1 Predicting Donor's Choose

1. What is the primary metric you care about in this task? Be sure to clearly state the question about why this is the case.
2. Is the column `teacher_number_of_previously_posted_projects` a good predictor for the approval of the project? Use both a `KNearestNeighbors` and `LogisticRegression` model. Which model performs better? Can you select the best parameters for each? The best penalty for `LogisticRegression`? Form a pipeline that includes a scaling transformation? Compare this to a `DummyClassifier`?
3. What are the top 8 states in terms of raw number approved? The lowest? Show me with a nice barplot.
4. Are these states different from the number of proportion of applications approved by state? Show me.
5. Does your model improve with the inclusion of the `teacher_prefix` column in a `LogisticRegression` model?
6. What is your best parameter for the training set with these inputs?
7. Construct a feature that is simply `STEM`, which is 1 if a scientific discipline is a part of the `subject_subcategory` column, or 0 if not. Did your model improve?
8. What if you include the `project_grade_category` column? Is your model improved?
9. What if your client only cares about what's happening in New York. Is there a difference in the performance of a `LogisticRegression` model? `KNearestNeighbors`?