We want to first visually explore the data to see if we can confirm some of our initial hypotheses as well as make new hypothesis about the problem we are trying to solve.

For this we will start by loading the data and understanding the data structure of the dataframe we have.

1 Importing essential libraries and reading the Dataframe

In [1]:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

In [2]:

```
# Set parameters for better visualization
plt.style.use('ggplot')
plt.rcParams['figure.figsize'] = (15, 10)
```

In [3]:

```
df = pd.read_excel('MonthWiseMarketArrivals_ChennaiClean.xlsx')
df.head()
```

Out[3]:

	market	month	year	quantity	priceMin	priceMax	priceMod	date
0	CHENNAI	January	2004	103400	798	1019	910	2004-01-01
1	CHENNAI	February	2004	87800	776	969	873	2004-02-01
2	CHENNAI	March	2004	102180	506	656	580	2004-03-01
3	CHENNAI	April	2004	83300	448	599	527	2004-04-01
4	CHENNAI	May	2004	84850	462	596	529	2004-05-01

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 204 entries, 0 to 203
Data columns (total 8 columns):
 #   Column    Non-Null Count  Dtype         
---  ------    --------------  -----         
 0   market    204 non-null    object        
 1   month     204 non-null    object        
 2   year      204 non-null    int64         
 3   quantity  204 non-null    int64         
 4   priceMin  204 non-null    int64         
 5   priceMax  204 non-null    int64         
 6   priceMod  204 non-null    int64         
 7   date      204 non-null    datetime64[ns]
dtypes: datetime64[ns](1), int64(5), object(2)
memory usage: 12.9+ KB
```

2 Changing the data type for date column

In [5]:

```
df.date = pd.DatetimeIndex(df.date)
df.dtypes
```

Out[5]:

```
market              object
month               object
year                 int64
quantity             int64
priceMin             int64
priceMax             int64
priceMod             int64
date        datetime64[ns]
dtype: object
```

In [6]:

```
pd.set_option("precision", 0)
df.describe()
```

Out[6]:

	year	quantity	priceMin	priceMax	priceMod
count	204	204	204	204	204
mean	2012	110667	1443	1791	1622
std	5	15078	1155	1324	1237
min	2004	63900	304	456	384
25%	2008	101550	753	1000	879
50%	2012	110450	1089	1446	1263
75%	2016	120275	1784	2138	1954
max	2020	150400	8696	11130	9876

In [7]:

```
df.index
```

Out[7]:

```
RangeIndex(start=0, stop=204, step=1)
```

3 Sorting the values by date

In [8]:

```
df = df.sort_values(by="date")
df.head()
```

Out[8]:

	market	month	year	quantity	priceMin	priceMax	priceMod	date
0	CHENNAI	January	2004	103400	798	1019	910	2004-01-01
1	CHENNAI	February	2004	87800	776	969	873	2004-02-01
2	CHENNAI	March	2004	102180	506	656	580	2004-03-01
3	CHENNAI	April	2004	83300	448	599	527	2004-04-01
4	CHENNAI	May	2004	84850	462	596	529	2004-05-01

In [9]:

```
df.index = pd.PeriodIndex(df.date, freq='M')
df.head()
```

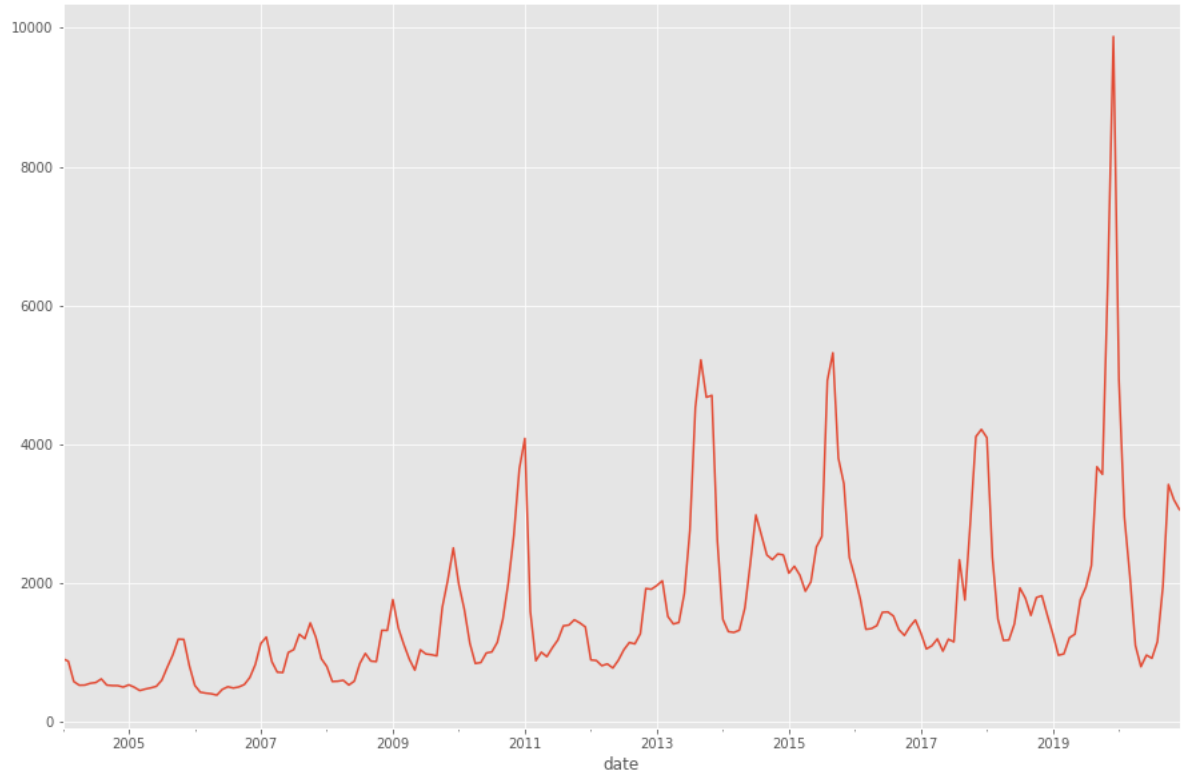
Out[9]:

date	market	month	year	quantity	priceMin	priceMax	priceMod	date
2004-01	CHENNAI	January	2004	103400	798	1019	910	2004-01-01
2004-02	CHENNAI	February	2004	87800	776	969	873	2004-02-01
2004-03	CHENNAI	March	2004	102180	506	656	580	2004-03-01
2004-04	CHENNAI	April	2004	83300	448	599	527	2004-04-01
2004-05	CHENNAI	May	2004	84850	462	596	529	2004-05-01

4 Date vs Modal Price

In [10]:

```
df.priceMod.plot()
plt.show()
```

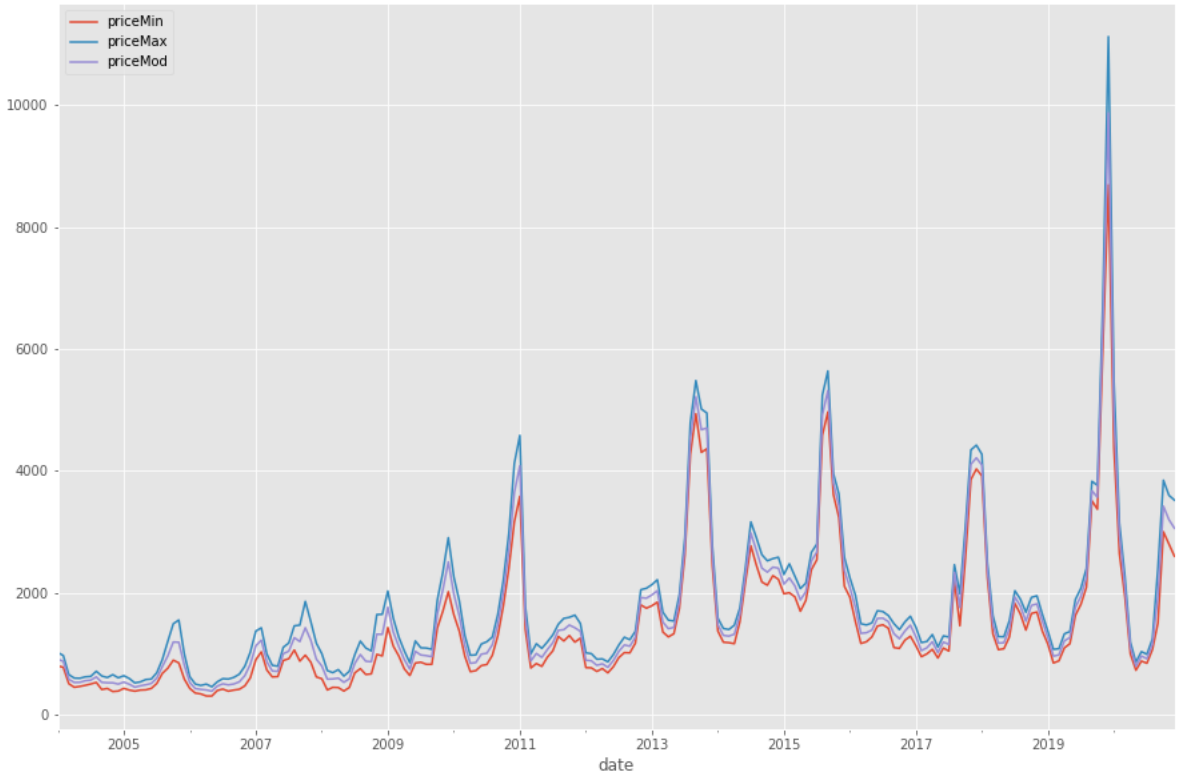


5 Date vs Min, Max, Mod price

In [11]:

```
plt.figure(figsize=(20, 10))
df.plot(kind="line", y=["priceMin" ,"priceMax" ,"priceMod"])
plt.show()
```

```
<Figure size 1440x720 with 0 Axes>
```



6 Price Difference vs Date

In [12]:

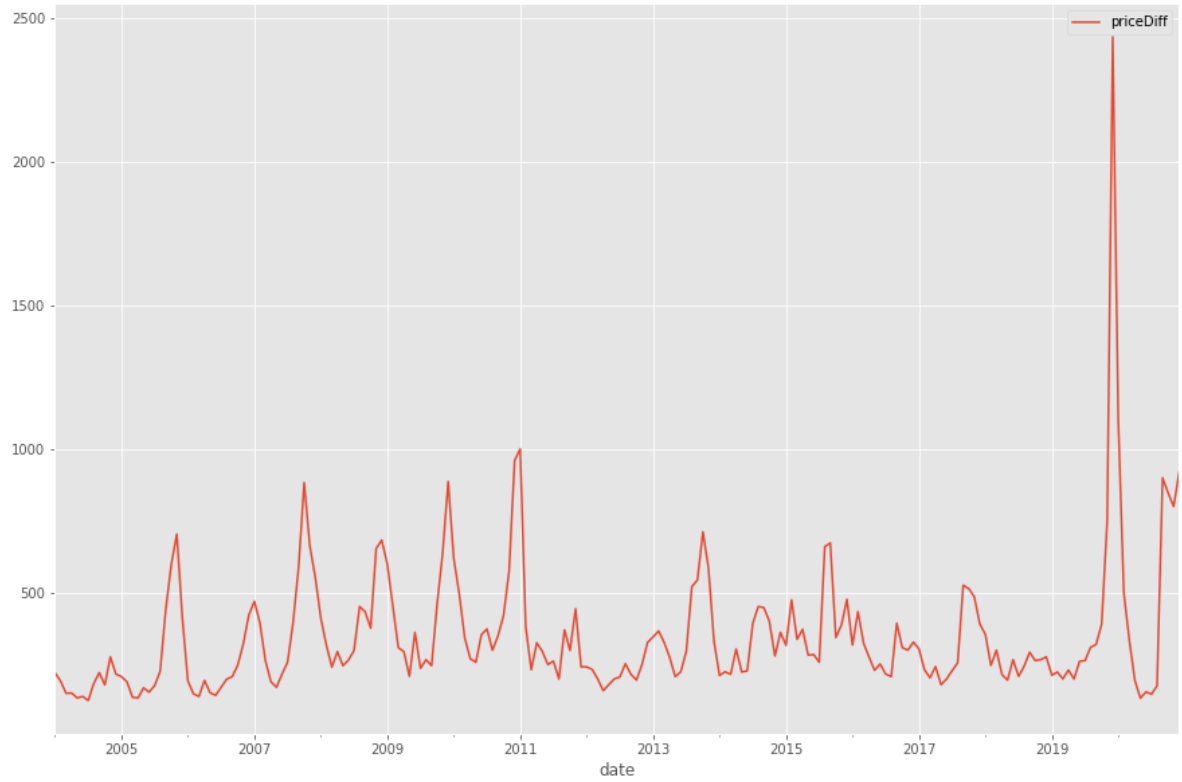
```
df["priceDiff"] = df["priceMax"] - df["priceMin"]
df.head()
```

Out[12]:

date	market	month	year	quantity	priceMin	priceMax	priceMod	date	priceDiff
2004-01	CHENNAI	January	2004	103400	798	1019	910	2004-01-01	221
2004-02	CHENNAI	February	2004	87800	776	969	873	2004-02-01	193
2004-03	CHENNAI	March	2004	102180	506	656	580	2004-03-01	150
2004-04	CHENNAI	April	2004	83300	448	599	527	2004-04-01	151
2004-05	CHENNAI	May	2004	84850	462	596	529	2004-05-01	134

In [13]:

```
df.plot(y="priceDiff")
plt.show()
```



7 Pivot table

In [14]:

```
df["monthVal"] = pd.DatetimeIndex(df["date"]).month
df.head()
```

Out[14]:

date	market	month	year	quantity	priceMin	priceMax	priceMod	date	priceDiff	mon
2004-01	CHENNAI	January	2004	103400	798	1019	910	2004-01-01	221	
2004-02	CHENNAI	February	2004	87800	776	969	873	2004-02-01	193	
2004-03	CHENNAI	March	2004	102180	506	656	580	2004-03-01	150	
2004-04	CHENNAI	April	2004	83300	448	599	527	2004-04-01	151	
2004-05	CHENNAI	May	2004	84850	462	596	529	2004-05-01	134	

In [15]:

