

S03T05

July 29, 2021

1 Exercici 1

```
[21]: import pandas as pd

coes = pd.read_csv('DelayedFlights.csv')
eee=coes[["Year","DepTime","ArrTime","AirTime","ArrDelay","Distance"]].head(5)
list(coes)
#coes[:3]
```

```
[21]: ['Unnamed: 0',
      'Year',
      'Month',
      'DayofMonth',
      'DayOfWeek',
      'DepTime',
      'CRSDepTime',
      'ArrTime',
      'CRSArrTime',
      'UniqueCarrier',
      'FlightNum',
      'TailNum',
      'ActualElapsedTime',
      'CRSElapsedTime',
      'AirTime',
      'ArrDelay',
      'DepDelay',
      'Origin',
      'Dest',
      'Distance',
      'TaxiIn',
      'TaxiOut',
      'Cancelled',
      'CancellationCode',
      'Diverted',
      'CarrierDelay',
      'WeatherDelay',
      'NASDelay',
```

```
'SecurityDelay',  
'LateAircraftDelay']
```

2 Exercici 2

```
[11]: eee.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3 entries, 0 to 2  
Data columns (total 6 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   Year        3 non-null      int64  
1   DepTime     3 non-null      float64  
2   ArrTime     3 non-null      float64  
3   AirTime     3 non-null      float64  
4   ArrDelay    3 non-null      float64  
5   Distance    3 non-null      int64  
dtypes: float64(4), int64(2)  
memory usage: 272.0 bytes
```

```
[21]: cses.isnull().sum()
```

```
[21]: Unnamed: 0          0  
Year                0  
Month              0  
DayofMonth         0  
DayOfWeek          0  
DepTime            0  
CRSDepTime         0  
ArrTime           7110  
CRSArrTime         0  
UniqueCarrier      0  
FlightNum          0  
TailNum           5  
ActualElapsedTime  8387  
CRSElapsedTime     198  
AirTime           8387  
ArrDelay          8387  
DepDelay           0  
Origin            0  
Dest              0  
Distance           0  
TaxiIn            7110  
TaxiOut           455  
Cancelled          0  
CancellationCode   0
```

```

Diverted          0
CarrierDelay      689270
WeatherDelay      689270
NASDelay          689270
SecurityDelay     689270
LateAircraftDelay 689270
dtype: int64

```

```

[15]: coses["VelocitatMitjanaDeVol"]=coses["Distance"]/coses["AirTime"]
      coses["NoCobraPerPuntualitat"]=coses["ArrDelay"]==0
      coses[:5]

```

```

[15]:   Unnamed: 0  Year  Month  DayofMonth  DayOfWeek  DepTime  CRSDepTime  \
0          0  2008     1           3           4    2003.0        1955
1          1  2008     1           3           4     754.0         735
2          2  2008     1           3           4     628.0         620
3          4  2008     1           3           4    1829.0        1755
4          5  2008     1           3           4    1940.0        1915

      ArrTime  CRSArrTime UniqueCarrier  ...  Cancelled  CancellationCode  \
0    2211.0         2225           WN  ...          0              N
1    1002.0         1000           WN  ...          0              N
2     804.0          750           WN  ...          0              N
3    1959.0         1925           WN  ...          0              N
4    2121.0         2110           WN  ...          0              N

      Diverted  CarrierDelay  WeatherDelay  NASDelay  SecurityDelay  \
0           0           NaN           NaN           NaN           NaN
1           0           NaN           NaN           NaN           NaN
2           0           NaN           NaN           NaN           NaN
3           0            2.0            0.0            0.0            0.0
4           0           NaN           NaN           NaN           NaN

      LateAircraftDelay  VelocitatMitjanaDeVol  NoCobraPerPuntualitat
0                NaN           6.982759           False
1                NaN           7.168142           False
2                NaN           6.776316           False
3            32.0           6.688312           False
4                NaN           7.908046           False

```

[5 rows x 32 columns]

```

[41]: #coses['ArrDelay'].unique()
      coses['UniqueCarrier'].value_counts()

```

```

[41]: WN    377602
      AA    191865

```

```

MQ      141920
UA      141426
OO      132433
DL      114238
XE      103663
CO      100195
US       98425
EV       81877
NW       79108
FL       71284
YV       67063
B6       55315
OH       52657
9E       51885
AS       39293
F9       28269
HA        7490
AQ         750
Name: UniqueCarrier, dtype: int64

```

```

[42]: cses[["Year","DepTime","ArrTime","AirTime","ArrDelay","Distance"]].
      ↪sort_values(by=['Distance'],ascending=[False]).head(5)

```

```

[42]:
      Year  DepTime  ArrTime  AirTime  ArrDelay  Distance
1255581  2008    2001.0    1126.0    543.0      -4.0     4962
1926346  2008    2333.0    1348.0    528.0     148.0     4962
718062   2008    1345.0    1826.0    620.0      16.0     4962
1928077  2008    2135.0    1148.0    523.0      28.0     4962
173695   2008    1326.0    1912.0    621.0      -8.0     4962

```

```

[43]: cses[["Year","DepTime","ArrTime","AirTime","ArrDelay","Distance"]].
      ↪sort_values(by=['ArrDelay'],ascending=[False]).head(5)

```

```

[43]:
      Year  DepTime  ArrTime  AirTime  ArrDelay  Distance
322516  2008    1117.0    2256.0    437.0    2461.0     3972
686014  2008     724.0     858.0    132.0    2453.0        930
839306  2008         2.0     254.0    145.0    1951.0     1085
1009553 2008    2140.0    2252.0     50.0    1707.0        304
1881639 2008    1602.0    1921.0    192.0    1655.0     1124

```

3 Exercici 3

```

[45]: #cses.to_excel('cses.xlsx')
      eee.to_excel('eee.xlsx')

```

```

[ ]:

```