

# **DonorsChoose.org**

# **Project Application Screening**



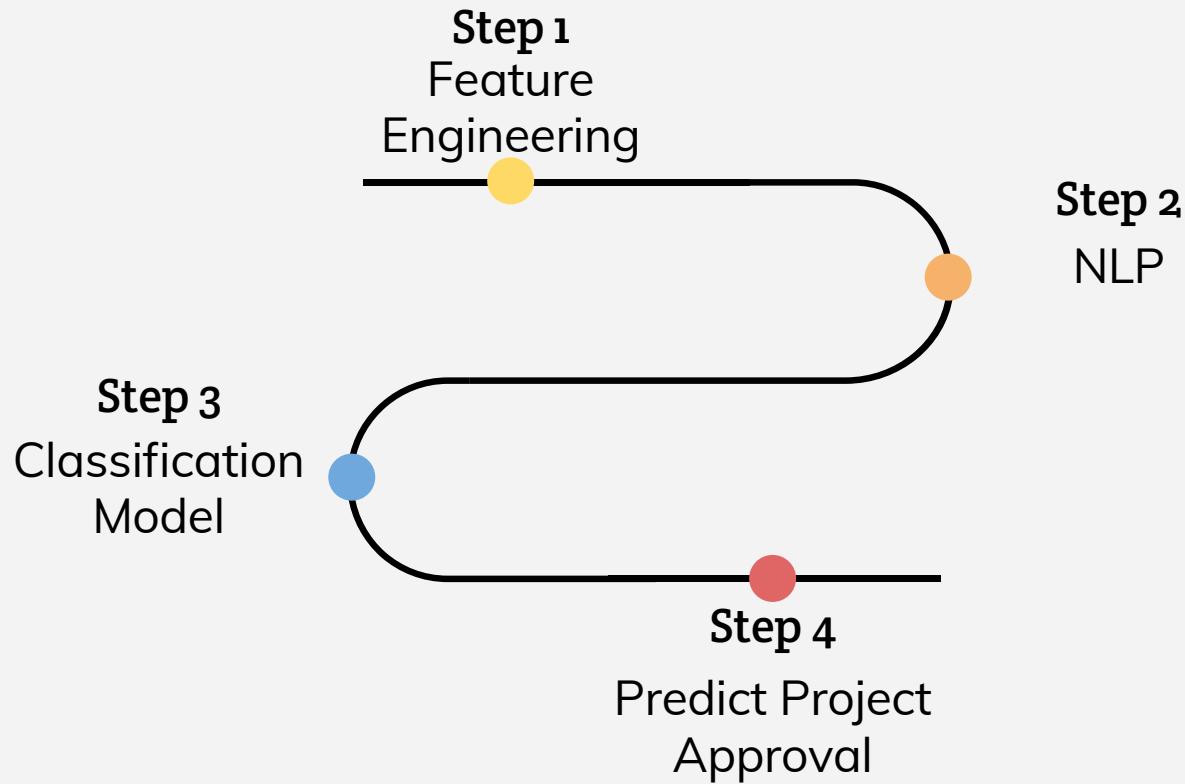
Carissa Sosa | Metis SF Fall 2019



## Overview

- DonorsChoose.org is public charity helping teachers fund classroom projects
  - Founded in 2000 - overall raised over \$900mm
  - Employ a lean staff of 135 and supplement with volunteers
  - All projects are hand-screened before they appear on website
-

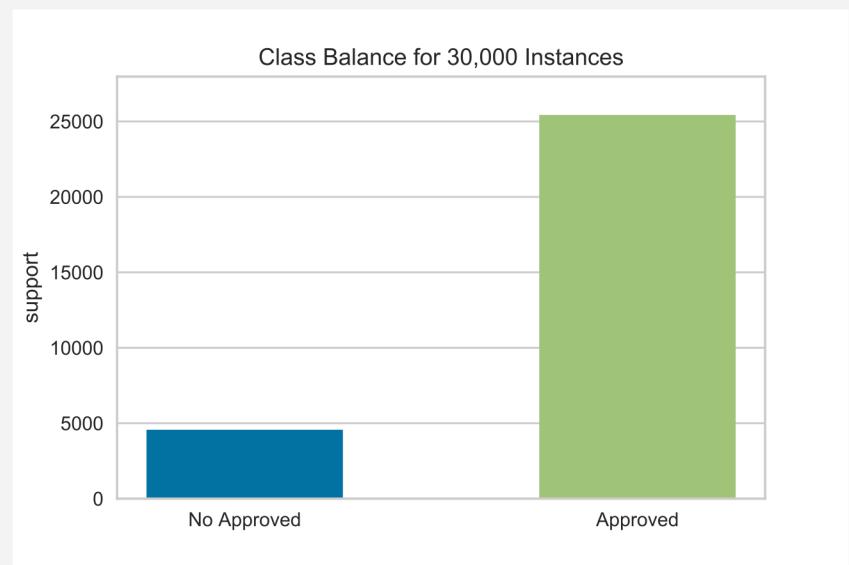
# Objective & Goals



# Methodology

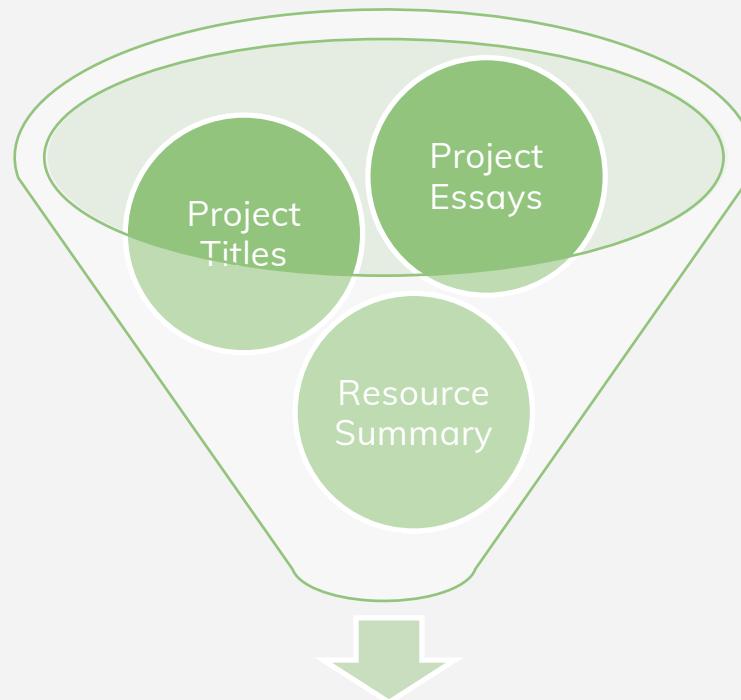
## Data

- DonorsChoose.org Dataset - 30,000 Applications
- Combination of text, categorical and continuous features
- Imbalanced Dataset

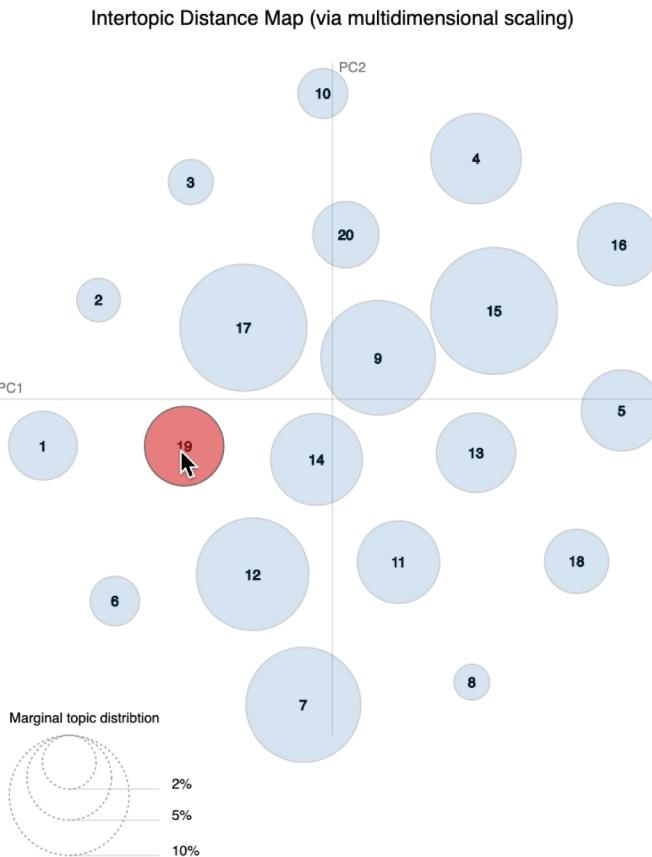


# Methodology

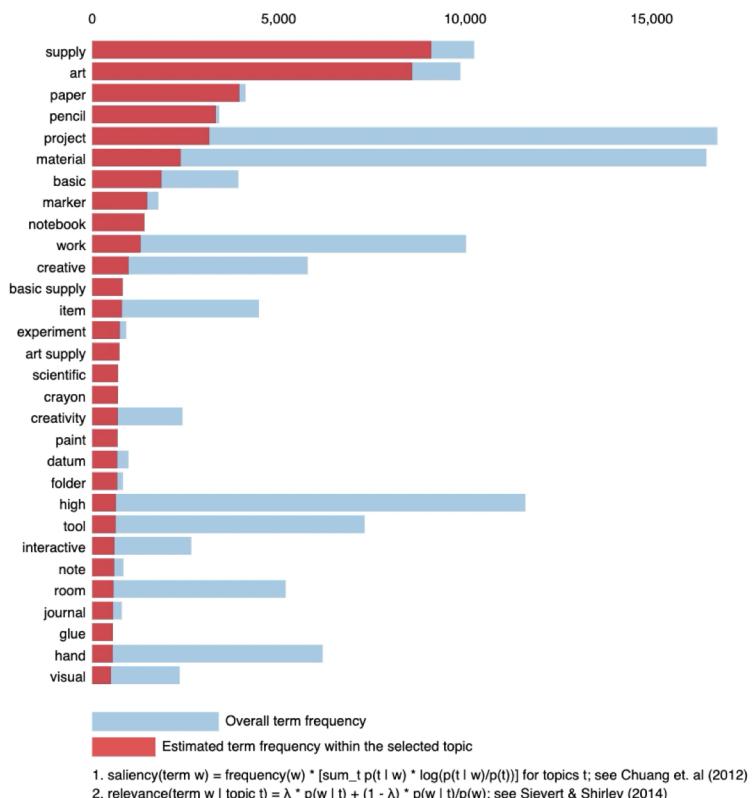
## Step 1 & 2



**Topic Modeling - LDA**



Top-30 Most Relevant Terms for Topic 19 (4.4% of tokens)



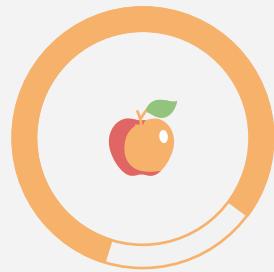
# Methodology

## Step 3



### Classification

RandomForest



### Features

Total Project Price  
Previous Submission  
Dominant Topics

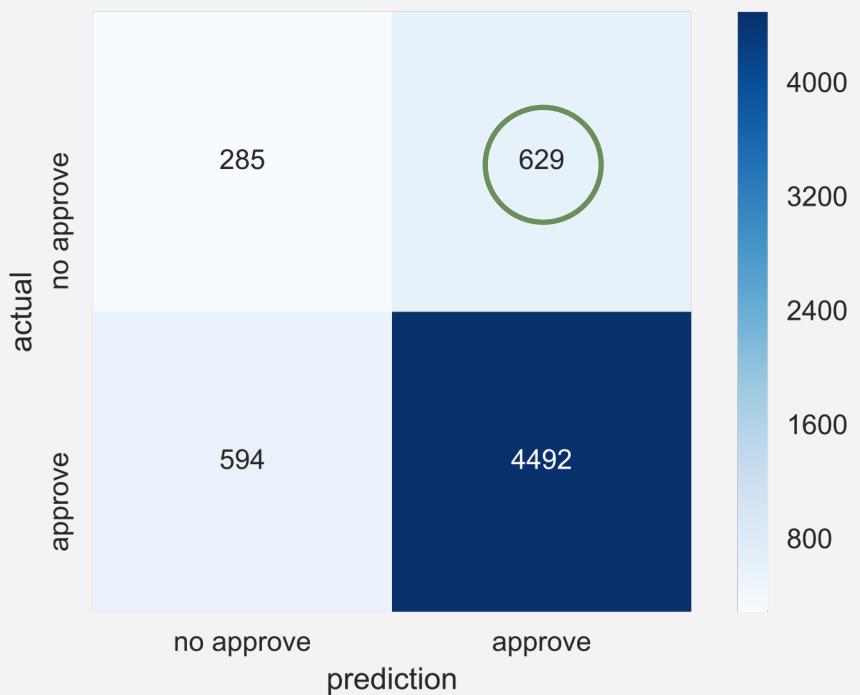


### Metrics

Recall  
AUC Score

# Results

## Step 4



0.47

Threshold

0.8732

Recall

0.68

AUC Score



## Recommendation

- Split project approval between model and human (80/20)
  - False Positives do additional hand screening
  - Free up resources to do more outreach to schools and partners
  - Continue to train model on new labeled data
-

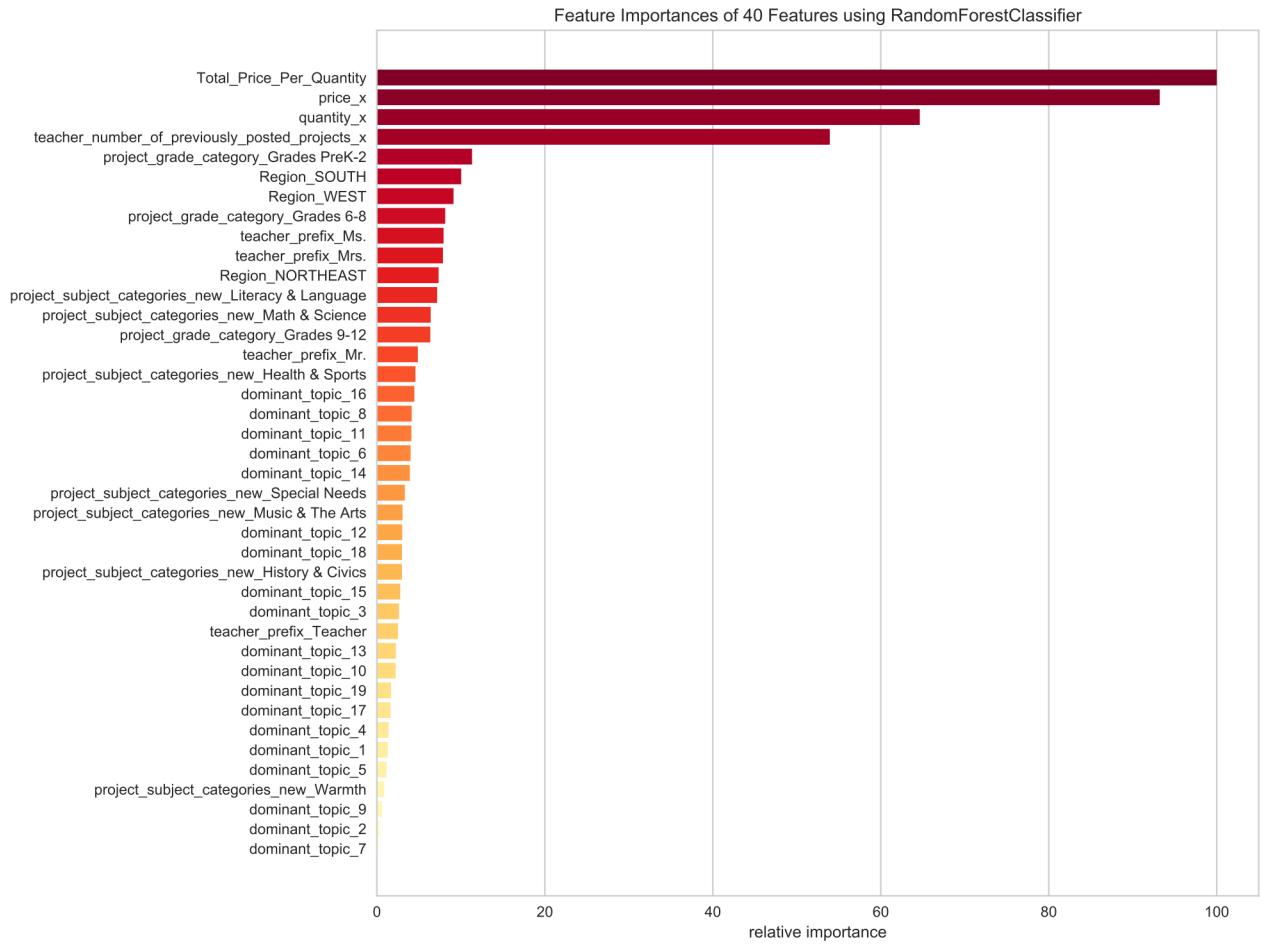
# Thanks!

