Sipna College of Engineering & Technology, Amravati.
Department of Computer Science & Engineering
Session 2022-2023

Branch :- Computer Sci. & Engg. **Class :- Final Year**
Subject :-Artificial Intelligence and Machine Learning **Sem :- VIII**
Teacher Manual

PRACTICAL NO 6

AIM: Demonstrate the concept of Bivariate and multivariate analysis.

S/W REQUIRED: Python

DATA SET USED: employee.csv

Bivariate Analysis:-

Bivariate analysis is one of the simplest forms of quantitative (statistical) analysis.
Bivariate analysis is stated to be an analysis of any concurrent relation between two variables or attributes. This study explores the relationship of two variables as well as the depth of this relationship to figure out if there are any discrepancies between two variables and any causes of this difference. Some of the examples are percentage table, scatter plot, etc.

Types of Bivariate Analysis

1. **Numerical and Numerical** – In this type, both the variables of bivariate data, independent and dependent, are having numerical values.
2. **Categorical and Categorical** – When both the variables are categorical.
3. **Numerical and Categorical** – When one variable is numerical and one is categorical.

Multivariate Analysis:-
multivariate analysis encompasses all statistical techniques that are used to analyze more than two variables at once. The aim is to find patterns and correlations between several variables simultaneously—allowing for a much deeper, more complex understanding of a given scenario than you'll get with bivariate analysis.

Multivariate data analysis techniques and examples

There are many different techniques for multivariate analysis, and they can be divided into two categories:

- Dependence techniques
- Interdependence techniques

Multivariate analysis techniques: Dependence vs. interdependence

When we use the terms “dependence” and “interdependence,” we’re referring to different types of relationships within the data. To give a brief explanation:

Dependence methods

Dependence methods are used when one or some of the variables are dependent on others. Dependence looks at cause and effect; in other words, can the values of two or more independent variables be used to explain, describe, or predict the value of another, dependent variable? To give a simple example, the dependent variable of “weight” might be predicted by independent variables such as “height” and “age.”

In machine learning, dependence techniques are used to build predictive models. The analyst enters input data into the model, specifying which variables are independent and which ones are dependent—in other words, which variables they want the model to predict, and which variables they want the model to use to make those predictions.

Interdependence methods

Interdependence methods are used to understand the structural makeup and underlying patterns within a dataset. In this case, no variables are dependent on others, so you’re not looking for causal relationships. Rather, interdependence methods seek to give meaning to a set of variables or to group them together in meaningful ways.

So: One is about the effect of certain variables on others, while the other is all about the structure of the dataset.

Implementation:

Bivariate Analysis:-

```
#Importing important libraries
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
#Reading the dataset
data=pd.read_csv('employee.csv')
data.shape
```

O/P:- `(1470, 35)`

```
# lets check the head of the dataset
data.head()
```

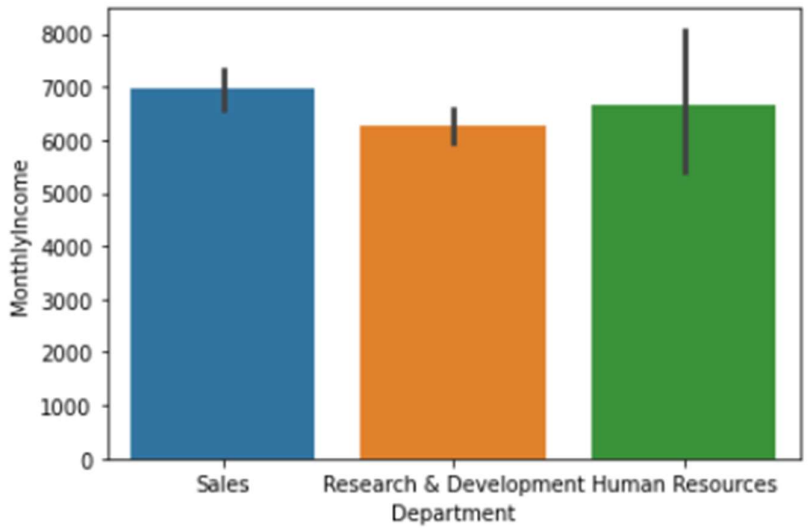
	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80

5 rows × 35 columns

Categorical vs Continuous Variables

```
sns.barplot(data['Department'], data['MonthlyIncome'])
```

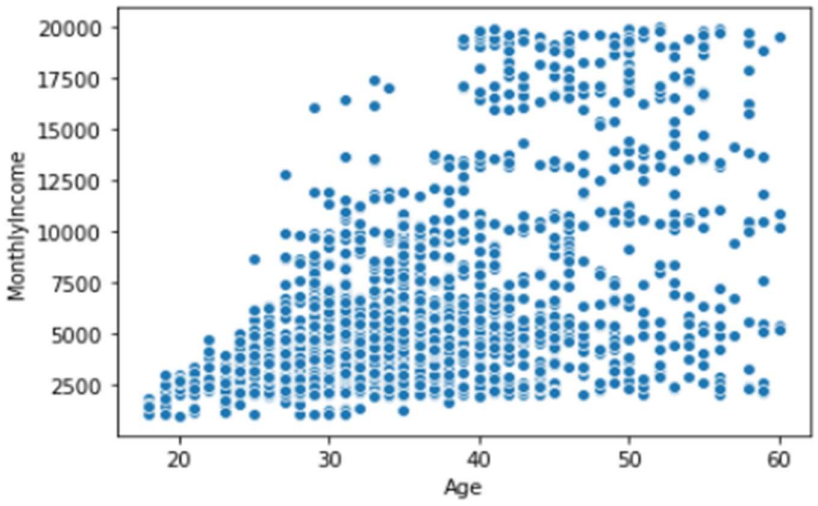
```
<matplotlib.axes._subplots.AxesSubplot at 0x1dd99744668>
```



Continous vs Continuous Variables

```
sns.scatterplot(data['Age'], data['MonthlyIncome'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x1dd9b827dd8>
```



Multivariate Analysis:-

```
#Importing important libraries
import pandas as pd
import seaborn as sns

#Reading the dataset
data=pd.read_csv('employee.csv')
data.shape
```

O/P:- `(1470, 35)`

```
#lets check the head of the dataset
data.head()
```

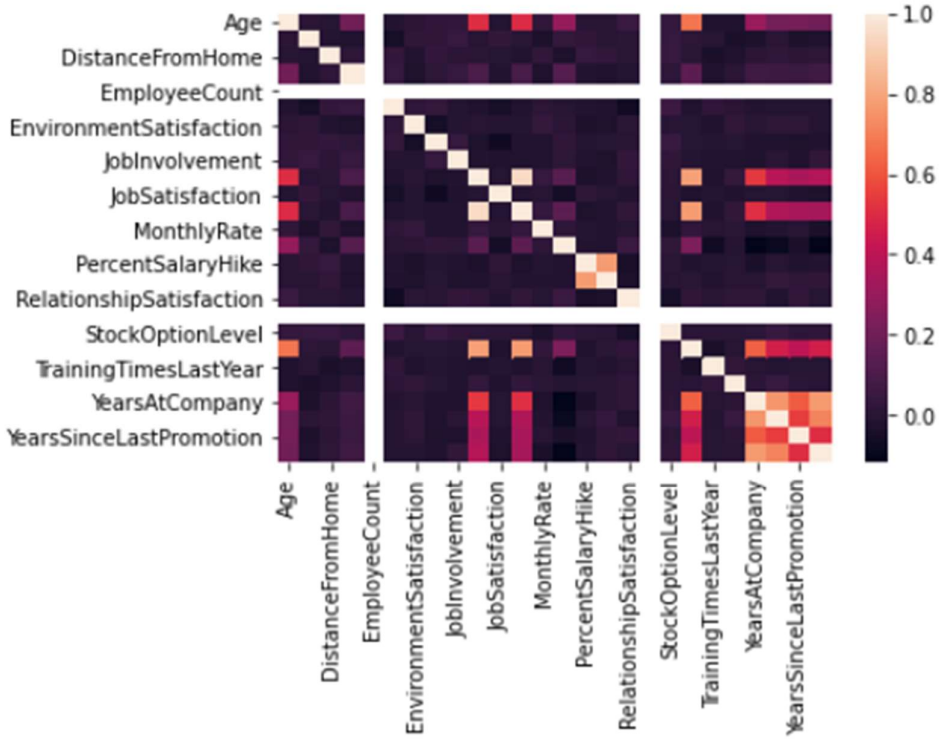
	Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	...	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears	TrainingTimesLastYear	WorkLifeBalance
0	41	Yes	Travel_Rarely	1102	Sales	1	2	Life Sciences	1	1	...	1	80	0	8	0	1
1	49	No	Travel_Frequently	279	Research & Development	8	1	Life Sciences	1	2	...	4	80	1	10	3	3
2	37	Yes	Travel_Rarely	1373	Research & Development	2	2	Other	1	4	...	2	80	0	7	3	3
3	33	No	Travel_Frequently	1392	Research & Development	3	4	Life Sciences	1	5	...	3	80	0	8	3	3
4	27	No	Travel_Rarely	591	Research & Development	2	1	Medical	1	7	...	4	80	1	6	3	3

5 rows × 35 columns

Correlation Heatmap

```
sns.heatmap(data.corr())
```

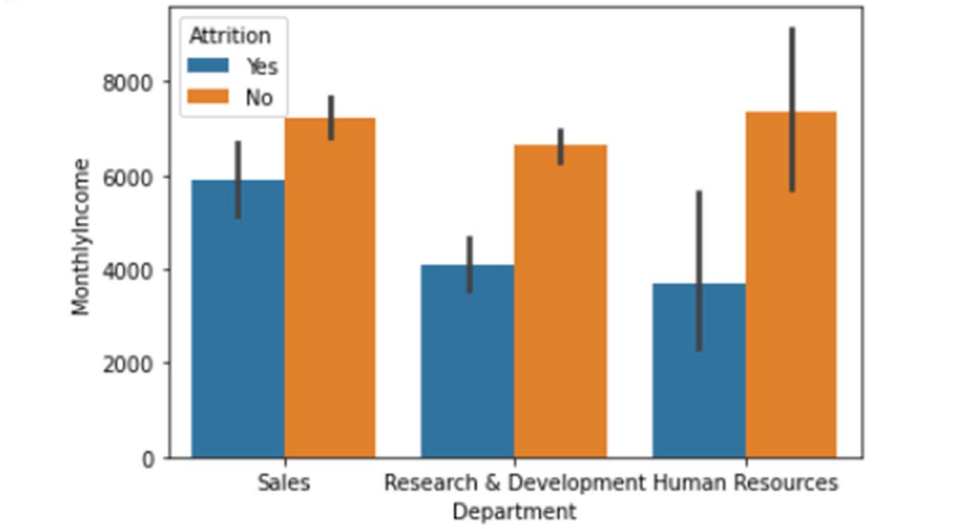
```
<matplotlib.axes._subplots.AxesSubplot at 0x18b97ed6b00>
```



Bar plot with Extra Variable

```
sns.barplot(data['Department'], data['MonthlyIncome'], data['Attrition'])
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x18b9a027550>
```



CONCLUSION: Thus we have implemented the concept of bivariate and multivariate analysis