In [77]:
```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
import plotly.express as px
import folium
import xgboost
import sklearn
from pylab import rcParams
from sorted_months_weekdays import Month_Sorted_Month, Weekday_Sorted_Week
import bokeh as bh

df = pd.read_csv('C:/Users/Bilal Aktas/Desktop/Git Lab/Fontys/Fontys Subjects/Semester 4/Applied Data Scienc
```

Data Preparation

In [78]:
```python
df.columns = ['Delivery date', 'Delivery time', 'Pharmacy number', 'Pharmacy Postcode (2)', 'Year of birth',
              'Gender', 'CNK', 'Product name', 'ATC code', 'Units', 'Price', 'Contribution']
```

In [79]:
```python
df.head()
```

Out[79]:

	Delivery date	Delivery time	Pharmacy number	Pharmacy Postcode (2)	Year of birth	Gender	CNK	Product name	ATC code	Units	Price	Contribution
0	01/01/2020	00:00	9105972	10	1925	2	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00
1	01/01/2020	00:00	9105972	10	1923	2	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00
2	01/01/2020	00:00	9105972	10	1924	2	736165	BURINEX COMP 1 X 5 MG	C03CA02	0	0.38	0.08
3	01/01/2020	00:00	9105972	10	1921	2	750695	XARELTO COMP 1 X 15 MG	B01AF01	0	2.38	0.10
4	01/01/2020	00:00	9105972	10	1924	2	750695	XARELTO COMP 1 X 15 MG	B01AF01	0	2.38	0.10

In [80]:
```python
TimeNull = df['Delivery time']
TimeNull.fillna('00:00',inplace=True)
```

In [81]:
```python
df = df[df['Gender'] != 0]
df = df[df['Gender'] != 3]
```

In [82]:
```python
df["Gender"].value_counts(sort=True)
```

Out[82]:
```
2    2856183
1    2203711
Name: Gender, dtype: int64
```

In [83]:
```python
df['Delivery date'] = pd.to_datetime(df['Delivery date'], infer_datetime_format=True)
```

In [84]:
```python
df.isnull().any()
```

Out[84]:
```
Delivery date            False
Delivery time            False
Pharmacy number          False
Pharmacy Postcode (2)    False
Year of birth            False
Gender                   False
CNK                      False
Product name             False
ATC code                 False
Units                    False
Price                    False
Contribution             False
dtype: bool
```

In [85]:
```python
df['ATCShort'] = df['ATC code'].str[:3]
df.head(50)
```

Out[85]:

	Delivery date	Delivery time	Pharmacy number	Pharmacy Postcode (2)	Year of birth	Gender	CNK	Product name	ATC code	Units	Price	Contribution	ATCSho
0	2020-01-01	00:00	9105972	10	1925	2	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
1	2020-01-01	00:00	9105972	10	1923	2	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
2	2020-01-01	00:00	9105972	10	1924	2	736165	BURINEX COMP 1 X 5 MG	C03CA02	0	0.38	0.08	C
3	2020-01-01	00:00	9105972	10	1921	2	750695	XARELTO COMP 1 X 15 MG	B01AF01	0	2.38	0.10	B
4	2020-01-01	00:00	9105972	10	1924	2	750695	XARELTO COMP 1 X 15 MG	B01AF01	0	2.38	0.10	B
5	2020-01-01	00:00	9105972	10	1921	2	7706310	ESCITALOPRAM TEVA COMP 1 X 10 MG	N06AB10	0	0.17	0.04	N
6	2020-01-01	00:00	9105972	10	1921	2	7706310	ESCITALOPRAM TEVA COMP 1 X 10 MG	N06AB10	0	0.17	0.04	N
7	2020-01-01	00:00	9105972	10	1924	2	743732	L THYROXINE CHRISTIAENS COMP 1 X 50 MCG	H03AA01	0	0.02	0.01	H
8	2020-01-01	00:00	9105972	10	1923	2	743732	L THYROXINE CHRISTIAENS COMP 1 X 50 MCG	H03AA01	0	0.02	0.01	H
9	2020-01-01	00:00	9105972	10	1923	2	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
10	2020-01-01	00:00	9105972	10	1925	1	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
11	2020-01-01	00:00	9105972	10	1924	1	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
12	2020-01-01	00:00	9105972	10	1926	1	702217	BURINEX COMP 1 X 1 MG	C03CA02	0	0.11	0.04	C
13	2020-01-01	00:00	9105972	10	1924	1	702217	BURINEX COMP 1 X 1 MG	C03CA02	0	0.11	0.04	C
14	2020-01-01	00:00	9105972	10	1928	2	708081	GLUCOPHAGE COMP 1 X 500 MG	A10BA02	0	0.03	0.00	A
15	2020-01-01	00:00	7983030	41	1927	2	2342269	ZUURSTOFCONC.KROBER HUUR + ONDERHOUD OXY...		0	92.80	0.00	
16	2020-01-01	00:00	7983030	41	1926	2	2342285	ZUURSTOFCONC.KROBER BEVOCHTIGER 1XGEBR OXY...		0	5.68	0.00	
17	2020-01-01	00:00	7983030	41	1927	2	4004941	HONORARIUM APOTHEKER OXYCONCENTRATOR COORDINATIE		0	12.98	0.00	
18	2020-01-01	00:00	9105972	10	1926	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
19	2020-01-01	00:00	9105972	10	1927	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
20	2020-01-01	00:00	9105972	10	1927	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.01	C
21	2020-01-01	00:00	9105972	10	1928	2	795187	OXYNORM INSTANT COMP 1 X 5 MG	N02AA05	0	0.14	0.05	N
22	2020-01-01	00:00	9105972	10	1929	2	795187	OXYNORM INSTANT COMP 1 X 5 MG	N02AA05	0	0.14	0.06	N
23	2020-01-01	00:00	9105972	10	1929	1	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
24	2020-01-01	00:00	9105972	10	1930	1	5520465	HONORARIUM PER WEEK PER RUSTOORDBEWONER ROB-RVT		0	3.20	0.00	
25	2020-01-01	00:00	9105972	10	1931	1	743294	L THYROXINE CHRISTIAENS COMP 1 X 100 MCG	H03AA01	0	0.04	0.01	H
26	2020-01-01	00:00	9105972	10	1933	1	743294	L THYROXINE CHRISTIAENS COMP 1 X 100 MCG	H03AA01	0	0.04	0.01	H
27	2020-01-01	00:00	9105972	10	1933	1	785105	OXYCONTIN COMP 1 X 5 MG	N02AA05	0	0.12	0.06	N
28	2020-01-01	00:00	9105972	10	1930	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
29	2020-01-01	00:00	9105972	10	1932	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
30	2020-01-01	00:00	9105972	10	1930	2	747626	TRITACE COMP 1 X 5 MG	C09AA05	0	0.22	0.09	C
31	2020-01-01	00:00	9105972	10	1930	2	747626	TRITACE COMP 1 X 5 MG	C09AA05	0	0.22	0.09	C
32	2020-01-01	00:00	9105972	10	1932	1	779975	SERTRALINE EG COMP 1 X 50 MG	N06AB06	0	0.22	0.05	N
33	2020-01-01	00:00	9105972	10	1930	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
34	2020-01-01	00:00	9105972	10	1930	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
35	2020-01-01	00:00	9105972	10	1930	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.00	C
36	2020-01-01	00:00	9105972	10	1934	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.01	C
37	2020-01-01	00:00	9105972	10	1934	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.01	C
38	2020-01-01	00:00	9105972	10	1933	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.01	C
39	2020-01-01	00:00	9105972	10	1934	2	784579	SIMVASTATINE EG COMP 1 X 20 MG	C10AA01	0	0.10	0.03	C
40	2020-01-01	00:00	9105972	10	1932	2	784579	SIMVASTATINE EG COMP 1 X 20 MG	C10AA01	0	0.10	0.04	C
41	2020-01-01	00:00	9105972	10	1933	2	791186	PANTOMED NYCOMED COMP 1 X 20MG	A02BC02	0	0.11	0.04	A
42	2020-01-01	00:00	9105972	10	1934	2	793372	LOSARTAN TEVA COMP 1 X 100 MG	C09CA01	0	0.16	0.05	C
43	2020-01-01	00:00	9105972	10	1935	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
44	2020-01-01	00:00	9105972	10	1931	1	789537	ASAFLOW COMP 1 X 80 MG	B01AC06	0	0.03	0.01	B
45	2020-01-01	00:00	9105972	10	1935	2	700351	ALDACTONE COMP 1 X 25 MG	C03DA01	0	0.06	0.01	C
46	2020-01-01	00:00	9105972	10	1934	1	7701188	DONEPEZIL TEVA COMP 1 X 10 MG	N06DA02	0	0.22	0.04	N
47	2020-01-01	00:00	9105972	10	1932	2	716282	REDOMEX DRAG 1 X 25 MG	N06AA09	0	0.03	0.00	N
48	2020-01-01	00:00	9105972	10	1932	2	716282	REDOMEX DRAG 1 X 25 MG	N06AA09	0	0.03	0.01	N
49	2020-01-01	00:00	9105972	10	1935	2	729301	FUROSEMIDE COMP 1 X 40 MG	C03CA01	0	0.04	0.01	C

In [86]:
```python
df['Delivery date'].min()
```

Out[86]:
```
Timestamp('2020-01-01 00:00:00')
```

In [87]:
```python
df['Delivery date'].max()
```

Out[87]:
```
Timestamp('2020-12-06 00:00:00')
```

Sales in 2020

In [52]:
```python
df = df.sort_values('Delivery date')
```

In [63]:
```python
df = df.groupby('Delivery date')['Price'].sum().reset_index()
```

In [64]:
```python
df
```

Out[64]:

	Delivery date	Price
0	2020-01-01	9940.83
1	2020-01-02	552784.38
2	2020-01-03	608643.46
3	2020-01-04	304993.12
4	2020-01-05	14375.33
...	...	...
237	2020-11-07	475.05
238	2020-12-02	3504.75
239	2020-12-03	3081.04
240	2020-12-05	4787.53
241	2020-12-06	908.60

242 rows × 2 columns

In [65]:
```python
df['month'] = df['Delivery date'].dt.month
```

In [66]:
```python
df['month']
```

Out[66]:
```
0       1
1       1
2       1
3       1
4       1
       ..
237    11
238    12
239    12
240    12
241    12
Name: month, Length: 242, dtype: int64
```

In [67]:
```python
import calendar
df['monthName'] = df['month'].apply(lambda x: calendar.month_name[x])
```

In [68]:
```python
df['monthName']
```

Out[68]:
```
0       January
1       January
2       January
3       January
4       January
         ...
237    November
238    December
239    December
240    December
241    December
Name: monthName, Length: 242, dtype: object
```

In [69]:
```python
ordered_months = ["January", "February", "March", "April", "May", "June",
        "July", "August", "September", "October", "November", "December"]
df['monthName'].apply(lambda x:ordered_months.index(x))
```

Out[69]:
```
0       0
1       0
2       0
3       0
4       0
       ..
237    10
238    11
239    11
240    11
241    11
Name: monthName, Length: 242, dtype: int64
```

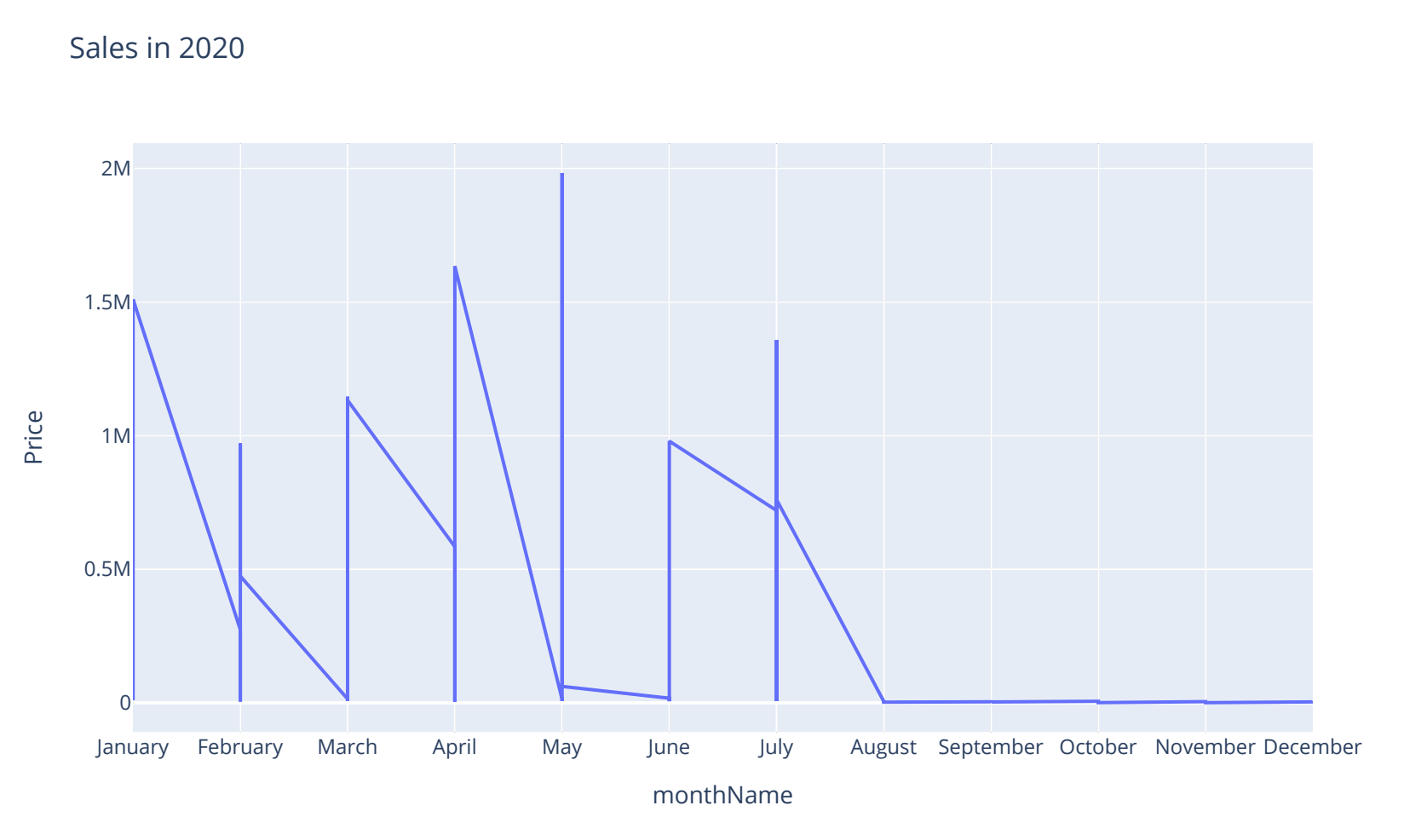
In [70]:
```python
df
```

Out[70]:

