Supermarket Data Analysis

import the pandas module

In [1]:

```
import pandas as pd
```

read the file into your pandas DataFrame

In [2]:

```
path='C:/Users/HP/Downloads/'
file='supermarket_sales.xlsx'
```

display the last 5 records of your DataFrame

In [3]:

```
data=pd.read_excel(path+file)
data.tail()
```

Out[3]:

	Invoice ID	Branch	City	Customer type	Gender	Product line	Unit price	Quantity	Tax 5%	Total	Date	Time	Payment	Ra
995	233-67-5758	C	Naypyitaw	Normal	Male	Health and beauty	40.35	1	2.0175	42.3675	1/29/2019	13:46:00	Ewallet	
996	303-96-2227	B	Mandalay	Normal	Female	Home and lifestyle	97.38	10	48.6900	1022.4900	2019-02-03 00:00:00	17:16:00	Ewallet	
997	727-02-1313	A	Yangon	Member	Male	Food and beverages	31.84	1	1.5920	33.4320	2019-09-02 00:00:00	13:22:00	Cash	
998	347-56-2442	A	Yangon	Normal	Male	Home and lifestyle	65.82	1	3.2910	69.1110	2/22/2019	15:33:00	Cash	
999	849-09-3807	A	Yangon	Member	Female	Fashion accessories	88.34	7	30.9190	649.2990	2/18/2019	13:28:00	Cash	

What is the most Popular Product line in the dataset?

In [4]:

```
l=list(data['Product line'])
max=0
res=l[0]
for i in l:
    f=l.count(i)
    if f>max:
        max=f
        res=i
print(res)
```

```
Fashion accessories
```

Does the retail chain get more Male customers than Female Customers?

In [11]:

```
if data.Gender.max()=='Male':
    print("Yes")
```

```
Yes
```

What is the Median Unit Price at this store?

In [10]:

```
data['Unit price'].median()
```

Out[10]:

```
55.230000000000004
```

What is the Average Unit Price at this store?

In [12]:

```
s=data['Unit price'].sum()
n=data['Unit price'].count()
print(s/n)
```

```
55.67213
```

What is earliest date captured in this data? (Hint: Think along the lines of minimum and maximum!)

In [13]:

```
print(data.Date.value_counts().min())
```

```
6
```

What is the most common rating received by this retail chain?

In [14]:

```
l=list(data['Rating'])
max=0
res=l[0]
for i in l:
    f=l.count(i)
    if f>max:
        max=f
        res=i
print(res)
```

```
6.0
```

What is the second most popular means of payment by the customers at this retail chain?

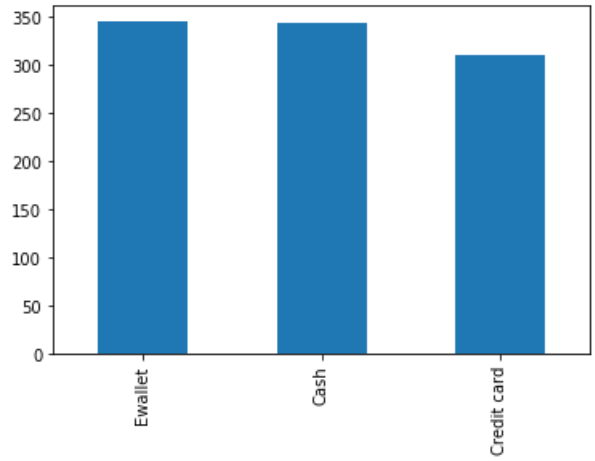
In [18]:

```
l=list(data['Payment'])
max=0
smax=0
res=l[0]
for i in l:
    f=l.count(i)
    if f>max:
        max=f
    else:
        smax=f
        res=i
print(res)
%matplotlib inline
data['Payment'].value_counts().plot.bar()
```

```
Cash
```

Out[18]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1ee5432a208>
```



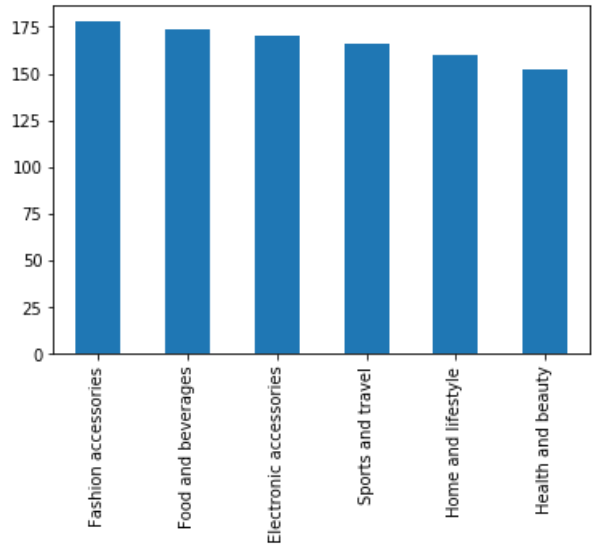
Create a Bar Plot showing the different product lines at the retail chain

In [17]:

```
%matplotlib inline
data['Product line'].value_counts().plot.bar()
```

Out[17]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x1ee53953b48>
```



In [ ]: