


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

```
df = pd.read_csv('/content/Amazon Sales data.csv')
df
```



	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925	255.28	159.42	2533654.00	1582243.50	95
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804	205.70	117.11	576782.80	328376.44	24
2	Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779	651.21	524.96	1158502.59	933903.84	22
3	Sub-Saharan Africa	Sao Tome and Principe	Fruits	Online	C	6/20/2014	514321792	7/5/2014	8102	9.33	6.92	75591.66	56065.84	1
4	Sub-Saharan Africa	Rwanda	Office Supplies	Offline	L	2/1/2013	115456712	2/6/2013	5062	651.21	524.96	3296425.02	2657347.52	63
...
95	Sub-Saharan Africa	Mali	Clothes	Online	M	7/26/2011	512878119	9/3/2011	888	109.28	35.84	97040.64	31825.92	6
96	Asia	Malaysia	Fruits	Offline	L	11/11/2011	810711038	12/28/2011	6267	9.33	6.92	58471.11	43367.64	1

df.head()

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804
2	Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779

df.dtypes

Region	object
Country	object
Item Type	object
Sales Channel	object
Order Priority	object
Order Date	object
Order ID	int64
Ship Date	object
Units Sold	int64
Unit Price	float64
Unit Cost	float64
Total Revenue	float64
Total Cost	float64
Total Profit	float64
dtype:	object

```
len(df)

100
```

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100 entries, 0 to 99
Data columns (total 14 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Region              100 non-null   object
1   Country             100 non-null   object
2   Item Type           100 non-null   object
3   Sales Channel       100 non-null   object
4   Order Priority       100 non-null   object
5   Order Date          100 non-null   object
6   Order ID            100 non-null   int64
7   Ship Date           100 non-null   object
8   Units Sold          100 non-null   int64
9   Unit Price          100 non-null   float64
10  Unit Cost           100 non-null   float64
11  Total Revenue       100 non-null   float64
12  Total Cost          100 non-null   float64
13  Total Profit        100 non-null   float64
dtypes: float64(5), int64(2), object(7)
memory usage: 11.1+ KB
```

```
df.isnull().sum()
```

```
Region      0
Country     0
Item Type    0
Sales Channel 0
Order Priority 0
Order Date  0
Order ID    0
Ship Date   0
Units Sold  0
Unit Price  0
Unit Cost   0
Total Revenue 0
Total Cost   0
Total Profit 0
dtype: int64
```

```
df = df.drop_duplicates()
df.head()
```

	Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold
0	Australia and Oceania	Tuvalu	Baby Food	Offline	H	5/28/2010	669165933	6/27/2010	9925
1	Central America and the Caribbean	Grenada	Cereal	Online	C	8/22/2012	963881480	9/15/2012	2804
2	Europe	Russia	Office Supplies	Offline	L	5/2/2014	341417157	5/8/2014	1779

San

```
len(df)
```

```
100
```

```
df.describe()
```

	Order ID	Units Sold	Unit Price	Unit Cost	Total Revenue	Total Cost	
count	1.000000e+02	100.000000	100.000000	100.000000	1.000000e+02	1.000000e+02	1.000
mean	5.550204e+08	5128.710000	276.761300	191.048000	1.373488e+06	9.318057e+05	4.416
std	2.606153e+08	2794.484562	235.592241	188.208181	1.460029e+06	1.083938e+06	4.385
min	1.146066e+08	124.000000	9.330000	6.920000	4.870260e+03	3.612240e+03	1.258
25%	3.389225e+08	2836.250000	81.730000	35.840000	2.687212e+05	1.688680e+05	1.214
50%	5.577086e+08	5382.500000	179.880000	107.275000	7.523144e+05	3.635664e+05	2.907
75%	7.907551e+08	7369.000000	437.200000	263.330000	2.212045e+06	1.613870e+06	6.358