```
from google.colab import files
uploaded = files.upload()
```

Choose Files school_stu…plicates.xlsx

school_students_duplicates.xlsx(application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) - 6222 bytes, last modified: 11/4/2025 - 100% done

Saving school_students_duplicates.xlsx to school_students_duplicates (1).xlsx

```
import pandas as pd
```

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

```
print("----- Original Data (with duplicates) -----")
display(df)
```

```
----- Original Data (with duplicates) -----
```

	Student_ID	Name	Class	Section	Age	Gender	Marks	Email	Phone
0	S001	Arjun Mehta	10	A	15.0	Male	88.0	arjun.mehta@gmail.com	9.876543e+09
1	S002	Priya Sharma	9	B	14.0	Female	92.0	priya.sharma@gmail.com	9.876543e+09
2	S003	arjun mehta	10	A	15.0	male	88.0	arjun.mehta@gmail.com	9.876543e+09
3	S004	Rahul Verma	10	A	16.0	Male	76.0	rahulv@gmail.com	9.876543e+09
4	S005	Sneha Reddy	9	B	14.0	Female	95.0	sneha_reddy@gmail.com	9.876543e+09
5	S006	Pooja Nair	8	C	NaN	Female	84.0	poojanair@gmail.com	9.876543e+09
6	S007	Aman Gupta	10	A	15.0	Male	78.0	amangupta@gmail.com	NaN
7	S008	Rahul Verma	10	A	16.0	Male	76.0	rahulv@gmail.com	9.876543e+09
8	S009	Priya Sharma	9	B	14.0	Female	NaN	priya.sharma@gmail.com	9.876543e+09
9	S010	Sneha Reddy	9	B	14.0	FEMALE	95.0	sneha_reddy@gmail.com	9.876543e+09
10	S011	Kiran Kumar	8	C	13.0	Male	82.0	kirank@gmail.com	9.876543e+09
11	S012	Arjun Mehta	10	A	15.0	Male	88.0	arjun.mehta@gmail.com	9.876543e+09

Next steps: Generate code with df New interactive sheet

```
df_cleaned = df.drop_duplicates()
```

```
print("----- Cleaned Data (duplicates removed) -----")
display(df_cleaned)
```

```
----- Cleaned Data (duplicates removed) -----
```

	Student_ID	Name	Class	Section	Age	Gender	Marks	Email	Phone
0	S001	Arjun Mehta	10	A	15.0	Male	88.0	arjun.mehta@gmail.com	9.876543e+09
1	S002	Priya Sharma	9	B	14.0	Female	92.0	priya.sharma@gmail.com	9.876543e+09
2	S003	arjun mehta	10	A	15.0	male	88.0	arjun.mehta@gmail.com	9.876543e+09
3	S004	Rahul Verma	10	A	16.0	Male	76.0	rahulv@gmail.com	9.876543e+09
4	S005	Sneha Reddy	9	B	14.0	Female	95.0	sneha_reddy@gmail.com	9.876543e+09
5	S006	Pooja Nair	8	C	NaN	Female	84.0	poojanair@gmail.com	9.876543e+09
6	S007	Aman Gupta	10	A	15.0	Male	78.0	amangupta@gmail.com	NaN
7	S008	Rahul Verma	10	A	16.0	Male	76.0	rahulv@gmail.com	9.876543e+09
8	S009	Priya Sharma	9	B	14.0	Female	NaN	priya.sharma@gmail.com	9.876543e+09
9	S010	Sneha Reddy	9	B	14.0	FEMALE	95.0	sneha_reddy@gmail.com	9.876543e+09
10	S011	Kiran Kumar	8	C	13.0	Male	82.0	kirank@gmail.com	9.876543e+09
11	S012	Arjun Mehta	10	A	15.0	Male	88.0	arjun.mehta@gmail.com	9.876543e+09

Next steps: Generate code with df_cleaned New interactive sheet

```
print("Rows before:", len(df))
print("Rows after :", len(df_cleaned))
```

```
Rows before: 12
Rows after : 12
```

```
df_cleaned.to_excel("school_students_cleaned.xlsx", index=False)
print("✅ Cleaned file saved as school_students_cleaned.xlsx")
```

```
✅ Cleaned file saved as school_students_cleaned.xlsx
```