```
pip install pandas
```

```
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-p
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-pa
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packag
```

```
from google.colab import files

uploaded = files.upload()
```

Choose Files   svamp_dataset.csv
svamp_dataset.csv(text/csv) - 258887 bytes, last modified: 10/31/2025 - 100% done
Saving svamp_dataset.csv to svamp_dataset.csv

```
import pandas as pd
df = pd.read_csv("svamp_dataset.csv")
df.head()
```

	ID	Body	Question	Equation	Answer	Type	question_concat
0	chal-777	There are 87 oranges and 290 bananas in Philip...	How big is each group of bananas?	(290.0 / 2.0)	145	Common-Division	There are 87 oranges and 290 bananas in Philip...
1	chal-508	Marco and his dad went strawberry picking. Mar...	How much did Marco's strawberries weigh?	(30.0 - 11.0)	19	Subtraction	Marco and his dad went strawberry picking. Mar...
2	chal-896	Edward spent $ 6 to buy 2 books each book cost...	How much did each book cost?	(6.0 / 2.0)	3	Common-Division	Edward spent $ 6 to buy 2 books each book cost...
3	chal-923	Frank was reading through his favorite book. T...	How many pages are in each chapter?	(594.0 / 3.0)	198	Common-Division	Frank was reading through his favorite book. T...
4	chal-34	There were 78 dollars in Olivia's wallet. She ...	How much money does she have left?	(78.0 - 15.0)	63	Subtraction	There were 78 dollars in Olivia's wallet. She ...

Next steps: Generate code with df   New interactive sheet

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 700 entries, 0 to 699
Data columns (total 7 columns):
 #   Column           Non-Null Count   Dtype
---  ------           --------------   -----
 0   ID               700 non-null     object
 1   Body             700 non-null     object
 2   Question         700 non-null     object
 3   Equation         700 non-null     object
 4   Answer           700 non-null     int64
 5   Type             700 non-null     object
 6   question_concat  700 non-null     object
```

```
dtypes: int64(1), object(6)
memory usage: 38.4+ KB
```

```
df.describe()
```

	Answer
count	7.000000e+02
mean	7.400224e+03
std	8.666324e+04
min	1.000000e+00
25%	4.750000e+00
50%	1.500000e+01
75%	6.500000e+01
max	1.538832e+06

```
#missing value
df.isnull().sum()
```

	0
ID	0
Body	0
Question	0
Equation	0
Answer	0
Type	0
question_concat	0

dtype: int64