```
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

```
#import library
import pandas as pd
import plotly.express as pr
import plotly.graph_objects as go

# Read the data
df = pd.read_csv('/content/drive/MyDrive/project_dataset.csv.xls')

# To display the data
df
```

```
       Unnamed: 0  Year  Quarter  Month  DayofMonth  DayOfWeek
FlightDate  \
0         1295781  1998        2      4           2          4  1998-
04-02
1         1125375  2013        2      5          13          1  2013-
05-13
2          118824  1993        3      9          25          6  1993-
09-25
3          634825  1994        4     11          12          6  1994-
11-12
4         1888125  2017        3      8          17          4  2017-
08-17
...           ...   ...      ...    ...         ...        ...
...
26995      821542  2017        1      1          24          2  2017-
01-24
26996     1910565  2013        2      6          27          4  2013-
06-27
26997        9055  2016        3      8          26          5  2016-
08-26
26998       84136  2009        3      8           8          6  2009-
08-08
26999      113029  1993        3      7          17          6  1993-
07-17

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
0                    AS                     19930
AS
1                    EV                     20366
EV
2                    UA                     19977
UA
3                    HP                     19991
HP
```

```
4                    UA                    19977
UA
...                 ...                      ...
...
26995                DL                    19790
DL
26996                B6                    20409
B6
26997                AA                    19805
AA
26998                YV                    20378
YV
26999                DL                    19790
DL

       ... Div4WheelsOff  Div4TailNum  Div5Airport  Div5AirportID  \
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
...    ...           ...          ...          ...            ...
26995  ...           NaN          NaN          NaN            NaN
26996  ...           NaN          NaN          NaN            NaN
26997  ...           NaN          NaN          NaN            NaN
26998  ...           NaN          NaN          NaN            NaN
26999  ...           NaN          NaN          NaN            NaN

       Div5AirportSeqID Div5WheelsOn Div5TotalGTime
Div5LongestGTime  \
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

...                 ...          ...            ...              ...

26995               NaN          NaN            NaN              NaN

26996               NaN          NaN            NaN              NaN

26997               NaN          NaN            NaN              NaN

26998               NaN          NaN            NaN              NaN
```

```
26999               NaN          NaN             NaN               NaN


       Div5WheelsOff Div5TailNum
0                NaN         NaN
1                NaN         NaN
2                NaN         NaN
3                NaN         NaN
4                NaN         NaN
...              ...         ...
26995            NaN         NaN
26996            NaN         NaN
26997            NaN         NaN
26998            NaN         NaN
26999            NaN         NaN

[27000 rows x 110 columns]
```

```
# For find the total Rows and columns
df.shape
```

```
(27000, 110)
```

```
# for copy file fone one file to another file
df1 = df.copy()
```

```
# To display values from Top to Bottom
df1.head(12)
```

```
    Unnamed: 0  Year  Quarter  Month  DayofMonth  DayOfWeek
FlightDate  \
0      1295781  1998        2      4           2          4  1998-04-
02
1      1125375  2013        2      5          13          1  2013-05-
13
2       118824  1993        3      9          25          6  1993-09-
25
3       634825  1994        4     11          12          6  1994-11-
12
4      1888125  2017        3      8          17          4  2017-08-
17
5      1133769  2014        2      4           2          3  2014-04-
02
6      1465620  2011        3      8          15          1  2011-08-
15
7       172607  2019        1      1           6          7  2019-01-
06
8      1975839  2002        1      1          25          5  2002-01-
25
9      1912303  2005        3      8           1          1  2005-08-
```

```
01
10      525812  2000        4     12          17           7  2000-12-
17
11     1963835  2005        3      7           5           2  2005-07-
05

   Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
0                  AS                     19930
AS
1                  EV                     20366
EV
2                  UA                     19977
UA
3                  HP                     19991
HP
4                  UA                     19977
UA
5                  AA                     19805
AA
6                  WN                     19393
WN
7                  EV                     20366
EV
8                  AA                     19805
AA
9                  AA                     19805
AA
10                 UA                     19977
UA
11                 WN                     19393
WN

