```
import pandas as pd
```

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_rem
```

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

Choose Files Advertising …d Sales.csv

Advertising Budget and Sales.csv(text/csv) - 4802 bytes, last modified: 12/25/2021 - 100% done

```
Saving Advertising Budget and Sales.csv to Advertising Budget and Sales (2).csv
```

```
df = pd.read_csv("Advertising Budget and Sales.csv")

df.head()
```

	Unnamed: 0	TV Ad Budget ($)	Radio Ad Budget ($)	Newspaper Ad Budget ($)	Sales ($)
0	1	230.1	37.8	69.2	22.1
1	2	44.5	39.3	45.1	10.4
2	3	17.2	45.9	69.3	9.3
3	4	151.5	41.3	58.5	18.5
4	5	180.8	10.8	58.4	12.9

Next steps: Generate code with df New interactive sheet

Data Preperation

```
df = df.rename(columns={"Sales ($)": "Sales(Units)"})

df.columns
```

```
Index(['Unnamed: 0', 'TV Ad Budget ($)', 'Radio Ad Budget ($)',
       'Newspaper Ad Budget ($)', 'Sales(Units)'],
      dtype='object')
```

```
if "Unnamed: 0" in df.columns:
    df = df.drop(columns=["Unnamed: 0"])

print("Missing Values:")
print(df.isnull().sum())

df = df.drop_duplicates()

df.info()
```

```
Missing Values:
TV Ad Budget ($)           0
Radio Ad Budget ($)        0
Newspaper Ad Budget ($)    0
Sales(Units)               0
dtype: int64
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 4 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   TV Ad Budget ($)        200 non-null    float64
 1   Radio Ad Budget ($)     200 non-null    float64
 2   Newspaper Ad Budget ($) 200 non-null    float64
 3   Sales(Units)            200 non-null    float64
dtypes: float64(4)
memory usage: 6.4 KB
```

Data Engineering

```
# Creating total advertising budget column

df["Total_Ad_Budget"] = (
    df["TV Ad Budget ($)"] +
    df["Radio Ad Budget ($)"] +
    df["Newspaper Ad Budget ($)"]
)

df = df.round(1)

df.head()
```

	TV Ad Budget ($)	Radio Ad Budget ($)	Newspaper Ad Budget ($)	Sales(Units)	Total_Ad_Budget
0	230.1	37.8	69.2	22.1	337.1
1	44.5	39.3	45.1	10.4	128.9
2	17.2	45.9	69.3	9.3	132.4
3	151.5	41.3	58.5	18.5	251.3
4	180.8	10.8	58.4	12.9	250.0

Next steps: Generate code with df | New interactive sheet

this column will allow me to see what each company is spending on all three ad budgets. this is probably the most important added column because of the amount of things I need to know what the total for all of these ad budgets.

```
# creating a channel share columns for each ad

df["TV_Share"] = df["TV Ad Budget ($)"] / df["Total_Ad_Budget"]
df["Radio_Share"] = df["Radio Ad Budget ($)"] / df["Total_Ad_Budget"]
df["Newspaper_Share"] = df["Newspaper Ad Budget ($)"] / df["Total_Ad_Budget"]

# round share columns and make them percentages

share_cols = ["TV_Share", "Radio_Share", "Newspaper_Share"]

df[share_cols] = (df[share_cols] * 100).round(2)
df[share_cols] = df[share_cols].astype(str) + "%"

df.head()
```

	TV Ad Budget ($)	Radio Ad Budget ($)	Newspaper Ad Budget ($)	Sales(Units)	Total_Ad_Budget	TV_Share	Radio_Share	Newspaper_Share
0	230.1	37.8	69.2	22.1	337.1	68.26%	11.21%	20.53%
1	44.5	39.3	45.1	10.4	128.9	34.52%	30.49%	34.99%
2	17.2	45.9	69.3	9.3	132.4	12.99%	34.67%	52.34%
3	151.5	41.3	58.5	18.5	251.3	60.29%	16.43%	23.28%

Next steps: Generate code with df | New interactive sheet

These three columns will help me see which ad budget has the most money invested in them, which will help me with answering my specific question which channel has the most influence on volume?

```
# making a sales effciency column

df["Sales_per_Dollar"] = df["Sales(Units)"] / df["Total_Ad_Budget"]

df.head()
```

	TV Ad Budget ($)	Radio Ad Budget ($)	Newspaper Ad Budget ($)	Sales(Units)	Total_Ad_Budget	TV_Share	Radio_Share	Newspaper_Share	Sales_per_Dollar
0	230.1	37.8	69.2	22.1	337.1	68.26%	11.21%	20.53%	0.065559
1	44.5	39.3	45.1	10.4	128.9	34.52%	30.49%	34.99%	0.080683
2	17.2	45.9	69.3	9.3	132.4	12.99%	34.67%	52.34%	0.070242
3	151.5	41.3	58.5	18.5	251.3	60.29%	16.43%	23.28%	0.073617

Next steps: Generate code with df New interactive sheet

This will help me with answer one of my specific questions, what is the marginal increase in units sold for every $10,000 spent?

```
new_order = [
    "TV Ad Budget ($)",
    "Radio Ad Budget ($)",
    "Newspaper Ad Budget ($)",
    "TV_Share",
    "Radio_Share",
    "Newspaper_Share",
    "Sales_per_Dollar",
    "Total_Ad_Budget",
    "Sales(Units)"
]

df = df[new_order]

df.head()
```

	TV Ad Budget ($)	Radio Ad Budget ($)	Newspaper Ad Budget ($)	TV_Share	Radio_Share	Newspaper_Share	Sales_per_Dollar	Total_Ad_Budget	Sales(Units)
0	230.1	37.8	69.2	68.26%	11.21%	20.53%	0.065559	337.1	22.1
1	44.5	39.3	45.1	34.52%	30.49%	34.99%	0.080683	128.9	10.4
2	17.2	45.9	69.3	12.99%	34.67%	52.34%	0.070242	132.4	9.3
3	151.5	41.3	58.5	60.29%	16.43%	23.28%	0.073617	251.3	18.5

Next steps: Generate code with df New interactive sheet

I had to reorder the columns because it makes the dataset easier to read and it looks a lot more organized when it is order like this.

```
df.to_csv("v_clean_advertising_dataset.csv", index=False)
```

```
from google.colab import files
files.download("v_clean_advertising_dataset.csv")
```

```
Start coding or generate with AI.
```