

The Data Cleaning :

1. Assignment -Missing Values (Quick Revision)

In [44]: `import pandas as pd`

```
data = {
    'Name': ['ramiz', 'love', 'koko', None, 'Max'],
    'Age': [21, 22, None, 20, 23],
    'course': ['ai', 'Data Science', 'game dev', None, 'Python'],
    'Marks': [100, 200, 300, None, None]
}
df = pd.DataFrame(data)
print(df)

print("The Showing the total Data   NaN:\n")
print(df.isnull())
print("\n")
print("Showing the total data sum :\n")

print(df.isnull().sum())

print("Fill the missing value in the data : ( Age , marks , 'course'):\n")

df['Marks'] = df['Marks'].fillna(df['Marks'].mean())

df['Age'] = df['Age'].fillna(0)

df['course'] = df['course'].fillna('Unkmon')

df_dropna = df.dropna(subset=['Name'])

print(df_dropna)
```

	Name	Age	course	Marks
0	ramiz	21.0	ai	100.0
1	love	22.0	Data Science	200.0
2	koko	NaN	game dev	300.0
3	None	20.0	None	NaN
4	Max	23.0	Python	NaN

The Showing the total Data NaN:

	Name	Age	course	Marks
0	False	False	False	False
1	False	False	False	False
2	False	True	False	False
3	True	False	True	True
4	False	False	False	True

Showing the total data sum :

```
Name      1
Age        1
course     1
Marks      2
dtype: int64
```

Fill the missing value in the data : (Age , marks , 'course'):

	Name	Age	course	Marks
0	ramiz	21.0	ai	100.0
1	love	22.0	Data Science	200.0
2	koko	0.0	game dev	300.0
4	Max	23.0	Python	200.0

2. Assignment - Duplicate & Replace & Rename

In [42]: `import pandas as pd`

```
data = {
    'Name': ["Ramiz", "Aman", "Neha", "Aman", "Zara"],
    'Age': [21, 22, 23, 22, 20],
```

```

    "Course": ["Python", "AI", "Data Science", "AI", "Web"],
    "Marks": [88, 76, 95, 76, 92]
}

df = pd.DataFrame(data)

print("Original DataFrame:")
print(df)

# 1. Find duplicates
print("\nDuplicate Rows:")
print(df[df.duplicated()])

# 2. Remove duplicates
df = df.drop_duplicates()

# 3. Replace Course name 'Web' → 'Web Development'
df['Course'] = df['Course'].replace("Web", "Web Development")

# 4. Rename column 'Marks' → 'Score'
df = df.rename(columns={"Marks": "Score"})

print("\nCleaned DataFrame:")
print(df)

```

Original DataFrame:

	Name	Age	Course	Marks
0	Ramiz	21	Python	88
1	Aman	22	AI	76
2	Neha	23	Data Science	95
3	Aman	22	AI	76
4	Zara	20	Web	92

Duplicate Rows:

	Name	Age	Course	Marks
3	Aman	22	AI	76

Cleaned DataFrame:

	Name	Age	Course	Score
0	Ramiz	21	Python	88
1	Aman	22	AI	76
2	Neha	23	Data Science	95
4	Zara	20	Web Development	92

The Mini Project on Two Assignment bases

In [43]: `import pandas as pd`

```

data = {

    'Name': ['Aman', 'Aman', 'Koko', 'Sahil', 'Ramiz', None],

    'Age': [22, 22, 23, None, 25, 26],

    'Course': ['ai', 'ai', 'web dev', None, "Game Dev", 'web dev'],

    'Marks': [100, 100, None, 200, 300, 400, ]
}

df =pd.DataFrame(data)

print("\n Chick the Nan in the Data : \n")

print(df.isnull())

print("\n Chick the Nan Count \n")

print(df.isnull().sum())

df ['Marks'] =df['Marks'].fillna(df['Marks'].mean ())

df['Age'] =df['Age'].fillna(0)

df ['Course'] =df['Course'].fillna("Unkomon")

df_dropna =df.dropna(subset=['Name'])

```

```

print(df_dropped)

print("Now Find the Duplicated value in the data : \n")
print(df[df.duplicated])

print(" Removed the Duplicated data: \n")

df =df.drop_duplicates()

df ['Course'] =df['Course'].replace('web dev', 'Web Development')

df =df.rename(columns={"Marks": "Score"})

print("The Hole Data Clining : \n", df)

```

Chick the Nan in the Data :

	Name	Age	Course	Marks
0	False	False	False	False
1	False	False	False	False
2	False	False	False	True
3	False	True	True	False
4	False	False	False	False
5	True	False	False	False

Chick the Nan Count

	Name	Age	Course	Marks
0	Ramiz	21.0	Python	88.000000
1	Aman	22.0	AI	91.666667
2	Neha	0.0	Data Science	95.000000
3	Zara	20.0	Unkmon	92.000000

dtype: int64

Now Find the Duplicated value in the data :

	Name	Age	Course	Marks
1	Aman	22.0	ai	100.0

Removed the Duplicated data:

The Hole Data Clining :

	Name	Age	Course	Score
0	Aman	22.0	ai	100.0
2	Koko	23.0	Web Development	220.0
3	Sahil	0.0	Unkomon	200.0
4	Ramiz	25.0	Game Dev	300.0
5	None	26.0	Web Development	400.0

In []: