

In [14]:

```
import pandas as pd
```

```
data1 = pd.read_csv('flights.csv', sep = ',')
```

```
data1.info()
```

C:\Users\chala\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:2785: DtypeWarning: Columns (7,8) have mixed types. Specify dtype option on import or set low_memory=False.

interactivity=interactivity, compiler=compiler, result=result)

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 5819079 entries, 0 to 5819078
```

```
Data columns (total 31 columns):
```

```
YEAR                int64
```

```
MONTH              int64
```

```
DAY               int64
```

```
DAY_OF_WEEK       int64
```

```
AIRLINE           object
```

```
FLIGHT_NUMBER     int64
```

```
TAIL_NUMBER       object
```

```
ORIGIN_AIRPORT    object
```

```
DESTINATION_AIRPORT object
```

```
SCHEDULED_DEPARTURE int64
```

```
DEPARTURE_TIME    float64
```

```
DEPARTURE_DELAY   float64
```

```
TAXI_OUT          float64
```

```
WHEELS_OFF        float64
```

```
SCHEDULED_TIME    float64
```

```
ELAPSED_TIME      float64
```

```
AIR_TIME          float64
```

```
DISTANCE          int64
```

```
WHEELS_ON         float64
```

```
TAXI_IN           float64
```

```
SCHEDULED_ARRIVAL int64
```

```
ARRIVAL_TIME      float64
```

```
ARRIVAL_DELAY     float64
```

```
DIVERTED          int64
```

```
CANCELLED         int64
```

```
CANCELLATION_REASON object
```

```
AIR_SYSTEM_DELAY  float64
```

```
SECURITY_DELAY    float64
```

```
AIRLINE_DELAY     float64
```

```
LATE_AIRCRAFT_DELAY float64
```

```
WEATHER_DELAY     float64
```

```
dtypes: float64(16), int64(10), object(5)
```

```
memory usage: 1.3+ GB
```

In [15]:

```
import pandas as pd
data2 = pd.read_csv('airlines.csv',sep =',')
data2.info()
```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 2 columns):
IATA_CODE 14 non-null object
AIRLINE 14 non-null object
dtypes: object(2)
memory usage: 304.0+ bytes

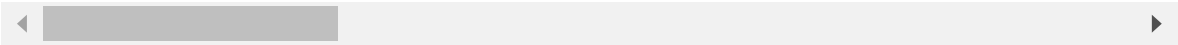
In [7]:

```
data1.head()
```

Out[7]:

	YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE	FLIGHT_NUMBER	TAIL_NUMBER	ORIGIN
0	2015	1	1	4	AS	98	N407AS	
1	2015	1	1	4	AA	2336	N3KUAA	
2	2015	1	1	4	US	840	N171US	
3	2015	1	1	4	AA	258	N3HYAA	
4	2015	1	1	4	AS	135	N527AS	

5 rows × 31 columns



In [8]:

```
data2.head()
```

Out[8]:

	IATA_CODE	AIRLINE
0	UA	United Air Lines Inc.
1	AA	American Airlines Inc.
2	US	US Airways Inc.
3	F9	Frontier Airlines Inc.
4	B6	JetBlue Airways

In [9]:

```
data1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5819079 entries, 0 to 5819078
Data columns (total 31 columns):
YEAR                int64
MONTH               int64
DAY                 int64
DAY_OF_WEEK         int64
AIRLINE             object
FLIGHT_NUMBER       int64
TAIL_NUMBER         object
ORIGIN_AIRPORT      object
DESTINATION_AIRPORT object
SCHEDULED_DEPARTURE int64
DEPARTURE_TIME      float64
DEPARTURE_DELAY     float64
TAXI_OUT            float64
WHEELS_OFF          float64
SCHEDULED_TIME      float64
ELAPSED_TIME        float64
AIR_TIME            float64
DISTANCE            int64
WHEELS_ON           float64
TAXI_IN             float64
SCHEDULED_ARRIVAL   int64
ARRIVAL_TIME        float64
ARRIVAL_DELAY       float64
DIVERTED            int64
CANCELLED           int64
CANCELLATION_REASON object
AIR_SYSTEM_DELAY    float64
SECURITY_DELAY      float64
AIRLINE_DELAY       float64
LATE_AIRCRAFT_DELAY float64
WEATHER_DELAY       float64
dtypes: float64(16), int64(10), object(5)
memory usage: 1.3+ GB
```

In [10]:

```
data2.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 14 entries, 0 to 13
Data columns (total 2 columns):
IATA_CODE    14 non-null object
AIRLINE      14 non-null object
dtypes: object(2)
memory usage: 304.0+ bytes
```

In [11]:

```
print("columns:" + str(data1.shape[1]))
```

columns:31

In [12]:

```
print("rows:"+str(data1.shape[0]))
```

rows:5819079

In [13]:

```
print("columns:"+str(data2.shape[1]))
```

columns:2

In [14]:

```
print("rows:"+str(data2.shape[0]))
```

rows:14

In [15]:

```
data1[(data1.CANCELLED == 1)].shape
```

Out[15]:

(89884, 31)

In [16]:

```
data1.groupby('AIRLINE')['CANCELLED'].count()
```

Out[16]:

AIRLINE

AA 725984

AS 172521

B6 267048

DL 875881

EV 571977

F9 90836

HA 76272

MQ 294632

NK 117379

OO 588353

UA 515723

US 198715

VX 61903

WN 1261855

Name: CANCELLED, dtype: int64

In [6]:

```
data3 = data1.groupby('AIRLINE')['CANCELLED'].count()
```

In [17]:

```
data1[['DEPARTURE_DELAY','ARRIVAL_DELAY']].describe()
```

Out[17]:

	DEPARTURE_DELAY	ARRIVAL_DELAY
count	5.732926e+06	5.714008e+06
mean	9.370158e+00	4.407057e+00
std	3.708094e+01	3.927130e+01
min	-8.200000e+01	-8.700000e+01
25%	-5.000000e+00	-1.300000e+01
50%	-2.000000e+00	-5.000000e+00
75%	7.000000e+00	8.000000e+00
max	1.988000e+03	1.971000e+03

In [4]:

```
nData = data1.dropna(subset=['DEPARTURE_DELAY','ARRIVAL_DELAY'])  
nData.shape
```

Out[4]:

(5714008, 31)

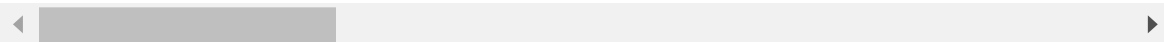
In [23]:

```
mdata = pd.merge(nData,data2, how='outer',
                  left_on='AIRLINE', right_on='IATA_CODE')
mdata = mdata.drop('IATA_CODE',1)
mdata.head()
```

Out[23]:

	YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE_x	FLIGHT_NUMBER	TAIL_NUMBER	ORIG
0	2015	1	1	4	AS	98	N407AS	
1	2015	1	1	4	AS	135	N527AS	
2	2015	1	1	4	AS	108	N309AS	
3	2015	1	1	4	AS	122	N413AS	
4	2015	1	1	4	AS	130	N457AS	

5 rows × 32 columns



In [25]:

```
averageDepartDelay = mdata.groupby('AIRLINE_x')['DEPARTURE_DELAY'].mean()
```

In [26]:

```
averageDepartDelay.sort_values(ascending=False)
```

Out[26]:

```
AIRLINE_x
NK    15.883101
UA    14.333056
F9    13.303352
B6    11.442467
WN    10.517183
MQ     9.967187
VX     8.993486
AA     8.826106
EV     8.615598
OO     7.736083
DL     7.313300
US     6.081000
AS     1.718926
HA     0.469918
Name: DEPARTURE_DELAY, dtype: float64
```

In [27]:

```
averageDepartDelay.sort_values(ascending=True)
```

Out[27]:

```
AIRLINE_x
HA    0.469918
AS    1.718926
US    6.081000
DL    7.313300
OO    7.736083
EV    8.615598
AA    8.826106
VX    8.993486
MQ    9.967187
WN   10.517183
B6   11.442467
F9   13.303352
UA   14.333056
NK   15.883101
Name: DEPARTURE_DELAY, dtype: float64
```

In [28]:

```
month_flight = mdata.groupby('MONTH')['FLIGHT_NUMBER'].count()
mdata.groupby('MONTH')['FLIGHT_NUMBER'].count()
```

Out[28]:

```
MONTH
1    457013
2    407663
3    492138
4    479251
5    489641
6    492847
7    514384
8    503956
9    462153
10   482878
11   462367
12   469717
Name: FLIGHT_NUMBER, dtype: int64
```

In [29]:

```
march = mdata[mdata.MONTH == 3]
```

In [31]:

```
march[['AIRLINE_x', 'FLIGHT_NUMBER', 'MONTH']].groupby(['AIRLINE_x', 'MONTH']).count().unstack().max()
```

Out[31]:

```
MONTH
FLIGHT_NUMBER 3    106854
dtype: int64
```

In [33]:

```
(mdata.groupby('ORIGIN_AIRPORT').size())
```


Out[33]:

ORIGIN_AIRPORT

10135	226
10136	181
10140	1702
10141	66
10146	81
10154	28
10155	135
10157	110
10158	229
10165	9
10170	26
10185	265
10208	213
10257	688
10268	54
10279	287
10299	1148
10333	48
10372	55
10397	30750
10408	248
10423	3764
10431	263
10434	106
10469	113
10529	1613
10551	79
10561	172
10577	58
10581	25

...

SRQ	3318
STC	77
STL	46181
STT	4171
STX	932
SUN	867
SUX	589
SWF	678
SYR	5447
TLH	3141
TOL	897
TPA	63077
TRI	1906
TTN	2771
TUL	13701
TUS	14922
TVC	2660
TWF	805
TXK	918
TYR	2199
TYS	6754
UST	144
VEL	200
VLD	925
VPS	4744
WRG	649
WYS	208

```
XNA    8963
YAK     650
YUM    1854
Length: 929, dtype: int64
```

In [34]:

```
neworigin = (mdata.groupby('ORIGIN_AIRPORT').size()).sort_values()
print(neworigin.tail(n=10))
```

```
ORIGIN_AIRPORT
MSP    111055
LAS    131937
IAH    144019
SFO    145491
PHX    145552
LAX    192003
DEN    193402
DFW    232647
ORD    276554
ATL    343506
dtype: int64
```

In [35]:

```
(mdata.groupby('DESTINATION_AIRPORT').size())
```

Out[35]:

DESTINATION_AIRPORT

10135	224
10136	183
10140	1706
10141	67
10146	82
10154	27
10155	135
10157	110
10158	230
10165	9
10170	26
10185	266
10208	212
10257	692
10268	54
10279	287
10299	1148
10333	47
10372	54
10397	30783
10408	247
10423	3741
10431	259
10434	107
10469	113
10529	1612
10551	79
10561	172
10577	58
10581	24

...

SRQ	3336
STC	77
STL	46273
STT	4306
STX	933
SUN	854
SUX	591
SWF	680
SYR	5475
TLH	3149
TOL	901
TPA	63157
TRI	1912
TTN	2761
TUL	13748
TUS	14956
TVC	2666
TWF	806
TXK	915
TYR	2199
TYS	6764
UST	146
VEL	197
VLD	921
VPS	4743
WRG	652
WYS	207

```
XNA      8986
YAK       652
YUM      1856
Length: 929, dtype: int64
```

In [36]:

```
neworigin = (mdata.groupby('DESTINATION_AIRPORT').size()).sort_values()
print(neworigin.tail(n=10))
```

```
DESTINATION_AIRPORT
MSP    111146
LAS    132124
IAH    143587
PHX    145378
SFO    145409
LAX    192136
DEN    193033
DFW    231764
ORD    275864
ATL    343076
dtype: int64
```

In [38]:

