Box-Plots for Education

The features in this dataset

Your goal is to predict the **probability** that a certain label is attached to a budget line item. Each row in the budget has mostly **free-form text** features, except for the two below that are noted as **float**. Any of the fields may or may not be empty

- `FTE` **float** - If an employee, the percentage of full-time that the employee works.
- `Facility_or_Department` - If expenditure is tied to a department/facility, that department/facility.
- `Function_Description` - A description of the function the expenditure was serving.
- `Fund_Description` - A description of the source of the funds.
- `Job_Title_Description` - If this is an employee, a description of that employee's job title.
- `Location_Description` - A description of where the funds were spent.
- `Object_Description` - A description of what the funds were used for.
- `Position_Extra` - Any extra information about the position that we have.
- `Program_Description` - A description of the program that the funds were used for.
- `SubFund_Description` - More detail on `Fund_Description`
- `Sub_Object_Description` - More detail on `Object_Description`
- `Text_1` - Any additional text supplied by the district.
- `Text_2` - Any additional text supplied by the district.
- `Text_3` - Any additional text supplied by the district.
- `Text_4` - Any additional text supplied by the district.
- `Total` **float** - The total cost of the expenditure.

Feature data example

For example, a single row in the dataset might have these values:

FTE	1
Facility_or_Department	NaN
Function_Description	Care and Upkeep of Building Services
Fund_Description	General
Job_Title_Description	CLEANER
Location_Description	BUILDING OPERATIONS SECTION
Object_Description	Regular *

Position_Extra	NaN
Program_Description	NaN
SubFund_Description	Operation and Maintenance of Plant Services
Sub_Object_Description	NaN
Text_1	FACILITIES ADMINISTRATION
Text_2	FACILITIES
Text_3	NaN
Text_4	NaN
Total	36664.44

The labels in this dataset

For each of these rows, ERS attaches one label from each of 9 different categories:

- Function:

  - Aides Compensation
  - Career & Academic Counseling
  - Communications
  - Curriculum Development
  - Data Processing & Information Services
  - Development & Fundraising
  - Enrichment
  - Extended Time & Tutoring
  - Facilities & Maintenance
  - Facilities Planning
  - Finance, Budget, Purchasing & Distribution
  - Food Services
  - Governance
  - Human Resources
  - Instructional Materials & Supplies
  - Insurance
  - Legal
  - Library & Media
  - NO_LABEL
  - Other Compensation
  - Other Non-Compensation
  - Parent & Community Relations
  - Physical Health & Services

- Professional Development
- Recruitment
- Research & Accountability
- School Administration
- School Supervision
- Security & Safety
- Social & Emotional
- Special Population Program Management & Support
- Student Assignment
- Student Transportation
- Substitute Compensation
- Teacher Compensation
- Untracked Budget Set-Aside
- Utilities

- Object_Type:

  - Base Salary/Compensation
  - Benefits
  - Contracted Services
  - Equipment & Equipment Lease
  - NO_LABEL
  - Other Compensation/Stipend
  - Other Non-Compensation
  - Rent/Utilities
  - Substitute Compensation
  - Supplies/Materials
  - Travel & Conferences

- Operating_Status:

  - Non-Operating
  - Operating, Not PreK-12
  - PreK-12 Operating

- Position_Type:

  - (Exec) Director
  - Area Officers
  - Club Advisor/Coach
  - Coordinator/Manager
  - Custodian

- Guidance Counselor
- Instructional Coach
- Librarian
- NO_LABEL
- Non-Position
- Nurse
- Nurse Aide
- Occupational Therapist
- Other
- Physical Therapist
- Principal
- Psychologist
- School Monitor/Security
- Sec/Clerk/Other Admin
- Social Worker
- Speech Therapist
- Substitute
- TA
- Teacher
- Vice Principal
- Pre_K:

  - NO_LABEL
  - Non PreK
  - PreK
- Reporting:

  - NO_LABEL
  - Non-School
  - School
- Sharing:

  - Leadership & Management
  - NO_LABEL
  - School Reported
  - School on Central Budgets
  - Shared Services
- Student_Type:

  - Alternative

- At Risk
  - ELL
  - Gifted
  - NO_LABEL
  - Poverty
  - PreK
  - Special Education
  - Unspecified
- Use:

  - Business Services
  - ISPD
  - Instruction
  - Leadership
  - NO_LABEL
  - O&M
  - Pupil Services & Enrichment
  - Untracked Budget Set-Aside

Note, there is a hierarchical relationship for these labels. If a line is marked as `Non-Operating` in the `Operating_Status` category, then all of the other labels should be marked as `NO_LABEL` since ERS does not analyze and compare non-operating budget items.

Label example

The row in the example above would have the following labels attached to it:

Function	Facilities & Maintenance
Object_Type	Base Salary/Compensation
Operating_Status	PreK-12 Operating
Position_Type	Custodian
Pre_K	Non PreK
Reporting	School
Sharing	School Reported
Student_Type	Unspecified
Use	O&M

Submission format

Your goal is to predict a probability for each possible label in the dataset given a row of new data. Each

of these probabilities goes in a separate column in the submission file. The submission must be `50064x104` where 50064 is the number of rows in the test dataset (excluding the header) and 104 is the number of columns (excluding a first column of row ids). The columns in the submission have the format `ColumnName__PossibleLabel`, which we have listed below for your convenience. This is simply a flattening of [the labels](#) that we listed above.

- `Function__Aides Compensation`
- `Function__Career & Academic Counseling`
- `Function__Communications`
- `Function__Curriculum Development`
- `Function__Data Processing & Information Services`
- `Function__Development & Fundraising`
- `Function__Enrichment`
- `Function__Extended Time & Tutoring`
- `Function__Facilities & Maintenance`
- `Function__Facilities Planning`
- `Function__Finance, Budget, Purchasing & Distribution`
- `Function__Food Services`
- `Function__Governance`
- `Function__Human Resources`
- `Function__Instructional Materials & Supplies`
- `Function__Insurance`
- `Function__Legal`
- `Function__Library & Media`
- `Function__NO_LABEL`
- `Function__Other Compensation`
- `Function__Other Non-Compensation`
- `Function__Parent & Community Relations`
- `Function__Physical Health & Services`
- `Function__Professional Development`
- `Function__Recruitment`
- `Function__Research & Accountability`
- `Function__School Administration`
- `Function__School Supervision`
- `Function__Security & Safety`
- `Function__Social & Emotional`
- `Function__Special Population Program Management & Support`
- `Function__Student Assignment`

- Function__Student Transportation
- Function__Substitute Compensation
- Function__Teacher Compensation
- Function__Untracked Budget Set-Aside
- Function__Utilities
- Object_Type__Base Salary/Compensation
- Object_Type__Benefits
- Object_Type__Contracted Services
- Object_Type__Equipment & Equipment Lease
- Object_Type__NO_LABEL
- Object_Type__Other Compensation/Stipend
- Object_Type__Other Non-Compensation
- Object_Type__Rent/Utilities
- Object_Type__Substitute Compensation
- Object_Type__Supplies/Materials
- Object_Type__Travel & Conferences
- Operating_Status__Non-Operating
- Operating_Status__Operating, Not PreK-12
- Operating_Status__PreK-12 Operating
- Position_Type__(Exec) Director
- Position_Type__Area Officers
- Position_Type__Club Advisor/Coach
- Position_Type__Coordinator/Manager
- Position_Type__Custodian
- Position_Type__Guidance Counselor
- Position_Type__Instructional Coach
- Position_Type__Librarian
- Position_Type__NO_LABEL
- Position_Type__Non-Position
- Position_Type__Nurse
- Position_Type__Nurse Aide
- Position_Type__Occupational Therapist
- Position_Type__Other
- Position_Type__Physical Therapist
- Position_Type__Principal
- Position_Type__Psychologist
- Position_Type__School Monitor/Security
- Position_Type__Sec/Clerk/Other Admin

- Position_Type__Social Worker
- Position_Type__Speech Therapist
- Position_Type__Substitute
- Position_Type__TA
- Position_Type__Teacher
- Position_Type__Vice Principal
- Pre_K__NO_LABEL
- Pre_K__Non PreK
- Pre_K__PreK
- Reporting__NO_LABEL
- Reporting__Non-School
- Reporting__School
- Sharing__Leadership & Management
- Sharing__NO_LABEL
- Sharing__School Reported
- Sharing__School on Central Budgets
- Sharing__Shared Services
- Student_Type__Alternative
- Student_Type__At Risk
- Student_Type__ELL
- Student_Type__Gifted
- Student_Type__NO_LABEL
- Student_Type__Poverty
- Student_Type__PreK
- Student_Type__Special Education
- Student_Type__Unspecified
- Use__Business Services
- Use__ISPD
- Use__Instruction
- Use__Leadership
- Use__NO_LABEL
- Use__O&M
- Use__Pupil Services & Enrichment
- Use__Untracked Budget Set-Aside