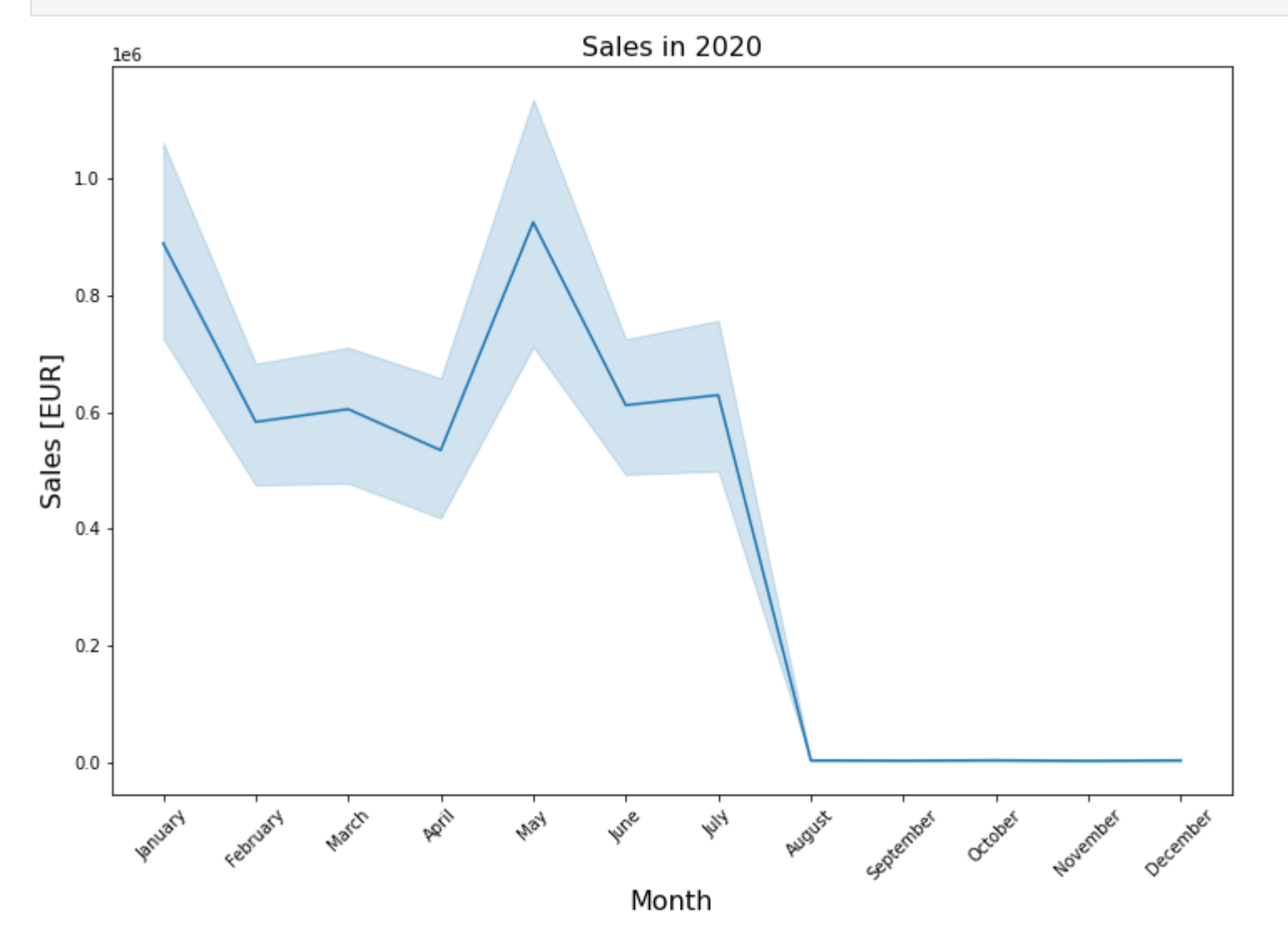
	Delivery date	Price	month	monthName
0	2020-01-01	9940.83	1	January
1	2020-01-02	552784.38	1	January
2	2020-01-03	608643.46	1	January
3	2020-01-04	304993.12	1	January
4	2020-01-05	14375.33	1	January
...	...	...	...	...
237	2020-11-07	475.05	11	November
238	2020-12-02	3504.75	12	December
239	2020-12-03	3081.04	12	December
240	2020-12-05	4787.53	12	December
241	2020-12-06	908.60	12	December

242 rows × 4 columns

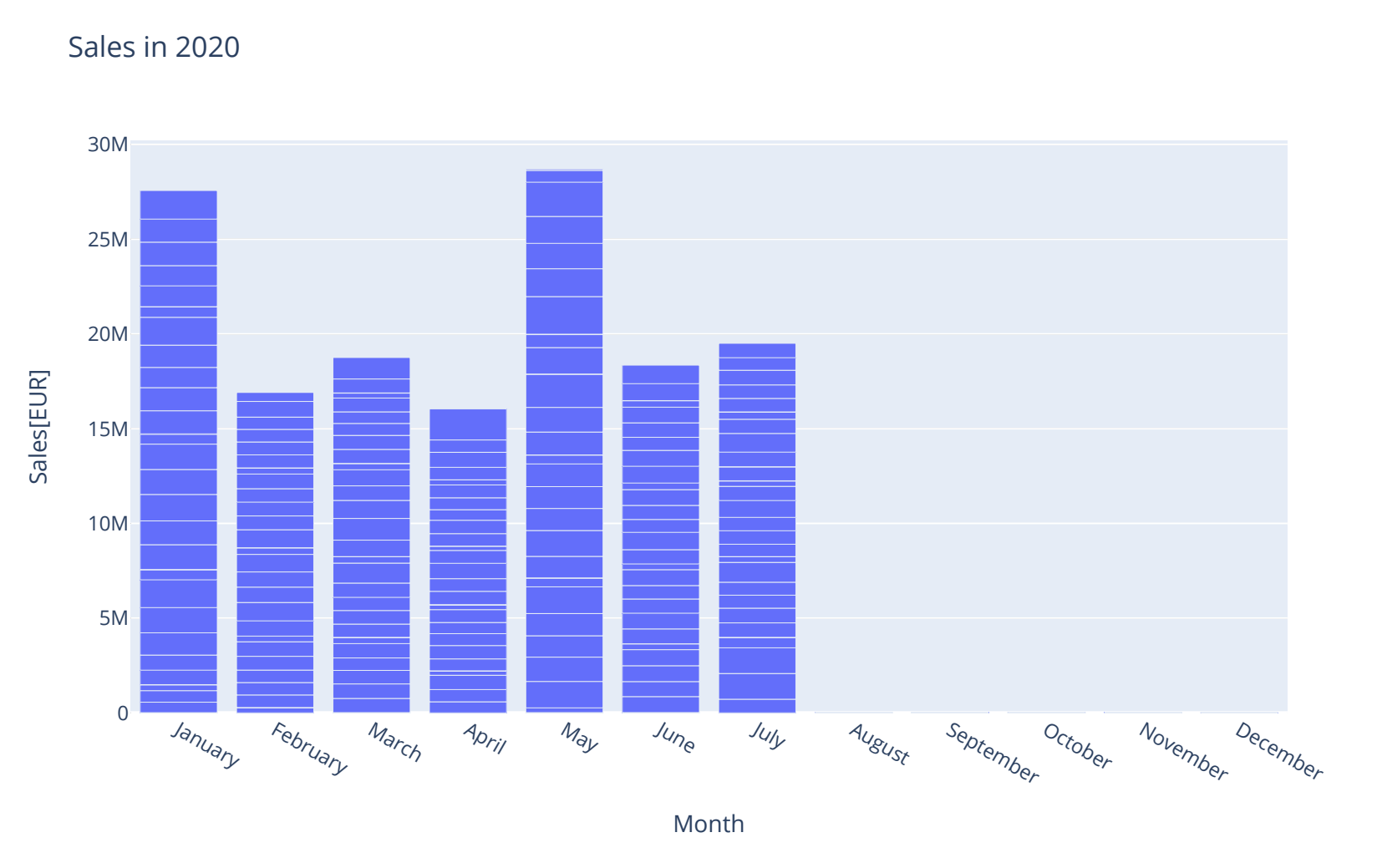
In [71]:
```python
import plotly.express as px
fig = px.line(df, x="monthName" , y="Price",title="Sales in 2020")
fig.show()
```



In [72]:
```python
plt.figure(figsize=(12, 8))
sns.lineplot(data=df,y="Price",x="monthName",markers=True,dashes=False)
plt.title("Sales in 2020", fontsize=16)
plt.xlabel("Month", fontsize=16)
plt.xticks(rotation=45)
plt.ylabel("Sales [EUR]", fontsize=16)
plt.show()
```



In [76]:
```python
fig = px.bar(df, x="monthName" , y="Price",
             title="Sales in 2020",labels={
                 "monthName": "Month",
                 "Price": "Sales[EUR]"})
fig.show()
```



Number of Deliveries in 2020

In [88]:
```python
df = df.sort_values('Delivery date')
```

In [89]:
```python
df = df.groupby('Delivery date')['ATC code'].count().reset_index()
```

In [90]:
```python
df['month'] = df['Delivery date'].dt.month
```

In [91]:
```python
import calendar
df['monthName'] = df['month'].apply(lambda x: calendar.month_name[x])
```

In [92]:
```python
ordered_months = ["January", "February", "March", "April", "May", "June",
        "July", "August", "September", "October", "November", "December"]
df['monthName'].apply(lambda x:ordered_months.index(x))
```

Out[92]:
```
0       0
1       0
2       0
3       0
4       0
       ..
237    10
238    11
239    11
240    11
241    11
Name: monthName, Length: 242, dtype: int64
```

