```
from google.colab import files

uploaded = files.upload()
```

Choose Files gsm8k_train.csv
gsm8k_train.csv(text/csv) - 3944420 bytes, last modified: 10/30/2025 - 100% done
Saving gsm8k_train.csv to gsm8k_train.csv

```
import pandas as pd
df=pd.read_csv("gsm8k_train.csv")
df.head()
```

	question	answer
0	Natalia sold clips to 48 of her friends in Apr...	Natalia sold 48/2 = <<48/2=24>>24 clips in May...
1	Weng earns $12 an hour for babysitting. Yester...	Weng earns 12/60 = $<<12/60=0.2>>0.2 per minut...
2	Betty is saving money for a new wallet which c...	In the beginning, Betty has only 100 / 2 = $<<...
3	Julie is reading a 120-page book. Yesterday, s...	Maila read 12 x 2 = <<12*2=24>>24 pages today....
4	James writes a 3-page letter to 2 different fr...	He writes each friend 3*2=<<3*2=6>>6 pages a w...

Next steps: (Generate code with df) (New interactive sheet)

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7473 entries, 0 to 7472
Data columns (total 2 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   question  7473 non-null   object
 1   answer    7473 non-null   object
dtypes: object(2)
memory usage: 116.9+ KB
```

```
df.describe()
```

	question	answer
count	7473	7473
unique	7473	7473
top	At 30, Anika is 4/3 the age of Maddie. What wo...	If Anika is 30 now, in 15 years, she'll be 30+...
freq	1	1

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

	0
question	0
answer	0

dtype: int64

```python
df.duplicated().sum()
```

```
np.int64(0)
```