


```
1 import pandas as pd
2
3 # Try reading the CSV file with the 'latin-1' encoding.
4 df = pd.read_csv("/Sample - Superstore.csv", encoding='latin-1')
5 print(df.head())
6
```



	Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer ID	\
0	1	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	
1	2	CA-2016-152156	11/8/2016	11/11/2016	Second Class	CG-12520	
2	3	CA-2016-138688	6/12/2016	6/16/2016	Second Class	DV-13045	
3	4	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	
4	5	US-2015-108966	10/11/2015	10/18/2015	Standard Class	SO-20335	

	Customer Name	Segment	Country	City	...	\
0	Claire Gute	Consumer	United States	Henderson	...	
1	Claire Gute	Consumer	United States	Henderson	...	
2	Darrin Van Huff	Corporate	United States	Los Angeles	...	
3	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	
4	Sean O'Donnell	Consumer	United States	Fort Lauderdale	...	

	Postal Code	Region	Product ID	Category	Sub-Category	\
0	42420	South	FUR-BO-10001798	Furniture	Bookcases	
1	42420	South	FUR-CH-10000454	Furniture	Chairs	
2	90036	West	OFF-LA-10000240	Office Supplies	Labels	
3	33311	South	FUR-TA-10000577	Furniture	Tables	
4	33311	South	OFF-ST-10000760	Office Supplies	Storage	

	Product Name	Sales	Quantity	\
0	Bush Somerset Collection Bookcase	261.9600	2	
1	Hon Deluxe Fabric Upholstered Stacking Chairs,...	731.9400	3	
2	Self-Adhesive Address Labels for Typewriters b...	14.6200	2	
3	Bretford CR4500 Series Slim Rectangular Table	957.5775	5	
4	Eldon Fold 'N Roll Cart System	22.3680	2	


	Discount	Profit
0	0.00	41.9136
1	0.00	219.5820
2	0.00	6.8714
3	0.45	-383.0310
4	0.20	2.5164

[5 rows x 21 columns]

▼ **Identify issues:**

Null Values:

```
1 print(df.isnull().sum())
2
```



Row ID	0
Order ID	0
Order Date	0
Ship Date	0
Ship Mode	0
Customer ID	0
Customer Name	0
Segment	0
Country	0
City	0
State	0
Postal Code	0
Region	0
Product ID	0
Category	0
Sub-Category	0
Product Name	0
Sales	0
Quantity	0
Discount	0
Profit	0
dtype:	int64


Duplicates:

```
1 print(df.duplicated().sum())
2
```

 0

Data Types:

```
1 print(df.dtypes)
2
```

 Row ID int64
Order ID object
Order Date object
Ship Date object
Ship Mode object
Customer ID object
Customer Name object
Segment object
Country object
City object
State object
Postal Code int64
Region object
Product ID object
Category object
Sub-Category object
Product Name object
Sales float64
Quantity int64
Discount float64
Profit float64
dtype: object

Remove Nulls:

```
1 df = df.dropna()
2
```

Drop Duplicates:

```
1 df = df.drop_duplicates()
2
```

Convert Data Types:

```
1 df['Order Date'] = pd.to_datetime(df['Order Date'])
2
```

Save the cleaned dataset:

```
1 df.to_csv("Cleaned_Superstore.csv", index=False)
2

1 import os
2 print(os.getcwd()) # This will show the directory where the file is saved
3
```

 /content

