

# US Air Comp

June 20, 2023

## 0.0.1 United States Airlines Analysis

```
[1]: import numpy as np
import pandas as pd
from bs4 import BeautifulSoup
import requests

import warnings
warnings.filterwarnings("ignore")
```

```
[2]: import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
```

```
[3]: from scipy.stats import ttest_1samp, ttest_ind, mannwhitneyu, levene, shapiro
from statsmodels.stats.power import ttest_power
from sklearn.preprocessing import OrdinalEncoder, StandardScaler
```

```
[4]: from sklearn.model_selection import StratifiedKFold, RandomizedSearchCV, \
    ↪train_test_split
from sklearn.pipeline import Pipeline
from sklearn.metrics import classification_report, accuracy_score
from sklearn.linear_model import SGDClassifier
from sklearn.tree import DecisionTreeClassifier
from xgboost import XGBClassifier
```

```
[5]: airlines = pd.read_excel('Airlines.xlsx')
```

```
[6]: airports = pd.read_excel('airports.xlsx')
```

```
[7]: runways = pd.read_excel('runways.xlsx')
```

```
[8]: print(airlines.shape)
print(airports.shape)
print(runways.shape)
```

(518556, 9)

(73805, 18)

(43977, 20)

```
[10]: airlines.head()
```

```
[10]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay
0	1	CO	269	SFO	IAH	3	15	205	1
1	2	US	1558	PHX	CLT	3	15	222	1
2	3	AA	2400	LAX	DFW	3	20	165	1
3	4	AA	2466	SFO	DFW	3	20	195	1
4	5	AS	108	ANC	SEA	3	30	202	0

```
[11]: airports.head()
```

```
[11]:
```

	id	ident	type	name	\
0	6523	00A	heliport	Total Rf Heliport	
1	323361	00AA	small_airport	Aero B Ranch Airport	
2	6524	00AK	small_airport	Lowell Field	
3	6525	00AL	small_airport	Epps Airpark	
4	6526	00AR	closed	Newport Hospital & Clinic Heliport	

	latitude_deg	longitude_deg	elevation_ft	continent	iso_country	iso_region	\
0	40.070801	-74.933601	11.0	NaN	US	US-PA	
1	38.704022	-101.473911	3435.0	NaN	US	US-KS	
2	59.947733	-151.692524	450.0	NaN	US	US-AK	
3	34.864799	-86.770302	820.0	NaN	US	US-AL	
4	35.608700	-91.254898	237.0	NaN	US	US-AR	

	municipality	scheduled_service	gps_code	iata_code	local_code	home_link	\
0	Bensalem	no	00A	NaN	00A	NaN	
1	Leoti	no	00AA	NaN	00AA	NaN	
2	Anchor Point	no	00AK	NaN	00AK	NaN	
3	Harvest	no	00AL	NaN	00AL	NaN	
4	Newport	no	NaN	NaN	NaN	NaN	

	wikipedia_link	keywords
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	00AR

```
[13]: runways.head()
```

```
[13]:
```

	id	airport_ref	airport_ident	length_ft	width_ft	surface	lighted	\
0	269408	6523	00A	80.0	80.0	ASPH-G	1	
1	255155	6524	00AK	2500.0	70.0	GRVL	0	
2	254165	6525	00AL	2300.0	200.0	TURF	0	

3	270932	6526	00AR	40.0	40.0	GRASS	0
4	322128	322127	00AS	1450.0	60.0	Turf	0

	closed	le_ident	le_latitude_deg	le_longitude_deg	le_elevation_ft	\
0	0	H1	NaN	NaN	NaN	
1	0	N	NaN	NaN	NaN	
2	0	1	NaN	NaN	NaN	
3	0	H1	NaN	NaN	NaN	
4	0	1	NaN	NaN	NaN	

	le_heading_degT	le_displaced_threshold_ft	he_ident	he_latitude_deg	\
0	NaN	NaN	NaN	NaN	
1	NaN	NaN	S	NaN	
2	NaN	NaN	19	NaN	
3	NaN	NaN	H1	NaN	
4	NaN	NaN	19	NaN	

	he_longitude_deg	he_elevation_ft	he_heading_degT	\
0	NaN	NaN	NaN	
1	NaN	NaN	NaN	
2	NaN	NaN	NaN	
3	NaN	NaN	NaN	
4	NaN	NaN	NaN	

	he_displaced_threshold_ft
0	NaN
1	NaN
2	NaN
3	NaN
4	NaN

## 0.0.2 1. Import and aggregate data:¶

- (a) Collect information related to flights, airports (e.g., type of airport and elevation), and runways (e.g., length\_ft, width\_ft, surface, and number of runways). Gather all fields you believe might cause avoidable delays in one dataset.

```
[13]: combined = pd.merge(airports,runways,left_on = 'ident', right_on = '
↳ 'airport_ident', how = 'left')
combined.head()
```

```
[13]:   id_x ident      type      name \
0   6523  00A  heliport  Total Rf Heliport
1  323361 00AA small_airport  Aero B Ranch Airport
2   6524 00AK small_airport  Lowell Field
3   6525 00AL small_airport  Epps Airpark
```

```

4      6526  00AR      closed  Newport Hospital & Clinic Heliport

      latitude_deg longitude_deg elevation_ft continent iso_country iso_region \
0      40.070801    -74.933601         11.0         NaN         US      US-PA
1      38.704022   -101.473911        3435.0         NaN         US      US-KS
2      59.947733   -151.692524         450.0         NaN         US      US-AK
3      34.864799   -86.770302         820.0         NaN         US      US-AL
4      35.608700   -91.254898         237.0         NaN         US      US-AR

      ... le_longitude_deg le_elevation_ft le_heading_degT \
0      ...              NaN              NaN              NaN
1      ...              NaN              NaN              NaN
2      ...              NaN              NaN              NaN
3      ...              NaN              NaN              NaN
4      ...              NaN              NaN              NaN

      le_displaced_threshold_ft he_ident he_latitude_deg he_longitude_deg \
0              NaN              NaN              NaN              NaN
1              NaN              NaN              NaN              NaN
2              NaN              S              NaN              NaN
3              NaN              19              NaN              NaN
4              NaN              H1              NaN              NaN

      he_elevation_ft he_heading_degT he_displaced_threshold_ft
0              NaN              NaN              NaN
1              NaN              NaN              NaN
2              NaN              NaN              NaN
3              NaN              NaN              NaN
4              NaN              NaN              NaN

```

[5 rows x 38 columns]

```
[14]: runway_count = pd.DataFrame(combined.ident.value_counts())
runway_count.head()
```

```
[14]:      ident
KORD      11
KNHU      10
JRA        9
SXS        8
1LA9       8

```

```
[15]: runway_count.rename(columns = {'ident':'runway_count'}, inplace = True)
runway_count['ident'] = runway_count.index
runway_count.head()
```

```
[15]:      runway_count ident
      KORD          11  KORD
      KNHU          10  KNHU
      JRA           9   JRA
      SXS           8   SXS
      1LA9          8   1LA9
```

```
[16]: runway_count.to_excel('runway_count.xlsx')
```

```
[17]: air_run = pd.merge(airports, runway_count, how = 'left', left_on = 'ident',
      ↪right_on = 'ident')[['iata_code', 'type', 'elevation_ft', 'runway_count']]
      air_run.head()
```

```
[17]:   iata_code      type  elevation_ft  runway_count
0      NaN      heliport          11.0             1
1      NaN  small_airport        3435.0             1
2      NaN  small_airport         450.0             1
3      NaN  small_airport         820.0             1
4      NaN      closed          237.0             1
```

```
[18]: air_run['iata_code'].unique()
```

```
[18]: array([nan, 'UTK', 'OCA', ..., 'YNJ', 'YKH', 'ZYW'], dtype=object)
```

```
[19]: air_run['iata_code'].nunique()
      air_run.to_csv('air_run.csv')
```

```
[20]: df = pd.merge(airlines, air_run, how = 'left', left_on = 'AirportFrom',
      ↪right_on = 'iata_code')
```

```
[21]: df.head(2)
```

```
[21]:   id  Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay  \
0   1      CO     269          SFO        IAH           3    15     205     1
1   2      US    1558          PHX        CLT           3    15     222     1

      iata_code      type  elevation_ft  runway_count
0      SFO  large_airport          13.0           4.0
1      PHX  large_airport        1135.0           3.0
```

```
[22]: df.rename(columns = {'iata_code': 'iata_code_source_airport', 'type':
      ↪'type_source_airport',
      ↪'elevation_ft':
      ↪'elevation_ft_source_airport', 'runway_count': 'runway_count_source_airport'},
      ↪inplace = True)
```

```
[23]: df.head()
```

```
[23]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	2	US	1558	PHX	CLT	3	15	222	1	
2	3	AA	2400	LAX	DFW	3	20	165	1	
3	4	AA	2466	SFO	DFW	3	20	195	1	
4	5	AS	108	ANC	SEA	3	30	202	0	

	iata_code_source_airport	type_source_airport	elevation_ft_source_airport	\
0	SFO	large_airport	13.0	
1	PHX	large_airport	1135.0	
2	LAX	large_airport	125.0	
3	SFO	large_airport	13.0	
4	ANC	large_airport	152.0	

	runway_count_source_airport
0	4.0
1	3.0
2	4.0
3	4.0
4	3.0

```
[24]: df.shape
```

```
[24]: (518556, 13)
```

```
[25]: df.columns
```

```
[25]: Index(['id', 'Airline', 'Flight', 'AirportFrom', 'AirportTo', 'DayOfWeek',
          'Time', 'Length', 'Delay', 'iata_code_source_airport',
          'type_source_airport', 'elevation_ft_source_airport',
          'runway_count_source_airport'],
         dtype='object')
```

```
[26]: df = pd.merge(df, air_run, how = 'left', left_on = 'AirportTo', right_on = 'iata_code')
```

```
[27]: df.head()
```

```
[27]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	2	US	1558	PHX	CLT	3	15	222	1	
2	3	AA	2400	LAX	DFW	3	20	165	1	
3	4	AA	2466	SFO	DFW	3	20	195	1	
4	5	AS	108	ANC	SEA	3	30	202	0	

	iata_code_source_airport	type_source_airport	elevation_ft_source_airport	\
0	SFO	large_airport	13.0	

1	PHX	large_airport	1135.0
2	LAX	large_airport	125.0
3	SFO	large_airport	13.0
4	ANC	large_airport	152.0

	runway_count_source_airport	iata_code	type	elevation_ft	\
0	4.0	IAH	large_airport	97.0	
1	3.0	CLT	large_airport	748.0	
2	4.0	DFW	large_airport	607.0	
3	4.0	DFW	large_airport	607.0	
4	3.0	SEA	large_airport	433.0	

	runway_count
0	5.0
1	4.0
2	7.0
3	7.0
4	4.0

```
[28]: df.rename(columns = {'iata_code': 'iata_code_dest_airport', 'type':
    ↳ 'type_dest_airport',
    ↳ 'elevation_ft': 'elevation_ft_dest_airport', 'runway_count':
    ↳ 'runway_count_dest_airport'}, inplace = True)
```

```
[29]: df.head()
```

```
[29]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	2	US	1558	PHX	CLT	3	15	222	1	
2	3	AA	2400	LAX	DFW	3	20	165	1	
3	4	AA	2466	SFO	DFW	3	20	195	1	
4	5	AS	108	ANC	SEA	3	30	202	0	

	iata_code_source_airport	type_source_airport	elevation_ft_source_airport	\
0	SFO	large_airport	13.0	
1	PHX	large_airport	1135.0	
2	LAX	large_airport	125.0	
3	SFO	large_airport	13.0	
4	ANC	large_airport	152.0	

	runway_count_source_airport	iata_code_dest_airport	type_dest_airport	\
0	4.0	IAH	large_airport	
1	3.0	CLT	large_airport	
2	4.0	DFW	large_airport	
3	4.0	DFW	large_airport	
4	3.0	SEA	large_airport	

	elevation_ft_dest_airport	runway_count_dest_airport
0	97.0	5.0
1	748.0	4.0
2	607.0	7.0
3	607.0	7.0
4	433.0	4.0

```
[30]: df.columns
```

```
[30]: Index(['id', 'Airline', 'Flight', 'AirportFrom', 'AirportTo', 'DayOfWeek',
        'Time', 'Length', 'Delay', 'iata_code_source_airport',
        'type_source_airport', 'elevation_ft_source_airport',
        'runway_count_source_airport', 'iata_code_dest_airport',
        'type_dest_airport', 'elevation_ft_dest_airport',
        'runway_count_dest_airport'],
        dtype='object')
```

```
[31]: df.
      ↪drop(['iata_code_source_airport', 'iata_code_dest_airport'], axis=1, inplace=True)
```

```
[32]: df.head(2)
```

```
[32]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	2	US	1558	PHX	CLT	3	15	222	1	

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	1135.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	3.0	large_airport	748.0	

	runway_count_dest_airport
0	5.0
1	4.0



0.0.3 1.(b) When it comes to on-time arrivals, different airlines perform differently based on the amount of experience they have. The major airlines in this field include US Airways Express (founded in 1967) Continental Airlines (founded in 1934), and Express Jet (founded in 19860. Pull such information specific to various airlines from the Wikipedia page link given below. [https://en.wikipedia.org/wiki/List\\_of\\_airlines\\_of\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_airlines_of_the_United_States).

```
[33]: wikiurl = 'https://en.wikipedia.org/wiki/List_of_airlines_of_the_United_States'
      table_class = "wikitable sortable jquery-tablesorter"
      response = requests.get(wikiurl)
      print(response.status_code)
```

200

```
[34]: soup = BeautifulSoup(response.text, 'html.parser')
      airline = soup.find_all('table', {"class": "wikitable", "class": "sortable"})
      print(airline[0])
```

```
<table class="wikitable sortable" style="border: 0; cellpadding: 2; cellspacing:
3;">
<tbody><tr style="vertical-align:middle;">
<th>Airline
</th>
<th>Image
</th>
<th><a class="mw-redirect" href="/wiki/IATA_airline_designator" title="IATA
airline designator">IATA</a>
</th>
<th><a class="mw-redirect" href="/wiki/ICAO_airline_designator" title="ICAO
airline designator">ICAO</a>
</th>
<th><a href="/wiki/Call_sign#Aviation" title="Call sign">Callsign</a>
</th>
<th>Primary hubs, <br/> <i>Secondary hubs</i>
</th>
<th>Founded
</th>
<th class="unsortable">Notes
</th></tr>
<tr>
<td><a href="/wiki/Alaska_Airlines" title="Alaska Airlines">Alaska Airlines</a>
</td>
<td><a class="image" href="/wiki/File:N615AS_Alaska_Airlines_2000_Boeing_737-790
_C_N_30344_(28850996478).jpg"></a>

</td>

<td>AS

</td>

<td>ASA

</td>

<td>ALASKA

</td>

<td><b><a class="mw-redirect" href="/wiki/Seattle-Tacoma\_International\_Airport" title="Seattle-Tacoma International Airport">Seattle/Tacoma</a></b><br><i><a href="/wiki/Ted\_Stevens\_Anchorage\_International\_Airport" title="Ted Stevens Anchorage International Airport">Anchorage</a></i><br><i><a href="/wiki/Portland\_International\_Airport" title="Portland International Airport">Portland (OR)</a></i><br><i><a href="/wiki/San\_Francisco\_International\_Airport" title="San Francisco International Airport">San Francisco</a></i><br><i><a href="/wiki/Los\_Angeles\_International\_Airport" title="Los Angeles International Airport">Los Angeles</a></i>

</td>

<td>1932

</td>

<td>Founded as McGee Airways and commenced operations in 1944 as Alaska Airlines.

</td></tr>

<tr>

<td><a href="/wiki/Allegiant\_Air" title="Allegiant Air">Allegiant Air</a>

</td>

<td><a class="image" href="/wiki/File:N221NV\_(33849198042).jpg"></a>

</td>

<td>G4

</td>

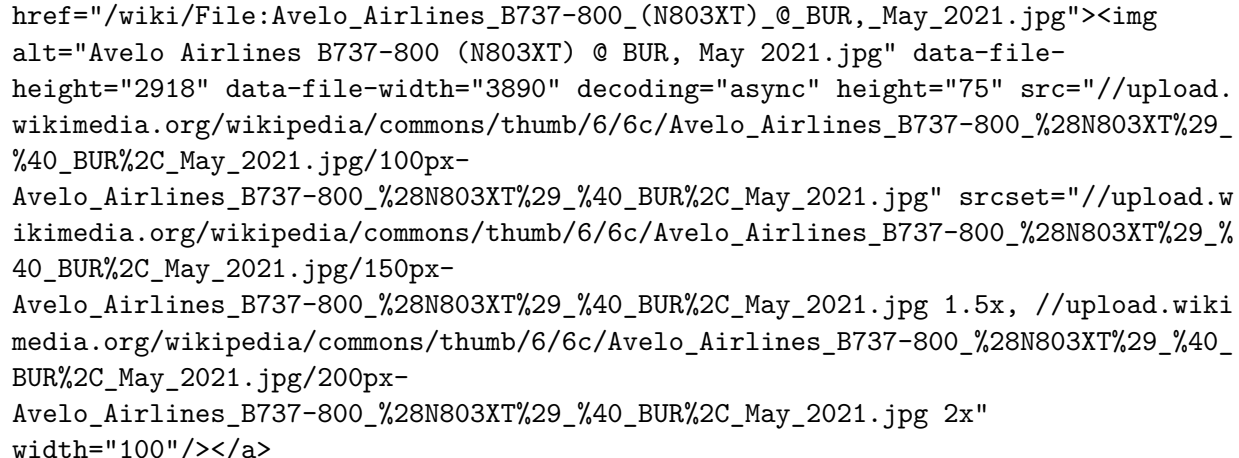
<td>AAY

</td>

<td>ALLEGIANT

</td>



|                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <p>AMERICAN</p> <p> <a href="/wiki/Dallas/Fort_Worth_International_Airport" title="Dallas/Fort Worth International Airport">Dallas/Fort Worth International Airport</a> <a href="/wiki/Charlotte_Douglas_International_Airport" title="Charlotte Douglas International Airport">Charlotte Douglas International Airport</a> <a href="/wiki/O'Hare_International_Airport" title="O'Hare International Airport">Chicago-O'Hare</a> <a href="/wiki/Los_Angeles_International_Airport" title="Los Angeles International Airport">Los Angeles</a> <a href="/wiki/Miami_International_Airport" title="Miami International Airport">Miami</a> <a href="/wiki/John_F._Kennedy_International_Airport" title="John F. Kennedy International Airport">New York-JFK</a> <a href="/wiki/LaGuardia_Airport" title="LaGuardia Airport">New York-LaGuardia</a> <a href="/wiki/Philadelphia_International_Airport" title="Philadelphia International Airport">Philadelphia</a> <a href="/wiki/Phoenix_Sky_Harbor_International_Airport" title="Phoenix Sky Harbor International Airport">Phoenix-Sky Harbor</a> <a href="/wiki/Ronald_Reagan_Washington_National_Airport" title="Ronald Reagan Washington National Airport">Washington-National</a> </p> <p>1926</p> <p>Founded as American Airways and commenced operations in 1936 as American Air Lines; largest airline in the world based on airline company revenue, scheduled passenger miles flown (per year), and fleet size.</p> <p> <a href="/wiki/Avelo_Airlines" title="Avelo Airlines">Avelo Airlines</a> </p> <p>  </p> <p>XP</p> |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|



title="Provo Municipal Airport">Provo</a></i><br/><i><a  
 href="/wiki/Tampa\_International\_Airport" title="Tampa International  
 Airport">Tampa</a></i>  
 </td>  
 <td>2018  
 </td>  
 <td>  
 </td></tr>  
 <tr>  
 <td><a href="/wiki/Delta\_Air\_Lines" title="Delta Air Lines">Delta Air Lines</a>  
 </td>  
 <td><a class="image" href="/wiki/File:N511DN\_(44338175191).jpg"></a>  
 </td>  
 <td>DL  
 </td>  
 <td>DAL  
 </td>  
 <td>DELTA  
 </td>  
 <td><b><a class="mw-redirect" href="/wiki/Hartsfield-  
 Jackson\_Atlanta\_International\_Airport" title="Hartsfield-Jackson Atlanta  
 International Airport">Atlanta</a></b><br/><i><a  
 href="/wiki/Logan\_International\_Airport" title="Logan International  
 Airport">Boston</a></i><br/><b><a class="mw-redirect"  
 href="/wiki/Detroit\_Metropolitan\_Wayne\_County\_Airport" title="Detroit  
 Metropolitan Wayne County Airport">Detroit</a></b><br/><i><a  
 href="/wiki/Los\_Angeles\_International\_Airport" title="Los Angeles International  
 Airport">Los Angeles</a></i><br/><b><a  
 href="/wiki/Minneapolis%E2%80%93Saint\_Paul\_International\_Airport"  
 title="Minneapolis-Saint Paul International Airport">Minneapolis/St.  
 Paul</a></b><br/><a href="/wiki/John\_F.\_Kennedy\_International\_Airport"  
 title="John F. Kennedy International Airport"><b>New York-JFK</b></a><br/><i><a  
 href="/wiki/LaGuardia\_Airport" title="LaGuardia Airport">New York-  
 LaGuardia</a></i><br/><a href="/wiki/Salt\_Lake\_City\_International\_Airport"  
 title="Salt Lake City International Airport"><b>Salt Lake City</b></a><br/><a  
 class="mw-redirect" href="/wiki/Seattle-Tacoma\_International\_Airport"  
 title="Seattle-Tacoma International Airport"><i>Seattle/Tacoma</i></a>  
 </td>  
 <td>1924  
 </td>  
 <td>Founded as Huff Daland Dusters and commenced operations in 1929 as Delta Air

Service.