    ... Div4WheelsOff  Div4TailNum  Div5Airport  Div5AirportID  \
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

    Div5AirportSeqID Div5WheelsOn Div5TotalGTime Div5LongestGTime  \
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

    Div5WheelsOff Div5TailNum
0             NaN         NaN
1             NaN         NaN
2             NaN         NaN
3             NaN         NaN
4             NaN         NaN
5             NaN         NaN
6             NaN         NaN
7             NaN         NaN
8             NaN         NaN
9             NaN         NaN
10            NaN         NaN
11            NaN         NaN

[12 rows x 110 columns]
```

```
# To check the dtat type
df1.dtypes
```

```
Unnamed: 0            int64
Year                  int64
Quarter               int64
Month                 int64
DayofMonth            int64
                     ...
Div5WheelsOn        float64
Div5TotalGTime      float64
Div5LongestGTime    float64
Div5WheelsOff       float64
Div5TailNum         float64
Length: 110, dtype: object
```

```
# display the information of the dataset
df1.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27000 entries, 0 to 26999
Columns: 110 entries, Unnamed: 0 to Div5TailNum
dtypes: float64(74), int64(19), object(17)
memory usage: 22.7+ MB
```

```
#To display dicription of the dataset
df1.describe()
```

```
         Unnamed: 0          Year       Quarter         Month
DayofMonth  \
count  2.700000e+04  27000.000000  27000.000000  27000.000000
27000.000000
mean   9.987625e+05   2004.304926      2.505889      6.518667
15.714370
std    5.767890e+05      9.189835      1.114589      3.444192
8.816548
min    1.480000e+02   1987.000000      1.000000      1.000000
1.000000
25%    5.029945e+05   1997.000000      2.000000      4.000000
8.000000
50%    9.964900e+05   2005.000000      3.000000      7.000000
16.000000
75%    1.500778e+06   2012.000000      3.000000      9.000000
23.000000
max    1.999844e+06   2020.000000      4.000000     12.000000
31.000000

          DayOfWeek  DOT_ID_Reporting_Airline  \
count  27000.000000              27000.000000
mean       3.939407              19923.175222
std        1.986853                367.912667
min        1.000000              19386.000000
25%        2.000000              19704.000000
50%        4.000000              19805.000000
75%        6.000000              20355.000000
max        7.000000              21171.000000

       Flight_Number_Reporting_Airline  OriginAirportID
OriginAirportSeqID  \
count                     27000.000000     27000.000000
2.700000e+04
mean                       1727.947963     12740.372481
1.274039e+06
std                        1658.882759      1543.896695
1.543894e+05
min                           1.000000     10135.000000
1.013501e+06
25%                         528.000000     11292.000000
1.129202e+06
50%                        1173.000000     12892.000000
1.289201e+06
75%                        2229.000000     14100.000000
1.410001e+06
max                        9576.000000     16440.000000
1.644001e+06
```

```
       ...  Div4WheelsOff  Div4TailNum  Div5Airport  Div5AirportID  \
count  ...            0.0          0.0          0.0            0.0
mean   ...            NaN          NaN          NaN            NaN
std    ...            NaN          NaN          NaN            NaN
min    ...            NaN          NaN          NaN            NaN
25%    ...            NaN          NaN          NaN            NaN
50%    ...            NaN          NaN          NaN            NaN
75%    ...            NaN          NaN          NaN            NaN
max    ...            NaN          NaN          NaN            NaN

       Div5AirportSeqID  Div5WheelsOn  Div5TotalGTime
Div5LongestGTime  \
count               0.0           0.0             0.0
0.0
mean                NaN           NaN             NaN
NaN
std                 NaN           NaN             NaN
NaN
min                 NaN           NaN             NaN
NaN
25%                 NaN           NaN             NaN
NaN
50%                 NaN           NaN             NaN
NaN
75%                 NaN           NaN             NaN
NaN
max                 NaN           NaN             NaN
NaN

       Div5WheelsOff  Div5TailNum
count            0.0          0.0
mean             NaN          NaN
std              NaN          NaN
min              NaN          NaN
25%              NaN          NaN
50%              NaN          NaN
75%              NaN          NaN
max              NaN          NaN

[8 rows x 93 columns]
```

```
#To check duplicate values
df1.duplicated()
```

```
0        False
1        False
2        False
3        False
4        False
```

```
        ...
26995    False
26996    False
26997    False
26998    False
26999    False
Length: 27000, dtype: bool
```

```
#To find NULL Values
df1.isna()
```

```
       Unnamed: 0   Year  Quarter  Month  DayofMonth  DayOfWeek
FlightDate  \
0           False  False    False  False       False      False
False
1           False  False    False  False       False      False
False
2           False  False    False  False       False      False
False
3           False  False    False  False       False      False
False
4           False  False    False  False       False      False
False
...           ...    ...      ...    ...         ...        ...
...
26995       False  False    False  False       False      False
False
26996       False  False    False  False       False      False
False
26997       False  False    False  False       False      False
False
26998       False  False    False  False       False      False
False
26999       False  False    False  False       False      False
False

       Reporting_Airline  DOT_ID_Reporting_Airline  \
0                  False                     False
1                  False                     False
2                  False                     False
3                  False                     False
4                  False                     False
...                  ...                       ...
26995              False                     False
26996              False                     False
26997              False                     False
26998              False                     False
26999              False                     False

       IATA_CODE_Reporting_Airline  ...  Div4WheelsOff  Div4TailNum  \
```

```
0                         False  ...          True         True
1                         False  ...          True         True
2                         False  ...          True         True
3                         False  ...          True         True
4                         False  ...          True         True
...                         ...  ...           ...          ...
26995                     False  ...          True         True
26996                     False  ...          True         True
26997                     False  ...          True         True
26998                     False  ...          True         True
26999                     False  ...          True         True

       Div5Airport  Div5AirportID  Div5AirportSeqID  Div5WheelsOn  \
0             True           True              True          True
1             True           True              True          True
2             True           True              True          True
3             True           True              True          True
4             True           True              True          True
...            ...            ...               ...           ...
26995         True           True              True          True
26996         True           True              True          True
26997         True           True              True          True
26998         True           True              True          True
26999         True           True              True          True

       Div5TotalGTime  Div5LongestGTime  Div5WheelsOff  Div5TailNum
0                True              True           True         True
1                True              True           True         True
2                True              True           True         True
3                True              True           True         True
4                True              True           True         True
...               ...               ...            ...          ...
26995            True              True           True         True
26996            True              True           True         True
26997            True              True           True         True
26998            True              True           True         True
26999            True              True           True         True

[27000 rows x 110 columns]
```

```
#calculate the total NULL values in each column
df1.isna().sum()
```

```
Unnamed: 0              0
Year                    0
Quarter                 0
Month                   0
DayofMonth              0
                     ...
Div5WheelsOn        27000
```

```
Div5TotalGTime      27000
Div5LongestGTime    27000
Div5WheelsOff       27000
Div5TailNum         27000
Length: 110, dtype: int64
```

```
df1.isna().sum()/df1.shape[0]
```

```
Unnamed: 0          0.0
Year                0.0
Quarter             0.0
Month               0.0
DayofMonth          0.0
                   ...
Div5WheelsOn        1.0
Div5TotalGTime      1.0
Div5LongestGTime    1.0
Div5WheelsOff       1.0
Div5TailNum         1.0
Length: 110, dtype: float64
```

```
df1.dropna(axis = 1 , inplace=True)
```

```
df1
```

```
       Unnamed: 0  Year  Quarter  Month  DayofMonth  DayOfWeek
FlightDate  \
0         1295781  1998        2      4           2          4  1998-
04-02
1         1125375  2013        2      5          13          1  2013-
05-13
2          118824  1993        3      9          25          6  1993-
09-25
3          634825  1994        4     11          12          6  1994-
11-12
4         1888125  2017        3      8          17          4  2017-
08-17
...           ...   ...      ...    ...         ...        ...
...
26995      821542  2017        1      1          24          2  2017-
01-24
26996     1910565  2013        2      6          27          4  2013-
06-27
26997        9055  2016        3      8          26          5  2016-
08-26
26998       84136  2009        3      8           8          6  2009-
08-08
26999      113029  1993        3      7          17          6  1993-
07-17

      Reporting_Airline  DOT_ID_Reporting_Airline
```

```
IATA_CODE_Reporting_Airline  \
0                    AS                        19930
AS
1                    EV                        20366
EV
2                    UA                        19977
UA
3                    HP                        19991
HP
4                    UA                        19977
UA
...                 ...                          ...
...
26995                DL                        19790
DL
26996                B6                        20409
B6
26997                AA                        19805
AA
26998                YV                        20378
YV
26999                DL                        19790
DL

       ...  DestWac  CRSDepTime  DepTimeBlk  CRSArrTime ArrTimeBlk
Cancelled  \
0      ...       93        1330   1300-1359        1426  1400-1459
0.0
1      ...       38        1301   1300-1359        1423  1400-1459
0.0
2      ...       41        1650   1600-1659        1730  1700-1759
0.0
3      ...       81        1245   1200-1259        1457  1400-1459
0.0
4      ...       82         755   0700-0759         902  0900-0959
0.0
...    ...      ...         ...         ...         ...        ...
...
26995  ...       63         910   0900-0959        1028  1000-1059
0.0
26996  ...       74        1320   1300-1359        1612  1600-1659
0.0
26997  ...       22         920   0900-0959        1405  1400-1459
0.0
26998  ...       82        1139   1100-1159        1217  1200-1259
0.0
26999  ...       74         615   0600-0659         721  0700-0759
0.0

       Diverted  Flights  Distance  DistanceGroup
```

```
0           0.0      1.0     224.0              1
1           0.0      1.0     277.0              2
2           0.0      1.0     130.0              1
3           0.0      1.0     370.0              2
4           0.0      1.0     692.0              3
...         ...      ...       ...            ...
26995       0.0      1.0     297.0              2
26996       0.0      1.0    1562.0              7
26997       0.0      1.0    1521.0              7
26998       0.0      1.0      73.0              1
26999       0.0      1.0     304.0              2

[27000 rows x 32 columns]
```

```
data = df1.sample(n=500 , random_state = 40)
```

```
data
```

```
       Unnamed: 0  Year  Quarter  Month  DayofMonth  DayOfWeek
FlightDate  \
12781      975274  2003        3      7          11          5  2003-
07-11
23751     1471411  1998        2      6           7          7  1998-
06-07
2796      1469316  2001        4     12           1          6  2001-
12-01
18633     1468783  1988        2      5          30          1  1988-
05-30
23187     1861806  2019        4     12           1          7  2019-
12-01
...           ...   ...      ...    ...         ...        ...
...
17188      428525  1995        2      6          17          6  1995-
06-17
24952     1300641  2018        1      3          26          1  2018-
03-26
17089     1410356  2015        1      3          16          1  2015-
03-16
11144      607318  1998        3      9          10          4  1998-
09-10
19762      301716  2005        3      7          10          7  2005-
07-10

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
12781                WN                     19393
WN
23751                US                     20355
US
2796                 UA                     19977
UA
```

```
18633                US                     20355
AL
23187                DL                     19790
DL
...                 ...                       ...
...
17188                US                     20355
US
24952                EV                     20366
EV
17089                AA                     19805
AA
11144                DL                     19790
DL
19762                OO                     20304
OO

       ...  DestWac  CRSDepTime  DepTimeBlk  CRSArrTime ArrTimeBlk
Cancelled  \
12781  ...       33        2045   2000-2059        2140  2100-2159
0.0
23751  ...       23        2035   2000-2059        2203  2200-2259
0.0
2796   ...       91        2230   2200-2259         535  0001-0559
0.0
18633  ...       91         630   0600-0659         723  0700-0759
0.0
23187  ...       85        1530   1500-1559        1757  1700-1759
0.0
...    ...      ...         ...         ...         ...        ...
...
17188  ...       23        1820   1800-1859        1926  1900-1959
0.0
24952  ...       71        1155   1100-1159        1328  1300-1359
0.0
17089  ...       74        1122   1100-1159        1325  1300-1359
0.0
11144  ...       72        1855   1800-1859        1922  1900-1959
0.0
19762  ...       41        1205   1200-1259        1229  1200-1259
0.0

       Diverted  Flights  Distance  DistanceGroup
12781       0.0      1.0     197.0              1
23751       0.0      1.0     448.0              2
2796        0.0      1.0    2615.0             11
18633       0.0      1.0     297.0              2
23187       0.0      1.0     867.0              4
...         ...      ...       ...            ...
17188       0.0      1.0     267.0              2
```

```
24952       0.0      1.0     374.0             2
17089       0.0      1.0     985.0             4
11144       0.0      1.0     425.0             2
19762       0.0      1.0     296.0             2

[500 rows x 32 columns]
```

```
#To find total number of rows and Columns
data.shape
```

```
(500, 32)
```

```
data.head
```

```
<bound method NDFrame.head of        Unnamed: 0  Year  Quarter  Month
DayofMonth  DayOfWeek  FlightDate  \
12781      975274  2003        3      7          11          5  2003-
07-11
23751     1471411  1998        2      6           7          7  1998-
06-07
2796      1469316  2001        4     12           1          6  2001-
12-01
18633     1468783  1988        2      5          30          1  1988-
05-30
23187     1861806  2019        4     12           1          7  2019-
12-01
...           ...   ...      ...    ...         ...        ...
...
17188      428525  1995        2      6          17          6  1995-
06-17
24952     1300641  2018        1      3          26          1  2018-
03-26
17089     1410356  2015        1      3          16          1  2015-
03-16
11144      607318  1998        3      9          10          4  1998-
09-10
19762      301716  2005        3      7          10          7  2005-
07-10

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
12781                WN                     19393
WN
23751                US                     20355
US
2796                 UA                     19977
UA
18633                US                     20355
AL
23187                DL                     19790
DL
```

```
...                 ...                       ...
...
17188                US                     20355
US
24952                EV                     20366
EV
17089                AA                     19805
AA
11144                DL                     19790
DL
19762                OO                     20304
OO

       ...  DestWac  CRSDepTime  DepTimeBlk  CRSArrTime ArrTimeBlk
Cancelled  \
12781  ...       33        2045   2000-2059        2140  2100-2159
0.0
23751  ...       23        2035   2000-2059        2203  2200-2259
0.0
2796   ...       91        2230   2200-2259         535  0001-0559
0.0
18633  ...       91         630   0600-0659         723  0700-0759
0.0
23187  ...       85        1530   1500-1559        1757  1700-1759
0.0
...    ...      ...         ...         ...         ...        ...
...
17188  ...       23        1820   1800-1859        1926  1900-1959
0.0
24952  ...       71        1155   1100-1159        1328  1300-1359
0.0
17089  ...       74        1122   1100-1159        1325  1300-1359
0.0
11144  ...       72        1855   1800-1859        1922  1900-1959
0.0
19762  ...       41        1205   1200-1259        1229  1200-1259
0.0

       Diverted  Flights  Distance  DistanceGroup
12781       0.0      1.0     197.0              1
23751       0.0      1.0     448.0              2
2796        0.0      1.0    2615.0             11
18633       0.0      1.0     297.0              2
23187       0.0      1.0     867.0              4
...         ...      ...       ...            ...
17188       0.0      1.0     267.0              2
24952       0.0      1.0     374.0              2
17089       0.0      1.0     985.0              4
11144       0.0      1.0     425.0              2
19762       0.0      1.0     296.0              2
```

```
[500 rows x 32 columns]>

data.tail

<bound method NDFrame.tail of        Unnamed: 0  Year  Quarter  Month
DayofMonth  DayOfWeek  FlightDate  \
12781      975274  2003        3      7          11          5  2003-
07-11
23751     1471411  1998        2      6           7          7  1998-
06-07
2796      1469316  2001        4     12           1          6  2001-
12-01
18633     1468783  1988        2      5          30          1  1988-
05-30
23187     1861806  2019        4     12           1          7  2019-
12-01
...           ...   ...      ...    ...         ...        ...
...
17188      428525  1995        2      6          17          6  1995-
06-17
24952     1300641  2018        1      3          26          1  2018-
03-26
17089     1410356  2015        1      3          16          1  2015-
03-16
11144      607318  1998        3      9          10          4  1998-
09-10
19762      301716  2005        3      7          10          7  2005-
07-10

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
12781                WN                     19393
WN
23751                US                     20355
US
2796                 UA                     19977
UA
18633                US                     20355
AL
23187                DL                     19790
DL
...                 ...                       ...
...
17188                US                     20355
US
24952                EV                     20366
EV
17089                AA                     19805
AA
```

```
11144                DL                     19790
DL
19762                OO                     20304
OO

       ...  DestWac  CRSDepTime  DepTimeBlk  CRSArrTime ArrTimeBlk
Cancelled  \
12781  ...       33        2045   2000-2059        2140  2100-2159
0.0
23751  ...       23        2035   2000-2059        2203  2200-2259
0.0
2796   ...       91        2230   2200-2259         535  0001-0559
0.0
18633  ...       91         630   0600-0659         723  0700-0759
0.0
23187  ...       85        1530   1500-1559        1757  1700-1759
0.0
...    ...      ...         ...         ...         ...        ...
...
17188  ...       23        1820   1800-1859        1926  1900-1959
0.0
24952  ...       71        1155   1100-1159        1328  1300-1359
0.0
17089  ...       74        1122   1100-1159        1325  1300-1359
0.0
11144  ...       72        1855   1800-1859        1922  1900-1959
0.0
19762  ...       41        1205   1200-1259        1229  1200-1259
0.0

       Diverted  Flights  Distance  DistanceGroup
12781       0.0      1.0     197.0              1
23751       0.0      1.0     448.0              2
2796        0.0      1.0    2615.0             11
18633       0.0      1.0     297.0              2
23187       0.0      1.0     867.0              4
...         ...      ...       ...            ...
17188       0.0      1.0     267.0              2
24952       0.0      1.0     374.0              2
17089       0.0      1.0     985.0              4
11144       0.0      1.0     425.0              2
19762       0.0      1.0     296.0              2

[500 rows x 32 columns]>

data.count

<bound method DataFrame.count of        Unnamed: 0  Year  Quarter
Month  DayofMonth  DayOfWeek  FlightDate  \
12781      975274  2003        3      7         11          5  2003-
```

```
07-11
23751     1471411  1998        2      6           7           7  1998-
06-07
2796      1469316  2001        4     12           1           6  2001-
12-01
18633     1468783  1988        2      5          30           1  1988-
05-30
23187     1861806  2019        4     12           1           7  2019-
12-01
...           ...   ...      ...    ...         ...         ...
...
17188      428525  1995        2      6          17           6  1995-
06-17
24952     1300641  2018        1      3          26           1  2018-
03-26
17089     1410356  2015        1      3          16           1  2015-
03-16
11144      607318  1998        3      9          10           4  1998-
09-10
19762      301716  2005        3      7          10           7  2005-
07-10

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
12781                WN                     19393
WN
23751                US                     20355
US
2796                 UA                     19977
UA
18633                US                     20355
AL
23187                DL                     19790
DL
...                 ...                       ...
...
17188                US                     20355
US
24952                EV                     20366
EV
17089                AA                     19805
AA
11144                DL                     19790
DL
19762                OO                     20304
OO

       ...  DestWac  CRSDepTime  DepTimeBlk  CRSArrTime ArrTimeBlk
Cancelled  \
12781  ...       33        2045   2000-2059        2140  2100-2159
```

```
0.0
23751  ...       23        2035   2000-2059        2203  2200-2259
0.0
2796   ...       91        2230   2200-2259         535  0001-0559
0.0
18633  ...       91         630   0600-0659         723  0700-0759
0.0
23187  ...       85        1530   1500-1559        1757  1700-1759
0.0
...    ...      ...         ...         ...         ...        ...
...
17188  ...       23        1820   1800-1859        1926  1900-1959
0.0
24952  ...       71        1155   1100-1159        1328  1300-1359
0.0
17089  ...       74        1122   1100-1159        1325  1300-1359
0.0
11144  ...       72        1855   1800-1859        1922  1900-1959
0.0
19762  ...       41        1205   1200-1259        1229  1200-1259
0.0

       Diverted  Flights  Distance  DistanceGroup
12781       0.0      1.0     197.0              1
23751       0.0      1.0     448.0              2
2796        0.0      1.0    2615.0             11
18633       0.0      1.0     297.0              2
23187       0.0      1.0     867.0              4
...         ...      ...       ...            ...
17188       0.0      1.0     267.0              2
24952       0.0      1.0     374.0              2
17089       0.0      1.0     985.0              4
11144       0.0      1.0     425.0              2
19762       0.0      1.0     296.0              2

[500 rows x 32 columns]>
```

```
data = data.drop(['Unnamed: 0'], axis=1)
```

```
data
```

```
       Year  Quarter  Month  DayofMonth  DayOfWeek  FlightDate  \
12781  2003        3      7          11          5  2003-07-11
23751  1998        2      6           7          7  1998-06-07
2796   2001        4     12           1          6  2001-12-01
18633  1988        2      5          30          1  1988-05-30
23187  2019        4     12           1          7  2019-12-01
...     ...      ...    ...         ...        ...         ...
17188  1995        2      6          17          6  1995-06-17
24952  2018        1      3          26          1  2018-03-26
17089  2015        1      3          16          1  2015-03-16
```

```
11144  1998        3      9          10          4  1998-09-10
19762  2005        3      7          10          7  2005-07-10

      Reporting_Airline  DOT_ID_Reporting_Airline
IATA_CODE_Reporting_Airline  \
12781                WN                     19393
WN
23751                US                     20355
US
2796                 UA                     19977
UA
18633                US                     20355
AL
23187                DL                     19790
DL
...                 ...                       ...
...
17188                US                     20355
US
24952                EV                     20366
EV
17089                AA                     19805
AA
11144                DL                     19790
DL
19762                OO                     20304
OO

       Flight_Number_Reporting_Airline  ...  DestWac  CRSDepTime
DepTimeBlk  \
12781                             2613  ...       33        2045
2000-2059
23751                              563  ...       23        2035
2000-2059
2796                                30  ...       91        2230
2200-2259
18633                             1546  ...       91         630
0600-0659
23187                             2793  ...       85        1530
1500-1559
...                                ...  ...      ...         ...
...
17188                              470  ...       23        1820
1800-1859
24952                             4388  ...       71        1155
1100-1159
17089                             2371  ...       74        1122
1100-1159
11144                              351  ...       72        1855
1800-1859
```

```
19762                           6982  ...       41        1205
1200-1259

      CRSArrTime ArrTimeBlk  Cancelled  Diverted  Flights  Distance  \
12781       2140  2100-2159        0.0       0.0      1.0     197.0
23751       2203  2200-2259        0.0       0.0      1.0     448.0
2796         535  0001-0559        0.0       0.0      1.0    2615.0
18633        723  0700-0759        0.0       0.0      1.0     297.0
23187       1757  1700-1759        0.0       0.0      1.0     867.0
...          ...        ...        ...       ...      ...       ...
17188       1926  1900-1959        0.0       0.0      1.0     267.0
24952       1328  1300-1359        0.0       0.0      1.0     374.0
17089       1325  1300-1359        0.0       0.0      1.0     985.0
11144       1922  1900-1959        0.0       0.0      1.0     425.0
19762       1229  1200-1259        0.0       0.0      1.0     296.0

      DistanceGroup
12781             1
23751             2
2796             11
18633             2
23187             4
...             ...
17188             2
24952             2
17089             4
11144             2
19762             2

[500 rows x 31 columns]
```

```
data.shape
```

```
(500, 31)
```

```
#To visualize Scatter plot for  Distance vs Departure time
fig = go.Figure(data = go.Scatter(x=df['Distance'],y =
df['DepTime'],mode='markers', marker=dict(color='red')))
fig.update_layout(title = 'Scatter Plot for Distance vs Departure
time', xaxis_title = 'Distance Time',yaxis_title ='Departure Time')
fig.show()
```

```
line_data = df.groupby('Month')['ArrDelay'].mean().reset_index()
line_data
```

```
    Month  ArrDelay
0       1  7.097321
1       2  6.530025
2       3  5.429533
3       4  4.040055
4       5  5.874889
```

```
5       6  9.347515
6       7  8.851436
7       8  7.402630
8       9  2.198975
9      10  3.242506
10     11  4.228961
11     12  9.474149
```

```
#To display Line plot of Monthly average delay plot for an year
fig = go.Figure(data = go.Scatter(x=line_data['Month'],y =
line_data['ArrDelay'],mode='lines', marker=dict(color='green')))
fig.update_layout(title = 'Monthly average delay plot for an year',
xaxis_title = 'Months',yaxis_title ='Delay of flights (in minutes)')
fig.show()
```

```
bar_data = df.groupby('DestState')['Flights'].sum().reset_index()
bar_data
```

```
   DestState  Flights
0         AK    192.0
1         AL    163.0
2         AR    118.0
3         AZ    842.0
4         CA   3174.0
5         CO    970.0
6         CT    122.0
7         FL   1804.0
8         GA   1553.0
9         HI    289.0
10        IA     84.0
11        ID     72.0
12        IL   1766.0
13        IN    192.0
14        KS     44.0
15        KY    463.0
16        LA    265.0
17        MA    484.0
18        MD    421.0
19        ME     34.0
20        MI    768.0
21        MN    591.0
22        MO    730.0
23        MS     67.0
24        MT     79.0
25        NC    907.0
26        ND     53.0
27        NE    106.0
28        NH     34.0
29        NJ    539.0
30        NM    142.0
31        NV    704.0
```

```
32         NY   1233.0
33         OH    538.0
34         OK    203.0
35         OR    298.0
36         PA    834.0
37         PR     92.0
38         RI     71.0
39         SC    118.0
40         SD     34.0
41         TN    612.0
42         TT      3.0
43         TX   3023.0
44         UT    439.0
45         VA    882.0
46         VI     16.0
47         VT     23.0
48         WA    545.0
49         WI    210.0
50         WV     20.0
51         WY     27.0
```

```
#To dispaly Bar chart of Total number of flights from each destination
states
fig = pr.bar(bar_data,x='DestState',y='Flights', title = 'Total number
of flights from each destination states')
fig.show()
```

```
bub_data = df.groupby('Reporting_Airline')
['Flights'].sum().reset_index()
bub_data
```

```
   Reporting_Airline  Flights
0                 9E    262.0
1                 AA   3151.0
2                 AS    710.0
3                 B6    499.0
4                 CO   1208.0
5                 DH     85.0
6                 DL   3530.0
7                 EA    120.0
8                 EV    895.0
9                 F9    154.0
10                FL    344.0
11                G4     36.0
12                HA    162.0
13                HP    503.0
14                KH     19.0
15           ML (1)      9.0
16                MQ   1096.0
17                NK    121.0
18                NW   1517.0
```

```
19                OH    321.0
20                OO   1518.0
21            PA (1)     41.0
22                PI    119.0
23                PS      9.0
24                TW    516.0
25                TZ     34.0
26                UA   2570.0
27                US   2342.0
28                VX     59.0
29                WN   4186.0
30                XE    469.0
31                YV    289.0
32                YX    106.0
```

```
#To dispaly Bubble chart of Total number of flights per reporting
airlines
fig = pr.scatter(bub_data,x='Reporting_Airline',y='Flights', title =
'Total number of flights per reporting airlines',size =
'Flights',hover_name='Reporting_Airline',size_max=60)
fig.show()
```