In [27]:
```
import pandas as pd
import matplotlib.pyplot as plt
df= pd.read_csv('flights.csv - flights.csv.csv')
df
```

Out[27]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
0	2013	1	1	517.0	515	2.0	830.0	819
1	2013	1	1	533.0	529	4.0	850.0	830
2	2013	1	1	542.0	540	2.0	923.0	850
3	2013	1	1	544.0	545	-1.0	1004.0	1022
4	2013	1	1	554.0	600	-6.0	812.0	837
...	...	...	...	...	...	...	...	...
336771	2013	9	30	NaN	1455	NaN	NaN	1634
336772	2013	9	30	NaN	2200	NaN	NaN	2312
336773	2013	9	30	NaN	1210	NaN	NaN	1330
336774	2013	9	30	NaN	1159	NaN	NaN	1344
336775	2013	9	30	NaN	840	NaN	NaN	1020

336776 rows × 19 columns

In [14]:
```
#Q1
b=df['arr_time']-df['sched_arr_time']
df1=df[(df['dep_delay']>1) & (b<30)]
df1
```

Out[14]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
0	2013	1	1	517.0	515	2.0	830.0	819
1	2013	1	1	533.0	529	4.0	850.0	830
26	2013	1	1	611.0	600	11.0	945.0	931
27	2013	1	1	613.0	610	3.0	925.0	921
31	2013	1	1	623.0	610	13.0	920.0	915
...	...	...	...	...	...	...	...	...
336752	2013	9	30	2142.0	2129	13.0	2250.0	2239
336754	2013	9	30	2147.0	2137	10.0	30.0	27
336759	2013	9	30	2207.0	2140	27.0	2257.0	2250
336763	2013	9	30	2235.0	2001	154.0	59.0	2249
336768	2013	9	30	2307.0	2255	12.0	2359.0	2358

67835 rows × 19 columns

In [15]:
```
#Q2
df2=df[df.dep_time<=600]
df2
```

Out[15]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
0	2013	1	1	517.0	515	2.0	830.0	819
1	2013	1	1	533.0	529	4.0	850.0	830
2	2013	1	1	542.0	540	2.0	923.0	850
3	2013	1	1	544.0	545	-1.0	1004.0	1022
4	2013	1	1	554.0	600	-6.0	812.0	837
...	...	...	...	...	...	...	...	...
335802	2013	9	30	557.0	600	-3.0	852.0	923
335803	2013	9	30	558.0	600	-2.0	815.0	829
335804	2013	9	30	558.0	600	-2.0	742.0	749
335805	2013	9	30	559.0	600	-1.0	NaN	715
335806	2013	9	30	600.0	600	0.0	844.0	856

9344 rows × 19 columns

In [11]:
```
#Q3
df3=df[df['dep_time'].isnull()]
df3
```

Out[11]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
838	2013	1	1	NaN	1630	NaN	NaN	1815
839	2013	1	1	NaN	1935	NaN	NaN	2240
840	2013	1	1	NaN	1500	NaN	NaN	1825
841	2013	1	1	NaN	600	NaN	NaN	901
1777	2013	1	2	NaN	1540	NaN	NaN	1747
...	...	...	...	...	...	...	...	...

In [16]:
```
df.isnull().sum()
```

Out[16]:
```
year                0
month               0
day                 0
dep_time         8255
sched_dep_time      0
dep_delay        8255
arr_time         8713
sched_arr_time      0
arr_delay        9430
carrier             0
flight              0
tailnum          2512
origin              0
dest                0
air_time         9430
distance            0
hour                0
minute              0
time_hour           0
dtype: int64
```

In [7]:
```
#Q4
df_asc =df.sort_values(by="dep_time")
df_asc
```

Out[7]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
250450	2013	7	1	1.0	2029	212.0	236.0	2359
109552	2013	12	30	1.0	2359	2.0	441.0	437
240026	2013	6	20	1.0	2359	2.0	340.0	350
212954	2013	5	22	1.0	1935	266.0	154.0	2140
215892	2013	5	25	1.0	2359	2.0	336.0	341
...	...	...	...	...	...	...	...	...
336771	2013	9	30	NaN	1455	NaN	NaN	1634
336772	2013	9	30	NaN	2200	NaN	NaN	2312
336773	2013	9	30	NaN	1210	NaN	NaN	1330
336774	2013	9	30	NaN	1159	NaN	NaN	1344
336775	2013	9	30	NaN	840	NaN	NaN	1020

