Diminos Case Study

In [1]:

```
import pandas as pd
import numpy as np
```

In [2]:

```
data = pd.read_csv('E:/file2/Downloads/diminos_data.csv')
data.head()
```

Out[2]:

	order_id	order_placed_at	order_delivered_at
0	1523111	2023-03-01 00:00:59	2023-03-01 00:18:07.443132
1	1523112	2023-03-01 00:03:59	2023-03-01 00:19:34.925241
2	1523113	2023-03-01 00:07:22	2023-03-01 00:22:28.291385
3	1523114	2023-03-01 00:07:47	2023-03-01 00:46:19.019399
4	1523115	2023-03-01 00:09:03	2023-03-01 00:25:13.619056

In [3]:

```
data.isnull().sum()
```

Out[3]:

```
order_id              0
order_placed_at       0
order_delivered_at    0
dtype: int64
```

In [4]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype 
---  ------              --------------  ----- 
 0   order_id            15000 non-null  int64 
 1   order_placed_at     15000 non-null  object
 2   order_delivered_at  15000 non-null  object
dtypes: int64(1), object(2)
memory usage: 351.7+ KB
```

In [5]:

```
data['order_placed_at']= pd.to_datetime(data['order_placed_at'])
data['order_delivered_at']=pd.to_datetime(data['order_delivered_at'])
```

In [6]:

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 15000 entries, 0 to 14999
Data columns (total 3 columns):
 #   Column              Non-Null Count  Dtype         
---  ------              --------------  -----         
 0   order_id            15000 non-null  int64         
 1   order_placed_at     15000 non-null  datetime64[ns]
 2   order_delivered_at  15000 non-null  datetime64[ns]
dtypes: datetime64[ns](2), int64(1)
memory usage: 351.7 KB
```

In [8]:

```
data['order_placed_at'].dt.dayofweek
```

Out[8]:

```
0        2
1        2
2        2
3        2
4        2
       ..
14995    0
14996    0
14997    0
14998    0
14999    0
Name: order_placed_at, Length: 15000, dtype: int64
```

In [11]:

```
data['dat_of_month'] = data['order_placed_at'].dt.day
```

In [13]:

```
data.tail()
```

Out[13]:

	order_id	order_placed_at	order_delivered_at	dat_of_month
14995	1538106	2023-03-27 23:37:05	2023-03-27 23:52:37.409378	27
14996	1538107	2023-03-27 23:47:38	2023-03-28 00:04:22.672912	27
14997	1538108	2023-03-27 23:50:16	2023-03-28 00:05:40.676238	27
14998	1538109	2023-03-27 23:52:44	2023-03-28 00:08:41.810358	27
14999	1538110	2023-03-27 23:58:20	2023-03-28 00:13:42.499311	27

In [15]:

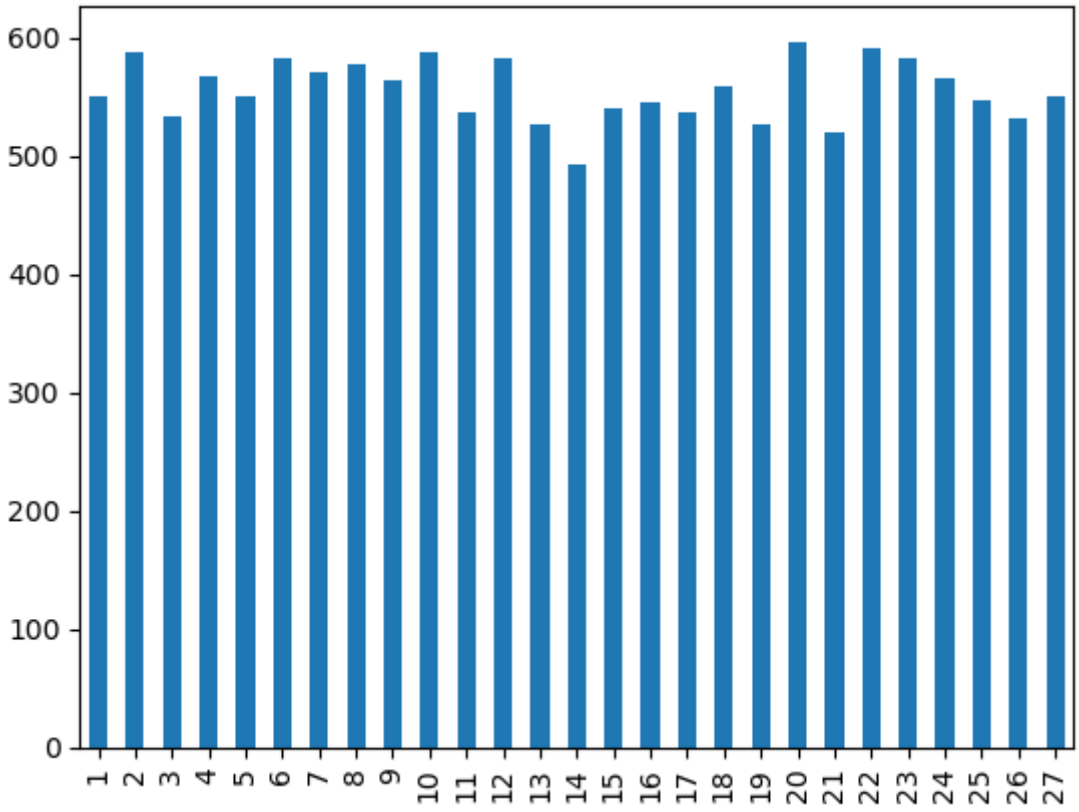
```
data['day_of_week']= data['order_placed_at'].dt.day_name()
```

In [18]:

```
data['dat_of_month'].value_counts().sort_index().plot(kind='bar')
```

Out[18]:

<AxesSubplot:>

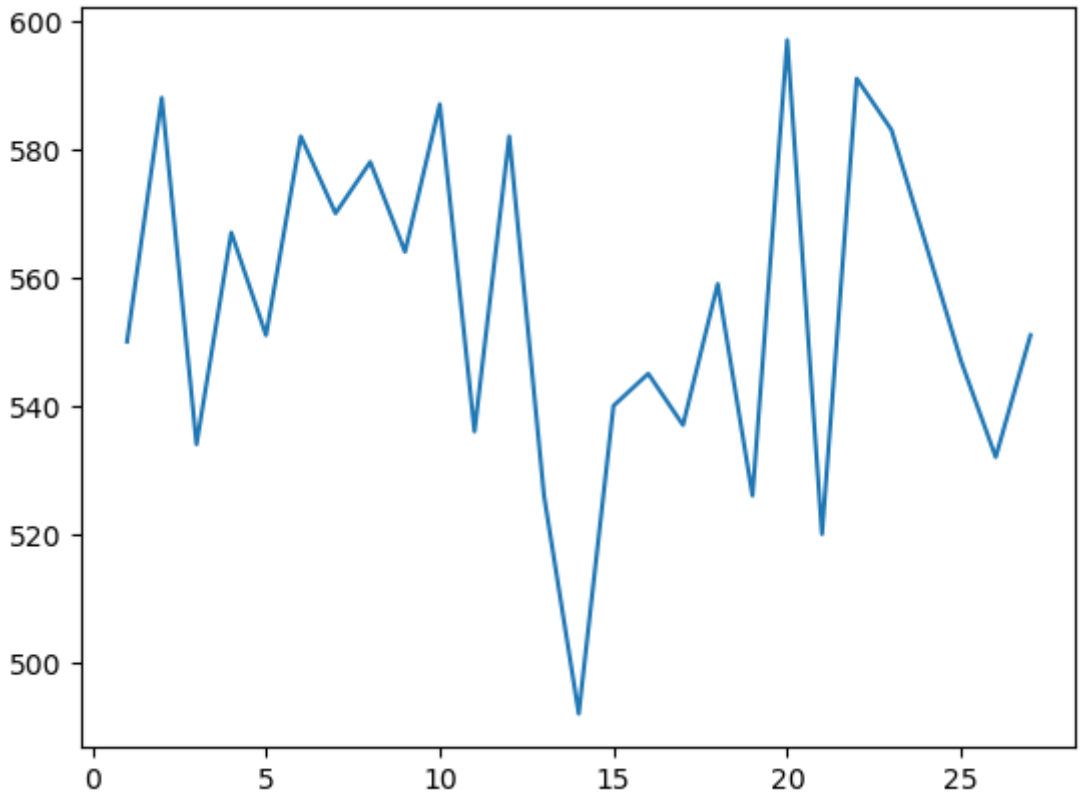


In [19]:

```
data['dat_of_month'].value_counts().sort_index().plot(kind='line')
```

Out[19]:

```
<AxesSubplot:>
```



In [20]:

```
data['delivery_time']=data['order_delivered_at']-data['order_placed_at']
```

In [21]:

```
data['delivery_time']
```

Out[21]:

```
0       0 days 00:17:08.443132
1       0 days 00:15:35.925241
2       0 days 00:15:06.291385
3       0 days 00:38:32.019399
4       0 days 00:16:10.619056
                 ...
14995   0 days 00:15:32.409378
14996   0 days 00:16:44.672912
14997   0 days 00:15:24.676238
14998   0 days 00:15:57.810358
14999   0 days 00:15:22.499311
Name: delivery_time, Length: 15000, dtype: timedelta64[ns]
```

In [23]:

```
data['delivery_time']=data['delivery_time'].dt.total_seconds()/60
```

In [25]:

```
data.head()
```

Out[25]:

	order_id	order_placed_at	order_delivered_at	dat_of_month	day_of_week	delivery_time
0	1523111	2023-03-01 00:00:59	2023-03-01 00:18:07.443132	1	Wednesday	17.140719
1	1523112	2023-03-01 00:03:59	2023-03-01 00:19:34.925241	1	Wednesday	15.598754
2	1523113	2023-03-01 00:07:22	2023-03-01 00:22:28.291385	1	Wednesday	15.104856
3	1523114	2023-03-01 00:07:47	2023-03-01 00:46:19.019399	1	Wednesday	38.533657
4	1523115	2023-03-01 00:09:03	2023-03-01 00:25:13.619056	1	Wednesday	16.176984

In [26]:

```
print('mean:',data['delivery_time'].mean())
print('median:',data['delivery_time'].median())
print('std:',data['delivery_time'].std())
```

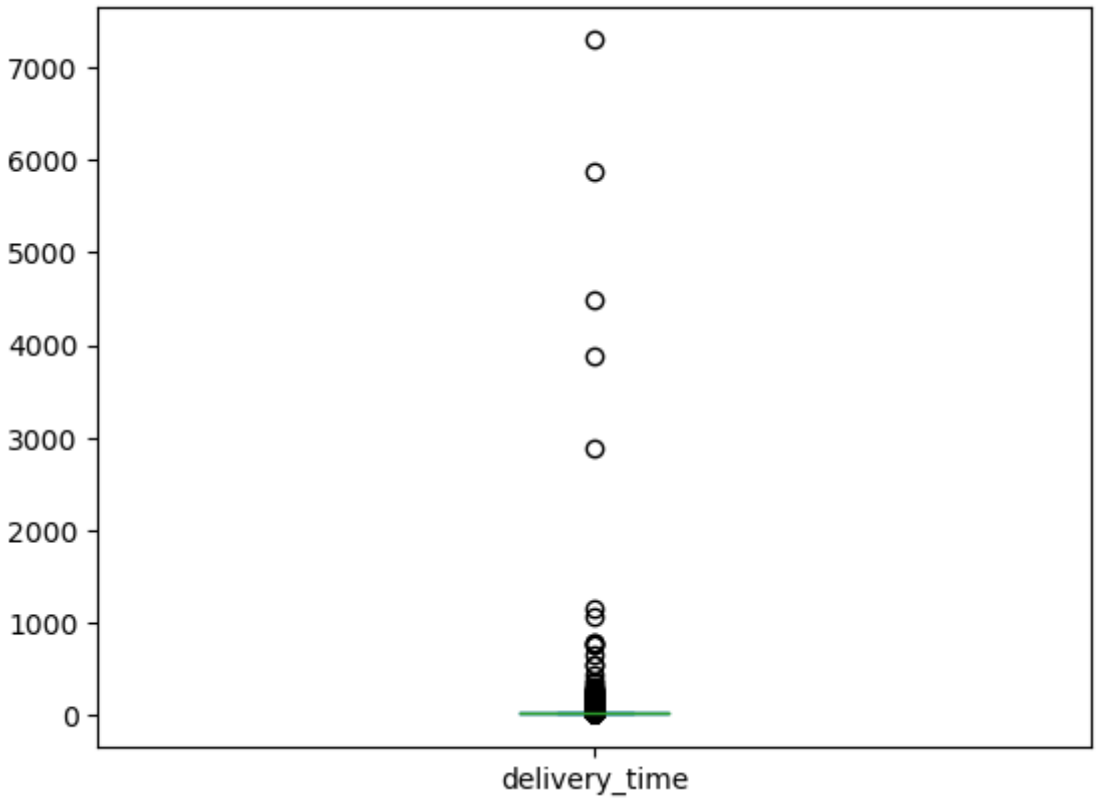
```
mean: 20.499389326467746
median: 15.797986066666668
std: 96.1603618021647
```

In [27]:

```
data['delivery_time'].plot(kind='box')
```

Out[27]:

```
<AxesSubplot:>
```



In [28]:

```
data=data[data['delivery_time']<40]
```

In [29]:

```
data['delivery_time'].plot(kind='box')
```

Out[29]:

```
<AxesSubplot:>
```

In [30]:

```
data['delivery_time'].quantile(0.95)
```

Out[30]:

```
23.596276583333314
```

In [ ]:

```

```