

Data Handling Using Pandas (Roman Urdu Notes)

1 Pandas ka Introduction

Pandas Python ki aik bohat powerful library hai jo data handling aur data analysis ke liye use hoti hai. Pandas ka use karke hum large aur messy data ko easily read, clean, filter aur analyze kar sakte hain. Real life me jo data hota hai wo aksar perfect nahi hota, is liye Pandas bohat useful tool hai.

2 Pandas Kyun Use Karte Hain

Pandas is liye use hoti hai kyun ke ye data ke sath kaam karna bohat easy bana deti hai. Is ki madad se hum Excel, CSV aur database files ka data read kar sakte hain. Pandas missing values handle karti hai, data ko sort karti hai aur analysis ke liye ready karti hai. Business, finance, healthcare aur education me Pandas ka bohat use hota hai.

3 Pandas Install aur Import Karna

Sab se pehle Pandas ko install karna hota hai. Ye kaam sirf aik dafa hota hai.

```
pip install pandas
```

Install karne ke baad Pandas ko Python me import kiya jata hai.

```
import pandas as pd
```

Yahan pd short name hai jo code ko short aur readable banata hai.

4 Pandas ke Data Structures

4.1 Series

Series Pandas ka aik one-dimensional data structure hota hai. Is ka matlab hai ke ye sirf aik column ka data store karta hai. Har value ke sath aik index hota hai jo us value ko

identify karta hai.

```
marks = pd.Series([85, 90, 78, 92])  
print(marks)
```

Series aksar tab use hoti hai jab sirf aik variable ka data handle karna ho.

4.2 DataFrame

DataFrame Pandas ka sab se important data structure hai. Ye two-dimensional hota hai, yani rows aur columns dono hotay hain. Ye bilkul Excel sheet jaisa hota hai.

```
data = {  
    "Name": ["Ali", "Sara", "Ahmed"],  
    "Marks": [85, 90, 78],  
    "Age": [20, 21, 19]  
}  
  
df = pd.DataFrame(data)  
print(df)
```

Zyada tar data analysis ka kaam DataFrame par hi hota hai.

5 Data Load Karna

Pandas different file formats se data load kar sakti hai.

5.1 CSV File Load Karna

```
df = pd.read_csv("data.csv")
```

5.2 Excel File Load Karna

```
df = pd.read_excel("data.xlsx")
```

6 Data ko Samajhna

Data load karne ke baad usko samajhna bohat zaroori hota hai.

```
df.head()  
df.tail()  
df.shape
```

```
df.columns
```

`head()` pehli rows dikhata hai, `tail()` last rows dikhata hai aur `shape` rows aur columns ki total count batata hai.

7 Data Information Check Karna

```
df.info()
```

Ye function batata hai ke kaun se columns hain, unka data type kya hai aur missing values kitni hain.

```
df.describe()
```

Ye numerical data ka statistical summary deta hai jaise mean, min aur max.

8 Missing Values Handle Karna

Real world data me missing values hona common baat hai.

8.1 Missing Values Check Karna

```
df.isna().sum()
```

Is se pata chalta hai ke kis column me kitni missing values hain.

8.2 Missing Values Remove Karna

```
df.dropna()
```

Ye un rows ko delete kar deta hai jahan missing values hoti hain.

8.3 Missing Values Fill Karna

```
df["Marks"].fillna(df["Marks"].mean(), inplace=True)
```

Mean se fill karna aik common aur acha tareeqa hai.

9 Columns aur Rows Select Karna

9.1 Column Select Karna

```
df["Marks"]  
df[["Name", "Marks"]]
```

9.2 Row Select Karna

```
df.loc[0]  
df.iloc[1]
```

loc label ke zariye aur iloc position ke zariye rows select karta hai.

10 Data Filtering

Filtering ka matlab hai condition ke basis par data nikalna.

```
df[df["Marks"] > 80]
```

```
df[(df["Marks"] > 80) & (df["Age"] > 20)]
```

Ye real analysis me bohat useful hota hai.

11 Data Sorting

```
df.sort_values("Marks")  
df.sort_values("Marks", ascending=False)
```

Sorting ranking aur comparison ke liye use hoti hai.

12 Grouping aur Analysis

```
df.groupby("Result").mean()
```

Grouping ka use data ko categories me divide karke summary nikalne ke liye hota hai.

13 Data Type Change Karna

```
df["Age"] = df["Age"].astype(int)
```

Correct data type analysis ke liye bohat zaroori hota hai.

14 Apply Function ka Use

```
df["Grade"] = df["Marks"].apply(  
    lambda x: "A" if x >= 85 else "B"  
)
```

Apply function custom logic lagane ke liye use hota hai.

15 Data Save Karna

```
df.to_csv("clean_data.csv", index=False)  
df.to_excel("output.xlsx", index=False)
```

Cleaned data ko save karna future use ke liye zaroori hota hai.

16 Conclusion

Pandas aik bohat important Python library hai jo data handling aur analysis ko bohat easy bana deti hai. Agar Pandas achi tarah samajh li jaye to data science aur analytics ke concepts bohat asaan ho jatay hain.