|                                                                                                                                                                                                                                     |                                                                                      |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------|
|                                                                                                                                                                                                                                     |                                                                                      |
| <a href="/wiki/Eastern_Airlines,_LLC" title="Eastern Airlines, LLC">Eastern Airlines</a>                                                                                                                                            |    |
| 2D                                                                                                                                                                                                                                  |                                                                                      |
| EAL                                                                                                                                                                                                                                 |                                                                                      |
| EASTERN                                                                                                                                                                                                                             |                                                                                      |
| <b> <a href="/wiki/Miami_International_Airport" title="Miami International Airport">Miami</a> </b><br><b> <a href="/wiki/John_F._Kennedy_International_Airport" title="John F. Kennedy International Airport">New York-JFK</a> </b> |                                                                                      |
| 2010                                                                                                                                                                                                                                |                                                                                      |
|                                                                                                                                                                                                                                     |                                                                                      |
| <a href="/wiki/Frontier_Airlines" title="Frontier Airlines">Frontier Airlines</a>                                                                                                                                                   |  |
| F9                                                                                                                                                                                                                                  |                                                                                      |

```

<td>FFT
</td>
<td><span class="nowrap">FRONTIER FLIGHT</span>
</td>
<td><b><a href="/wiki/Denver_International_Airport" title="Denver International
Airport">Denver</a></b><br/><i><a
href="/wiki/Hartsfield%E2%80%93Jackson_Atlanta_International_Airport"
title="Hartsfield–Jackson Atlanta International
Airport">Atlanta</a></i><br/><i><a href="/wiki/O'Hare_International_Airport"
title="O'Hare International Airport">Chicago–O'Hare</a></i><br/><i><a
href="/wiki/Cincinnati/Northern_Kentucky_International_Airport"
title="Cincinnati/Northern Kentucky International
Airport">Cincinnati</a></i><br/><i><a
href="/wiki/Cleveland_Hopkins_International_Airport" title="Cleveland Hopkins
International Airport">Cleveland</a></i><br/><i><a class="mw-redirect"
href="/wiki/McCarran_International_Airport" title="McCarran International
Airport">Las Vegas</a></i><br/><i><a href="/wiki/Miami_International_Airport"
title="Miami International Airport">Miami</a></i><br/><i><a
href="/wiki/Orlando_International_Airport" title="Orlando International
Airport">Orlando</a></i><br/><i><a
href="/wiki/Philadelphia_International_Airport" title="Philadelphia
International Airport">Philadelphia</a></i><br/><i><a
href="/wiki/Raleigh%E2%80%93Durham_International_Airport" title="Raleigh–Durham
International Airport">Raleigh/Durham</a></i>
</td>
<td>1994
</td>
<td>
</td></tr>
<tr>
<td><a href="/wiki/Hawaiian_Airlines" title="Hawaiian Airlines">Hawaiian
Airlines</a>
</td>
<td><a class="image"
href="/wiki/File:Hawaiian_Airlines,_Airbus_A330-243,_N395HA_-_SEA_(19056912980).jpg"></a>

```



```

</td>
<td>HA
</td>
<td>HAL
</td>
<td>HAWAIIAN
</td>
<td><b><a href="/wiki/Daniel_K._Inouye_International_Airport" title="Daniel K.
Inouye International Airport">Honolulu</a></b><br><i><a
href="/wiki/Kahului_Airport" title="Kahului Airport">Kahului</a></i>
</td>
<td>1929
</td>
<td>Founded as Inter-Island Airways in early 1929 and commenced operations in
late 1929 as Hawaiian Airlines
</td></tr>
<tr>
<td><a href="/wiki/JetBlue" title="JetBlue">JetBlue</a>
</td>
<td><a class="image" href="/wiki/File:JetBlue_Airways_Airbus_A321-271NX_N2016J_t
axiing_at_JFK_Airport.jpg"></a>
</td>
<td>B6
</td>
<td>JBU
</td>
<td>JETBLUE
</td>
<td><b><a href="/wiki/John_F._Kennedy_International_Airport" title="John F.
Kennedy International Airport">New York-JFK</a></b><br><i><a
href="/wiki/Logan_International_Airport" title="Logan International
Airport">Boston</a></i><br><i><a href="/wiki/Los_Angeles_International_Airport"
title="Los Angeles International Airport">Los Angeles</a></i><br><i><a
href="/wiki/Fort_Lauderdale%E2%80%93Hollywood_International_Airport" title="Fort
Lauderdale-Hollywood International Airport">Fort Lauderdale</a></i><br><i><a
href="/wiki/Orlando_International_Airport" title="Orlando International
Airport">Orlando</a></i>

```

```

</td>
<td>1998
</td>
<td>Founded as New Air and commenced operations in 2000 as JetBlue Airways.
</td></tr>
<tr>
<td><a href="/wiki/Southwest_Airlines" title="Southwest Airlines">Southwest
Airlines</a>
</td>
<td><a class="image"
href="/wiki/File:N8712L_Southwest_Airlines_Boeing_737-8_MAX_s-
n_36930_(24896397167).jpg"></a>
</td>
<td>WN
</td>
<td>SWA
</td>
<td>SOUTHWEST
</td>
<td><b><a href="/wiki/Dallas_Love_Field" title="Dallas Love Field">Dallas-
Love</a></b><br/><i><a
href="/wiki/Hartsfield%E2%80%93Jackson_Atlanta_International_Airport"
title="Hartsfield-Jackson Atlanta International
Airport">Atlanta</a></i><br/><i><a class="mw-redirect"
href="/wiki/Baltimore/Washington_International_Thurgood_Marshall_Airport"
title="Baltimore/Washington International Thurgood Marshall
Airport">Baltimore</a></i><br/><i><a href="/wiki/Midway_International_Airport"
title="Midway International Airport">Chicago-Midway</a></i><br/><i><a
href="/wiki/Denver_International_Airport" title="Denver International
Airport">Denver</a></i><br/><i><a href="/wiki/William_P._Hobby_Airport"
title="William P. Hobby Airport">Houston-Hobby</a> </i><br/><i><a class="mw-
redirect" href="/wiki/McCarran_International_Airport" title="McCarran
International Airport">Las Vegas</a></i><br/><i><a
href="/wiki/Los_Angeles_International_Airport" title="Los Angeles International
Airport">Los Angeles</a></i><br/><i><a
href="/wiki/Oakland_International_Airport" title="Oakland International
Airport">Oakland</a></i><br/><i><a href="/wiki/Orlando_International_Airport"

```

[Orlando](/wiki/Orlando_International_Airport "Orlando International Airport")  
[Phoenix-Sky Harbor](/wiki/Phoenix_Sky_Harbor_International_Airport "Phoenix Sky Harbor International Airport")  
[Nashville International Airport](/wiki/Nashville_International_Airport "Nashville International Airport")

1967

Founded as Air Southwest and commenced operations in 1971 as Southwest Airlines; largest airline in the world based on number of passengers served per year.

[Spirit Airlines](/wiki/Spirit_Airlines "Spirit Airlines")



NK

NKS

SPIRIT WINGS

**[Atlantic City International Airport](/wiki/Atlantic_City_International_Airport "Atlantic City International Airport")**
**[Atlanta](/wiki/Hartsfield%E2%80%93Jackson_Atlanta_International_Airport "Hartsfield-Jackson Atlanta International Airport")**
  
[Detroit](/wiki/Detroit_Metropolitan_Wayne_County_Airport "Detroit Metropolitan Wayne County Airport")
  
[Las Vegas](/wiki/McCarran_International_Airport "McCarran International Airport")
  
[Fort Lauderdale](/wiki/Fort_Lauderdale%E2%80%93Hollywood_International_Airport "Fort Lauderdale-Hollywood International Airport")
  
[Chicago-O'Hare](/wiki/O'Hare_International_Airport "O'Hare International Airport")
  
[Dallas/Fort Worth](/wiki/Dallas/Fort_Worth_International_Airport "Dallas/Fort Worth International Airport")
  
[Baltimore](/wiki/Baltimore/Washington_International_Thurgood_Marshall_Airport "Baltimore/Washington International Thurgood Marshall Airport")
  
[Orlando](/wiki/Orlando_International_Airport "Orlando International Airport")

[Houston-Intercontinental](/wiki/George_Bush_Intercontinental_Airport "George Bush Intercontinental Airport")

1980

Founded as Charter One.

[Sun Country Airlines](/wiki/Sun_Country_Airlines "Sun Country Airlines")



SY

SCX

SUN COUNTRY

**[Minneapolis/St. Paul](/wiki/Minneapolis-Saint_Paul_International_Airport "Minneapolis-Saint Paul International Airport")**
  
*[Dallas/Fort Worth](/wiki/Dallas/Fort_Worth_International_Airport "Dallas/Fort Worth International Airport")*
  
*[Las Vegas](/wiki/McCarran_International_Airport "McCarran International Airport")*

1982

Commenced operations in 1983. Operates some [Amazon Air](/wiki/Amazon_Air "Amazon Air") cargo flights.

[United Airlines](/wiki/United_Airlines "United Airlines")



```

/5/5e/United_Airlines_Boeing_787-10_Dreamliner_N14001_approaching_Newark_Liberty
_International_Airport.jpg/100px-United_Airlines_Boeing_787-10_Dreamliner_N14001
_approaching_Newark_Liberty_International_Airport.jpg" srcset="//upload.wikimedi
a.org/wikipedia/commons/thumb/5/5e/United_Airlines_Boeing_787-10_Dreamliner_N140
01_approaching_Newark_Liberty_International_Airport.jpg/150px-United_Airlines_Bo
eing_787-10_Dreamliner_N14001_approaching_Newark_Liberty_International_Airport.j
pg 1.5x, //upload.wikimedia.org/wikipedia/commons/thumb/5/5e/United_Airlines_Boe
ing_787-10_Dreamliner_N14001_approaching_Newark_Liberty_International_Airport.jp
g/200px-United_Airlines_Boeing_787-10_Dreamliner_N14001_approaching_Newark_Liber
ty_International_Airport.jpg 2x" width="100"/></a>
</td>
<td>UA
</td>
<td>UAL
</td>
<td>UNITED
</td>
<td><b><a href="/wiki/O%27Hare_International_Airport" title="O'Hare
International Airport">Chicago-O'Hare</a></b><br/><a
href="/wiki/Denver_International_Airport" title="Denver International
Airport"><b>Denver</b></a><br/><i><a
href="/wiki/Antonio_B._Won_Pat_International_Airport" title="Antonio B. Won Pat
International Airport">Guam</a></i><br/><span class="nowrap"><b><a
href="/wiki/George_Bush_Intercontinental_Airport" title="George Bush
Intercontinental Airport">Houston-Intercontinental</a></b></span><br/><i><a
href="/wiki/Los_Angeles_International_Airport" title="Los Angeles International
Airport">Los Angeles</a></i><br/><b><a
href="/wiki/Newark_Liberty_International_Airport" title="Newark Liberty
International Airport">Newark</a></b><br/><a
href="/wiki/San_Francisco_International_Airport" title="San Francisco
International Airport"><b>San Francisco</b></a><br/><a class="mw-redirect"
href="/wiki/Washington_Dulles_International_Airport" title="Washington Dulles
International Airport"><b>Washington-Dulles</b></a>
</td>
<td>1926
</td>
<td>Founded as Varney Air Lines and commenced operations in 1931 as United Air
Lines.
</td></tr></tbody></table>

```

```

[35]: tables = pd.read_html(str(airline[0]))
      print(tables[0])

```

|   | Airline           | Image | IATA | ICAO | Callsign \ |
|---|-------------------|-------|------|------|------------|
| 0 | Alaska Airlines   | NaN   | AS   | ASA  | ALASKA     |
| 1 | Allegiant Air     | NaN   | G4   | AAV  | ALLEGIANTE |
| 2 | American Airlines | NaN   | AA   | AAL  | AMERICAN   |
| 3 | Avelo Airlines    | NaN   | XP   | VXP  | AVELO      |

|    |                      |     |    |     |                 |
|----|----------------------|-----|----|-----|-----------------|
| 4  | Breeze Airways       | NaN | MX | MXY | MOXY            |
| 5  | Delta Air Lines      | NaN | DL | DAL | DELTA           |
| 6  | Eastern Airlines     | NaN | 2D | EAL | EASTERN         |
| 7  | Frontier Airlines    | NaN | F9 | FFT | FRONTIER FLIGHT |
| 8  | Hawaiian Airlines    | NaN | HA | HAL | HAWAIIAN        |
| 9  | JetBlue              | NaN | B6 | JBU | JETBLUE         |
| 10 | Southwest Airlines   | NaN | WN | SWA | SOUTHWEST       |
| 11 | Spirit Airlines      | NaN | NK | NKS | SPIRIT WINGS    |
| 12 | Sun Country Airlines | NaN | SY | SCX | SUN COUNTRY     |
| 13 | United Airlines      | NaN | UA | UAL | UNITED          |

|    | Primary hubs,        | Secondary hubs                              | Founded | \ |
|----|----------------------|---------------------------------------------|---------|---|
| 0  | Seattle/Tacoma       | AnchoragePortland (OR)San Franci...         | 1932    |   |
| 1  | Las Vegas            | CincinnatiFort Walton BeachIndianapol...    | 1997    |   |
| 2  | Dallas/Fort Worth    | CharlotteChicago-O'HareLos An...            | 1926    |   |
| 3  | Burbank              | New HavenOrlandoRaleigh/DurhamWilmington... | 1987    |   |
| 4  | Charleston           | HartfordNew OrleansNorfolkProvidence...     | 2018    |   |
| 5  | Atlanta              | BostonDetroitLos AngelesMinneapolis/St...   | 1924    |   |
| 6  |                      | MiamiNew York-JFK                           | 2010    |   |
| 7  | Denver               | AtlantaChicago-O'HareCincinnatiCleveland... | 1994    |   |
| 8  |                      | HonoluluKahului                             | 1929    |   |
| 9  | New York-JFK         | BostonLos AngelesFort LauderdaleOr...       | 1998    |   |
| 10 | Dallas-Love          | AtlantaBaltimoreChicago-MidwayDenve...      | 1967    |   |
| 11 | Atlantic City        | AtlantaDetroitLas VegasFort Laude...        | 1980    |   |
| 12 | Minneapolis/St. Paul | Dallas/Fort WorthLas Vegas                  | 1982    |   |
| 13 | Chicago-O'Hare       | DenverGuamHouston-Intercontinent...         | 1926    |   |

|    | Notes                                             |
|----|---------------------------------------------------|
| 0  | Founded as McGee Airways and commenced operati... |
| 1  | Founded as WestJet Express and commenced opera... |
| 2  | Founded as American Airways and commenced oper... |
| 3  | First did business as Casino Express Airlines ... |
| 4  | NaN                                               |
| 5  | Founded as Huff Daland Dusters and commenced o... |
| 6  | NaN                                               |
| 7  | NaN                                               |
| 8  | Founded as Inter-Island Airways in early 1929 ... |
| 9  | Founded as New Air and commenced operations in... |
| 10 | Founded as Air Southwest and commenced operati... |
| 11 | Founded as Charter One.                           |
| 12 | Commenced operations in 1983.Operates some Ama... |
| 13 | Founded as Varney Air Lines and commenced oper... |

```
[36]: air = []
      for i in range(0,7):
          tables = pd.read_html(str(airline[i]))
          air.append(pd.DataFrame(tables[0]))
```

```
[37]: air[0].columns
```

```
[37]: Index(['Airline', 'Image', 'IATA', 'ICAO', 'Callsign',  
        'Primary hubs', 'Secondary hubs', 'Founded', 'Notes'],  
        dtype='object')
```

```
[38]: air[0].drop(['Image', 'Callsign', 'Primary hubs', 'Secondary_  
        ↳hubs', 'Notes'], axis=1, inplace=True)
```

```
[39]: air[0].head()
```

```
[39]:
```

|   | Airline           | IATA | ICAO | Founded |
|---|-------------------|------|------|---------|
| 0 | Alaska Airlines   | AS   | ASA  | 1932    |
| 1 | Allegiant Air     | G4   | AAY  | 1997    |
| 2 | American Airlines | AA   | AAL  | 1926    |
| 3 | Avelo Airlines    | XP   | VXP  | 1987    |
| 4 | Breeze Airways    | MX   | MXV  | 2018    |

```
[40]: mainline_airlines = air[0]
```

```
[41]: mainline_airlines.isna().sum()
```

```
[41]: Airline      0  
      IATA      0  
      ICAO      0  
      Founded   0  
      dtype: int64
```

```
[42]: regional_airlines = air[1]  
regional_airlines.drop(['Image', 'Callsign', 'Primary Hubs', 'Secondary_  
        ↳Hubs', 'Notes'], axis=1, inplace=True)  
regional_airlines.head()
```

```
[42]:
```

|   | Airline          | IATA | ICAO | Founded |
|---|------------------|------|------|---------|
| 0 | Air Wisconsin    | ZW   | AWI  | 1965    |
| 1 | Cape Air         | 9K   | KAP  | 1988    |
| 2 | CommutAir        | C5   | UCA  | 1989    |
| 3 | Contour Airlines | LF   | VTE  | 1982    |
| 4 | Elite Airways    | 7Q   | MNU  | 2006    |

```
[43]: commuter_airlines = air[2]  
commuter_airlines.drop(['Image', 'Callsign', 'Primary Hubs', 'Secondary_  
        ↳Hubs', 'Notes'], axis=1, inplace=True)  
commuter_airlines.head()
```

```
[43]:
```

|   | Airline      | IATA | ICAO | Founded |
|---|--------------|------|------|---------|
| 0 | Advanced Air | AN   | WSN  | 2005    |

|   |              |    |     |      |
|---|--------------|----|-----|------|
| 1 | Air Sunshine | YI | RSI | 1982 |
| 2 | Bering Air   | 8E | BRG | 1979 |
| 3 | Boutique Air | 4B | BTQ | 2007 |
| 4 | Everts Air   | 5V | VTB | 1978 |

```
[44]: charter_airlines = air[3]
charter_airlines.drop(['Image', 'Callsign', 'Primary Hubs, Secondary_
↳Hubs', 'Notes'], axis=1, inplace=True)
charter_airlines.head()
```

```
[44]:
```

|   | Airline             | IATA | ICAO | Founded |
|---|---------------------|------|------|---------|
| 0 | Air Charter Bahamas | NaN  | NaN  | NaN     |
| 1 | Air Flight Charters | NaN  | FLL  | 1987.0  |
| 2 | Airshare            | NaN  | XSR  | 2000.0  |
| 3 | Berry Aviation      | NaN  | BYA  | 1983.0  |
| 4 | Bighorn Airways     | NaN  | BHR  | 1947.0  |

```
[45]: cargo_airlines = air[4]
cargo_airlines.drop(['Image', 'Callsign', 'Primary Hubs, Secondary_
↳Hubs', 'Notes'], axis=1, inplace=True)
cargo_airlines.head()
```

```
[45]:
```

|   | Airline                     | IATA | ICAO | Founded |
|---|-----------------------------|------|------|---------|
| 0 | 21 Air                      | 2I   | CSB  | 2014.0  |
| 1 | ABX Air                     | GB   | ABX  | 1980.0  |
| 2 | Air Cargo Carriers          | 2Q   | SNC  | 1986.0  |
| 3 | AirNet Express              | NaN  | USC  | 1974.0  |
| 4 | Air Transport International | 8C   | ATN  | 1978.0  |

```
[46]: air_ambulances = air[5]
air_ambulances.drop(['Image', 'Callsign', 'Primary Hubs, Secondary_
↳Hubs', 'Notes'], axis=1, inplace=True)
air_ambulances.head()
```

```
[46]:
```

|   | Airline               | IATA | ICAO | Founded |
|---|-----------------------|------|------|---------|
| 0 | AirMed International  | NaN  | NaN  | 1987.0  |
| 1 | Air Methods           | NaN  | NaN  | 1980.0  |
| 2 | Critical Air Medicine | NaN  | NaN  | 1984.0  |
| 3 | Lifestar              | NaN  | NaN  | NaN     |
| 4 | Life Lion             | NaN  | NaN  | NaN     |

```
[47]: state_run_airlines = air[6]
state_run_airlines.drop(['Image', 'Callsign', 'Primary Hubs, Secondary_
↳Hubs', 'Notes'], axis=1, inplace=True)
state_run_airlines.head()
```



```
[47]:
```

|   | Airline                                          | IATA | ICAO | Founded |
|---|--------------------------------------------------|------|------|---------|
| 0 | Comco                                            | NaN  | NaN  | 2002    |
| 1 | Janet                                            | NaN  | WWW  | 1972    |
| 2 | Justice Prisoner and Alien Transportation System | NaN  | JUD  | 1980    |

```
[48]: all_airline = pd.
      ↪concat([mainline_airlines,regional_airlines,commuter_airlines,charter_airlines,cargo_airlin
              state_run_airlines],axis=0)
all_airline.shape
```

```
[48]: (140, 4)
```

```
[49]: all_airline.head(10)
```

```
[49]:
```

|   | Airline           | IATA | ICAO | Founded |
|---|-------------------|------|------|---------|
| 0 | Alaska Airlines   | AS   | ASA  | 1932.0  |
| 1 | Allegiant Air     | G4   | AAV  | 1997.0  |
| 2 | American Airlines | AA   | AAL  | 1926.0  |
| 3 | Avelo Airlines    | XP   | VXP  | 1987.0  |
| 4 | Breeze Airways    | MX   | MXV  | 2018.0  |
| 5 | Delta Air Lines   | DL   | DAL  | 1924.0  |
| 6 | Eastern Airlines  | 2D   | EAL  | 2010.0  |
| 7 | Frontier Airlines | F9   | FFT  | 1994.0  |
| 8 | Hawaiian Airlines | HA   | HAL  | 1929.0  |
| 9 | JetBlue           | B6   | JBU  | 1998.0  |

```
[50]: significant_airlines = pd.merge(df[['Airline']].
      ↪drop_duplicates(),all_airline[['IATA', 'Founded']].drop_duplicates(),
      how = 'left', left_on = 'Airline', right_on =
      ↪'IATA')
significant_airlines
```

```
[50]:
```

|    | Airline | IATA | Founded |
|----|---------|------|---------|
| 0  | CO      | NaN  | NaN     |
| 1  | US      | NaN  | NaN     |
| 2  | AA      | AA   | 1926.0  |
| 3  | AS      | AS   | 1932.0  |
| 4  | DL      | DL   | 1924.0  |
| 5  | B6      | B6   | 1998.0  |
| 6  | HA      | HA   | 1929.0  |
| 7  | OO      | OO   | 1972.0  |
| 8  | 9E      | 9E   | 1985.0  |
| 9  | OH      | OH   | 1979.0  |
| 10 | EV      | NaN  | NaN     |
| 11 | XE      | XE   | 2016.0  |
| 12 | YV      | YV   | 1980.0  |
| 13 | UA      | UA   | 1926.0  |

|    |    |    |        |
|----|----|----|--------|
| 14 | MQ | MQ | 1984.0 |
| 15 | F9 | F9 | 1994.0 |
| 16 | WN | WN | 1967.0 |

**0.0.4 3.(d)** The total passenger traffic may also contribute to flight delays. The term hub refers to busy commercial airports. Large hubs are airports that account for at least 1 percent of the total passenger enplanements in the United States. Airports that account for 0.25 percent to 1 percent of total passenger enplanements are considered medium hubs. Pull passenger traffic data from the Wikipedia page given below using web scraping and collate it in a table. [https://en.wikipedia.org/wiki/List\\_of\\_the\\_busiest\\_airports\\_in\\_the\\_United\\_States](https://en.wikipedia.org/wiki/List_of_the_busiest_airports_in_the_United_States)

```
[51]: wikiurl = 'https://en.wikipedia.org/wiki/
↳List_of_the_busiest_airports_in_the_United_States'
table_class = "wikitable sortable jquery-tablesorter"
response = requests.get(wikiurl)
print(response.status_code)
```

200

```
[52]: soup = BeautifulSoup(response.text, 'html.parser')
hub = soup.find_all('table', {"class": "wikitable", "class": "sortable"})
print(hub[0])
```

```
<table class="wikitable sortable">
<tbody><tr>
<th>Rank<br/>(2021)</th>
<th>Airports (large hubs)</th>
<th>IATA<br/>Code</th>
<th>Major cities served</th>
<th>State
</th>
<th>2021<sup class="reference" id="cite_ref-:0_3-0"><a
href="#cite_note-:0-3">[3]</a></sup>
</th>
<th>2020<sup class="reference" id="cite_ref-CY20_4-0"><a href="#cite_note-
CY20-4">[4]</a></sup></th>
<th>2019<sup class="reference" id="cite_ref-CY19_5-0"><a href="#cite_note-
CY19-5">[5]</a></sup></th>
<th>2018<sup class="reference" id="cite_ref-CY18_6-0"><a href="#cite_note-
CY18-6">[6]</a></sup></th>
<th>2017<sup class="reference" id="cite_ref-CY17_7-0"><a href="#cite_note-
CY17-7">[7]</a></sup></th>
<th>2016<sup class="reference" id="cite_ref-CY16_8-0"><a href="#cite_note-
CY16-8">[8]</a></sup></th>
<th>2015<sup class="reference" id="cite_ref-CY15_9-0"><a href="#cite_note-
```

CY15-9">[9]</a></sup></th>  
<th>2014<sup class="reference" id="cite\_ref-CY14\_10-0"><a href="#cite\_note-CY14-10">[10]</a></sup></th>  
<th>2013<sup class="reference" id="cite\_ref-CY13\_11-0"><a href="#cite\_note-CY13-11">[11]</a></sup></th>  
<th>2012<sup class="reference" id="cite\_ref-CY12\_12-0"><a href="#cite\_note-CY12-12">[12]</a></sup>  
</th></tr>  
<tr>  
<td>1  
</td>  
<td><a href="/wiki/Hartsfield%E2%80%93Jackson\_Atlanta\_International\_Airport" title="Hartsfield-Jackson Atlanta International Airport">Hartsfield-Jackson Atlanta International Airport</a>  
</td>  
<td>ATL  
</td>  
<td><a href="/wiki/Atlanta" title="Atlanta">Atlanta</a>  
</td>  
<td>GA  
</td>  
<td>36,676,010  
</td>  
<td>20,559,866  
</td>  
<td>53,505,795  
</td>  
<td>51,865,797  
</td>  
<td>50,251,964  
</td>  
<td>50,501,858  
</td>  
<td>49,340,732  
</td>  
<td>46,604,273  
</td>  
<td>45,308,407  
</td>  
<td>45,798,928  
</td></tr>  
<tr>  
<td>2  
</td>  
<td><a class="mw-redirect" href="/wiki/Dallas/Fort\_Worth\_International\_Airport" title="Dallas/Fort Worth International Airport">Dallas/Fort Worth International Airport</a>  
</td>

DFW	
	<a href="/wiki/Dallas" title="Dallas">Dallas</a> & <a href="/wiki/Fort_Worth,_Texas" title="Fort Worth, Texas">Fort Worth</a>
TX	
30,005,266	
18,593,421	
35,778,573	
32,821,799	
31,816,933	
31,283,579	
31,589,839	
30,804,567	
29,038,128	
28,022,904	
3	<a href="/wiki/Denver_International_Airport" title="Denver International Airport">Denver International Airport</a>
DEN	
	<a href="/wiki/Denver" title="Denver">Denver</a>
CO	
28,645,527	
16,243,216	
33,592,945	
31,362,941	
29,809,097	

```

</td>
<td>28,267,394
</td>
<td>26,280,043
</td>
<td>26,000,591
</td>
<td>25,496,885
</td>
<td>25,799,841
</td></tr>
<tr>
<td>4
</td>
<td><a href="/wiki/0%27Hare_International_Airport" title="O'Hare International
Airport">O'Hare International Airport</a>
</td>
<td>ORD
</td>
<td><a href="/wiki/Chicago" title="Chicago">Chicago</a>
</td>
<td>IL
</td>
<td>26,350,976
</td>
<td>14,606,034
</td>
<td>40,871,223
</td>
<td>39,873,927
</td>
<td>38,593,028
</td>
<td>37,589,899
</td>
<td>36,305,668
</td>
<td>33,843,426
</td>
<td>32,317,835
</td>
<td>32,171,795
</td></tr>
<tr>
<td>5
</td>
<td><a href="/wiki/Los_Angeles_International_Airport" title="Los Angeles
International Airport">Los Angeles International Airport</a>

```

```

</td>
<td>LAX
</td>
<td><a href="/wiki/Los_Angeles" title="Los Angeles">Los Angeles</a>
</td>
<td>CA
</td>
<td>23,663,410
</td>
<td>14,055,777
</td>
<td>42,939,104
</td>
<td>42,624,050
</td>
<td>41,232,432
</td>
<td>39,636,042
</td>
<td>36,351,272
</td>
<td>34,314,197
</td>
<td>32,425,892
</td>
<td>31,326,268
</td></tr>
<tr>
<td>6
</td>
<td><a href="/wiki/Charlotte_Douglas_International_Airport" title="Charlotte
Douglas International Airport">Charlotte Douglas International Airport</a>
</td>
<td>CLT
</td>
<td><a href="/wiki/Charlotte,_North_Carolina" title="Charlotte, North
Carolina">Charlotte</a>
</td>
<td>NC
</td>
<td>20,900,875
</td>
<td>12,952,869
</td>
<td>24,199,688
</td>
<td>22,281,949
</td>

```

<td>22,011,251
</td>
<td>21,511,880
</td>
<td>21,913,166
</td>
<td>21,537,725
</td>
<td>21,346,601
</td>
<td>20,033,816
</td></tr>
<tr>
<td>7
</td>
<td><a href="/wiki/Orlando_International_Airport" title="Orlando International Airport">Orlando International Airport</a>
</td>
<td>MCO
</td>
<td><a href="/wiki/Orlando,_Florida" title="Orlando, Florida">Orlando</a>
</td>
<td>FL
</td>
<td>19,618,838
</td>
<td>10,467,728
</td>
<td>24,562,271
</td>
<td>23,202,480
</td>
<td>21,565,448
</td>
<td>20,283,541
</td>
<td>18,759,938
</td>
<td>17,278,608
</td>
<td>16,884,524
</td>
<td>17,159,427
</td></tr>
<tr>
<td>8
</td>
<td><a href="/wiki/Harry_Reid_International_Airport" title="Harry Reid

International Airport">Harry Reid International Airport</a>
</td>
<td>LAS
</td>
<td><a href="/wiki/Las_Vegas" title="Las Vegas">Las Vegas</a>
</td>
<td>NV
</td>
<td>19,160,342
</td>
<td>10,584,059
</td>
<td>24,728,361
</td>
<td>23,795,012
</td>
<td>23,364,393
</td>
<td>22,833,267
</td>
<td>21,857,693
</td>
<td>20,620,248
</td>
<td>19,946,179
</td>
<td>19,959,651
</td></tr>
<tr>
<td>9
</td>
<td><a href="/wiki/Phoenix_Sky_Harbor_International_Airport" title="Phoenix Sky Harbor International Airport">Phoenix Sky Harbor International Airport</a>
</td>
<td>PHX
</td>
<td><a href="/wiki/Phoenix,_Arizona" title="Phoenix, Arizona">Phoenix</a>
</td>
<td>AZ
</td>
<td>18,940,287
</td>
<td>10,531,436
</td>
<td>22,433,552
</td>
<td>21,622,580
</td>



<td>21,185,458
</td>
<td>20,896,265
</td>
<td>21,351,504
</td>
<td>20,344,867
</td>
<td>19,525,109
</td>
<td>19,560,870
</td></tr>
<tr>
<td>10
</td>
<td><a href="/wiki/Miami_International_Airport" title="Miami International Airport">Miami International Airport</a>
</td>
<td>MIA
</td>
<td><a href="/wiki/Miami" title="Miami">Miami</a>
</td>
<td>FL
</td>
<td>17,500,096
</td>
<td>8,786,007
</td>
<td>21,421,031
</td>
<td>21,021,640
</td>
<td>20,709,225
</td>
<td>20,875,813
</td>
<td>20,986,349
</td>
<td>19,471,466
</td>
<td>19,420,089
</td>
<td>18,987,488
</td></tr>
<tr>
<td>11
</td>
<td><a href="/wiki/Seattle%E2%80%93Tacoma_International_Airport"

```

title="Seattle-Tacoma International Airport">Seattle-Tacoma International
Airport</a>
</td>
<td>SEA
</td>
<td><a href="/wiki/Seattle" title="Seattle">Seattle</a>
</td>
<td>WA
</td>
<td>17,430,195
</td>
<td>9,462,411
</td>
<td>25,001,762
</td>
<td>24,024,908
</td>
<td>22,639,124
</td>
<td>21,887,110
</td>
<td>20,148,980
</td>
<td>17,888,080
</td>
<td>16,690,295
</td>
<td>16,121,123
</td></tr>
<tr>
<td>12
</td>
<td><a href="/wiki/George_Bush_Intercontinental_Airport" title="George Bush
Intercontinental Airport">George Bush Intercontinental Airport</a>
</td>
<td>IAH
</td>
<td><a href="/wiki/Houston" title="Houston">Houston</a>
</td>
<td>TX
</td>
<td>16,242,821
</td>
<td>8,682,558
</td>
<td>21,905,309
</td>
<td>21,157,398

```

```

</td>
<td>19,603,731
</td>
<td>20,062,072
</td>
<td>20,595,881
</td>
<td>19,772,087
</td>
<td>18,952,840
</td>
<td>19,039,000
</td></tr>
<tr>
<td>13
</td>
<td><a href="/wiki/John_F._Kennedy_International_Airport" title="John F. Kennedy
International Airport">John F. Kennedy International Airport</a>
</td>
<td>JFK
</td>
<td><a href="/wiki/New_York_City" title="New York City">New York City</a>
</td>
<td>NY
</td>
<td>15,273,342
</td>
<td>8,269,819
</td>
<td>31,036,655
</td>
<td>30,620,769
</td>
<td>29,533,154
</td>
<td>29,239,151
</td>
<td>27,782,369
</td>
<td>26,244,928
</td>
<td>25,036,358
</td>
<td>24,520,981
</td></tr>
<tr>
<td>14
</td>

```

<a href="/wiki/Newark_Liberty_International_Airport" title="Newark Liberty International Airport">Newark Liberty International Airport</a>	EWR	NJ	14,514,049	7,985,474	23,160,763	22,797,602	21,571,198	19,923,009	18,684,818	17,773,405	17,546,506	17,055,993
<a href="/wiki/Fort_Lauderdale%E2%80%93Hollywood_International_Airport" title="Fort Lauderdale-Hollywood International Airport">Fort Lauderdale-Hollywood International Airport</a>	FLL	<a href="/wiki/Fort_Lauderdale,_Florida" title="Fort Lauderdale, Florida">Fort Lauderdale</a> & <a href="/wiki/Hollywood,_Florida" title="Hollywood, Florida">Hollywood</a>	FL	13,598,994	8,015,744							

<td>17,950,989
</td>
<td>17,612,331
</td>
<td>15,817,043
</td>
<td>14,263,270
</td>
<td>13,061,632
</td>
<td>12,031,860
</td>
<td>11,538,140
</td>
<td>11,445,103
</td></tr>
<tr>
<td>16
</td>
<td><a href="/wiki/Minneapolis%E2%80%93Saint_Paul_International_Airport" title="Minneapolis-Saint Paul International Airport">Minneapolis-Saint Paul International Airport</a>
</td>
<td>MSP
</td>
<td><a href="/wiki/Minneapolis" title="Minneapolis">Minneapolis</a> & <a href="/wiki/Saint_Paul,_Minnesota" title="Saint Paul, Minnesota">Saint Paul</a>
</td>
<td>MN
</td>
<td>12,211,409
</td>
<td>7,069,720
</td>
<td>19,192,917
</td>
<td>18,361,942
</td>
<td>18,409,704
</td>
<td>18,123,844
</td>
<td>17,634,273
</td>
<td>16,972,678
</td>
<td>16,280,835
</td>

15,943,878
17
<a href="/wiki/San_Francisco_International_Airport" title="San Francisco International Airport">San Francisco International Airport</a>
SFO
<a href="/wiki/San_Francisco" title="San Francisco">San Francisco</a>
CA
11,725,347
7,745,057
27,779,230
27,790,717
26,900,048
25,707,101
24,190,560
22,770,783
21,704,626
21,284,236
18
<a href="/wiki/Detroit_Metropolitan_Airport" title="Detroit Metropolitan Airport">Detroit Metropolitan Airport</a>
DTW
<a href="/wiki/Detroit" title="Detroit">Detroit</a>
MI
11,517,696

```

<td>6,822,324
</td>
<td>18,143,040
</td>
<td>17,436,837
</td>
<td>17,036,092
</td>
<td>16,847,135
</td>
<td>16,255,520
</td>
<td>15,775,941
</td>
<td>15,683,523
</td>
<td>15,599,879
</td></tr>
<tr>
<td>19
</td>
<td><a href="/wiki/Logan_International_Airport" title="Logan International
Airport">Logan International Airport</a>
</td>
<td>BOS
</td>
<td><a href="/wiki/Boston" title="Boston">Boston</a>
</td>
<td>MA
</td>
<td>10,909,817
</td>
<td>6,035,452
</td>
<td>20,699,377
</td>
<td>20,006,521
</td>
<td>18,759,742
</td>
<td>17,759,044
</td>
<td>16,290,362
</td>
<td>15,507,561
</td>
<td>14,810,153
</td>

```

14,293,695
------------

20
----

<a href="/wiki/Salt_Lake_City_International_Airport" title="Salt Lake City International Airport">Salt Lake City International Airport</a>
--------------------------------------------------------------------------------------------------------------------------------------------

SLC
-----

<a href="/wiki/Salt_Lake_City" title="Salt Lake City">Salt Lake City</a>
--------------------------------------------------------------------------

UT
----

10,795,906
------------

5,753,239
-----------

12,840,841
------------

12,226,730
------------

11,615,954
------------

11,143,738
------------

10,634,538
------------

10,139,065
------------

9,668,048
-----------

9,579,840
-----------

21
----

<a href="/wiki/Philadelphia_International_Airport" title="Philadelphia International Airport">Philadelphia International Airport</a>
--------------------------------------------------------------------------------------------------------------------------------------

PHL
-----

<a href="/wiki/Philadelphia" title="Philadelphia">Philadelphia</a>
--------------------------------------------------------------------

PA
----

9,820,222
-----------



```

<td>5,753,239
</td>
<td>16,006,389
</td>
<td>15,292,670
</td>
<td>14,271,243
</td>
<td>14,564,419
</td>
<td>15,101,349
</td>
<td>14,792,339
</td>
<td>14,727,945
</td>
<td>14,589,337
</td></tr>
<tr>
<td>22
</td>
<td><a href="/wiki/Baltimore/Washington_International_Airport"
title="Baltimore/Washington International Airport">Baltimore/Washington
International Airport</a>
</td>
<td>BWI
</td>
<td><a href="/wiki/Baltimore" title="Baltimore">Baltimore</a> & <a
href="/wiki/Washington,_D.C." title="Washington, D.C.">Washington, D.C.</a>
</td>
<td>MD
</td>
<td>9,253,561
</td>
<td>5,451,355
</td>
<td>13,284,687
</td>
<td>13,371,816
</td>
<td>12,976,554
</td>
<td>12,340,972
</td>
<td>11,738,845
</td>
<td>11,022,200
</td>

```

11,132,731
11,186,444
23
<a href="/wiki/Tampa_International_Airport" title="Tampa International Airport">Tampa International Airport</a>
TPA
<a href="/wiki/Tampa,_Florida" title="Tampa, Florida">Tampa</a>
FL
8,847,197
4,966,775
10,978,756
10,368,514
9,548,580
9,194,994
9,150,458
8,531,561
8,267,752
8,218,487
24
<a href="/wiki/San_Diego_International_Airport" title="San Diego International Airport">San Diego International Airport</a>
SAN
<a href="/wiki/San_Diego" title="San Diego">San Diego</a>
CA

<td>7,836,360
</td>
<td>4,637,856
</td>
<td>12,648,692
</td>
<td>12,174,224
</td>
<td>11,139,933
</td>
<td>10,340,164
</td>
<td>9,985,763
</td>
<td>9,333,152
</td>
<td>8,878,772
</td>
<td>8,686,621
</td></tr>
<tr>
<td>25
</td>
<td><a href="/wiki/LaGuardia_Airport" title="LaGuardia Airport">LaGuardia Airport</a>
</td>
<td>LGA
</td>
<td><a href="/wiki/New_York_City" title="New York City">New York City</a>
</td>
<td>NY
</td>
<td>7,827,307
</td>
<td>4,147,116
</td>
<td>15,393,601
</td>
<td>15,058,501
</td>
<td>14,614,802
</td>
<td>14,762,593
</td>
<td>14,319,924
</td>
<td>13,535,372
</td>

```

<td>13,372,269
</td>
<td>12,818,717
</td></tr>
<tr>
<td>26
</td>
<td><a class="mw-redirect" href="/wiki/Chicago_Midway_International_Airport"
title="Chicago Midway International Airport">Midway International Airport</a>
</td>
<td>MDW
</td>
<td><a href="/wiki/Chicago" title="Chicago">Chicago</a>
</td>
<td>IL
</td>
<td>7,680,617
</td>
<td>4,236,603
</td>
<td>10,081,781
</td>
<td>10,678,018
</td>
<td>10,912,074
</td>
<td>11,044,387
</td>
<td>10,830,850
</td>
<td>10,311,996
</td>
<td>9,915,646
</td>
<td>9,436,387
</td></tr>
<tr>
<td>27
</td>
<td><a href="/wiki/Nashville_International_Airport" title="Nashville
International Airport">Nashville International Airport</a>
</td>
<td>BNA
</td>
<td><a href="/wiki/Nashville,_Tennessee" title="Nashville,
Tennessee">Nashville</a>
</td>
<td>TN

```

```

</td>
<td>7,594,049
</td>
<td>4,013,995
</td>
<td>8,935,654
</td>
<td>8,017,347
</td>
<td>6,902,771
</td>
<td>6,338,517
</td>
<td>5,715,205
</td>
<td>5,396,958
</td>
<td>5,050,989
</td>
<td>4,797,102
</td></tr>
<tr>
<td>28
</td>
<td><a class="mw-redirect" href="/wiki/Washington_Dulles_International_Airport"
title="Washington Dulles International Airport">Washington Dulles International
Airport</a>
</td>
<td>IAD
</td>
<td><a href="/wiki/Washington,_D.C." title="Washington, D.C.">Washington,
D.C.</a>
</td>
<td>VA
</td>
<td>7,227,875
</td>
<td>3,862,658
</td>
<td>11,884,117
</td>
<td>11,621,623
</td>
<td>11,024,306
</td>
<td>10,596,942
</td>
<td>10,363,974

```

```

</td>
<td>10,415,948
</td>
<td>10,570,993
</td>
<td>10,816,216
</td></tr>
<tr>
<td>29
</td>
<td><a href="/wiki/Ronald_Reagan_Washington_National_Airport" title="Ronald
Reagan Washington National Airport">Ronald Reagan Washington National
Airport</a>
</td>
<td>DCA
</td>
<td><a href="/wiki/Washington,_D.C." title="Washington, D.C.">Washington,
D.C.</a>
</td>
<td>VA
</td>
<td>6,731,737
</td>
<td>3,573,489
</td>
<td>11,595,454
</td>
<td>11,367,176
</td>
<td>11,506,310
</td>
<td>11,470,854
</td>
<td>11,242,375
</td>
<td>10,057,794
</td>
<td>9,838,034
</td>
<td>9,462,231
</td></tr>
<tr>
<td>30
</td>
<td><a href="/wiki/Austin%E2%80%93Bergstrom_International_Airport"
title="Austin-Bergstrom International Airport">Austin-Bergstrom International
Airport</a>
</td>

```

```

<td>AUS
</td>
<td><a href="/wiki/Austin,_Texas" title="Austin, Texas">Austin</a>
</td>
<td>TX
</td>
<td>6,666,215
</td>
<td>3,141,505
</td>
<td>8,683,711
</td>
<td>7,921,797
</td>
<td>6,973,115
</td>
<td>6,095,545
</td>
<td>5,797,547
</td>
<td>5,219,982
</td>
<td>4,900,959
</td>
<td>4,606,252
</td></tr>
<tr>
<td>31
</td>
<td><a href="/wiki/Dallas_Love_Field" title="Dallas Love Field">Dallas Love
Field</a>
</td>
<td>DAL
</td>
<td><a href="/wiki/Dallas" title="Dallas">Dallas</a>
</td>
<td>TX
</td>
<td>6,487,563
</td>
<td>3,669,930
</td>
<td>8,408,457
</td>
<td>8,134,848
</td>
<td>7,876,769
</td>

```

<div> <div>&lt;td&gt;7,554,596</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;7,040,921</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,522,341</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,023,779</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;3,902,628</div> <div>&lt;/td&gt;&lt;/tr&gt;</div> </div>	<div>&lt;tr&gt;</div> <div>&lt;td&gt;32</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a href="/wiki/Daniel_K._Inouye_International_Airport" title="Daniel K. Inouye International Airport"&gt;Daniel K. Inouye International Airport&lt;/a&gt;</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;HNL</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a href="/wiki/Honolulu" title="Honolulu"&gt;Honolulu&lt;/a&gt;</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;HI</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,830,928</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;3,126,391</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,988,678</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,578,505</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,743,989</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,656,340</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,656,340</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,463,000</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,466,995</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;9,225,848</div> <div>&lt;/td&gt;&lt;/tr&gt;</div>
<div>&lt;tr&gt;</div> <div>&lt;td&gt;33</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a href="/wiki/Portland_International_Airport" title="Portland International Airport"&gt;Portland International Airport&lt;/a&gt;</div> <div>&lt;/td&gt;</div>	



```

<td>PDX
</td>
<td><a href="/wiki/Portland,_Oregon" title="Portland, Oregon">Portland</a>
</td>
<td>OR
</td>
<td>5,759,879
</td>
<td>3,455,877
</td>
<td>9,797,408
</td>
<td>9,940,866
</td>
<td>9,435,473
</td>
<td>9,071,154
</td>
<td>8,340,234
</td>
<td>7,878,760
</td>
<td>7,452,603
</td>
<td>7,142,620
</td></tr>
<tr>
<td>34
</td>
<td><a href="/wiki/William_P._Hobby_Airport" title="William P. Hobby
Airport">William P. Hobby Airport</a>
</td>
<td>HOU
</td>
<td><a href="/wiki/Houston" title="Houston">Houston</a>
</td>
<td>TX
</td>
<td>5,560,780
</td>
<td>3,127,178
</td>
<td>7,069,614
</td>
<td>6,937,061
</td>
<td>6,741,870
</td>

```

<div> <div>&lt;td&gt;6,285,181</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,937,944</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,800,726</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,377,050</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,043,737</div> <div>&lt;/td&gt;&lt;/tr&gt;</div> </div>	<div>&lt;tr&gt;</div> <div>&lt;td&gt;35</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a href="/wiki/Southwest_Florida_International_Airport" title="Southwest Florida International Airport"&gt;Southwest Florida International Airport&lt;/a&gt;</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;RSW</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a href="/wiki/Fort_Myers,_Florida" title="Fort Myers, Florida"&gt;Fort Myers&lt;/a&gt;</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;FL</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,080,805</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;2,947,139</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;5,144,467</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,719,568</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,461,304</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,350,650</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,231,134</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;4,025,959</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;3,788,870</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;3,634,152</div> <div>&lt;/td&gt;&lt;/tr&gt;</div>
<div>&lt;tr&gt;</div> <div>&lt;td&gt;36</div> <div>&lt;/td&gt;</div> <div>&lt;td&gt;&lt;a class="mw-redirect" href="/wiki/Lambert-St._Louis_International_Airport" title="Lambert-St. Louis International Airport"&gt;St. Louis Lambert International</div>	

```

Airport</a>
</td>
<td>STL
</td>
<td><a href="/wiki/St._Louis" title="St. Louis">St. Louis</a>
</td>
<td>MO
</td>
<td>5,070,471
</td>
<td>3,041,765
</td>
<td>7,946,986
</td>
<td>7,822,274
</td>
<td>7,372,805
</td>
<td>6,793,076
</td>
<td>6,239,231
</td>
<td>6,108,758
</td>
<td>6,216,104
</td>
<td>6,208,750
</td></tr>
<tr>
<td>37
</td>
<td><a href="/wiki/Sacramento_International_Airport" title="Sacramento
International Airport">Sacramento International Airport</a>
</td>
<td>SMF
</td>
<td><a href="/wiki/Sacramento,_California" title="Sacramento,
California">Sacramento</a>
</td>
<td>CA
</td>
<td>4,760,275
</td>
<td>2,710,342
</td>
<td>6,454,413
</td>
<td>6,031,630

```

```

</td>
<td>5,460,526
</td>
<td>4,969,366
</td>
<td>4,816,440
</td>
<td>4,384,616
</td>
<td>4,255,145
</td>
<td>4,357,899
</td></tr>
<tr>
<td>38
</td>
<td><a href="/wiki/Luis_Mu%C3%B1oz_Mar%C3%ADn_International_Airport" title="Luis
Muñoz Marín International Airport">Luis Muñoz Marín International Airport</a>
</td>
<td>SJU
</td>
<td><a href="/wiki/San_Juan,_Puerto_Rico" title="San Juan, Puerto Rico">San
Juan</a>
</td>
<td>PR
</td>
<td>4,738,725
</td>
<td>2,362,851
</td>
<td>4,590,117
</td>
<td>4,033,412
</td>
<td>4,163,587
</td>
<td>4,343,354
</td>
<td>4,218,785
</td>
<td>4,150,828
</td>
<td>4,103,197
</td>
<td>4,204,478
</td></tr>
<tr>
<td>39

```

```

</td>
<td><a class="mw-redirect" href="/wiki/Raleigh-Durham_International_Airport"
title="Raleigh-Durham International Airport">Raleigh-Durham International
Airport</a>
</td>
<td>RDU
</td>
<td><a href="/wiki/Raleigh,_North_Carolina" title="Raleigh, North
Carolina">Raleigh</a>
</td>
<td>NC
</td>
<td>4,311,049
</td>
<td>2,337,496
</td>
<td>6,919,429
</td>
<td>6,416,822
</td>
<td>5,851,004
</td>
<td>5,401,714
</td>
<td>4,954,717
</td>
<td>4,673,869
</td>
<td>4,482,016
</td>
<td>4,490,374
</td></tr>
<tr>
<td>40
</td>
<td><a href="/wiki/Louis_Armstrong_New_Orleans_International_Airport"
title="Louis Armstrong New Orleans International Airport">Louis Armstrong New
Orleans International Airport</a>
</td>
<td>MSY
</td>
<td><a class="mw-redirect" href="/wiki/New_Orleans,_Louisiana" title="New
Orleans, Louisiana">New Orleans</a>
</td>
<td>LA
</td>
<td>4,017,147
</td>

```

```

<td>2,632,606
</td>
<td>6,717,105
</td>
<td>6,565,482
</td>
<td>6,005,527
</td>
<td>5,569,705
</td>
<td>5,329,696
</td>
<td>4,870,569
</td>
<td>4,576,539
</td>
<td>4,293,624
</td></tr>
<tr>
<td>41
</td>
<td><a href="/wiki/Oakland_International_Airport" title="Oakland International
Airport">Oakland International Airport</a>
</td>
<td>OAK
</td>
<td><a href="/wiki/Oakland,_California" title="Oakland, California">Oakland</a>
</td>
<td>CA
</td>
<td>4,011,953
</td>
<td>2,271,294
</td>
<td>6,560,230
</td>
<td>6,798,321
</td>
<td>6,530,308
</td>
<td>5,934,639
</td>
<td>5,506,672
</td>
<td>5,069,257
</td>
<td>4,770,716
</td>

```

4,926,683
42
<a href="/wiki/John_Wayne_Airport" title="John Wayne Airport">John Wayne Airport</a>
SNA
<a href="/wiki/Orange_County,_California" title="Orange County, California">Orange County</a>
CA
3,807,205
1,824,836
5,153,276
5,317,149
5,195,047
5,217,242
4,945,175
4,584,147
4,540,628
4,381,172
43
<a href="/wiki/Kansas_City_International_Airport" title="Kansas City International Airport">Kansas City International Airport</a>
MCI
<a href="/wiki/Kansas_City,_Missouri" title="Kansas City, Missouri">Kansas City</a>
MO

```

<td>3,795,290
</td>
<td>2,167,616
</td>
<td>5,759,419
</td>
<td>5,935,131
</td>
<td>5,744,918
</td>
<td>5,391,557
</td>
<td>5,135,127
</td>
<td>4,982,722
</td>
<td>4,836,221
</td>
<td>4,866,850
</td></tr>
<tr>
<td>44
</td>
<td><a href="/wiki/San_Antonio_International_Airport" title="San Antonio
International Airport">San Antonio International Airport</a>
</td>
<td>SAT
</td>
<td><a class="mw-redirect" href="/wiki/San_Antonio,_Texas" title="San Antonio,
Texas">San Antonio</a>
</td>
<td>TX
</td>
<td>3,677,643
</td>
<td>1,919,958
</td>
<td>5,022,980
</td>
<td>4,844,427
</td>
<td>4,521,611
</td>
<td>4,179,994
</td>
<td>4,091,389
</td>
<td>4,046,856

```



</td>	
<td>4,005,874	
</td>	
<td>4,036,625	
</td></tr>	
<tr>	
<td>45	
</td>	
<td><a href="/wiki/San_Jose_International_Airport" title="San Jose International Airport">Norman Y. Mineta San José International Airport</a>	
</td>	
<td>SJC	
</td>	
<td><a href="/wiki/San_Jose,_California" title="San Jose, California">San Jose</a>	
</td>	
<td>CA	
</td>	
<td>3,619,690	
</td>	
<td>2,283,186	
</td>	
<td>7,828,885	
</td>	
<td>7,140,616	
</td>	
<td>6,225,148	
</td>	
<td>5,321,603	
</td>	
<td>4,885,690	
</td>	
<td>4,621,003	
</td>	
<td>4,315,839	
</td>	
<td>4,077,654	
</td></tr>	
<tr>	
<td>46	
</td>	
<td><a href="/wiki/Cleveland_Hopkins_International_Airport" title="Cleveland Hopkins International Airport">Cleveland Hopkins International Airport</a>	
</td>	
<td>CLE	
</td>	
<td><a class="mw-redirect" href="/wiki/Cleveland,_Ohio" title="Cleveland, Ohio">Cleveland</a>	

```

</td>
<td>0H
</td>
<td>3,552,402
</td>
<td>1,990,156
</td>
<td>4,894,541
</td>
<td>4,836,580
</td>
<td>4,562,740
</td>
<td>4,205,739
</td>
<td>4,083,476
</td>
<td>3,686,315
</td>
<td>4,375,448
</td>
<td>4,346,941
</td></tr>
<tr>
<td>47
</td>
<td><a href="/wiki/Indianapolis_International_Airport" title="Indianapolis
International Airport">Indianapolis International Airport</a>
</td>
<td>IND
</td>
<td><a class="mw-redirect" href="/wiki/Indianapolis,_Indiana"
title="Indianapolis, Indiana">Indianapolis</a>
</td>
<td>IN
</td>
<td>3,487,100
</td>
<td>1,989,126
</td>
<td>4,709,183
</td>
<td>4,695,040
</td>
<td>4,376,432
</td>
<td>4,216,766
</td>

```

<div> <div>3,889,567</div> <div></div> </div> <div> <div>3,605,908</div> <div></div> </div> <div> <div>3,535,015</div> <div></div> </div> <div> <div>3,586,422</div> <div></div> </div>	
<div>48</div> <div> <a href="/wiki/Pittsburgh_International_Airport" title="Pittsburgh International Airport">Pittsburgh International Airport</a> </div> <div>PIT</div> <div> <a class="mw-redirect" href="/wiki/Pittsburgh,_Pennsylvania" title="Pittsburgh, Pennsylvania">Pittsburgh</a> </div> <div>PA</div> <div>3,069,259</div> <div>1,742,406</div> <div>4,715,947</div> <div>4,670,033</div> <div>4,327,431</div> <div>3,986,114</div> <div>3,890,677</div> <div>3,827,860</div> <div>3,812,460</div> <div>3,892,338</div>	
<div>49</div> <div> <a href="/wiki/Cincinnati/Northern_Kentucky_International_Airport" title="Cincinnati/Northern Kentucky International Airport">Cincinnati/Northern Kentucky International Airport</a> </div>	

<td>CVG	
</td>	
<td><a href="/wiki/Cincinnati" title="Cincinnati">Cincinnati</a> & <a href="/wiki/Covington,_Kentucky" title="Covington, Kentucky">Covington</a>	
</td>	
<td>KY	
</td>	
<td>3,050,597	
</td>	
<td>1,729,395	
</td>	
<td>4,413,457	
</td>	
<td>4,269,258	
</td>	
<td>3,926,158	
</td>	
<td>3,269,979	
</td>	
<td>3,036,697	
</td>	
<td>2,874,684	
</td>	
<td>2,776,377	
</td>	
<td>2,937,850	
</td></tr>	
<tr>	
<td>50	
</td>	
<td><a href="/wiki/Kahului_Airport" title="Kahului Airport">Kahului Airport</a>	
</td>	
<td>OGG	
</td>	
<td><a href="/wiki/Kahului,_Hawaii" title="Kahului, Hawaii">Kahului</a>	
</td>	
<td>HI	
</td>	
<td>2,933,315	
</td>	
<td>1,135,141	
</td>	
<td>3,791,807	
</td>	
<td>3,572,133	
</td>	
<td>3,442,189	
</td>	

<td>3,352,813
</td>
<td>3,220,753
</td>
<td>3,019,338
</td>
<td>2,955,304
</td>
<td>2,861,278
</td></tr>
<tr>
<td>51
</td>
<td><a href="/wiki/John_Glenn_Columbus_International_Airport" title="John Glenn Columbus International Airport">John Glenn Columbus International Airport</a>
</td>
<td>CMH
</td>
<td><a href="/wiki/Columbus,_Ohio" title="Columbus, Ohio">Columbus</a>
</td>
<td>OH
</td>
<td>2,825,259
</td>
<td>1,577,596
</td>
<td>4,172,067
</td>
<td>4,054,572
</td>
<td>3,765,007
</td>
<td>3,567,864
</td>
<td>3,312,496
</td>
<td>3,115,501
</td>
<td>3,063,822
</td>
<td>3,095,575
</td></tr>
<tr>
<td>52
</td>
<td><a href="/wiki/Palm_Beach_International_Airport" title="Palm Beach International Airport">Palm Beach International Airport</a>
</td>

<div>PBI</div> <div> <a href="/wiki/West_Palm_Beach,_Florida" title="West Palm Beach, Florida">West Palm Beach</a> </div> <div>FL</div> <div>2,567,897</div> <div>1,518,732</div> <div>3,460,429</div> <div>3,263,042</div> <div>3,166,532</div> <div>3,100,624</div> <div>3,113,485</div> <div>2,926,242</div> <div>2,844,507</div> <div>2,796,359</div>
<div>53</div> <div> <a href="/wiki/Jacksonville_International_Airport" title="Jacksonville International Airport">Jacksonville International Airport</a> </div> <div>JAX</div> <div> <a href="/wiki/Jacksonville,_Florida" title="Jacksonville, Florida">Jacksonville</a> </div> <div>FL</div> <div>2,425,685</div> <div>1,367,501</div> <div>3,479,923</div> <div>3,118,540</div>

```

<td>2,759,067
</td>
<td>2,799,587
</td>
<td>2,716,465
</td>
<td>2,589,198
</td>
<td>2,549,070
</td>
<td>2,579,023
</td></tr>
<tr>
<td>54
</td>
<td><a href="/wiki/Bradley_International_Airport" title="Bradley International
Airport">Bradley International Airport</a>
</td>
<td>BDL
</td>
<td><a href="/wiki/Hartford,_Connecticut" title="Hartford,
Connecticut">Hartford</a>
</td>
<td>CT
</td>
<td>2,273,259
</td>
<td>1,150,033
</td>
<td>3,323,614
</td>
<td>3,330,734
</td>
<td>3,214,976
</td>
<td>2,982,194
</td>
<td>2,926,047
</td>
<td>2,913,380
</td>
<td>2,681,181
</td>
<td>2,647,610
</td></tr>
<tr>
<td>55
</td>

```

<a href="/wiki/Milwaukee_Mitchell_International_Airport" title="Milwaukee Mitchell International Airport">Milwaukee Mitchell International Airport</a>	
MKE	
<a class="mw-redirect" href="/wiki/Milwaukee,_Wisconsin" title="Milwaukee, Wisconsin">Milwaukee</a>	
WI	
2,231,010	
1,263,385	
3,374,073	
3,548,817	
3,452,544	
3,383,271	
3,229,876	
3,228,607	
3,214,811	
3,710,384	
56	
<a href="/wiki/Ontario_International_Airport" title="Ontario International Airport">Ontario International Airport</a>	
ONT	
<a href="/wiki/Ontario,_California" title="Ontario, California">Ontario</a>	
CA	
2,201,528	
1,237,946	
2,723,002	



```

<td>2,499,171
</td>
<td>2,247,645
</td>
<td>2,127,387
</td>
<td>2,089,801
</td>
<td>2,037,346
</td>
<td>1,970,538
</td>
<td>2,142,393
</td></tr>
<tr>
<td>57
</td>
<td><a href="/wiki/Ted_Stevens_Anchorage_International_Airport" title="Ted
Stevens Anchorage International Airport">Ted Stevens Anchorage International
Airport</a>
</td>
<td>ANC
</td>
<td><a href="/wiki/Anchorage,_Alaska" title="Anchorage, Alaska">Anchorage</a>
</td>
<td>AK
</td>
<td>2,184,959
</td>
<td>1,157,301
</td>
<td>2,713,843
</td>
<td>2,642,901
</td>
<td>2,556,188
</td>
<td>2,563,524
</td>
<td>2,525,876
</td>
<td>2,381,826
</td>
<td>2,325,030
</td>
<td>2,249,717
</td></tr>
<tr>

```

58	<a href="/wiki/Charleston_International_Airport" title="Charleston International Airport">Charleston International Airport</a>	CHS	<a href="/wiki/Charleston,_South_Carolina" title="Charleston, South Carolina">Charleston</a>	SC	2,015,277	944,660	2,375,868	2,192,893	1,945,699	1,811,695	1,669,988	1,539,326	1,441,415	1,283,970
59	<a href="/wiki/Hollywood_Burbank_Airport" title="Hollywood Burbank Airport">Hollywood Burbank Airport</a>	BUR	<a href="/wiki/Burbank,_California" title="Burbank, California">Burbank</a>	CA	1,942,417	1,056,838								

<td>2,988,720
</td>
<td>2,680,240
</td>
<td>2,402,106
</td>
<td>2,077,892
</td>
<td>1,973,897
</td>
<td>1,928,491
</td>
<td>1,918,011
</td>
<td>2,027,203
</td></tr>
<tr>
<td>60
</td>
<td><a href="/wiki/Eppley_Airfield" title="Eppley Airfield">Eppley Airfield</a>
</td>
<td>OMA
</td>
<td><a href="/wiki/Omaha,_Nebraska" title="Omaha, Nebraska">Omaha</a>
</td>
<td>NE
</td>
<td>1,829,912
</td>
<td>1,036,245
</td>
<td>2,455,274
</td>
<td>2,457,087
</td>
<td>2,303,223
</td>
<td>2,127,387
</td>
<td>2,046,155
</td>
<td>2,020,354
</td>
<td>1,975,339
</td>
<td>2,018,738
</td></tr>
<tr>

61	<a href="/wiki/Boise_Airport" title="Boise Airport">Boise Airport</a> B0I <a href="/wiki/Boise,_Idaho" title="Boise, Idaho">Boise</a> ID 1,809,000 991,241 2,057,750 1,943,181 1,777,642 1,633,507 1,487,777 1,378,352 1,313,741 1,307,505
62	<a href="/wiki/Memphis_International_Airport" title="Memphis International Airport">Memphis International Airport</a> MEM <a href="/wiki/Memphis,_Tennessee" title="Memphis, Tennessee">Memphis</a> TN 1,793,073 1,015,981 2,318,442

```

<td>2,213,083
</td>
<td>2,102,739
</td>
<td>2,016,089
</td>
<td>1,873,716
</td>
<td>1,800,268
</td>
<td>2,301,003
</td>
<td>3,359,668
</td></tr>
<tr>
<td>63
</td>
<td><a href="/wiki/Reno%E2%80%93Tahoe_International_Airport" title="Reno-Tahoe
International Airport">Reno-Tahoe International Airport</a>
</td>
<td>RNO
</td>
<td><a href="/wiki/Reno,_Nevada" title="Reno, Nevada">Reno</a>
</td>
<td>NV
</td>
<td>1,781,785
</td>
<td>976,937
</td>
<td>2,162,250
</td>
<td>2,048,916
</td>
<td>1,953,028
</td>
<td>1,771,864
</td>
<td>1,669,876
</td>
<td>1,611,572
</td>
<td>1,671,926
</td>
<td>1,685,333
</td></tr>
<tr>
<td>64

```

```

</td>
<td><a href="/wiki/Albuquerque_International_Sunport" title="Albuquerque
International Sunport">Albuquerque International Sunport</a>
</td>
<td>ABQ
</td>
<td><a href="/wiki/Albuquerque,_New_Mexico" title="Albuquerque, New
Mexico">Albuquerque</a>
</td>
<td>NM
</td>
<td>1,688,646
</td>
<td>868,922
</td>
<td>2,641,450
</td>
<td>2,647,269
</td>
<td>2,412,328
</td>
<td>2,341,719
</td>
<td>2,323,883
</td>
<td>2,354,184
</td>
<td>2,477,783
</td>
<td>2,630,574
</td></tr>
<tr>
<td>65
</td>
<td><a href="/wiki/Norfolk_International_Airport" title="Norfolk International
Airport">Norfolk International Airport</a>
</td>
<td>ORF
</td>
<td><a href="/wiki/Norfolk,_Virginia" title="Norfolk, Virginia">Norfolk</a>
</td>
<td>VA
</td>
<td>1,658,024
</td>
<td>884,882
</td>
<td>1,990,864

```

```

</td>
<td>1,846,031
</td>
<td>1,694,329
</td>
<td>1,602,631
</td>
<td>1,515,200
</td>
<td>1,488,114
</td>
<td>1,560,754
</td>
<td>1,651,440
</td></tr></tbody></table>

```

```

[53]: tables = pd.read_html(str(hub[0]))
      print(tables[0])

```

	Rank(2021)	Airports (large hubs)	IATACode	\
0	1	Hartsfield-Jackson Atlanta International Airport	ATL	
1	2	Dallas/Fort Worth International Airport	DFW	
2	3	Denver International Airport	DEN	
3	4	O'Hare International Airport	ORD	
4	5	Los Angeles International Airport	LAX	
..	...			
60	61	Boise Airport	BOI	
61	62	Memphis International Airport	MEM	
62	63	Reno-Tahoe International Airport	RNO	
63	64	Albuquerque International Sunport	ABQ	
64	65	Norfolk International Airport	ORF	

	Major cities served	State	2021[3]	2020[4]	2019[5]	2018[6]	\
0	Atlanta	GA	36676010	20559866	53505795	51865797	
1	Dallas & Fort Worth	TX	30005266	18593421	35778573	32821799	
2	Denver	CO	28645527	16243216	33592945	31362941	
3	Chicago	IL	26350976	14606034	40871223	39873927	
4	Los Angeles	CA	23663410	14055777	42939104	42624050	
..	...	...	...	...	...	...	
60	Boise	ID	1809000	991241	2057750	1943181	
61	Memphis	TN	1793073	1015981	2318442	2213083	
62	Reno	NV	1781785	976937	2162250	2048916	
63	Albuquerque	NM	1688646	868922	2641450	2647269	
64	Norfolk	VA	1658024	884882	1990864	1846031	

	2017[7]	2016[8]	2015[9]	2014[10]	2013[11]	2012[12]
0	50251964	50501858	49340732	46604273	45308407	45798928
1	31816933	31283579	31589839	30804567	29038128	28022904

```

2  29809097  28267394  26280043  26000591  25496885  25799841
3  38593028  37589899  36305668  33843426  32317835  32171795
4  41232432  39636042  36351272  34314197  32425892  31326268
..      ...      ...      ...      ...      ...
60  1777642  1633507  1487777  1378352  1313741  1307505
61  2102739  2016089  1873716  1800268  2301003  3359668
62  1953028  1771864  1669876  1611572  1671926  1685333
63  2412328  2341719  2323883  2354184  2477783  2630574
64  1694329  1602631  1515200  1488114  1560754  1651440

```

[65 rows x 15 columns]

```

[54]: air_hub = []
      for i in range(0,2):
          tables = pd.read_html(str(hub[i]))
          air_hub.append(pd.DataFrame(tables[0]))

```

```

[55]: large_hubs = air_hub[0]
      large_hubs.head()

```

```

[55]: Rank(2021)                Airports (large hubs) IATACode \
0      1  Hartsfield-Jackson Atlanta International Airport  ATL
1      2      Dallas/Fort Worth International Airport      DFW
2      3              Denver International Airport          DEN
3      4      O'Hare International Airport                  ORD
4      5      Los Angeles International Airport              LAX

```

```

      Major cities served State  2021[3]  2020[4]  2019[5]  2018[6] \
0      Atlanta      GA  36676010  20559866  53505795  51865797
1  Dallas & Fort Worth  TX  30005266  18593421  35778573  32821799
2      Denver      CO  28645527  16243216  33592945  31362941
3      Chicago      IL  26350976  14606034  40871223  39873927
4      Los Angeles  CA  23663410  14055777  42939104  42624050

```

```

      2017[7]  2016[8]  2015[9]  2014[10]  2013[11]  2012[12]
0  50251964  50501858  49340732  46604273  45308407  45798928
1  31816933  31283579  31589839  30804567  29038128  28022904
2  29809097  28267394  26280043  26000591  25496885  25799841
3  38593028  37589899  36305668  33843426  32317835  32171795
4  41232432  39636042  36351272  34314197  32425892  31326268

```

```

[56]: large_hubs['Hub Size'] = "Large Hub"

```

```

[57]: large_hubs.rename(columns = {'Airports (large hubs)': 'airport'}, inplace = True)

```

```

[58]: large_hubs.rename(columns = {'Major cities served': 'city'}, inplace = True)

```



```
[59]: large_hubs
```

```
[59]:      Rank(2021)                                airport IATACode \
0          1  Hartsfield-Jackson Atlanta International Airport      ATL
1          2      Dallas/Fort Worth International Airport          DFW
2          3              Denver International Airport            DEN
3          4              O'Hare International Airport            ORD
4          5      Los Angeles International Airport              LAX
..         ...
60         61              Boise Airport                        BOI
61         62      Memphis International Airport              MEM
62         63      Reno-Tahoe International Airport            RNO
63         64      Albuquerque International Sunport          ABQ
64         65      Norfolk International Airport              ORF
```

```

              city State  2021[3]  2020[4]  2019[5]  2018[6] \
0          Atlanta   GA  36676010  20559866  53505795  51865797
1  Dallas & Fort Worth TX  30005266  18593421  35778573  32821799
2          Denver   CO  28645527  16243216  33592945  31362941
3          Chicago  IL  26350976  14606034  40871223  39873927
4      Los Angeles  CA  23663410  14055777  42939104  42624050
..         ...     ...
60          Boise   ID  1809000   991241   2057750   1943181
61          Memphis TN  1793073   1015981   2318442   2213083
62          Reno    NV  1781785   976937   2162250   2048916
63      Albuquerque NM  1688646   868922   2641450   2647269
64          Norfolk VA  1658024   884882   1990864   1846031
```

```

      2017[7]  2016[8]  2015[9]  2014[10]  2013[11]  2012[12]  Hub Size
0  50251964  50501858  49340732  46604273  45308407  45798928  Large Hub
1  31816933  31283579  31589839  30804567  29038128  28022904  Large Hub
2  29809097  28267394  26280043  26000591  25496885  25799841  Large Hub
3  38593028  37589899  36305668  33843426  32317835  32171795  Large Hub
4  41232432  39636042  36351272  34314197  32425892  31326268  Large Hub
..         ...     ...
60  1777642  1633507  1487777  1378352  1313741  1307505  Large Hub
61  2102739  2016089  1873716  1800268  2301003  3359668  Large Hub
62  1953028  1771864  1669876  1611572  1671926  1685333  Large Hub
63  2412328  2341719  2323883  2354184  2477783  2630574  Large Hub
64  1694329  1602631  1515200  1488114  1560754  1651440  Large Hub
```

```
[65 rows x 16 columns]
```

```
[60]: medium_hubs = air_hub[1]
      medium_hubs.head()
```

```
[60]:
```

	Rank	Rank change		Airport name \
	Rank	Rank change		Airport name
0	1	NaN	Hartsfield-Jackson Atlanta International Airport	
1	2	NaN	Dallas/Fort Worth International Airport	
2	3	NaN	Denver International Airport	
3	4	NaN	O'Hare International Airport	
4	5	NaN	Los Angeles International Airport	

	Location	IATA Code	Traffic	Aircraft \
	Location	IATA Code	Passengers % chg.2019/20	Movements
0	College Park, Georgia	ATL	93699630	23.8 NaN
1	Irving, Texas	DFW	73362946	17.4 NaN
2	Denver, Colorado	DEN	69286461	17.8 NaN
3	Chicago, Illinois	ORD	68340619	26.5 NaN
4	Los Angeles, California	LAX	65924298	37.3 NaN

```
% chg.2019/20
```

0	0.0
1	NaN
2	NaN
3	NaN
4	NaN

```
[61]: medium_hubs['Hub Size'] = "Medium Hub"
```

```
[75]: medium_hubs.rename(columns = {'Airports (medium hubs)': 'airport'}, inplace = True)
medium_hubs.rename(columns = {'City served': 'city'}, inplace = True)
medium_hubs.head()
```

```
[75]:
```

	Rank	Rank change		Airport name \
	Rank	Rank change		Airport name
0	1	NaN	Hartsfield-Jackson Atlanta International Airport	
1	2	NaN	Dallas/Fort Worth International Airport	
2	3	NaN	Denver International Airport	
3	4	NaN	O'Hare International Airport	
4	5	NaN	Los Angeles International Airport	

	Location	IATA Code	Traffic	Aircraft \
	Location	IATA Code	Passengers % chg.2019/20	Movements
0	College Park, Georgia	ATL	93699630	23.8 NaN
1	Irving, Texas	DFW	73362946	17.4 NaN
2	Denver, Colorado	DEN	69286461	17.8 NaN
3	Chicago, Illinois	ORD	68340619	26.5 NaN
4	Los Angeles, California	LAX	65924298	37.3 NaN

	% chg.2019/20	Hub Size
0	0.0	Medium Hub
1	NaN	Medium Hub
2	NaN	Medium Hub
3	NaN	Medium Hub
4	NaN	Medium Hub

```
[61]: final_hub_data = pd.concat([large_hubs, medium_hubs])
final_hub_data
```

```
[61]: Rank(2021)                                airport IATACode \
0      1  Hartsfield-Jackson Atlanta International Airport      ATL
1      2      Dallas/Fort Worth International Airport          DFW
2      3      Denver International Airport                    DEN
3      4      O'Hare International Airport                    ORD
4      5      Los Angeles International Airport                LAX
..      ..
30     61      Boise Airport                                  BOI
31     62      Memphis International Airport                  MEM
32     63      Reno-Tahoe International Airport               RNO
33     64      Albuquerque International Sunport              ABQ
34     65      Norfolk International Airport                  ORF
```

	city	State	2021[3]	2020[4]	2019[5]	2018[6]	\
0	Atlanta	GA	36676010	20559866	53505795	51865797	
1	Dallas & Fort Worth	TX	30005266	18593421	35778573	32821799	
2	Denver	CO	28645527	16243216	33592945	31362941	
3	Chicago	IL	26350976	14606034	40871223	39873927	
4	Los Angeles	CA	23663410	14055777	42939104	42624050	
..	...	...	...	...	...	...	
30	Boise	ID	1809000	991241	2057750	1943181	
31	Memphis	TN	1793073	1015981	2318442	2213083	
32	Reno	NV	1781785	976937	2162250	2048916	
33	Albuquerque	NM	1688646	868922	2641450	2647269	
34	Norfolk	VA	1658024	884882	1990864	1846031	

	2017[7]	2016[8]	2015[9]	2014[10]	2013[11]	2012[12]	Hub Size
0	50251964	50501858	49340732	46604273	45308407	45798928	Large Hub
1	31816933	31283579	31589839	30804567	29038128	28022904	Large Hub
2	29809097	28267394	26280043	26000591	25496885	25799841	Large Hub
3	38593028	37589899	36305668	33843426	32317835	32171795	Large Hub
4	41232432	39636042	36351272	34314197	32425892	31326268	Large Hub
..	...	...	...	...	...	...	
30	1777642	1633507	1487777	1378352	1313741	1307505	Medium Hub
31	2102739	2016089	1873716	1800268	2301003	3359668	Medium Hub
32	1953028	1771864	1669876	1611572	1671926	1685333	Medium Hub

```

33  2412328  2341719  2323883  2354184  2477783  2630574  Medium Hub
34  1694329  1602631  1515200  1488114  1560754  1651440  Medium Hub

```

[65 rows x 16 columns]

```

[62]: column_temp = final_hub_data.columns.str.split('([ ]').str[0].str.strip().str.
      ↪lower().str.replace(' ','_').values
      column_temp[list(map( lambda x : x.isnumeric(), column_temp))] = 'data_' +
      ↪column_temp[list(map( lambda x : x.isnumeric(), column_temp))]
      final_hub_data.columns = column_temp
      final_hub_data.columns

```

```

[62]: Index(['rank', 'airport', 'iatacode', 'city', 'state', 'data_2021',
            'data_2020', 'data_2019', 'data_2018', 'data_2017', 'data_2016',
            'data_2015', 'data_2014', 'data_2013', 'data_2012', 'hub_size'],
            dtype='object')

```

```

[63]: final_hub_data.head(3)

```

```

[63]:      rank      airport iatacode \
0      1  Hartsfield-Jackson Atlanta International Airport      ATL
1      2      Dallas/Fort Worth International Airport      DFW
2      3      Denver International Airport      DEN

      city state  data_2021  data_2020  data_2019  data_2018 \
0      Atlanta  GA   36676010   20559866   53505795   51865797
1  Dallas & Fort Worth  TX   30005266   18593421   35778573   32821799
2      Denver  CO   28645527   16243216   33592945   31362941

      data_2017  data_2016  data_2015  data_2014  data_2013  data_2012  hub_size
0   50251964   50501858   49340732   46604273   45308407   45798928  Large Hub
1   31816933   31283579   31589839   30804567   29038128   28022904  Large Hub
2   29809097   28267394   26280043   26000591   25496885   25799841  Large Hub

```

```

[64]: final_hub_data.isna().sum()

```

```

[64]: rank      0
      airport    0
      iatacode    0
      city      0
      state     0
      data_2021  0
      data_2020  0
      data_2019  0
      data_2018  0
      data_2017  0
      data_2016  0

```

```

data_2015    0
data_2014    0
data_2013    0
data_2012    0
hub_size     0
dtype: int64

```

```
[65]: final_hub_data.to_csv('hub_data.csv')
```

```
[65]: master_df = pd.merge(df, final_hub_data[['iatacode', 'data_2021']],how = 'left'
↳, left_on = 'AirportFrom', right_on = 'iatacode')
```

```
[66]: master_df.rename(columns = {'iatacode': 'iatacode_source' , 'data_2021':
↳ 'data_2021_source_airport'}, inplace = True)
```

```
[67]: master_df = pd.merge(master_df, final_hub_data[['iatacode', 'data_2021']],how =
↳ 'left' , left_on = 'AirportTo', right_on = 'iatacode')
```

```
[68]: master_df.rename(columns = {'iatacode': 'iatacode_dest' , 'data_2021':
↳ 'data_2021_dest_airport'}, inplace = True)
```

```
[69]: master_df.head(3)
```

```
[69]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	2	US	1558	PHX	CLT	3	15	222	1	
2	3	AA	2400	LAX	DFW	3	20	165	1	

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	1135.0	
2	large_airport	125.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	3.0	large_airport	748.0	
2	4.0	large_airport	607.0	

	runway_count_dest_airport	iatacode_source	data_2021_source_airport	\
0	5.0	SFO	11725347.0	
1	4.0	PHX	18940287.0	
2	7.0	LAX	23663410.0	

	iatacode_dest	data_2021_dest_airport
0	IAH	16242821.0
1	CLT	20900875.0
2	DFW	30005266.0

```
[70]: master_df.drop(['iatacode_source', 'iatacode_dest'], axis=1, inplace=True)
```

```
[71]: significant_airlines
```

```
[71]:
```

	Airline	IATA	Founded
0	CO	NaN	NaN
1	US	NaN	NaN
2	AA	AA	1926.0
3	AS	AS	1932.0
4	DL	DL	1924.0
5	B6	B6	1998.0
6	HA	HA	1929.0
7	OO	OO	1972.0
8	9E	9E	1985.0
9	OH	OH	1979.0
10	EV	NaN	NaN
11	XE	XE	2016.0
12	YV	YV	1980.0
13	UA	UA	1926.0
14	MQ	MQ	1984.0
15	F9	F9	1994.0
16	WN	WN	1967.0

```
[72]: master_df = pd.merge(master_df, significant_airlines[['Airline', 'Founded']],
    ↪ on = 'Airline')
```

```
[73]: master_df.head()
```

```
[73]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	6	CO	1094	LAX	IAH	3	30	181	1	
2	11	CO	223	ANC	SEA	3	49	201	1	
3	18	CO	1496	LAS	IAH	3	60	162	0	
4	20	CO	507	ONT	IAH	3	75	167	0	

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	125.0	
2	large_airport	152.0	
3	large_airport	2181.0	
4	large_airport	944.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	4.0	large_airport	97.0	
2	3.0	large_airport	433.0	
3	4.0	large_airport	97.0	

4	2.0	large_airport	97.0
---	-----	---------------	------

	runway_count_dest_airport	data_2021_source_airport \
0	5.0	11725347.0
1	5.0	23663410.0
2	4.0	2184959.0
3	5.0	19160342.0
4	5.0	2201528.0

	data_2021_dest_airport	Founded
0	16242821.0	NaN
1	16242821.0	NaN
2	17430195.0	NaN
3	16242821.0	NaN
4	16242821.0	NaN

**0.1 2.** You should then examine the missing values in each field, perform missing value treatment, and justify your actions.

```
[74]: master_df.isna().sum()
```

```
[74]: id                0
      Airline           0
      Flight           0
      AirportFrom       0
      AirportTo         0
      DayOfWeek         0
      Time              0
      Length            0
      Delay             0
      type_source_airport 31
      elevation_ft_source_airport 31
      runway_count_source_airport 31
      type_dest_airport 31
      elevation_ft_dest_airport 31
      runway_count_dest_airport 31
      data_2021_source_airport 83582
      data_2021_dest_airport 83531
      Founded          83601
      dtype: int64
```

```
[75]: master_df[master_df.type_source_airport.isna()].AirportFrom.unique()
```

```
[75]: array(['CYS'], dtype=object)
```

```
[76]: master_df[master_df.type_dest_airport.isna()].AirportTo.unique()
```

```
[76]: array(['CYS'], dtype=object)
```

```
[77]: data_dict = pd.read_excel('Data Dictionary.xlsx', sheet_name = 'airlines', header = 29, usecols = [0,1])
data_dict.head()
```

```
[77]:
```

	Aiport ID	Description
0	ABE	RAF Calveley
1	ABE	Bisho Airport
2	ABE	Beica Airport
3	ABE	Lehigh Valley International Airport
4	ABE	Bethel Airport

```
[78]: name = data_dict[data_dict['Aiport ID'] == 'CYS'].Description.values[0]
name.lower()
```

```
[78]: 'cheyenne regional jerry olson field'
```

```
[79]: air_miss = airports.loc[name.lower() == airports.name.str.lower(), ['ident', 'name', 'iata_code', 'type', 'elevation_ft']]
```

```
[80]: air_miss
```

```
[80]:
```

	ident	name	iata_code	type	\
34675	KCYS	Cheyenne Regional Jerry Olson Field	NaN	medium_airport	

	elevation_ft
34675	6159.0

```
[81]: air_miss_comb = pd.merge(air_miss, runways[['airport_ident', 'id']], how = 'left', left_on = 'ident', right_on = 'airport_ident')
runway_count_miss = air_miss_comb.groupby('ident')[['id']].count().sort_values(by = 'id', ascending = False).reset_index()
runway_count_miss
```

```
[81]:
```

	ident	id
0	KCYS	2

```
[82]: air_miss_data = pd.merge(air_miss, runway_count_miss).rename(columns = {'id' : 'runway_count'})[['iata_code', 'type', 'elevation_ft', 'runway_count']]
```

```
[83]: air_miss_data
```

```
[83]:
```

	iata_code	type	elevation_ft	runway_count
0	NaN	medium_airport	6159.0	2



```
[84]: master_df.loc[master_df.AirportFrom == 'CYS', 'type_source_airport'] =_
      ↪air_miss_data.type.values[0]
      master_df.loc[master_df.AirportFrom == 'CYS', 'elevation_ft_source_airport'] =_
      ↪air_miss_data.elevation_ft.values[0]
      master_df.loc[master_df.AirportFrom == 'CYS', 'runway_count_source_airport'] =_
      ↪air_miss_data.runway_count.values[0]
```

```
[85]: master_df.loc[master_df.AirportTo == 'CYS', 'type_dest_airport'] =_
      ↪air_miss_data.type.values[0]
      master_df.loc[master_df.AirportTo == 'CYS', 'elevation_ft_dest_airport'] =_
      ↪air_miss_data.elevation_ft.values[0]
      master_df.loc[master_df.AirportTo == 'CYS', 'runway_count_dest_airport'] =_
      ↪air_miss_data.runway_count.values[0]
```

```
[86]: master_df.isna().sum().sort_values(ascending = False)
```

```
[86]: Founded                83601
      data_2021_source_airport  83582
      data_2021_dest_airport   83531
      Airline                  0
      runway_count_dest_airport 0
      elevation_ft_dest_airport 0
      type_dest_airport         0
      runway_count_source_airport 0
      elevation_ft_source_airport 0
      id                       0
      Delay                    0
      Length                   0
      Time                     0
      DayOfWeek                0
      AirportTo                0
      AirportFrom              0
      Flight                   0
      type_source_airport       0
      dtype: int64
```

```
[87]: miss_founded = master_df[master_df.Founded.isna()].Airline.unique()
      miss_founded
```

```
[87]: array(['C0', 'US', 'EV'], dtype=object)
```

```
[88]: airline_dict = pd.read_excel('Data Dictionary.xlsx', sheet_name =_
      ↪'airlines', header = 10, usecols = [0,1])
      airline_dict.head()
```

```
[88]: Airlines ID      Description
      0      WN      Southwest
      1      DL      Delta
      2      OO      Skywest
      3      AA      American Airlines
      4      MQ      Envoy
```

```
[89]: print(airline_dict[airline_dict['Airlines ID'].isin( ['EV', 'CO', 'US'])])
```

```
Airlines ID      Description
5      US      PSA (initially US Airway Express)
7      EV      ExpressJet
9      CO      United Airlines (initially CO)
```

### 0.1.1 use google for these years

- US - 1967
- CO - 1934
- EV - 1986

```
[90]: miss_val = {'US' : 1967, 'CO' : 1934, 'EV' : 1986}
      for aline in miss_founded:
          master_df.loc[(master_df.Founded.isna()) &
                        (master_df.Airline == aline), 'Founded'] = miss_val[aline]
```

```
[91]: master_df.isna().sum().sort_values(ascending = False)
```

```
[91]: data_2021_source_airport      83582
      data_2021_dest_airport      83531
      id                          0
      Airline                     0
      runway_count_dest_airport    0
      elevation_ft_dest_airport    0
      type_dest_airport            0
      runway_count_source_airport  0
      elevation_ft_source_airport  0
      type_source_airport          0
      Delay                       0
      Length                     0
      Time                       0
      DayOfWeek                   0
      AirportTo                   0
      AirportFrom                 0
      Flight                     0
      Founded                     0
      dtype: int64
```

```
[92]: (master_df.isna().sum().sort_values(ascending = False)/master_df.shape[0])*100
```

```
[92]: data_2021_source_airport      16.118221
      data_2021_dest_airport      16.108386
      id                          0.000000
      Airline                     0.000000
      runway_count_dest_airport   0.000000
      elevation_ft_dest_airport   0.000000
      type_dest_airport           0.000000
      runway_count_source_airport 0.000000
      elevation_ft_source_airport 0.000000
      type_source_airport         0.000000
      Delay                      0.000000
      Length                     0.000000
      Time                       0.000000
      DayOfWeek                  0.000000
      AirportTo                  0.000000
      AirportFrom                0.000000
      Flight                     0.000000
      Founded                    0.000000
      dtype: float64
```

```
[93]: master_df.groupby('type_source_airport')[['data_2021_source_airport']].median()
```

```
[93]: data_2021_source_airport
      type_source_airport
      large_airport      14514049.0
      medium_airport     2273259.0
      small_airport      NaN
```

```
[94]: med_val = master_df.
      ↳groupby('type_source_airport')[['data_2021_source_airport']].median()
      med_val
```

```
[94]: data_2021_source_airport
      type_source_airport
      large_airport      14514049.0
      medium_airport     2273259.0
      small_airport      NaN
```

```
[95]: for typ in master_df.type_source_airport.unique():
      master_df.loc[(master_df.type_source_airport == typ)& (master_df.
      ↳data_2021_source_airport.isna()),
      'data_2021_source_airport'] = med_val.loc[typ].values[0]
```

```
[96]: med_val_dest = master_df.
      ↳groupby('type_dest_airport')[['data_2021_dest_airport']].median()
```

```
med_val_dest
```

```
[96]: data_2021_dest_airport
type_dest_airport
large_airport      14514049.0
medium_airport     2273259.0
small_airport      NaN
```

```
[97]: for typ in master_df.type_source_airport.unique():
      master_df.loc[(master_df.type_dest_airport == typ)& (master_df.
      ↪data_2021_dest_airport.isna()),
                  'data_2021_dest_airport'] = med_val.loc[typ].values[0]
```

```
[98]: master_df.head(2)
```

```
[98]:   id  Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay  \
0    1      CO     269          SFO         IAH           3    15     205      1
1    6      CO    1094          LAX         IAH           3    30     181      1

      type_source_airport  elevation_ft_source_airport  \
0      large_airport              13.0
1      large_airport             125.0

      runway_count_source_airport  type_dest_airport  elevation_ft_dest_airport  \
0              4.0      large_airport              97.0
1              4.0      large_airport              97.0

      runway_count_dest_airport  data_2021_source_airport  \
0              5.0              11725347.0
1              5.0              23663410.0

      data_2021_dest_airport  Founded
0      16242821.0      1934.0
1      16242821.0      1934.0
```

```
[99]: (master_df.isna().sum().sort_values(ascending = False)/master_df.shape[0])*100
```

```
[99]: data_2021_source_airport      0.226205
data_2021_dest_airport        0.224855
id                             0.000000
Airline                       0.000000
runway_count_dest_airport      0.000000
elevation_ft_dest_airport      0.000000
type_dest_airport              0.000000
runway_count_source_airport    0.000000
elevation_ft_source_airport    0.000000
type_source_airport            0.000000
```

```

Delay          0.000000
Length         0.000000
Time           0.000000
DayOfWeek      0.000000
AirportTo      0.000000
AirportFrom    0.000000
Flight         0.000000
Founded        0.000000
dtype: float64

```

```
[100]: master_df = master_df.dropna()
```

```
[101]: (master_df.isna().sum().sort_values(ascending = False)/master_df.shape[0])*100
```

```

[101]: id          0.0
Airline          0.0
data_2021_dest_airport  0.0
data_2021_source_airport  0.0
runway_count_dest_airport  0.0
elevation_ft_dest_airport  0.0
type_dest_airport  0.0
runway_count_source_airport  0.0
elevation_ft_source_airport  0.0
type_source_airport  0.0
Delay           0.0
Length          0.0
Time            0.0
DayOfWeek       0.0
AirportTo       0.0
AirportFrom     0.0
Flight          0.0
Founded         0.0
dtype: float64

```

## 0.2 3. Perform data visualization and share your insights on the following points:

### 0.2.1 (a) According to the data provided, approximately 70% of Southwest Airlines flights are delayed. Visualize it to compare it with the data of other airlines.

```
[102]: master_df.head()
```

```

[102]:   id  Airline  Flight  AirportFrom  AirportTo  DayOfWeek  Time  Length  Delay  \
0    1      CO     269          SFO         IAH           3    15     205      1
1    6      CO    1094          LAX         IAH           3    30     181      1
2   11      CO     223          ANC         SEA           3    49     201      1

```

3	18	CO	1496	LAS	IAH	3	60	162	0
4	20	CO	507	ONT	IAH	3	75	167	0

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	125.0	
2	large_airport	152.0	
3	large_airport	2181.0	
4	large_airport	944.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	4.0	large_airport	97.0	
2	3.0	large_airport	433.0	
3	4.0	large_airport	97.0	
4	2.0	large_airport	97.0	

	runway_count_dest_airport	data_2021_source_airport	\
0	5.0	11725347.0	
1	5.0	23663410.0	
2	4.0	2184959.0	
3	5.0	19160342.0	
4	5.0	2201528.0	

	data_2021_dest_airport	Founded
0	16242821.0	1934.0
1	16242821.0	1934.0
2	17430195.0	1934.0
3	16242821.0	1934.0
4	16242821.0	1934.0

```
[103]: airline_dict.head()
```

```
[103]: Airlines ID      Description
0      WN      Southwest
1      DL      Delta
2      OO      Skywest
3      AA      American Airlines
4      MQ      Envoy
```

```
[104]: master_df = pd.
↳merge(master_df,airline_dict,left_on="Airline",right_on="Airlines_
↳ID",how='left')
master_df.head()
```

```
[104]: id Airline Flight AirportFrom AirportTo DayOfWeek Time Length Delay \
0 1 CO 269 SFO IAH 3 15 205 1
```

1	6	CO	1094	LAX	IAH	3	30	181	1
2	11	CO	223	ANC	SEA	3	49	201	1
3	18	CO	1496	LAS	IAH	3	60	162	0
4	20	CO	507	ONT	IAH	3	75	167	0

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	125.0	
2	large_airport	152.0	
3	large_airport	2181.0	
4	large_airport	944.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	4.0	large_airport	97.0	
2	3.0	large_airport	433.0	
3	4.0	large_airport	97.0	
4	2.0	large_airport	97.0	

	runway_count_dest_airport	data_2021_source_airport	\
0	5.0	11725347.0	
1	5.0	23663410.0	
2	4.0	2184959.0	
3	5.0	19160342.0	
4	5.0	2201528.0	

	data_2021_dest_airport	Founded	Airlines	ID	\
0	16242821.0	1934.0		CO	
1	16242821.0	1934.0		CO	
2	17430195.0	1934.0		CO	
3	16242821.0	1934.0		CO	
4	16242821.0	1934.0		CO	

	Description
0	United Airlines (initially CO)
1	United Airlines (initially CO)
2	United Airlines (initially CO)
3	United Airlines (initially CO)
4	United Airlines (initially CO)

```
[105]: id_airline = airline_dict.loc[airline_dict['Description'].str.strip().str.
      ↪lower() == 'southwest', 'Airlines ID'].values[0]
      id_airline
```

```
[105]: 'WN'
```

```
[106]: def percent_Delay(x):  
        return round(x.sum()/x.size * 100,2)
```

```
[107]: master_df.groupby('Airline')['Delay'].agg(percent_Delay)
```

```
[107]: Airline  
9E      39.78  
AA      38.84  
AS      33.93  
B6      46.70  
CO      56.58  
DL      45.05  
EV      40.20  
F9      44.92  
HA      32.02  
MQ      34.80  
OH      27.83  
OO      45.44  
UA      32.39  
US      33.60  
WN      69.78  
XE      37.90  
YV      24.37  
Name: Delay, dtype: float64
```

```
[108]: delay_perc = master_df.groupby('Airline')['Delay'].agg(percent_Delay)  
delay_perc = delay_perc.reset_index()
```

```
[109]: delay_perc
```

```
[109]:   Airline  Delay  
0      9E   39.78  
1      AA   38.84  
2      AS   33.93  
3      B6   46.70  
4      CO   56.58  
5      DL   45.05  
6      EV   40.20  
7      F9   44.92  
8      HA   32.02  
9      MQ   34.80  
10     OH   27.83  
11     OO   45.44  
12     UA   32.39  
13     US   33.60  
14     WN   69.78  
15     XE   37.90
```



16 YV 24.37

```
[110]: delay_data = pd.  
      ↪merge(delay_perc,airline_dict,left_on="Airline",right_on="Airlines ID",  
      ↪how='left')[['Airline','Delay','Description']]
```

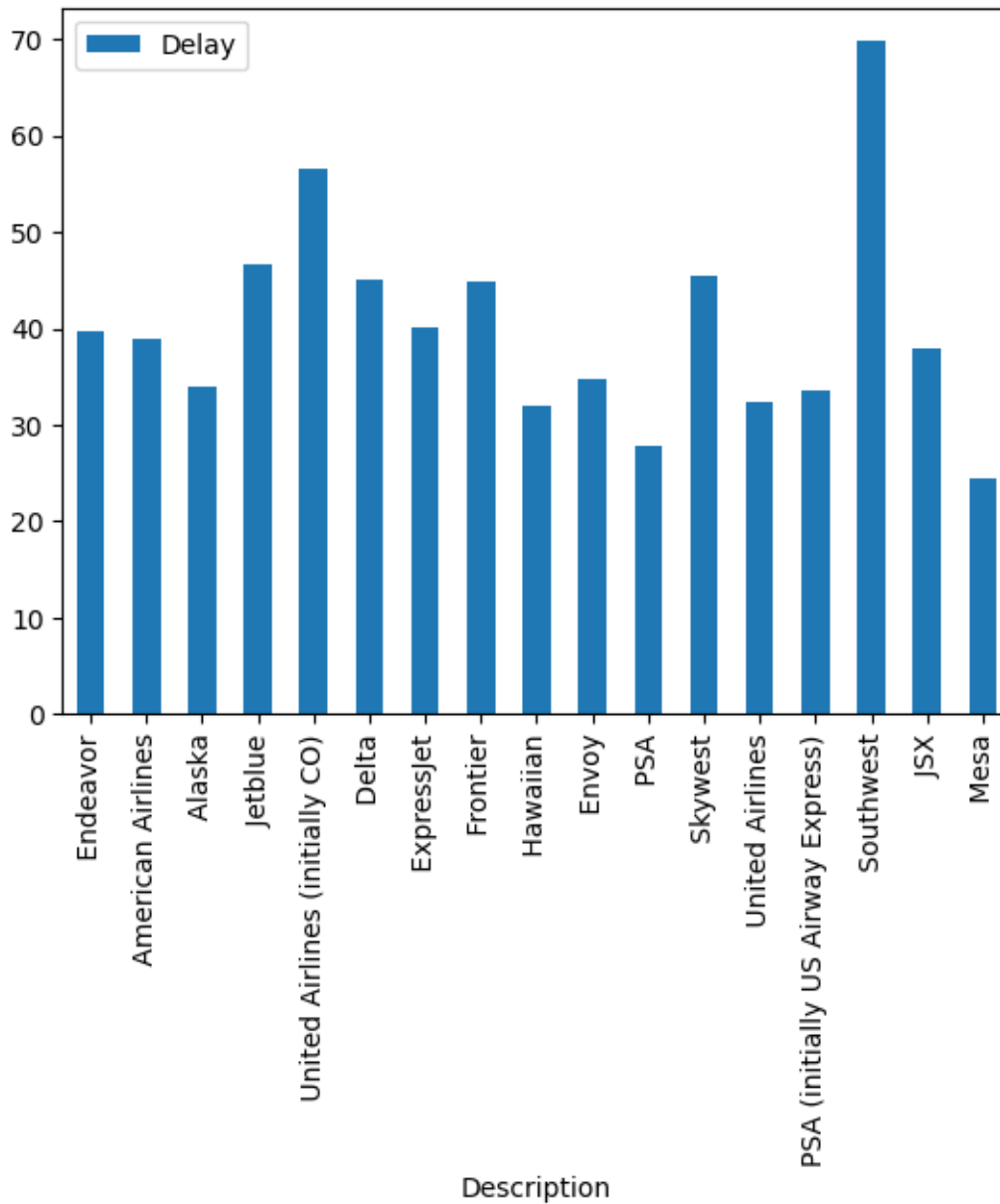
```
[111]: delay_data
```

```
[111]:
```

	Airline	Delay	Description
0	9E	39.78	Endeavor
1	AA	38.84	American Airlines
2	AS	33.93	Alaska
3	B6	46.70	Jetblue
4	CO	56.58	United Airlines (initially CO)
5	DL	45.05	Delta
6	EV	40.20	ExpressJet
7	F9	44.92	Frontier
8	HA	32.02	Hawaiian
9	MQ	34.80	Envoy
10	OH	27.83	PSA
11	OO	45.44	Skywest
12	UA	32.39	United Airlines
13	US	33.60	PSA (initially US Airway Express)
14	WN	69.78	Southwest
15	XE	37.90	JSX
16	YV	24.37	Mesa

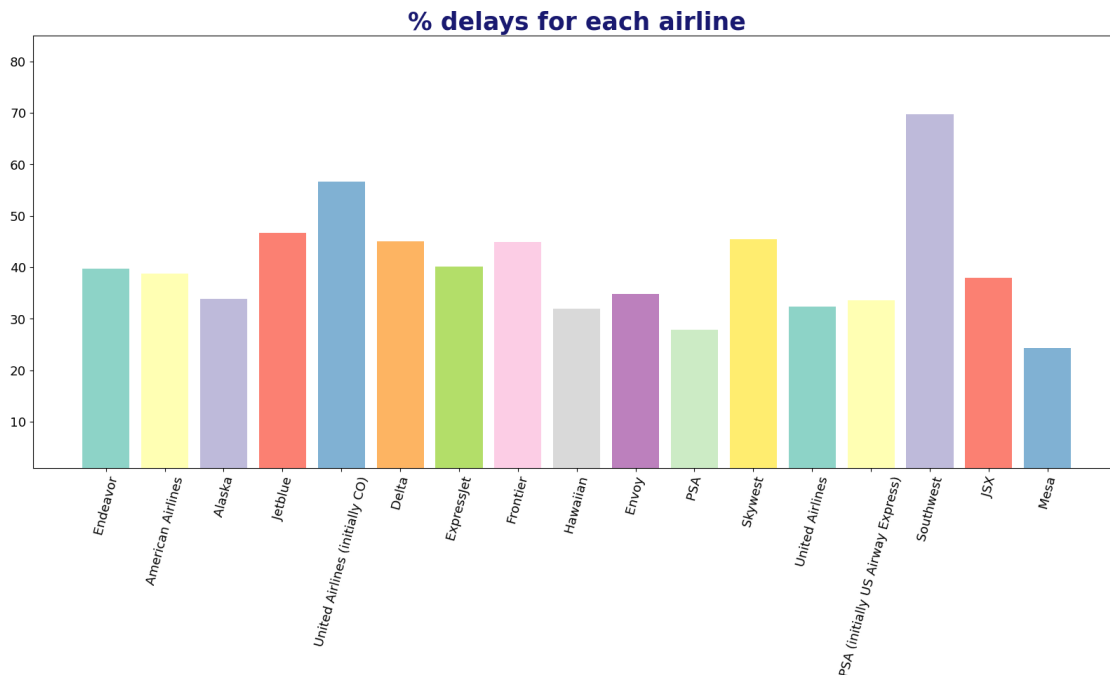
```
[112]: delay_data.plot(x='Description',y='Delay',kind='bar')
```

```
[112]: <AxesSubplot:xlabel='Description'>
```



```
[113]: plt.figure(figsize = (20,8))
plt.bar(delay_data.Description, height = delay_data.Delay, color = plt.
        ↳get_cmap('Set3').colors)
plt.ylim(1,85)
plt.xticks(size = 13, rotation = 75)
plt.yticks(size = 13)
plt.title('% delays for each airline', size = 25, color = 'midnightblue',
        ↳weight = 'heavy', family = 'times')
plt.show()
```

findfont: Font family ['times'] not found. Falling back to DejaVu Sans.



**0.2.2 b. Flights were delayed on various weekdays. Which day of the week is the safest for travel?**

```
[114]: master_df.head()
```

```
[114]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	6	CO	1094	LAX	IAH	3	30	181	1	
2	11	CO	223	ANC	SEA	3	49	201	1	
3	18	CO	1496	LAS	IAH	3	60	162	0	
4	20	CO	507	ONT	IAH	3	75	167	0	

	type_source_airport	elevation_ft_source_airport	\
0	large_airport	13.0	
1	large_airport	125.0	
2	large_airport	152.0	
3	large_airport	2181.0	
4	large_airport	944.0	

	runway_count_source_airport	type_dest_airport	elevation_ft_dest_airport	\
0	4.0	large_airport	97.0	
1	4.0	large_airport	97.0	

2	3.0	large_airport	433.0
3	4.0	large_airport	97.0
4	2.0	large_airport	97.0

	runway_count_dest_airport	data_2021_source_airport \
0	5.0	11725347.0
1	5.0	23663410.0
2	4.0	2184959.0
3	5.0	19160342.0
4	5.0	2201528.0

	data_2021_dest_airport	Founded	Airlines ID \
0	16242821.0	1934.0	CO
1	16242821.0	1934.0	CO
2	17430195.0	1934.0	CO
3	16242821.0	1934.0	CO
4	16242821.0	1934.0	CO

	Description
0	United Airlines (initially CO)
1	United Airlines (initially CO)
2	United Airlines (initially CO)
3	United Airlines (initially CO)
4	United Airlines (initially CO)

```
[115]: master_df['DayOfWeek'].value_counts()
```

```
[115]: 4    87583
      3    86101
      5    81413
      1    69693
      2    68422
      7    66908
      6    56097
      Name: DayOfWeek, dtype: int64
```

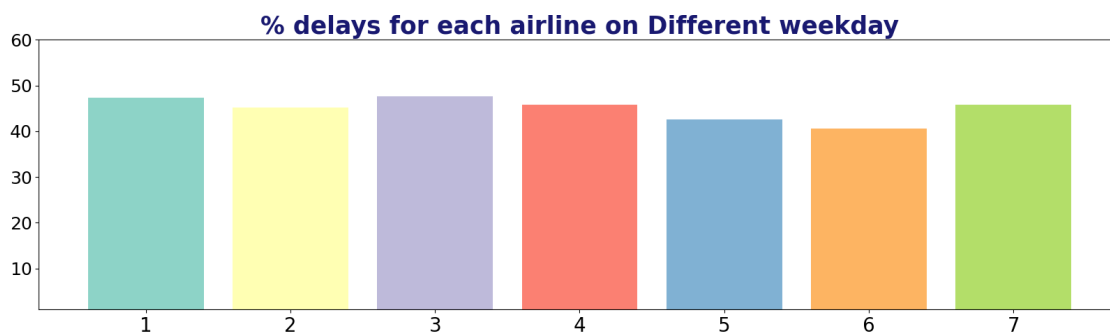
```
[116]: master_df['DayOfWeek'].value_counts()/len(master_df)*100
```

```
[116]: 4    16.966315
      3    16.679226
      5    15.771081
      1    13.500718
      2    13.254503
      7    12.961216
      6    10.866942
      Name: DayOfWeek, dtype: float64
```

```
[117]: delay_perc_weekday = master_df.groupby('DayOfWeek')['Delay'].agg(percent_Delay)
delay_perc_weekday
```

```
[117]: DayOfWeek
1      47.28
2      45.25
3      47.63
4      45.84
5      42.58
6      40.57
7      45.77
Name: Delay, dtype: float64
```

```
[118]: plt.figure(figsize = (20,5))
plt.bar(delay_perc_weekday.index, height = delay_perc_weekday.values, color = _
→plt.get_cmap('Set3').colors)
plt.ylim(1,60)
plt.xticks(size = 20)
plt.yticks(size = 18)
plt.title('% delays for each airline on Different weekday', size = 25, color = _
→'midnightblue', weight = 'heavy', family = 'times')
plt.show()
```



### 0.2.3 (c) Which airlines should be recommended for short-, medium-, and long-distance travel?

```
[119]: duration_data = master_df[['Airline', 'Length', 'Delay']].copy()
```

```
[120]: duration_data['duration'] = pd.cut(duration_data.Length, 3, labels = ['short', _
→'medium', 'long'])
duration_data_grp = duration_data.groupby(['Airline', 'duration'])['Delay'].
→agg(percent_Delay).reset_index().pivot(index = 'Airline',
columns = 'duration').fillna(0)['Delay']
```

```
duration_data_grp.columns = duration_data_grp.columns.astype(str)
duration_data_grp.reset_index()
```

```
[120]: duration Airline  short  medium  long
0          9E  39.78    0.00  0.00
1          AA  37.61   43.25  60.40
2          AS  32.58   38.17  0.00
3          B6  45.70   51.05  0.00
4          CO  52.82   64.95  66.87
5          DL  43.87   50.24  48.62
6          EV  40.20   50.00  0.00
7          F9  45.04   43.56  0.00
8          HA  30.16   40.48  0.00
9          MQ  34.81   27.42  0.00
10         OH  27.71   39.20  0.00
11         OO  45.40   53.03  0.00
12         UA  29.92   37.10  39.26
13         US  31.96   40.72  0.00
14         WN  69.12   77.61  0.00
15         XE  37.87   53.70  0.00
16         YV  24.37   25.86  0.00
```

```
[121]: airline_dict.Description = airline_dict.Description.str.strip()
duration_data_grp = pd.merge(duration_data_grp,airline_dict[['Airlines ID',
↳ 'Description']],
                                left_on = 'Airline', right_on = 'Airlines ID', how_
↳ 'left')
duration_data_grp
```

```
[121]:      short  medium  long Airlines ID      Description
0    39.78    0.00  0.00      9E      Endeavor
1    37.61   43.25  60.40      AA    American Airlines
2    32.58   38.17  0.00      AS      Alaska
3    45.70   51.05  0.00      B6      Jetblue
4    52.82   64.95  66.87      CO    United Airlines (initially CO)
5    43.87   50.24  48.62      DL      Delta
6    40.20   50.00  0.00      EV    ExpressJet
7    45.04   43.56  0.00      F9      Frontier
8    30.16   40.48  0.00      HA      Hawaiian
9    34.81   27.42  0.00      MQ      Envoy
10   27.71   39.20  0.00      OH      PSA
11   45.40   53.03  0.00      OO      Skywest
12   29.92   37.10  39.26      UA    United Airlines
13   31.96   40.72  0.00      US    PSA (initially US Airway Express)
14   69.12   77.61  0.00      WN      Southwest
15   37.87   53.70  0.00      XE      JSX
16   24.37   25.86  0.00      YV      Mesa
```

```
[122]: long = duration_data_grp[duration_data_grp.long == duration_data_grp.long.
↳min()].Description.values.tolist()
print('Airlines with no delays for long flights :\n', ', '.join(long))
medium = duration_data_grp[duration_data_grp.medium == duration_data_grp.medium.
↳min()].Description.values.tolist()
print('\nAirlines with no delays for medium flights :\n', ', '.join(medium))
short = duration_data_grp[duration_data_grp.short == duration_data_grp.short.
↳min()].Description.values.tolist()
print('\nAirlines with no delays for short flights :\n', ', '.join(short) )
```

Airlines with no delays for long flights :

Endeavor, Alaska, Jetblue, ExpressJet, Frontier, Hawaiian, Envoy, PSA, Skywest,  
PSA (initially US Airway Express), Southwest, JSX, Mesa

Airlines with no delays for medium flights :

Endeavor

Airlines with no delays for short flights :

Mesa

#### 0.2.4 (d) Do you notice any patterns in the departure times of long-duration flights?

```
[123]: master_df['duration'] = pd.cut(master_df.Length, 3, labels = ['short', 'medium', 'long'])
master_df.head(3)
```

```
[123]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	6	CO	1094	LAX	IAH	3	30	181	1	
2	11	CO	223	ANC	SEA	3	49	201	1	

	type_source_airport	...	runway_count_source_airport	type_dest_airport	\
0	large_airport	...	4.0	large_airport	
1	large_airport	...	4.0	large_airport	
2	large_airport	...	3.0	large_airport	

	elevation_ft_dest_airport	runway_count_dest_airport	\
0	97.0	5.0	
1	97.0	5.0	
2	433.0	4.0	

	data_2021_source_airport	data_2021_dest_airport	Founded	Airlines	ID	\
0	11725347.0	16242821.0	1934.0		CO	
1	23663410.0	16242821.0	1934.0		CO	
2	2184959.0	17430195.0	1934.0		CO	

	Description	duration
0	United Airlines (initially CO)	short
1	United Airlines (initially CO)	short
2	United Airlines (initially CO)	short

[3 rows x 21 columns]

```
[124]: pd.crosstab(master_df.Time, master_df.duration)['long']
```

```
[124]: Time
10      0
15      0
20      0
21      0
25      0
..
1428    0
1430    0
1431    0
1435    0
1439    0
Name: long, Length: 1131, dtype: int64
```

```
[125]: y = pd.crosstab(master_df.Time, master_df.duration)['long'].index
x = pd.crosstab(master_df.Time, master_df.duration)['long'].values
```

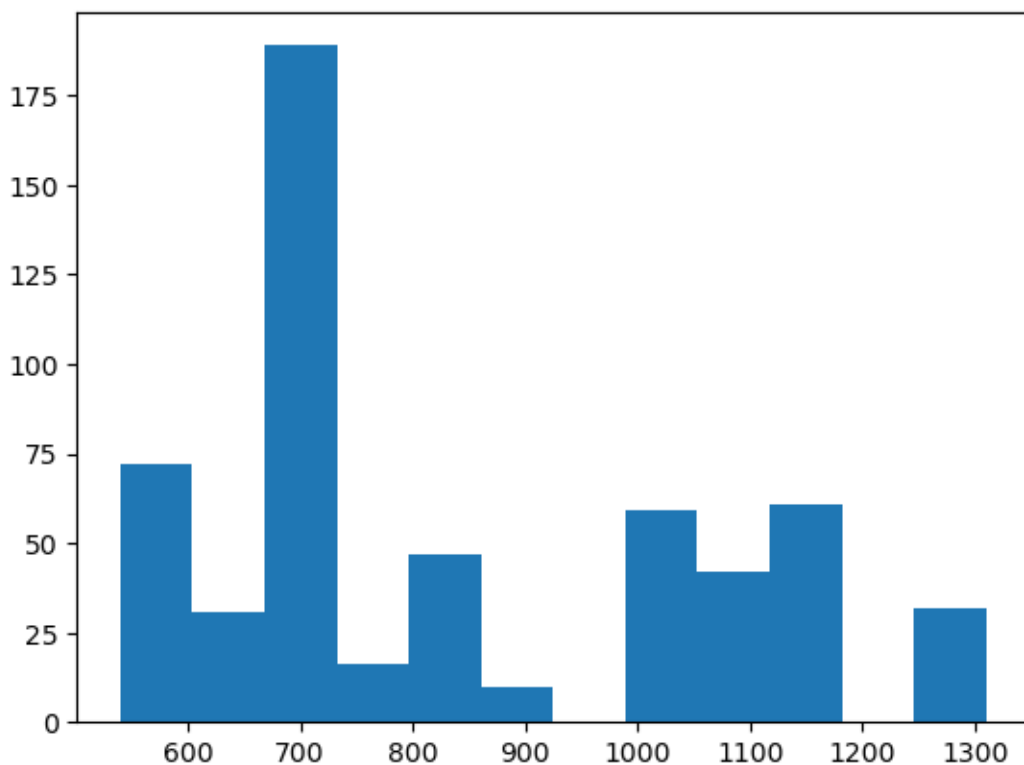
```
[126]: filter_data = master_df.loc[master_df.duration == 'long', ['Time', 'duration']]
```

```
[127]: filter_data.Time.describe()
```

```
[127]: count      559.000000
mean       840.635063
std        221.020092
min        540.000000
25%        670.000000
50%        717.000000
75%       1045.000000
max       1310.000000
Name: Time, dtype: float64
```

```
[128]: plt.hist(filter_data.Time, bins = 12)
plt.show()
```





```
[129]: # majority of long distant flights depart in the morning between 5 and 9.
```

**0.3 4. How many flights were delayed at large hubs compared to medium hubs?**  
Use appropriate visualization to represent your findings.

```
[130]: master_df.head()
```

```
[130]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	C0	269	SFO	IAH	3	15	205	1	
1	6	C0	1094	LAX	IAH	3	30	181	1	
2	11	C0	223	ANC	SEA	3	49	201	1	
3	18	C0	1496	LAS	IAH	3	60	162	0	
4	20	C0	507	ONT	IAH	3	75	167	0	

	type_source_airport	...	runway_count_source_airport	type_dest_airport	\
0	large_airport	...	4.0	large_airport	
1	large_airport	...	4.0	large_airport	
2	large_airport	...	3.0	large_airport	
3	large_airport	...	4.0	large_airport	
4	large_airport	...	2.0	large_airport	

	elevation_ft_dest_airport	runway_count_dest_airport	\
0	97.0	5.0	
1	97.0	5.0	
2	433.0	4.0	
3	97.0	5.0	
4	97.0	5.0	

	data_2021_source_airport	data_2021_dest_airport	Founded	Airlines	ID	\
0	11725347.0	16242821.0	1934.0		C0	
1	23663410.0	16242821.0	1934.0		C0	
2	2184959.0	17430195.0	1934.0		C0	
3	19160342.0	16242821.0	1934.0		C0	
4	2201528.0	16242821.0	1934.0		C0	

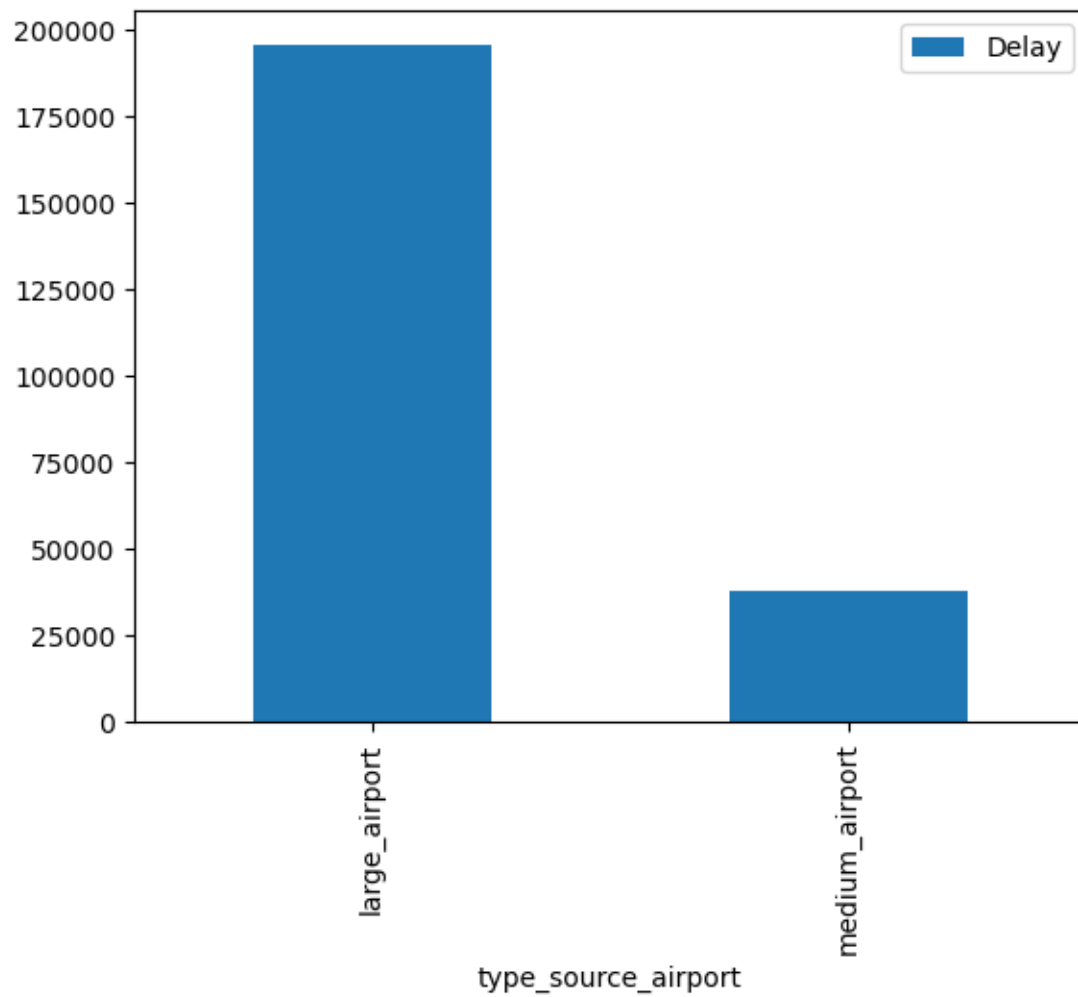
  

	Description	duration
0	United Airlines (initially C0)	short
1	United Airlines (initially C0)	short
2	United Airlines (initially C0)	short
3	United Airlines (initially C0)	short
4	United Airlines (initially C0)	short

[5 rows x 21 columns]

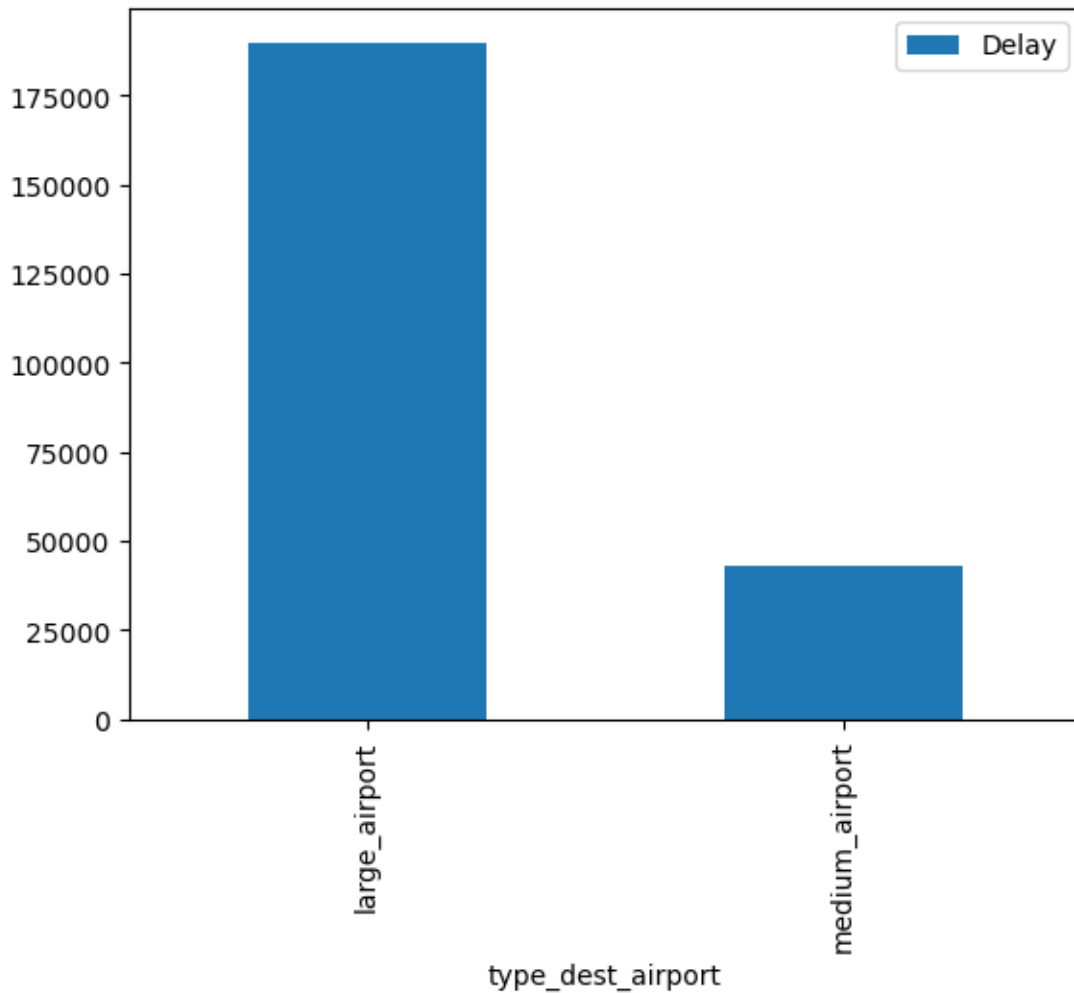
```
[131]: master_df.groupby('type_source_airport')[['Delay']].agg('sum').plot.bar()
```

```
[131]: <AxesSubplot:xlabel='type_source_airport'>
```



```
[132]: master_df.groupby('type_dest_airport')[['Delay']].agg('sum').plot.bar()
```

```
[132]: <AxesSubplot:xlabel='type_dest_airport'>
```



#### 0.4 5. Use hypothesis testing strategies to discover:

##### 0.4.1 (a) If the airport's altitude has anything to do with flight delays for incoming and departing flights

For outgoing flights

```
[133]: master_df.columns
```

```
[133]: Index(['id', 'Airline', 'Flight', 'AirportFrom', 'AirportTo', 'DayOfWeek',
        'Time', 'Length', 'Delay', 'type_source_airport',
        'elevation_ft_source_airport', 'runway_count_source_airport',
        'type_dest_airport', 'elevation_ft_dest_airport',
        'runway_count_dest_airport', 'data_2021_source_airport',
        'data_2021_dest_airport', 'Founded', 'Airlines ID', 'Description',
        'duration'],
```

```
dtype='object')
```

```
[134]: Ho = "Avg elevation for Delayed flights and not Delayed flights not_
      ↪significantly different"
      Ha = "Avg elevation for Delayed flights and not Delayed flights are different"
```

```
[135]: x = np.array(master_df[master_df['Delay']==1].elevation_ft_source_airport)
      x.shape
```

```
[135]: (233117,)
```

```
[136]: x
```

```
[136]: array([ 13., 125., 152., ..., 2181., 13., 125.])
```

```
[137]: y=np.array(master_df[master_df['Delay']==0].elevation_ft_source_airport)
      y
```

```
[137]: array([2181., 944., 107., ..., 1026., 17., 96.])
```

```
[138]: len(y)
```

```
[138]: 283100
```

```
[139]: t,p_value = ttest_ind(x,y,axis=0)
```

```
[140]: p_value
```

```
[140]: 1.909387572702128e-19
```

```
[141]: if p_value<0.05:
      print(f'{Ha} as the p_value({p_value})<0.05')
      else:
      print(f'{Ho} as the p_value({p_value})>0.05')
```

Avg elevation for Delayed flights and not Delayed flights are different as the p\_value(1.909387572702128e-19)<0.05

#### For incoming flights

```
[143]: Ho = "Avg elevation for Delayed flights and not Delayed flights not_
      ↪significantly different"
      Ha = "Avg elevation for Delayed flights and not Delayed flights are different"
```

```
[144]: x = np.array(master_df[master_df['Delay']==1].elevation_ft_dest_airport)
      x.shape
```

```
[144]: (233117,)
```

```
[145]: y=np.array(master_df[master_df['Delay']==0].elevation_ft_dest_airport)
y
```

```
[145]: array([ 97.,  97.,  97., ..., 620., 1135.,   9.])
```

```
[146]: t,p_value = ttest_ind(x,y,axis=0)
```

```
[147]: p_value
```

```
[147]: 2.7948729392515513e-21
```

```
[148]: if p_value<0.05:
        print(f'{Ha} as the p_value({p_value})<0.05')
    else:
        print(f'{Ho} as the p_value({p_value})>0.05')
```

Avg elevation for Delayed flights and not Delayed flights are different as the  $p\_value(2.7948729392515513e-21) < 0.05$

#### 0.4.2 (b) If the number of runways at an airport affects flight delays.

##### For Outgoing Flights

```
[149]: Ho = "Number of runways at an airport does not affects flight delays"
Ha = "Number of runways at an airport definetly affects flight delays"
```

```
[150]: x = np.array(master_df[master_df['Delay']==1].runway_count_source_airport)
x.shape
```

```
[150]: (233117,)
```

```
[151]: x
```

```
[151]: array([4., 4., 3., ..., 4., 4., 4.])
```

```
[152]: y=np.array(master_df[master_df['Delay']==0].runway_count_source_airport)
y
```

```
[152]: array([4., 2., 2., ..., 3., 1., 4.])
```

```
[153]: len(y)
```

```
[153]: 283100
```

```
[154]: t,p_value = ttest_ind(x,y,axis=0)
```

```
[155]: p_value
```

```
[155]: 4.208877752611582e-97
```

```
[156]: if p_value<0.05:  
        print(f'{Ha} as the p_value({p_value})<0.05')  
    else:  
        print(f'{Ho} as the p_value({p_value})>0.05')
```

Number of runways at an airport definitely affects flight delays as the  
`p_value(4.208877752611582e-97)<0.05`

### For Incoming Flights

```
[157]: Ho = "Number of runways at an airport does not affects flight delays"  
      Ha = "Number of runways at an airport definitely affects flight delays"
```

```
[158]: x = np.array(master_df[master_df['Delay']==1].runway_count_dest_airport)  
      x.shape
```

```
[158]: (233117,)
```

```
[159]: x
```

```
[159]: array([5., 5., 4., ..., 3., 4., 4.])
```

```
[160]: y=np.array(master_df[master_df['Delay']==0].runway_count_dest_airport)  
      y
```

```
[160]: array([5., 5., 5., ..., 5., 3., 3.])
```

```
[161]: len(y)
```

```
[161]: 283100
```

```
[162]: t,p_value = ttest_ind(x,y,axis=0)
```

```
[163]: p_value
```

```
[163]: 0.0
```

```
[164]: if p_value<0.05:  
        print(f'{Ha} as the p_value({p_value})<0.05')  
    else:  
        print(f'{Ho} as the p_value({p_value})>0.05')
```

Number of runways at an airport definitely affects flight delays as the  
`p_value(0.0)<0.05`

### 0.4.3 (c) If the duration of a flight (length) affects flight delays

```
[165]: Ho = "Duration of a flight (length) does not affects flight delays"
Ha = "Duration of a flight (length) affects flight delays"
```

```
[166]: x = np.array(master_df[master_df['Delay']==1].Length)
x.shape
```

```
[166]: (233117,)
```

```
[167]: x
```

```
[167]: array([205, 181, 201, ..., 75, 75, 75], dtype=int64)
```

```
[168]: y=np.array(master_df[master_df['Delay']==0].Length)
y
```

```
[168]: array([162, 167, 64, ..., 85, 70, 55], dtype=int64)
```

```
[169]: len(y)
```

```
[169]: 283100
```

```
[170]: if p_value<0.05:
        print(f'{Ha} as the p_value({p_value})<0.05')
    else:
        print(f'{Ho} as the p_value({p_value})>0.05')
```

Duration of a flight (length) affects flight delays as the p\_value(0.0)<0.05

### 0.5 6. Find the correlation matrix between the flight delay predictors, create a heatmap to visualize this, and share your findings

```
[171]: master_df.head()
```

```
[171]:
```

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	Delay	\
0	1	CO	269	SFO	IAH	3	15	205	1	
1	6	CO	1094	LAX	IAH	3	30	181	1	
2	11	CO	223	ANC	SEA	3	49	201	1	
3	18	CO	1496	LAS	IAH	3	60	162	0	
4	20	CO	507	ONT	IAH	3	75	167	0	

	type_source_airport	...	runway_count_source_airport	type_dest_airport	\
0	large_airport	...	4.0	large_airport	
1	large_airport	...	4.0	large_airport	
2	large_airport	...	3.0	large_airport	



3	large_airport	...	4.0	large_airport
4	large_airport	...	2.0	large_airport

	elevation_ft_dest_airport	runway_count_dest_airport	\
0	97.0	5.0	
1	97.0	5.0	
2	433.0	4.0	
3	97.0	5.0	
4	97.0	5.0	

	data_2021_source_airport	data_2021_dest_airport	Founded	Airlines ID	\
0	11725347.0	16242821.0	1934.0	CO	
1	23663410.0	16242821.0	1934.0	CO	
2	2184959.0	17430195.0	1934.0	CO	
3	19160342.0	16242821.0	1934.0	CO	
4	2201528.0	16242821.0	1934.0	CO	

	Description	duration
0	United Airlines (initially CO)	short
1	United Airlines (initially CO)	short
2	United Airlines (initially CO)	short
3	United Airlines (initially CO)	short
4	United Airlines (initially CO)	short

[5 rows x 21 columns]

```
[172]: corr_matrix = master_df.corr()
corr_matrix
```

```
[172]:
```

	id	Flight	DayOfWeek	Time	\
id	1.000000	-0.006202	-0.055079	0.028489	
Flight	-0.006202	1.000000	0.001468	-0.001958	
DayOfWeek	-0.055079	0.001468	1.000000	0.001092	
Time	0.028489	-0.001958	0.001092	1.000000	
Length	0.020984	-0.346860	0.013248	-0.020629	
Delay	0.140434	-0.057371	-0.025832	0.149801	
elevation_ft_source_airport	0.003951	0.141999	0.001177	0.013952	
runway_count_source_airport	-0.003162	0.012212	0.000033	0.091551	
elevation_ft_dest_airport	0.003102	0.144422	0.000999	-0.010090	
runway_count_dest_airport	-0.009681	0.013726	-0.000335	-0.115846	
data_2021_source_airport	0.001409	-0.057653	0.004348	0.124963	
data_2021_dest_airport	-0.006601	-0.055555	0.004086	-0.128112	
Founded	0.009168	0.434772	-0.005205	0.004272	

	Length	Delay	elevation_ft_source_airport	\
id	0.020984	0.140434	0.003951	
Flight	-0.346860	-0.057371	0.141999	

DayOfWeek	0.013248	-0.025832	0.001177
Time	-0.020629	0.149801	0.013952
Length	1.000000	0.040162	-0.086144
Delay	0.040162	1.000000	0.012551
elevation_ft_source_airport	-0.086144	0.012551	1.000000
runway_count_source_airport	0.050683	0.029099	0.095883
elevation_ft_dest_airport	-0.050448	0.013180	0.231182
runway_count_dest_airport	0.090177	-0.061431	-0.017638
data_2021_source_airport	0.099096	0.020315	0.162831
data_2021_dest_airport	0.158075	-0.051622	-0.022168
Founded	-0.332935	-0.002103	0.020167

	runway_count_source_airport \
id	-0.003162
Flight	0.012212
DayOfWeek	0.000033
Time	0.091551
Length	0.050683
Delay	0.029099
elevation_ft_source_airport	0.095883
runway_count_source_airport	1.000000
elevation_ft_dest_airport	-0.018412
runway_count_dest_airport	-0.212869
data_2021_source_airport	0.615706
data_2021_dest_airport	-0.242414
Founded	-0.101429

	elevation_ft_dest_airport \
id	0.003102
Flight	0.144422
DayOfWeek	0.000999
Time	-0.010090
Length	-0.050448
Delay	0.013180
elevation_ft_source_airport	0.231182
runway_count_source_airport	-0.018412
elevation_ft_dest_airport	1.000000
runway_count_dest_airport	0.095766
data_2021_source_airport	-0.024008
data_2021_dest_airport	0.162782
Founded	0.019975

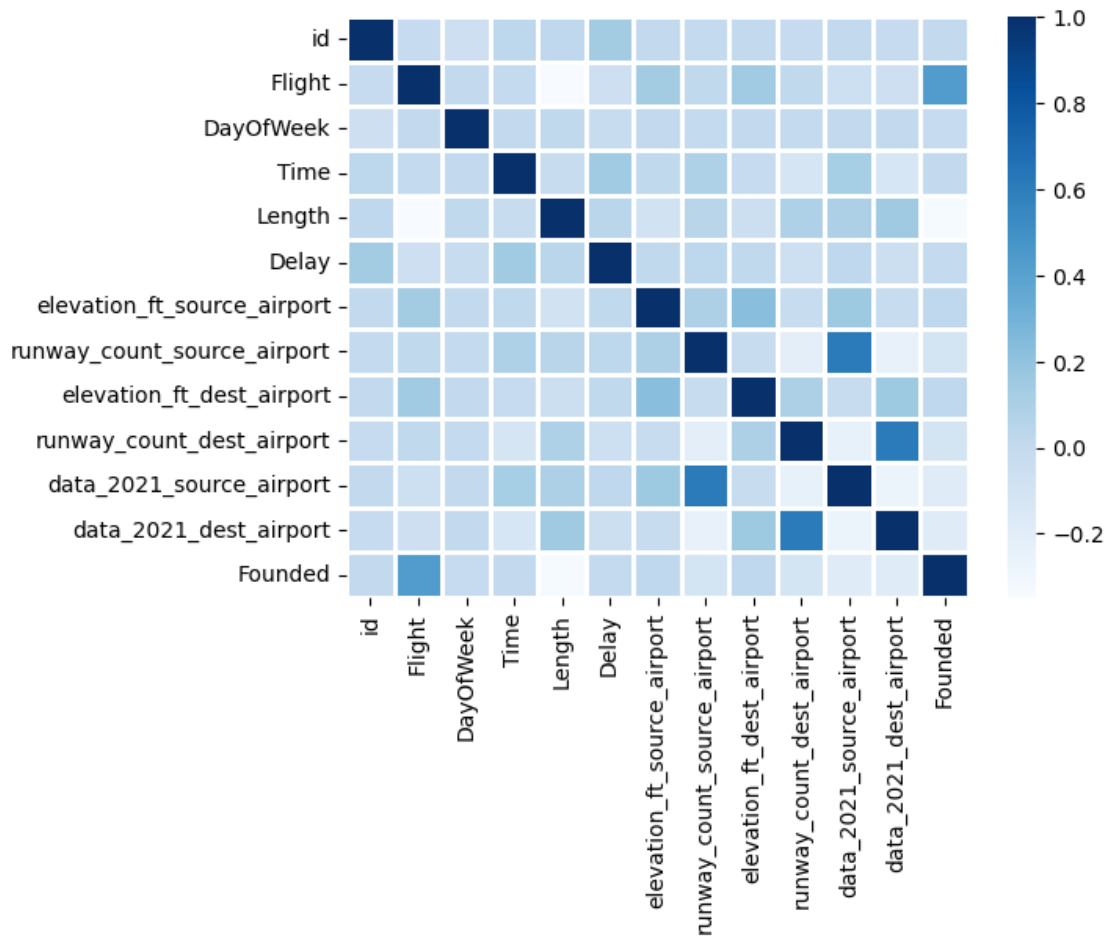
	runway_count_dest_airport \
id	-0.009681
Flight	0.013726
DayOfWeek	-0.000335
Time	-0.115846

Length	0.090177
Delay	-0.061431
elevation_ft_source_airport	-0.017638
runway_count_source_airport	-0.212869
elevation_ft_dest_airport	0.095766
runway_count_dest_airport	1.000000
data_2021_source_airport	-0.240883
data_2021_dest_airport	0.615722
Founded	-0.100473

	data_2021_source_airport	data_2021_dest_airport \
id	0.001409	-0.006601
Flight	-0.057653	-0.055555
DayOfWeek	0.004348	0.004086
Time	0.124963	-0.128112
Length	0.099096	0.158075
Delay	0.020315	-0.051622
elevation_ft_source_airport	0.162831	-0.022168
runway_count_source_airport	0.615706	-0.242414
elevation_ft_dest_airport	-0.024008	0.162782
runway_count_dest_airport	-0.240883	0.615722
data_2021_source_airport	1.000000	-0.269936
data_2021_dest_airport	-0.269936	1.000000
Founded	-0.180238	-0.179615

	Founded
id	0.009168
Flight	0.434772
DayOfWeek	-0.005205
Time	0.004272
Length	-0.332935
Delay	-0.002103
elevation_ft_source_airport	0.020167
runway_count_source_airport	-0.101429
elevation_ft_dest_airport	0.019975
runway_count_dest_airport	-0.100473
data_2021_source_airport	-0.180238
data_2021_dest_airport	-0.179615
Founded	1.000000

```
[173]: sns.heatmap(corr_matrix, cmap='Blues',linecolor='white', linewidths=2)
plt.show()
```



## 1 Project Task: Machine Learning

### 1.1 1. Use OneHotEncoder and OrdinalEncoder to deal with categorical variables.

```
[175]: master_df.isna().sum()
```

```
[175]: id          0
       Airline     0
       Flight      0
       AirportFrom  0
       AirportTo    0
       DayOfWeek    0
       Time         0
       Length       0
       Delay        0
```

```

type_source_airport      0
elevation_ft_source_airport  0
runway_count_source_airport  0
type_dest_airport        0
elevation_ft_dest_airport  0
runway_count_dest_airport  0
data_2021_source_airport   0
data_2021_dest_airport     0
Founded                   0
Airlines ID               0
Description                0
duration                  0
dtype: int64

```

```

[174]: master_df.to_csv('Master_Dataset.csv')
       master_df.to_excel('Master_Dataset.xlsx')

```

```

[177]: master_df.drop(columns = ['id', 'Flight', 'duration', 'Airline', 'AirportFrom', 'AirportTo'], inplace = True)
       master_df.head(3)

```

```

[177]:
   DayOfWeek  Time  Length  Delay type_source_airport \
0          3    15    205     1      large_airport
1          3    30    181     1      large_airport
2          3    49    201     1      large_airport

   elevation_ft_source_airport  runway_count_source_airport  type_dest_airport \
0                        13.0                             4.0      large_airport
1                       125.0                             4.0      large_airport
2                       152.0                             3.0      large_airport

   elevation_ft_dest_airport  runway_count_dest_airport \
0                        97.0                             5.0
1                        97.0                             5.0
2                       433.0                             4.0

   data_2021_source_airport  data_2021_dest_airport  Founded Airlines ID \
0          11725347.0          16242821.0    1934.0         CO
1          23663410.0          16242821.0    1934.0         CO
2          2184959.0          17430195.0    1934.0         CO

   Description
0  United Airlines (initially CO)
1  United Airlines (initially CO)
2  United Airlines (initially CO)

```

```
[178]: master_df.drop(columns = ['Airlines ID', 'Description'], inplace = True)
master_df.head(3)
```

```
[178]:   DayOfWeek  Time  Length  Delay type_source_airport \
0         3    15    205     1      large_airport
1         3    30    181     1      large_airport
2         3    49    201     1      large_airport

      elevation_ft_source_airport  runway_count_source_airport type_dest_airport \
0                          13.0                        4.0      large_airport
1                         125.0                        4.0      large_airport
2                         152.0                        3.0      large_airport

      elevation_ft_dest_airport  runway_count_dest_airport \
0                          97.0                        5.0
1                          97.0                        5.0
2                         433.0                        4.0

      data_2021_source_airport  data_2021_dest_airport  Founded
0                11725347.0                16242821.0   1934.0
1                23663410.0                16242821.0   1934.0
2                2184959.0                 17430195.0   1934.0
```

```
[232]: master_df.type_dest_airport.unique()
```

```
[232]: array(['large_airport', 'medium_airport'], dtype=object)
```

```
[233]: master_df.type_source_airport.unique()
```

```
[233]: array(['large_airport', 'medium_airport'], dtype=object)
```

```
[237]: ordinal = OrdinalEncoder(categories=[['medium_airport', 'large_airport'],
      → ['large_airport'], ['medium_airport', 'large_airport']])
ordinal.fit(master_df[['type_source_airport', 'type_dest_airport']])
```

```
[237]: OrdinalEncoder(categories=[['medium_airport', 'large_airport'],
      → ['medium_airport', 'large_airport']])
```

```
[238]: master_df[['type_source_airport', 'type_dest_airport']] = ordinal.
      → transform(master_df[['type_source_airport', 'type_dest_airport']])
```

```
[239]: master_df.head()
```

```
[239]:   DayOfWeek  Time  Length  Delay  type_source_airport \
0         3    15    205     1                1.0
1         3    30    181     1                1.0
2         3    49    201     1                1.0
```

3	3	60	162	0	1.0
4	3	75	167	0	1.0

	elevation_ft_source_airport	runway_count_source_airport	\
0	13.0	4.0	
1	125.0	4.0	
2	152.0	3.0	
3	2181.0	4.0	
4	944.0	2.0	

	type_dest_airport	elevation_ft_dest_airport	runway_count_dest_airport	\
0	1.0	97.0	5.0	
1	1.0	97.0	5.0	
2	1.0	433.0	4.0	
3	1.0	97.0	5.0	
4	1.0	97.0	5.0	

	data_2021_source_airport	data_2021_dest_airport	Founded
0	11725347.0	16242821.0	1934.0
1	23663410.0	16242821.0	1934.0
2	2184959.0	17430195.0	1934.0
3	19160342.0	16242821.0	1934.0
4	2201528.0	16242821.0	1934.0

```
[240]: master_df['years_of_operation'] = 2023 - master_df['Founded']
```

```
[242]: master_df.drop(['Founded'],axis=1,inplace=True)
```

```
[243]: master_df.head()
```

```
[243]:
```

	DayOfWeek	Time	Length	Delay	type_source_airport	\
0	3	15	205	1	1.0	
1	3	30	181	1	1.0	
2	3	49	201	1	1.0	
3	3	60	162	0	1.0	
4	3	75	167	0	1.0	

	elevation_ft_source_airport	runway_count_source_airport	\
0	13.0	4.0	
1	125.0	4.0	
2	152.0	3.0	
3	2181.0	4.0	
4	944.0	2.0	

	type_dest_airport	elevation_ft_dest_airport	runway_count_dest_airport	\
0	1.0	97.0	5.0	
1	1.0	97.0	5.0	

2	1.0	433.0	4.0
3	1.0	97.0	5.0
4	1.0	97.0	5.0

	data_2021_source_airport	data_2021_dest_airport	years_of_operation
0	11725347.0	16242821.0	89.0
1	23663410.0	16242821.0	89.0
2	2184959.0	17430195.0	89.0
3	19160342.0	16242821.0	89.0
4	2201528.0	16242821.0	89.0

```
[244]: master_df.shape
```

```
[244]: (516217, 13)
```

## 1.2 2. Perform the following model building steps:

### 1.2.1 (a) Apply logistic regression (use stochastic gradient descent optimizer) and decision tree models

### 1.2.2 (b) Use the stratified five-fold method to build and validate the models

Note: Make sure you use standardization effectively, ensuring no data leakage and leverage pipelines to have a cleaner code ### (c) Use RandomizedSearchCV for hyperparameter tuning, and use k-fold for cross-validation ### (d) Keep a few data points (10%) for prediction purposes to evaluate how you would make the final prediction, and do not use this data for testing or validation Note: The final prediction will be based on the voting (majority class by 5 models created using the stratified 5-fold method) ### (g) Compare the results of logistic regression and decision tree classifier

```
[245]: model = master_df.copy()
model.head()
```

```
[245]:   DayOfWeek  Time  Length  Delay  type_source_airport  \
0          3    15    205     1          1.0
1          3    30    181     1          1.0
2          3    49    201     1          1.0
3          3    60    162     0          1.0
4          3    75    167     0          1.0
```

	elevation_ft_source_airport	runway_count_source_airport	\
0	13.0	4.0	
1	125.0	4.0	
2	152.0	3.0	
3	2181.0	4.0	
4	944.0	2.0	



	type_dest_airport	elevation_ft_dest_airport	runway_count_dest_airport	\
0	1.0	97.0	5.0	
1	1.0	97.0	5.0	
2	1.0	433.0	4.0	
3	1.0	97.0	5.0	
4	1.0	97.0	5.0	

	data_2021_source_airport	data_2021_dest_airport	years_of_operation
0	11725347.0	16242821.0	89.0
1	23663410.0	16242821.0	89.0
2	2184959.0	17430195.0	89.0
3	19160342.0	16242821.0	89.0
4	2201528.0	16242821.0	89.0

```
[250]: X = model.drop(['Delay'],axis=1)
       y = model.Delay
```

```
[251]: print(X.shape)
       print(y.shape)
```

```
(516217, 12)
(516217,)
```

```
[252]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
       ↪random_state=42)
       print(X_train.shape)
       print(X_test.shape)
       print(y_train.shape)
       print(y_test.shape)
```

```
(412973, 12)
(103244, 12)
(412973,)
(103244,)
```

```
[253]: # We will use X_train in K-fold Cross validation and X_test during final
       ↪prediction.
```

```
[254]: folds = StratifiedKFold(n_splits=5, shuffle = True, random_state=12)
       accuracy_train = {}
       accuracy_test = {}
       final_predictions_sgd = {}
       i = 1
       for train_index, test_index in folds.split(X_train,y_train):
           print('iter ', i)
           train,test = model.loc[train_index,], model.loc[test_index,]
```

```

sc = StandardScaler()
sgd = SGDClassifier()

# define search space

space = dict()
space['sgd__penalty'] = ['l1', 'l2', 'elasticnet']
space['sgd__l1_ratio'] = [0,.1,.2,.8,1]
space['sgd__alpha'] = [1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100, 1000,10000]
space['sgd__learning_rate'] = ['constant', 'adaptive']
space['sgd__eta0']=[1e-5, 1e-4, 1e-3, 1e-2, 1e-1 , 2e-1, 3e-1, 5e-1, 8e-1,
↪4e-1, 8e-1, 1, 10, 100]

pipe = Pipeline([('sc',sc), ('sgd', sgd)])

# define search
search = RandomizedSearchCV( pipe, space, scoring='accuracy',
                             cv=5, refit=True, return_train_score = True,
                             random_state = 12, n_jobs = -1, n_iter = 2
                             )

# execute search
fold_X_train = train.drop(columns = 'Delay')
fold_y_train = train.Delay

result = search.fit(fold_X_train, fold_y_train)

train_pred = result.predict(fold_X_train)

fold_X_test = test.drop(columns = 'Delay')
fold_y_test = test.Delay
test_pred = result.predict(fold_X_test)
final_predictions_sgd.update({'Fold{}'.format(i):result.predict(X_test)})

# get rmse for each fold for train data
accuracy_train.update({'Fold{}'.format(i): round(accuracy_score(y_true =
↪fold_y_train, y_pred = train_pred)*100,3)})
accuracy_test.update({'Fold{}'.format(i): round(accuracy_score(y_true =
↪fold_y_test, y_pred = test_pred) * 100,3)})
i += 1

```

```

iter  1
iter  2
iter  3
iter  4
iter  5

```

```

[256]: folds = StratifiedKFold(n_splits=5, shuffle = True, random_state=12)
dt_accuracy_train = {}
dt_accuracy_test = {}
final_predictions_dt = {}
i = 1
for train_index, test_index in folds.split(X_train,y_train):
    print('iter ', i)

    train, test = model.loc[train_index,], model.loc[test_index,]

    sc = StandardScaler()
    dt = DecisionTreeClassifier()

    # define search space
    space = dict()
    space['dt__min_samples_split'] = [25000, 30000, 35000, 40000, 45000, 50000,
↪60000 ]
    space['dt__min_samples_leaf'] = [10000, 15000, 20000]

    pipe = Pipeline([('sc',sc), ('dt', dt)])

    # define search
    search = RandomizedSearchCV( pipe, space, scoring='accuracy',
                                cv=5, refit=True, return_train_score = True,
                                random_state = 12, n_jobs = -1, n_iter = 2
                                )

    # execute search
    fold_X_train = train.drop(columns = 'Delay')
    fold_y_train = train.Delay

    result = search.fit(fold_X_train, fold_y_train)

    train_pred = result.predict(fold_X_train)

    fold_X_test = test.drop(columns = 'Delay')
    fold_y_test = test.Delay
    test_pred = result.predict(fold_X_test)
    final_predictions_dt.update({'Fold{}'.format(i):result.predict(X_test)})

    # get rmse for each fold for train data
    dt_accuracy_train.update({'Fold{}'.format(i): round(accuracy_score(y_true =
↪fold_y_train, y_pred = train_pred)*100,3)})
    dt_accuracy_test.update({'Fold{}'.format(i): round(accuracy_score(y_true =
↪fold_y_test, y_pred = test_pred) * 100,3)})
    i += 1

```

```
iter 1
iter 2
iter 3
iter 4
iter 5
```

```
[257]: # compare results :
train_results = pd.DataFrame ({'sgd' : accuracy_train.values(), 'dt':
    ↳dt_accuracy_train.values() },
                                index = ['Fold {}'.format(i) for i in range(1,6)])
train_results
```

```
[257]:
```

	sgd	dt
Fold 1	60.380	61.110
Fold 2	60.433	61.275
Fold 3	60.432	61.160
Fold 4	60.455	61.276
Fold 5	60.477	61.256

```
[258]: test_results = pd.DataFrame ({'sgd' : accuracy_test.values(), 'dt':
    ↳dt_accuracy_test.values() },
                                index = ['Fold {}'.format(i) for i in range(1,6)])
test_results
```

```
[258]:
```

	sgd	dt
Fold 1	60.659	61.298
Fold 2	60.444	61.165
Fold 3	60.450	61.015
Fold 4	60.355	61.234
Fold 5	60.268	61.242

```
[259]: final_predictions_dt
```

```
[259]: {'Fold1': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold2': array([1, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold3': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold4': array([1, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold5': array([1, 0, 0, ..., 0, 0, 0], dtype=int64)}
```

```
[260]: final_predictions_sgd
```

```
[260]: {'Fold1': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold2': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold3': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold4': array([0, 0, 0, ..., 0, 0, 0], dtype=int64),
'Fold5': array([0, 0, 0, ..., 0, 0, 0], dtype=int64)}
```

### 1.3 3. Use the stratified five-fold method to build and validate the models using the XGB classifier, compare all methods, and share your findings

```
[261]: folds = StratifiedKFold(n_splits=5, shuffle = True, random_state=12)
xgb_accuracy_train = {}
xgb_accuracy_test = {}
final_predictions_xgb = []

i = 1
for train_index, test_index in folds.split(X_train,y_train):
    print('iter ', i)
    train, test = model.loc[train_index,], model.loc[test_index,]
    sc = StandardScaler()
    xgb_r = XGBClassifier(random_state = 12, use_label_encoder = False)

    # define search space
    space = dict()
    space['xgb_r__n_estimators'] = [40,50,60]
    space['xgb_r__max_depth'] = [3,4,5]
    space['xgb_r__colsample_bytree']: [0.4,.5,.6]
    space['xgb_r__lambda'] = [.0001,.002,.0004,.0003]
    space['xgb_r__alpha'] = [.01,.02,.1,.4]

    pipe = Pipeline([('sc',sc), ('xgb_r', xgb_r)])

    # define search
    search = RandomizedSearchCV( pipe, space,
    →scoring='neg_root_mean_squared_error',
                                cv=5, refit=True, return_train_score = True,
                                random_state = 12, n_jobs = -1, n_iter = 2
                                )

    # execute search
    fold_X_train = train.drop(columns = 'Delay')
    fold_y_train = train.Delay

    result = search.fit(fold_X_train, fold_y_train)

    train_pred = result.predict(fold_X_train)

    fold_X_test = test.drop(columns = 'Delay')
    fold_y_test = test.Delay
    test_pred = result.predict(fold_X_test)

    final_predictions_xgb.append(result.predict(X_test))

    # get rmse for each fold for train data
```

```

xgb_accuracy_train.update({'Fold{}'.format(i): round(accuracy_score(y_true_
↪= fold_y_train, y_pred = train_pred),3)})
xgb_accuracy_test.update({'Fold{}'.format(i): round(accuracy_score(y_true =
↪fold_y_test, y_pred = test_pred),3)})
i += 1

```

```

iter 1
[19:49:31] WARNING: C:\Windows\Temp\abs_557yfx6311\croots\recipe\xgboost-
split_1659548953302\work\src\learner.cc:1115: Starting in XGBoost 1.3.0, the
default evaluation metric used with the objective 'binary:logistic' was changed
from 'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore
the old behavior.
iter 2
[19:50:26] WARNING: C:\Windows\Temp\abs_557yfx6311\croots\recipe\xgboost-
split_1659548953302\work\src\learner.cc:1115: Starting in XGBoost 1.3.0, the
default evaluation metric used with the objective 'binary:logistic' was changed
from 'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore
the old behavior.
iter 3
[19:51:26] WARNING: C:\Windows\Temp\abs_557yfx6311\croots\recipe\xgboost-
split_1659548953302\work\src\learner.cc:1115: Starting in XGBoost 1.3.0, the
default evaluation metric used with the objective 'binary:logistic' was changed
from 'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore
the old behavior.
iter 4
[19:52:35] WARNING: C:\Windows\Temp\abs_557yfx6311\croots\recipe\xgboost-
split_1659548953302\work\src\learner.cc:1115: Starting in XGBoost 1.3.0, the
default evaluation metric used with the objective 'binary:logistic' was changed
from 'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore
the old behavior.
iter 5
[19:53:43] WARNING: C:\Windows\Temp\abs_557yfx6311\croots\recipe\xgboost-
split_1659548953302\work\src\learner.cc:1115: Starting in XGBoost 1.3.0, the
default evaluation metric used with the objective 'binary:logistic' was changed
from 'error' to 'logloss'. Explicitly set eval_metric if you'd like to restore
the old behavior.

```

```

[266]: train_results['xgb'] = xgb_accuracy_train.values()
       test_results['xgb'] = xgb_accuracy_test.values()

```

```

[267]: train_results

```

```

[267]:
      sgd      dt      xgb
Fold 1  60.380  61.110  0.642
Fold 2  60.433  61.275  0.640
Fold 3  60.432  61.160  0.643
Fold 4  60.455  61.276  0.643

```

Fold 5 60.477 61.256 0.638

```
[269]: train_results['xgb']=train_results['xgb']*100
```

```
[270]: train_results
```

```
[270]:
```

	sgd	dt	xgb
Fold 1	60.380	61.110	64.2
Fold 2	60.433	61.275	64.0
Fold 3	60.432	61.160	64.3
Fold 4	60.455	61.276	64.3
Fold 5	60.477	61.256	63.8

```
[271]: test_results['xgb']=test_results['xgb']*100
```

```
[272]: test_results
```

```
[272]:
```

	sgd	dt	xgb
Fold 1	60.659	61.298	64.0
Fold 2	60.444	61.165	64.0
Fold 3	60.450	61.015	63.8
Fold 4	60.355	61.234	64.1
Fold 5	60.268	61.242	63.5