This notebook contains Weather data for four cities: Chicago (ORD), Denver (DEN), Newark (EWR), and Washington (IAD)

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

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

import necessary libraries

```
import pandas as pd
```

Read all the csv files into dataframe

FOR WASHINGTON

```
d1=pd.read_csv('/content/drive/MyDrive/ML Project/washington.csv')
```

```
d1
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike
0	IAD	1/1/22	17.5	12.3	14.6	17.5	12.3	14.6
1	IAD	2/1/22	17.1	8.8	14.2	17.1	5.5	13.9
2	IAD	3/1/22	7.8	-3.3	0.4	4.1	-9.9	-5.0
3	IAD	4/1/22	1.1	-3.7	-1.5	0.9	-8.9	-4.4
4	IAD	5/1/22	4.9	-2.2	2.1	2.9	-6.8	-1.4
...	...	...	...	...	...	...	...	...
360	IAD	27/12/22	6.1	-1.4	1.7	6.1	-3.2	0.7
361	IAD	28/12/22	10.0	-1.8	2.5	10.0	-4.7	0.6
362	IAD	29/12/22	11.3	-1.3	4.7	11.3	-1.6	3.2
363	IAD	30/12/22	17.9	0.6	7.8	17.9	-0.4	7.4
364	IAD	31/12/22	13.4	6.1	10.6	13.4	4.7	10.4

365 rows × 33 columns

split datetime column into day month year columns

```
d1[["day", "month", "year"]] = d1["datetime"].str.split("/", expand = True)
```

```
d1
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	h
0	IAD	1/1/22	17.5	12.3	14.6	17.5	12.3	14.6	13.3	
1	IAD	2/1/22	17.1	8.8	14.2	17.1	5.5	13.9	10.9	
2	IAD	3/1/22	7.8	-3.3	0.4	4.1	-9.9	-5.0	-3.0	
3	IAD	4/1/22	1.1	-3.7	-1.5	0.9	-8.9	-4.4	-7.1	
4	IAD	5/1/22	4.9	-2.2	2.1	2.9	-6.8	-1.4	-1.7	
...	...	...	...	...	...	...	...	...	...	
360	IAD	27/12/22	6.1	-1.4	1.7	6.1	-3.2	0.7	-6.9	
361	IAD	28/12/22	10.0	-1.8	2.5	10.0	-4.7	0.6	-4.0	
362	IAD	29/12/22	11.3	-1.3	4.7	11.3	-1.6	3.2	-1.7	
363	IAD	30/12/22	17.9	0.6	7.8	17.9	-0.4	7.4	3.1	
364	IAD	31/12/22	13.4	6.1	10.6	13.4	4.7	10.4	10.2	

365 rows × 36 columns

drop unnecessary columns

```
d1=d1.drop(columns=['datetime','stations', 'description', 'sunrise', 'sunset','conditions','preciptype'])
```

check for null values

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

```
Origin Airport      0
tempmax             0
tempmin             0
temp                0
feelslikemax        0
feelslikemin        0
feelslike           0
dew                 0
humidity            0
precip              0
precipprob          0
precipcover         0
snow                0
snowdepth           0
windgust            3
windspeed           0
winddir             0
sealevelpressure    0
cloudcover          0
visibility          0
```

```
solarradiation      0
solarenergy         0
uvindex             0
severerisk          9
moonphase           0
icon                0
day                 0
month               0
year                0
dtype: int64
```

drop null values

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

FOR DENVER

```
d2=pd.read_csv('/content/drive/MyDrive/ML Project/denver .csv')
```

```
d2
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	DEN	1/1/22	-11.7	-18.0	-15.1	-13.5	-26.5	-20.2	-17.5
1	DEN	1/2/22	5.1	-14.7	-4.1	2.0	-22.5	-9.7	-12.6
2	DEN	1/3/22	10.2	-3.9	2.5	10.2	-9.1	-0.9	-9.7
3	DEN	1/4/22	8.6	-5.2	2.7	4.7	-11.0	-1.4	-10.9
4	DEN	1/5/22	3.7	-17.0	-7.8	1.0	-24.1	-13.3	-11.2
...	...	...	...	...	...	...	...	...	...
360	DEN	12/27/22	16.9	4.2	9.7	16.9	0.4	7.9	-5.6
361	DEN	12/28/22	9.7	-1.1	5.1	9.7	-6.3	2.2	-1.9
362	DEN	12/29/22	2.9	-6.3	-2.4	2.9	-13.5	-7.0	-7.1
363	DEN	12/30/22	3.6	-8.1	-2.2	1.8	-15.0	-6.8	-10.4
364	DEN	12/31/22	11.5	-2.2	3.2	11.5	-6.6	0.1	-4.5

365 rows × 33 columns

```
d2[["month", "day", "year"]] = d2["datetime"].str.split("/", expand = True)
```

```
d2
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	DEN	1/1/22	-11.7	-18.0	-15.1	-13.5	-26.5	-20.2	-17.5
1	DEN	1/2/22	5.1	-14.7	-4.1	2.0	-22.5	-9.7	-12.6
2	DEN	1/3/22	10.2	-3.9	2.5	10.2	-9.1	-0.9	-9.7
3	DEN	1/4/22	8.6	-5.2	2.7	4.7	-11.0	-1.4	-10.9
4	DEN	1/5/22	3.7	-17.0	-7.8	1.0	-24.1	-13.3	-11.2
...	...	...	...	...	...	...	...	...	...
360	DEN	12/27/22	16.9	4.2	9.7	16.9	0.4	7.9	-5.6
361	DEN	12/28/22	9.7	-1.1	5.1	9.7	-6.3	2.2	-1.9
362	DEN	12/29/22	2.9	-6.3	-2.4	2.9	-13.5	-7.0	-7.1
363	DEN	12/30/22	3.6	-8.1	-2.2	1.8	-15.0	-6.8	-10.4
364	DEN	12/31/22	11.5	-2.2	3.2	11.5	-6.6	0.1	-4.5

365 rows × 36 columns

```
d2=d2.drop(columns=['datetime','stations', 'description', 'sunrise', 'sunset','conditions','preciptype'])
```

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

```
Origin Airport      0
tempmax             0
tempmin             0
temp                0
feelslikemax        0
feelslikemin        0
feelslike           0
dew                 0
humidity            0
precip              0
precipprob          0
precipcover         0
snow                0
snowdepth           0
windgust            3
windspeed           0
winddir             0
sealevelpressure    0
cloudcover          0
visibility          0
solarradiation      0
solarenergy         0
uvindex             0
severerisk          9
```

```
moonphase           0
icon                0
month               0
day                 0
year                0
dtype: int64
```

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

FOR NEWARK

```
d3=pd.read_csv('/content/drive/MyDrive/ML Project/Newark.csv')
```

```
d3
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	EWR	1/1/22	12.9	9.4	11.2	12.9	8.8	11.1	10.8
1	EWR	1/2/22	14.4	2.9	10.4	14.4	-2.4	9.3	8.0
2	EWR	1/3/22	2.2	-5.1	-1.4	-2.1	-11.9	-7.8	-10.1
3	EWR	1/4/22	1.2	-7.7	-3.2	-1.1	-14.3	-7.5	-12.1
4	EWR	1/5/22	7.2	-3.2	1.9	5.3	-6.7	-0.7	-0.1
...	...	...	...	...	...	...	...	...	...
360	EWR	12/27/22	2.1	-2.1	0.1	-1.3	-7.0	-4.2	-10.1
361	EWR	12/28/22	9.2	-1.6	3.8	6.2	-4.2	1.3	-5.4
362	EWR	12/29/22	11.0	0.6	5.4	11.0	-2.3	3.5	-3.1
363	EWR	12/30/22	16.6	4.2	9.4	16.6	2.3	8.7	0.8
364	EWR	12/31/22	12.0	7.6	10.4	12.0	6.4	10.2	9.0

365 rows × 33 columns

```
d3[["month", "day", "year"]] = d3["datetime"].str.split("/", expand = True)
```

```
d3
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	EWR	1/1/22	12.9	9.4	11.2	12.9	8.8	11.1	10.8
1	EWR	1/2/22	14.4	2.9	10.4	14.4	-2.4	9.3	8.0
2	EWR	1/3/22	2.2	-5.1	-1.4	-2.1	-11.9	-7.8	-10.1
3	EWR	1/4/22	1.2	-7.7	-3.2	-1.1	-14.3	-7.5	-12.1
4	EWR	1/5/22	7.2	-3.2	1.9	5.3	-6.7	-0.7	-0.1
...	...	...	...	...	...	...	...	...	...
360	EWR	12/27/22	2.1	-2.1	0.1	-1.3	-7.0	-4.2	-10.1
361	EWR	12/28/22	9.2	-1.6	3.8	6.2	-4.2	1.3	-5.4
362	EWR	12/29/22	11.0	0.6	5.4	11.0	-2.3	3.5	-3.1
363	EWR	12/30/22	16.6	4.2	9.4	16.6	2.3	8.7	0.8
364	EWR	12/31/22	12.0	7.6	10.4	12.0	6.4	10.2	9.0

365 rows × 36 columns

```
d3=d3.drop(columns=['datetime','stations', 'description', 'sunrise', 'sunset','conditions','preciptype'])
```

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

```
Origin Airport      0
tempmax             0
tempmin             0
temp                0
feelslikemax        0
feelslikemin        0
feelslike           0
dew                 0
humidity            0
precip              0
precipprob          0
precipcover         0
snow                0
snowdepth           0
windgust            2
windspeed           0
winddir             0
sealevelpressure    0
cloudcover          0
visibility          0
solarradiation      0
solarenergy         0
uvindex             0
severerisk          9
moonphase           0
icon                0
month               0
day                 0
year                0
dtype: int64
```

```
d3=d3.dropna()
```

FOR CHICAGO

```
d4=pd.read_csv('/content/drive/MyDrive/ML Project/Chicago,United States 2022-01-01 to 2022-12-31.csv')
```

```
d4
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	ORD	1/1/22	5.0	-1.3	1.6	0.8	-8.5	-4.3	-1.3
1	ORD	1/2/22	-1.6	-9.0	-4.7	-6.3	-14.0	-10.6	-8.9
2	ORD	1/3/22	-4.7	-12.0	-8.2	-10.5	-17.9	-13.9	-14.6
3	ORD	1/4/22	1.9	-4.8	-1.2	-3.8	-10.7	-6.8	-6.9
4	ORD	1/5/22	1.5	-10.3	-6.8	-4.3	-21.3	-15.7	-12.0
...	...	...	...	...	...	...	...	...	...
360	ORD	12/27/22	-3.7	-13.2	-8.2	-11.4	-19.4	-14.9	-12.2
361	ORD	12/28/22	7.2	-3.7	2.6	2.8	-10.8	-2.8	-3.1
362	ORD	12/29/22	14.2	6.6	10.8	14.2	2.5	9.4	7.6
363	ORD	12/30/22	13.9	0.1	6.1	13.9	-2.9	4.0	2.6
364	ORD	12/31/22	3.7	-2.5	0.5	3.1	-5.3	-1.2	-3.0

365 rows × 33 columns

```
d4[["month", "day", "year"]] = d4["datetime"].str.split("/", expand = True)
```

```
d4
```

	Origin Airport	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew
0	ORD	1/1/22	5.0	-1.3	1.6	0.8	-8.5	-4.3	-1.3
1	ORD	1/2/22	-1.6	-9.0	-4.7	-6.3	-14.0	-10.6	-8.9
2	ORD	1/3/22	-4.7	-12.0	-8.2	-10.5	-17.9	-13.9	-14.6
3	ORD	1/4/22	1.9	-4.8	-1.2	-3.8	-10.7	-6.8	-6.9
4	ORD	1/5/22	1.5	-10.3	-6.8	-4.3	-21.3	-15.7	-12.0
...	...	...	...	...	...	...	...	...	...
360	ORD	12/27/22	-3.7	-13.2	-8.2	-11.4	-19.4	-14.9	-12.2
361	ORD	12/28/22	7.2	-3.7	2.6	2.8	-10.8	-2.8	-3.1
362	ORD	12/29/22	14.2	6.6	10.8	14.2	2.5	9.4	7.6

```
d4=d4.drop(columns=['datetime','stations', 'description', 'sunrise', 'sunset','conditions','preciptype'])
```

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

```
Origin Airport      0
tempmax             0
tempmin             0
temp                0
feelslikemax        0
feelslikemin        0
feelslike           0
dew                 0
humidity            0
precip              0
precipprob          0
precipcover         0
snow                0
snowdepth           0
windgust            1
windspeed           0
winddir             0
sealevelpressure    0
cloudcover          0
visibility          0
solarradiation      0
solarenergy         0
uvindex             0
severerisk          9
moonphase           0
icon                0
month               0
day                 0
year                0
dtype: int64
```

```
d4=d4.dropna()
```

Combine dataframes for all cities into one

```
df = pd.concat([d1,d2,d3,d4], axis=0)
```

```
df
```

	Origin Airport	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity
9	IAD	6.7	-1.7	1.6	2.4	-8.3	-3.6	-9.8	43.1
10	IAD	-1.8	-5.0	-3.5	-5.3	-12.7	-8.5	-15.0	41.1
11	IAD	8.3	-5.1	1.4	5.4	-10.5	-2.5	-8.5	49.1
12	IAD	10.4	-0.3	4.3	10.4	-0.3	3.5	-2.5	63.1
13	IAD	9.2	2.0	5.5	5.3	-3.5	1.6	-2.0	59.0
...	...	...	...	...	...	...	...	...	...
360	ORD	-3.7	-13.2	-8.2	-11.4	-19.4	-14.9	-12.2	73.7
361	ORD	7.2	-3.7	2.6	2.8	-10.8	-2.8	-3.1	67.0

```
df.describe()
```

	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	
count	1424.000000	1424.000000	1424.000000	1424.000000	1424.000000	1424.000000	1424.00(
mean	18.304284	8.110604	13.026896	17.564537	5.834129	11.597121	3.71
std	11.027321	10.582234	10.575506	12.355199	12.991435	12.476315	10.67
min	-20.600000	-28.100000	-24.100000	-30.600000	-37.500000	-33.800000	-28.90(
25%	10.000000	0.200000	4.700000	9.975000	-3.800000	2.000000	-5.00(
50%	19.300000	8.600000	13.900000	19.300000	6.500000	13.500000	3.90(
75%	27.700000	17.525000	22.500000	27.525000	17.525000	22.500000	13.00(
max	38.100000	28.400000	32.500000	41.500000	30.800000	35.500000	23.50(

8 rows × 24 columns

Convert the dataframe into csv file

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