

Project Step - 2

Data Collection

Harsh Siddhapura
Shivkesh Madasu
Shrenika Soma
Jhansi Alugoju

02/11/2024

“Forecasting Philanthropy: A Predictive Analysis for Donors Supporting Various School Projects in the USA”

Project: Our team will build a regression model that forecasts/predicts analysis for donors to support various projects in different schools in the United States of America. The title of the project is, “Forecasting Philanthropy: A Predictive Analysis for Donors Supporting Various School Projects in the USA”.

Google Drive Dataset Link:

<https://drive.google.com/drive/folders/1fSaWJX43KUwbZgV0Atm0fDUwaCpfTOU9?usp=sharing>

+++++

1. Data Sources:

- Link - 1: Donations.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Donations.csv>
- Link - 2: Donors.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Donors.csv>
- Link - 3: Projects.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Projects.csv>
- Link - 4: Resources.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Resources.csv>
- Link - 5: Schools.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Schools.csv>
- Link - 6: Teachers.csv
<https://www.kaggle.com/datasets/perkymaster/school-donations/data?select=Teachers.csv>

+++++

2. Number of data points collected (How many rows):

- Donations.csv: **4687884** Rows
- Donors.csv: **2122640** Rows
- Projects.csv: **1110017** Rows
- Resources.csv: **7210448** Rows
- Schools.csv: **72993** Rows
- Teachers.csv: **402900** Rows

+++++

3. Number of attributes collected (How many columns):

- **Donations.csv: 7 Columns**
 - Project ID
 - Donation ID
 - Donor ID
 - Donation Included Optional Donation
 - Donation Amount
 - Donor Cart Sequence
 - Donation Received Date
- **Donors.csv: 5 Columns**
 - Donor ID
 - Donor City
 - Donor State
 - Donor Is Teacher
 - Donor Zip
- **Projects.csv: 18 Columns**
 - Project ID
 - School ID
 - Teacher ID
 - Teacher Project Posted Sequence
 - Project Type
 - Project Title
 - Project Essay
 - Project Short Description

- Project Need Statement
- Project Subject Category Tree
- Project Subject Subcategory Tree
- Project Grade Level Category
- Project Resource Category
- Project Cost
- Project Posted Date
- Project Expiration Date
- Project Current Status
- Project Fully Funded Date
- **Resources.csv: 5 Columns**
 - Project ID
 - Resource Item Name
 - Resource Quantity
 - Resource Unit Price
 - Resource Vendor Name
- **Schools.csv: 9 Columns**
 - School ID
 - School Name
 - School Metro Type
 - School Percentage Free Lunch
 - School State
 - School Zip
 - School City
 - School County
 - School District
- **Teachers.csv: 3 Columns**
 - Teacher ID
 - Teacher Prefix
 - Teacher First Project Posted Date

Total: 47 Columns

+++++

4. A brief explanation for what each column represents.

○ Donations.csv: 7 Columns

- **Project ID:** A unique identifier for the project.
- **Donation ID:** A unique identifier for the donation.
- **Donor ID:** A unique identifier for the donor who made the donation.
- **Donation Included Optional Donation:** Indicates whether the donation included an optional donation to support the platform.
- **Donation Amount:** The amount of money donated.
- **Donor Cart Sequence:** The sequence number of the donation in the donor's cart.
- **Donation Received Date:** The date and time when the donation was received.

○ Donors.csv: 5 Columns

- **Donor ID:** A unique identifier for the donor.
- **Donor City:** The city where the donor is located.
- **Donor State:** The state where the donor is located.
- **Donor Is Teacher:** Indicates whether the donor is also a teacher (Yes/No).
- **Donor Zip:** The zip code of the donor's location.

○ Projects.csv: 18 Columns

- **Project ID:** A unique identifier for the project.
- **School ID:** A unique identifier for the school where the project is being implemented.
- **Teacher ID:** A unique identifier for the teacher who posted the project.
- **Teacher Project Posted Sequence:** The sequence number of the project posted by the same teacher.
- **Project Type:** The type of the project (e.g., Teacher-Led).
- **Project Title:** The title of the project.
- **Project Essay:** A detailed description of the project.
- **Project Short Description:** A brief description of the project.
- **Project Need Statement:** A statement that explains what the project needs.
- **Project Subject Category Tree:** The main category of the project subject.
- **Project Subject Subcategory Tree:** The subcategory of the project subject.
- **Project Grade Level Category:** The grade level for which the project is intended.
- **Project Resource Category:** The category of resources required for the project.
- **Project Cost:** The total cost of the project.
- **Project Posted Date:** The date when the project was posted.
- **Project Expiration Date:** The date when the project will expire.

- **Project Current Status:** The current status of the project (e.g., Fully Funded, Expired).
- **Project Fully Funded Date:** The date when the project was fully funded.
- **Resources.csv: 5 Columns**
 - **Project ID:** A unique identifier for the project.
 - **Resource Item Name:** The name of the resource item required for the project.
 - **Resource Quantity:** The quantity of the resource item required.
 - **Resource Unit Price:** The price per unit of the resource item.
 - **Resource Vendor Name:** The name of the vendor from whom the resource item is purchased.
- **Schools.csv: 9 Columns**
 - **School ID:** A unique identifier for the school.
 - **School Name:** The name of the school.
 - **School Metro Type:** The type of area where the school is located (e.g., rural, urban).
 - **School Percentage Free Lunch:** The percentage of students at the school who receive free lunch.
 - **School State:** The state where the school is located.
 - **School Zip:** The zip code of the school's location.
 - **School City:** The city where the school is located.
 - **School County:** The county where the school is located.
 - **School District:** The school district to which the school belongs.
- **Teachers.csv: 3 Columns**
 - **Teacher ID:** A unique identifier for the teacher.
 - **Teacher Prefix:** The prefix of the teacher's name (e.g., Mrs., Mr.).
 - **Teacher First Project Posted Date:** The date when the teacher posted their first project.

+++++

5. A brief explanation for what each row represents.

- **Donations.csv:** Each row in the data represents a unique donation made to a specific project by a donor. For example, the first row represents a donation of \$178.37 made by the donor with ID 1f4b5b6e68445c6c4a0509b3aca93f38 to the project with ID 000009891526c0ade7180f8423792063. This donation did not include an optional

donation to support the platform, and it was the 11th donation in the donor's cart. The donation was received on 2016-08-23 13:15:57.

- **Donors.csv:** Each row in the data represents a unique donor with all their details as described by the columns. For example, the first row represents a donor with ID 00000ce845c00cbf0686c992fc369df4 who is located in Evanston, Illinois, is not a teacher, and has the zip code 602. Similarly, the second row represents another donor with different details.
- **Projects.csv:** Each row in the data represents a unique project with all the details as described by the columns. For example, the first row represents a project titled "Stand Up to Bullying: Together We Can!" which was posted by a teacher (with the ID 4ee5200e89d9e2998ec8baad8a3c5968) at a school (with the ID e180c7424cb9c68cb49f141b092a988f). The project was fully funded on 2013-01-11. The project details, its needs, cost, and other related information are also provided in the same row.
- **Resources.csv:** Each row in the data represents a unique resource required for a specific project. For example, the first row represents a resource named "chair move and store cart" required for the project with ID 000009891526c0ade7180f8423792063. This resource has a unit price of \$350.00, and 1 unit of this resource is required for the project. The vendor's name is not provided in this case (represented as NaN). Similarly, the second and third rows represent other resources required for different projects with their respective details.
- **Schools.csv:** Each row in the data represents a unique school with all its details as described by the columns. For example, the first row represents Capon Bridge Middle School, which is located in a rural area in Capon Bridge, Hampshire, West Virginia, with zip code 26711. The school belongs to the Hampshire Co School District, and 56% of its students receive free lunch. Similarly, the second row represents another school with different details.
- **Teachers.csv:** Each row in the data represents a unique teacher with all their details as described by the columns. For example, the first row represents a teacher with ID 00000f7264c27ba6fea0c837ed6aa0aa who uses the prefix "Mrs." and posted their first project on 2013-08-21. Similarly, the second and third rows represent other teachers with different details.

+++++

6. The list of rows/columns collected by each team member.

- **Donations.csv: 7 Columns**
 - Project ID: Shrenika Soma
 - Donation ID: Jhansi Alugoju
 - Donor ID: Shivkesh Madasu
 - Donation Included Optional Donation: Harsh Siddhapura
 - Donation Amount: Shrenika Soma
 - Donor Cart Sequence: Jhansi Alugoju
 - Donation Received Date: Harsh Siddhapura
- **Donors.csv: 5 Columns**
 - Donor ID: Shivkesh Madasu
 - Donor City: Shrenika Soma
 - Donor State: Harsh Siddhapura
 - Donor Is Teacher: Jhansi Alugoju
 - Donor Zip: Harsh Siddhapura
- **Projects.csv: 18 Columns**
 - Project ID: Harsh Siddhapura
 - School ID: Shrenika Soma
 - Teacher ID: Shivkesh Madasu
 - Teacher Project Posted Sequence: Jhansi Alugoju
 - Project Type: Shrenika Soma: Harsh Siddhapura
 - Project Title: Shivkesh Madasu
 - Project Essay: Jhansi Alugoju
 - Project Short Description: Harsh Siddhapura
 - Project Need Statement: Shrenika Soma
 - Project Subject Category Tree: Shivkesh Madasu
 - Project Subject Subcategory Tree: Jhansi Alugoju
 - Project Grade Level Category: Harsh Siddhapura
 - Project Resource Category: Shivkesh Madasu
 - Project Cost: Harsh Siddhapura
 - Project Posted Date: Shrenika Soma
 - Project Expiration Date: Harsh Siddhapura
 - Project Current Status: Jhansi Alugoju
 - Project Fully Funded Date: Shivkesh Madasu
- **Resources.csv: 5 Columns**
 - Project ID: Jhansi Alugoju

- Resource Item Name: Shivkesh Madasu
- Resource Quantity: Shrenika Soma
- Resource Unit Price: Harsh Siddhapura
- Resource Vendor Name: Shivkesh Madasu

○ **Schools.csv: 9 Columns**

- School ID: Shivkesh Madasu
- School Name: Jhansi Alugoju
- School Metro Type: Shrenika Soma
- School Percentage Free Lunch: Harsh Siddhapura
- School State: Shivkesh Madasu
- School Zip: Shrenika Soma
- School City: Shivkesh Madasu
- School County: Jhansi Alugoju
- School District: Harsh Siddhapura

○ **Teachers.csv: 3 Columns**

- Teacher ID: Shrenika Soma
- Teacher Prefix: Jhansi Alugoju
- Teacher First Project Posted Date: Harsh Siddhapura

+++++

References

1. Kaggle: <https://www.kaggle.com/datasets/perkymaster/school-donations/data>
2. DonorsChoose: <https://www.donorschoose.org/>
3. Greater Cedar Rapids Community Foundation:
<https://www.gcrcf.org/teachers-students/donors-supporting-education/>