336776 rows × 19 columns

In [9]:
```
#Q4
df_dsc =df.sort_values(by="dep_time",ascending=False)
df_dsc
```

Out[9]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
80973	2013	11	27	2400.0	2359	1.0	515.0	445
156854	2013	3	22	2400.0	2358	2.0	339.0	338
54966	2013	10	30	2400.0	2359	1.0	327.0	337
256601	2013	7	7	2400.0	1950	250.0	107.0	2130
159422	2013	3	25	2400.0	2355	5.0	339.0	340
...	...	...	...	...	...	...	...	...
336771	2013	9	30	NaN	1455	NaN	NaN	1634
336772	2013	9	30	NaN	2200	NaN	NaN	2312
336773	2013	9	30	NaN	1210	NaN	NaN	1330
336774	2013	9	30	NaN	1159	NaN	NaN	1344
336775	2013	9	30	NaN	840	NaN	NaN	1020

336776 rows × 19 columns

In [10]:
```
#Q5
delayed=df.sort_values(by="dep_delay",ascending=False)
delayed
```

Out[10]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
7072	2013	1	9	641.0	900	1301.0	1242.0	1530
235778	2013	6	15	1432.0	1935	1137.0	1607.0	2120
8239	2013	1	10	1121.0	1635	1126.0	1239.0	1810
327043	2013	9	20	1139.0	1845	1014.0	1457.0	2210
270376	2013	7	22	845.0	1600	1005.0	1044.0	1815
...	...	...	...	...	...	...	...	...
336771	2013	9	30	NaN	1455	NaN	NaN	1634
336772	2013	9	30	NaN	2200	NaN	NaN	2312
336773	2013	9	30	NaN	1210	NaN	NaN	1330
336774	2013	9	30	NaN	1159	NaN	NaN	1344
336775	2013	9	30	NaN	840	NaN	NaN	1020

336776 rows × 19 columns

In [17]:
```
#Q5
earliest=df.sort_values(by="dep_time",ascending=False)
earliest
```

Out[17]:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time
80973	2013	11	27	2400.0	2359	1.0	515.0	445
156854	2013	3	22	2400.0	2358	2.0	339.0	338
54966	2013	10	30	2400.0	2359	1.0	327.0	337
256601	2013	7	7	2400.0	1950	250.0	107.0	2130
159422	2013	3	25	2400.0	2355	5.0	339.0	340
...	...	...	...	...	...	...	...	...
336771	2013	9	30	NaN	1455	NaN	NaN	1634
336772	2013	9	30	NaN	2200	NaN	NaN	2312
336773	2013	9	30	NaN	1210	NaN	NaN	1330
336774	2013	9	30	NaN	1159	NaN	NaN	1344
336775	2013	9	30	NaN	840	NaN	NaN	1020

336776 rows × 20 columns

In [16]:
```
#Q6
df['speed']=df["distance"]/df["arr_time"]
fastest=df.sort_values("speed",ascending=False)
print(fastest["speed"])
```

```
103361    2586.0
77963     2586.0
252266    2586.0
238856    2586.0
245577    2586.0
           ...
336771       NaN
336772       NaN
336773       NaN
336774       NaN
336775       NaN
Name: speed, Length: 336776, dtype: float64
```

In [38]:
```
# #Q7
# df7=df.sort_values(by="distance")
# df7.head() #travelled the shortest
# df7.head() #travelled the fastest
```

In [39]:
```
farthest_flight = df.loc[df['distance'].idxmax()]
farthest_flight
```

Out[39]:
```
year                            2013
month                              1
day                                1
dep_time                       857.0
sched_dep_time                   900
dep_delay                       -3.0
arr_time                      1516.0
sched_arr_time                   1530
arr_delay                      -14.0
carrier                           HA
flight                            51
tailnum                       N380HA
origin                           JFK
dest                             HNL
air_time                       659.0
distance                        4983
hour                               9
minute                             0
time_hour          2013-01-01 09:00:00
Name: 162, dtype: object
```

In [40]:
```
shortest_flight = df.loc[df['distance'].idxmin()]
shortest_flight
```

Out[40]:
```
year                            2013
month                              7
day                               27
dep_time                         NaN
sched_dep_time                   106
dep_delay                        NaN
arr_time                         NaN
sched_arr_time                   245
arr_delay                        NaN
carrier                           US
flight                          1632
tailnum                          NaN
origin                           EWR
dest                             LGA
air_time                         NaN
distance                          17
hour                               1
minute                             6
time_hour          2013-07-27 01:00:00
Name: 275945, dtype: object
```

In [19]:
```
#Q8
df8=df[(df["arr_delay"].notnull())].groupby("dest").size()
df8
```

Out[19]:
```
dest
ABQ      254
ACK      264
ALB      418
ANC        8
ATL    16837
       ...
TPA     7390
TUL      294
TVC       95
TYS      578
XNA      992
Length: 104, dtype: int64
```

In [41]:
```
#Q9
cancelled=df.groupby(['year','month',"day"]).size()
cancelled
```

Out[41]:
```
year  month  day
2013  1      1      842
             2      943
             3      914
             4      915
             5      720
                   ...
      12     27     963
             28     814
             29     888
             30     968
             31     776
Length: 365, dtype: int64
```

In [ ]:

In [ ]: