In [1]:

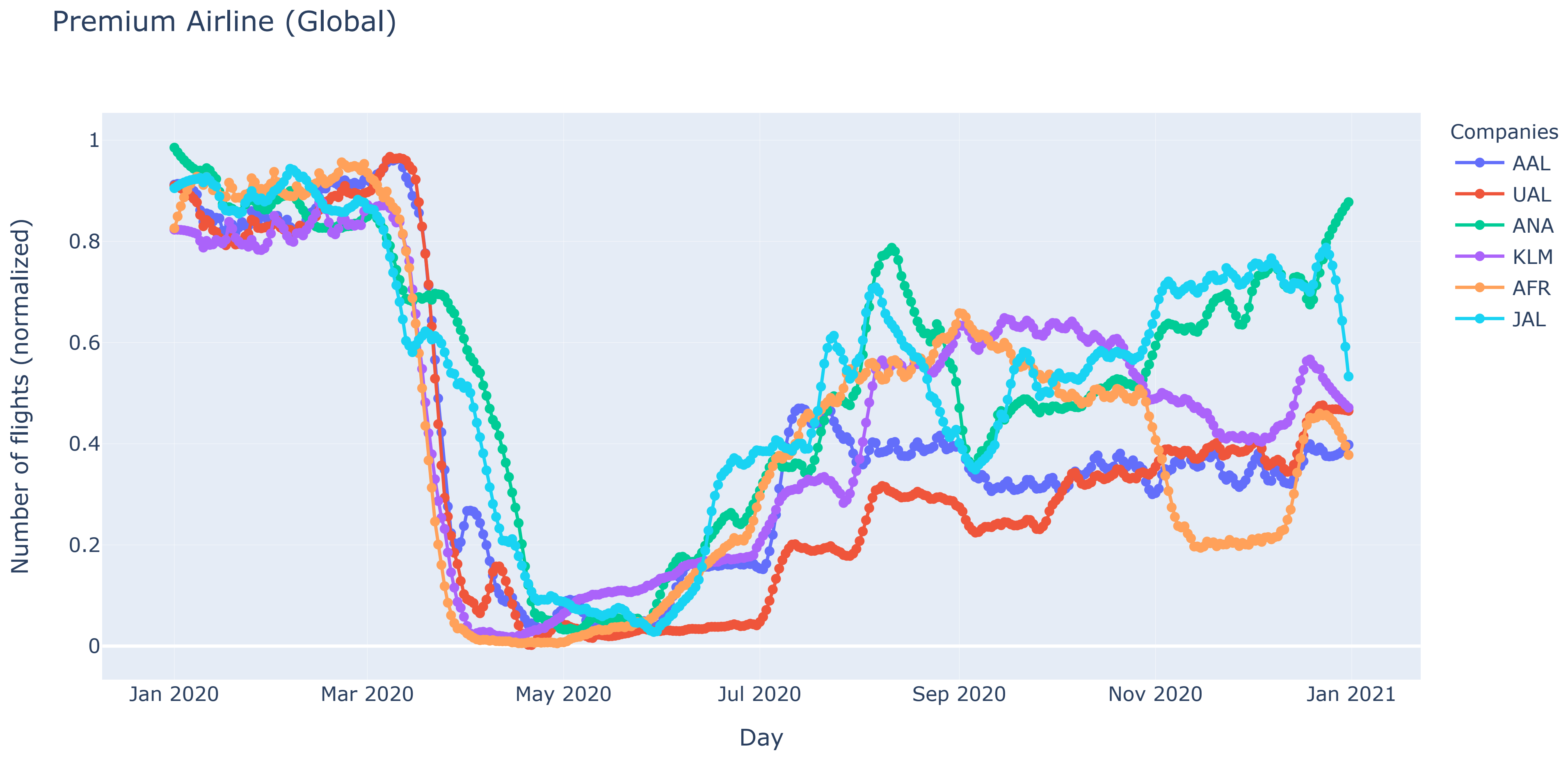
```
import plotly.graph_objects as go
import plotly.express as px
import pandas as pd
import datetime
import numpy as np
from scipy import signal

def normalize(df):
    '''

    Takes the dataframe as an input and normalizes the value in all the columns based on
    min-max normalization

    paramater : df : dataframe

    '''

    assert isinstance(df,pd.DataFrame)

    result = df.copy()
    column_list = list(df.columns)
    comp_col    = column_list[2:len(df.columns)]
    print(comp_col)
    for feature_name in comp_col:
        max_value = df[feature_name].max()
        min_value = df[feature_name].min()
        result[feature_name] = (df[feature_name] - min_value) / (max_value - min_value)
    return result

def company_wise_line_plot(input_file, airline_list, airline_type):

    '''
    This function takes the path to cleaned dataset having company wise data set for 2020
    and the list of airlines premium/small/cargo as an input and makes the line chart for
    the entire year

    input_file   : string : file path to the cleaned data set
    airline_list : list   : list of airlines
    airline_type : string : premium/small/cargo

    '''

    assert isinstance(input_file,str)
    assert isinstance(airline_list,list)
    assert isinstance(airline_type,str)

    airlines_data = pd.read_csv(input_file)
    airlines_data.fillna(0, inplace=True)
    airlines_data['day'] = airlines_data['day'].apply(lambda x: datetime.datetime.strptime(x.split()[0], '%Y-%m-%d'))
    airlines_data['day'] = airlines_data['day'].apply(lambda x: x.date()).apply(str)

    data = normalize(airlines_data)

    fig = go.Figure()

    for airline in airline_list:
        fig.add_trace(go.Scatter(x=data['day'], y=signal.savgol_filter(data[airline],15,2),mode='lines+markers',name=airline))


    fig.update_xaxes(title_text="Day")
    fig.update_yaxes(title_text="Number of flights (normalized)")
    title_string = airline_type + " (Global)"
    fig.update_layout(title_text = title_string)
    fig.update_layout(legend_title = "Companies")

    fig.show()


def bar_based_comparision(input_file_2020, input_file_2019):

    '''
    This function takes the path to cleaned dataset having company wise data set for 2020
    2019 and depending on the type of airlines (premium/small/cargo), it plots a bar graph comparision
    of percentage reduction in number of flights from 2019 to 2020 for each of the airline categories

    input_file_2020   : string : file path to the cleaned data set from 2019
    input_file_2019   : string : file path to the cleaned data set form 2020

    '''

    assert isinstance(input_file_2020,str)
    assert isinstance(input_file_2019,str)

    airlines_data_2020 = pd.read_csv(input_file_2020)
    airlines_data_2020['day'] = airlines_data_2020['day'].apply(lambda x: datetime.datetime.strptime(x.split()[0], '%Y-%m-%d'))
    airlines_data_2020['month'] = airlines_data_2020['day'].dt.month
    by_month = airlines_data_2020.groupby('month').sum()
    by_month_2020 = by_month.drop('Unnamed: 0',axis=1)

    airlines_data_2019 = pd.read_csv(input_file_2019)
    airlines_data_2019['day'] = airlines_data_2019['day'].apply(lambda x: datetime.datetime.strptime(x.split()[0], '%Y-%m-%d'))
    airlines_data_2019['month'] = airlines_data_2019['day'].dt.month
    by_month = airlines_data_2019.groupby('month').sum()
    by_month_2019 = by_month.drop('Unnamed: 0',axis=1)

    by_month_2019.fillna(0, inplace=True)
    by_month_2020.fillna(0, inplace=True)


    airlines_2020 = pd.DataFrame(columns=['Cargo', 'Premium', 'Small'])
    airlines_2020['Cargo'] = by_month_2020[cargo_airlines].T.sum()
    airlines_2020['Premium'] = by_month_2020[premium_airlines].T.sum()
    airlines_2020['Small'] = by_month_2020[small_airlines].T.sum()

    airlines_2019 = pd.DataFrame(columns=['Cargo', 'Premium', 'Small'])
    airlines_2019['Cargo'] = by_month_2019[cargo_airlines].T.sum()
    airlines_2019['Premium'] = by_month_2019[premium_airlines].T.sum()
    airlines_2019['Small'] = by_month_2019[small_airlines].T.sum()

    diff = ((airlines_2020 - airlines_2019)/airlines_2019)*100
    diff['Month'] = ['Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec']


    fig = px.bar(diff, x=diff.index, y=['Cargo','Premium','Small'], range_y = (-100,30), barmode='group')
    fig.add_trace(px.line(diff, x=np.arange(0.7,12), y='Cargo').data[0])
    line2 = px.line(diff, x=np.arange(1,13), y='Premium').data[0]
    line2['line']['color'] = '#EF553B'
    fig.add_trace(line2)
    line3 = px.line(diff, x=np.arange(1.3,13.3), y='Small').data[0]
    line3['line']['color'] = '#00cc96'
    fig.add_trace(line3)
    fig.update_yaxes(title_text="Percentage Change")
    fig.update_xaxes(title_text="Month")
    fig.update_layout(title_text="Flight Comparision with 2019")
    fig.update_layout(
        xaxis = dict(
        tickmode = 'array',
        tickvals = [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12],
        ticktext = ['Jan','Feb','Mar','Apr','May','Jun','Jul','Aug','Sep','Oct','Nov','Dec']
        )
    )
    fig.show()
```

```
C:\ProgramData\Anaconda3\lib\site-packages\scipy\__init__.py:137: UserWarning: NumPy 1.16.5 or above is required for this version of SciPy (detected version 1.16.4)
  UserWarning)
```

In [2]:

```
## Premium Airline Plots ##
path_to_data = 'Data/company_wise_data_2020.csv'
premium_airlines = ["AAL", "UAL", "ANA", "KLM", "AFR", "JAL"]
airline_type = "Premium Airline"
company_wise_line_plot(path_to_data, premium_airlines,airline_type)
```

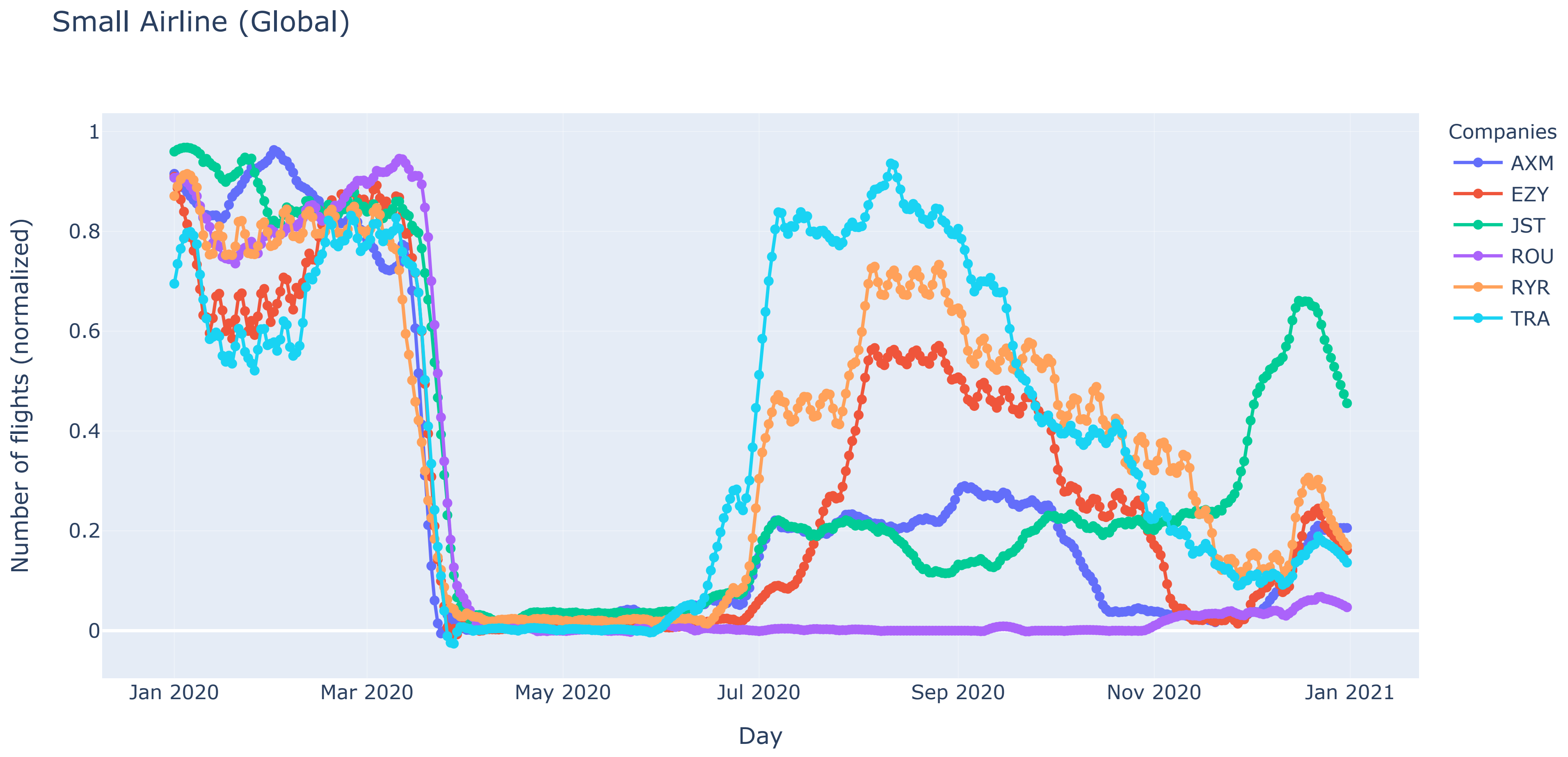
```
['AAL', 'UAL', 'ANA', 'KLM', 'AFR', 'JAL', 'AXM', 'EZY', 'JST', 'ROU', 'RYR', 'TRA', 'CLX', 'FDX', 'GEC', 'GTI', 'UPS']
```



In [3]:

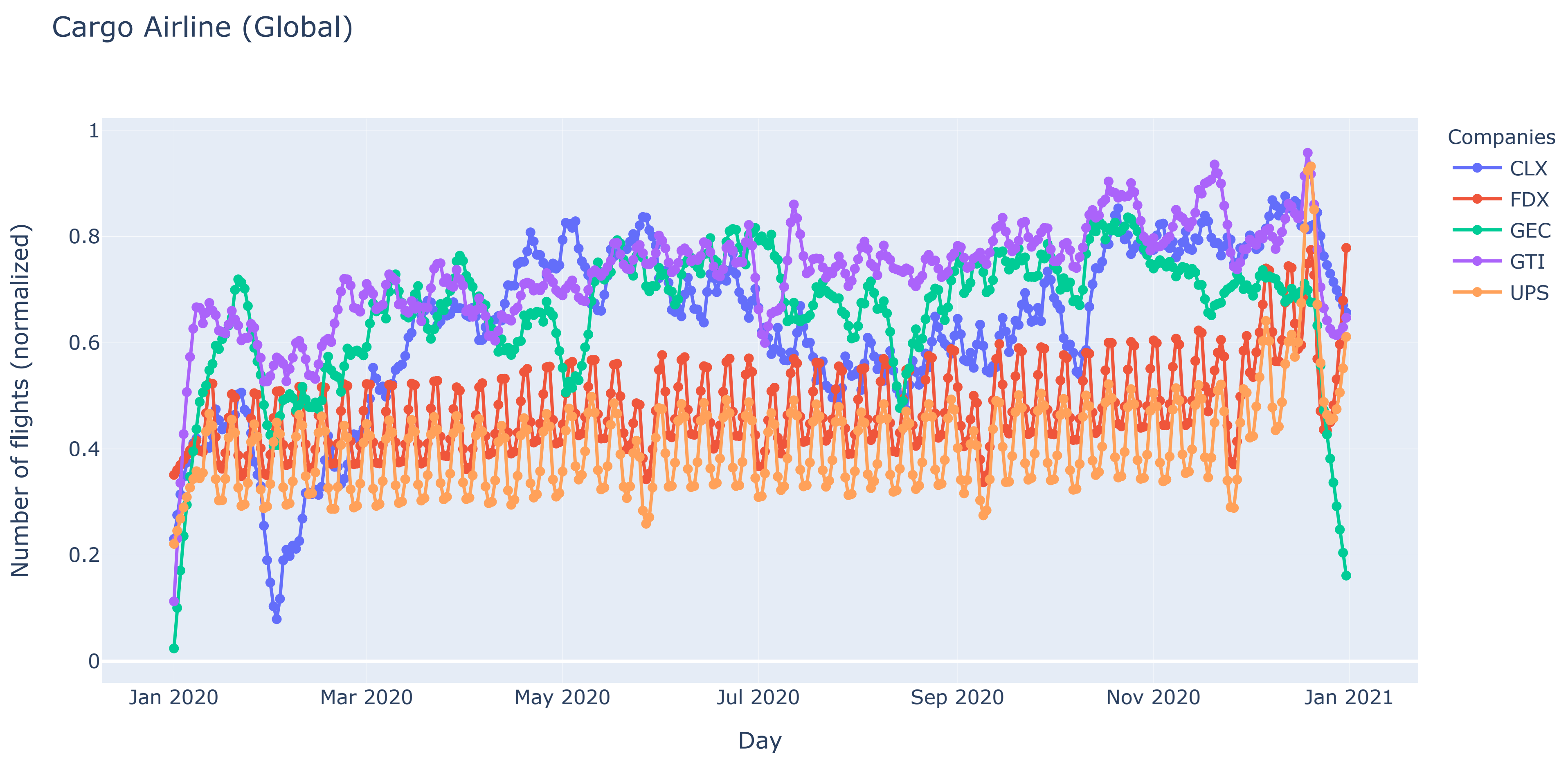
```
## Small Airline Plots ##
path_to_data = 'Data/company_wise_data_2020.csv'
small_airlines = ["AXM", "EZY", "JST", "ROU", "RYR", "TRA"]
airline_type = "Small Airline"
company_wise_line_plot(path_to_data, small_airlines,airline_type)
```

```
['AAL', 'UAL', 'ANA', 'KLM', 'AFR', 'JAL', 'AXM', 'EZY', 'JST', 'ROU', 'RYR', 'TRA', 'CLX', 'FDX', 'GEC', 'GTI', 'UPS']
```

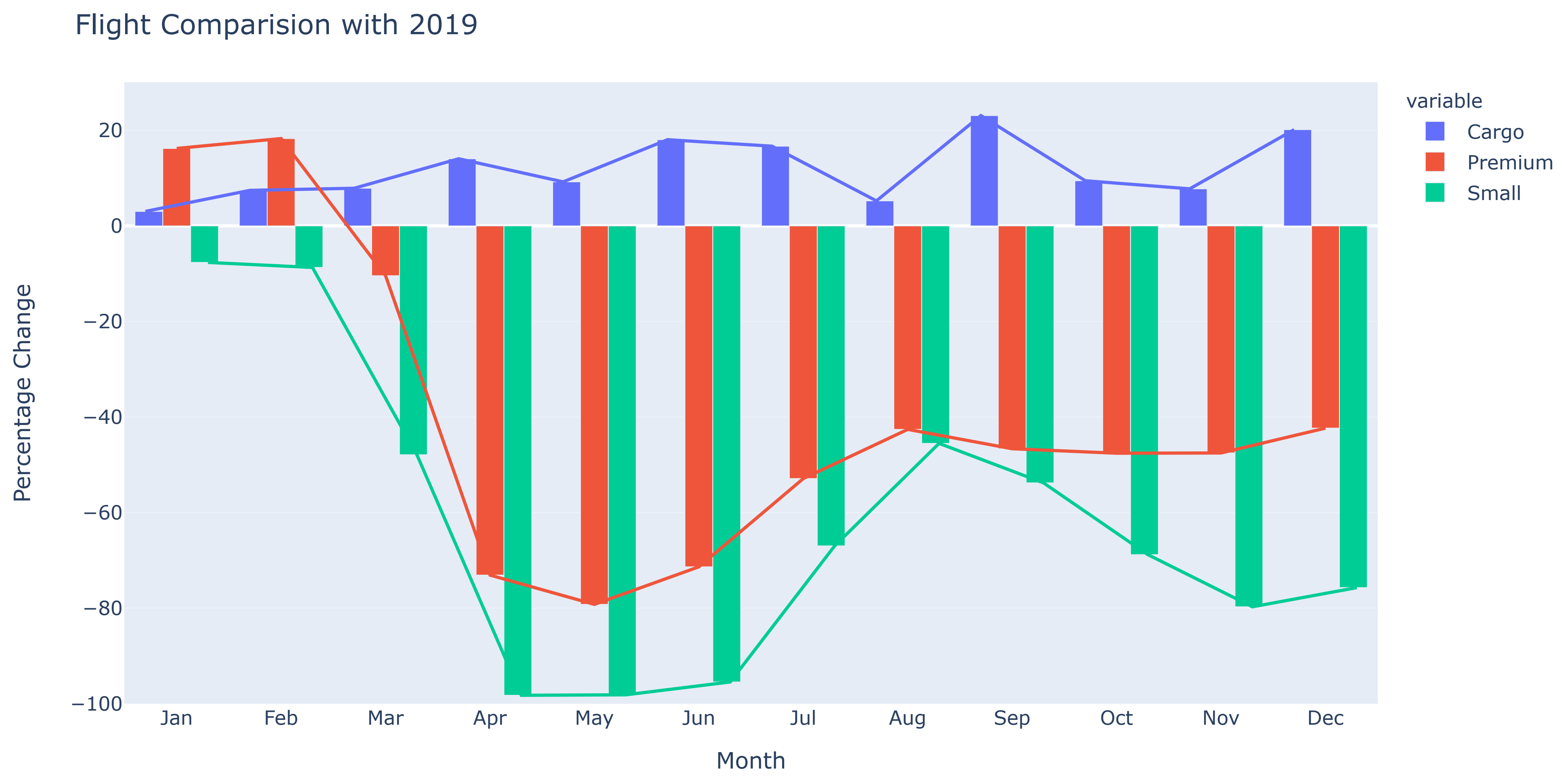


In [4]:
```
## Cargo Airline Plots ##
path_to_data = 'Data/company_wise_data_2020.csv'
cargo_airlines = ["CLX", "FDX", "GEC", "GTI", "UPS"]
airline_type = "Cargo Airline"
company_wise_line_plot(path_to_data, cargo_airlines,airline_type)
```

```
['AAL', 'UAL', 'ANA', 'KLM', 'AFR', 'JAL', 'AXM', 'EZY', 'JST', 'ROU', 'RYR', 'TRA', 'CLX', 'FDX', 'GEC', 'GTI', 'UPS']
```



In [5]:
```
#Bar Chart Comparision from 2019 to 2020
data_2019 = 'Data/company_wise_data_2019.csv'
data_2020 = 'Data/company_wise_data_2020.csv'
bar_based_comparision(data_2020,data_2019)
```


Flight Comparision with 2019
Percentage Change
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
variable
Cargo
Premium
Small