Pandas

In [58]:
```
import pandas as pd

# Create a dictionary of data
data = {'Name': ['Alice','Alice', 'Bob', 'Charlie', 'David','David'],
        'Age': [25,20, 30, 30, 40, 40],
        'City': ['New York','New York', 'San Francisco', 'Los Angeles'

# Create a DataFrame from the dictionary
df = pd.DataFrame(data)

# Display the DataFrame
print(df)
```

```
      Name  Age           City
0    Alice   25       New York
1    Alice   20       New York
2      Bob   30  San Francisco
3  Charlie   30    Los Angeles
4    David   40        Chicago
5    David   40        Chicago
```

In [59]:
```
# Removing duplicates based on one column
df_no_duplicates = df.drop_duplicates(subset=['Name'])
print(df_no_duplicates)
```

```
      Name  Age           City
0    Alice   25       New York
2      Bob   30  San Francisco
3  Charlie   30    Los Angeles
4    David   40        Chicago
```

In [60]:
```

# Removing duplicates based on multiple columns
df_no_duplicates = df.drop_duplicates(subset=['Name', 'City'])
print(df_no_duplicates)
```

```
      Name  Age           City
0    Alice   25       New York
2      Bob   30  San Francisco
3  Charlie   30    Los Angeles
4    David   40        Chicago
```

In [61]:
```
# Removing duplicates based on all columns
df_no_duplicates = df.drop_duplicates(data)
print(df_no_duplicates)
```

```
      Name  Age           City
0    Alice   25       New York
1    Alice   20       New York
2      Bob   30  San Francisco
3  Charlie   30    Los Angeles
4    David   40        Chicago
```

In [62]:
```
data_1=df.duplicated(data)
```

In [63]:
```
df.duplicated(subset=['Name'])
```

Out[63]:
```
0    False
1     True
2    False
3    False
4    False
5     True
dtype: bool
```

In [64]:
```

# Removing duplicates and keeping one occurrence
df_no_duplicates = df[df.duplicated(subset=['Name'], keep=False) ]

print(df_no_duplicates)

```

```
    Name  Age      City
0  Alice   25  New York
1  Alice   20  New York
4  David   40   Chicago
5  David   40   Chicago
```

In [65]:
```
df_no_duplicates = df.drop_duplicates(subset=['Name'])
print(df_no_duplicates)
```

```
      Name  Age           City
0    Alice   25       New York
2      Bob   30  San Francisco
3  Charlie   30    Los Angeles
4    David   40        Chicago
```

In [ ]: