**EXERCISES ON DATAFRAMES (30.08.2024)**

**Handling Missing Values\*\*  
#1. Create a DataFrame with missing values:   
# ```python  
# data = {  
# "Name": ["Amit", "Neha", "Raj", "Priya"],  
# "Age": [28, None, 35, 29],  
# "City": ["Delhi", "Mumbai", None, "Chennai"]  
# }  
# df = pd.DataFrame(data)  
# ```  
#2. Fill missing values in the `"Age"` column with the average age.  
#3. Drop rows where any column has missing data.  
import pandas as pd**  
  
data = {  
 "Name": ["Amit", "Neha", "Raj", "Priya"],  
 "Age": [28, None, 35, 29],  
 "City": ["Delhi", "Mumbai", None, "Chennai"]  
}  
df = pd.DataFrame(data)  
print(df)  
  
avg\_age = df['Age'].mean()  
df['Age']=df["Age"].fillna(avg\_age)  
df\_cleaned = df.dropna()  
print(df\_cleaned)  
  
**2.Adding and Removing Columns\*\*  
#1. Add a new column `"Salary"` with the following values: `[50000, 60000, 70000, 65000]`.  
#2. Remove the `"City"` column from the DataFrame.**  
df['Salary'] = [50000, 60000, 70000, 65000]  
df.drop(columns=['City'], inplace=True)  
print(df)  
  
**3.Sorting Data  
#1. Sort the DataFrame by `"Age"` in ascending order.  
#2. Sort the DataFrame first by `"City"` and then by `"Age"` in descending order.**  
  
df\_sorted\_by\_age = df.sort\_values(by='Age', ascending=True)  
df\_sorted\_by\_city\_and\_age = df.sort\_values(by=['City', 'Age'], ascending=[True, False])

**4. Grouping and Aggregation\*\*  
#1. Group the DataFrame by `"City"` and calculate the average `"Age"` for each city.  
#2. Group the DataFrame by `"City"` and `"Age"`, and count the number of occurrences for each group.**avg\_age\_by\_city = df.groupby('City')['Age'].mean()  
count\_by\_city\_age = df.groupby(['City', 'Age']).size()  
  
**5.Merging DataFrames\*\*  
#1. Create two DataFrames:A  
# ```python  
# df1 = pd.DataFrame({  
# "Name": ["Amit", "Neha", "Raj"],  
# "Department": ["HR", "IT", "Finance"]  
# })  
  
# df2 = pd.DataFrame({  
# "Name": ["Neha", "Raj", "Priya"],  
# "Salary": [60000, 70000, 65000]  
# })  
#.2 Merge `df1` and `df2` on the `"Name"` column using an inner join.  
#3. Merge the same DataFrames using a left join.**  
  
import pandas as pd  
  
df1 = pd.DataFrame({  
 "Name": ["Amit", "Neha", "Raj"],  
 "Department": ["HR", "IT", "Finance"]  
})  
  
df2 = pd.DataFrame({  
 "Name": ["Neha", "Raj", "Priya"],  
 "Salary": [60000, 70000, 65000]  
})  
  
merged\_inner = pd.merge(df1, df2, on="Name", how="inner")  
merged\_left = pd.merge(df1, df2, on="Name", how="left")