In [1]:

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

In [2]:

```
vstable = pd.read_csv("Video_Store_3.csv", index_col=0, na_values=["?"])

vstable.shape
```

Out[2]:

```
(50, 7)
```

In [3]:

```
vstable.head(10)
```

Out[3]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.0	32	2.5	Action	Yes
2	F	54000	33.0	12	3.4	Drama	No
3	F	32000	NaN	42	1.6	Comedy	No
4	NaN	59000	70.0	16	4.2	Drama	Yes
5	M	37000	35.0	25	3.2	Action	Yes
6	M	18000	20.0	29	1.7	Action	No
7	F	29000	NaN	19	3.8	Drama	No
8	M	74000	25.0	31	2.4	Action	Yes
9	NaN	38000	21.0	18	2.1	Comedy	No
10	F	65000	40.0	21	3.3	Drama	No

In [4]:

```
vstable.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 50 entries, 1 to 50
Data columns (total 7 columns):
 #   Column         Non-Null Count  Dtype  
---  ------         --------------  -----  
 0   Gender         45 non-null     object 
 1   Income         50 non-null     int64  
 2   Age            43 non-null     float64
 3   Rentals        50 non-null     int64  
 4   Avg Per Visit  50 non-null     float64
 5   Genre          50 non-null     object 
 6   Incidentals    50 non-null     object 
dtypes: float64(2), int64(2), object(3)
memory usage: 3.1+ KB
```

In [5]:

```
vstable.describe(include="all")
```

Out[5]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
count	45	50.000000	43.000000	50.000000	50.000000	50	50
unique	2	NaN	NaN	NaN	NaN	3	2
top	F	NaN	NaN	NaN	NaN	Drama	Yes
freq	23	NaN	NaN	NaN	NaN	20	26
mean	NaN	42300.000000	30.930233	26.320000	2.748000	NaN	NaN
std	NaN	21409.753642	11.650455	10.047723	0.898125	NaN	NaN
min	NaN	1000.000000	16.000000	9.000000	1.100000	NaN	NaN
25%	NaN	26750.000000	22.000000	19.000000	2.125000	NaN	NaN
50%	NaN	41000.000000	29.000000	25.000000	2.750000	NaN	NaN
75%	NaN	56750.000000	35.000000	32.000000	3.375000	NaN	NaN
max	NaN	89000.000000	70.000000	48.000000	4.700000	NaN	NaN

In [6]:

```
vstable.isnull()[0:10]
```

Out[6]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	True	False	False	False	False
4	True	False	False	False	False	False	False
5	False	False	False	False	False	False	False
6	False	False	False	False	False	False	False
7	False	False	True	False	False	False	False
8	False	False	False	False	False	False	False
9	True	False	False	False	False	False	False
10	False	False	False	False	False	False	False

In [7]:

```
vstable[vstable.isnull().any(axis=1)]
```

Out[7]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
3	F	32000	NaN	42	1.6	Comedy	No
4	NaN	59000	70.0	16	4.2	Drama	Yes
7	F	29000	NaN	19	3.8	Drama	No
9	NaN	38000	21.0	18	2.1	Comedy	No
14	M	45000	NaN	24	2.7	Drama	No
15	NaN	68000	30.0	36	2.7	Comedy	Yes
23	F	2000	NaN	30	2.5	Comedy	No
25	NaN	1000	16.0	25	1.4	Comedy	Yes
31	F	49000	NaN	15	3.2	Comedy	No
33	NaN	23000	25.0	28	2.7	Action	No
41	F	50000	NaN	17	1.4	Drama	No
46	F	57000	NaN	9	1.1	Drama	No

In [8]:

```
vstable[vstable.Gender.isnull()]
```

Out[8]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
4	NaN	59000	70.0	16	4.2	Drama	Yes
9	NaN	38000	21.0	18	2.1	Comedy	No
15	NaN	68000	30.0	36	2.7	Comedy	Yes
25	NaN	1000	16.0	25	1.4	Comedy	Yes
33	NaN	23000	25.0	28	2.7	Action	No

In [9]:

```
vstable[vstable.Age.isnull()]
```

Out[9]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
3	F	32000	NaN	42	1.6	Comedy	No
7	F	29000	NaN	19	3.8	Drama	No
14	M	45000	NaN	24	2.7	Drama	No
23	F	2000	NaN	30	2.5	Comedy	No
31	F	49000	NaN	15	3.2	Comedy	No
41	F	50000	NaN	17	1.4	Drama	No
46	F	57000	NaN	9	1.1	Drama	No

In [10]:

```
age_mean = vstable.Age.mean()
vstable.Age.fillna(age_mean, axis=0, inplace=True)
```

In [11]:

```
vstable.head(10)
```

Out[11]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.000000	32	2.5	Action	Yes
2	F	54000	33.000000	12	3.4	Drama	No
3	F	32000	30.930233	42	1.6	Comedy	No
4	NaN	59000	70.000000	16	4.2	Drama	Yes
5	M	37000	35.000000	25	3.2	Action	Yes
6	M	18000	20.000000	29	1.7	Action	No
7	F	29000	30.930233	19	3.8	Drama	No
8	M	74000	25.000000	31	2.4	Action	Yes
9	NaN	38000	21.000000	18	2.1	Comedy	No
10	F	65000	40.000000	21	3.3	Drama	No

In [12]:

```
vstable.drop(vstable[vstable.Gender.isnull()].index, axis=0).head(10)
```

Out[12]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.000000	32	2.5	Action	Yes
2	F	54000	33.000000	12	3.4	Drama	No
3	F	32000	30.930233	42	1.6	Comedy	No
5	M	37000	35.000000	25	3.2	Action	Yes
6	M	18000	20.000000	29	1.7	Action	No
7	F	29000	30.930233	19	3.8	Drama	No
8	M	74000	25.000000	31	2.4	Action	Yes
10	F	65000	40.000000	21	3.3	Drama	No
11	F	41000	22.000000	48	2.3	Drama	Yes
12	F	26000	22.000000	32	2.9	Action	Yes

In [13]:

```
vstable.head(10)
```

Out[13]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.000000	32	2.5	Action	Yes
2	F	54000	33.000000	12	3.4	Drama	No
3	F	32000	30.930233	42	1.6	Comedy	No
4	NaN	59000	70.000000	16	4.2	Drama	Yes
5	M	37000	35.000000	25	3.2	Action	Yes
6	M	18000	20.000000	29	1.7	Action	No
7	F	29000	30.930233	19	3.8	Drama	No
8	M	74000	25.000000	31	2.4	Action	Yes
9	NaN	38000	21.000000	18	2.1	Comedy	No
10	F	65000	40.000000	21	3.3	Drama	No

To permanently remove the rows with NaN Gender values, use the "inplace" parameter in the "drop" function.

In [14]:

```
vstable.drop(vstable[vstable.Gender.isnull()].index, axis=0, inplace=True)
vstable.head(10)
```

Out[14]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.000000	32	2.5	Action	Yes
2	F	54000	33.000000	12	3.4	Drama	No
3	F	32000	30.930233	42	1.6	Comedy	No
5	M	37000	35.000000	25	3.2	Action	Yes
6	M	18000	20.000000	29	1.7	Action	No
7	F	29000	30.930233	19	3.8	Drama	No
8	M	74000	25.000000	31	2.4	Action	Yes
10	F	65000	40.000000	21	3.3	Drama	No
11	F	41000	22.000000	48	2.3	Drama	Yes
12	F	26000	22.000000	32	2.9	Action	Yes

It is also possible to use the "dropna" function to drop all rows that have one or more NaN values.

In [15]:

```
vstable2 = pd.read_csv("Video_Store_3.csv", index_col=0, na_values=["?"])
```

In [16]:

```
vstable2.head(10)
```

Out[16]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.0	32	2.5	Action	Yes
2	F	54000	33.0	12	3.4	Drama	No
3	F	32000	NaN	42	1.6	Comedy	No
4	NaN	59000	70.0	16	4.2	Drama	Yes
5	M	37000	35.0	25	3.2	Action	Yes
6	M	18000	20.0	29	1.7	Action	No
7	F	29000	NaN	19	3.8	Drama	No
8	M	74000	25.0	31	2.4	Action	Yes
9	NaN	38000	21.0	18	2.1	Comedy	No
10	F	65000	40.0	21	3.3	Drama	No

In [17]:

```
vstable2.dropna(axis=0, inplace=True)
vstable2.shape
```

Out[17]:

```
(38, 7)
```

In [18]:

```
vstable2.head(10)
```

Out[18]:

	Gender	Income	Age	Rentals	Avg Per Visit	Genre	Incidentals
Cust ID							
1	M	45000	25.0	32	2.5	Action	Yes
2	F	54000	33.0	12	3.4	Drama	No
5	M	37000	35.0	25	3.2	Action	Yes
6	M	18000	20.0	29	1.7	Action	No
8	M	74000	25.0	31	2.4	Action	Yes
10	F	65000	40.0	21	3.3	Drama	No
11	F	41000	22.0	48	2.3	Drama	Yes
12	F	26000	22.0	32	2.9	Action	Yes
13	M	83000	46.0	14	3.6	Comedy	No
16	M	17000	19.0	26	2.2	Action	Yes