场是红色突出显示的机场之一。

在航空领域，枢纽指的是作为以下活动的中心连接点的机场航班，乘客可在多个国内和国际目的地之间转机（CTI, n.d.）。例如，由于 DFW 是美国航空公司的主要枢纽，如果有人想乘坐美国航空公司的航班从纳什维尔 (BNA) 飞往东京 (HND)，很可能的路线是 BNA-DFW，在达拉斯中转，然后 DFW-HND。您可以在附录 II 中查看哪些航空公司在您指定的机场设有枢纽。

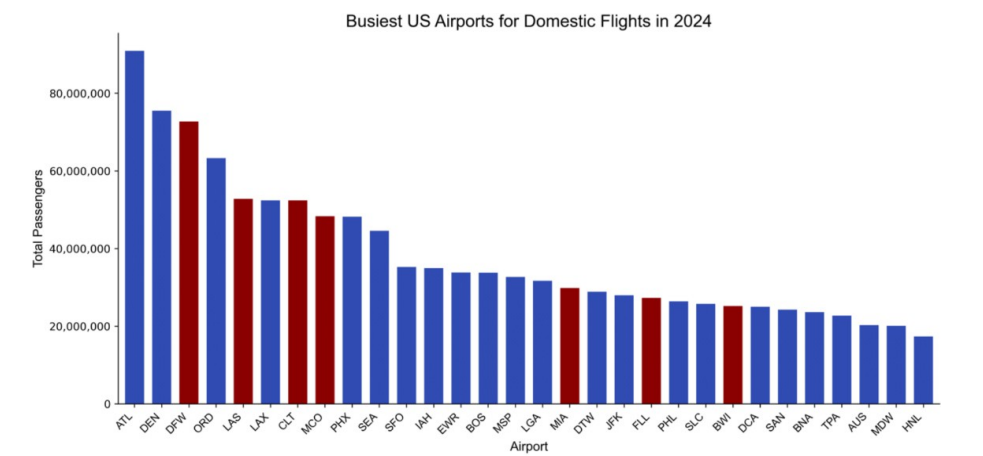


Figure 2. Busiest US airports for domestic flights, by total passengers.

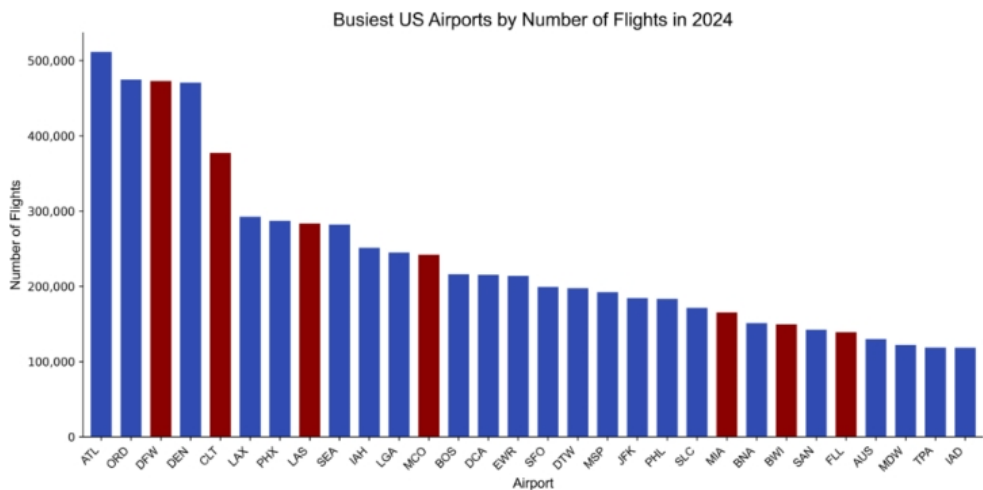


Figure 3. Busiest US airports for domestic flights, by number of flights.

