Notebook for Code demonstration

Author: Jayden Lee

In [1]:
```
import csv
import pandas as pd
import numpy as np
import os
import sys
from main import main
```

In [2]:
```
# Run main.py in Jupyter IDE
fp1 = 'data/sample_input_1.csv'
fp2 = 'data/sample_input_2.csv'

main(fp1, fp2)
```

```
Raw CSV file contents parsed successfully!
Raw CSV file contents parsed successfully!
Columns with at least one missing values are: ['len', 'val']
Columns with at least one missing values are: ['len', 'spd']
Fill-in success!
Fill-in success!
Answer to question 3a: Column 'rd' in sample datasets contain non-tabular, or
non-normalized, data
Normalization complete!
Normalization complete!
Concatenation success!
Datasets parsed, transformed, and wrote out successfully!
```

In [3]:
```
# Output demonstration

output_df = pd.read_csv('output/processed_data.csv')
display(output_df.head(10))
print(f'\nOutput dataframe has dimension of {output_df.shape[0]} rows and {outp
```

	id	spd	len	dec	rd	dt	val	rd_i	rd_s	rd_l	rd_v
0	0	195.0	6.8	kdmmoXxkfT	i=0;l=6.8;v=F;s=195	9/29/2025	F	0	195	6.8	F
1	1	123.0	8.7	iWatxPHTPT	l=8.7;v=T;s=123;i=1	11/23/2025	T	1	123	8.7	T
2	2	222.0	8.4	pordUQ3vBa	v=T;i=2;l=8.4;s=222	11/19/2025	T	2	222	8.4	T
3	3	82.0	9.5	16LkodUACt	v=F;s=82;i=3;l=9.5	8/25/2025	F	3	82	9.5	F
4	4	24.0	7.2	SIsDeqgHgz	s=24;i=4;v=F;l=7.2	8/23/2025	F	4	24	7.2	F
5	5	110.0	10.7	MLgwDMCqXa	l=10.7;i=5;v=T;s=110	12/18/2025	T	5	110	10.7	T
6	6	20.0	6.4	FAlVl6RmrU	i=6;v=T;l=6.4;s=20	11/4/2025	T	6	20	6.4	T
7	7	236.0	6.9	VmpRQc8zgm	s=236;v=T;l=6.9;i=7	10/29/2025	T	7	236	6.9	T
8	8	41.0	9.7	RymAKXoTI5	v=T;i=8;l=9.7;s=41	8/2/2025	T	8	41	9.7	T
9	9	34.0	10.0	tLd7twElfS	s=34;v=F;l=10.0;i=9	9/3/2025	F	9	34	10.0	F

```
Output dataframe has dimension of 1000 rows and 11 columns
```