EX NO : 4

DATE : 22/8/24

Demonstrate the operations of crosstab, joint probability, marginal probability, conditional probability using pandas

AIM:

To demonstrate the operations of crosstab, joint probability, marginal probability, conditional probability using pandas in jupyter notebook

Coffee shop sales dataset

```
data = {
  'SalesAmount': [5.00, 7.50, 3.75, 6.00, 8.00, 7.50, 4.50, 9.00, 10.00, 5.50],
  'CoffeeType': ['Espresso', 'Latte', 'Espresso', 'Latte', 'Cappuccino', 'Latte', 'Espresso', 'Cappuccino', 'Latte', 'Espresso'],
  'CustomerType': ['Regular', 'New', 'Regular', 'Regular', 'New', 'New', 'Regular', 'New', 'Regular', 'Regular'],
  'DayOfWeek': ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday', 'Monday', 'Tuesday', 'Wednesday'],
  'Quantity': [1, 2, 1, 2, 1, 3, 1, 1, 2, 1]
}
```

QUESTIONS:

1.Create a cross-tabulation of CoffeeType and CustomerType.

```
import pandas as pd
data = {
    'SalesAmount': [5.00, 7.50, 3.75, 6.00, 8.00, 7.50, 4.50, 9.00, 10.00, 5.50],
    'CoffeeType': ['Espresso', 'Latte', 'Espresso', 'Latte', 'Cappuccino', 'Latte', 'Espresso', 'Cappuccino', 'Latte', 'Espresso'],
    'CustomerType': ['Regular', 'New', 'Regular', 'Regular', 'New', 'New', 'Regular', 'New', 'Regular', 'Regular'],
    'DayOfWeek': ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday', 'Sunday', 'Monday', 'Tuesday', 'Wednesday'],
    'Quantity': [1, 2, 1, 2, 1, 3, 1, 1, 2, 1]
}

df = pd.DataFrame(data)
cross_tab = pd.crosstab(df['CoffeeType'], df['CustomerType'])
print(cross_tab)
```

```
CustomerType  New  Regular
CoffeeType
Cappuccino      2        0
Espresso        0        4
Latte           2        2
```

2. Calculate the joint probability of a sale being Latte and New customer.

```
joint_count = len(df[(df['CoffeeType'] == 'Latte') & (df['CustomerType'] == 'New')])
joint_probability = joint_count / len(df)
print(joint_probability)
```

```
0.2
```

3. Find the marginal probability of a sale amount being greater than 6.00

```
marginal_prob = len(df[df['SalesAmount'] > 6.00]) / len(df)
print(marginal_prob)
```

```
0.5
```

4. What is the conditional probability of a sale being Cappuccino given the SalesAmount is greater than 5.00?

```
cappuccino_given_sales = len(df[(df['CoffeeType'] == 'Cappuccino') & (df['SalesAmount'] > 5.00)]) / len(df[df['SalesAmount'] > 5.00])
print(cappuccino_given_sales)
```

```
0.2857142857142857
```

5. Compute the Pearson correlation coefficient between SalesAmount and Quantity.

```
correlation = df['SalesAmount'].corr(df['Quantity'])
print(correlation)
```

```
0.394489811164956
```

6. Create a cross-tabulation showing the average SalesAmount by CoffeeType and DayOfWeek

```
pivot_table = pd.pivot_table(df, values='SalesAmount', index='CoffeeType', columns='DayOfWeek', aggfunc='mean')
print(pivot_table)
```

```
DayOfWeek   Friday  Monday  Saturday  Sunday  Thursday  Tuesday  Wednesday
CoffeeType
Cappuccino     8.0     9.0       NaN     NaN       NaN      NaN        NaN
Espresso       NaN     5.0       NaN     4.5       NaN      NaN      4.625
Latte          NaN     NaN       7.5     NaN       6.0     8.75        NaN
```

7. Calculate the joint probability of a sale being on Wednesday and Latte.

```
joint_wednesday_latte = len(df[(df['DayOfWeek'] == 'Wednesday') & (df['CoffeeType'] == 'Latte')]) / len(df)
print(joint_wednesday_latte)
```

```
0.0
```

8. Find the marginal probability of having a Quantity of 2.

```
marginal_quantity_2 = len(df[df['Quantity'] == 2]) / len(df)
print(marginal_quantity_2)
```

```
0.3
```

9. What is the conditional probability of a sale being Espresso given that Quantity is 1?

```
espresso_given_quantity_1 = len(df[(df['CoffeeType'] == 'Espresso') & (df['Quantity'] == 1)]) / len(df[df['Quantity'] == 1])
print(espresso_given_quantity_1)
```

```
0.6666666666666666
```

10. Calculate the Pearson correlation coefficient between SalesAmount and encoded

```
from sklearn.preprocessing import LabelEncoder
label_encoder = LabelEncoder()
df['CustomerTypeEncoded'] = label_encoder.fit_transform(df['CustomerType'])
correlation_encoded = df['SalesAmount'].corr(df['CustomerTypeEncoded'])
print(correlation_encoded)
```

```
-0.5586225033606335
```

RUBRICS

Problem Understanding (10)	Implementation (20)	Viva (10)	Time Management (10)	Total (50)

RESULT

Thus the operations of crosstab, joint probability, marginal probability, conditional probability using pandas in jupyter notebook was successfully demonstrated and the output was verified