在航空领域，航向是指飞机相对于磁场的指向。

北，以 0 和 360 度为单位。航向为 0 或 360 表示机场朝北，90 表示朝东，180 表示朝南，270 表示朝西。

跑道的名称介于 01 和 36 之间，是将跑道的航向四舍五入到最接近的 10 度。例如，一条朝向 87° 的跑道被命名为跑道 09，而朝向 274° 的跑道则被命名为跑道 27。由于飞机可以从任一方向起飞和降落，因此每条跑道都有两个名称，每个进场方向一个。在有平行跑道的大型机场，跑道上会分别标有 "L"（左）、"C"（中）和 "R"（右），以示区分（飞行员协会，2025 年）。您可以在附录 III 中查看分配给您的机场地图，以及机场跑道和航站楼。

航班延误可以从两个不同的角度进行分析。出发延误是指飞机未能在规定时间内离开出发机场，通常是由于出发机场控制的因素造成的。当飞机晚点 15 分钟以上时，就会发生到达延误。不过，到达延误通常不是到达机场的责任。例如，如果飞机从某机场起飞时延误很长时间，它就会晚点到达。因此，即使延误源于其他地方，也会被算作 "到达延误"。

图 4 显示了美国排名前 30 位的机场在出发延误和到达延误方面的情况。您的指定机场交通繁忙，但准点率并不高。然而，其他交通量更大的机场（如 ATL）却能保持较好的准点率。

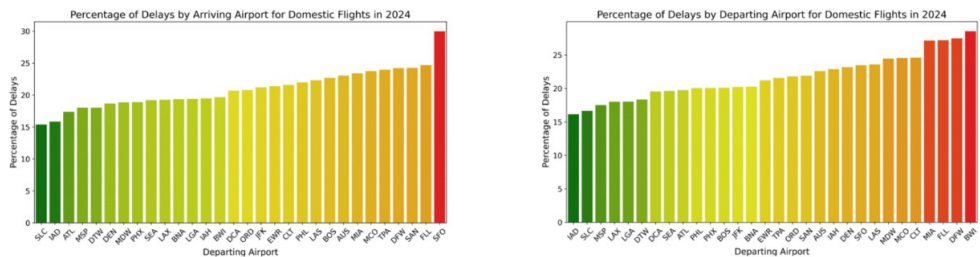


Figure 4. Percentage of delays by arriving and departing airport, respectively, of the top 30 American airports, for domestic flights.

在机场控制这些延误并不容易，因为这通常不是机场的责任。

fautt.唯一可以 "归因于 "机场效率管理的延误类型是 NAS 延误，而且完全不是。从附表 D 中可以看出，这只占延误分钟数的很小一部分。

利用机器学习预测延迟

机场的既得利益在于让乘客在通过其设施时保持愉快的心情。通常情况下，机场无法控制航班延误，但他们可以改善

当机场能够预测航班何时以及哪些航班会延误时，就能提升乘客体验，从而增加收入。

你是一家大型机场分析创新团队的业务分析师。您的

该团队成立于六个月前，是机场《2025 战略计划》倡议的一部分，旨在利用先进的分析技术改善运营和财务状况。

您的团队收集了 2023 年和 2024 年的机场数据，包括计划时间、实际起飞和到达时间、运营航班的航空公司、航班号、出发地和目的地、天气以及执飞该航线的飞机信息。

此外，您还可以获得 2000 年以来在机场进出飞机的乘客人数数据。有了这些数据，您就可以预测下一年的乘客人数。请参见附录 E 中的数据字典。

航班延误对客户体验、零售消费和机场效率有何影响？

识别高风险航班是否有助于优化登机口分配，甚至创造能带来额外收入的优质服务？更重要的是，如何将预测分析作为一种战略优势而不仅仅是一种运营工具加以利用？

附录 A

衡量航空公司规模的方法主要有两种。第一，一年的 乘客运输量；第二，每家航空公司一年的航班数量。图 A1 和 A2 显示了 2024 年的这两种方法。

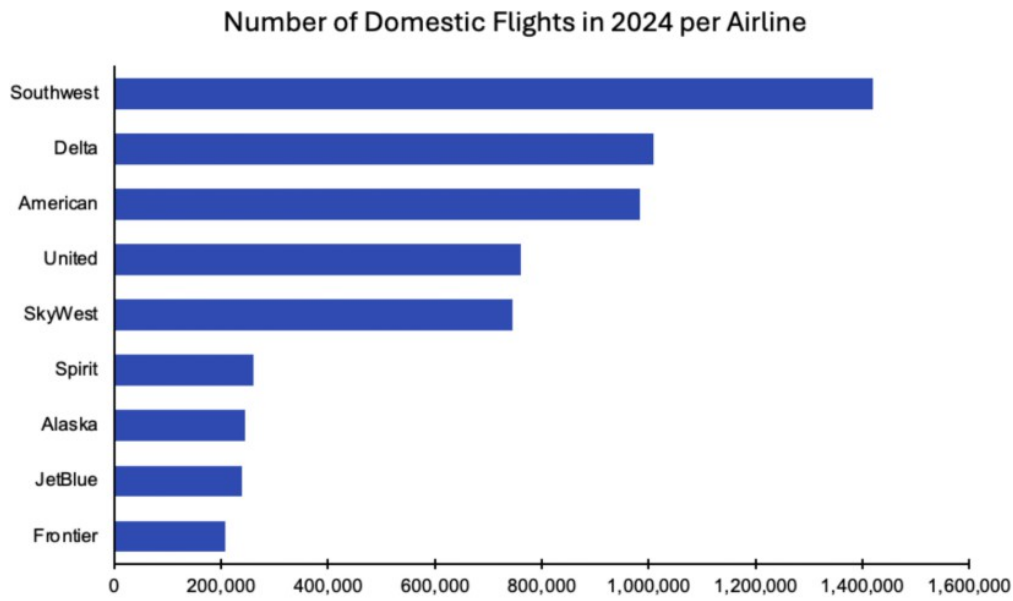


Figure A1. Number of US domestic flights, during 2024, per airline.

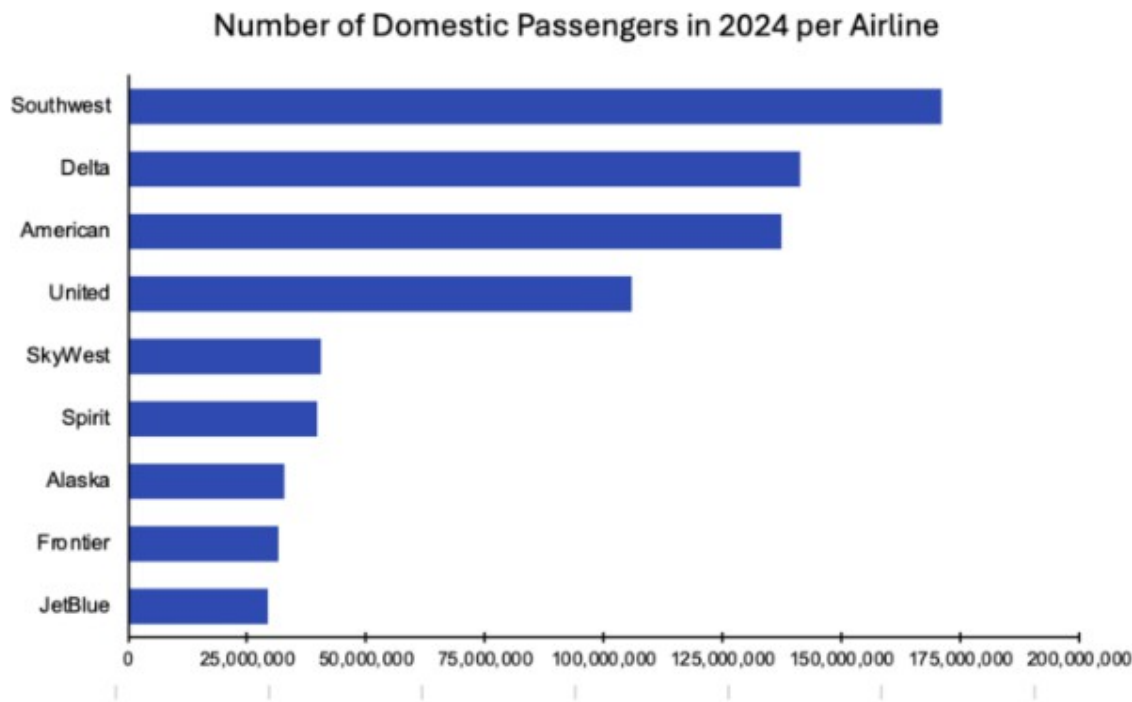


Figure A2. Number of US domestic passengers carried by the main airlines.

附录 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.).

附录 C

用于商业航空的飞机一般分为窄体飞机和宽体飞机。

宽体飞机。窄体飞机只有一条过道，用于国内或地区航班。相比之下，宽体飞机有两个过道，设计用于长途航线。有时，宽体飞机因其运载能力较强，可用于国内航线和高需求航线。

主要的飞机制造商有波音（美国）、空中客车（欧洲）、巴西航空工业公司（巴西）和庞巴迪（加拿大）。空中客车公司和波音公司主导短途和长途市场，而巴西航空工业公司和庞巴迪公司则专门生产小型支线喷气式飞机。

庞巴迪公司的加航支线喷气机（CRJ）系列包括 CRJ-100、200、440、550、700、705、900 和 1000。它们都是支线喷气机，载客量在 50 到 100 人之间。

巴西航空工业公司的支线喷气机（ERJ）系列包括 ERJ-135、140 和 145，都是载客量低于 50 人的支线喷气机。巴西航空工业公司还生产 E-Jet 系列，包括 E170、E175、E190 和 E195 型号，载客量在 50 至 100 人之间，航程比 ERJ 系列更远。

在支线喷气式飞机中，飞机制造业的两大巨头是

是波音公司和空中客车公司。短途航班的主要机型是 B737 系列和 A320 系列，以及 B737 和 A320 的下一代机型 "MAX "系列和 "neo "系列。B757 系列是一种窄体、中程、运力较大的飞机，在从海岸到海岸的航线上非常常见。此外，还有一些 B717 仍在飞行，这是波音公司的支线飞机。

波音公司有 B767、B777 和 B787 系列，空中客车公司有 A330 和 A350 系列。

此外，庞巴迪公司最初开发了一个短途窄体飞机系列，即 C 系列。但后来被空中客车公司收购，并重新命名为今天的 A220 系列（Hartley，2024 年）。

Model	Range (km)	Capacity	Period	Model	Range	Capacity	Period
B717-200	2,620	100-120	1999-2006	A220-100	6,390	100-135	2018-pr
B737-700	5,575	126-149	1997-2020	A220-300	6,297	120-160	2018-pr
B737-800	5,436	162-189	1998-2020	A319-100	6,950	110-156	1996-2021
B737-900	5,463	178-220	2001-2019	A320-200	6,200	140-180	1988-2020
B737-MAX8	6,480	162-210	2017-pr	A321-200	5,950	170-220	1994-2021
B737-MAX9	6,110	178-220	2018-pr	A320neo	6,300	150-194	2016-pr
B757-200	7,572	200-228	1983-2005	A321neo	7,400	180-244	2017-pr
B757-300	5,287	243-280	1999-2005	A330-200	15,094	220-406	1998-2019
B767-300	10,002	218-350	1986-2014	A330-300	11,750	250-440	1994-2020
B767-400	9,688	245-375	2000-2009	A330-900	13,334	260-465	2018-pr
B777-200	10,730	276-364	1995-2021	A350-900	15,742	300-440	2014-pr
B777-300	14,685	290-304	1998-2024	A350-1000	16,482	350-480	2018-pr
B787-8	13,530	210-248	2011-pr				
B787-9	14,010	250-296	2014-pr				
B787-10	11,730	300-336	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.).

在数据集中，出现了多个飞机。在表 C1 和 C2 中，显示了这些飞机的典型航程、正常运力和交货期等基本信息。

Nodef	距离（公里）	容量	期间	诺德尔	范围	容量	期间
CRJ100	2,417	50	1992-2006	ERJ135	3,100	37	1999-2005
CRJ200	2,275	50	1994-2006	ERJ140	3,000	44	2M1-200Z
CRJ460	2,275	44	1994-200B	ERJ145	2,000	50	1996-2011
CRJ550	1,850	50	2001-2020	E170	3,100	70-78	2004-pr
CRJ700	2,650	66-78	2001-2020	E175	2,800	y8-86	200Spr
CRJ705	3,140	75	2001-2020	E190	3,200	94-106	2005-pr
CRJ900	2,500	81-90	2003-2021	E195	2,650	106-118	2006pr
CRJ1000	3,050	97-104	2010-2020				

表 c2.庞巴迪和巴西航空工业公司商用飞机的运力、航程和交付期。某些型号有 "ER "版本，即增程型。所示航程只是一个概念。所有数据摘自 Embreer (n.d.) end Airlines Inform (n.d.)。

附录 D

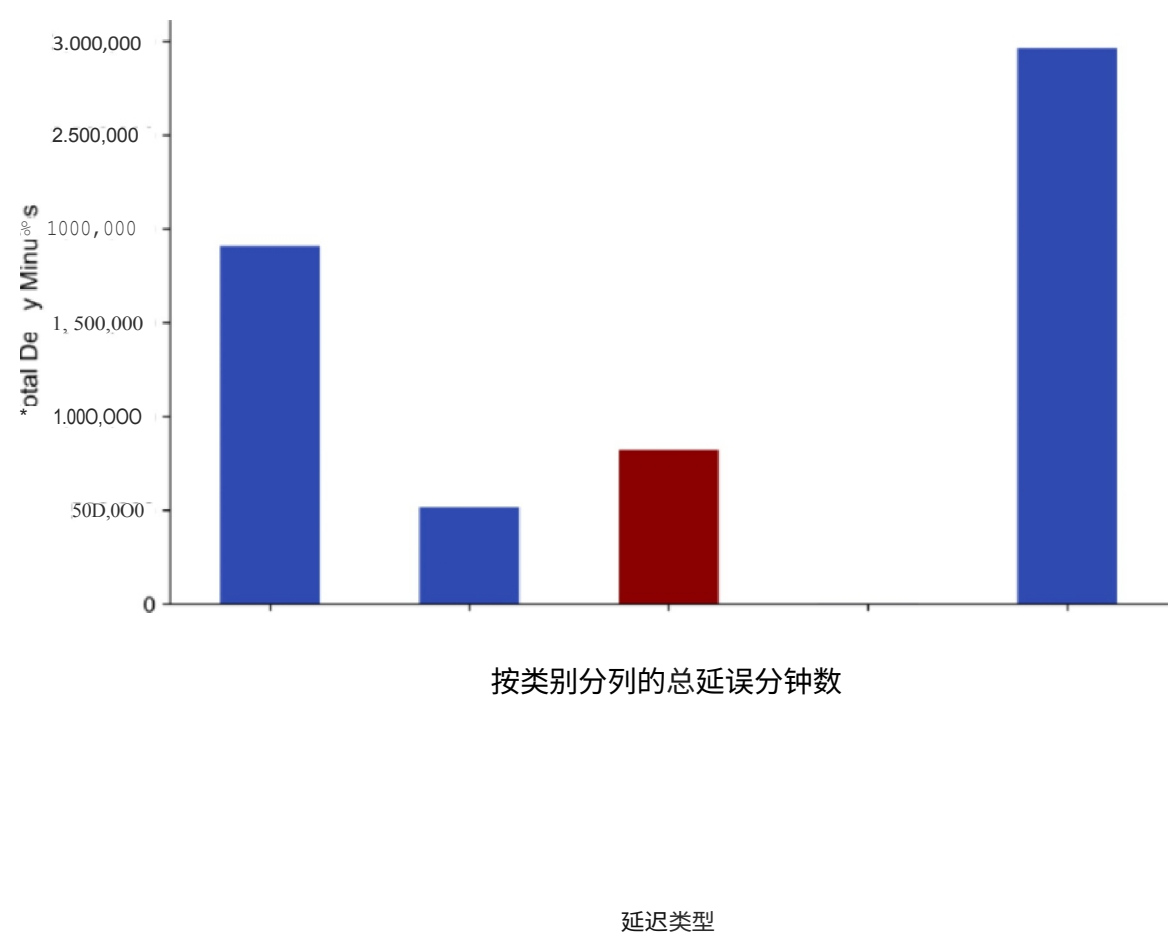


图 D1.按类别分列的 202A 国内航班延误总时长。

附录 E

您将可以访问五个文件。第一个文件名为 CDE_data.osv，其中 CDE 是指定机场的 IATA 代码。数据集的结构及其列的说明详见表 E1。该数据取自 BTS (n.d.-b) 和 National Weather Service (n.d.)。

年	飞行年份
季度	1: 1 月 1 日至 3 月 31 日；2: 4 月 1 日至 6 月 30 日；3: 7 月 1 日至 9 月 30 日；4: 10 月 1 月至 12 月 31 日。
月份	月份的数字代码（如一月为 1）。
月日	每月的哪一天
星期	星期几的数字代码（a-g，星期一为 1）。
FL_DATE	飞行日期，格式为 "*YYYY-MM-DD"。
市场唯一载体	销售航空公司（即出售机票的航空公司）的 JATA 代码。
唯一运算符载体	运营航空公司的 IATA 代码，即运营航班（提供飞机和机组人员）的航空公司。
尾号	执行飞行任务的飞机的尾号（相当于飞机的牌照）。
运载火箭编号	运营航空公司航班号（只有数字）。
origin_airport_id	由美国交通部分配的认识号码，用于识别一个独特的机场。
origin airport seq_id	美国交通部分配的一个标识号，用于在特定时间点标识一个唯一的机场。机场属性（如机场名称或坐标）可能会随着时间的推移而改变。
原产地城市市场代码	城市市场 ID 是美国交通部分配的一个认识号码，用于识别城市市场。
起源	始发机场的 IATA 代码。
城市名称	城市名称的由来。
原籍国	原籍州的双字母缩写（例如，DFW 就是 TX）。

dest_airport_id	由美国交通部分配的认识号码，用于识别一个独特的机场。
dest_airport_seq_id	由美国交通部分配的一个标识号，用于在特定时间点标识一个唯一的机场。 。机场属性（如机场名称或坐标）可能会随着时间的推移而改变。
毁灭城市市场	城市市场 ID 是美国交通部分配的一个认识号码，用于识别城市市场。
DEST	目的地机场的 JATA 代码。
DEST_CITY_NAME	目的地城市名称。
DEST_STATE_ABR	目的地州的双字母缩写（例如，如果是 DFW，则为 TX）
CRS_DEP_TIME	预定起飞时间（当地时间）。
DEP_TIME	实际出发时间（当地时间）。
DEP_DELAY	计划起飞时间与实际起飞时间的分钟差。提前出发的航班显示负数。
DEP_DELAY_NEW	计划起飞时间与实际起飞时间的分钟差。提前出发时间设为 0。
DEP_DEL15	出发延迟指示器，15 分钟或以上（1= 是）。
TAXI_OUT	滑出时间（分钟）。出闸时间与起飞（车轮关闭）时间之间的持续时间。
TAXI_IN	出租车进站时间，以分钟为单位。着陆（机轮打开）时间与登机口进站时间之间的持续时间
CRS_ARR_TIME	预定抵达时间（当地时间）。
ARR_TIME	实际抵达时间（当地时间）。
延迟	计划到达时间与实际到达时间的分钟差。提前到达者显示负数。
ARR_DELAY NEW	计划到达时间与实际到达时间之差，以分钟为单位。提前到达设置为 0。
ARR DEL15	到达延迟指示器，15 分钟或以上（1= 是）。

取消	取消飞行指示器（1=是）。
取消代码	说明取消的原因。-"表示未知，"A "表示运营商，"B "表示天气，"C "表示NAS，"D "表示安全
分割	改道飞行指示器（1=是）。
距离	机场之间的距离（英里）。
CARRIER_DELAY	运营商延迟，以分钟为单位。
天气延误	天气延误，以分钟为单位。
NAS 延迟	NAS 延迟，以分钟为单位。
安全延迟	安全延迟，以分钟为单位。
飞机晚点	飞机晚点，以分钟为单位。
制造商	运营飞机制造商。"空中客车"、*波音*。 *投弹手*，或*安布雷拉*。
型号	运行的飞机型号。
年龄飞行	飞行时飞机的年龄。
路线	航班路线，以及机场的 IATA 代码（如 DFW-LAX）。
型号	操作飞机的一般模式。
航空公司全称	运营商全称。
周日满	星期日期 全名。
日期	时间戳（年、月、日和小时），与指定机场的计划到达或起飞时间相对应。如果指定机场是目的地，则反映到达时间；如果是出发地，则反映出发时间。
签证	指定机场在到达（如果是目的地）或离开（如果是出发地）时的能见度记录，单位为英里（mile）。
精密	指定机场在到达（如果是目的地）或离开（如果是出发地）时的一小时降水量，单位为英寸（in）。

Windspd	指定机场在到达（适合目的地）或离开（如果是出发地）时的风速，单位为英里/小时（mph）。
古斯特	指定机场在到达（如果是目的地）或离开（如果是出发地）时记录的最大阵风，单位为英里/小时（mph）。如果没有阵风，数值为*NaN"。
风向标	指定机场在到达（如果是目的地）或离开（如果是出发地）时的风向，单位为度（°）。
温度	指定机场在到达（如果是目的地）或离开（如果是出发地）时的记录温度，单位为华氏度（PF）。

表 E1.CDE_data.csvcolumn 的结构和说明。

您还可以访问 2023_data.csv 和 2024_data.csv。这些文件分别包含 2023 年和 2024 年的所有国内商业航班。结构与表 E1 相同，但不包括与天气有关的最后七列。

虽然这些文件不必用于机器学习模型，但您可以将指定机场的性能指标与其他机场进行比较。CDE_data.cs 数据集是通过过滤这两个数据集并添加天气数据而获得的。

关于乘客人数，您将获得 I100.csv，其中包含所有美国国内航班按航线和航空公司的每月乘客人数。这将是您预测乘客人数的主要数据来源。它包括 2000 年至 2024 年的数据，表 E2 详细说明了其中各列的情况。

列名	说明
乘客	乘客人数。
运输	登机货物（磅）。
邮件	邮件装机量（磅）。
距离	机场之间的距离（英里）。
UNIQUE_CARRIER	航空公司的 IATA 代码（如果存在）。如果没有，则使用不同的唯一航空公司代码。
航空公司标识	由美国运输部分配的认识号码，用于识别一家独特的航空公司（承运商）。
唯一的运营商名称	航空公司全称。
承运商	航空公司的 IATA 代码（如果存在）。如果没有，则使用不同的唯一航空公司代码。
始发机场 ID	由美国交通部分配的认识号码，用于识别一个独特的机场。
origin_airport_seq_id	由美国交通部分配的认识号码，用于在特定时间点识别唯一的机场。机场属性（如机场名称或坐标）可能会随着时间的推移而改变。
原产地城市市场 ID	城市市场 ID 是美国交通部分配的一个认识号码，用于识别城市市场。
起源	始发机场的 IATA 代码

城市名称	原籍城市名称
原籍国	原籍州的双字母缩写（例如，DFW 就是 TX）。
目的地机场 ID	由美国交通部分配的认识号码，用于识别一个独特的机场。
dest_airport_ser_id	美国交通部分配的一个标识号，用于在特定时间点标识一个唯一的机场。机场属性（如机场名称或坐标）可能会随着时间的推移而改变。
目的地城市市场 ID	城市市场 ID 是美国交通部分配的一个认识号码，用于识别城市市场。
DEST	目的地机场的 IATA 代码。
DEST_CITY_NAME	目的地城市名称。
DEST_STATE_ABR	目的地州的双字母缩写（例如，DFW 是 TX）。

Table E2. T100 .csv column structure and description.

最后，CDE_Appendix.pdf 将包含指定机场的具体信息。同样，CDE 是指定机场的 IATA 代码。