```
import numpy as np
mdata["Delays"] = np.where(mdata['DEPARTURE_DELAY' or 'ARRIVAL_DELAY'] > 0, 'D','NotD')
print(mdata)
```

	YEAR	MONTH	DAY	DAY_OF_WEEK	AIRLINE_x	FLIGHT_NUMBER	TAIL_NUMBER \
0	2015	1	1	4	AS	98	N407AS
1	2015	1	1	4	AS	135	N527AS
2	2015	1	1	4	AS	108	N309AS
3	2015	1	1	4	AS	122	N413AS
4	2015	1	1	4	AS	130	N457AS
5	2015	1	1	4	AS	134	N464AS
6	2015	1	1	4	AS	144	N514AS
7	2015	1	1	4	AS	114	N303AS
8	2015	1	1	4	AS	695	N607AS
9	2015	1	1	4	AS	730	N423AS
10	2015	1	1	4	AS	81	N577AS
11	2015	1	1	4	AS	162	N792AS
12	2015	1	1	4	AS	200	N767AS
13	2015	1	1	4	AS	342	N440AS
14	2015	1	1	4	AS	406	N589AS
15	2015	1	1	4	AS	477	N453AS
16	2015	1	1	4	AS	631	N512AS
17	2015	1	1	4	AS	683	N618AS
18	2015	1	1	4	AS	699	N532AS
19	2015	1	1	4	AS	320	N622AS
20	2015	1	1	4	AS	602	N407AS
21	2015	1	1	4	AS	345	N469AS
22	2015	1	1	4	AS	454	N442AS
23	2015	1	1	4	AS	504	N627AS
24	2015	1	1	4	AS	545	N519AS
25	2015	1	1	4	AS	124	N593AS
26	2015	1	1	4	AS	157	N762AS
27	2015	1	1	4	AS	499	N551AS
28	2015	1	1	4	AS	573	N467AS
29	2015	1	1	4	AS	240	N569AS

...
5713978	2015	12	31	4	VX	1488	N838VA
5713979	2015	12	31	4	VX	919	N627VA
5713980	2015	12	31	4	VX	97	N526VA
5713981	2015	12	31	4	VX	241	N621VA
5713982	2015	12	31	4	VX	2	N843VA
5713983	2015	12	31	4	VX	755	N636VA
5713984	2015	12	31	4	VX	963	N853VA
5713985	2015	12	31	4	VX	221	N855VA
5713986	2015	12	31	4	VX	315	N835VA
5713987	2015	12	31	4	VX	882	N846VA
5713988	2015	12	31	4	VX	945	N639VA
5713989	2015	12	31	4	VX	367	N836VA
5713990	2015	12	31	4	VX	717	N523VA
5713991	2015	12	31	4	VX	936	N840VA
5713992	2015	12	31	4	VX	777	N530VA
5713993	2015	12	31	4	VX	47	N281VA
5713994	2015	12	31	4	VX	960	N839VA
5713995	2015	12	31	4	VX	211	N628VA
5713996	2015	12	31	4	VX	347	N852VA
5713997	2015	12	31	4	VX	918	N283VA
5713998	2015	12	31	4	VX	938	N527VA
5713999	2015	12	31	4	VX	1905	N848VA
5714000	2015	12	31	4	VX	593	N625VA
5714001	2015	12	31	4	VX	941	N854VA
5714002	2015	12	31	4	VX	337	N631VA
5714003	2015	12	31	4	VX	769	N622VA
5714004	2015	12	31	4	VX	357	N284VA
5714005	2015	12	31	4	VX	1916	N853VA
5714006	2015	12	31	4	VX	490	N840VA

5714007 2015 12 31 4 VX 48 N281VA

	ORIGIN_AIRPORT	DESTINATION_AIRPORT	SCHEDULED_DEPARTURE	...	\
0	ANC	SEA	5	...	
1	SEA	ANC	25	...	
2	ANC	SEA	45	...	
3	ANC	PDX	50	...	
4	FAI	SEA	115	...	
5	ANC	SEA	155	...	
6	ANC	PDX	200	...	
7	ANC	SEA	220	...	
8	GEG	SEA	500	...	
9	ANC	SEA	505	...	
10	SEA	ANC	600	...	
11	FAI	ANC	600	...	
12	SEA	SJC	600	...	
13	SEA	OAK	600	...	
14	PDX	SJC	600	...	
15	LAX	SEA	600	...	
16	PHX	SEA	600	...	
17	DEN	SEA	600	...	
18	GEG	SEA	600	...	
19	SEA	SFO	605	...	
20	SEA	LAS	605	...	
21	OAK	SEA	610	...	
22	SEA	LAX	610	...	
23	SEA	SNA	615	...	
24	ONT	SEA	620	...	
25	FAI	SEA	625	...	
26	ANC	OTZ	630	...	
27	SAN	SEA	630	...	
28	SAN	PDX	630	...	
29	SEA	SAN	635	...	
...	
5713978	LAX	LAS	1700	...	
5713979	LAS	SFO	1700	...	
5713980	IAD	LAX	1700	...	
5713981	ORD	LAX	1710	...	
5713982	DCA	SFO	1710	...	
5713983	SEA	SFO	1720	...	
5713984	SAN	SFO	1725	...	
5713985	AUS	SFO	1730	...	
5713986	MCO	LAX	1730	...	
5713987	LAX	DAL	1730	...	
5713988	LAX	SFO	1735	...	
5713989	BOS	LAX	1745	...	
5713990	DAL	SFO	1745	...	
5713991	SFO	LAX	1745	...	
5713992	DAL	LAS	1800	...	
5713993	SFO	OGG	1805	...	
5713994	SFO	SAN	1805	...	
5713995	ORD	SFO	1830	...	
5713996	FLL	SFO	1830	...	
5713997	SFO	LAS	1830	...	
5713998	SFO	LAX	1830	...	
5713999	LAS	SFO	1835	...	
5714000	PSP	SFO	1840	...	
5714001	LAX	SFO	1850	...	
5714002	FLL	LAX	1855	...	
5714003	LGA	DAL	1855	...	
5714004	BOS	SFO	1859	...	

5714005	SFO	LAS	1940 ...
5714006	LAX	LAS	1950 ...
5714007	OGG	SFO	2320 ...

	DIVERTED	CANCELLED	CANCELLATION_REASON	AIR_SYSTEM_DELAY \
0	0	0	NaN	NaN
1	0	0	NaN	NaN
2	0	0	NaN	NaN
3	0	0	NaN	NaN
4	0	0	NaN	NaN
5	0	0	NaN	NaN
6	0	0	NaN	NaN
7	0	0	NaN	NaN
8	0	0	NaN	NaN
9	0	0	NaN	NaN
10	0	0	NaN	NaN
11	0	0	NaN	NaN
12	0	0	NaN	NaN
13	0	0	NaN	NaN
14	0	0	NaN	NaN
15	0	0	NaN	NaN
16	0	0	NaN	NaN
17	0	0	NaN	NaN
18	0	0	NaN	NaN
19	0	0	NaN	NaN
20	0	0	NaN	NaN
21	0	0	NaN	NaN
22	0	0	NaN	NaN
23	0	0	NaN	NaN
24	0	0	NaN	NaN
25	0	0	NaN	NaN
26	0	0	NaN	NaN
27	0	0	NaN	NaN
28	0	0	NaN	NaN
29	0	0	NaN	NaN
...
5713978	0	0	NaN	0.0
5713979	0	0	NaN	NaN
5713980	0	0	NaN	NaN
5713981	0	0	NaN	NaN
5713982	0	0	NaN	NaN
5713983	0	0	NaN	NaN
5713984	0	0	NaN	NaN
5713985	0	0	NaN	NaN
5713986	0	0	NaN	NaN
5713987	0	0	NaN	0.0
5713988	0	0	NaN	NaN
5713989	0	0	NaN	NaN
5713990	0	0	NaN	0.0
5713991	0	0	NaN	NaN
5713992	0	0	NaN	7.0
5713993	0	0	NaN	NaN
5713994	0	0	NaN	NaN
5713995	0	0	NaN	NaN
5713996	0	0	NaN	NaN
5713997	0	0	NaN	NaN
5713998	0	0	NaN	NaN
5713999	0	0	NaN	0.0
5714000	0	0	NaN	NaN
5714001	0	0	NaN	NaN
5714002	0	0	NaN	NaN

5714003	0	0	NaN	NaN
5714004	0	0	NaN	NaN
5714005	0	0	NaN	NaN
5714006	0	0	NaN	NaN
5714007	0	0	NaN	NaN

	SECURITY_DELAY	AIRLINE_DELAY	LATE_AIRCRAFT_DELAY	WEATHER_DELAY \
0	NaN	NaN	NaN	NaN
1	NaN	NaN	NaN	NaN
2	NaN	NaN	NaN	NaN
3	NaN	NaN	NaN	NaN
4	NaN	NaN	NaN	NaN
5	NaN	NaN	NaN	NaN
6	NaN	NaN	NaN	NaN
7	NaN	NaN	NaN	NaN
8	NaN	NaN	NaN	NaN
9	NaN	NaN	NaN	NaN
10	NaN	NaN	NaN	NaN
11	NaN	NaN	NaN	NaN
12	NaN	NaN	NaN	NaN
13	NaN	NaN	NaN	NaN
14	NaN	NaN	NaN	NaN
15	NaN	NaN	NaN	NaN
16	NaN	NaN	NaN	NaN
17	NaN	NaN	NaN	NaN
18	NaN	NaN	NaN	NaN
19	NaN	NaN	NaN	NaN
20	NaN	NaN	NaN	NaN
21	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	NaN
23	NaN	NaN	NaN	NaN
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	NaN
27	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	NaN
29	NaN	NaN	NaN	NaN
...
5713978	0.0	0.0	82.0	0.0
5713979	NaN	NaN	NaN	NaN
5713980	NaN	NaN	NaN	NaN
5713981	NaN	NaN	NaN	NaN
5713982	NaN	NaN	NaN	NaN
5713983	NaN	NaN	NaN	NaN
5713984	NaN	NaN	NaN	NaN
5713985	NaN	NaN	NaN	NaN
5713986	NaN	NaN	NaN	NaN
5713987	0.0	0.0	80.0	0.0
5713988	NaN	NaN	NaN	NaN
5713989	NaN	NaN	NaN	NaN
5713990	0.0	0.0	42.0	0.0
5713991	NaN	NaN	NaN	NaN
5713992	0.0	0.0	29.0	0.0
5713993	NaN	NaN	NaN	NaN
5713994	NaN	NaN	NaN	NaN
5713995	NaN	NaN	NaN	NaN
5713996	NaN	NaN	NaN	NaN
5713997	NaN	NaN	NaN	NaN
5713998	NaN	NaN	NaN	NaN
5713999	0.0	16.0	0.0	0.0
5714000	NaN	NaN	NaN	NaN

5714001	NaN	NaN	NaN	NaN
5714002	NaN	NaN	NaN	NaN
5714003	NaN	NaN	NaN	NaN
5714004	NaN	NaN	NaN	NaN
5714005	NaN	NaN	NaN	NaN
5714006	NaN	NaN	NaN	NaN
5714007	NaN	NaN	NaN	NaN

	AIRLINE_y	Delays
0	Alaska Airlines Inc.	NotD
1	Alaska Airlines Inc.	NotD
2	Alaska Airlines Inc.	NotD
3	Alaska Airlines Inc.	NotD
4	Alaska Airlines Inc.	NotD
5	Alaska Airlines Inc.	NotD
6	Alaska Airlines Inc.	NotD
7	Alaska Airlines Inc.	NotD
8	Alaska Airlines Inc.	NotD
9	Alaska Airlines Inc.	NotD
10	Alaska Airlines Inc.	NotD
11	Alaska Airlines Inc.	NotD
12	Alaska Airlines Inc.	NotD
13	Alaska Airlines Inc.	NotD
14	Alaska Airlines Inc.	NotD
15	Alaska Airlines Inc.	NotD
16	Alaska Airlines Inc.	NotD
17	Alaska Airlines Inc.	NotD
18	Alaska Airlines Inc.	NotD
19	Alaska Airlines Inc.	NotD
20	Alaska Airlines Inc.	NotD
21	Alaska Airlines Inc.	NotD
22	Alaska Airlines Inc.	D
23	Alaska Airlines Inc.	NotD
24	Alaska Airlines Inc.	NotD
25	Alaska Airlines Inc.	NotD
26	Alaska Airlines Inc.	NotD
27	Alaska Airlines Inc.	NotD
28	Alaska Airlines Inc.	NotD
29	Alaska Airlines Inc.	NotD

...
5713978	Virgin America	D
5713979	Virgin America	D
5713980	Virgin America	NotD
5713981	Virgin America	NotD
5713982	Virgin America	NotD
5713983	Virgin America	NotD
5713984	Virgin America	NotD
5713985	Virgin America	NotD
5713986	Virgin America	NotD
5713987	Virgin America	D
5713988	Virgin America	NotD
5713989	Virgin America	NotD
5713990	Virgin America	D
5713991	Virgin America	D
5713992	Virgin America	D
5713993	Virgin America	D
5713994	Virgin America	NotD
5713995	Virgin America	NotD
5713996	Virgin America	NotD
5713997	Virgin America	D
5713998	Virgin America	NotD

5713999	Virgin America	D
5714000	Virgin America	NotD
5714001	Virgin America	NotD
5714002	Virgin America	D
5714003	Virgin America	NotD
5714004	Virgin America	NotD
5714005	Virgin America	NotD
5714006	Virgin America	NotD
5714007	Virgin America	NotD

[5714008 rows x 33 columns]

In [39]:

```
mdata.groupby('Delays').size()
```

Out[39]:

```
Delays
D      2115049
NotD   3598959
dtype: int64
```

In [40]:

```
data1[['DEPARTURE_DELAY', 'ARRIVAL_DELAY']].dropna(how='any').shape
```

Out[40]:

```
(5714008, 2)
```

In [41]:

```
averageDepartDelay = mdata.groupby('AIRLINE_x')['DEPARTURE_DELAY'].mean()
```

In [47]:

```
averageDepartDelay.min()
```

Out[47]:

```
0.4699175444825818
```

In [48]:

```
print(averageDepartDelay.idxmin())
print(averageDepartDelay.min())
```

```
HA
0.4699175444825818
```

In []: