This notebook has the flight departure details from Chicago (ORD), Denver (DEN), Newark (EWR), and Washington (IAD) Airports and the flight arrival details of Syracuse(SYR) Airport

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

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

import necessary libraries for processing

```
import pandas as pd
import numpy as np
```

Read csv files

```
d1=pd.read_csv('/content/drive/MyDrive/ML Project/Detailed_Statistics_Departures-CHICAGOO.csv')
d2=pd.read_csv('/content/drive/MyDrive/ML Project/Detailed_Statistics_Departures-DEN.csv')
d3=pd.read_csv('/content/drive/MyDrive/ML Project/Detailed_Statistics_Departures-EWR.csv')
d4=pd.read_csv('/content/drive/MyDrive/ML Project/Detailed_Statistics_Departures-IAD.csv')
d5=pd.read_csv('/content/drive/MyDrive/ML Project/Detailed_Statistics_Arrivals-2.csv')
```

```
d1
```

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time
0	UA	1/1/22	66.0	N26906	KOA	9:40	9:37
1	UA	1/1/22	202.0	N27901	OGG	8:50	8:51
2	UA	1/1/22	219.0	N2140U	HNL	10:00	10:00
3	UA	1/1/22	221.0	N15710	EGE	9:00	9:04
4	UA	1/1/22	224.0	NaN	EWR	12:54	0:00
...	...	...	...	...	...	...	...
69088	UA	12/31/22	2668.0	N14242	FLL	10:35	10:29
69089	UA	12/31/22	2677.0	N17245	IAD	13:28	13:56
69090	UA	12/31/22	2678.0	N39423	SFO	6:34	9:58
69091	NaN	NaN	NaN	NaN	NaN	NaN	NaN
69092	SOURCE: Bureau of Transportation Statistics	NaN	NaN	NaN	NaN	NaN	NaN

69093 rows × 17 columns

```
d2
```

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Tax (Mir
0	UA	1/1/22	243.0	NaN	IAH	17:25	0:00	145.0	0.0	0.0	0:00	
1	UA	1/1/22	250.0	N34131	IAH	10:00	10:47	147.0	220.0	47.0	12:16	

d3

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Tax (Mir
0	UA	1/1/22	42.0	N665UA	OGG	8:30	8:37	670.0	710.0	7.0	9:02	
1	UA	1/1/22	205.0	N69888	SJU	12:00	12:41	225.0	202.0	41.0	12:51	
2	UA	1/1/22	215.0	N440UA	DFW	17:29	18:07	252.0	261.0	38.0	18:25	
3	UA	1/1/22	238.0	N873UA	FLL	10:59	11:07	196.0	167.0	8.0	11:24	
4	UA	1/1/22	239.0	N69804	RSW	21:15	22:10	195.0	187.0	55.0	22:44	
...	...	...	...	...	...	...	...	...	...	...	...	
61585	UA	12/31/22	2631.0	N67827	DFW	10:02	10:03	248.0	258.0	1.0	10:48	
61586	UA	12/31/22	2667.0	N77259	MSY	12:40	12:39	205.0	205.0	-1.0	12:58	
61587	UA	12/31/22	2674.0	N76515	IAH	11:53	11:57	248.0	243.0	4.0	12:14	
61588	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
61589	SOURCE: Bureau of Transportation Statistics	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

61590 rows × 17 columns

d4

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Tax (Mir
0	UA	1/1/22	277.0	N818UA	SFO	6:35	6:55	374.0	356.0	20.0	7:08	
1	UA	1/1/22	291.0	N69824	SMF	19:18	19:38	362.0	333.0	20.0	19:53	
2	UA	1/1/22	299.0	N34460	ORF	22:20	22:32	60.0	65.0	12.0	22:58	
3	UA	1/1/22	326.0	N66803	SEA	17:39	21:49	360.0	338.0	250.0	22:04	
4	UA	1/1/22	345.0	N66056	HNL	10:40	11:03	650.0	649.0	23.0	11:15	
...	...	...	...	...	...	...	...	...	...	...	...	
29570	UA	12/31/22	2639.0	N810UA	BDL	12:35	12:31	77.0	74.0	-4.0	12:46	
29571	UA	12/31/22	2642.0	N76532	ATL	8:35	8:34	118.0	129.0	-1.0	8:57	
29572	UA	12/31/22	2657.0	N415UA	EWR	8:41	8:40	79.0	73.0	-1.0	9:01	
29573	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
29574	SOURCE: Bureau of Transportation Statistics	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

29575 rows × 17 columns

Check for Null Values

```
d1.isna().sum()
```

```
Carrier Code                        1
Date (MM/DD/YYYY)                   2
Flight Number                       2
Tail Number                      1236
```

```
Destination Airport                          2
Scheduled departure time                     2
Actual departure time                        2
Scheduled elapsed time (Minutes)             2
Actual elapsed time (Minutes)                2
Departure delay (Minutes)                    2
Wheels-off time                              2
Taxi-Out time (Minutes)                      2
Delay Carrier (Minutes)                      2
Delay Weather (Minutes)                      2
Delay National Aviation System (Minutes)     2
Delay Security (Minutes)                     2
Delay Late Aircraft Arrival (Minutes)        2
dtype: int64
```

```
d2.isna().sum()
```

```
Carrier Code                                 1
Date (MM/DD/YYYY)                            2
Flight Number                                2
Tail Number                               1042
Destination Airport                          2
Scheduled departure time                     2
Actual departure time                        2
Scheduled elapsed time (Minutes)             2
Actual elapsed time (Minutes)                2
Departure delay (Minutes)                    2
Wheels-off time                              2
Taxi-Out time (Minutes)                      2
Delay Carrier (Minutes)                      2
Delay Weather (Minutes)                      2
Delay National Aviation System (Minutes)     2
Delay Security (Minutes)                     2
Delay Late Aircraft Arrival (Minutes)        2
dtype: int64
```

```
d3.isna().sum()
```

```
Carrier Code                                 1
Date (MM/DD/YYYY)                            2
Flight Number                                2
Tail Number                               2167
Destination Airport                          2
Scheduled departure time                     2
Actual departure time                        2
Scheduled elapsed time (Minutes)             2
Actual elapsed time (Minutes)                2
Departure delay (Minutes)                    2
Wheels-off time                              2
Taxi-Out time (Minutes)                      2
Delay Carrier (Minutes)                      2
Delay Weather (Minutes)                      2
Delay National Aviation System (Minutes)     2
Delay Security (Minutes)                     2
Delay Late Aircraft Arrival (Minutes)        2
dtype: int64
```

```
d4.isna().sum()
```

```
Carrier Code                                 1
Date (MM/DD/YYYY)                            2
Flight Number                                2
Tail Number                                485
Destination Airport                          2
Scheduled departure time                     2
Actual departure time                        2
Scheduled elapsed time (Minutes)             2
Actual elapsed time (Minutes)                2
Departure delay (Minutes)                    2
Wheels-off time                              2
Taxi-Out time (Minutes)                      2
Delay Carrier (Minutes)                      2
Delay Weather (Minutes)                      2
Delay National Aviation System (Minutes)     2
Delay Security (Minutes)                     2
Delay Late Aircraft Arrival (Minutes)        2
dtype: int64
```

Remove all null values

```
d11=d1.dropna()
```

```
d22=d2.dropna()
```

```
d33=d3.dropna()

d44=d4.dropna()

d55=d5.dropna()
```

Combine the departure dataframes using concatenation

```
flightsinfo = pd.concat([d11,d22,d33,d44], axis=0)

flightsinfo
```

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Ou tim (Minutes
0	UA	1/1/22	66.0	N26906	KOA	9:40	9:37	560.0	537.0	-3.0	10:11	34.(
1	UA	1/1/22	202.0	N27901	OGG	8:50	8:51	545.0	522.0	1.0	9:09	18.(
2	UA	1/1/22	219.0	N2140U	HNL	10:00	10:00	558.0	547.0	0.0	10:20	20.(
3	UA	1/1/22	221.0	N15710	EGE	9:00	9:04	180.0	0.0	4.0	9:41	37.(
5	UA	1/1/22	262.0	N429UA	SEA	19:50	23:07	288.0	281.0	197.0	23:29	22.(
...	...	...	...	...	...	...	...	...	...	...	...	.
29568	UA	12/31/22	2608.0	N17245	ORD	17:31	18:00	133.0	121.0	29.0	18:20	20.(
29569	UA	12/31/22	2618.0	N66831	SFO	22:15	22:12	366.0	355.0	-3.0	22:27	15.(
29570	UA	12/31/22	2639.0	N810UA	BDL	12:35	12:31	77.0	74.0	-4.0	12:46	15.(
29571	UA	12/31/22	2642.0	N76532	ATL	8:35	8:34	118.0	129.0	-1.0	8:57	23.(
29572	UA	12/31/22	2657.0	N415UA	EWR	8:41	8:40	79.0	73.0	-1.0	9:01	21.(

236434 rows × 17 columns

Filter rows for flights going to Syracuse Airport

```
flightsinfo = flightsinfo[flightsinfo['Destination Airport'].str.contains('SYR')]
flightsinfo.head(20)
```

	Carrier Code	Date (MM/DD/YYYY)	Flight Number	Tail Number	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Out time (Minutes)
6034	UA	2/11/22	1094.0	N489UA	SYR	18:10	18:39	109.0	109.0	29.0	19:06	27.0
6195	UA	2/12/22	1094.0	N838UA	SYR	18:10	19:09	109.0	98.0	59.0	19:25	16.0

Merge the departure dataframe and arrival dataframe on Date and Flight number

```
df=pd.merge(flightsinfo,d5,on=['Date (MM/DD/YYYY)','Flight Number'])
```

```
df
```

	Carrier Code_x	Date (MM/DD/YYYY)	Flight Number	Tail Number_x	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	...	Scheduled Elapsed Time (Minutes)	(
0	UA	2/11/22	1094.0	N489UA	SYR	18:10	18:39	109.0	109.0	29.0	...	109.0	
1	UA	2/12/22	1094.0	N838UA	SYR	18:10	19:09	109.0	98.0	59.0	...	109.0	
2	UA	2/13/22	1094.0	N880UA	SYR	18:10	18:08	109.0	104.0	-2.0	...	109.0	
3	UA	2/14/22	1094.0	N823UA	SYR	18:10	18:08	109.0	99.0	-2.0	...	109.0	
4	UA	2/15/22	1094.0	N825UA	SYR	18:10	18:07	109.0	105.0	-3.0	...	109.0	
...	...	...	...	...	...	...	...	...	...	...	...	...	
1017	UA	10/25/22	2198.0	N873UA	SYR	22:20	22:13	71.0	67.0	-7.0	...	71.0	
1018	UA	10/26/22	2198.0	N845UA	SYR	22:20	22:16	71.0	67.0	-4.0	...	71.0	
1019	UA	10/27/22	2198.0	N813UA	SYR	22:20	22:14	71.0	68.0	-6.0	...	71.0	
1020	UA	10/28/22	2198.0	N845UA	SYR	22:20	22:07	71.0	63.0	-13.0	...	71.0	
1021	UA	10/29/22	2198.0	N447UA	SYR	22:20	22:10	71.0	66.0	-10.0	...	71.0	

1022 rows × 32 columns

```
df.columns
```

```
Index(['Carrier Code_x', 'Date (MM/DD/YYYY)', 'Flight Number', 'Tail Number_x',
       'Destination Airport', 'Scheduled departure time',
       'Actual departure time', 'Scheduled elapsed time (Minutes)',
       'Actual elapsed time (Minutes)', 'Departure delay (Minutes)',
       'Wheels-off time', 'Taxi-Out time (Minutes)',
       'Delay Carrier (Minutes)_x', 'Delay Weather (Minutes)_x',
       'Delay National Aviation System (Minutes)_x',
       'Delay Security (Minutes)_x', 'Delay Late Aircraft Arrival (Minutes)_x',
       'Carrier Code_y', 'Tail Number_y', 'Origin Airport',
       'Scheduled Arrival Time', 'Actual Arrival Time',
       'Scheduled Elapsed Time (Minutes)', 'Actual Elapsed Time (Minutes)',
       'Arrival Delay (Minutes)', 'Wheels-on Time', 'Taxi-In time (Minutes)',
       'Delay Carrier (Minutes)_y', 'Delay Weather (Minutes)_y',
       'Delay National Aviation System (Minutes)_y',
       'Delay Security (Minutes)_y',
       'Delay Late Aircraft Arrival (Minutes)_y'],
      dtype='object')
```

Remove duplicate columns

```
df = df.T.drop_duplicates().T
```

```
df
```

	Carrier Code_x	Date (MM/DD/YYYY)	Flight Number	Tail Number_x	Destination Airport	Scheduled departure time	Actual departure time	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	...	Delay Weather (Minutes)_x
0	UA	2/11/22	1094.0	N489UA	SYR	18:10	18:39	109.0	109.0	29.0	...	0.0
1	UA	2/12/22	1094.0	N838UA	SYR	18:10	19:09	109.0	98.0	59.0	...	0.0
2	UA	2/13/22	1094.0	N880UA	SYR	18:10	18:08	109.0	104.0	-2.0	...	0.0
3	UA	2/14/22	1094.0	N823UA	SYR	18:10	18:08	109.0	99.0	-2.0	...	0.0
4	UA	2/15/22	1094.0	N825UA	SYR	18:10	18:07	109.0	105.0	-3.0	...	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...

```
df.columns
```

```
Index(['Carrier Code_x', 'Date (MM/DD/YYYY)', 'Flight Number', 'Tail Number_x',
       'Destination Airport', 'Scheduled departure time',
       'Actual departure time', 'Scheduled elapsed time (Minutes)',
       'Actual elapsed time (Minutes)', 'Departure delay (Minutes)',
       'Wheels-off time', 'Taxi-Out time (Minutes)',
       'Delay Carrier (Minutes)_x', 'Delay Weather (Minutes)_x',
       'Delay National Aviation System (Minutes)_x',
       'Delay Security (Minutes)_x', 'Delay Late Aircraft Arrival (Minutes)_x',
       'Origin Airport', 'Scheduled Arrival Time', 'Actual Arrival Time',
       'Arrival Delay (Minutes)', 'Wheels-on Time', 'Taxi-In time (Minutes)'],
      dtype='object')
```

Drop unnecessary columns

```
df=df.drop(columns=['Scheduled Arrival Time', 'Actual Arrival Time','Carrier Code_x','Tail Number_x','Scheduled departure time',
       'Actual departure time'])
```

```
df
```

	Date (MM/DD/YYYY)	Flight Number	Destination Airport	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Out time (Minutes)	Delay Carrier (Minutes)_x	Delay Weather (Minutes)_x	Delay National Aviation System (Minutes)_x
0	2/11/22	1094.0	SYR	109.0	109.0	29.0	19:06	27.0	29.0	0.0	0.0
1	2/12/22	1094.0	SYR	109.0	98.0	59.0	19:25	16.0	39.0	0.0	0.0
2	2/13/22	1094.0	SYR	109.0	104.0	-2.0	18:26	18.0	0.0	0.0	0.0
3	2/14/22	1094.0	SYR	109.0	99.0	-2.0	18:24	16.0	0.0	0.0	0.0
4	2/15/22	1094.0	SYR	109.0	105.0	-3.0	18:36	29.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...
1017	10/25/22	2198.0	SYR	71.0	67.0	-7.0	22:28	15.0	0.0	0.0	0.0
1018	10/26/22	2198.0	SYR	71.0	67.0	-4.0	22:30	14.0	0.0	0.0	0.0
1019	10/27/22	2198.0	SYR	71.0	68.0	-6.0	22:31	17.0	0.0	0.0	0.0
1020	10/28/22	2198.0	SYR	71.0	63.0	-13.0	22:25	18.0	0.0	0.0	0.0
1021	10/29/22	2198.0	SYR	71.0	66.0	-10.0	22:25	15.0	0.0	0.0	0.0

1022 rows × 17 columns

Change the date column into seperate day month year columns

```
df[["month", "day", "year"]] = df["Date (MM/DD/YYYY)"].str.split("/", expand = True)
```

```
df
```

	Date (MM/DD/YYYY)	Flight Number	Destination Airport	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Out time (Minutes)	Delay Carrier (Minutes)_x	Delay Weather (Minutes)_x	Delay National Aviation System (Minutes)_x
0	2/11/22	1094.0	SYR	109.0	109.0	29.0	19:06	27.0	29.0	0.0	0.0
1	2/12/22	1094.0	SYR	109.0	98.0	59.0	19:25	16.0	39.0	0.0	0.0
2	2/13/22	1094.0	SYR	109.0	104.0	-2.0	18:26	18.0	0.0	0.0	0.0
3	2/14/22	1094.0	SYR	109.0	99.0	-2.0	18:24	16.0	0.0	0.0	0.0
4	2/15/22	1094.0	SYR	109.0	105.0	-3.0	18:36	29.0	0.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...
1017	10/25/22	2198.0	SYR	71.0	67.0	-7.0	22:28	15.0	0.0	0.0	0.0

```
df.isna().sum()
```

```
Date (MM/DD/YYYY)                           0
Flight Number                               0
Destination Airport                         0
Scheduled elapsed time (Minutes)            0
Actual elapsed time (Minutes)               0
Departure delay (Minutes)                   0
Wheels-off time                             0
Taxi-Out time (Minutes)                     0
Delay Carrier (Minutes)_x                   0
Delay Weather (Minutes)_x                   0
Delay National Aviation System (Minutes)_x  0
Delay Security (Minutes)_x                  0
Delay Late Aircraft Arrival (Minutes)_x     0
Origin Airport                              0
Arrival Delay (Minutes)                     0
Wheels-on Time                              0
Taxi-In time (Minutes)                      0
month                                       0
day                                         0
year                                        0
dtype: int64
```

```
df.describe()
```

	Date (MM/DD/YYYY)	Flight Number	Destination Airport	Scheduled elapsed time (Minutes)	Actual elapsed time (Minutes)	Departure delay (Minutes)	Wheels-off time	Taxi-Out time (Minutes)	Delay Carrier (Minutes)_x	Delay Weather (Minutes)_x	Dela Nationa Aviatio Syste (Minutes)_
count	1022	1022.0	1022	1022.0	1022.0	1022.0	1022	1022.0	1022.0	1022.0	1022.
unique	363	38.0	1	28.0	143.0	125.0	353	50.0	69.0	15.0	37.
top	10/28/22	604.0	SYR	114.0	67.0	-3.0	10:01	14.0	0.0	0.0	0.
freq	5	210.0	1022	133.0	35.0	79.0	21	89.0	873.0	1008.0	928.

Convert the final dataframe to csv file

```
df.to_csv('flightinfo.csv')
```