United States Airlines Analysis

January 30, 2023

[1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

1 Project Task: Week 1

1.1 Applied data science with Python

1.2 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. Hint: In this case, you would have to determine the keys to join the tables. A data description will be useful.

[2]:
```
airport = pd.read_excel('airports.xlsx')
airlines = pd.read_excel('Airlines.xlsx')
runways = pd.read_excel('runways.xlsx')
```

[3]:
```
airport.head()
```

[3]:

	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
```

[4]:
```
airport.shape
```

[4]:
```
(73805, 18)
```

[5]:
```
airlines.head()
```

[5]:
```
   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
```

[6]:
```
airlines.shape
```

[6]:
```
(518556, 9)
```

[7]:
```
runways.head()
```

[7]:
```
       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
```

[8]:
```
runways.shape
```

[8]:
```
(43977, 20)
```

Before merging the data lets drop the columns that will not play an important role in the model building

[10]:
```
runways.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 43977 entries, 0 to 43976
Data columns (total 20 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   id                        43977 non-null  int64
 1   airport_ref               43977 non-null  int64
 2   airport_ident             43977 non-null  object
 3   length_ft                 43753 non-null  float64
 4   width_ft                  41088 non-null  float64
 5   surface                   43520 non-null  object
 6   lighted                   43977 non-null  int64
 7   closed                    43977 non-null  int64
 8   le_ident                  43793 non-null  object
 9   le_latitude_deg           15016 non-null  float64
 10  le_longitude_deg          15000 non-null  float64
```

```
 11  le_elevation_ft            12781 non-null  float64
 12  le_heading_degT            14624 non-null  float64
 13  le_displaced_threshold_ft  2883 non-null   float64
 14  he_ident                   37332 non-null  object
 15  he_latitude_deg            14971 non-null  float64
 16  he_longitude_deg           14973 non-null  float64
 17  he_elevation_ft            12620 non-null  float64
 18  he_heading_degT            16428 non-null  float64
 19  he_displaced_threshold_ft  3176 non-null   float64
dtypes: float64(12), int64(4), object(4)
memory usage: 6.7+ MB
```

[11]:
```
runways.drop(['le_ident', 'le_latitude_deg','le_longitude_deg',␣
↪'le_elevation_ft', 'le_heading_degT',
      'le_displaced_threshold_ft', 'he_ident',␣
↪'he_latitude_deg','he_longitude_deg', 'he_elevation_ft', 'he_heading_degT',
      'he_displaced_threshold_ft'], axis = 1,inplace=True)
```

[12]:
```
runways
```

[12]:

	id	airport_ref	airport_ident	length_ft	width_ft	surface	lighted	closed
0	269408	6523	00A	80.0	80.0	ASPH-G	1	0
1	255155	6524	00AK	2500.0	70.0	GRVL	0	0
2	254165	6525	00AL	2300.0	200.0	TURF	0	0
3	270932	6526	00AR	40.0	40.0	GRASS	0	0
4	322128	322127	00AS	1450.0	60.0	Turf	0	0
…	…	…	…	…	…		…	…
43972	235186	27243	ZYTX	10499.0	148.0	CON	1	0
43973	235169	27244	ZYYJ	8530.0	148.0	CON	1	0
43974	354997	317861	ZYYK	8202.0	NaN	NaN	0	0
43975	346789	346788	ZZ-0003	1800.0	15.0	Turf	0	0
43976	313663	313629	ZZZZ	1713.0	82.0	concrete	0	0

[43977 rows x 8 columns]

Now lets remove the feature from the airpot data that is not usefull.

[14]:
```
airport.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 73805 entries, 0 to 73804
Data columns (total 18 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   id                73805 non-null  int64
 1   ident             73805 non-null  object
 2   type              73805 non-null  object
 3   name              73805 non-null  object
 4   latitude_deg      73805 non-null  float64
 5   longitude_deg     73805 non-null  float64
 6   elevation_ft      59683 non-null  float64
 7   continent         38086 non-null  object
 8   iso_country       73546 non-null  object
 9   iso_region        73805 non-null  object
 10  municipality      68739 non-null  object
 11  scheduled_service 73805 non-null  object
 12  gps_code          42996 non-null  object
 13  iata_code         9160 non-null   object
 14  local_code        32975 non-null  object
 15  home_link         3492 non-null   object
 16  wikipedia_link    10705 non-null  object
 17  keywords          13951 non-null  object
dtypes: float64(3), int64(1), object(14)
memory usage: 10.1+ MB
```

[15]:
```
airport.drop(['continent', 'iso_country', 'iso_region','municipality',␣
↪'gps_code','local_code', 'home_link',
           'wikipedia_link', 'keywords'], axis=1, inplace=True)
airport
```

[15]:

	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	
…	…	…	…	…	
73800	46378	ZZ-0001	heliport	Sealand Helipad	
73801	307326	ZZ-0002	small_airport	Glorioso Islands Airstrip	
73802	346788	ZZ-0003	small_airport	Fainting Goat Airport	
73803	342102	ZZZW	closed	Scandium City Heliport	
73804	313629	ZZZZ	small_airport	Satsuma I çjima Airport	

```
       latitude_deg  longitude_deg  elevation_ft scheduled_service iata_code
0         40.070801     -74.933601          11.0                 no       NaN
1         38.704022    -101.473911        3435.0                 no       NaN
2         59.947733    -151.692524         450.0                 no       NaN
3         34.864799     -86.770302         820.0                 no       NaN
4         35.608700     -91.254898         237.0                 no       NaN
…               …              …             …                  …        …
73800     51.894444       1.482500          40.0                 no       NaN
73801    -11.584278      47.296389          11.0                 no       NaN
73802     32.110587     -97.356312         690.0                 no       NaN
73803     69.355287    -138.939310           4.0                 no       ZYW
73804     30.784722     130.270556         338.0                 no       NaN

[73805 rows x 9 columns]
```

Now lets merge the runways and airport data.

[17]:
```
airport_runway = pd.merge(airport, runways, left_on = "ident", right_on =␣
↪"airport_ident")
```

[19]:
```
airport_runway.drop(['id_x', 'id_y'], axis=1, inplace=True)
```

[20]:
```
airport_runway
```

[20]:
```
         ident            type                                    name  \
0          00A        heliport                     Total Rf Heliport
1         00AK   small_airport                          Lowell Field
2         00AL   small_airport                          Epps Airpark
3         00AR          closed      Newport Hospital & Clinic Heliport
4         00AS   small_airport                        Fulton Airport
…          …              …                                       …
43972     ZYTX   large_airport  Shenyang Taoxian International Airport
43973     ZYYJ  medium_airport             Yanji Chaoyangchuan Airport
43974     ZYYK  medium_airport                  Yingkou Lanqi Airport
43975  ZZ-0003   small_airport                  Fainting Goat Airport
43976     ZZZZ   small_airport                Satsuma I çjima Airport

       latitude_deg  longitude_deg  elevation_ft scheduled_service iata_code  \
0         40.070801     -74.933601          11.0                 no       NaN
1         59.947733    -151.692524         450.0                 no       NaN
2         34.864799     -86.770302         820.0                 no       NaN
3         35.608700     -91.254898         237.0                 no       NaN
4         34.942803     -97.818019        1100.0                 no       NaN
…               …              …             …                  …        …
43972     41.639801     123.483002         198.0                yes       SHE
43973     42.882801     129.451004         624.0                yes       YNJ
43974     40.542524     122.358600           NaN                yes       YKH
```

```
43975     32.110587     -97.356312         690.0                no       NaN
43976     30.784722     130.270556         338.0                no       NaN

       airport_ref airport_ident  length_ft  width_ft   surface  lighted  \
0             6523           00A       80.0      80.0    ASPH-G        1
1             6524          00AK     2500.0      70.0      GRVL        0
2             6525          00AL     2300.0     200.0      TURF        0
3             6526          00AR       40.0      40.0     GRASS        0
4           322127          00AS     1450.0      60.0      Turf        0
…              …             …          …         …         …      …
43972        27243           ZYTX    10499.0     148.0       CON        1
43973        27244           ZYYJ     8530.0     148.0       CON        1
43974       317861           ZYYK     8202.0       NaN       NaN        0
43975       346788        ZZ-0003     1800.0      15.0      Turf        0
43976       313629           ZZZZ     1713.0      82.0  concrete        0

       closed
0           0
1           0
2           0
3           0
4           0
…         …
43972       0
43973       0
43974       0
43975       0
43976       0

[43977 rows x 15 columns]
```

Now lets merge the final column airline.

[22]:
```
final_df = pd.merge(airlines,airport_runway,how = "inner", left_on =␣
↪"AirportFrom", right_on = "iata_code" )
```

[23]:
```
final_df.drop_duplicates(subset=['id'], keep='first', inplace=True)
final_df
```

[23]:
```
              id Airline  Flight AirportFrom AirportTo  DayOfWeek  Time  \
0              1      CO     269         SFO       IAH          3    15
4              4      AA    2466         SFO       DFW          3    20
8              9      DL    2606         SFO       MSP          3    35
12           129      DL    1580         SFO       DTW          3   345
16           150      UA     756         SFO       DEN          3   348
…          …      …      …           …         …          …   …
2160266   451344      CO       2         GUM       HNL          1   400
```

```
2160268  469866      CO       2         GUM       HNL          2   400
2160270  488365      CO       2         GUM       HNL          3   400
2160272  506855      CO       2         GUM       HNL          4   400
2160274  525138      CO       2         GUM       HNL          5   400

         Length  Delay ident  … elevation_ft scheduled_service  iata_code  \
0           205      1  KSFO  …         13.0               yes        SFO
4           195      1  KSFO  …         13.0               yes        SFO
8           216      1  KSFO  …         13.0               yes        SFO
12          270      0  KSFO  …         13.0               yes        SFO
16          158      0  KSFO  …         13.0               yes        SFO
…          …    …    …  …           …                …         …
2160266     430      1  PGUM  …        298.0               yes        GUM
2160268     430      1  PGUM  …        298.0               yes        GUM
2160270     430      0  PGUM  …        298.0               yes        GUM
2160272     430      1  PGUM  …        298.0               yes        GUM
2160274     430      1  PGUM  …        298.0               yes        GUM

         airport_ref  airport_ident length_ft width_ft  surface lighted  \
0               3878           KSFO    7500.0    200.0      ASP       1
4               3878           KSFO    7500.0    200.0      ASP       1
8               3878           KSFO    7500.0    200.0      ASP       1
12              3878           KSFO    7500.0    200.0      ASP       1
16              3878           KSFO    7500.0    200.0      ASP       1
…              …             …       …       …      …     …
2160266         5433           PGUM   12015.0    150.0  asphalt       1
2160268         5433           PGUM   12015.0    150.0  asphalt       1
2160270         5433           PGUM   12015.0    150.0  asphalt       1
2160272         5433           PGUM   12015.0    150.0  asphalt       1
2160274         5433           PGUM   12015.0    150.0  asphalt       1

         closed
0             0
4             0
8             0
12            0
16            0
…         …
2160266       0
2160268       0
2160270       0
2160272       0
2160274       0

[518525 rows x 24 columns]
```

1.2.1 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.

Hint: Here, you should use web scraping to learn how long an airline has been operating.

```
# Now lets use the web scrapping to import the data frome the wikipedia.
```

[24]:
```
url = "https://en.wikipedia.org/wiki/List_of_airlines_of_the_United_States"
tables = pd.read_html(url)
```

[25]:
```
print(tables)
```

```
[                Airline  Image IATA ICAO         Callsign  \
0        Alaska Airlines    NaN   AS  ASA           ALASKA
1          Allegiant Air    NaN   G4  AAY        ALLEGIANT
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/TacomaAnchoragePortland (OR)San Franci…     1932
1   Las VegasCincinnatiFort Walton BeachIndianapol…     1997
2   Dallas/Fort WorthCharlotteChicago-O'HareLos An…     1926
3                             BurbankNew HavenOrlando     1987
4      CharlestonHartfordNew OrleansNorfolkProvoTampa     2018
5   AtlantaBostonDetroitLos AngelesMinneapolis/St…     1924
6                                  MiamiNew York-JFK     2010
7   DenverAtlantaChicago-O'HareCincinnatiCleveland…     1994
8                                    HonoluluKahului     1929
9   New York-JFKBostonLos AngelesFort LauderdaleOr…     1998
10  Dallas-LoveAtlantaBaltimoreChicago-MidwayDenve…     1967
11  Atlantic CityDetroitLas VegasFort LauderdaleCh…     1980
12     Minneapolis/St. PaulDallas/Fort WorthLas Vegas     1982
13  Chicago-O'HareDenverGuamHouston-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…  ,               Airline
Image IATA ICAO      Callsign  \
0       Air Wisconsin    NaN   ZW  AWI     WISCONSIN
1            Cape Air    NaN   9K  KAP          CAIR
2           CommutAir    NaN   C5  UCA     COMMUTAIR
3    Contour Airlines    NaN   LF  VTE     VOLUNTEER
4       Elite Airways    NaN   7Q  MNU        MAINER
5        Endeavor Air    NaN   9E  EDV      ENDEAVOR
6           Envoy Air    NaN   MQ  ENY         ENVOY
7      GoJet Airlines    NaN   G7  GJS     LINDBERGH
8         Horizon Air    NaN   QX  QXE       HORIZON
9       Mesa Airlines    NaN   YV  ASH   AIR SHUTTLE
10  Piedmont Airlines    NaN   PT  PDT      PIEDMONT
11       PSA Airlines    NaN   OH  JIA   BLUE STREAK
12   Republic Airways    NaN   YX  RPA     BRICKYARD
13     Silver Airways    NaN   3M  SIL  SILVER WINGS
14   SkyWest Airlines    NaN   OO  SKW       SKYWEST

                         Primary Hubs, Secondary Hubs  Founded  \
0   AppletonChicago-O'HareColumbiaMilwaukeeWashing…     1965
1   HyannisBillingsBostonNantucketSt. LouisSan Jua…     1988
2                     DenverNewarkWashington-Dulles      1989
3                                     Smyrna (TN)      1982
4            Melbourne/OrlandoNewarkPortland (Maine)     2006
5   Minneapolis/St. PaulAtlanta CincinnatiDetroitN…     1985
6              Dallas/Fort WorthChicago-O'Hare Miami     1984
7                              Chicago-O'HareDenver      2004
8                       Seattle/TacomaPortland (OR)      1981
9   As American Eagle:Phoenix-Sky HarborDallas/For…     1980
10          CharlottePhiladelphiaWashington-National     1961
11          CharlottePhiladelphiaWashington-National     1979
12  As American Eagle:IndianapolisColumbus (OH)Kan…     1998
13                       Fort LauderdaleOrlandoTampa     2011
14  As Delta Connection:AtlantaBoiseColorado Sprin…     1972
```

	Notes
0	Operates as United Express
1	NaN
2	Operates as United Express.
3	NaN
4	Commenced operations in 2014.
5	Founded as Express Airlines I. Operates as Del…
6	Founded as American Eagle Airlines. Operates a…
7	Commenced operations in 2005. Operates as Unit…
8	Operates as Alaska Airlines.
9	Founded as Mesa Air Shuttle. All but one aircr…
10	Founded as Henson Aviation and commenced opera…
11	Founded as Vee Neal Airlines. Operates as Amer…
12	Commenced operations in 2005. Operates as Amer…
13	NaN
14	Operates as Delta Connection, United Express, … ,

	Airline	Image	IATA	ICAO	Callsign	\
0	Advanced Air		NaN	AN	WSN	WINGSPAN
1	Air Sunshine		NaN	YI	RSI	AIR SUNSHINE
2	Bering Air		NaN	8E	BRG	BERING AIR
3	Boutique Air		NaN	4B	BTQ	BOUTIQUE
4	Everts Air		NaN	5V	VTS	EVERTS
5	Gem Air		NaN	NaN	NaN	NaN
6	Grand Canyon Airlines		NaN	YR	CVU	CANYON VIEW
7	Grand Canyon Scenic Airlines		NaN	YR	SCE	SCENIC
8	Grant Aviation		NaN	GV	GUN	HOOT
9	Griffing Flying Service		NaN	NaN	NaN	NaN
10	Island Airways		NaN	NaN	NaN	NaN
11	JSX		NaN	XE	JSX	BIGSTRIPE
12	Kenmore Air		NaN	M5	KEN	KENMORE
13	Key Lime Air		NaN	KG	LYM	KEY LIME
14	Mokulele Airlines		NaN	MW	MHO	MAHALO
15	New England Airlines		NaN	EJ	NEA	NEW ENGLAND
16	Penobscot Island Air		NaN	NaN	NaN	NaN
17	Reliant Air		NaN	NaN	RLI	RELIANT
18	San Juan Airlines		NaN	NaN	NaN	SKYFERRY
19	Servant Air		NaN	8D	NaN	NaN
20	Southern Airways Express		NaN	9X	FDY	FRIENDLY
21	Surf Air		NaN	NaN	UF	SURFAIR
22	Taquan Air		NaN	K3	TQN	TAQUAN
23	Tradewind Aviation		NaN	TJ	GPD	GOODSPEED
24	Ultimate Air Shuttle		NaN	UE	UJC	ULTIMATE
25	Utah Airways		NaN	NaN	NaN	NaN
26	Warbelow's Air Ventures		NaN	4W	WAV	WARBELOW
27	Wright Air Service		NaN	8V	WRF	WRIGHT FLYER

	Primary Hubs, Secondary Hubs	Founded	\

0	Hawthorne	2005
1	San Juan	1982
2	NomeKotzebueUnalakleet	1979
3	Dallas/Fort WorthDenverPhoenix-Sky Harbor	2007
4	FairbanksAnchorage	1978
5	Salmon	2014
6	Boulder CityGrand CanyonPage	1927
7	Grand Canyon	1967
8	AnchorageBethelCold BayDillinghamEmmonakKenaiK…	1971
9	Port Clinton	1937
10	Charlevoix	1945
11	BurbankOaklandLas VegasSanta AnaPhoenixConcord	2016
12	KenmoreSeattle-Lake UnionSeattle-Boeing	1946
13	Denver-CentennialDenverDenver-Rocky MountainGr…	1997
14	Kailua-KonaKahului	1994
15	Westerly	1970
16	Rockland	2004
17	Danbury	1988
18	Bellingham	2002
19	Kodiak	2003
20	MemphisDestinPittsburghWashington-Dulles	2013
21	HawthorneOaklandSan CarlosSanta BarbaraTruckee	2012
22	Ketchikan Harbor	1977
23	Oxford (CT)San Juan White Plains	2001
24	Cincinnati-Lunken	2009
25	Ogden	2015
26	Fairbanks	1958
27	Fairbanks	1966

	Notes
0	Has the EAS contract to serve Grant County Air…
1	NaN
2	NaN
3	NaN
4	Founded as Tatonduk Flying Service.
5	NaN
6	Founded as Scenic Airways.
7	Founded as Scenic Airlines.
8	Founded as Delta Air Services.
9	NaN
10	Founded as McPhillips Flying Service.
11	NaN
12	Founded as Mines Collins Munro.
13	Operates as Denver Air Connection.
14	Founded as Mokulele Flight Service.
15	NaN
16	NaN
17	NaN

```
18                                                 NaN
19                                                 NaN
20                                                 NaN
21                                                 NaN
22                                                 NaN
23                                                 NaN
24                                                 NaN
25                                                 NaN
26                                                 NaN
27                                                 NaN  ,
Airline  Image IATA ICAO        Callsign  \
0                  Air Charter Bahamas    NaN  NaN  NaN             NaN
1                  Air Flight Charters    NaN  NaN  FLL             NaN
2                              Airshare    NaN  NaN  XSR        AIRSHARE
3                       Berry Aviation    NaN  NaN  BYA           BERRY
4                     Bighorn Airways     NaN  NaN  BHR     BIGHORN AIR
5                 Charter Air Transport    NaN   VC  SRY        STINGRAY
6                       Choice Airways     NaN  NaN  CSX      CHOICE AIR
7                            ExcelAire     NaN  NaN  XLS       EXCELAIRE
8             Global Crossing Airlines     NaN   G6  GXA          GEMINI
9                       Great Lakes Air    NaN  NaN  NaN             NaN
10                    Gryphon Airlines     NaN   Y3  VOS             NaN
11                       IAero Airways     NaN   WQ  SWQ     SWIFTFLIGHT
12                         IBC Airways     NaN   II  CSQ         CHASQUI
13  L-3 Flight International Aviation     NaN  NaN  RTD         RIPTIDE
14              Liberty Jet Management     NaN  NaN  LRT     LIBERTY JET
15                               NetJets    NaN   1I  EJA         EXECJET
16              Omni Air International     NaN   X9  OAE    OMNI-EXPRESS
17                  Omni Air Transport     NaN  NaN  DRL         DRILLER
18                   Pacific Coast Jet     NaN  NaN  PXT      PACK COAST
19                  Pentastar Aviation     NaN  NaN  DCX           TANGO
20                         Phoenix Air     NaN  NaN  PHA       GRAY BIRD
21                          PlaneSense     NaN  NaN  CNS         CHRONOS
22                Presidential Airways     NaN  NaN  PRD    PRESIDENTIAL
23             Sierra Pacific Airlines     NaN   SI  SPA  SIERRA PACIFIC
24                               Skymax    NaN  NaN  SMX          SKYMAX
25                   Songbird Airways     NaN   SK  SGB        SONGBIRD
26                  Stampede Aviation     NaN  NaN  NaN             NaN
27                Superior Air Charter    NaN  NaN  RSP       REDSTRIPE
28                   Superior Aviation    NaN   SO  HKA       SPEND AIR
29                  Talkeetna Air Taxi    NaN  NaN  NaN             NaN
30                Tropic Ocean Airways    NaN  NaN  NaN             NaN
31             World Atlantic Airlines    NaN   K8  WAL  WORLD ATLANTIC
32                  XOJET Aviation LLC    NaN  NaN  XOJ           XOJET

   Primary Hubs, Secondary Hubs  Founded  \
0                           NaN      NaN
1                Fort Lauderdale   1987.0
```

2	NaN	2000.0
3	San Marcos	1983.0
4	Sheridan	1947.0
5	Cleveland-Lakefront	1997.0
6	Fort Lauderdale-Executive	2009.0
7	Long Island/Islip	1993.0
8	Atlantic CityLas VegasMiami	2019.0
9	St. Ignace	NaN
10	NaN	NaN
11	Miami	1997.0
12	Fort Lauderdale	1991.0
13	Newport News	1972.0
14	Long Island/Islip	2006.0
15	Columbus	1964.0
16	Tulsa	1993.0
17	Tulsa	NaN
18	NaN	2006.0
19	Waterford	1964.0
20	Cartersville	1978.0
21	Portsmouth (NH)	1992.0
22	Melbourne/Orlando	NaN
23	Tucson	1970.0
24	Fort Lauderdale	1997.0
25	Miami	1990.0
26	Healy/Denali NP	2011.0
27	NaN	2006.0
28	Lansing	1979.0
29	Talkeetna	1947.0
30	Fort Lauderdale	2009.0
31	Miami	2002.0
32	Sacramento-McClellan	2006.0

	Notes
0	NaN
1	NaN
2	Founded as Executive Flight Services
3	NaN
4	NaN
5	NaN
6	NaN
7	NaN
8	NaN
9	NaN
10	NaN
11	Founded as Swift Air
12	NaN
13	NaN
14	NaN

```
15                Founded as Executive Jets Aviation.
16                                                 NaN
17                                                 NaN
18                                                 NaN
19            Founded as Chrysler Air Transportation.
20                                                 NaN
21                                                 NaN
22                                                 NaN
23                      Commenced operations in 1971.
24                      Commenced operations in 2013.
25                                                 NaN
26                                                 NaN
27                                                 NaN
28                                                 NaN
29               Founded as Talkeetna Flying Service.
30                                                 NaN
31  Founded as Caribbean Sun Airlines and commence…
32                                                 NaN  ,
Airline  Image IATA  ICAO         Callsign  \
0                        21 Air    NaN   2I   CSB      CARGO SOUTH
1                       ABX Air    NaN   GB   ABX             ABEX
2            Air Cargo Carriers    NaN   2Q   SNC      NIGHT CARGO
3                AirNet Express    NaN  NaN   USC       STAR CHECK
4   Air Transport International    NaN   8C   ATN    AIR TRANSPORT
5         Alaska Central Express    NaN   KO   AER          ACE AIR
6                 Aloha Air Cargo    NaN   KH   AAH            ALOHA
7             Alpine Air Express    NaN   5A   AIP       ALPINE AIR
8                     Amazon Air    NaN  AFW  KAFW       AMAZON AIR
9                    Ameriflight    NaN   A8   AMF         AMFLIGHT
10       Amerijet International    NaN   M6   AJT         AMERIJET
11        Ameristar Jet Charter    NaN   7Z   AJI        AMERISTAR
12        Asia Pacific Airlines    NaN   P9   MGE         MAGELLAN
13                    Atlas Air    NaN   5Y   GTI            GIANT
14              Bemidji Airlines    NaN   CH   BMJ          BEMIDJI
15               Castle Aviation    NaN  NaN   CSJ           CASTLE
16                 Corporate Air    NaN  NaN   CPT          AIRSPUR
17                       CSA Air    NaN  NaN   IRO         IRON AIR
18               Empire Airlines    NaN   EM   CFS           EMPIRE
19              Everts Air Cargo    NaN   5V   VTS           EVERTS
20                 FedEx Express    NaN   FX   FDX            FEDEX
21      Freight Runners Express    NaN  NaN   FRG  FREIGHT RUNNERS
22                     IFL Group    NaN   IF   IFL           EIFFEL
23                   Kalitta Air    NaN   K4   CKS           CONNIE
24              Kalitta Charters    NaN   CB   KFS          KALITTA
25              Lynden Air Cargo    NaN   L2   LYC           LYNDEN
26                    Martinaire    NaN  NaN   MRA           MARTEX
27                Merlin Airways    NaN  NaN   MEI           AVALON
28            Mountain Air Cargo    NaN   C2   MTN         MOUNTAIN
```

29	National Airlines	NaN	N8	NCR	NATIONAL CARGO
30	Northern Air Cargo	NaN	NC	NAC	YUKON
31	Polar Air Cargo	NaN	PO	PAC	POLAR
32	Royal Air Freight	NaN	NaN	RAX	AIR ROYAL
33	Ryan Air Services	NaN	7S	RYA	RYAN AIR
34	Sky Lease Cargo	NaN	GG	KYE	SKY CUBE
35	Skyway Enterprises	NaN	KI	SKZ	SKYWAY-INC
36	Strat Air	NaN	NaN	NaN	NaN
37	Tepper Aviation	NaN	NaN	NaN	NaN
38	Trans Executive Airlines	NaN	KH	MUI	RHOADES EXPRESS
39	UPS Airlines	NaN	5X	UPS	UPS
40	USA Jet Airlines	NaN	UJ	JUS	JET USA
41	West Air	NaN	NaN	PCM	PAC VALLEY
42	Western Global Airlines	NaN	KD	WGN	WESTERN GLOBAL
43	Wiggins Airways	NaN	WG	WIG	WIGGINS AIRWAYS

	Primary Hubs, Secondary Hubs	Founded	\
0	Miami	2014.0	
1	Wilmington (OH)CincinnatiMiami	1980.0	
2	MilwaukeeCincinnati	1986.0	
3	Columbus-Rickenbacker	1974.0	
4	Wilmington (OH)Cincinnati	1978.0	
5	Anchorage	1996.0	
6	Honolulu	1946.0	
7	ProvoBillingsSioux Falls	1971.0	
8	Fort Worth/AllianceCincinnatiLeipzig/HalleSan …	2015.0	
9	Dallas/Fort Worth	1968.0	
10	MiamiPort of Spain	1974.0	
11	Dallas-AddisonEl PasoWillow Run	2000.0	
12	GuamHonolulu	1998.0	
13	New York-JFKAnchorageCincinnatiHoustonHuntsvil…	1992.0	
14	BemidjiMinneapolis/St. Paul	1946.0	
15	Akron/Canton	1986.0	
16	Billings	1981.0	
17	Iron Mountain	1998.0	
18	Coeur d' AleneSpokane	1977.0	
19	FairbanksAnchorage	1995.0	
20	MemphisAnchorageCologne/BonnDubaiFort WorthGre…	1971.0	
21	Milwaukee	1985.0	
22	WaterfordMiami	1983.0	
23	YpsilantiAnchorageBahrainCincinnatiHong KongNe…	1967.0	
24	Ypsilanti	NaN	
25	Anchorage	1995.0	
26	Addison	1978.0	
27	BillingsMiamiSan Juan	1983.0	
28	Kinston	1974.0	
29	Orlando/Sanford	1985.0	
30	AnchorageMiami	1956.0	

31	AnchorageCincinnatiHong KongHonoluluLos Angele…	1993.0	
32	Waterford	1961.0	
33	AnchorageAniakBethelEmmonakKotzebueNomeSt. Mar…	1953.0	
34	Miami	1969.0	
35	NaN	1981.0	
36	Miami	2018.0	
37	Crestview	1987.0	
38	Honolulu	1982.0	
39	LouisvilleChicago/RockfordCologne/BonnColumbia…	1988.0	
40	YpsilantiLaredo	1994.0	
41	Las VegasOaklandOntarioSacramentoSan Diego	1988.0	
42	Miami Liege, Belgium; AnchorageFort Myers, FL	2013.0	
43	Manchester	1929.0	

	Notes
0	NaN
1	Founded as Airborne Express. Operates some Ama…
2	Commenced operations in 1980.
3	Founded as Financial Air Express.
4	Founded as US Airways and commenced operations…
5	NaN
6	Founded as Trans-Pacific Airlines and separate…
7	NaN
8	Formerly Amazon Prime Air
9	Founded as California Air Charter.
10	NaN
11	NaN
12	NaN
13	Commenced operations in 1993. Operates some Am…
14	Commenced operations in 1947.
15	NaN
16	NaN
17	NaN
18	NaN
19	NaN
20	Founded as Federal Express and commenced opera…
21	NaN
22	Founded as Air Contract Cargo.
23	Founded as American International Airways.
24	NaN
25	NaN
26	NaN
27	NaN
28	NaN
29	Commenced operations in 1986.
30	NaN
31	NaN
32	NaN

```
33                    Founded as Unalakleet Air Taxi.
34  Founded as Wrangler Aviation and commenced ope…
35                       Commenced operations in 1983.
36                                                NaN
37                                                NaN
38                                                NaN
39                                                NaN
40                                                NaN
41                                                NaN
42                                                NaN
43                                                NaN  ,
Airline  Image  IATA  ICAO  Callsign  \
0   AirMed International    NaN   NaN   NaN       NaN
1            Air Methods    NaN   NaN   NaN       NaN
2  Critical Air Medicine    NaN   NaN   NaN       NaN
3               Lifestar    NaN   NaN   NaN       NaN
4              Life Lion    NaN   NaN   NaN       NaN

  Primary Hubs, Secondary Hubs  Founded                              Notes
0     Birmingham-Shuttlesworth   1987.0  Founded as MEDjet International.
1            Denver-Centennial   1980.0                               NaN
2                          NaN   1984.0                               NaN
3                          NaN      NaN                               NaN
4                          NaN      NaN                               NaN  ,
Airline  Image  IATA ICAO  \
0                                          Comco    NaN   NaN  NaN
1                                          Janet    NaN   NaN  WWW
2  Justice Prisoner and Alien Transportation System    NaN   NaN  JUD

  Callsign Primary Hubs, Secondary Hubs  Founded  \
0      NaN                          NaN     2002
1    JANET                    Las Vegas     1972
2  JUSTICE                Oklahoma City     1980

                         Notes
0                          NaN
1                          NaN
2  Commenced operations in 1995.  ,                                vteLists of
airlines  \
0                                  By airline codes
1                                      By continent
2                                        By country
3                           vteExpand for full list
4   A Abkhazia Afghanistan Akrotiri and Dhekelia Å…
5                                                 A
6                                                 B
7                                                 C
8                                                 D
```

```
9                                                    E
10                                                   F
11                                                   G
12                                                   H
13                                                   I
14                                                   J
15                                                   K
16                                                   L
17                                                   M
18                                                   N
19                                                   O
20                                                   P
21                                                   Q
22                                                   R
23                                                   S
24                                                   T
25 U Uganda Ukraine United Arab Emirates United K…
26                                                   U
27                                                   V
28                                                   W
29                                                   Y
30                                                   Z
31                                          See also

                              vteLists of airlines.1
0  All 0-9 A B C D E F G H I J K L M N O P Q R S …
1                Africa Americas Asia Europe Oceania
2  vteExpand for full listA Abkhazia Afghanistan …
3                             vteExpand for full list
4  A Abkhazia Afghanistan Akrotiri and Dhekelia Å…
5  Abkhazia Afghanistan Akrotiri and Dhekelia Åla…
6  The Bahamas Bahrain Bangladesh Barbados Belaru…
7  Cambodia Cameroon Canada Cape Verde Cayman Isl…
8  Denmark Dhekelia Djibouti Dominica Dominican R…
9  East Timor Ecuador Egypt El Salvador Equatoria…
10 Falkland Islands Faroe Islands Fiji Finland Fr…
11 Gabon The Gambia Georgia Germany Ghana Gibralt…
12                  Haiti Honduras Hong Kong Hungary
13 Iceland India Indonesia Iran Iraq Ireland Isra…
14                       Jamaica Japan Jersey Jordan
15 Kazakhstan Kenya Kiribati North Korea South Ko…
16 Laos Latvia Lebanon Lesotho Liberia Libya Liec…
17 Macau Macedonia, Republic of Madagascar Malawi…
18 Namibia Nauru Nepal Netherlands Netherlands An…
19                                                Oman
20 Pakistan Palau Palestine Panama Papua New Guin…
21                                               Qatar
22                             Romania Russia Rwanda
```

```
23  Sahrawi Arab Democratic Republic Saint Barthél…
24  Taiwan Tajikistan Tanzania Thailand Togo Tokel…
25  U Uganda Ukraine United Arab Emirates United K…
26  Uganda Ukraine United Arab Emirates United Kin…
27  Vanuatu Vatican City Venezuela Vietnam British…
28                                 Wallis and Futuna
29                                             Yemen
30                                Zambia Zimbabwe
31  List of airline holding companies List of airl…  ,
vteExpand for full list  \
0   A Abkhazia Afghanistan Akrotiri and Dhekelia Å…
1                                                  A
2                                                  B
3                                                  C
4                                                  D
5                                                  E
6                                                  F
7                                                  G
8                                                  H
9                                                  I
10                                                 J
11                                                 K
12                                                 L
13                                                 M
14                                                 N
15                                                 O
16                                                 P
17                                                 Q
18                                                 R
19                                                 S
20                                                 T
21  U Uganda Ukraine United Arab Emirates United K…
22                                                 U
23                                                 V
24                                                 W
25                                                 Y
26                                                 Z

                          vteExpand for full list.1
0   A Abkhazia Afghanistan Akrotiri and Dhekelia Å…
1   Abkhazia Afghanistan Akrotiri and Dhekelia Åla…
2   The Bahamas Bahrain Bangladesh Barbados Belaru…
3   Cambodia Cameroon Canada Cape Verde Cayman Isl…
4   Denmark Dhekelia Djibouti Dominica Dominican R…
5   East Timor Ecuador Egypt El Salvador Equatoria…
6   Falkland Islands Faroe Islands Fiji Finland Fr…
7   Gabon The Gambia Georgia Germany Ghana Gibralt…
8                   Haiti Honduras Hong Kong Hungary
```

```
9   Iceland India Indonesia Iran Iraq Ireland Isra…
10                        Jamaica Japan Jersey Jordan
11  Kazakhstan Kenya Kiribati North Korea South Ko…
12  Laos Latvia Lebanon Lesotho Liberia Libya Liec…
13  Macau Macedonia, Republic of Madagascar Malawi…
14  Namibia Nauru Nepal Netherlands Netherlands An…
15                                              Oman
16  Pakistan Palau Palestine Panama Papua New Guin…
17                                            Qatar
18                                Romania Russia Rwanda
19  Sahrawi Arab Democratic Republic Saint Barthél…
20  Taiwan Tajikistan Tanzania Thailand Togo Tokel…
21  U Uganda Ukraine United Arab Emirates United K…
22  Uganda Ukraine United Arab Emirates United Kin…
23  Vanuatu Vatican City Venezuela Vietnam British…
24                                   Wallis and Futuna
25                                               Yemen
26                                  Zambia Zimbabwe  ,     0
1
0   A  Abkhazia Afghanistan Akrotiri and Dhekelia Åla…
1   B  The Bahamas Bahrain Bangladesh Barbados Belaru…
2   C  Cambodia Cameroon Canada Cape Verde Cayman Isl…
3   D  Denmark Dhekelia Djibouti Dominica Dominican R…
4   E  East Timor Ecuador Egypt El Salvador Equatoria…
5   F  Falkland Islands Faroe Islands Fiji Finland Fr…
6   G  Gabon The Gambia Georgia Germany Ghana Gibralt…
7   H                    Haiti Honduras Hong Kong Hungary
8   I  Iceland India Indonesia Iran Iraq Ireland Isra…
9   J                         Jamaica Japan Jersey Jordan
10  K  Kazakhstan Kenya Kiribati North Korea South Ko…
11  L  Laos Latvia Lebanon Lesotho Liberia Libya Liec…
12  M  Macau Macedonia, Republic of Madagascar Malawi…
13  N  Namibia Nauru Nepal Netherlands Netherlands An…
14  O                                               Oman
15  P  Pakistan Palau Palestine Panama Papua New Guin…
16  Q                                             Qatar
17  R                                 Romania Russia Rwanda
18  S  Sahrawi Arab Democratic Republic Saint Barthél…
19  T  Taiwan Tajikistan Tanzania Thailand Togo Tokel…,    0
1
0  U  Uganda Ukraine United Arab Emirates United Kin…
1  V  Vanuatu Vatican City Venezuela Vietnam British…
2  W                                   Wallis and Futuna
3  Y                                               Yemen
4  Z                                  Zambia Zimbabwe,
vteAirlines of the United States  \
0                                          Mainline
1                                          Regional
```

```
2                                        Affiliated
3                                       Independent
4                                             Cargo
5                                           Charter
6                                 Air taxi and tours
7                                      Air ambulance
8                                        Government
9  List of airline holding companies List of airl…

                  vteAirlines of the United States.1
0  Alaska Airlines Allegiant Air American Airline…
1  Affiliated Air Wisconsin CommutAir Endeavor Ai…
2  Air Wisconsin CommutAir Endeavor Air Envoy Air…
3  Advanced Air Air Flamenco Air Sunshine Bering …
4  ABX Air Air Cargo Carriers Air Transport Inter…
5  Air Charter Bahamas Airstream Jets Alerion Avi…
6  Gem Air Grand Canyon Scenic Airlines Griffing …
7  Air Evac Lifeteam AirMed International Air Met…
8                  Comco Janet JPATS Patriot Express
9  List of airline holding companies List of airl…  ,              0
1
0   Affiliated  Air Wisconsin CommutAir Endeavor Air Envoy Air…
1  Independent  Advanced Air Air Flamenco Air Sunshine Bering …,
vteList of airlines of the Americas  \
0  North America Caribbean Central America Northe…
1                                  Sovereign states
2                 Dependencies andother territories

               vteList of airlines of the Americas.1  \
0  North America Caribbean Central America Northe…
1  Antigua and Barbuda Argentina Bahamas Barbados…
2  Anguilla Aruba Bermuda Bonaire British Virgin …

               vteList of airlines of the Americas.2
0  North America Caribbean Central America Northe…
1                                              NaN
2                                              NaN  ,
0                    1
0  Authority control: National libraries  Israel United States]
```

[26]:
```
tables[0]
```

[26]:

	Airline	Image	IATA	ICAO	Callsign	\
0	Alaska Airlines	NaN	AS	ASA	ALASKA	
1	Allegiant Air	NaN	G4	AAY	ALLEGIANT	
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/TacomaAnchoragePortland (OR)San Franci…     1932
1   Las VegasCincinnatiFort Walton BeachIndianapol…     1997
2   Dallas/Fort WorthCharlotteChicago-O'HareLos An…     1926
3                              BurbankNew HavenOrlando     1987
4      CharlestonHartfordNew OrleansNorfolkProvoTampa     2018
5   AtlantaBostonDetroitLos AngelesMinneapolis/St…     1924
6                                   MiamiNew York-JFK     2010
7   DenverAtlantaChicago-O'HareCincinnatiCleveland…     1994
8                                     HonoluluKahului     1929
9   New York-JFKBostonLos AngelesFort LauderdaleOr…     1998
10  Dallas-LoveAtlantaBaltimoreChicago-MidwayDenve…     1967
11  Atlantic CityDetroitLas VegasFort LauderdaleCh…     1980
12     Minneapolis/St. PaulDallas/Fort WorthLas Vegas     1982
13  Chicago-O'HareDenverGuamHouston-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…
```

[32]:
```
tables[6]
```

[32]:
```
                                      Airline  Image  IATA ICAO  \
0                                       Comco    NaN   NaN  NaN
```

```
1                                          Janet    NaN   NaN  WWW
2  Justice Prisoner and Alien Transportation System    NaN   NaN  JUD

  Callsign Primary Hubs, Secondary Hubs  Founded  \
0      NaN                         NaN     2002
1    JANET                   Las Vegas     1972
2  JUSTICE               Oklahoma City     1980

                          Notes
0                           NaN
1                           NaN
2  Commenced operations in 1995.
```

[33]:
```
# Lets first merge all wikipedia table.
wiki_table =␣
 ↪[tables[0],tables[1],tables[2],tables[3],tables[4],tables[5],tables[6]]
```

[34]:
```
wiki_tables = pd.concat(wiki_table, ignore_index=True)
```

[35]:
```
wiki_tables
```

[35]:
```
                                       Airline  Image IATA ICAO  \
0                                Alaska Airlines    NaN   AS  ASA
1                                  Allegiant Air    NaN   G4  AAY
2                               American Airlines    NaN   AA  AAL
3                                  Avelo Airlines    NaN   XP  VXP
4                                  Breeze Airways    NaN   MX  MXY
..                                            …    …  …  …
137                                    Lifestar    NaN  NaN  NaN
138                                   Life Lion    NaN  NaN  NaN
139                                       Comco    NaN  NaN  NaN
140                                       Janet    NaN  NaN  WWW
141  Justice Prisoner and Alien Transportation System    NaN  NaN  JUD

      Callsign                        Primary hubs, Secondary hubs  Founded  \
0       ALASKA  Seattle/TacomaAnchoragePortland (OR)San Franci…   1932.0
1    ALLEGIANT  Las VegasCincinnatiFort Walton BeachIndianapol…   1997.0
2     AMERICAN  Dallas/Fort WorthCharlotteChicago-O'HareLos An…   1926.0
3        AVELO                              BurbankNew HavenOrlando   1987.0
4         MOXY     CharlestonHartfordNew OrleansNorfolkProvoTampa   2018.0
..         …                                                  …       …
137        NaN                                                NaN      NaN
138        NaN                                                NaN      NaN
139        NaN                                                NaN   2002.0
140      JANET                                                NaN   1972.0
141    JUSTICE                                                NaN   1980.0
```

```
                                                  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
..                                                 …
137                                                NaN
138                                                NaN
139                                                NaN
140                                                NaN
141                        Commenced operations in 1995.

    Primary Hubs, Secondary Hubs
0                            NaN
1                            NaN
2                            NaN
3                            NaN
4                            NaN
..                           …
137                          NaN
138                          NaN
139                          NaN
140                    Las Vegas
141                Oklahoma City

[142 rows x 9 columns]
```

1.2.2 c. You should then get all the information gathered so far in one place.

First we got only that column from wiki pedia table that we need to merge.

[36]:
```
wiki_df = wiki_tables[['IATA', "Founded"]]
wiki_df
```

[36]:
```
    IATA  Founded
0     AS   1932.0
1     G4   1997.0
2     AA   1926.0
3     XP   1987.0
4     MX   2018.0
..    …       …
137  NaN      NaN
138  NaN      NaN
139  NaN   2002.0
140  NaN   1972.0
```

```
141  NaN   1980.0

[142 rows x 2 columns]
```

[37]:
```
# Now we gather all the information that we got from wiki pedia link and the␣
 ↪data that we have.
df = final_df.merge(wiki_df, left_on ='Airline', right_on = "IATA")
```

[38]:
```
df
```

[38]:

	id	Airline	Flight	AirportFrom	AirportTo	DayOfWeek	Time	Length	\
0	4	AA	2466	SFO	DFW	3	20	195	
1	231	AA	526	SFO	DFW	3	360	215	
2	234	AA	552	SFO	MIA	3	360	315	
3	905	AA	810	SFO	ORD	3	385	255	
4	1739	AA	24	SFO	JFK	3	425	325	
…	…	…	…	…	…	…	…	…	
434919	497838	9E	4292	LWB	JFK	3	890	110	
434920	516333	9E	4292	LWB	JFK	4	890	110	
434921	534123	9E	4292	LWB	JFK	5	890	110	
434922	69058	9E	3752	ABR	MSP	7	410	76	
434923	189396	9E	3752	ABR	MSP	7	410	76	

	Delay	ident	…	iata_code	airport_ref	airport_ident	length_ft	\
0	1	KSFO	…	SFO	3878	KSFO	7500.0	
1	0	KSFO	…	SFO	3878	KSFO	7500.0	
2	1	KSFO	…	SFO	3878	KSFO	7500.0	
3	0	KSFO	…	SFO	3878	KSFO	7500.0	
4	1	KSFO	…	SFO	3878	KSFO	7500.0	
…	…	…	…	…	…	…	…	
434919	1	KLWB	…	LWB	20390	KLWB	7004.0	
434920	0	KLWB	…	LWB	20390	KLWB	7004.0	
434921	0	KLWB	…	LWB	20390	KLWB	7004.0	
434922	1	KABR	…	ABR	3358	KABR	6901.0	
434923	0	KABR	…	ABR	3358	KABR	6901.0	

	width_ft	surface	lighted	closed	IATA	Founded
0	200.0	ASP	1	0	AA	1926.0
1	200.0	ASP	1	0	AA	1926.0
2	200.0	ASP	1	0	AA	1926.0
3	200.0	ASP	1	0	AA	1926.0
4	200.0	ASP	1	0	AA	1926.0
…	…	…	…	…	…	…
434919	150.0	ASP	1	0	9E	1985.0
434920	150.0	ASP	1	0	9E	1985.0
434921	150.0	ASP	1	0	9E	1985.0
434922	100.0	CON	1	0	9E	1985.0

```
434923     100.0     CON       1       0   9E   1985.0

[434924 rows x 26 columns]
```

1.2.3 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

[39]:
```
# Now lets use the web scrapping to import the data frome the wikipedia.
url2 = "https://en.wikipedia.org/wiki/
 ↪List_of_the_busiest_airports_in_the_United_States"
table = pd.read_html(url2)
```

[40]:
```
table
```

[40]:
```
[    Rank(2021)                          Airports (large hubs) IATACode  \
 0             1  Hartsfield-Jackson Atlanta International Airport      ATL
 1             2                 Los Angeles International Airport      LAX
 2             3              Chicago O'Hare International Airport      ORD
 3             4           Dallas/Fort Worth International Airport      DFW
 4             5                      Denver International Airport      DEN
 5             6             John F. Kennedy International Airport      JFK
 6             7               San Francisco International Airport      SFO
 7             8               Seattle-Tacoma International Airport      SEA
 8             9                     Orlando International Airport      MCO
 9            10                  Harry Reid International Airport      LAS
 10           11           Charlotte-Douglas International Airport      CLT
 11           12              Newark Liberty International Airport      EWR
 12           13          Phoenix Sky Harbor International Airport      PHX
 13           14              George Bush Intercontinental Airport      IAH
 14           15                       Miami International Airport      MIA
 15           16                Boston Logan International Airport      BOS
 16           17     Minneapolis-Saint Paul International Airport      MSP
 17           18                      Detroit Metropolitan Airport      DTW
 18           19   Fort Lauderdale-Hollywood International Airport      FLL
 19           20                Philadelphia International Airport      PHL
 20           21                         New York LaGuardia Airport      LGA
 21           22        Baltimore/Washington International Airport      BWI
 22           23              Salt Lake City International Airport      SLC
 23           24                   San Diego International Airport      SAN
 24           25           Washington Dulles International Airport      IAD
```

25	26	Ronald Reagan Washington National Airport	DCA
26	27	Tampa International Airport	TPA
27	28	Chicago Midway International Airport	MDW

	Major cities served	State	2021	2020[3]	2019[4]	\
0	Atlanta	GA	53485010	20559866	53505795	
1	Los Angeles	CA	42880266	18593421	35778573	
2	Chicago	IL	40872527	16243216	33592945	
3	Dallas/Fort Worth	TX	35764976	14606034	40871223	
4	Denver	CO	33575410	14055777	42939104	
5	New York	NY	31036875	12952869	24199688	
6	San Francisco	CA	27701838	10467728	24562271	
7	Seattle	WA	24962342	10584059	24728361	
8	Orlando	FL	24550287	10531436	22433552	
9	Las Vegas	NV	24411096	8786007	21421031	
10	Charlotte	NC	24181195	9462411	25001762	
11	Newark	NJ	23140821	8682558	21905309	
12	Phoenix	AZ	22411342	8269819	31036655	
13	Houston	TX	21897049	7985474	23160763	
14	Miami	FL	21292994	8015744	17950989	
15	Boston	MA	20679409	7069720	19192917	
16	Minneapolis	MN	19152347	7745057	27779230	
17	Detroit	MI	18122696	6822324	18143040	
18	Fort Lauderdale	FL	17936817	6035452	20699377	
19	Philadelphia	PA	15989906	5753239	12840841	
20	New York	NY	15392968	5753239	16006389	
21	Baltimore & Washington, D.C.	MD	9253561	5451355	13284687	
22	Salt Lake City	UT	12829197	4966775	10978756	
23	San Diego	CA	12641360	4637856	12648692	
24	Washington, D.C.	VA	11857307	4147116	15393601	
25	Arlington	VA	11576617	4236603	10081781	
26	Tampa	FL	10921049	4013995	8935654	
27	Chicago	IL	10064875	3862658	11884117	

	2018[5]	2017[6]	2016[7]	2015[8]	2014[9]	2013[10]	2012[11]	\
0	51865797	50251964	50501858	49340732	46604273	45308407	45798928	
1	32821799	31816933	31283579	31589839	30804567	29038128	28022904	
2	31362941	29809097	28267394	26280043	26000591	25496885	25799841	
3	39873927	38593028	37589899	36305668	33843426	32317835	32171795	
4	42624050	41232432	39636042	36351272	34314197	32425892	31326268	
5	22281949	22011251	21511880	21913166	21537725	21346601	20033816	
6	23202480	21565448	20283541	18759938	17278608	16884524	17159427	
7	23795012	23364393	22833267	21857693	20620248	19946179	19959651	
8	21622580	21185458	20896265	21351504	20344867	19525109	19560870	
9	21021640	20709225	20875813	20986349	19471466	19420089	18987488	
10	24024908	22639124	21887110	20148980	17888080	16690295	16121123	
11	21157398	19603731	20062072	20595881	19772087	18952840	19039000	

```
12  30620769  29533154  29239151  27782369  26244928  25036358  24520981
13  22797602  21571198  19923009  18684818  17773405  17546506  17055993
14  17612331  15817043  14263270  13061632  12031860  11538140  11445103
15  18361942  18409704  18123844  17634273  16972678  16280835  15943878
16  27790717  26900048  25707101  24190560  22770783  21704626  21284236
17  17436837  17036092  16847135  16255520  15775941  15683523  15599879
18  20006521  18759742  17759044  16290362  15507561  14810153  14293695
19  12226730  11615954  11143738  10634538  10139065   9668048   9579840
20  15292670  14271243  14564419  15101349  14792339  14727945  14589337
21  13371816  12976554  12340972  11738845  11022200  11132731  11186444
22  10368514   9548580   9194994   9150458   8531561   8267752   8218487
23  12174224  11139933  10340164   9985763   9333152   8878772   8686621
24  15058501  14614802  14762593  14319924  13535372  13372269  12818717
25  10678018  10912074  11044387  10830850  10311996   9915646   9436387
26   8017347   6902771   6338517   5715205   5396958   5050989   4797102
27  11621623  11024306  10596942  10363974  10415948  10570993  10816216

    2011[12]
0   44414121
1   27518358
2   25667499
3   31892301
4   30528737
5   19022535
6   17250415
7   19872617
8   19750306
9   18342158
10  15971676
11  19306660
12  23664832
13  16814092
14  11332466
15  15895653
16  20056568
17  15716865
18  14180730
19   9701756
20  14883180
21  11067319
22   8174194
23   8465683
24  11989227
25   9134576
26   4673047
27  11044383  ,
    Rank(2020)                          Airports (medium hubs) IATACode  \
```

0	29	Daniel K. Inouye International Airport	HNL
1	30	Portland International Airport	PDX
2	31	Nashville International Airport	BNA
3	32	Austin-Bergstrom International Airport	AUS
4	33	Dallas Love Field	DAL
5	34	St. Louis Lambert International Airport	STL
6	35	Norman Y. Mineta San Jose International Airport	SJC
7	36	William P. Hobby Airport	HOU
8	37	Raleigh-Durham International Airport	RDU
9	38	Louis Armstrong New Orleans International Airport	MSY
10	39	Oakland International Airport	OAK
11	40	Sacramento International Airport	SMF
12	41	Kansas City International Airport	MCI
13	42	John Wayne Airport	SNA
14	43	Fort Myers International Airport	RSW
15	44	San Antonio International Airport	SAT
16	45	Cleveland Hopkins International Airport	CLE
17	46	Indianapolis International Airport	IND
18	47	Pittsburgh International Airport	PIT
19	48	Luis Munoz Marin International Airport	SJU
20	49	Cincinnati/Northern Kentucky International Air…	CVG
21	50	John Glenn Columbus International Airport	CMH
22	51	Kahului Airport	OGG
23	52	Jacksonville International Airport	JAX
24	53	Palm Beach International Airport	PBI
25	54	General Mitchell International Airport	MKE
26	55	Bradley International Airport	BDL
27	56	Hollywood Burbank Airport	BUR
28	57	Ontario International Airport	ONT
29	58	Ted Stevens Anchorage International Airport	ANC
30	59	Albuquerque International Sunport	ABQ
31	60	Omaha Eppley Airfield	OMA
32	61	Buffalo Niagara International Airport	BUF
33	62	Charleston International Airport	CHS
34	63	Memphis International Airport	MEM
35	64	Richmond International Airport	RIC

	City served	State	2020	2019	2018	2017	\
0	Honolulu	HI	9893930	8408457	8134848.0	7876769.0	
1	Portland	OR	9790489	11595454	11367176.0	11506310.0	
2	Nashville	TN	8498877	9797408	9940866.0	9435473.0	
3	Austin	TX	3141505	8683711	7921797.0	6973115.0	
4	Dallas	TX	8069178	7069614	6937061.0	6741870.0	
5	St Louis	MO	7750391	9988678	9578505.0	9743989.0	
6	San Jose	CA	7676765	7946986	7822274.0	7372805.0	
7	Houston	TX	7057139	5144467	4719568.0	4461304.0	
8	Raleigh	NC	6907342	6454413	6031630.0	5460526.0	

9	New Orleans	LA	6858606	6717105	6565482.0	6005527.0
10	Oakland	CA	6540851	4590117	4033412.0	4163587.0
11	Sacramento	CA	6448496	6919429	6416822.0	5851004.0
12	Kansas City	MO	5748186	7828885	7140616.0	6225148.0
13	Santa Ana	CA	5150294	6560230	6798321.0	6530308.0
14	Fort Myers	FL	5039616	5759419	5935131.0	5744918.0
15	San Antonio	TX	5021156	4894541	4836580.0	4562740.0
16	Cleveland	OH	4884126	4709183	4695040.0	4376432.0
17	Indianapolis	IN	4684958	5022980	4844427.0	4521611.0
18	Pittsburgh	PA	4675836	5153276	5317149.0	5195047.0
19	San Juan	Puerto Rico	4543406	4715947	4670033.0	4327431.0
20	Cincinnati	OH/KY	4396395	4413457	4269258.0	3926158.0
21	Columbus	OH	4155596	4172067	4054572.0	3765007.0
22	Kahului	HI	3776732	3460429	3263042.0	3166532.0
23	Jacksonville	FL	3472501	3479923	3118540.0	2759067.0
24	Palm Beach	FL	3453385	3374073	3548817.0	3452544.0
25	Milwaukee	WI	3358946	2723002	2499171.0	2247645.0
26	Hartford	CT	3318301	2713843	2642901.0	2556188.0
27	Burbank	CA	2985033	3323614	3330734.0	3214976.0
28	Ontario	CA	2717141	3791807	3572133.0	3442189.0
29	Anchorage	AK	2988720	2654702	NaN	NaN
30	Albuquerque	NM	2638245	2455274	2457087.0	2303223.0
31	Omaha	NB	2450981	2318442	NaN	NaN
32	Buffalo	NY	2448241	2057750	NaN	NaN
33	Charleston	SC	2375944	2375660	NaN	NaN
34	Memphis	TN	2307944	2375868	NaN	NaN
35	Richmond	VA	2187142	2142156	NaN	NaN

	2016	2015[7]	2014[1]
0	7554596.0	7040921.0	4522341.0
1	11470854.0	11242375.0	10057794.0
2	9071154.0	8340234.0	7878760.0
3	6095545.0	5797547.0	5219982.0
4	6285181.0	5937944.0	5800726.0
5	9656340.0	9656340.0	9463000.0
6	6793076.0	6239231.0	6108758.0
7	4350650.0	4231134.0	4025959.0
8	4969366.0	4816440.0	4384616.0
9	5569705.0	5329696.0	4870569.0
10	4343354.0	4218785.0	4150828.0
11	5401714.0	4954717.0	4673869.0
12	5321603.0	4885690.0	4621003.0
13	5934639.0	5506672.0	5069257.0
14	5391557.0	5135127.0	4982722.0
15	4205739.0	4083476.0	3686315.0
16	4216766.0	3889567.0	3605908.0
17	4179994.0	4091389.0	4046856.0

```
18   5217242.0   4945175.0   4584147.0
19   3986114.0   3890677.0   3827860.0
20   3269979.0   3036697.0   2874684.0
21   3567864.0   3312496.0   3115501.0
22   3100624.0   3113485.0   2926242.0
23   2799587.0   2716465.0   2589198.0
24   3383271.0   3229876.0   3228607.0
25   2127387.0   2089801.0   2037346.0
26   2563524.0   2525876.0   2381826.0
27   2982194.0   2926047.0   2913380.0
28   3352813.0   3220753.0   3019338.0
29         NaN         NaN         NaN
30   2127387.0   2046155.0   2020354.0
31         NaN         NaN         NaN
32         NaN         NaN         NaN
33         NaN         NaN         NaN
34         NaN         NaN         NaN
35         NaN         NaN         NaN  ,
   Rank Rank change                                    Airport name  \
   Rank Rank change                                    Airport name
0     1         NaN   Hartsfield-Jackson Atlanta International Airport
1     2         2.0            Dallas/Fort Worth International Airport
2     3         2.0                      Denver International Airport
3     4         1.0                      O'Hare International Airport
4     5         3.0                 Los Angeles International Airport
5     6         5.0            Charlotte Douglas International Airport
6     7         2.0                  Harry Reid International Airport
7     8         5.0           Phoenix Sky Harbor International Airport
8     9         1.0                     Orlando International Airport
9    10         2.0                Seattle-Tacoma International Airport
10   11         3.0                       Miami International Airport
11   12         3.0                George Bush Intercontinental Airport
12   13         7.0               John F. Kennedy International Airport
13   14         5.0    Fort Lauderdale-Hollywood International Airport
14   15         8.0                 San Francisco International Airport
15   16         4.0                Newark Liberty International Airport
16   17         NaN        Minneapolis-Saint Paul International Airport
17   18         NaN                       Detroit Metropolitan Airport
18   19         3.0  General Edward Lawrence Logan International Ai…
19   20         3.0                Salt Lake City International Airport
20   21         1.0                  Philadelphia International Airport
21   22         NaN          Baltimore/Washington International Airport
22   23         4.0                       Tampa International Airport
23   24         NaN                   San Diego International Airport
24   25         4.0                Chicago Midway International Airport
25   26         1.0             Washington Dulles International Airport
26   27         4.0                     Nashville International Airport
```

27	28	7.0	LaGuardia Airport
28	29	4.0	Dallas Love Field
29	30	4.0	Ronald Reagan Washington National Airport
30	31	1.0	Portland International Airport
31	32	4.0	Daniel K. Inouye International Airport
32	33	NaN	William P. Hobby Airport
33	34	2.0	Austin-Bergstrom International Airport
34	35	1.0	St. Louis Lambert International Airport

	Location	IATA Code	Traffic		Aircraft
\					
	Location	IATA Code	Passengers	% chg.2019/20	Movements
0	College Park, Georgia	ATL	42918685	61.2	NaN
1	Irving, Texas	DFW	39364990	47.6	NaN
2	Denver, Colorado	DEN	33741129	51.1	NaN
3	Chicago, Illinois	ORD	30860251	63.5	NaN
4	Los Angeles, California	LAX	28779527	67.3	NaN
5	Charlotte, North Carolina	CLT	27205082	45.8	NaN
6	Paradise, Nevada	LAS	22201479	56.9	NaN
7	Phoenix, Arizona	PHX	21978708	52.5	NaN
8	Orlando, Florida	MCO	21617803	57.3	NaN
9	SeaTac, Washington	SEA	20061507	61.3	NaN
10	Miami, Florida	MIA	18663858	59.4	NaN
11	Houston, Texas	IAH	18213571	59.8	NaN
12	Queens, New York	JFK	16630642	73.4	NaN
13	Fort Lauderdale, Florida	FLL	16484132	55.1	NaN
14	San Mateo County, California	SFO	16409625	71.5	NaN
15	Newark, New Jersey	EWR	15892892	65.7	NaN
16	Minneapolis, Minnesota	MSP	14851289	59.8	NaN
17	Romulus, Michigan	DTW	14105007	61.6	NaN
18	Boston, Massachusetts	BOS	12618128	70.3	NaN
19	Salt Lake City, Utah	SLC	12559026	53.2	NaN
20	Philadelphia, Pennsylvania	PHL	11865006	64.1	NaN
21	Linthicum Heights, Maryland	BWI	11204511	58.5	NaN
22	Tampa, Florida	TPA	10238151	54.5	NaN
23	San Diego, California	SAN	8991533	64.3	NaN
24	Chicago, Illinois	MDW	8853948	57.5	NaN
25	Dulles, Virginia	IAD	8333460	66.4	NaN
26	Nashville, Tennessee	BNA	8284570	54.7	NaN
27	Queens, New York	LGA	8245192	73.5	NaN
28	Dallas, Texas	DAL	7684653	54.2	NaN
29	Arlington, Virginia	DCA	7574966	68.4	NaN
30	Portland, Oregon	PDX	7084543	64.4	NaN
31	Honolulu, Hawaii	HNL	6656825	69.6	NaN
32	Houston, Texas	HOU	6476309	55.2	NaN
33	Austin, Texas	AUS	6472579	62.7	NaN
34	St Louis, Missouri	STL	6302402	60.3	NaN

```
   % chg.2019/20
0            0.0
1            NaN
2            NaN
3            NaN
4            NaN
5            NaN
6            NaN
7            NaN
8            NaN
9            NaN
10           NaN
11           NaN
12           NaN
13           NaN
14           NaN
15           NaN
16           NaN
17           NaN
18           NaN
19           NaN
20           NaN
21           NaN
22           NaN
23           NaN
24           NaN
25           NaN
26           NaN
27           NaN
28           NaN
29           NaN
30           NaN
31           NaN
32           NaN
33           NaN
34           NaN  ,
  Location of 35 busiest airports in the United States
0  .mw-parser-output .locmap .od{position:absolut…  ,
   Rank Rank change                                      Airport name  \
   Rank Rank change                                      Airport name
0     1         NaN   Hartsfield-Jackson Atlanta International Airport
1     2         NaN              Los Angeles International Airport[13]
2     3         NaN                        O'Hare International Airport
3     4         NaN            Dallas/Fort Worth International Airport
4     5         NaN                        Denver International Airport
```

5	6	NaN	John F. Kennedy International Airport[14]
6	7	NaN	San Francisco International Airport
7	8	NaN	Seattle-Tacoma International Airport[15]
8	9	NaN	Harry Reid International Airport[16]
9	10	NaN	Orlando International Airport
10	11	NaN	Charlotte Douglas International Airport
11	12	NaN	Newark Liberty International Airport[17]
12	13	1.0	Phoenix Sky Harbor International Airport[18]
13	14	1.0	Miami International Airport
14	15	NaN	George Bush Intercontinental Airport[19]
15	16	NaN	General Edward Lawrence Logan International Ai…
16	17	NaN	Minneapolis-Saint Paul International Airport[21]
17	18	1.0	Detroit Metropolitan Airport[22]
18	19	1.0	Fort Lauderdale-Hollywood International Airpor…
19	20	NaN	Philadelphia International Airport
20	21	NaN	LaGuardia Airport[24]
21	22	NaN	Baltimore/Washington International Airport
22	23	NaN	Salt Lake City International Airport[25]
23	24	NaN	San Diego International Airport[26]
24	25	NaN	Washington Dulles International Airport
25	26	NaN	Ronald Reagan Washington National Airport
26	27	1.0	Tampa International Airport[27]
27	28	1.0	Daniel K. Inouye International Airport[28]
28	29	2.0	Chicago Midway International Airport
29	30	NaN	Portland International Airport[29]
30	31	1.0	Nashville International Airport[30]
31	32	1.0	Austin-Bergstrom International Airport
32	33	2.0	Dallas Love Field[31]
33	34	NaN	St. Louis Lambert International Airport[32]
34	35	NaN	Norman Y. Mineta San Jose International Airpor…

	Location	IATA Code	Traffic		Aircraft
\					
	Location	IATA Code	Passengers	% chg.2018/19	Movements
0	College Park, Georgia	ATL	110531300	2.3	904301.0
1	Los Angeles, California	LAX	88068013	0.6	691257.0
2	Chicago, Illinois	ORD	84649115	1.7	919704.0
3	Irving, Texas	DFW	75066956	8.6	720007.0
4	Denver, Colorado	DEN	69015703	7.0	640098.0
5	Queens, New York	JFK	62551072	1.5	456060.0
6	San Mateo County, California	SFO	57488023	0.5	458496.0
7	SeaTac, Washington	SEA	51829239	4.0	450487.0
8	Paradise, Nevada	LAS	51537638	3.7	552962.0
9	Orlando, Florida	MCO	50613072	6.1	357689.0
10	Charlotte, North Carolina	CLT	50168783	8.0	578263.0
11	Newark, New Jersey	EWR	46336452	1.0	446320.0
12	Phoenix, Arizona	PHX	46288337	3.0	438891.0

13	Miami, Florida	MIA	45924466	2.0	416773.0
14	Houston, Texas	IAH	45264059	3.3	478070.0
15	Boston, Massachusetts	BOS	42522411	3.9	427176.0
16	Minneapolis, Minnesota	MSP	39555035	4.0	406076.0
17	Romulus, Michigan	DTW	36769279	4.3	396909.0
18	Fort Lauderdale, Florida	FLL	36747622	2.2	331447.0
19	Philadelphia, Pennsylvania	PHL	33018886	4.2	390321.0
20	Queens, New York	LGA	31084894	3.3	373356.0
21	Linthicum Heights, Maryland	BWI	26993896	0.6	262597.0
22	Salt Lake City, Utah	SLC	26808014	4.9	344715.0
23	San Diego, California	SAN	25216947	4.0	231354.0
24	Dulles, Virginia	IAD	24817677	3.1	285042.0
25	Arlington, Virginia	DCA	23945527	1.8	292682.0
26	Tampa, Florida	TPA	22497953	5.7	217360.0
27	Honolulu, Hawaii	HNL	21870691	4.2	326832.0
28	Chicago, Illinois	MDW	20844860	5.4	232084.0
29	Portland, Oregon	PDX	19891365	0.0	238384.0
30	Nashville, Tennessee	BNA	18273434	14.2	NaN
31	Austin, Texas	AUS	17343729	9.6	209726.0
32	Dallas, Texas	DAL	16780158	3.4	231879.0
33	St Louis, Missouri	STL	15878527	1.6	193925.0
34	San Jose, California	SJC	15650444	9.3	207111.0

	% chg.2018/19
0	1.0
1	2.3
2	1.8
3	7.9
4	6.1
5	0.1
6	2.5
7	2.8
8	2.4
9	2.9
10	5.1
11	1.6
12	1.1
13	0.2
14	2.4
15	0.7
16	0.3
17	0.8
18	0.6
19	2.8
20	0.4
21	1.5

```
22          2.2
23          2.8
24          3.9
25          0.4
26          5.0
27         10.7
28          4.6
29          1.9
30          NaN
31          0.2
32          0.3
33          0.2
34         19.4  ,
   Rank Rank change                                  Airport name  \
   Rank Rank change                                  Airport name
0     1         NaN  Hartsfield-Jackson Atlanta International Airpo…
1     2         NaN              Los Angeles International Airport[35]
2     3         NaN                   O'Hare International Airport[36]
3     4         NaN        Dallas/Fort Worth International Airport[37]
4     5         NaN                  Denver International Airport[38]
5     6         NaN          John F. Kennedy International Airport[39]
6     7         NaN            San Francisco International Airport[40]
7     8         1.0           Seattle-Tacoma International Airport[41]
8     9         1.0               Harry Reid International Airport[42]
9    10         2.0                  Orlando International Airport[43]
10   11         1.0        Charlotte Douglas International Airport[44]
11   12         1.0           Newark Liberty International Airport[45]
12   13         1.0                    Miami International Airport[46]
13   14         1.0       Phoenix Sky Harbor International Airport[47]
14   15         NaN           George Bush Intercontinental Airport[48]
15   16         NaN  General Edward Lawrence Logan International Ai…
16   17         NaN   Minneapolis-Saint Paul International Airport[50]
17   18         1.0  Fort Lauderdale-Hollywood International Airpor…
18   19         1.0                  Detroit Metropolitan Airport[52]
19   20         NaN             Philadelphia International Airport[53]
20   21         NaN                                LaGuardia Airport[54]
21   22         NaN     Baltimore/Washington International Airport[55]
22   23         NaN           Salt Lake City International Airport[56]
23   24         2.0                San Diego International Airport[57]
24   25         NaN        Washington Dulles International Airport[58]
25   26         2.0      Ronald Reagan Washington National Airport[59]
26   27         NaN           Chicago Midway International Airport[60]
27   28         NaN                    Tampa International Airport[61]
28   29         NaN             Daniel K. Inouye International Airport
29   30         NaN                 Portland International Airport[62]
30   31         NaN                                Dallas Love Field[63]
31   32         1.0                Nashville International Airport[64]
```

```
32   33         1.0         Austin-Bergstrom International Airport[65]
33   34         2.0        St. Louis Lambert International Airport[66]
34   35         NaN  Norman Y. Mineta San Jose International Airpor…
```

	Location	IATA Code	Traffic	\
	Location	IATA Code	Passengers	% chg.2017/18
0	College Park, Georgia	ATL	107394029	3.3
1	Los Angeles, California	LAX	87534384	3.5
2	Chicago, Illinois	ORD	83245472	4.3
3	Irving, Texas	DFW	69112607	3.0
4	Denver, Colorado	DEN	64494613	5.1
5	Queens, New York	JFK	61909148	3.9
6	South San Francisco, California	SFO	57793313	3.5
7	SeaTac, Washington	SEA	49849520	6.2
8	Las Vegas, Nevada	LAS	49716584	2.5
9	Orlando, Florida	MCO	47696627	6.9
10	Charlotte, North Carolina	CLT	46444380	1.2
11	Newark, New Jersey	EWR	46065175	6.6
12	Miami, Florida	MIA	45044312	2.2
13	Phoenix, Arizona	PHX	44943686	2.3
14	Houston, Texas	IAH	43807539	7.6
15	Boston, Massachusetts	BOS	40941925	6.6
16	Minneapolis, Minnesota	MSP	38037381	0.0
17	Fort Lauderdale, Florida	FLL	35963370	10.6
18	Romulus, Michigan	DTW	35236676	1.5
19	Philadelphia, Pennsylvania	PHL	31691956	7.1
20	Queens, New York	LGA	30094074	1.8
21	Linthicum Heights, Maryland	BWI	27145831	2.9
22	Salt Lake City, Utah	SLC	25554244	5.6
23	San Diego, California	SAN	24238300	9.3
24	Dulles, Virginia	IAD	24060709	5.1
25	Arlington, Virginia	DCA	23464618	1.8
26	Chicago, Illinois	MDW	22027737	1.9
27	Tampa, Florida	TPA	21289390	8.5
28	Honolulu, Hawaii	HNL	20990932	1.1
29	Portland, Oregon	PDX	19882788	4.2
30	Dallas, Texas	DAL	16229151	3.2
31	Nashville, Tennessee	BNA	15996194	13.2
32	Austin, Texas	AUS	15819912	13.9
33	St Louis, Missouri	STL	15632586	5.9
34	San Jose, California	SJC	14319292	14.7

	Aircraft Movements	% chg.2017/18
0	895682	01.7
1	707833	01.1
2	903747	04.2

```
3     667213          02.0
4     603403          03.6
5     455529          01.6
6     470164          02.1
7     438391          05.4
8     539857          00.6
9     347672          05.1
10    550013          00.4
11    458674          04.6
12    416032          00.7
13    434252          00.8
14    466738          03.6
15    424024          05.6
16    407476          02.1
17    329662          05.4
18    393681          00.4
19    379665          02.6
20    372025          00.8
21    266569          01.9
22    337276          03.1
23    225058          07.5
24    274281          03.6
25    293827          00.2
26    243322          03.2
27    206938          05.9
28    295233          5.30
29    233993          02.2
30    231110          01.6
31    216966          05.2
32    210080          05.2
33         -             -
34    173389         011.3  ,
   Rank                                    Airport name  \
   Rank                                    Airport name
0     1   Hartsfield-Jackson Atlanta International Airport
1     2                  Los Angeles International Airport
2     3                       O'Hare International Airport
3     4            Dallas/Fort Worth International Airport
4     5              John F. Kennedy International Airport
5     6                       Denver International Airport
6     7                San Francisco International Airport
7     8                   Harry Reid International Airport
8     9               Seattle-Tacoma International Airport
9    10                        Miami International Airport
10   11            Charlotte Douglas International Airport
11   12           Phoenix Sky Harbor International Airport
12   13                      Orlando International Airport
```

13	14	George Bush Intercontinental Airport
14	15	Newark Liberty International Airport
15	16	Minneapolis-Saint Paul International Airport
16	17	General Edward Lawrence Logan International Ai…
17	18	Detroit Metropolitan Airport
18	19	Philadelphia International Airport
19	20	LaGuardia Airport
20	21	Fort Lauderdale-Hollywood International Airport
21	22	Baltimore/Washington International Airport
22	23	Ronald Reagan Washington National Airport
23	24	Salt Lake City International Airport
24	25	Chicago Midway International Airport
25	26	Washington Dulles International Airport
26	27	San Diego International Airport
27	28	Honolulu International Airport
28	29	Tampa International Airport
29	30	Portland International Airport
30	31	Dallas Love Field
31	32	St. Louis Lambert International Airport
32	33	Nashville International Airport
33	34	William P. Hobby Airport
34	35	Austin-Bergstrom International Airport
35	36	Oakland International Airport

	Location	IATA Code	Traffic	\
	Location	IATA Code	Passengers	% chg.2015/16
0	College Park, Georgia	ATL	104171935	02.6
1	Los Angeles, California	LAX	80921527	08.0
2	Chicago, Illinois	ORD	77960588	01.3
3	Irving, Texas	DFW	65670697	00.2
4	Queens, New York	JFK	59105513	03.9
5	Denver, Colorado	DEN	58266515	07.9
6	South San Francisco, California	SFO	53099282	06.1
7	Las Vegas, Nevada	LAS	47496614	04.5
8	SeaTac, Washington	SEA	45736700	08.0
9	Miami, Florida	MIA	44584603	00.5
10	Charlotte, North Carolina	CLT	44422022	01.0
11	Phoenix, Arizona	PHX	43302381	01.6
12	Orlando, Florida	MCO	41923399	08.0
13	Houston, Texas	IAH	41622594	03.3
14	Newark, New Jersey	EWR	40563285	08.2
15	Minneapolis, Minnesota	MSP	37413728	02.3
16	Boston, Massachusetts	BOS	36356917	08.5
17	Romulus, Michigan	DTW	34401254	02.9
18	Philadelphia, Pennsylvania	PHL	30155090	04.1
19	Queens, New York	LGA	29786769	04.7
20	Fort Lauderdale, Florida	FLL	29205002	08.4

21	Linthicum Heights, Maryland	BWI	25122651	05.4
22	Arlington, Virginia	DCA	23568586	02.4
23	Salt Lake City, Utah	SLC	23157445	04.5
24	Chicago, Illinois	MDW	22677859	02.1
25	Dulles, Virginia	IAD	21817340	01.5
26	San Diego, California	SAN	20725801	03.2
27	Honolulu, Hawaii	HNL	19878659	- 00.0
28	Tampa, Florida	TPA	18931922	00.6
29	Portland, Oregon	PDX	18352767	08.9
30	Dallas, Texas	DAL	15562738	07.3
31	St Louis, Missouri	STL	13959126	09.5
32	Nashville, Tennessee	BNA	12979803	011.2
33	Houston, Texas	HOU	12909075	06.1
34	Austin, Texas	AUS	12436849	04.5
35	Oakland, California	OAK	12070967	07.7

	Aircraft Movements	% chg.2015/16
0	898356	01.8
1	697138	06.3
2	867635	00.9
3	672748	01.3
4	452415	03.0
5	565503	04.5
6	450388	04.8
7	541428	02.1
8	412170	08.1
9	414234	00.3
10	545742	00.3
11	440643	00.1
12	316981	02.9
13	470780	06.4
14	435907	05.3
15	412872	02.0
16	372930	02.5
17	393427	03.7
18	394022	04.2
19	369987	02.7
20	290239	04.4
21	248585	00.9
22	295038	00.8
23	320137	02.7
24	253046	00.2
25	265743	01.5
26	197132	01.5
27	316154	01.1
28	-	-

```
29    227709          04.4
30    224193          03.7
31    190560          02.5
32    194758          05.6
33    200741          00.1
34    192032          00.4
35    222771          03.3  ,
    Rank                                      Airport name  \
0      1              John F. Kennedy International Airport
1      2                        Miami International Airport
2      3                   Los Angeles International Airport
3      4               George Bush Intercontinental Airport
4      5               Newark Liberty International Airport
5      6            Dallas/Fort Worth International Airport
6      7   Hartsfield-Jackson Atlanta International Airport
7      8                       O'Hare International Airport
8      9    Fort Lauderdale-Hollywood International Airport
9     10            Washington Dulles International Airport
10    11                San Francisco International Airport
11    12  General Edward Lawrence Logan International Ai…
12    13            Charlotte Douglas International Airport
13    14                        Denver International Airport
14    15                      Orlando International Airport
15    16                Seattle-Tacoma International Airport
16    17           Phoenix Sky Harbor International Airport
17    18                 Philadelphia International Airport
18    19          Detroit Metropolitan Wayne County Airport
19    20                    Harry Reid International Airport
20    22       Minneapolis-Saint Paul International Airport
21    38              Daniel K. Inouye International Airport

                          Location IATA Code  2021[68]  2020[69]  2019[70]
0                 Queens, New York       JFK  12466165   8219317  33432159
1                   Miami, Florida       MIA  11592445   6565834  20735658
2           Los Angeles, California       LAX   7862532   6246602  25210140
3                   Houston, Texas       IAH   6458473   3491935  10764589
4                Newark, New Jersey       EWR   6250880   3688541  14087622
5                     Irving, Texas       DFW   5852397   3268822   9103438
6             College Park, Georgia       ATL   5474264   3347184  12268779
7                 Chicago, Illinois       ORD   5148494   3481860  13412885
8           Fort Lauderdale, Florida       FLL   4016553   2839383   8524251
9                   Dulles, Virginia       IAD   3230027   1917510   7990292
10  South San Francisco, California       SFO   3139041   3210024  14357960
11             Boston, Massachusetts       BOS   2046561   1574712   7534504
12         Charlotte, North Carolina       CLT   1989704   1069001   3405907
13                  Denver, Colorado       DEN   1856124    934563   3037012
14                  Orlando, Florida       MCO   1837706   1525177   6957048
```

```
 15              SeaTac, Washington       SEA   1393603   1273179   5392147
 16                Phoenix, Arizona       PHX   1223856    750138   1958468
 17       Philadelphia, Pennsylvania       PHL    988733    682030   3847253
 18                Romulus, Michigan       DTW    966375    873744   3717775
 19                 Paradise, Nevada       LAS    738257    711614   3462627
 20           Minneapolis, Minnesota       MSP    673759    835721   3144386
 21                 Honolulu, Hawaii       HNL    167502   1052697   5207875  ,
   Rank                                      Airport name  \
   Rank                                      Airport name
 0    1                      Memphis International Airport
 1    2       Ted Stevens Anchorage International Airport
 2    3                    Louisville International Airport
 3    4                        O'Hare International Airport
 4    5                         Miami International Airport
 5    6                   Los Angeles International Airport
 6    7  Cincinnati/Northern Kentucky International Air…
 7    8                  Indianapolis International Airport
 8    9            Dallas/Fort Worth International Airport
 9   10                      Ontario International Airport

                  Location IATA code        Cargo
                  Location IATA code         lbs. % chg.2017/16
 0       Memphis, Tennessee       MEM  23949525780        00.35%
 1        Anchorage, Alaska       ANC  17337337377        02.79%
 2     Louisville, Kentucky       SDF  13403682652        04.68%
 3        Chicago, Illinois       ORD  10373559593       010.84%
 4           Miami, Florida       MIA   7963988407        00.82%
 5  Los Angeles, California       LAX   7197930264        03.85%
 6         Hebron, Kentucky       CVG   5700282994       033.32%
 7    Indianapolis, Indiana       IND   5138500318       0-3.58%
 8            Irving, Texas       DFW   4155362297        07.65%
 9      Ontario, California       ONT   3522510318       015.81%  ,
   .mw-parser-output .navbar{display:inline;font-size:88%;font-
weight:normal}.mw-parser-output .navbar-collapse{float:left;text-align:left}.mw-
parser-output .navbar-boxtext{word-spacing:0}.mw-parser-output .navbar
ul{display:inline-block;white-space:nowrap;line-height:inherit}.mw-parser-output
.navbar-brackets::before{margin-right:-0.125em;content:"[ "}.mw-parser-output
.navbar-brackets::after{margin-left:-0.125em;content:" ]"}.mw-parser-output
.navbar li{word-spacing:-0.125em}.mw-parser-output .navbar a>span,.mw-parser-
output .navbar a>abbr{text-decoration:inherit}.mw-parser-output .navbar-mini
abbr{font-variant:small-caps;border-bottom:none;text-
decoration:none;cursor:inherit}.mw-parser-output .navbar-ct-full{font-
size:114%;margin:0 7em}.mw-parser-output .navbar-ct-mini{font-size:114%;margin:0
4em}vteMajor airports in the United States  \
 0  Atlanta (Hartsfield-Jackson - ATL) Baltimore (…
 1                                      Statistics
```

```
   .mw-parser-output .navbar{display:inline;font-size:88%;font-
weight:normal}.mw-parser-output .navbar-collapse{float:left;text-align:left}.mw-
parser-output .navbar-boxtext{word-spacing:0}.mw-parser-output .navbar
ul{display:inline-block;white-space:nowrap;line-height:inherit}.mw-parser-output
.navbar-brackets::before{margin-right:-0.125em;content:"[ "}.mw-parser-output
.navbar-brackets::after{margin-left:-0.125em;content:" ]"}.mw-parser-output
.navbar li{word-spacing:-0.125em}.mw-parser-output .navbar a>span,.mw-parser-
output .navbar a>abbr{text-decoration:inherit}.mw-parser-output .navbar-mini
abbr{font-variant:small-caps;border-bottom:none;text-
decoration:none;cursor:inherit}.mw-parser-output .navbar-ct-full{font-
size:114%;margin:0 7em}.mw-parser-output .navbar-ct-mini{font-size:114%;margin:0
4em}vteMajor airports in the United States.1
 0  Atlanta (Hartsfield-Jackson - ATL) Baltimore (…
 1                                          Statistics
,
   vteList of the busiest airports in North America  \
 0                                Sovereign states
 1               Dependencies andother territories

   vteList of the busiest airports in North America.1
 0  Antigua and Barbuda Bahamas Barbados Belize Ca…
 1  Anguilla Aruba Bermuda Bonaire British Virgin …  ,
       vteLists of the busiest airports by continent  \
 0  Africa Asia Europe North America Oceania South…

     vteLists of the busiest airports by continent.1
 0  Africa Asia Europe North America Oceania South…  ,
                       vteAviation statistics  \
 0                         Airports worldwide
 1   Busiest airports by continent and country
 2                                     Africa
 3                                       Asia
 4                                     Europe
 5                              North America
 6                                    Oceania
 7                              South America
 8                                  By region
 9                                   Airlines
 10                                    Routes

                            vteAviation statistics.1
 0   Busiest airports by continent By aircraft move…
 1   Africa Morocco South Africa Asia China (exclud…
 2                               Morocco South Africa
 3   China (excluding Hong Kong and Macau) India In…
 4   Austria Belgium Bulgaria Croatia France German…
 5   Canada Dominican Republic Mexico United States…
```

```
6                              Australia New Zealand
7   Argentina Brazil Chile Colombia Ecuador Paragu…
8   Baltic Caribbean Central America Latin America…
9   World's largest airlines Airline holding compa…
10  Busiest passenger air routes General aviation …  ,
               0                                                 1
0         Africa                                Morocco South Africa
1           Asia  China (excluding Hong Kong and Macau) India In…
2         Europe  Austria Belgium Bulgaria Croatia France German…
3  North America  Canada Dominican Republic Mexico United States…
4        Oceania                               Australia New Zealand
5  South America  Argentina Brazil Chile Colombia Ecuador Paragu…
6      By region  Baltic Caribbean Central America Latin America…]
```

[41]:
```
table[0] = table[0].drop(['2021', '2013[10]', '2012[11]', '2011[12]'], axis=1)
```

[42]:
```
table[0].head()
```

[42]:
```
   Rank(2021)                          Airports (large hubs) IATACode  \
0           1  Hartsfield-Jackson Atlanta International Airport      ATL
1           2                 Los Angeles International Airport      LAX
2           3             Chicago O'Hare International Airport      ORD
3           4          Dallas/Fort Worth International Airport      DFW
4           5                      Denver International Airport      DEN

  Major cities served State   2020[3]   2019[4]   2018[5]   2017[6]   2016[7]  \
0             Atlanta    GA  20559866  53505795  51865797  50251964  50501858
1         Los Angeles    CA  18593421  35778573  32821799  31816933  31283579
2             Chicago    IL  16243216  33592945  31362941  29809097  28267394
3   Dallas/Fort Worth    TX  14606034  40871223  39873927  38593028  37589899
4              Denver    CO  14055777  42939104  42624050  41232432  39636042

    2015[8]   2014[9]
0  49340732  46604273
1  31589839  30804567
2  26280043  26000591
3  36305668  33843426
4  36351272  34314197
```

[43]:
```
table[0]['traffic_Chg19_20'] = table[0]['2020[3]'] - table[0]['2019[4]']
```

[44]:
```
table[0]['traffic_Chg18_19'] = table[0]['2019[4]'] - table[0]['2018[5]']
table[0]['hubs'] = str('large_hub')
```

[45]:
```
table[0] = table[0][['IATACode', 'traffic_Chg19_20', 'traffic_Chg18_19',␣
↪'hubs']]
table[0]
```

[45]:

	IATACode	traffic_Chg19_20	traffic_Chg18_19	hubs
0	ATL	-32945929	1639998	large_hub
1	LAX	-17185152	2956774	large_hub
2	ORD	-17349729	2230004	large_hub
3	DFW	-26265189	997296	large_hub
4	DEN	-28883327	315054	large_hub
5	JFK	-11246819	1917739	large_hub
6	SFO	-14094543	1359791	large_hub
7	SEA	-14144302	933349	large_hub
8	MCO	-11902116	810972	large_hub
9	LAS	-12635024	399391	large_hub
10	CLT	-15539351	976854	large_hub
11	EWR	-13222751	747911	large_hub
12	PHX	-22766836	415886	large_hub
13	IAH	-15175289	363161	large_hub
14	MIA	-9935245	338658	large_hub
15	BOS	-12123197	830975	large_hub
16	MSP	-20034173	-11487	large_hub
17	DTW	-11320716	706203	large_hub
18	FLL	-14663925	692856	large_hub
19	PHL	-7087602	614111	large_hub
20	LGA	-10253150	713719	large_hub
21	BWI	-7833332	-87129	large_hub
22	SLC	-6011981	610242	large_hub
23	SAN	-8010836	474468	large_hub
24	IAD	-11246485	335100	large_hub
25	DCA	-5845178	-596237	large_hub
26	TPA	-4921659	918307	large_hub
27	MDW	-8021459	262494	large_hub

[46]:
```
table[1].head()
```

[46]:

	Rank(2020)	Airports (medium hubs)	IATACode	City served	\
0	29	Daniel K. Inouye International Airport	HNL	Honolulu	
1	30	Portland International Airport	PDX	Portland	
2	31	Nashville International Airport	BNA	Nashville	
3	32	Austin-Bergstrom International Airport	AUS	Austin	
4	33	Dallas Love Field	DAL	Dallas	

	State	2020	2019	2018	2017	2016	2015[7]	\
0	HI	9893930	8408457	8134848.0	7876769.0	7554596.0	7040921.0	
1	OR	9790489	11595454	11367176.0	11506310.0	11470854.0	11242375.0	
2	TN	8498877	9797408	9940866.0	9435473.0	9071154.0	8340234.0	
3	TX	3141505	8683711	7921797.0	6973115.0	6095545.0	5797547.0	
4	TX	8069178	7069614	6937061.0	6741870.0	6285181.0	5937944.0	

	2014[1]

```
0   4522341.0
1  10057794.0
2   7878760.0
3   5219982.0
4   5800726.0
```

[47]:
```
table[1]['traffic_Chg19_20'] = table[1]['2020'] - table[1]['2019']
table[1]['traffic_Chg18_19'] = table[1]['2019'] - table[1]['2018']
table[1]['hubs'] = str('Medium_hub')
```

[48]:
```
table[1] = table[1][['IATACode', 'traffic_Chg19_20', 'traffic_Chg18_19','hubs']]
table[1]
```

[48]:

	IATACode	traffic_Chg19_20	traffic_Chg18_19	hubs
0	HNL	1485473	273609.0	Medium_hub
1	PDX	-1804965	228278.0	Medium_hub
2	BNA	-1298531	-143458.0	Medium_hub
3	AUS	-5542206	761914.0	Medium_hub
4	DAL	999564	132553.0	Medium_hub
5	STL	-2238287	410173.0	Medium_hub
6	SJC	-270221	124712.0	Medium_hub
7	HOU	1912672	424899.0	Medium_hub
8	RDU	452929	422783.0	Medium_hub
9	MSY	141501	151623.0	Medium_hub
10	OAK	1950734	556705.0	Medium_hub
11	SMF	-470933	502607.0	Medium_hub
12	MCI	-2080699	688269.0	Medium_hub
13	SNA	-1409936	-238091.0	Medium_hub
14	RSW	-719803	-175712.0	Medium_hub
15	SAT	126615	57961.0	Medium_hub
16	CLE	174943	14143.0	Medium_hub
17	IND	-338022	178553.0	Medium_hub
18	PIT	-477440	-163873.0	Medium_hub
19	SJU	-172541	45914.0	Medium_hub
20	CVG	-17062	144199.0	Medium_hub
21	CMH	-16471	117495.0	Medium_hub
22	OGG	316303	197387.0	Medium_hub
23	JAX	-7422	361383.0	Medium_hub
24	PBI	79312	-174744.0	Medium_hub
25	MKE	635944	223831.0	Medium_hub
26	BDL	604458	70942.0	Medium_hub
27	BUR	-338581	-7120.0	Medium_hub
28	ONT	-1074666	219674.0	Medium_hub
29	ANC	334018	NaN	Medium_hub
30	ABQ	182971	-1813.0	Medium_hub
31	OMA	132539	NaN	Medium_hub
32	BUF	390491	NaN	Medium_hub

```
33      CHS               284               NaN  Medium_hub
34      MEM            -67924               NaN  Medium_hub
35      RIC             44986               NaN  Medium_hub
```

[49]:
```
# Lets first merge all wikipedia table.
wiki_data = [table[0],table[1]]
wiki_data = pd.concat(wiki_data, ignore_index=True)
wiki_data
```

[49]:
```
   IATACode  traffic_Chg19_20  traffic_Chg18_19        hubs
0       ATL          -32945929         1639998.0   large_hub
1       LAX          -17185152         2956774.0   large_hub
2       ORD          -17349729         2230004.0   large_hub
3       DFW          -26265189          997296.0   large_hub
4       DEN          -28883327          315054.0   large_hub
..      …                 …                …          …
59      OMA             132539               NaN  Medium_hub
60      BUF             390491               NaN  Medium_hub
61      CHS                284               NaN  Medium_hub
62      MEM             -67924               NaN  Medium_hub
63      RIC              44986               NaN  Medium_hub

[64 rows x 4 columns]
```

[50]:
```
# Now we gather all the information that we got from wiki pedia link and the␣
 ↪data that we have.
final_df = df.merge(wiki_data, left_on ='iata_code', right_on = "IATACode")
```

[51]:
```
final_df
```

[51]:
```
            id Airline  Flight AirportFrom AirportTo  DayOfWeek  Time  Length  \
0            4      AA    2466         SFO       DFW          3    20     195
1          231      AA     526         SFO       DFW          3   360     215
2          234      AA     552         SFO       MIA          3   360     315
3          905      AA     810         SFO       ORD          3   385     255
4         1739      AA      24         SFO       JFK          3   425     325
…          …      …      …          …        …         …    …      …
363125  506267      9E    4052         DAL       MEM          4   370      90
363126  512858      9E    3704         DAL       MEM          4   705      92
363127  518247      9E    4060         DAL       MEM          4   990      90
363128  524678      9E    4052         DAL       MEM          5   370      90
363129  530841      9E    3704         DAL       MEM          5   705      92

        Delay ident  … width_ft surface  lighted  closed  IATA Founded  \
0           1  KSFO  …    200.0     ASP        1       0    AA  1926.0
1           0  KSFO  …    200.0     ASP        1       0    AA  1926.0
2           1  KSFO  …    200.0     ASP        1       0    AA  1926.0
```

```
3           0  KSFO  …    200.0     ASP         1       0    AA  1926.0
4           1  KSFO  …    200.0     ASP         1       0    AA  1926.0
…       …   …  …       …      …       …     …   …       …
363125      0  KDAL  …    150.0     CON         1       0    9E  1985.0
363126      1  KDAL  …    150.0     CON         1       0    9E  1985.0
363127      0  KDAL  …    150.0     CON         1       0    9E  1985.0
363128      1  KDAL  …    150.0     CON         1       0    9E  1985.0
363129      0  KDAL  …    150.0     CON         1       0    9E  1985.0

       IATACode  traffic_Chg19_20 traffic_Chg18_19        hubs
0           SFO         -14094543        1359791.0   large_hub
1           SFO         -14094543        1359791.0   large_hub
2           SFO         -14094543        1359791.0   large_hub
3           SFO         -14094543        1359791.0   large_hub
4           SFO         -14094543        1359791.0   large_hub
…         …                 …                …         …
363125      DAL            999564         132553.0  Medium_hub
363126      DAL            999564         132553.0  Medium_hub
363127      DAL            999564         132553.0  Medium_hub
363128      DAL            999564         132553.0  Medium_hub
363129      DAL            999564         132553.0  Medium_hub

[363130 rows x 30 columns]
```

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

[52]:
```
# Now we have the final data first we remove some column that is not useable.
final_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 363130 entries, 0 to 363129
Data columns (total 30 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   id                 363130 non-null  int64
 1   Airline            363130 non-null  object
 2   Flight             363130 non-null  int64
 3   AirportFrom        363130 non-null  object
 4   AirportTo          363130 non-null  object
 5   DayOfWeek          363130 non-null  int64
 6   Time               363130 non-null  int64
 7   Length             363130 non-null  int64
 8   Delay              363130 non-null  int64
 9   ident              363130 non-null  object
 10  type               363130 non-null  object
```

```
 11  name                363130 non-null  object
 12  latitude_deg        363130 non-null  float64
 13  longitude_deg       363130 non-null  float64
 14  elevation_ft        363130 non-null  float64
 15  scheduled_service   363130 non-null  object
 16  iata_code           363130 non-null  object
 17  airport_ref         363130 non-null  int64
 18  airport_ident       363130 non-null  object
 19  length_ft           363130 non-null  float64
 20  width_ft            363130 non-null  float64
 21  surface             363130 non-null  object
 22  lighted             363130 non-null  int64
 23  closed              363130 non-null  int64
 24  IATA                363130 non-null  object
 25  Founded             363130 non-null  float64
 26  IATACode            363130 non-null  object
 27  traffic_Chg19_20    363130 non-null  int64
 28  traffic_Chg18_19    351555 non-null  float64
 29  hubs                363130 non-null  object
dtypes: float64(7), int64(10), object(13)
memory usage: 85.9+ MB
```

[53]:
```
final_df.
 ↪drop(['id','AirportFrom','airport_ident','iata_code','AirportTo','surface',␣
 ↪'ident',
                            'IATA', 'IATACode','name'], axis=1,inplace=True)
```

