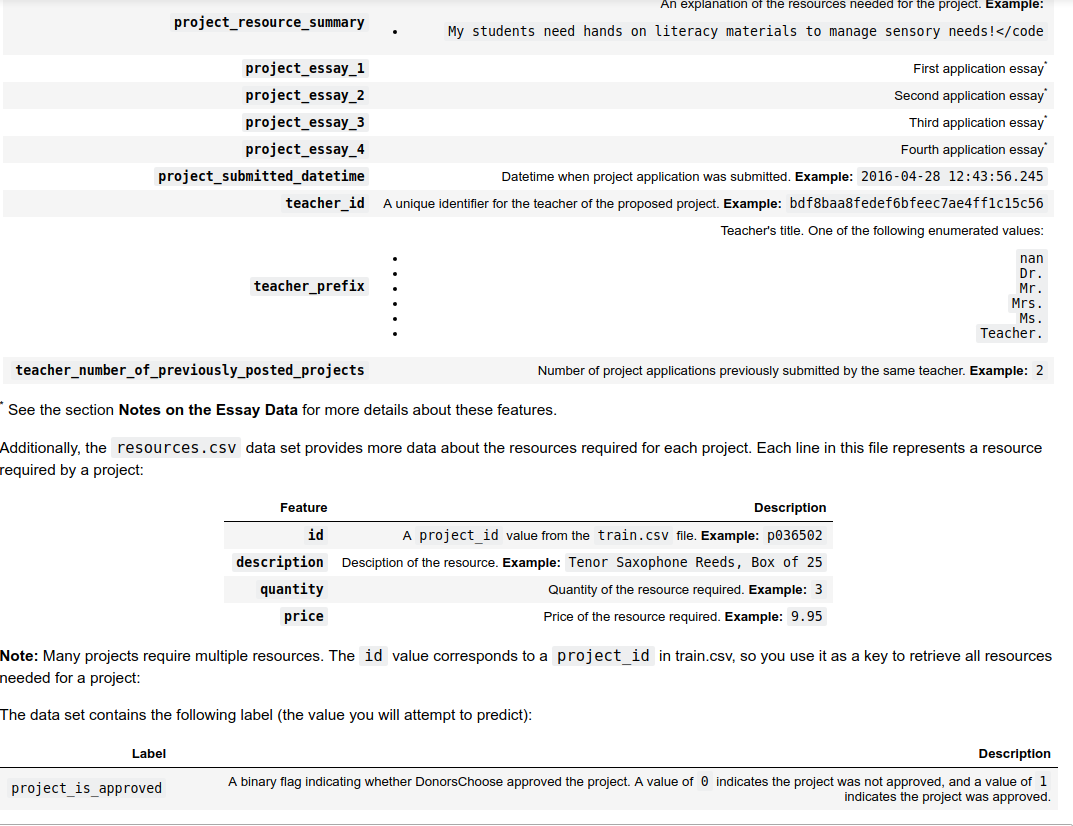
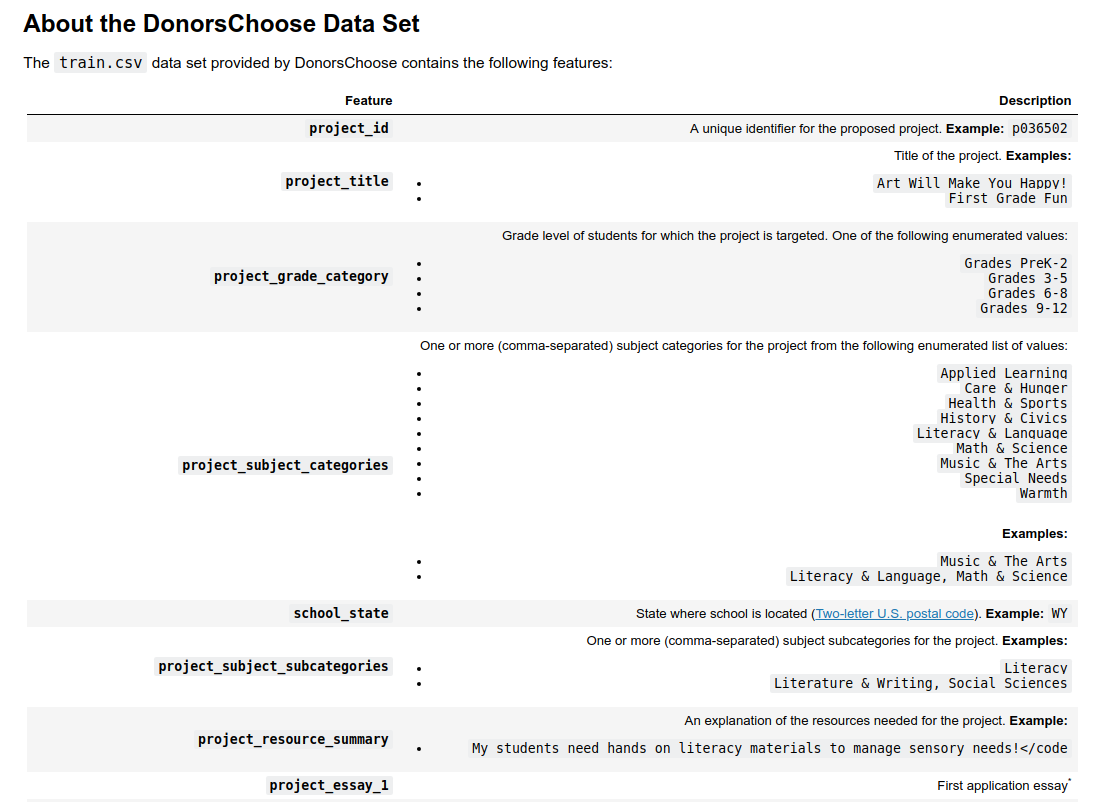
# **DONORS CHOOSE : NAIVE BAYES**

**EDA : DONORS CHOOSE**

This is the dataset and it’s columns

**DATASET DESCRIPTION**

**STEPS FOR EDA**

1 ) Calculate the shape / No. of data points you have.

2 ) **Data Analysis** : We calculate the total no of projects and the percentage of approved

And not approved projects.

3 ) **Univariate Analysis : School State .** We groupby School State column and calculate the mean of projects\_approved label for each state which is basically percentage of projects approved from each state since projects\_approved has only 0’s and 1’s[[1]](#footnote-0).

So basically we are calculating the percentage of projects approved and also total number of submissions from a given state. Plotting the bar-plots as well.

Ex : CA 86.78 % (Projects approved) 12980 (submissions)

4 ) **Univariate Analysis : Teacher prefix**  Same as above. We are calculating the percentage of projects approved and also total number of submissions from a given Teacher Prefix. Plotting the bar-plots as well.

EX :- Mrs. 84.48 % (Projects approved) 18739 (submissions)

5 ) **Univariate Analysis : project\_grade category**  Same as above but for project\_grade category. Ex : Grades 3- 5. 81.22 % (Projects approved) 19421 (submissions)

6 )

1. If the dataset has only 1’s and 0’s then it’s mean is the percentage of 1’s [↑](#footnote-ref-0)