

```
from google.colab import files
```

```
uploaded = files.upload()
```

[Choose Files](#) MGSM\_train.csv

**MGSM\_train.csv**(text/csv) - 2668 bytes, last modified: 10/31/2025 - 100% done

Saving MGSM\_train.csv to MGSM\_train (1).csv

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

	question	answer	answer_number	equation_solution
0	Question: Roger has 5 tennis balls. He buys 2 ...	Step-by-Step Answer: Roger started with 5 ball...	11	5 + 6 = 11.
1	Question: There were nine computers in the ser...	Step-by-Step Answer: There are 4 days from mon...	29	4 * 5 = 20. 9 + 20 = 29.
2	Question: Leah had 32 chocolates and her siste...	Step-by-Step Answer: Leah had 32 chocolates an...	39	32 + 42 = 74. 74 - 35 = 39.
3	Question: Shawn has five toys. For Christmas, ...	Step-by-Step Answer: He has 5 toys. He got 2 f...	9	5 + 2 = 7. 7 + 2 = 9.
4	Question: Michael had 58 golf balls. On tuesda...	Step-by-Step Answer: Michael started with 58 g...	33	58 - 23 = 35. 35 - 2 = 33.

Next steps:

[Generate code with df](#)

[New interactive sheet](#)

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8 entries, 0 to 7
Data columns (total 4 columns):
#   Column          Non-Null Count  Dtype
---  -
0   question        8 non-null     object
1   answer          8 non-null     object
```

```

2  answer_number      8 non-null    int64
3  equation_solution  8 non-null    object
dtypes: int64(1), object(3)
memory usage: 388.0+ bytes

```

```

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

```



```

          0
question  0
answer    0
answer_number  0
equation_solution  0

```

```
dtype: int64
```

```
df.describe()
```

	answer_number	
count	8.000000	
mean	17.750000	
std	13.551489	
min	5.000000	
25%	8.000000	
50%	10.000000	
75%	30.000000	
max	39.000000	

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

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