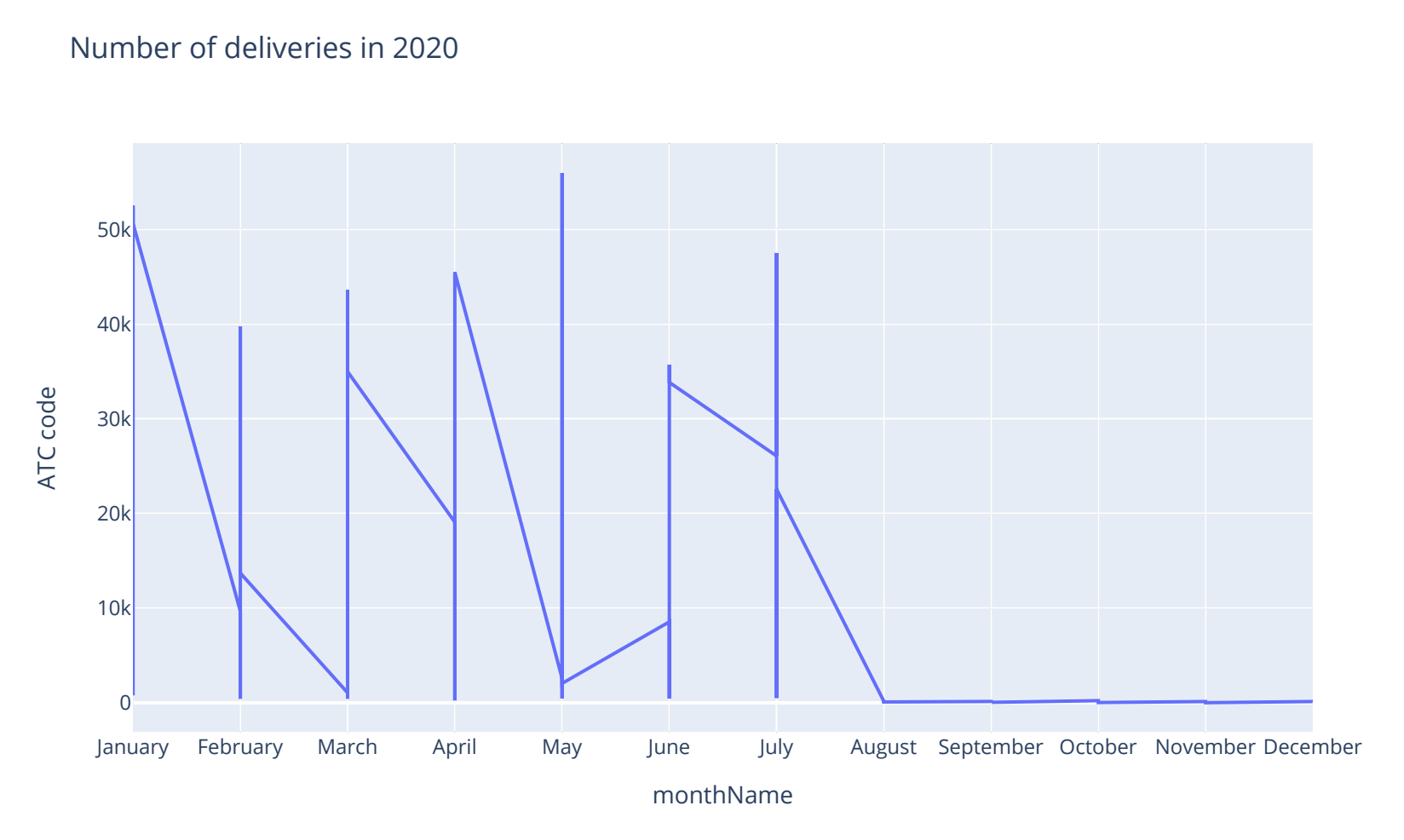
In [93]:
```python
df
```

Out[93]:

	Delivery date	ATC code	month	monthName
0	2020-01-01	788	1	January
1	2020-01-02	19607	1	January
2	2020-01-03	24267	1	January
3	2020-01-04	9404	1	January
4	2020-01-05	901	1	January
...	...	...	...	...
237	2020-11-07	13	11	November
238	2020-12-02	137	12	December
239	2020-12-03	98	12	December
240	2020-12-05	125	12	December
241	2020-12-06	47	12	December

242 rows × 4 columns

In [94]:
```python
fig = px.line(df, x="monthName" , y="ATC code",title="Number of deliveries in 2020")
fig.show()
```



In [95]:
```python
plt.figure(figsize=(12, 8))
sns.lineplot(data=df,y="ATC code",x="monthName",markers=True,dashes=False)
plt.title("Number of deliveries in 2020", fontsize=16)
plt.xlabel("Month", fontsize=16)
plt.xticks(rotation=45)
plt.ylabel("Number of deliveries", fontsize=16)
plt.show()
```