```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from matplotlib import style

df = pd.read_csv(r"D:\cwh\EDA\airline_delay.csv")
df
```

```
      year  month carrier            carrier_name airport  \
0     2020     12      9E       Endeavor Air Inc.     ABE
1     2020     12      9E       Endeavor Air Inc.     ABY
2     2020     12      9E       Endeavor Air Inc.     AEX
3     2020     12      9E       Endeavor Air Inc.     AGS
4     2020     12      9E       Endeavor Air Inc.     ALB
...    ...    ...     ...                     ...     ...
3346  2019     12      EV  ExpressJet Airlines LLC     AMA
3347  2019     12      EV  ExpressJet Airlines LLC     ATL
3348  2019     12      EV  ExpressJet Airlines LLC     AUS
3349  2019     12      EV  ExpressJet Airlines LLC     AVL
3350  2019     12      EV  ExpressJet Airlines LLC     AZO

                                             airport_name
arr_flights  \
0        Allentown/Bethlehem/Easton, PA: Lehigh Valley ...        44.0

1                  Albany, GA: Southwest Georgia Regional        90.0

2                  Alexandria, LA: Alexandria International        88.0

3              Augusta, GA: Augusta Regional at Bush Field       184.0

4                        Albany, NY: Albany International        76.0

...                                                   ...         ...

3346  Amarillo, TX: Rick Husband Amarillo International        56.0

3347  Atlanta, GA: Hartsfield-Jackson Atlanta Intern...        76.0

3348         Austin, TX: Austin - Bergstrom International         7.0

3349                     Asheville, NC: Asheville Regional        12.0

3350  Kalamazoo, MI: Kalamazoo/Battle Creek Internat...         9.0


      arr_del15  carrier_ct  weather_ct  ...  security_ct
late_aircraft_ct  \
0           3.0        1.63        0.00  ...          0.0
1.25
1           1.0        0.96        0.00  ...          0.0
```

```
0.00
2             8.0           5.75       0.00  ...           0.0
0.65
3             9.0           4.17       0.00  ...           0.0
3.00
4            11.0           4.78       0.00  ...           0.0
1.00
...           ...            ...        ...  ...           ...
...
3346          8.0           2.20       1.00  ...           0.0
0.80
3347         17.0           7.51       0.24  ...           0.0
5.12
3348          1.0           0.00       0.48  ...           0.0
0.00
3349          1.0           0.00       0.00  ...           0.0
0.00
3350          1.0           0.87       0.00  ...           0.0
0.00

      arr_cancelled  arr_diverted  arr_delay  carrier_delay  \
weather_delay  \
0               0.0           1.0       89.0           56.0
0.0
1               0.0           0.0       23.0           22.0
0.0
2               0.0           1.0      338.0          265.0
0.0
3               0.0           0.0      508.0          192.0
0.0
4               1.0           0.0      692.0          398.0
0.0
...             ...           ...        ...            ...
...
3346            0.0           1.0      353.0          165.0
19.0
3347            0.0           0.0     1880.0         1516.0
25.0
3348            0.0           0.0       96.0            0.0
46.0
3349            0.0           0.0       23.0            0.0
0.0
3350            0.0           0.0       23.0           20.0
0.0

      nas_delay  security_delay  late_aircraft_delay
0           3.0             0.0                 30.0
1           1.0             0.0                  0.0
2          45.0             0.0                 28.0
```

```
3            92.0              0.0              224.0
4           178.0              0.0              116.0
...           ...              ...                ...
3346        135.0              0.0               34.0
3347        200.0              0.0              139.0
3348         50.0              0.0                0.0
3349         23.0              0.0                0.0
3350          3.0              0.0                0.0

[3351 rows x 21 columns]


df.describe()

              year    month    arr_flights     arr_del15    carrier_ct  \
count  3351.000000   3351.0    3343.000000   3343.000000   3343.000000
mean   2019.459266     12.0     298.271014     50.995214     16.065337
std       0.498412      0.0     852.436335    146.484456     41.759516
min    2019.000000     12.0       1.000000      0.000000      0.000000
25%    2019.000000     12.0      35.000000      5.000000      1.490000
50%    2019.000000     12.0      83.000000     12.000000      4.750000
75%    2020.000000     12.0     194.500000     33.000000     12.255000
max    2020.000000     12.0   19713.000000   2289.000000    697.000000

        weather_ct          nas_ct   security_ct   late_aircraft_ct
arr_cancelled  \
count   3343.000000    3343.000000   3343.000000        3343.000000
3343.000000
mean       1.443144      16.183383      0.137320          17.166069
2.884535
std        4.821657      56.423008      0.646479          55.447043
10.126658
min        0.000000       0.000000      0.000000           0.000000
0.000000
25%        0.000000       0.820000      0.000000           0.900000
0.000000
50%        0.060000       2.980000      0.000000           3.280000
0.000000
75%        1.010000       8.870000      0.000000          10.240000
2.000000
max       89.420000    1039.540000     17.310000         819.660000
224.000000

        arr_diverted       arr_delay   carrier_delay   weather_delay
nas_delay  \
count   3343.000000     3343.000000     3343.000000     3343.000000
3343.00000
mean       0.575830     3333.868083     1144.763087      177.591385
749.57942
```

```
std          2.097884    10284.926623     3371.103512      734.343542
3190.50916
min          0.000000        0.000000        0.000000        0.000000
0.00000
25%          0.000000      230.000000       68.500000        0.000000
21.50000
50%          0.000000      746.000000      272.000000        3.000000
106.00000
75%          0.000000     2095.500000      830.500000       82.000000
362.00000
max         42.000000   160383.000000    55215.000000    14219.000000
82064.00000

       security_delay  late_aircraft_delay
count     3343.000000          3343.000000
mean         5.400838          1256.533353
std         27.161402          4184.451426
min          0.000000             0.000000
25%          0.000000            31.000000
50%          0.000000           205.000000
75%          0.000000           724.000000
max        553.000000         75179.000000
```

df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3351 entries, 0 to 3350
Data columns (total 21 columns):
 #   Column             Non-Null Count  Dtype
---  ------             --------------  -----
 0   year               3351 non-null   int64
 1   month              3351 non-null   int64
 2   carrier            3351 non-null   object
 3   carrier_name       3351 non-null   object
 4   airport            3351 non-null   object
 5   airport_name       3351 non-null   object
 6   arr_flights        3343 non-null   float64
 7   arr_del15          3343 non-null   float64
 8   carrier_ct         3343 non-null   float64
 9   weather_ct         3343 non-null   float64
 10  nas_ct             3343 non-null   float64
 11  security_ct        3343 non-null   float64
 12  late_aircraft_ct   3343 non-null   float64
 13  arr_cancelled      3343 non-null   float64
 14  arr_diverted       3343 non-null   float64
 15  arr_delay          3343 non-null   float64
 16  carrier_delay      3343 non-null   float64
 17  weather_delay      3343 non-null   float64
 18  nas_delay          3343 non-null   float64
 19  security_delay     3343 non-null   float64
```

```
 20  late_aircraft_delay  3343 non-null    float64
dtypes: float64(15), int64(2), object(4)
memory usage: 549.9+ KB

df["carrier"].unique()

array(['9E', 'AA', 'AS', 'B6', 'DL', 'F9', 'G4', 'HA', 'MQ', 'NK', 'OH',
        'OO', 'YX', 'UA', 'WN', 'YV', 'EV'], dtype=object)

df["airport"].nunique()

