Timeseries

November 17, 2023

[1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import gc
import os
import seaborn as sns
```

[2]:
```
airports = ['Atlanta', 'Chicago', 'Dallas', 'Denver', 'LA', 'JFK']
basepath = '/Users/ketansand/Downloads/Flight_Data'
years = np.arange(2012,2023).astype(str)
```

[59]:
```
file_name = '/Users/ketansand/Downloads/Flight_Data/2019/Atlanta.csv'
data_air = pd.read_csv(file_name)
```

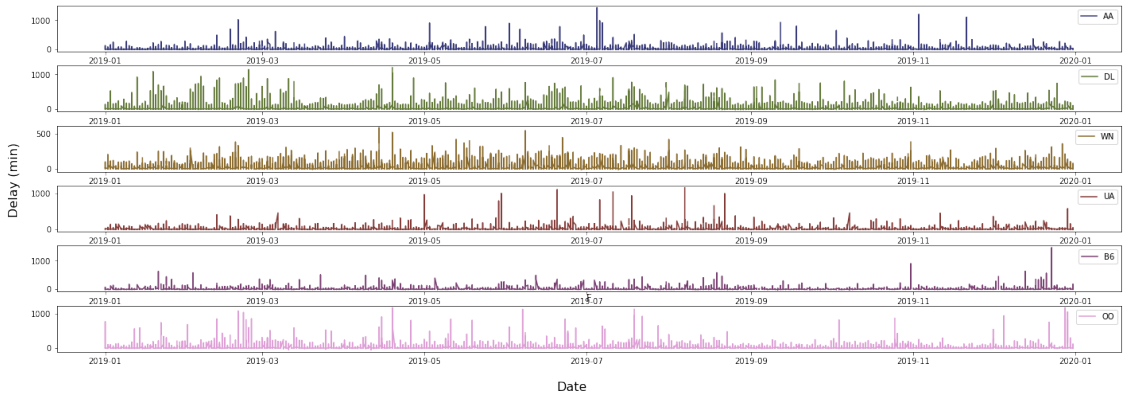
[60]:
```
data_air['Date'] = pd.to_datetime(data_air['FL_DATE'])
```

[61]:
```
print(data_air.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395009 entries, 0 to 395008
Data columns (total 17 columns):
 #   Column               Non-Null Count   Dtype
---  ------               --------------   -----
 0   Unnamed: 0           395009 non-null  int64
 1   FL_DATE              395009 non-null  object
 2   OP_UNIQUE_CARRIER    395009 non-null  object
 3   ORIGIN_AIRPORT_ID    395009 non-null  int64
 4   DEST_AIRPORT_ID      395009 non-null  int64
 5   CRS_DEP_TIME         395009 non-null  int64
 6   DEP_DELAY            392654 non-null  float64
 7   CANCELLED            395009 non-null  float64
 8   CANCELLATION_CODE    2406 non-null    object
 9   AIR_TIME             391701 non-null  float64
 10  CARRIER_DELAY        61519 non-null   float64
 11  WEATHER_DELAY        61519 non-null   float64
 12  NAS_DELAY            61519 non-null   float64
 13  SECURITY_DELAY       61519 non-null   float64
 14  LATE_AIRCRAFT_DELAY  61519 non-null   float64
```

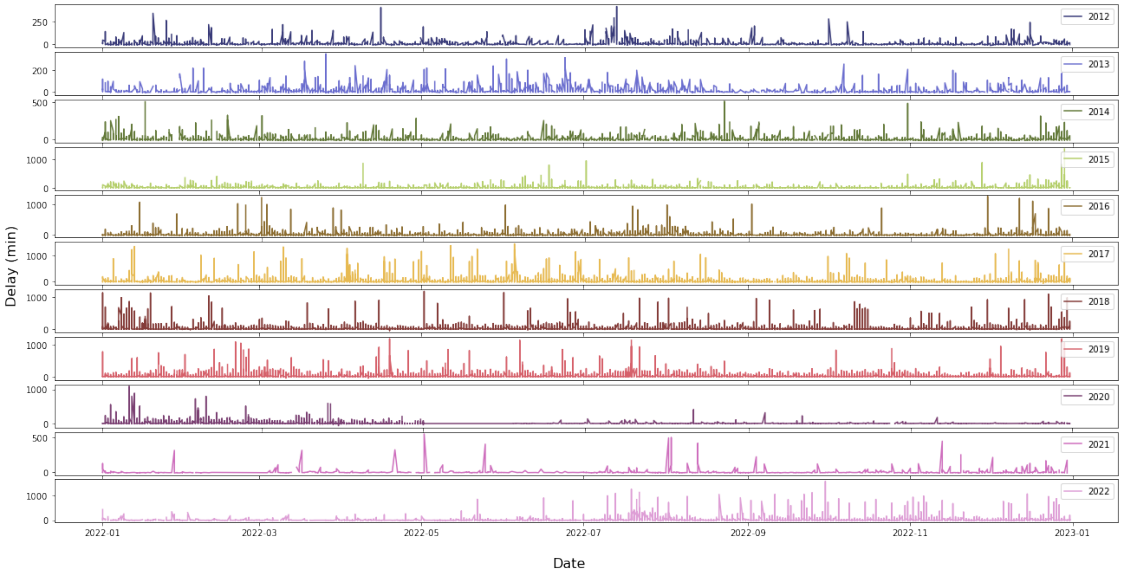
```
 15  Month                 395009 non-null  int64
 16  Date                  395009 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(8), int64(5), object(3)
memory usage: 51.2+ MB
None
```

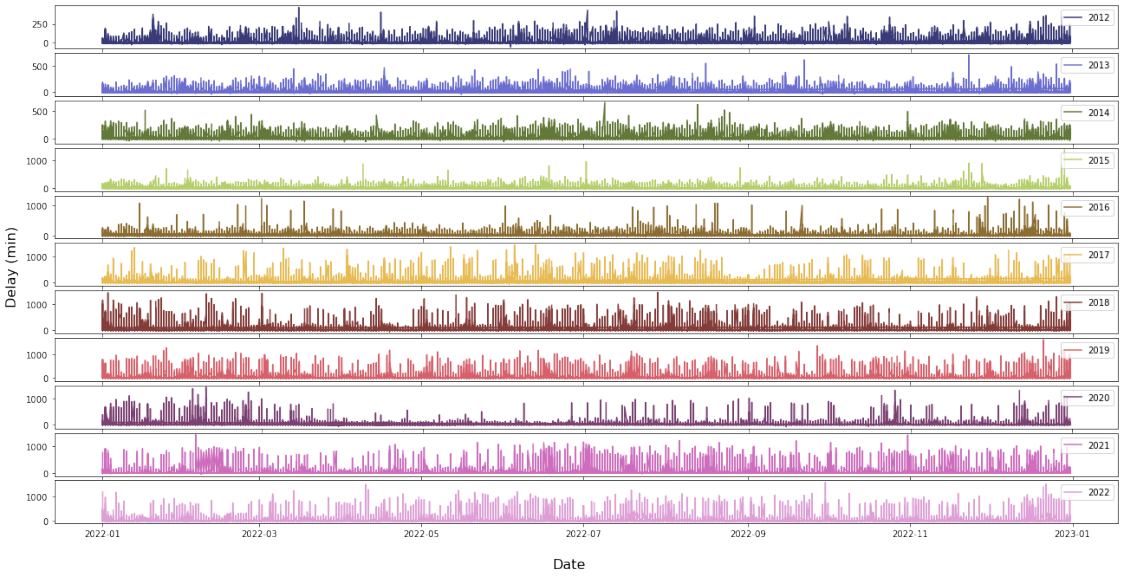
[86]:
```
gc.collect()



#carriers = data_air['OP_UNIQUE_CARRIER'].unique()
carriers=['AA', 'DL','WN', 'UA', 'B6', 'OO']

fig, axs = plt.subplots(nrows = len(carriers), ncols=1, figsize=(24, 8))
plt.subplots_adjust(hspace=0.3)
colors = plt.cm.get_cmap('tab20b', len(carriers))


for i in range(len(carriers)):
    subset = data_air[data_air['OP_UNIQUE_CARRIER'] == carriers[i]]
    #plt.fill_between(subset['Date'], subset['DEP_DELAY'], label=carrier)
    axs[i].plot(subset['Date'], subset['DEP_DELAY'], color = colors(i),␣
 ↪label=carriers[i])
    axs[i].legend()

fig.text(0.5, 0.04, 'Date', ha='center', fontsize=16)
fig.text(0.09, 0.5, 'Delay (min)', va='center', rotation='vertical',␣
 ↪fontsize=16)
plt.show()
```

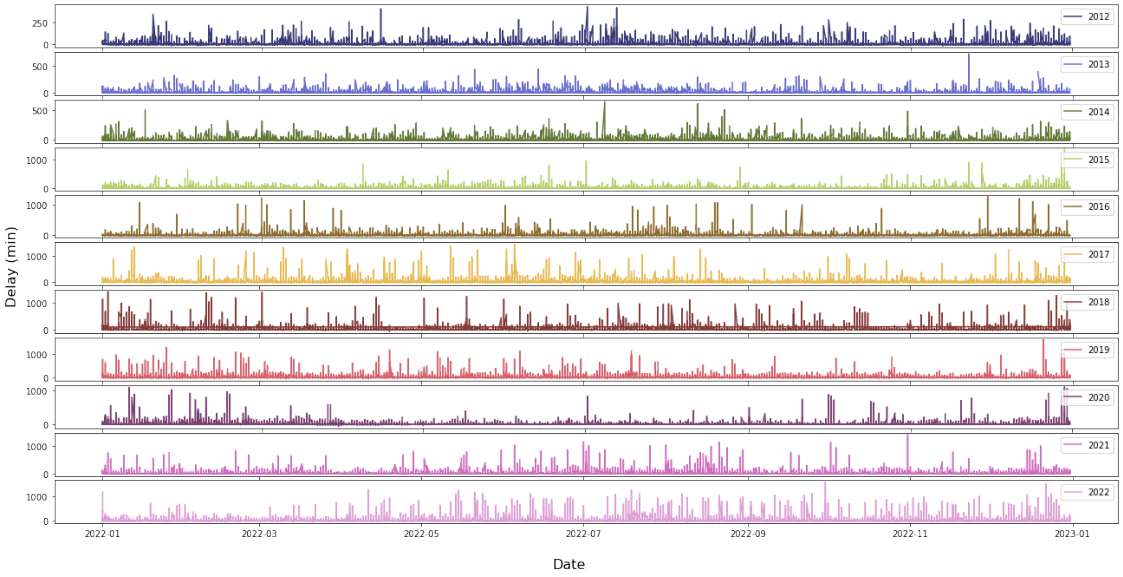


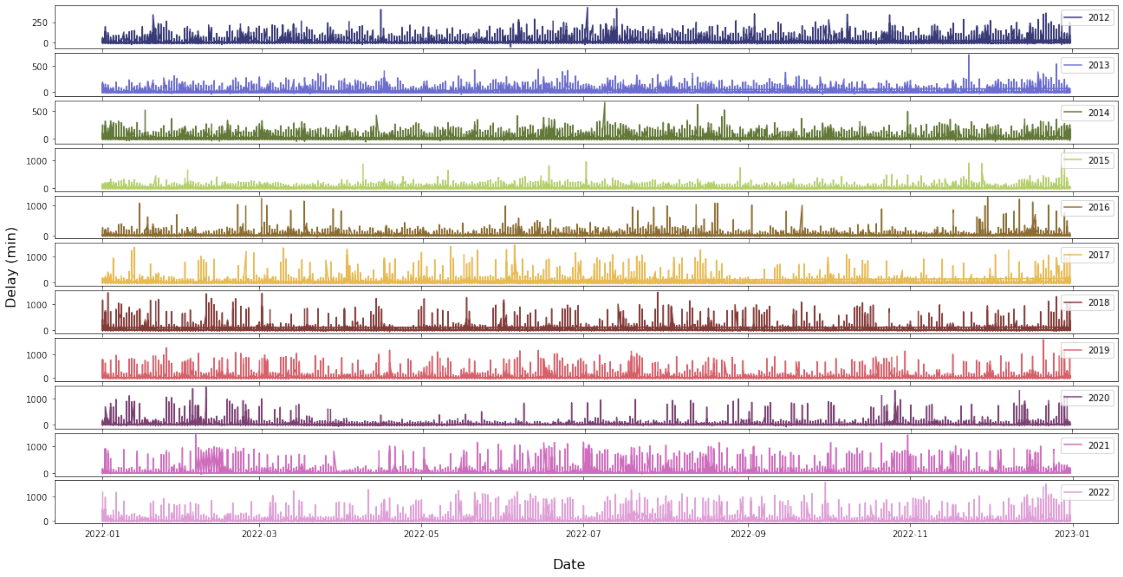
[107]:
```
#### For American Airlines #####
carrier='OO'
```

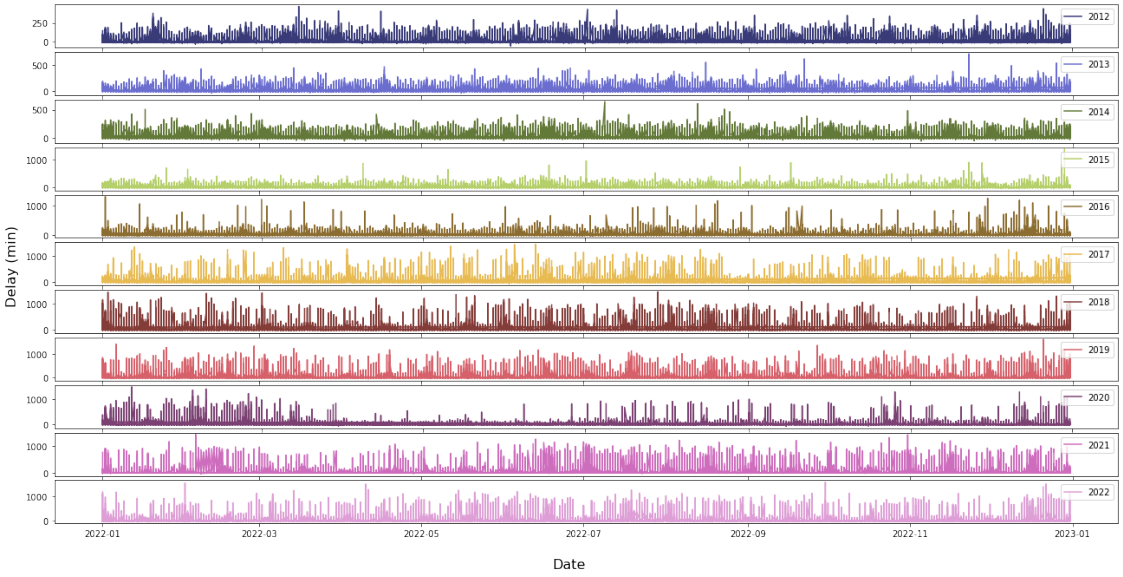
```
for j in range(len(airports)):
    fig, axs = plt.subplots(nrows = len(years), ncols=1, figsize=(22, 11))
    plt.subplots_adjust(hspace=0.1)
    colors = plt.cm.get_cmap('tab20b', len(years))
    for i in range(len(years)):
        file_name = '/Users/ketansand/Downloads/Flight_Data/{}/{}.csv'.
 ↪format(years[i],airports[j])
        data_air = pd.read_csv(file_name)
        data_air['Date'] = pd.to_datetime(data_air['FL_DATE'])
        subset = data_air[data_air['OP_UNIQUE_CARRIER'] == carrier]
#         subset = data_air

        #plt.fill_between(subset['Date'], subset['DEP_DELAY'], label=carrier)
        axs[i].plot(subset['Date'], subset['DEP_DELAY'], color = colors(i),␣
 ↪label=years[i])
        axs[i].legend(loc='upper right')
        gc.collect()

    fig.text(0.5, 0.06, 'Date', ha='center', fontsize=16)
    fig.text(0.09, 0.5, 'Delay (min)', va='center', rotation='vertical',␣
 ↪fontsize=16)
    oname = '/Users/ketansand/Downloads/Flight_Data/Plots/Timeseries/Skywest/
 ↪{}_{}.jpg'.format(airports[j],carrier)
    plt.savefig(oname, dpi=400, bbox_inches='tight')
    plt.show()
```



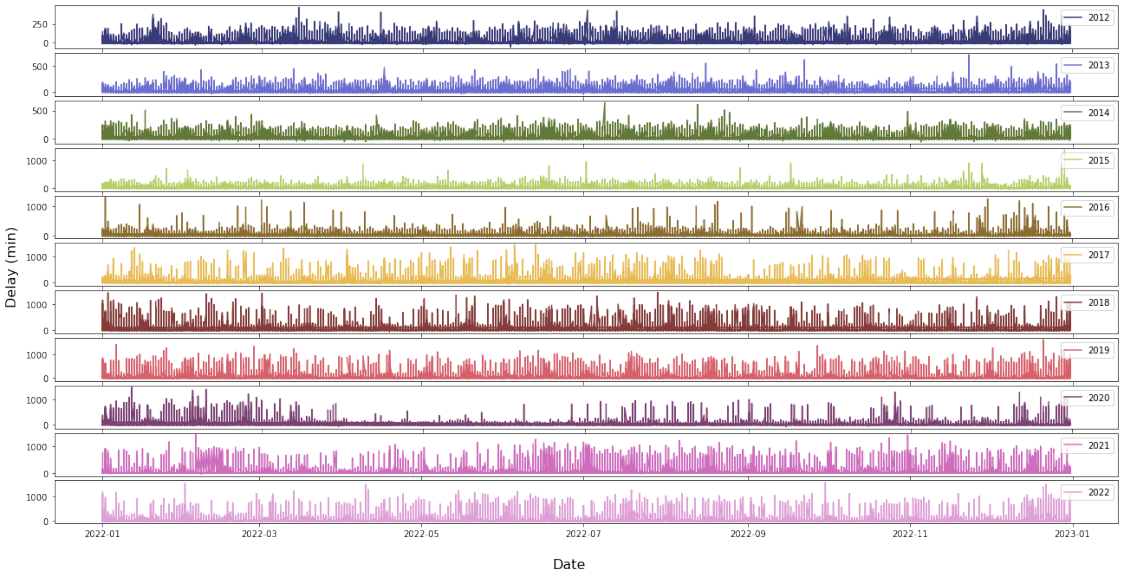

Delay (min)
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
Date


Delay (min)
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
Date






```
/Users/ketansand/opt/anaconda3/lib/python3.8/site-
packages/IPython/core/interactiveshell.py:3165: DtypeWarning: Columns (8) have
mixed types.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```


Delay (min)
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2022-01
2022-03
2022-05
2022-07
2022-09
2022-11
2023-01
Date

[ ]: