

Python Unit III & IV - 5 Mark Theory Answers

1. What is Pandas? Explain its main data structures.

Pandas is a Python library used for working with structured/tabular data. It allows cleaning, filtering, analyzing, and transforming data.

Main structures:

- Series: 1D labeled data (like a column)
- DataFrame: 2D labeled table (like Excel)

Example:

```
import pandas as pd  
  
df = pd.DataFrame({'Name': ['Rayees'], 'Marks': [90]})
```

2. What is data cleaning and how is missing data handled in Pandas?

Data cleaning fixes incorrect or missing values.

Methods:

- dropna(): remove rows with missing values
- fillna(value): fill missing values with a fixed value

Example:

```
df = df.fillna(0)  
  
df = df.dropna()
```

3. Difference between Matplotlib and Seaborn.

Matplotlib is used for basic plotting; Seaborn is used for better-looking statistical plots.

Matplotlib: more code, low-level

Seaborn: simpler syntax, high-level

Python Unit III & IV - 5 Mark Theory Answers

Example:

```
plt.plot([1,2],[3,4])
```

```
sns.barplot(x=['A','B'], y=[10,20])
```

4. What is indexing and filtering in Pandas?

Indexing lets you access specific rows/columns.

Filtering lets you select data based on conditions.

Example:

```
df['Name'] # column
```

```
df.loc[0] # row
```

```
df[df['Marks'] > 85] # filter
```

5. What is data transformation in Pandas?

Changing data into better form using functions like `apply()`, `astype()`.

Example:

```
df['Marks'] = df['Marks'].apply(lambda x: x + 5)
```

```
df['Marks'] = df['Marks'].astype(float)
```

6. What is EDA? Why is it important?

EDA (Exploratory Data Analysis) helps understand the data better using stats and visuals.

Important for:

- Detecting errors/outliers
- Understanding data distribution

Example:

```
df.describe(), sns.boxplot(y=df['Marks'])
```

Python Unit III & IV - 5 Mark Theory Answers

7. What is correlation? How is it visualized?

Correlation shows how features are related (positive, negative, or none).

Example:

```
df.corr()  
  
sns.heatmap(df.corr(), annot=True)
```

8. What is data preprocessing and feature scaling?

Preprocessing prepares data for ML. Scaling puts values in the same range.

Example:

```
from sklearn.preprocessing import StandardScaler  
  
scaler = StandardScaler()  
  
scaled = scaler.fit_transform(df[['Marks']])
```

9. How to build and train a machine learning model?

Steps:

- Prepare data
- Split into train/test
- Fit model
- Predict

Example:

```
model = LinearRegression()  
  
model.fit(X_train, y_train)  
  
pred = model.predict(X_test)
```

10. How to evaluate a machine learning model?

Use metrics like MSE (for regression) to check how good the model is.

Python Unit III & IV - 5 Mark Theory Answers

Example:

```
from sklearn.metrics import mean_squared_error  
  
print(mean_squared_error(y_test, pred))
```