```
df_Pivot = pd.pivot_table(df, values = "priceDiff",
                          columns = "year", index = "monthVal")
df_Pivot
```

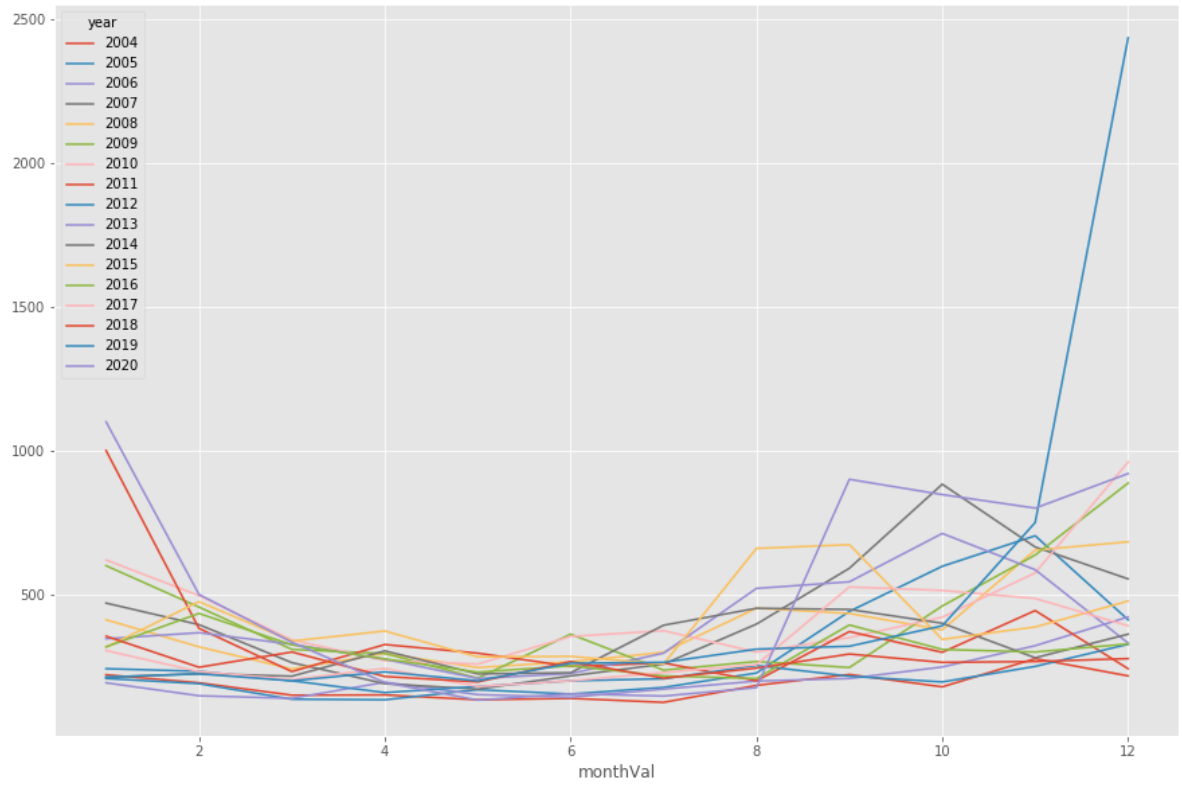
Out[15]:

monthVal	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	201
1	221	208	193	470	412	600	620	1000	242	346	212	317	318	30
2	193	190	148	395	318	456	496	382	233	367	225	475	434	23
3	150	136	139	263	241	309	342	232	200	327	216	338	325	20
4	151	134	195	190	295	295	270	326	159	275	304	373	276	24
5	134	169	152	171	246	209	258	296	180	208	224	283	230	18
6	139	154	143	217	266	362	354	250	200	225	228	285	252	20
7	125	177	171	258	299	237	374	262	207	296	393	258	217	22
8	184	227	199	397	452	267	300	200	253	521	452	660	208	25
9	222	441	208	591	434	246	350	371	216	544	448	673	394	52
10	179	598	248	883	377	460	421	299	196	712	400	344	309	51
11	277	704	322	665	654	637	575	444	250	586	280	387	300	48
12	217	412	422	554	683	887	960	242	327	331	362	477	328	39

8 Month vs Price Diff

In [16]:

```
df_Pivot.plot()
plt.show()
```



9 Price trend every year

In [17]:

```
df_Pivot.plot(subplots = True, figsize=(15, 15), layout=(5, 4), sharey=True)
plt.show()
```