Any Questions?

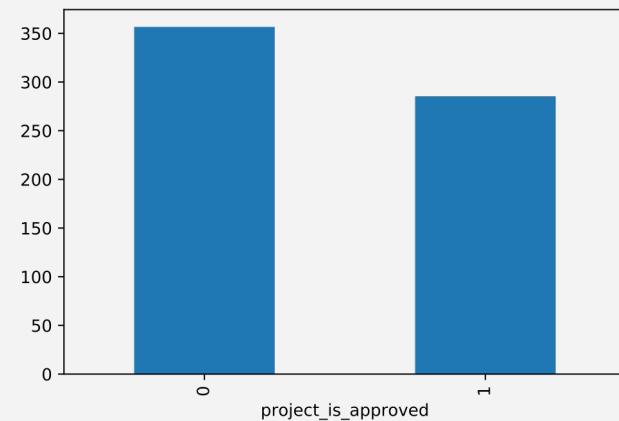


# Appendix

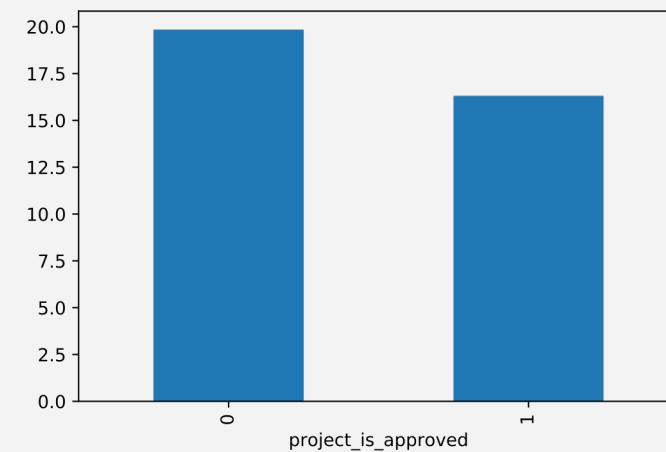




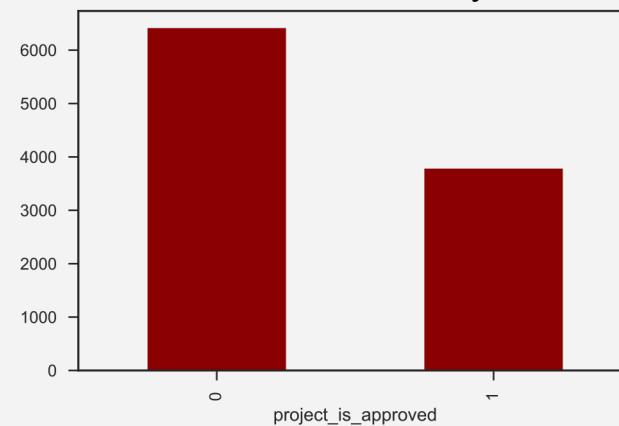
Price Mean



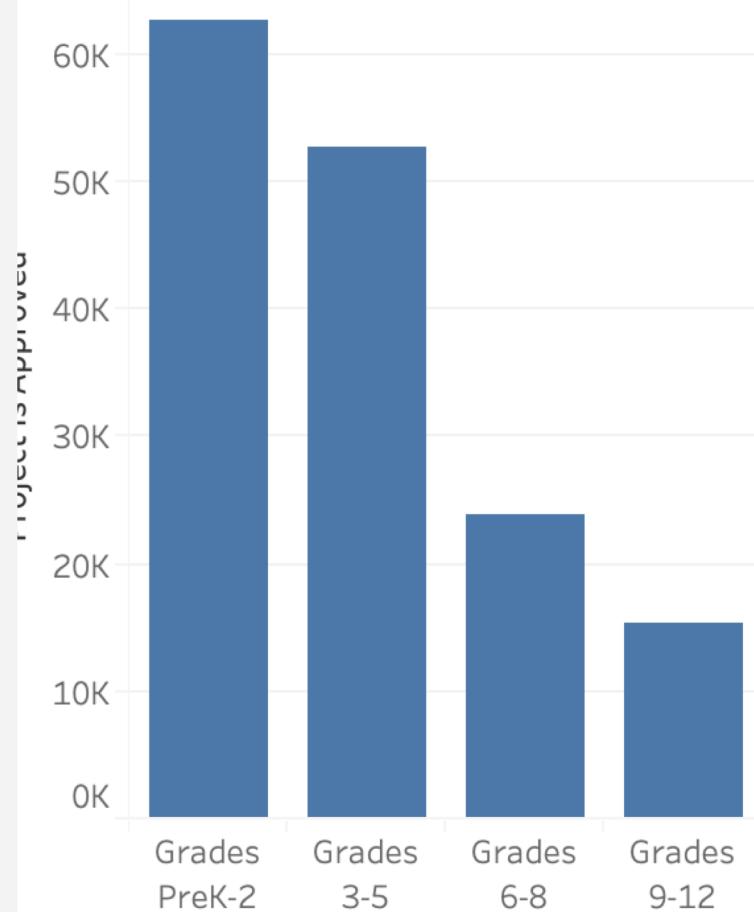
Quantity Mean



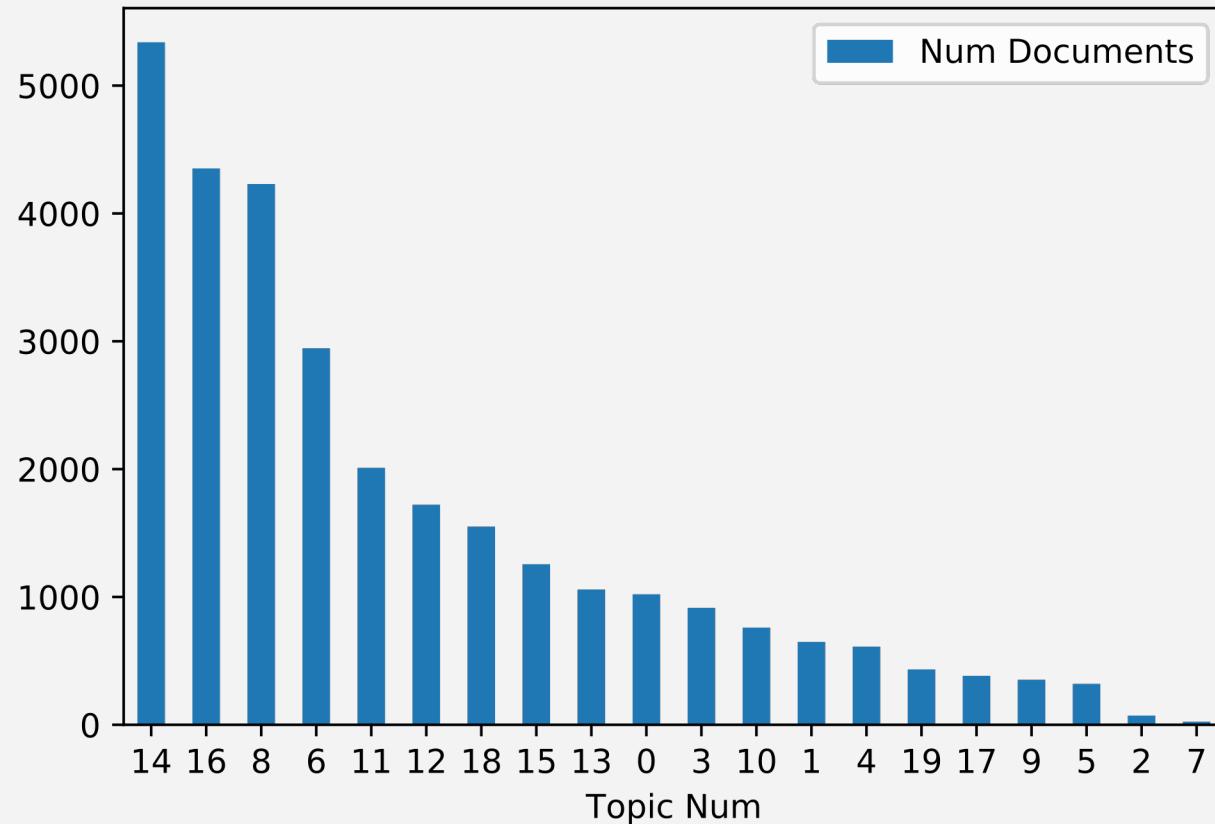
Total Price \* Quantity Mean

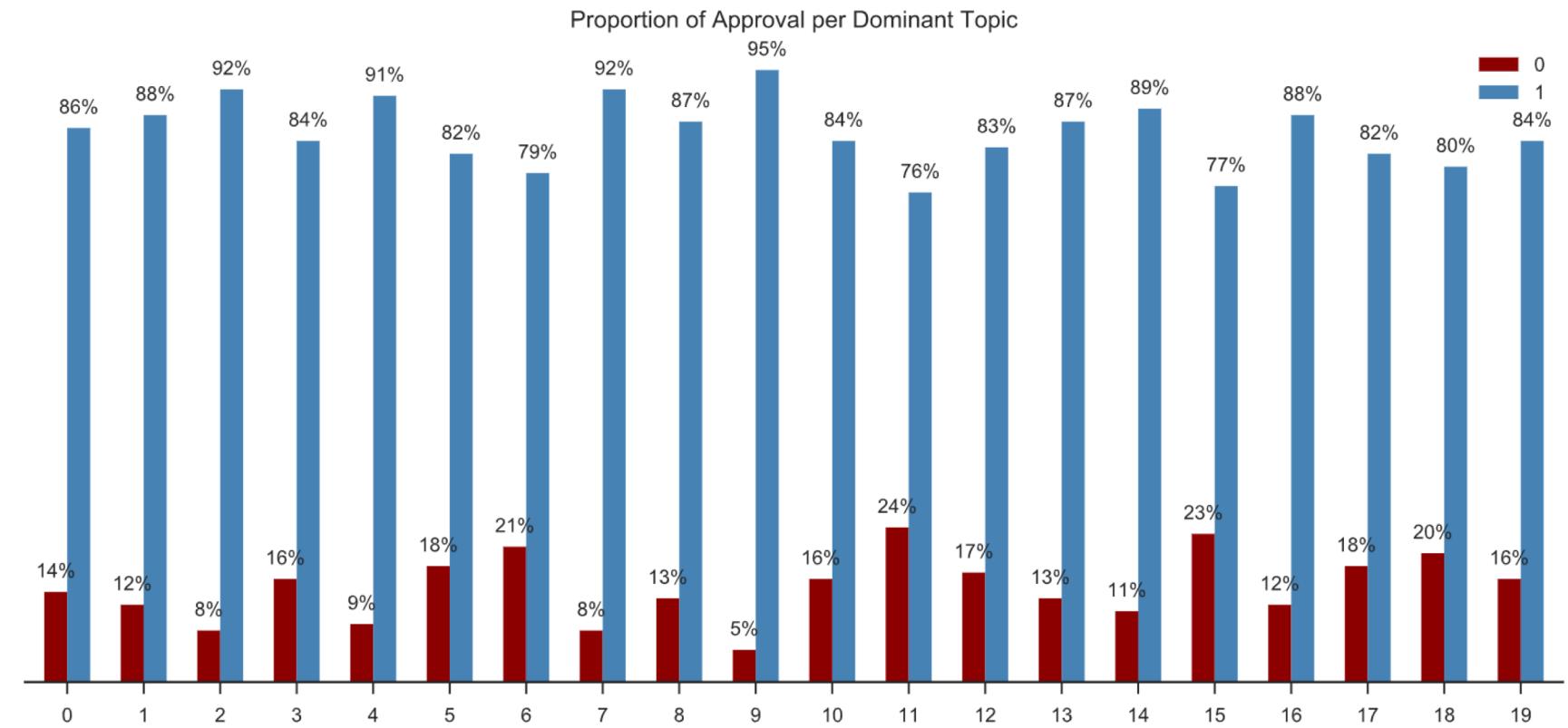


**Project Grade Category**

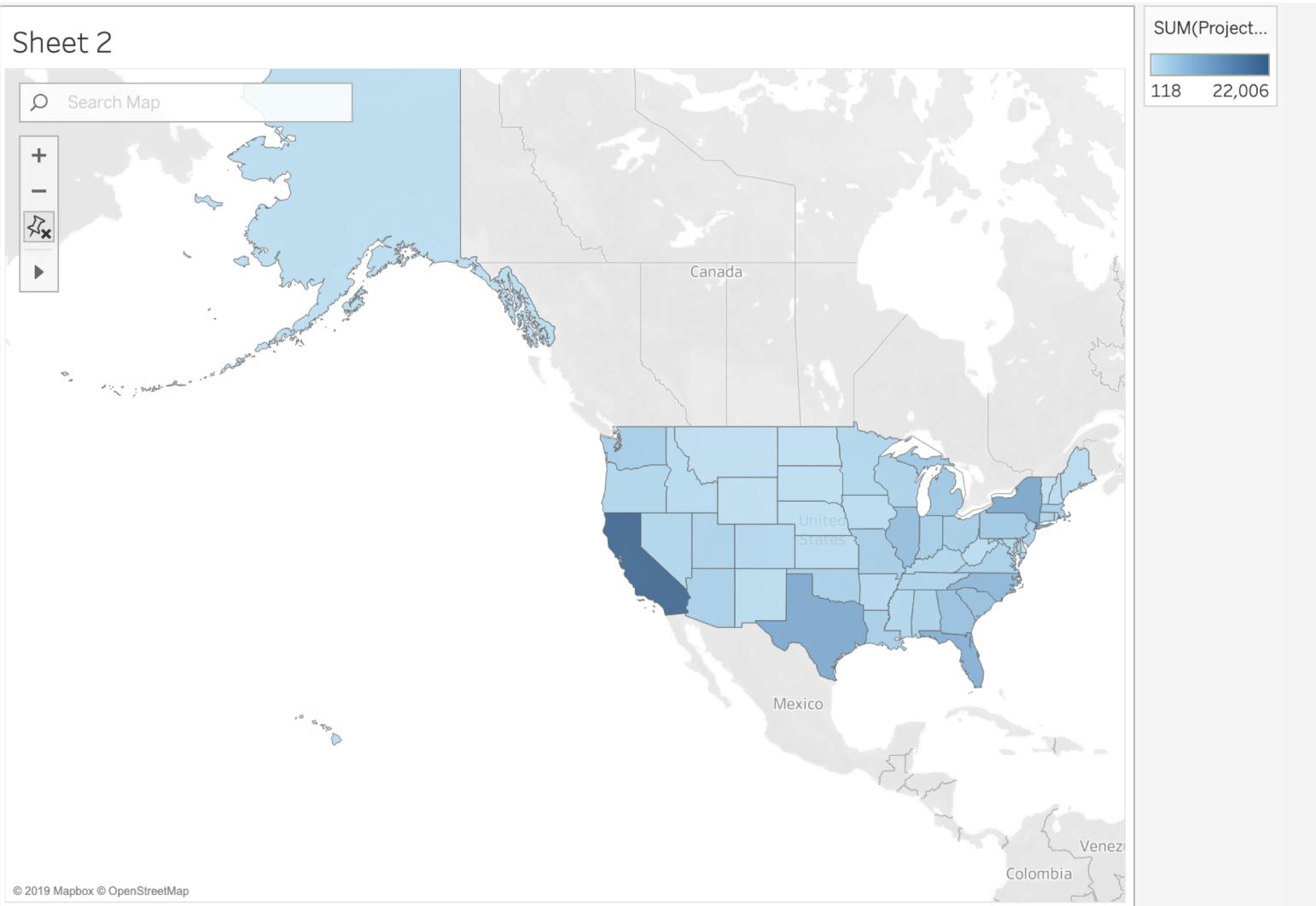


Count of Dominant Topics in Dataset





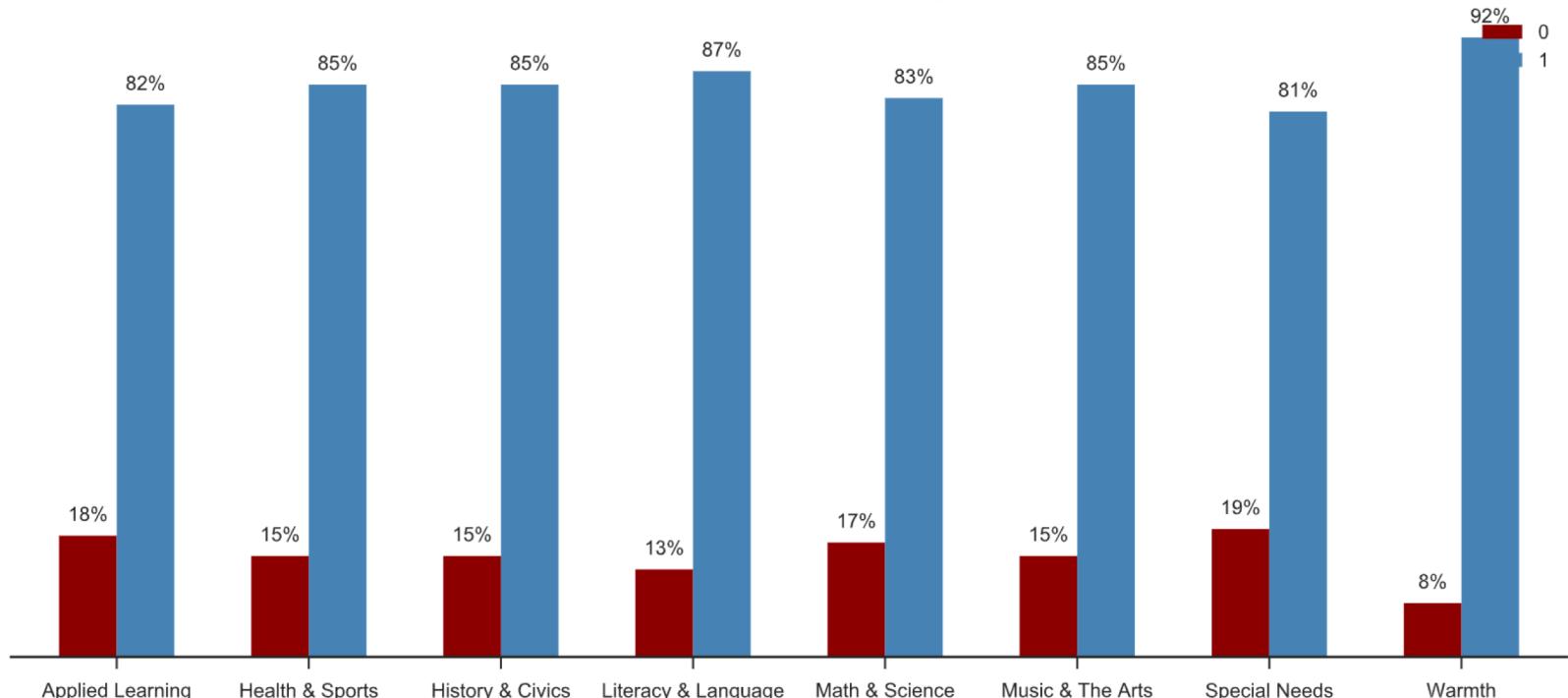
States  
w/most  
project  
approvals

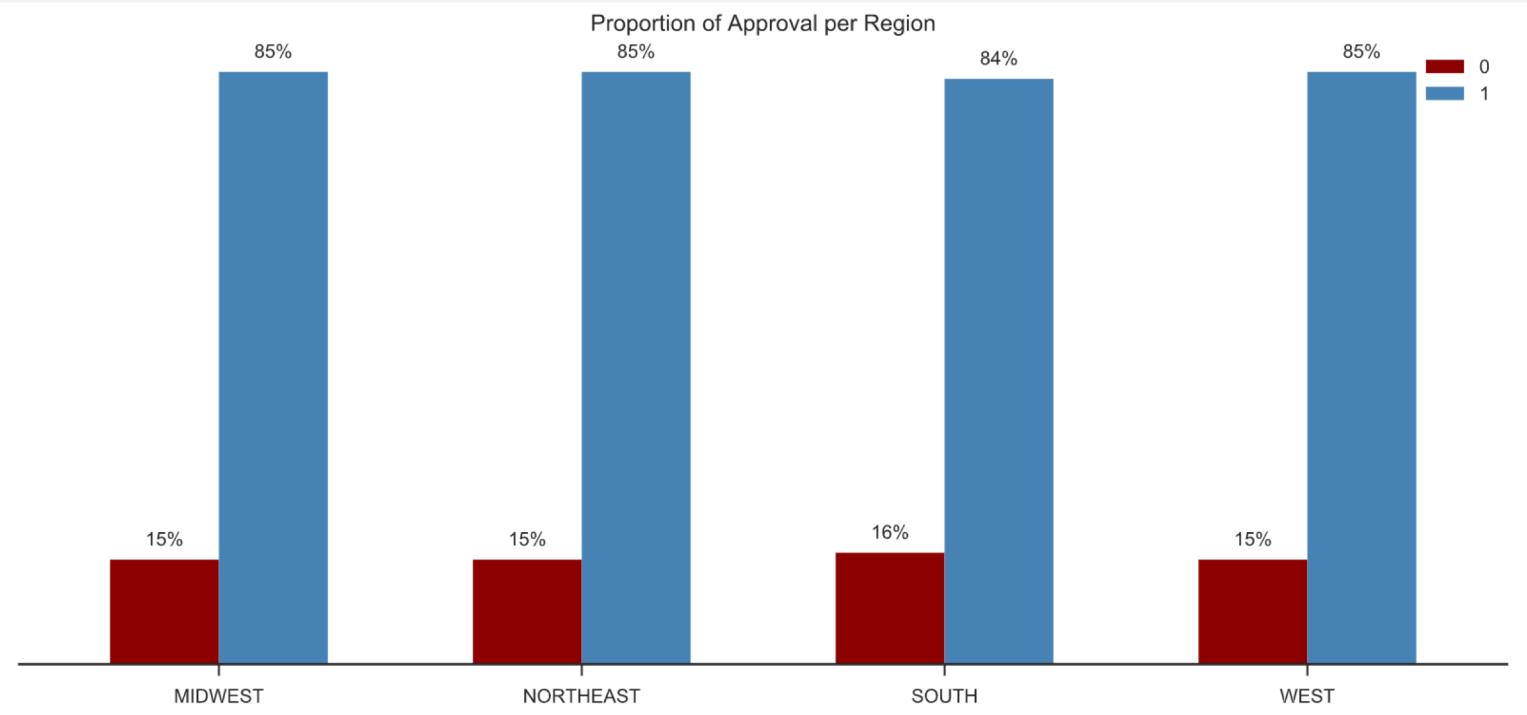


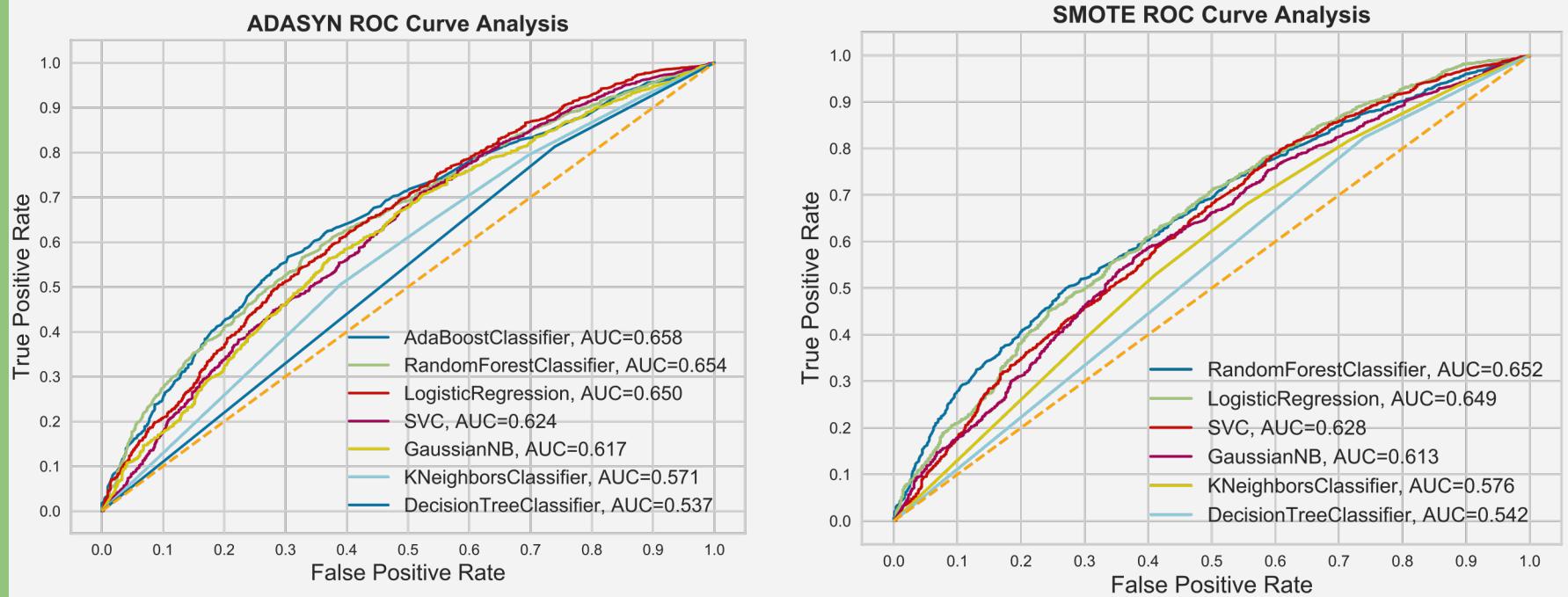
### Mean Word Count

Approved	289
Not Approved	277

Proportion of Approval per category







## Precision and Recall Curves

