Physicochemical Properties of Protein Tertiary Structure

This data set was pulled from UCI. This dataset contains the following information for each instance:

1. RMSD - Size of the residue.
2. F1 - Total surface area.
3. F2 - Non polar exposed area.
4. F3 - Fractional area of exposed non polar residue.
5. F4 - Fractional area of exposed non polar part of residue.
6. F5 - Molecular mass weighted exposed area.
7. F6 - Average deviation from standard exposed area of residue.
8. F7 - Euclidian distance.
9. F8 - Secondary structure penalty.
10. F9 - Spacial Distribution constraints (N,K Value).

This data is good for regression. You can implement in several ways, but can have students experiment with attributes F1-F9 to predict RMSD.

You can use the following code to load the data into google colaboratory:

```
import pandas as pd
url = "https://raw.githubusercontent.com/the-codingschool/TRAIN-datasets/main/protein/protein_tertiary.csv"
df1 = pd.read_csv(url)
```