In [ ]:
```
import pandas as pd
```

In [ ]:
```
df = pd.read_csv('https://raw.githubusercontent.com/April03exo/BSI
```

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

Out [7]:

	age	income	gender	purchased
0	32	NaN	male	1.0
1	23	12323.0	female	1.0
2	23	4423.0	NaN	0.0
3	44	1345.0	female	NaN
4	43	54523.0	NaN	1.0

In [ ]:
```
df.head(2)
```

Out [8]:

	age	income	gender	purchased
0	32	NaN	male	1.0
1	23	12323.0	female	1.0

In [ ]:
```
df.tail(1)
```

Out [10]:

	age	income	gender	purchased
5	12	2312.0	female	1.0

In [ ]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6 entries, 0 to 5
Data columns (total 4 columns):
 #   Column     Non-Null Count  Dtype  
---  ------     --------------  -----  
 0   age        6 non-null      int64  
 1   income     5 non-null      float64
 2   gender     4 non-null      object 
 3   purchased  5 non-null      float64
dtypes: float64(2), int64(1), object(1)
memory usage: 320.0+ bytes
```

In [ ]:
```
df.describe()
```

Out [12]:

	age	income	purchased
count	6.000000	5.000000	5.000000
mean	29.500000	14985.200000	0.800000
std	12.565827	22519.713924	0.447214

	age	income	purchased
min	12.000000	1345.000000	0.000000
25%	23.000000	2312.000000	1.000000
50%	27.500000	4423.000000	1.000000
75%	40.250000	12323.000000	1.000000
max	44.000000	54523.000000	1.000000

In [ ]:
```
df.shape
```

Out [13]:
```
(6, 4)
```

In [ ]:
```
df.columns
```

Out [14]:
```
Index(['age', 'income', 'gender', 'purchased'], dtype='object')
```

In [ ]:
```
df = df.drop("age" , axis = 1)
```

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

Out [17]:

	income	gender	purchased
0	NaN	male	1.0
1	12323.0	female	1.0
2	4423.0	NaN	0.0
3	1345.0	female	NaN
4	54523.0	NaN	1.0

In [ ]:
```
df_filna = df.fillna('n/a')
```

In [ ]:
```
df_filna.head()
```

Out [4]:

	age	income	gender	purchased
0	32	n/a	male	1.0
1	23	12323.0	female	1.0
2	23	4423.0	n/a	0.0
3	44	1345.0	female	n/a
4	43	54523.0	n/a	1.0

In [ ]:
```
df = df.dropna()
```

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

Out [24]:

	income	gender	purchased
1	12323.0	female	1.0
5	2312.0	female	1.0

In [ ]:
```
df_sort = df.sort_values(by = 'gender', ascending = True)
```

In [ ]:
```
df_sort.head()
```

Out [7]:

	age	income	gender	purchased
1	23	12323.0	female	1.0
5	12	2312.0	female	1.0

In [ ]:
```
df_grouped = df.groupby('gender').mean()
```

In [ ]:
```
df_grouped.head()
```

Out [9]:

	age	income	purchased
gender			
female	17.5	7317.5	1.0

In [ ]:
```
df_apply = df['gender'] = df['age'].apply (lambda x: x*2)
```

In [ ]:
```
df_apply.head()
```

Out [14]:

	age
1	46
5	24

dtype: int64

In [ ]:
```
df1 = pd.DataFrame({'key': ['age','gender'] , 'values' : [44, 'fem
df2 = pd.DataFrame({'key': ['age','gender'] , 'values' : [32, 'mal
merged_df = pd.merge(df1, df2, on = 'key')
```

In [ ]:
```
merged_df.head()
```

Out [22]:

	Key	values_x	values_y
0	age	44	32
1	gender	female	male

In [ ]:
```
pivot_df = df.pivot_table(values = 'gender' , index = 'income', ag
```

In [ ]:
```
pivot_df.head()
```

Out [30]:

income	gender
2312.0	24.0
12323.0	46.0