[54]:
```
final_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 363130 entries, 0 to 363129
Data columns (total 20 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Airline             363130 non-null  object
 1   Flight              363130 non-null  int64
 2   DayOfWeek           363130 non-null  int64
 3   Time                363130 non-null  int64
 4   Length              363130 non-null  int64
 5   Delay               363130 non-null  int64
 6   type                363130 non-null  object
 7   latitude_deg        363130 non-null  float64
 8   longitude_deg       363130 non-null  float64
 9   elevation_ft        363130 non-null  float64
 10  scheduled_service   363130 non-null  object
 11  airport_ref         363130 non-null  int64
 12  length_ft           363130 non-null  float64
 13  width_ft            363130 non-null  float64
```

```
 14  lighted            363130 non-null  int64
 15  closed             363130 non-null  int64
 16  Founded            363130 non-null  float64
 17  traffic_Chg19_20   363130 non-null  int64
 18  traffic_Chg18_19   351555 non-null  float64
 19  hubs               363130 non-null  object
dtypes: float64(7), int64(9), object(4)
memory usage: 58.2+ MB
```

[55]:
```
# Now lets check the null value and treat them.
final_df.isnull().sum()
```

[55]:
```
Airline                0
Flight                 0
DayOfWeek              0
Time                   0
Length                 0
Delay                  0
type                   0
latitude_deg           0
longitude_deg          0
elevation_ft           0
scheduled_service      0
airport_ref            0
length_ft              0
width_ft               0
lighted                0
closed                 0
Founded                0
traffic_Chg19_20       0
traffic_Chg18_19     11575
hubs                   0
dtype: int64
```

[62]:
```
final_df
```

[62]:

	Airline	Flight	DayOfWeek	Time	Length	Delay	type
0	AA	2466	3	20	195	1	large_airport
1	AA	526	3	360	215	0	large_airport
2	AA	552	3	360	315	1	large_airport
3	AA	810	3	385	255	0	large_airport
4	AA	24	3	425	325	1	large_airport
…	…	…	…	…	…	…	…
363125	9E	4052	4	370	90	0	medium_airport
363126	9E	3704	4	705	92	1	medium_airport
363127	9E	4060	4	990	90	0	medium_airport
363128	9E	4052	5	370	90	1	medium_airport

```
363129      9E    3704          5   705      92      0  medium_airport

        latitude_deg  longitude_deg  elevation_ft scheduled_service  \
0          37.618999    -122.375000          13.0               yes
1          37.618999    -122.375000          13.0               yes
2          37.618999    -122.375000          13.0               yes
3          37.618999    -122.375000          13.0               yes
4          37.618999    -122.375000          13.0               yes
…                …             …             …                …
363125     32.847099     -96.851799         487.0               yes
363126     32.847099     -96.851799         487.0               yes
363127     32.847099     -96.851799         487.0               yes
363128     32.847099     -96.851799         487.0               yes
363129     32.847099     -96.851799         487.0               yes

        airport_ref  length_ft  width_ft  lighted  closed  Founded  \
0              3878     7500.0     200.0        1       0   1926.0
1              3878     7500.0     200.0        1       0   1926.0
2              3878     7500.0     200.0        1       0   1926.0
3              3878     7500.0     200.0        1       0   1926.0
4              3878     7500.0     200.0        1       0   1926.0
…               …        …       …      …     …      …
363125         3479     7752.0     150.0        1       0   1985.0
363126         3479     7752.0     150.0        1       0   1985.0
363127         3479     7752.0     150.0        1       0   1985.0
363128         3479     7752.0     150.0        1       0   1985.0
363129         3479     7752.0     150.0        1       0   1985.0

        traffic_Chg19_20  traffic_Chg18_19        hubs
0               -14094543         1359791.0   large_hub
1               -14094543         1359791.0   large_hub
2               -14094543         1359791.0   large_hub
3               -14094543         1359791.0   large_hub
4               -14094543         1359791.0   large_hub
…                   …               …          …
363125             999564          132553.0  Medium_hub
363126             999564          132553.0  Medium_hub
363127             999564          132553.0  Medium_hub
363128             999564          132553.0  Medium_hub
363129             999564          132553.0  Medium_hub

[351555 rows x 20 columns]
```

Only one `traffic_Chg18_19` column contain the null value which is 3.18% of total data.

So we will drop that rows of null value because we have plenty of data.

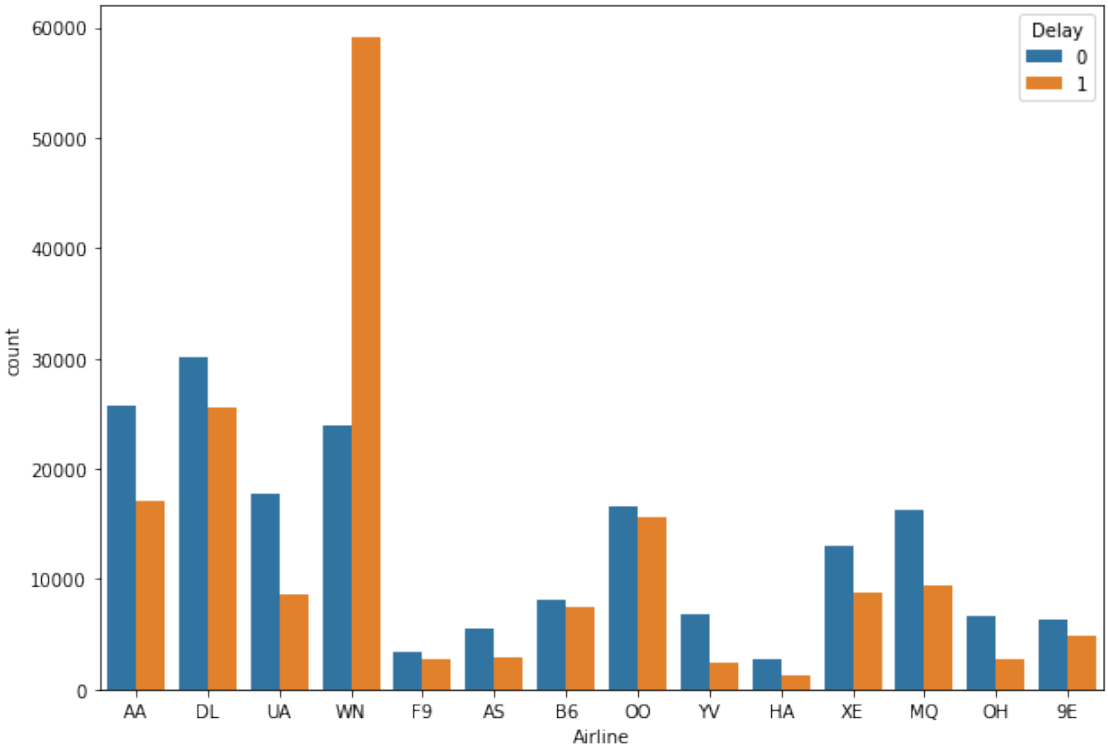
[58]:
```
final_df.dropna(axis=0,inplace=True)
```

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

1.3.1 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.

[60]:
```
plt.figure(figsize=(10,7))
sns.countplot(final_df['Airline'], hue= final_df['Delay'])
```

[60]: <AxesSubplot:xlabel='Airline', ylabel='count'>

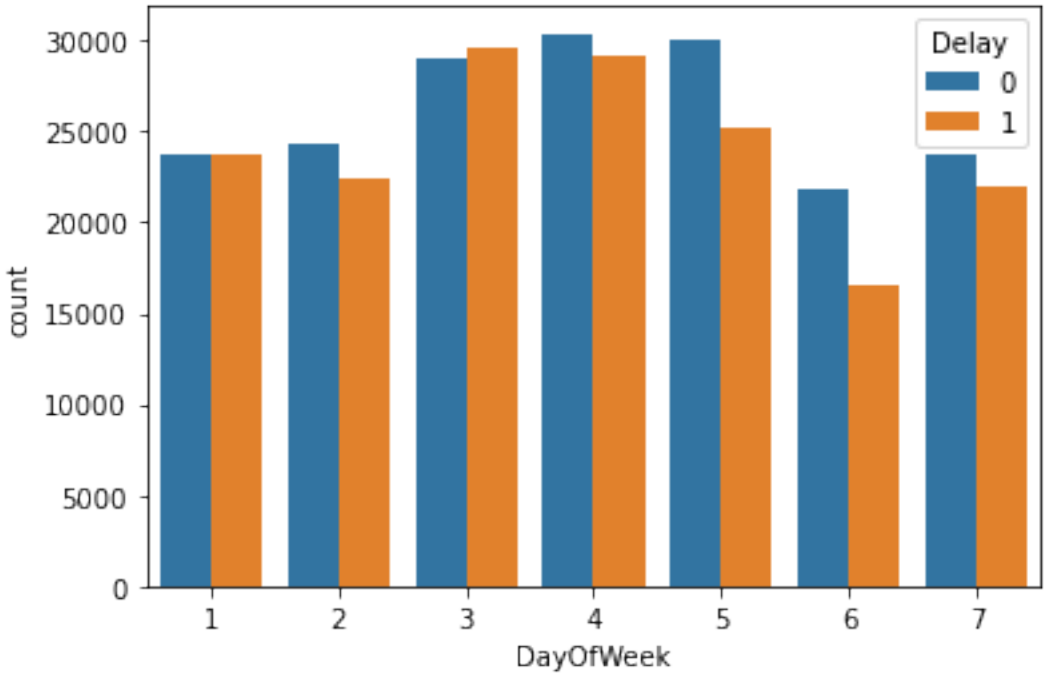


Airline code WN represent the southwest airlines. The graph clear show that 70% of flight of south west airline is delayed

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

[61]:
```
sns.countplot(final_df['DayOfWeek'],hue= final_df['Delay'])
```

[61]:
```
<AxesSubplot:xlabel='DayOfWeek', ylabel='count'>
```



We can from the above graph that on 6th day of the week we have least delayed flight.

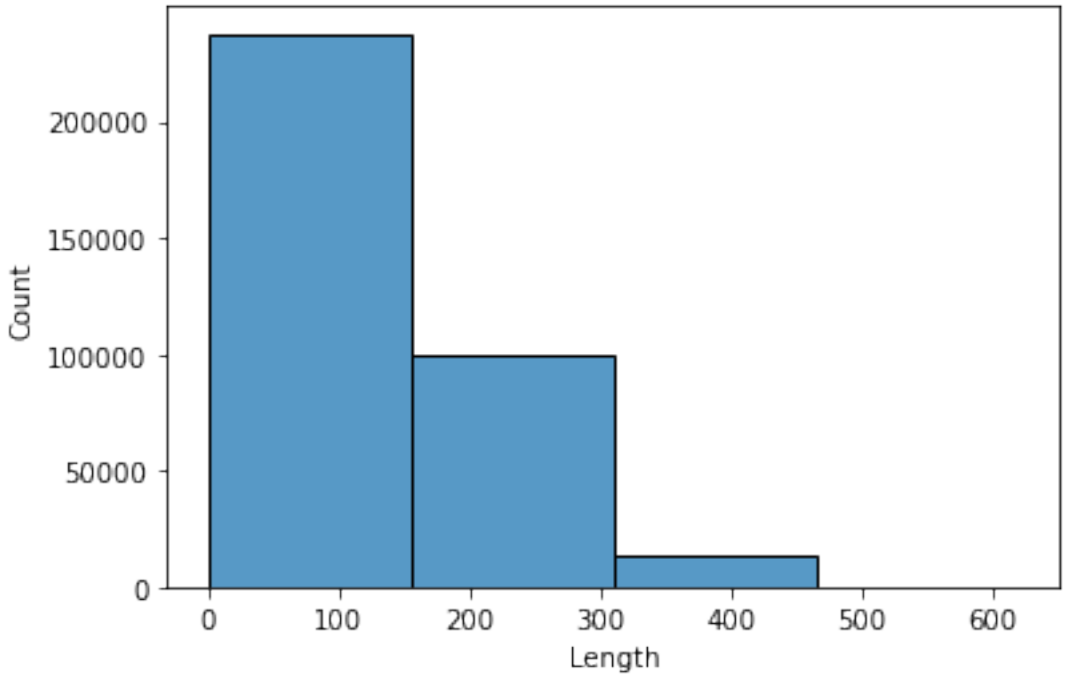
1.3.3 c. Which airlines should be recommended for short-, medium-, and long-distance travel?

[69]:
```
final_df['Length'].max()
```

[69]:
```
620
```

[78]:
```
sns.histplot(final_df['Length'], bins = 4)
```

[78]:
```
<AxesSubplot:xlabel='Length', ylabel='Count'>
```



[80]:
```
final_df['Airline'][final_df['Length']<180].value_counts()
```

[80]:
```
WN    70485
DL    38483
OO    30767
MQ    24954
AA    24370
XE    20807
UA    15534
9E    11145
B6     9766
YV     9152
OH     8989
AS     5386
F9     5105
HA     3034
Name: Airline, dtype: int64
```

The above airlines are recommended for Short distance as flight last anywhere from 30 minutes to 3 hours.

[86]:
```
final_df['Airline'][(final_df['Length']>180) & (final_df['Length']<360)].
↪value_counts()
```

[86]:
```
DL    16180
AA    15959
WN    11333
UA     9380
B6     4847
AS     2354
OO     1473
F9     1075
XE      911
HA      751
MQ      512
OH      423
YV      185
9E       47
Name: Airline, dtype: int64
```

The above airlines are recommended for Medium distance as flight last anywhere from 3 hours to 6 hours.

[87]:
```
final_df['Airline'][final_df['Length']>360].value_counts()
```

[87]:
```
UA    1304
AA    1081
DL     842
B6     822
AS     496
HA     252
WN      52
Name: Airline, dtype: int64
```

The above airlines are recommended for Long distance as flights are extend beyond 6 hours.

d. Do you notice any patterns in the departure times of long-duration flights?

[114]:
```
final_df[final_df['Length']>360].describe().T
```

[114]:

	count	mean	std	min	\
Flight	4849.0	5.563933e+02	7.537248e+02	1.000000e+00	
DayOfWeek	4849.0	4.004125e+00	1.925699e+00	1.000000e+00	
Time	4849.0	8.292021e+02	2.868094e+02	1.000000e+02	
Length	4849.0	3.959827e+02	4.084421e+01	3.610000e+02	
Delay	4849.0	4.188492e-01	4.934215e-01	0.000000e+00	
latitude_deg	4849.0	3.873399e+01	6.170430e+00	1.843940e+01	
longitude_deg	4849.0	-8.772557e+01	2.456466e+01	-1.579242e+02	
elevation_ft	4849.0	2.393364e+02	7.504192e+02	8.000000e+00	
airport_ref	4849.0	3.726874e+03	4.652847e+02	3.384000e+03	
length_ft	4849.0	1.004975e+04	2.111061e+03	4.892000e+03	
width_ft	4849.0	1.718602e+02	2.509397e+01	1.000000e+02	

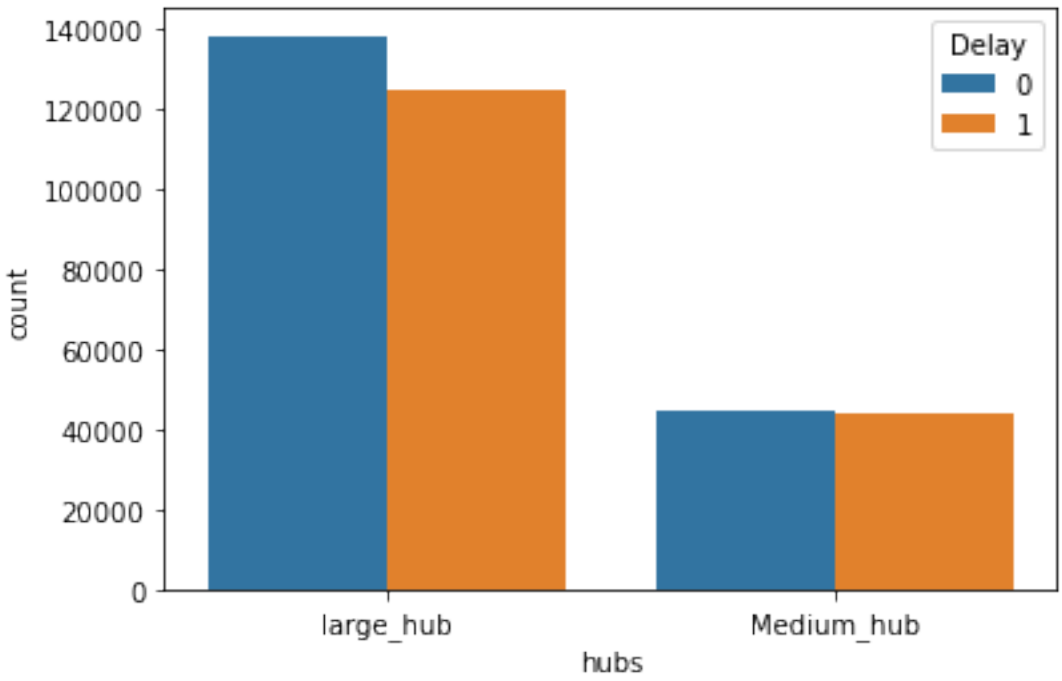
```
lighted           4849.0  9.509177e-01  2.160621e-01  0.000000e+00
closed            4849.0  0.000000e+00  0.000000e+00  0.000000e+00
Founded           4849.0  1.939067e+03  2.703800e+01  1.924000e+03
traffic_Chg19_20  4849.0 -1.167185e+07  5.257852e+06 -3.294593e+07
traffic_Chg18_19  4849.0  1.297978e+06  7.706599e+05 -8.712900e+04

                         25%           50%           75%           max
Flight            5.900000e+01  2.010000e+02  8.460000e+02  3.760000e+03
DayOfWeek         2.000000e+00  4.000000e+00  6.000000e+00  7.000000e+00
Time              5.500000e+02  8.850000e+02  1.080000e+03  1.435000e+03
Length            3.750000e+02  3.850000e+02  4.000000e+02  6.200000e+02
Delay             0.000000e+00  0.000000e+00  1.000000e+00  1.000000e+00
latitude_deg      3.894450e+01  4.063945e+01  4.078575e+01  4.744916e+01
longitude_deg    -9.703800e+01 -7.377932e+01 -7.377932e+01 -6.600180e+01
elevation_ft      1.300000e+01  1.800000e+01  1.250000e+02  5.431000e+03
airport_ref       3.602000e+03  3.622000e+03  3.632000e+03  6.384000e+03
length_ft         7.861000e+03  1.100000e+04  1.207900e+04  1.207900e+04
width_ft          1.500000e+02  1.500000e+02  2.000000e+02  2.000000e+02
lighted           1.000000e+00  1.000000e+00  1.000000e+00  1.000000e+00
closed            0.000000e+00  0.000000e+00  0.000000e+00  0.000000e+00
Founded           1.926000e+03  1.926000e+03  1.932000e+03  1.998000e+03
traffic_Chg19_20 -1.212320e+07 -1.124682e+07 -1.124682e+07  1.485473e+06
traffic_Chg18_19  7.479110e+05  9.972960e+05  1.917739e+06  2.956774e+06
```

[ ]:

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

[105]:
```
sns.countplot(final_df['hubs'], hue = final_df['Delay'])
```

[105]:
```
<AxesSubplot:xlabel='hubs', ylabel='count'>
```



From the large hubs its clear approx 120000 filght is delayed but from the medium hubs aprrox 40000 is delayed.

1.3.5 5. Use hypothesis testing strategies to discover:

a. If the airport's altitude has anything to do with flight delays for incoming and departing flights

[115]:
```
from scipy.stats import chi2_contingency
table = [final_df['latitude_deg'],final_df['Delay']]
stat, p, dof, expected = chi2_contingency(table)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

```
stat=186602.569, p=1.000
Probably independent
```

So its clear from the above hypothesis testing that altitude is nothing to do with the flight delay

b. If the number of runways at an airport affects flight delays

[117]:
```
from scipy.stats import chi2_contingency
table = [final_df['airport_ref'],final_df['Delay']]
stat, p, dof, expected = chi2_contingency(table)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

```
stat=192200.911, p=1.000
Probably independent
```

So its clear from the above hypothesis testing that no of runway is nothing to do with the flight delay

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

[118]:
```
from scipy.stats import spearmanr
data1 = final_df['Length']
data2 = final_df['Delay']
stat, p = spearmanr(data1, data2)
print('stat=%.3f, p=%.3f' % (stat, p))
if p > 0.05:
    print('Probably independent')
else:
    print('Probably dependent')
```

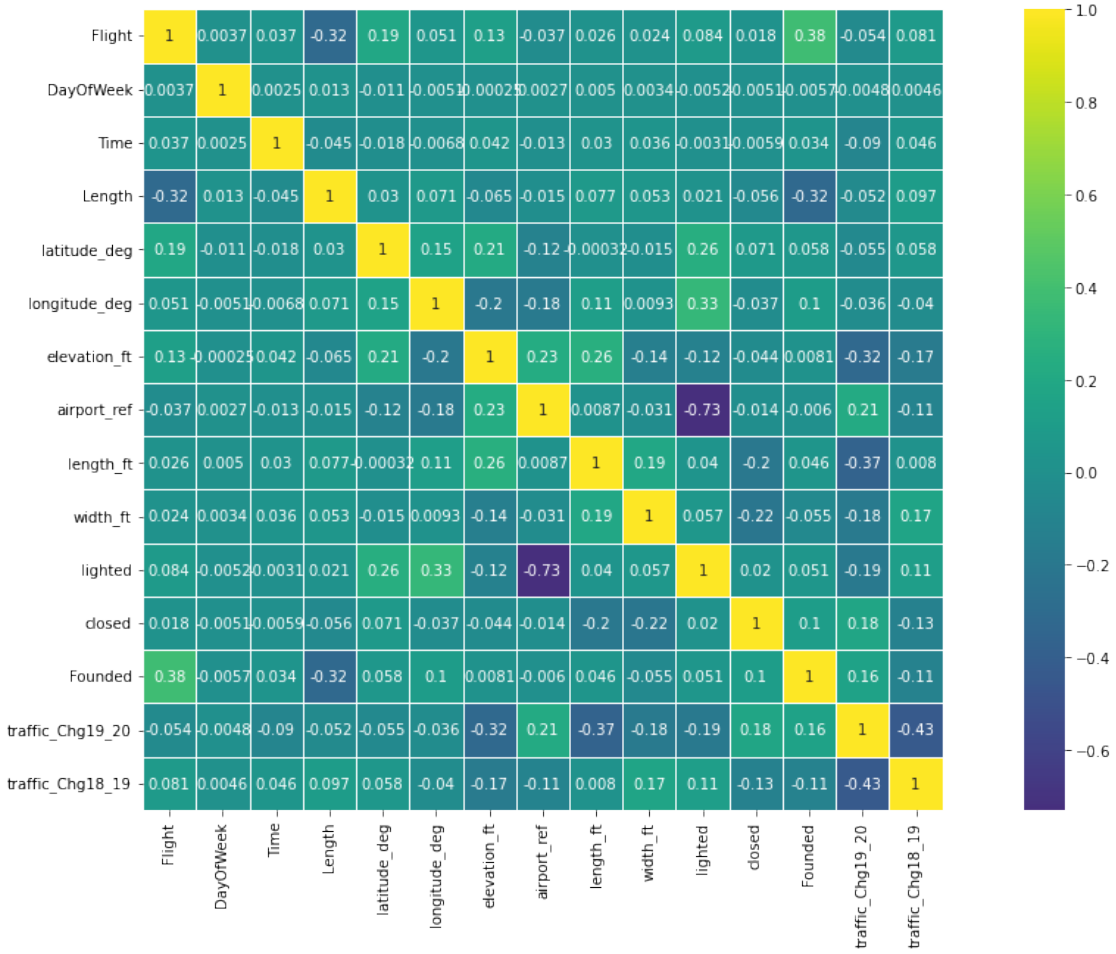
```
stat=-0.002, p=0.179
Probably independent
```

Both the variable are independent so that length of the flight is not affecting directly the delay.

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

[121]:
```
predictor = final_df.drop(['Delay'], axis=1)
corr = predictor.corr()
```

[126]:
```
plt.figure(figsize=(20,10))
sns.heatmap(corr, center=0, linewidths=.5, square = True , annot = True ,␣
 ↪cmap='viridis')
plt.show()
```



2 Project Task: Week 1

2.1 Machine learning

2.1.1 1. Use OneHotEncoder and OrdinalEncoder to deal with categorical variables

[127]:
```
# Before applying the one hot encodding or the label encoding first we check␣
↪all feature data type.
final_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 351555 entries, 0 to 363129
Data columns (total 20 columns):
 #   Column              Non-Null Count   Dtype
---  ------              --------------   -----
 0   Airline             351555 non-null  object
```

```
 1   Flight             351555 non-null  int64
 2   DayOfWeek          351555 non-null  int64
 3   Time               351555 non-null  int64
 4   Length             351555 non-null  int64
 5   Delay              351555 non-null  int64
 6   type               351555 non-null  object
 7   latitude_deg       351555 non-null  float64
 8   longitude_deg      351555 non-null  float64
 9   elevation_ft       351555 non-null  float64
 10  scheduled_service  351555 non-null  object
 11  airport_ref        351555 non-null  int64
 12  length_ft          351555 non-null  float64
 13  width_ft           351555 non-null  float64
 14  lighted            351555 non-null  int64
 15  closed             351555 non-null  int64
 16  Founded            351555 non-null  float64
 17  traffic_Chg19_20   351555 non-null  int64
 18  traffic_Chg18_19   351555 non-null  float64
 19  hubs               351555 non-null  object
dtypes: float64(7), int64(9), object(4)
memory usage: 56.3+ MB
```

[128]:
```
final_df['Airline'].value_counts()
```

[128]:
```
WN    82903
DL    55724
AA    42841
OO    32315
UA    26303
MQ    25698
XE    21733
B6    15497
9E    11192
OH     9440
YV     9337
AS     8355
F9     6180
HA     4037
Name: Airline, dtype: int64
```

[129]:
```
final_df['type'].value_counts()
```

[129]:
```
large_airport     334982
medium_airport     16573
Name: type, dtype: int64
```

[130]:
```
final_df['scheduled_service'].value_counts()
```

[130]:
```
yes    351555
Name: scheduled_service, dtype: int64
```

[131]:
```
final_df['hubs'].value_counts()
```

[131]:
```
large_hub     262540
Medium_hub     89015
Name: hubs, dtype: int64
```

The scheduled_service column throught has same value so it will not help in prediction so lets remove it and other three object column we will change through label encoder.

[132]:
```
final_df = final_df.drop(['scheduled_service'], axis=1)
```

[133]:
```
# Now using the ordinal encoder.
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
```

[134]:
```
final_df['Airline'] = le.fit_transform(final_df['Airline'])
final_df['type'] = le.fit_transform(final_df['type'])
final_df['hubs'] = le.fit_transform(final_df['hubs'])
```

[135]:
```
final_df.head()
```

[135]:

	Airline	Flight	DayOfWeek	Time	Length	Delay	type	latitude_deg	\
0	1	2466	3	20	195	1	0	37.618999	
1	1	526	3	360	215	0	0	37.618999	
2	1	552	3	360	315	1	0	37.618999	
3	1	810	3	385	255	0	0	37.618999	
4	1	24	3	425	325	1	0	37.618999	

	longitude_deg	elevation_ft	airport_ref	length_ft	width_ft	lighted	\
0	-122.375	13.0	3878	7500.0	200.0	1	
1	-122.375	13.0	3878	7500.0	200.0	1	
2	-122.375	13.0	3878	7500.0	200.0	1	
3	-122.375	13.0	3878	7500.0	200.0	1	
4	-122.375	13.0	3878	7500.0	200.0	1	

	closed	Founded	traffic_Chg19_20	traffic_Chg18_19	hubs
0	0	1926.0	-14094543	1359791.0	1
1	0	1926.0	-14094543	1359791.0	1
2	0	1926.0	-14094543	1359791.0	1
3	0	1926.0	-14094543	1359791.0	1
4	0	1926.0	-14094543	1359791.0	1

2.1.2 2. Perform the following model building steps:

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

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 crossvalidation #### 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

[136]:
```
# Lets first seperate the features and the target.
x = final_df.drop(['Delay'], axis= 1)
y = final_df["Delay"]
```

[137]:
```
from sklearn import preprocessing
scaler = preprocessing.MinMaxScaler()
x = scaler.fit_transform(x)
```

[138]:
```
# First Split the data into the training and testing set before performing the␣
 ↪further operation.
from sklearn.model_selection import train_test_split
```

[139]:
```
x_train, x_test, y_train, y_test = train_test_split(x, y, train_size=0.9,␣
 ↪random_state=10)
```

LogisticRegression

[140]:
```
# lets apply the logistic regression with the randomsearchcv hypermeter tunning.
from sklearn.linear_model import LogisticRegression
lr = LogisticRegression()
```

[141]:
```
from sklearn.model_selection import RandomizedSearchCV
```

[142]:
```
params = {"penalty": ["l1","l2"],
          'solver': ['newton-cg', 'liblinear']}

# Cross Validation
folds = 5

rscv = RandomizedSearchCV(estimator = lr,
                          param_distributions = params,
                          scoring = "accuracy",
                          verbose = 1,
                          cv= folds)
```

```
rscv.fit(x_train, y_train)
```

```
Fitting 5 folds for each of 4 candidates, totalling 20 fits
```

[142]:
```
RandomizedSearchCV(cv=5, estimator=LogisticRegression(),
                   param_distributions={'penalty': ['l1', 'l2'],
                                        'solver': ['newton-cg', 'liblinear']},
                   scoring='accuracy', verbose=1)
```

[143]:
```
print(rscv.best_params_)
print(rscv.best_score_)
```

```
{'solver': 'liblinear', 'penalty': 'l1'}
0.5921478987792488
```

[144]:
```
lr = LogisticRegression(penalty= 'l2', solver= 'newton-cg')
lr.fit(x_train,y_train).score(x_train,y_train)
```

[144]:
```
0.5923280414919136
```

[145]:
```
lr.score(x_test, y_test)
```

[145]:
```
0.593013994766185
```

DecisionTreeClassifier

[146]:
```
from sklearn.tree import DecisionTreeClassifier

dt = DecisionTreeClassifier()

params = {'criterion': ["gini", "entropy"],
          'min_samples_leaf' : [2,3,4,5,6,7,8,9],
         "max_depth": [2,3,4,5,6,7,8,9]}

rscv = RandomizedSearchCV(estimator = dt,
                          param_distributions= params,
                          scoring = "accuracy",
                          cv= 5,
                          verbose=1)
rscv.fit(x_train, y_train)
```

```
Fitting 5 folds for each of 10 candidates, totalling 50 fits
```

[146]:
```
RandomizedSearchCV(cv=5, estimator=DecisionTreeClassifier(),
                   param_distributions={'criterion': ['gini', 'entropy'],
                                        'max_depth': [2, 3, 4, 5, 6, 7, 8, 9],
                                        'min_samples_leaf': [2, 3, 4, 5, 6, 7,
                                                             8, 9]},
```

```
                    scoring='accuracy', verbose=1)
```

[147]:
```
print(rscv.best_params_)
print(rscv.best_score_)
```

```
{'min_samples_leaf': 6, 'max_depth': 9, 'criterion': 'entropy'}
0.6469110137916109
```

[148]:
```
dtc = DecisionTreeClassifier(max_depth= 9, criterion=␣
 ↪'entropy',min_samples_leaf= 6)
dtc.fit(x_train, y_train).score(x_train, y_train)
```

[148]: 0.6539464410443775

[149]:
```
dtc.score(x_test, y_test)
```

[149]: 0.649049948799636

After seeing the result its clear decision tree has good accuracy.

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

[150]:
```
from xgboost import XGBClassifier

# Create the parameter grid: gbm_param_grid
gbm_param_grid = {
                  'n_estimators': range(8, 20),
                  'max_depth': range(6, 10),
                  'learning_rate': [.4, .45, .5, .55, .6],
                  'colsample_bytree': [.6, .7, .8, .9, 1]
                   }

# Instantiate the regressor: gbm
gbm = XGBClassifier()

# Perform random search: grid_mse
xgb_random = RandomizedSearchCV(param_distributions=gbm_param_grid,
                                    estimator = gbm, scoring = "accuracy",
                                    verbose = 1, n_iter = 50, cv = 3)

# Fit randomized_mse to the data
xgb_random.fit(x_train, y_train)

# Print the best parameters and lowest RMSE
print("Best parameters found: ", xgb_random.best_params_)
```

```
print("Best accuracy found: ", xgb_random.best_score_)
```

```
Fitting 3 folds for each of 50 candidates, totalling 150 fits
Best parameters found:  {'n_estimators': 16, 'max_depth': 8, 'learning_rate':
0.5, 'colsample_bytree': 0.9}
Best accuracy found:  0.6608301551327947
```

[151]:
```
xgb = XGBClassifier(n_estimators=14, max_depth=9, learning_rate=0.45,␣
 ↪colsample_bytree=0.9)
xgb.fit(x_train,y_train).score(x_train,y_train)
```

[151]:
```
0.6860830786443699
```

[152]:
```
# Now lets compare the all method.
print(lr.score(x_test, y_test))
print(dtc.score(x_test, y_test))
print(xgb.score(x_test, y_test))
```

```
0.593013994766185
0.649049948799636
0.6630447149846399
```

After comparing the accuracy of the diffrent model the best result we getting from the XGBclassifier.