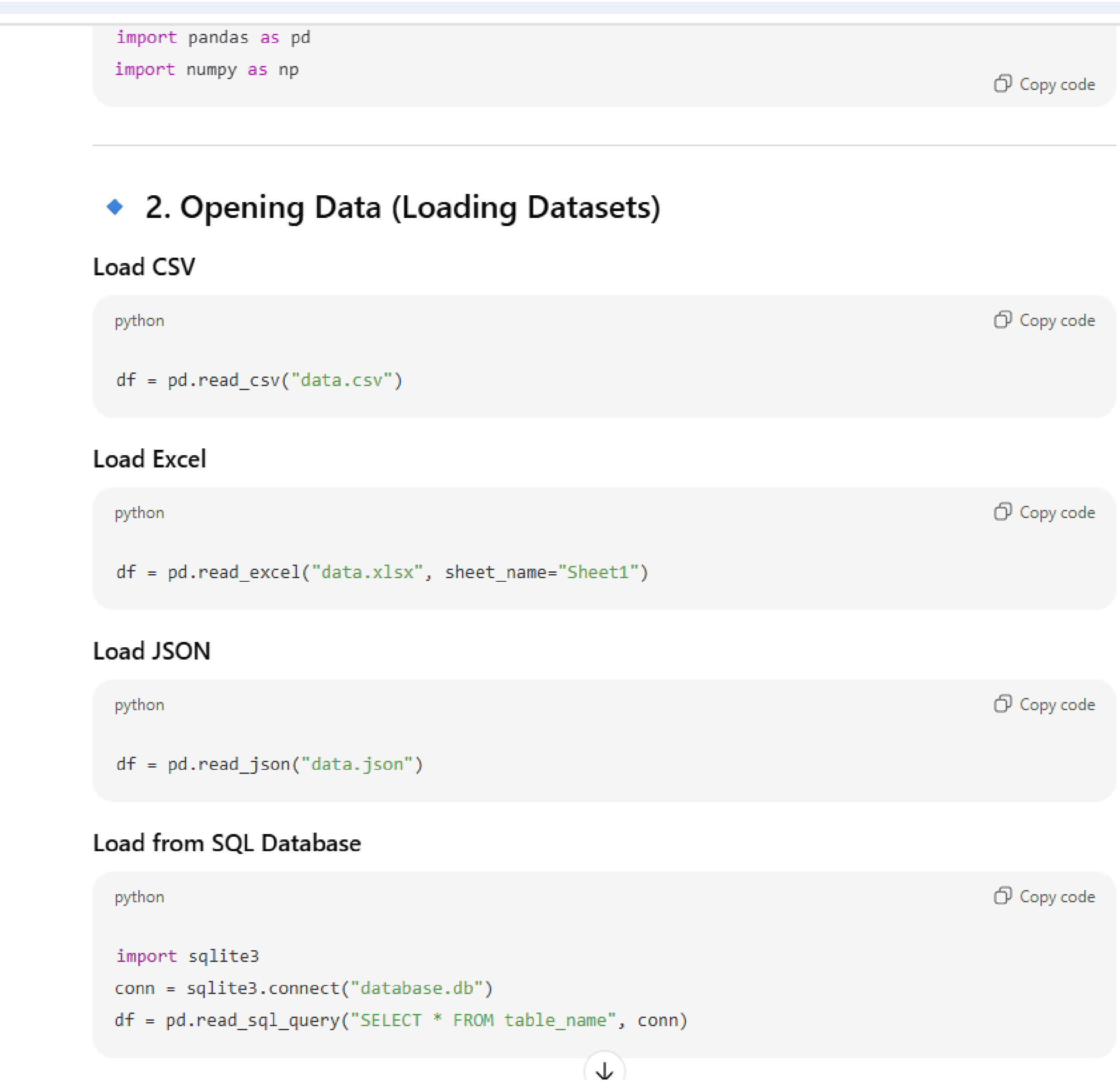


```python
import pandas as pd
import numpy as np
```

◆ 2. Opening Data (Loading Datasets)

Load CSV

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

Load Excel

```python
df = pd.read_excel("data.xlsx", sheet_name="Sheet1")
```

Load JSON

```python
df = pd.read_json("data.json")
```

Load from SQL Database

```python
import sqlite3
conn = sqlite3.connect("database.db")
df = pd.read_sql_query("SELECT * FROM table_name", conn)
```

Load from Clipboard (Copy-Paste data)

python Copy code

```python
df = pd.read_clipboard()
```

Save Data

python Copy code

```python
df.to_csv("cleaned_data.csv", index=False)
df.to_excel("cleaned_data.xlsx", index=False)
```

🔹 3. Basic Data Exploration

python Copy code

```python
df.head()        # first 5 rows
df.tail()        # last 5 rows
df.shape         # (rows, columns)
df.info()        # column info
df.describe()    # summary statistics
df.columns       # column names
df.dtypes        # data types
```

◆ 4. Selecting & Filtering Data

Select Column

python Copy code

```python
df['column_name']       # single column
df[['col1', 'col2']]    # multiple columns
```

Select Row by Index

python Copy code

```python
df.iloc[0]              # first row
df.iloc[0:5]            # first 5 rows
```

Select Row by Condition

python Copy code

```python
df[df['age'] > 30]                      # filter by condition
df[(df['age'] > 30) & (df['gender']=="M")]  # multiple conditions
```

◆ 5. Sorting

python Copy code

```python
df.sort_values("age")                # ascending
df.sort_values("age", ascending=False) # descending
```

◆ 6. Handling Missing Data

python

Copy code

```python
df.isnull().sum()                # check missing values
df.dropna(inplace=True)           # remove missing rows
df.fillna(0, inplace=True)        # replace NaN with 0
df['age'].fillna(df['age'].mean(), inplace=True)  # replace with mean
```

◆ 7. Adding / Modifying Columns

python

Copy code

```python
df['new_col'] = df['col1'] + df['col2']   # create new column
df['age_group'] = np.where(df['age'] > 30, "Above 30", "Below 30")
```

◆ 8. Removing Columns & Rows

python

Copy code

```python
df.drop("col_name", axis=1, inplace=True)   # drop column
df.drop([0,1], axis=0, inplace=True)        # drop rows by index
```

◆ 9. Grouping & Aggregation

python Copy code

```python
df.groupby("gender")["age"].mean()
df.groupby("department").agg({"salary":"mean", "age":"max"})
```

◆ 10. Joining & Merging

python Copy code

```python
pd.merge(df1, df2, on="id")                # inner join
pd.merge(df1, df2, on="id", how="left")    # left join
```

◆ 11. Pivot Tables

python Copy code

```python
df.pivot_table(values="sales", index="region", columns="year", aggfunc="sum")
```

◆ 12. Apply Functions

python Copy code

```python
df['name'] = df['name'].apply(lambda x: x.upper())
```

◆ 13. Unique & Value Counts

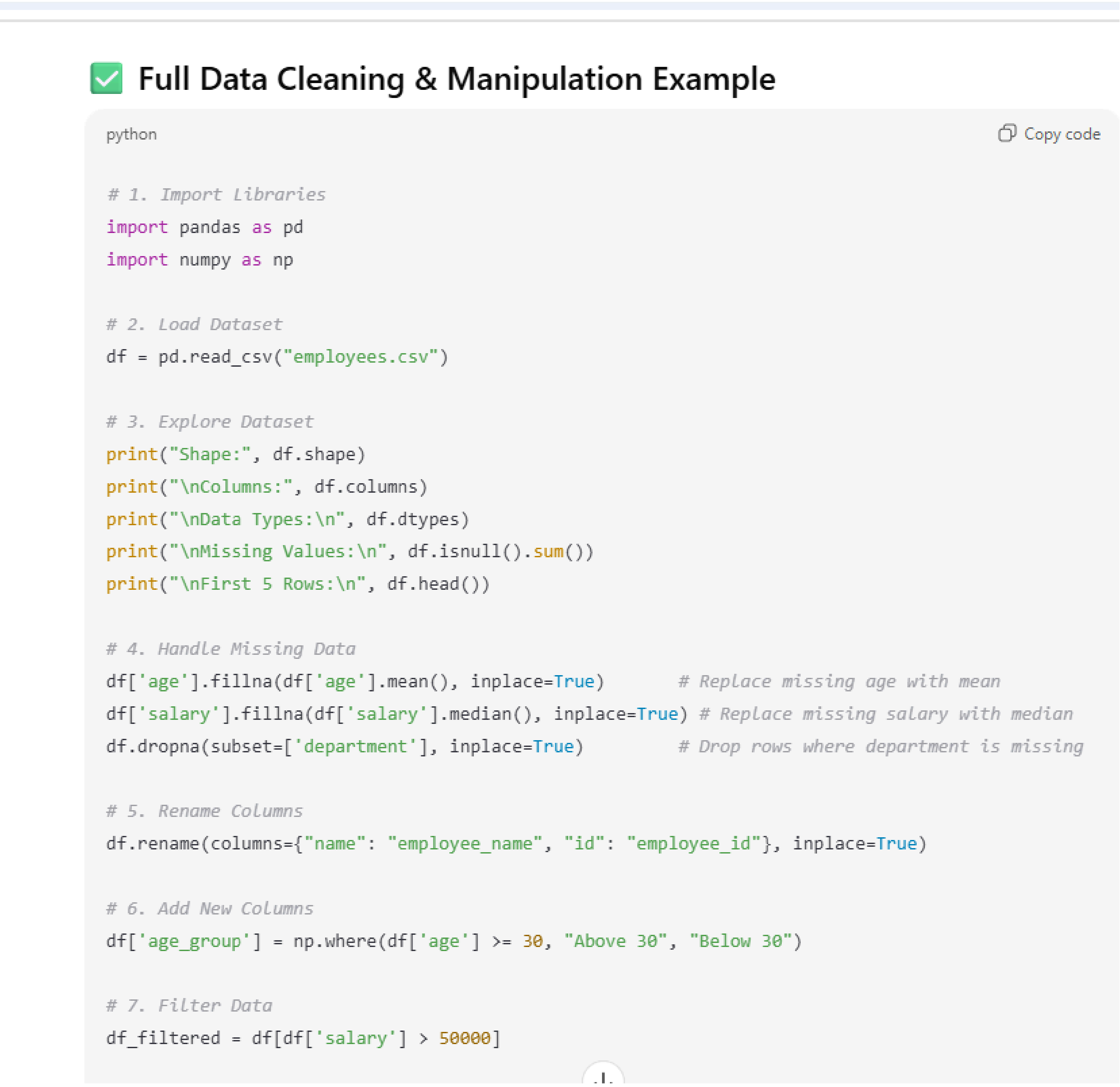
```python
df['gender'].unique()
df['gender'].value_counts()
```

◆ 14. Rename Columns

```python
df.rename(columns={"old_name":"new_name"}, inplace=True)
```

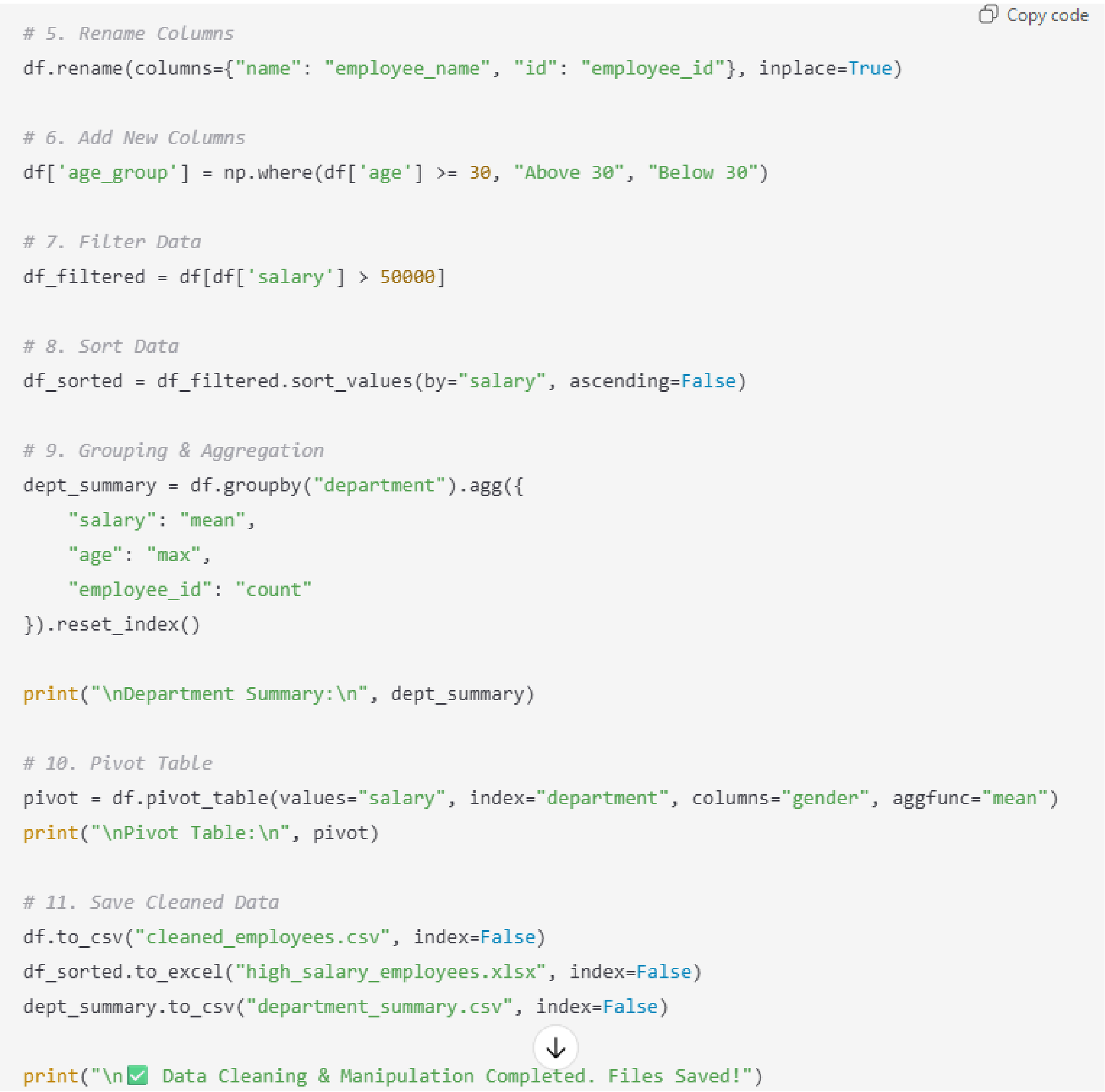
◆ 15. Export Clean Data

```python
df.to_csv("final_data.csv", index=False)
df.to_excel("final_data.xlsx", index=False)
```



✅ Full Data Cleaning & Manipulation Example

```python
# 1. Import Libraries
import pandas as pd
import numpy as np

# 2. Load Dataset
df = pd.read_csv("employees.csv")

# 3. Explore Dataset
print("Shape:", df.shape)
print("\nColumns:", df.columns)
print("\nData Types:\n", df.dtypes)
print("\nMissing Values:\n", df.isnull().sum())
print("\nFirst 5 Rows:\n", df.head())

# 4. Handle Missing Data
df['age'].fillna(df['age'].mean(), inplace=True)       # Replace missing age with mean
df['salary'].fillna(df['salary'].median(), inplace=True) # Replace missing salary with median
df.dropna(subset=['department'], inplace=True)         # Drop rows where department is missing

# 5. Rename Columns
df.rename(columns={"name": "employee_name", "id": "employee_id"}, inplace=True)

# 6. Add New Columns
df['age_group'] = np.where(df['age'] >= 30, "Above 30", "Below 30")

# 7. Filter Data
df_filtered = df[df['salary'] > 50000]
```



```python
# 5. Rename Columns
df.rename(columns={"name": "employee_name", "id": "employee_id"}, inplace=True)

# 6. Add New Columns
df['age_group'] = np.where(df['age'] >= 30, "Above 30", "Below 30")

# 7. Filter Data
df_filtered = df[df['salary'] > 50000]

# 8. Sort Data
df_sorted = df_filtered.sort_values(by="salary", ascending=False)

# 9. Grouping & Aggregation
dept_summary = df.groupby("department").agg({
    "salary": "mean",
    "age": "max",
    "employee_id": "count"
}).reset_index()

print("\nDepartment Summary:\n", dept_summary)

# 10. Pivot Table
pivot = df.pivot_table(values="salary", index="department", columns="gender", aggfunc="mean")
print("\nPivot Table:\n", pivot)

# 11. Save Cleaned Data
df.to_csv("cleaned_employees.csv", index=False)
df_sorted.to_excel("high_salary_employees.xlsx", index=False)
dept_summary.to_csv("department_summary.csv", index=False)

print("\n✅ Data Cleaning & Manipulation Completed. Files Saved!")
```