Business Problem

Donors choosing serves hundreds of thousands of project proposals each year for classroom projects in need of funding. Right now, a teacher's success in getting their project funded is strongly dependent on their persuasive skills in addition to the merit of the project.

Next year, Donors Choose wants to receive close to 500,000 project proposals. As a result, there are three main problems they need to address:

- How to write a short review process and resources to screen 500,000 projects so that they can be evaluated quickly and as effectively as possible.
- How to ensure the success of project ideas among very different volunteers to improve the experience for teachers.
- How to focus volunteer time on the applications that need the most assistance.

The goal of the competition is to predict whether or not a Donors Choose proposal submitted by a teacher will be approved, taking into account the teacher's project proposal. Kaggle is hosting this competition for the Donors Choose organization so that very little information is needed by volunteers when they to read and review teacher's project proposals.

Dataset Information

The TEAMS.CSV data set contains the following fields:

Feature	Description
project_id	A unique identifier for the proposed project. **Example:** `p036502`
project_title	Title of the project. **Examples:** • Art Supplies for Frida • Kyle's Reading Room
project_grade_category	Grade level of students for which the project is targeted. One of the following enumerated values: • Grades PreK-2 • Grades 3-5 • Grades 6-8 • Grades 9-12
project_subject_categories	One or two (comma-separated) subject categories for the project from the following enumerated list of values: • Applied Sciences • Care & Hygiene • Health & Life Skills • History & Civics • Literacy & Language • Math & Science • Music & The Arts • Special Needs • Visiting Students **Examples:** • `Music & The Arts` • `Literacy & Language, Math & Science`
school_state	State where school is located (two-letter USPS abbreviation). **Example:** `NJ`
project_subject_subcategories	One or two (comma-separated) subject subcategories for the project. **Examples:** • `Literature & Writing, Social Sciences`
project_resource_summary	An explanation of the resources needed for the project. **Example:** • `My students need hands on literacy materials to manipulate sounds`
project_essay_1	"Introduce us to your classroom"
project_essay_2	"Tell your story"
project_essay_3	"Title I schools only"
project_essay_4	"Title I schools only"
project_submitted_datetime	Datetime when project application was submitted. **Example:** `2016-04-28 12:43:56.245`
teacher_id	A unique identifier for the teacher of the proposed project. **Example:** `bdf8baa8fedef6bfeec7ae4ff1c15ca9`
teacher_prefix	Teacher's title. One of the following enumerated values: • `nan` • `Dr.` • `Mr.` • `Mrs.` • `Ms.` • `Teacher.`
teacher_number_of_previously_posted_projects	Number of previously posted applications previously submitted by the same teacher. **Example:** `2`

See the section **Notes on the Data Fields** for more details about these features.

Additionally, the `resources.csv` data set provides more data about the resources required for each project. Each line in this file represents a resource required by a project:

Feature	Description
`id`	A project identifier
`description`	Description of the resource. **Examples:** `Tenkiri Storyboards Bins, set 20`
`quantity`	Quantity of the resource required. **Example:** `3`
`price`	Price of the resource required. **Example:** `$119.80`

Note: Many projects require multiple resources. The `id` value corresponds to a `project_id` in `train.csv`, so you can join it to the data in `train.csv` to know more about the resources needed for a project.

The data set contains the following label (the value you will attempt to predict):

Label	Description
`project_is_approved`	A binary flag indicating whether Donors Choose approved the project. A value of `0` indicates the project was not approved, and a value of `1` indicates the project was approved.

Load Libraries

```
numpy.linalg import LinalgError
numpy.random import randn
sklearn import ensemble
numpy import np
numpy import nan as NA
matplotlib import rcParams
matplotlib import matplotlib
matplotlib import matplotlib.pyplot as plt
matplotlib import matplotlib.mlab as mlab
matplotlib import matplotlib.ticker as ticker
matplotlib import matplotlib.gridspec as gridspec
matplotlib import matplotlib.colors
from collections import Counter, defaultdict, OrderedDict
from mpl_toolkits.mpl_bar3d import bar3d
from mpl_toolkits.mplot3d import proj3d
from mpl_toolkits.mplot3d import Axes3D
from mpl_toolkits.mplot3d import art3d
from matplotlib import offsetbox
from matplotlib import offsetbox TextArea, DrawingArea, OffsetImage, AnnotationBbox
from matplotlib import offsetbox VPacker, HPacker, PaddedBox
from matplotlib import matplotlib.patches as patches
from matplotlib import matplotlib.path as path
from matplotlib import matplotlib.ticker
from matplotlib import matplotlib.cm as cm
from matplotlib import matplotlib.colors as mcolors, colorConverter
from matplotlib import matplotlib.dates
from matplotlib import matplotlib.lines as lines
from matplotlib import matplotlib.text
from matplotlib import matplotlib.axes
from matplotlib import matplotlib.axes.Axes
from matplotlib import matplotlib.axes.Axes.hist
from matplotlib import matplotlib.axes.Axes.boxplot
from matplotlib import matplotlib.axes.Axes.scatter
from matplotlib import matplotlib.axes.Axes.bar
from matplotlib import matplotlib.axes.Axes.errorbar
from matplotlib import matplotlib.axes.Axes.pcolor
from matplotlib import matplotlib.axes.Axes.pcolorfast
from matplotlib import matplotlib.axes.Axes.imshow
from matplotlib import matplotlib.axes.Axes.text
from matplotlib import matplotlib.axes.Axes.fill_between
from matplotlib import matplotlib.axes.Axes.stackplot
from matplotlib import matplotlib.axes.Axes.table
from matplotlib import matplotlib.axes.Axes.plot_date
from matplotlib import matplotlib.axes.Axes.pie
from matplotlib import matplotlib.axes.Axes.annotate
from matplotlib import matplotlib.axes.Axes.legend
from matplotlib import matplotlib.axes.Axes.axhline
from matplotlib import matplotlib.axes.Axes.axvline
from matplotlib import matplotlib.axes.Axes.axhspan
from matplotlib import matplotlib.axes.Axes.axvspan
from matplotlib import matplotlib.axes.Axes.spy
from matplotlib import matplotlib.axes.Axes.matshow
from matplotlib import matplotlib.axes.Axes.set_xticklabels
from matplotlib import matplotlib.axes.Axes.set_yticklabels
from matplotlib import matplotlib.axes.Axes.grid
from matplotlib import matplotlib.axes.Axes.set_xlim
from matplotlib import matplotlib.axes.Axes.set_ylim
from matplotlib import matplotlib.axes.Axes.set_xlabel
from matplotlib import matplotlib.axes.Axes.set_ylabel
from matplotlib import matplotlib.axes.Axes.set_title
from matplotlib import matplotlib.axes.Axes.get_xticklabels
from matplotlib import matplotlib.axes.Axes.get_yticklabels
from matplotlib import matplotlib.axes.Axes.get_xticklines
from matplotlib import matplotlib.axes.Axes.get_yticklines
from matplotlib import matplotlib.axes.Axes.get_xgridlines
from matplotlib import matplotlib.axes.Axes.get_ygridlines
from matplotlib import matplotlib.axes.Axes.get_xaxis
from matplotlib import matplotlib.axes.Axes.get_yaxis
from matplotlib import matplotlib.axes.Axes.get_xlim
from matplotlib import matplotlib.axes.Axes.get_ylim
from matplotlib import matplotlib.axes.Axes.get_xscale
from matplotlib import matplotlib.axes.Axes.get_yscale
from matplotlib import matplotlib.axes.Axes.get_xticks
from matplotlib import matplotlib.axes.Axes.get_yticks
from matplotlib import matplotlib.axes.Axes.get_xlabel
from matplotlib import matplotlib.axes.Axes.get_ylabel
from matplotlib import matplotlib.axes.Axes.get_title
from matplotlib import matplotlib.axes.Axes.get_legend_handles_labels
from matplotlib import matplotlib.axes.Axes.get_lines
from matplotlib import matplotlib.axes.Axes.get_images
from matplotlib import matplotlib.axes.Axes.get_patches
from matplotlib import matplotlib.axes.Axes.get_texts
from matplotlib import matplotlib.axes.Axes.get_tables
from matplotlib import matplotlib.axes.Axes.get_collections
from matplotlib import matplotlib.axes.Axes.get_containers
from matplotlib import matplotlib.axes.Axes.get_children
from matplotlib import matplotlib.axes.Axes.get_frame
from matplotlib import matplotlib.axes.Axes.get_axes
from matplotlib import matplotlib.axes.Axes.get_figure
from matplotlib import matplotlib.axes.Axes.get_aspect
from matplotlib import matplotlib.axes.Axes.get_adjustable
from matplotlib import matplotlib.axes.Axes.get_anchor
from matplotlib import matplotlib.axes.Axes.get_navigate
from matplotlib import matplotlib.axes.Axes.get_navigate_mode
from matplotlib import matplotlib.axes.Axes.get_interactive_warn
from matplotlib import matplotlib.axes.Axes.get_cursor_properties
from matplotlib import matplotlib.axes.Axes.get_mouseover
from matplotlib import matplotlib.axes.Axes.get_default_bbox_extra_artists
from matplotlib import matplotlib.axes.Axes.get_tightbbox
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsmajor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_xticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
from matplotlib import matplotlib.axes.Axes.get_yticklabelsminor_offset_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text_text
```

1.1 Reading Data

```
train_df = pd.read_csv('../input/train.csv', low_memory=False, parse_dates=['project_submitted_datetime'])
print('Number of rows in train_df', len(train_df))
print('Number of rows in train_df', len(train_df.columns))
train_df.shape, train_df.dtypes, train_df.head(2), train_df.tail(2), train_df.describe(), train_df.describe(include=['O'])
# Resources
resources_df = pd.read_csv('../input/resources.csv', low_memory=False)
print('Number of rows in resources_df', len(resources_df))
print('Number of rows in resources_df', len(resources_df.columns))
resources_df.shape, resources_df.dtypes, resources_df.head(2), resources_df.tail(2), resources_df.describe(), resources_df.describe(include=['O'])
```

Categorical Feature Transformation

2.1 School state

```
train_df['SchoolState'] = train_df['school_state']
# Number of unique school states in each dataset
print('Number of unique school states in train', len(train_df['SchoolState'].unique()))
```

2.3 Teacher_prefix

```
train_df['teacher_prefix'] = train_df['teacher_prefix'].fillna('')
# Number of unique teacher prefixes in each dataset
print('Number of unique teacher prefixes in train', len(train_df['teacher_prefix'].unique()))
```

2.4 project grade category

```
train_df['project_grade_category'] = train_df['project_grade_category']
# Number of unique project grade categories in each dataset
print('Number of unique project grade categories in train', len(train_df['project_grade_category'].unique))
```

2.5 categories

```
train_df['project_subject_categories'] = train_df['project_subject_categories']
# Number of unique project subject categories in each dataset
print('Number of unique project subject categories in train', len(train_df['project_subject_categories'].unique))
```

2.6 sub categories

```
train_df['project_subject_subcategories'] = train_df['project_subject_subcategories']
# Number of unique project subject subcategories in each dataset
print('Number of unique project subject subcategories in train', len(train_df['project_subject_subcategories'].unique))
```

2 Numerical Feature Transformation

2.1 Price Transformation

```
# Price in train and resources
# Price in train and resources, sum, mean, median, min, max, count
# Price in train and resources, sum, mean, median, min, max, count, nunique
train_df['price'] = train_df['price'].fillna(0)
# Price in train and resources, sum, mean, median, min, max, count, nunique
# Price in train and resources, sum, mean, median, min, max, count, nunique, std
# Price in train and resources, sum, mean, median, min, max, count, nunique, std, skew
# Price in train and resources, sum, mean, median, min, max, count, nunique, std, skew, kurt
```

2.2 teacher_number_of_previously_posted_projects Transformation

```
train_df['teacher_number_of_previously_posted_projects'] = train_df['teacher_number_of_previously_posted_projects']
# Teacher_number_of_previously_posted_projects in train and resources, sum, mean, median, min, max, count, nunique, std, skew, kurt
# Teacher_number_of_previously_posted_projects in train and resources, sum, mean, median, min, max, count, nunique, std, skew, kurt, var
train_df['teacher_number_of_previously_posted_projects'] = train_df['teacher_number_of_previously_posted_projects'].fillna(0)
```

3. Vectorizing Text data

3.1 TFIDF vectorizer

```
CountVectorizer, TfidfVectorizer, HashingVectorizer
# Count vectorizer
cv = CountVectorizer(lowercase=True, stop_words='english', ngram_range=(1,1), max_features=10000)
cv_train = cv.fit_transform(train_df['clean_essays'])
print('Shape of sparse matrix with count vectorizer', cv_train.shape)
print('Number of non-zero elements in count vectorizer', cv_train.nnz)
# Tfidf vectorizer
tfidf = TfidfVectorizer(lowercase=True, stop_words='english', ngram_range=(1,1), max_features=10000)
tfidf_train = tfidf.fit_transform(train_df['clean_essays'])
print('Shape of sparse matrix with tfidf vectorizer', tfidf_train.shape)
print('Number of non-zero elements in tfidf vectorizer', tfidf_train.nnz)
```

3.2 TfIdf weighted W2V

```
# glove vectors
glove_words = set(w for line in open('../input/glove.6B.300d.txt', 'r', encoding='utf-8').split(" ")[0] for line in open('../input/glove.6B.300d.txt', 'r', encoding='utf-8'))
print('Number of words in glove vectors', len(glove_words))
# Word vectors
word_vectors = {}
ss = time.time()
for line in open('../input/glove.6B.300d.txt', 'r', encoding='utf-8'):
    split_line = line.split(" ")
    word = split_line[0]
    try:
        word_vectors[word] = np.array([float(i) for i in split_line[1:]])
    except ValueError:
        print('ValueError', word)
    if (time.time() - ss) > 60:
        print('Break while loop for 60 seconds')
        break
print('Number of words with glove vectors', len(word_vectors))
# Word vectors for resources
word_vectors = {}
ss = time.time()
for line in open('../input/glove.6B.300d.txt', 'r', encoding='utf-8'):
    split_line = line.split(" ")
    word = split_line[0]
    try:
        word_vectors[word] = np.array([float(i) for i in split_line[1:]])
    except ValueError:
        print('ValueError', word)
    if (time.time() - ss) > 60:
        print('Break while loop for 60 seconds')
        break
print('Number of words with glove vectors', len(word_vectors))
```

3.3 Analyse Essay Sentiments

```
train_df['sentiment_score_essay'] = train_df['clean_essays'].apply(lambda x: sia.polarity_scores(x))
# Sentiment score for resources
resources_df['sentiment_score_resources'] = resources_df['description'].fillna('').apply(lambda x: sia.polarity_scores(x))
```

3.4 Downcasting Column Data Type

```
train_df[numeric_cols] = train_df[numeric_cols].apply(pd.to_numeric)
# Downcasting numeric columns
train_df[numeric_cols] = train_df[numeric_cols].apply(pd.to_numeric, downcast='float')
# Downcasting numeric columns
train_df[numeric_cols] = train_df[numeric_cols].apply(pd.to_numeric, downcast='integer')
# Downcasting numeric columns
train_df[numeric_cols] = train_df[numeric_cols].apply(pd.to_numeric, downcast='unsigned')
```

```
# Fit the model on the training data.
model_dt = DecisionTreeClassifier(max_depth=3)
model_dt.fit(text_clf_dt[0][0], text_clf_dt[0][1])
# Predicting on the test data.
y_pred = model_dt.predict(text_clf_dt[0][2])
# Evaluating the model.
from sklearn.metrics import classification_report
print(f"Decision Tree Report:\n", classification_report(text_clf_dt[0][3], y_pred))
```

```
Decision Tree Report:
              precision    recall  f1-score   support

           0       0.00      0.00      0.00       1.0
           1       0.00      0.00      0.00       1.0
           2       0.00      0.00      0.00       1.0
           3       0.00      0.00      0.00       1.0
           4       0.00      0.00      0.00       1.0
           5       0.00      0.00      0.00       1.0
           6       0.00      0.00      0.00       1.0
           7       0.00      0.00      0.00       1.0
```