360
```

Carrier and number of delay
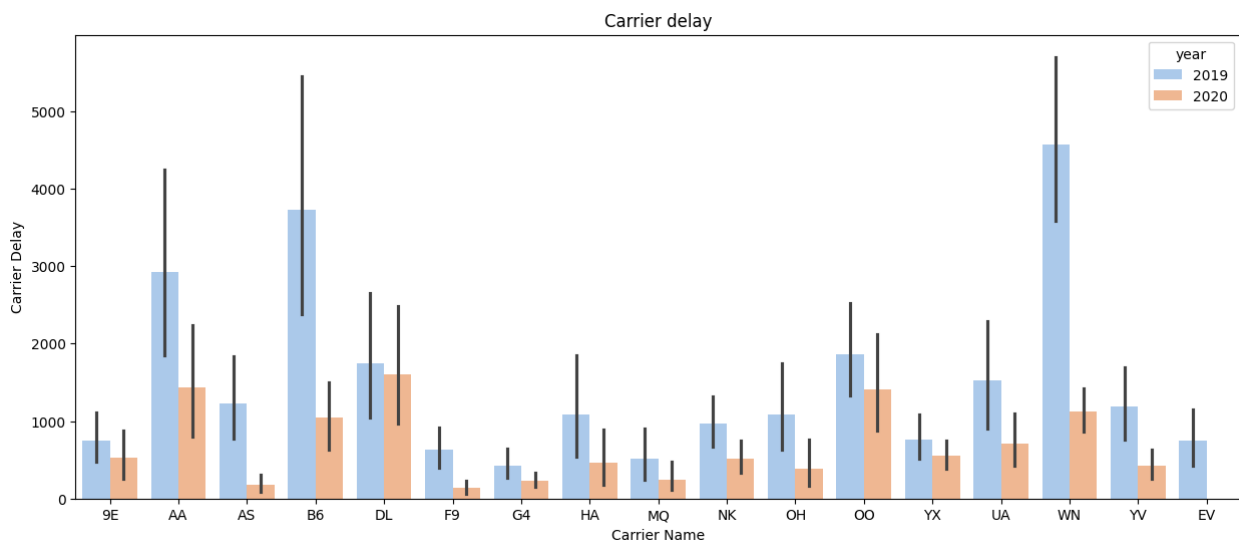
```python
x = "carrier"
y = "carrier_delay"

plt.figure(figsize=(15, 6))

sns.barplot(x =x, y =y, data = df, palette = "pastel", hue ="year")

plt.title("Carrier delay")
plt.xlabel("Carrier Name")
plt.ylabel("Carrier Delay")

plt.show()
```



```python
sns.scatterplot(x = 'carrier', data = df,y = 'arr_delay', palette = "pastel",hue = 'year')
```

```
plt.xlabel("carrier")
plt.ylabel("Delay Time (minutes)")
```

```
Text(0, 0.5, 'Delay Time (minutes)')
```
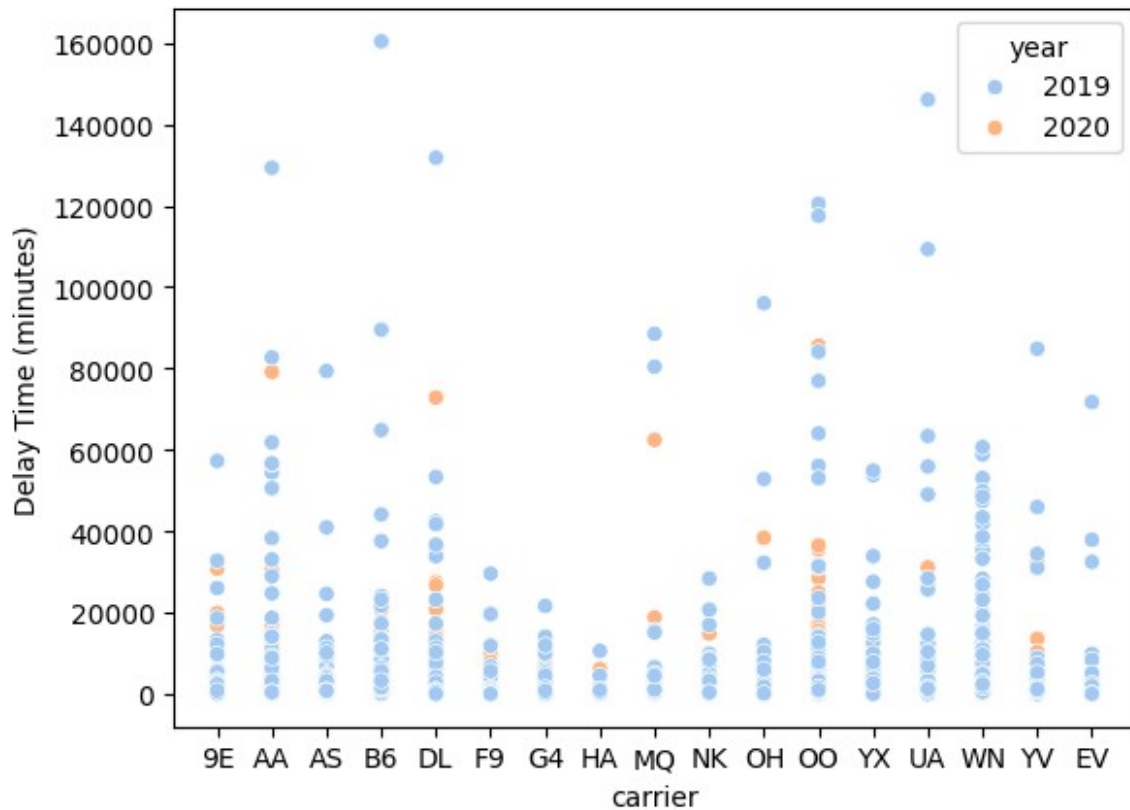


```
#delays =
df[["carrier_delay","weather_delay","security_delay","nas_delay","late
_aircraft_delay"]]

#delays.plot(kind = "bar", stacked = True, colormap ="viridis",
figsize = (20,6))

sns.scatterplot(data = df, x = 'carrier' y = )

#plt.title("Distribution of Delays by Category")
#plt.xlabel("Index")
#plt.ylabel("Delay Time (minutes)")

#plt.show()
```

```
  Cell In[9], line 5
    sns.scatterplot(data = df, x = 'carrier' y = )
                                              ^
```

```
SyntaxError: invalid syntax. Perhaps you forgot a comma?

df

"""i =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]
j = ['9E', 'AA', 'AS', 'B6', 'DL', 'F9', 'G4', 'HA', 'MQ', 'NK',
'OH','OO', 'YX', 'UA', 'WN', 'YV', 'EV']

for flight in j:
    for max in (df.loc[j]):
        x = 0
        if max>x:
            x = x+max
            continue"""

"""i =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]
j = ['9E', 'AA', 'AS', 'B6', 'DL', 'F9', 'G4', 'HA', 'MQ', 'NK',
'OH','OO', 'YX', 'UA', 'WN', 'YV', 'EV']

for flight in j:
    for max in j(axis = 0):
        x = 0
        if max>x:
            x = x+max
            continue"""

"""i =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]

x = "carrier"
y = i.avg()

plt.figure(figsize=(15, 6))

sns.scatterplot(x=x,y=y, data = df, palette = "pastel")
plt.show()"""

"""b =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]
df['max_delay'] = df[b].idxmax(axis=1)
print(df[['carrier', 'max_delay']])"""

b =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]
df['max_delay'] = df[b].idxmax(axis=1)
#print(df[['carrier', 'max_delay']])
ss =
df[["weather_delay","nas_delay","security_delay","late_aircraft_delay"
```

```
]]

plt.figure(figsize=(15, 2))

sns.barplot(x = 'carrier', y ='max_delay', data = df, hue
='max_delay')
plt.title("Carrier with Maximum Delay Type")
plt.xlabel("Carrier")
plt.ylabel("Delay Type")
plt.show()
```

```
C:\Users\Anushka\AppData\Local\Temp\ipykernel_13968\980311725.py:2:
FutureWarning: The behavior of DataFrame.idxmax with all-NA values, or
any-NA and skipna=False, is deprecated. In a future version this will
raise ValueError
  df['max_delay'] = df[b].idxmax(axis=1)
```
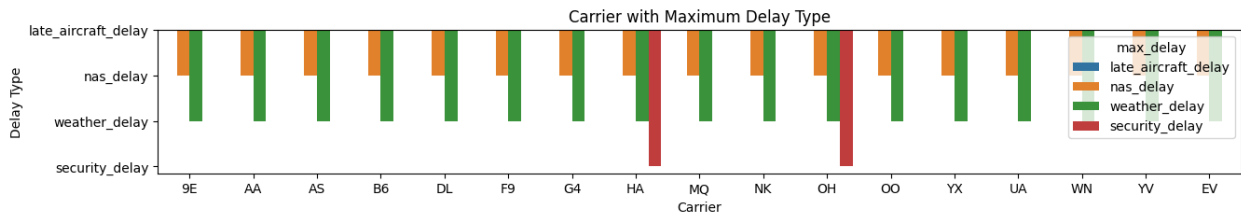

Carrier with Maximum Delay Type

```
"""b =
["weather_delay","nas_delay","security_delay","late_aircraft_delay"]
df['max_delay'] = df[b].idxmax(axis=1)
#print(df[['carrier', 'max_delay']])

j = ['9E', 'AA', 'AS', 'B6', 'DL', 'F9', 'G4', 'HA', 'MQ', 'NK',
'OH','OO', 'YX', 'UA', 'WN', 'YV', 'EV']


for fl in j:
    for carr in j:
        print(j(df['max_delay']))"""
```