

Utilizing Data Mining Techniques to Link Dropout Rates and Societal Variables

Laura Morales
Advisor: Dr. Stefan Robila
Department of Computer Science
Montclair State University
Montclair, NJ 07043

Master Project
Fall 2016

Abstract

Education is a very important factor that can help improve and enhance numerous lives. Within our society and communities, we have great number of young individuals who are dropping out of high school for various reasons. This unfortunate event tends to have a long lasting impact on

their lives, lowering the amount of opportunities available to them, and possibly lowering the quality of life they can have in the future. Having statistics available to determine how many individuals are dropping out of high school is not enough. We need to start focusing on other variables related to why individuals are dropping out of high school. With this information, we can begin the process of determining possible root causes of this unfavorable decision and start providing guidance, help, or information to the affected or at risk students or families. This process can begin here, by developing easy to use accessible tools for data processing and visualization.

For my project, I developed a web-based tool focused on processing educational retention data sets. This site is publicly accessible (<https://msuproject.herokuapp.com/pages/index>) and allows an individual to upload user specific information according to your needs, such as data related to their state/city/county/school, clean (data mine) the information using OpenRefine, download and save the clean dataset, generate a graph to help summarize and correlate variables to help draw meaningful conclusions, and lastly download the generated graph for their records. The development of the tool allowed me to learn new software building techniques such as writing in HTML, implementing API's to generate graphs, and learning how to customize the graphs to display data in different ways.

Acknowledgements:

I consider this Master's Project an opportunity to try to help and give back to my community by helping discover patterns or variables that may potentially lead to students dropping out of high school. It is my opinion that education is such an important factor in our lives, which can impact our future opportunities and living environments. There are a large amount of adolescents that dropout every year for varies reasons, and although the overall dropout rate has started to decrease, we still need to work together to help find these gaps and bridge them not only for their futures, but for our society.

I want to thank my Advisor, Dr. Stefan Robila, for giving me this opportunity to work not his project which can hopefully be used to help give back to those adolescents and communities that need our help, and educate and deter students from dropping out.

Lastly, I want to thank my mother for instilling the importance of education in me at an early age. Growing up in the low/medium class, with a single mother and two other siblings allowed me to experience first hand some of the challenges and disadvantages that can hinder the impact education can have in one's life, or the ability to complete education. Seeing my mother struggle and sacrifice for our family, gave me the motivation and drive to aspire to become someone; someone who challenges herself and does not settle, and more importantly someone that can one day help those children and adolescents who are now struggling due to the limitations that they are facing, regardless of what those limitations are.

Table of Contents

Abstract	2
Acknowledgements:	3
Table of Contents	5
List of Figures	7
1. Introduction	8
2. Literature and Data	10
2.1. Previous work	10
2.2. Data Used	15
3. Application Design and Structure	18
3.1. Application Design	18
3.2. Code Structure	25
3.2.1 Ruby on Rails	26
3.2.2. Bootstrap	29
3.2.3. OpenRefine	32
3.2.4. Highcharts	36
3.2.5. Github, Amazon S3, and Heroku	40
4. Project Results	43
4.1. Field Information	43
5. Implementation	51
6. Future Advancements	53
6.1. Chart Options and Enhancements	54
6.2. Website Design and Layout Improvements	56
7. Conclusion	58
References	59

	5
Appendix I - Creation of Web Application	62
Appendix II - Source Code	66

List of Figures

Figure 1: Child Trends Website which offers report with charts on Mothers noting competing high school.	10
Figure 2: The State of New Jersey Department of Education: downloadable Excel sheets for graduation rates	11
Figure 3: Different online tools that are available within the US Census Website	12
Figure 4: The Master's Project homepage within the web application.	17
Figure 5: The About Page within the Web Application.	18
Figure 6: The Project Results page from the web application.	19
Figure 7: The Data Mining page from the web application.	20
Figure 8: Generate Graph page in web application.	21
Figure 8.1: Generate Graph page in web application, with an external CSV file uploaded and chart generated.	21
Figure 9: Contact page in web application.	22
Figure 10: Bootstrap script requirements for implementation within web application.	26
Figure 11: This image reflects how the layouts of the web application will be called upon in the main application.html layout page.	27
Figure 12: OpenRefine upload file option.	28
Figure 13: OpenRefine initial uploaded file changes/edits.	29
Figure 14: An example of OpenRefine edit options.	30
Figure 15: This image reflects the standard code that was used to generate the Project Results chart using the Highchart API.	32
Figure 16: Highcharts view	33
Figure 17: Github repository updated code with comments and dates.	35
Figure 18: Amazon S3 uploaded file and file properties.	36
Figure 19: Raw CVS data (above)	44
Figure 20: CSV finalized data after data mining/data cleaning (above)	44
Figure 21: Highchart example of how a user can select the options to print or download the	

1. Introduction

Dropout and Graduation rates are statistics that are constantly updated, reviewed, and used to determine how many adolescents are able to successfully complete their high school education [13]. This information is generally made public and often mentioned in topics related to social and community concerns. There are a vast amount of variables that impact these statistics, with the most common factor related to economic, race, and social status' [13]. It is my opinion, that among these variables, there are numerous other elements that can potentially play a large role and can have an enormous impact on graduation or dropout rates for those individuals. Having the ability to obtain, manipulate , and analyze this information, can potentially help determine or find new patterns or variables that impact the dropout rates for students within certain the specified communities. In turn, as these newly found patterns are confirmed, this will allow members of society to step in to help address these issue and improve the graduation rates of the upcoming generations. At this time, unfortunately not all of this information is currently being collected or is readily available for analysis.

For my project, I developed a web-based tool focused on processing educational retention data sets. This site is publicly accessible and allows an individual to upload user specific information according to your needs, such as data related to their state/city/county/school, clean (data mine)

the information using OpenRefine, download and save the clean dataset, generate a graph to help summarize and correlate variables to help draw meaningful conclusions, and lastly download the generated graph for their records.

This report is organized as follows. In Section 2 I describe the research that was conducted to help determine what type of information is currently available, and was used for this project. Section 3 covers the details regarding the design and structure for this web application. In Section 4 discusses the results I was able to conclude based on the information I analyzed for the Virginia Department of Education. The implementation of this web application in regards to my research and how this tool can be used online, is covered in Section 5. Section 6, reviews the future implementations and improvements that can be completed on this web application, to help make this tool more productive and effective. Lastly, Section 7 covers the conclusions that were made from the research that was completed in this project and the creation of the online web application as a future tool.

2. Literature and Data

2.1. Previous work

After conducting extensive research, I was unable to locate any online tools that can help analyze or visualize high school dropout data. The current data that is available, is presented as a report or downloadable CSV/Excel file, depending on the resource. The interpretation and conclusions from the provided data is generally outlined within the data source, or is opened to interpretation from the user, assuming they are able to understand the data at hand. Kindly see below for a list of the different types of data I was able to find during the research for this project and tool:

- A. Child Trends, an online source that has databank trends available for various subjects, provides a downloadable PDF for “Parental Education”, which has a report on information that was already collected and analyzed, about mothers who did not complete high school between 1974-2015 [23]. Kindly see Figure 1 below for an example of this site.

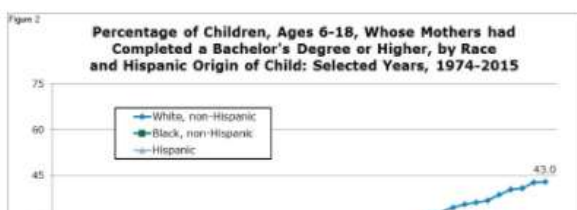
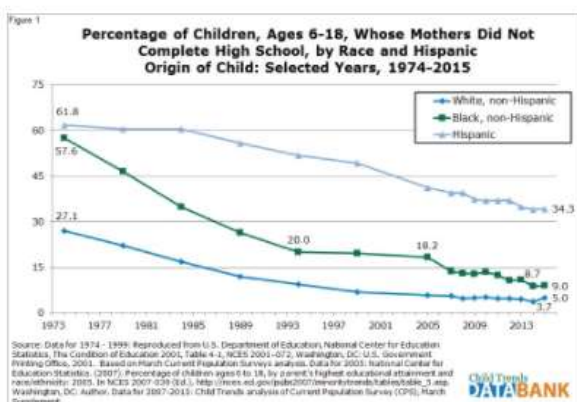


Figure 1: Child Trends Website which offers report with charts on Mothers noting competing high school.

B. The State of New Jersey Department of Education website, the Maryland State Department of Education website, and the Virginia Department of Education website all provides downloadable excel sheets, CSV files, and/or reports of data, pertaining to either the graduation or dropout rates for review and analyzation, (please see Figure 2 below for an example of what The State of New Jersey Department of Education website provides).

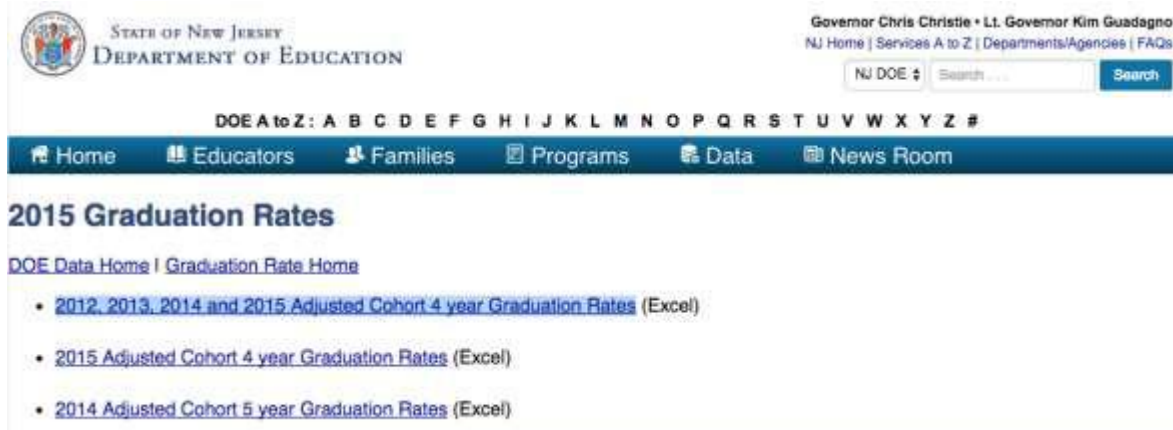


Figure 2: The State of New Jersey Department of Education: downloadable Excel sheets for graduation rates

C. The US Census website is the closest option I was able to find, which allows users to review and manipulate some of the data that has been collected, such as community information, economic data, and business statistics. However, I was unable to find a tool that allows the user to review/visualize any data pertaining to high school graduation or dropout rates. Please see Figure 3 below, for a screenshot of the tools available within the US Census Website).

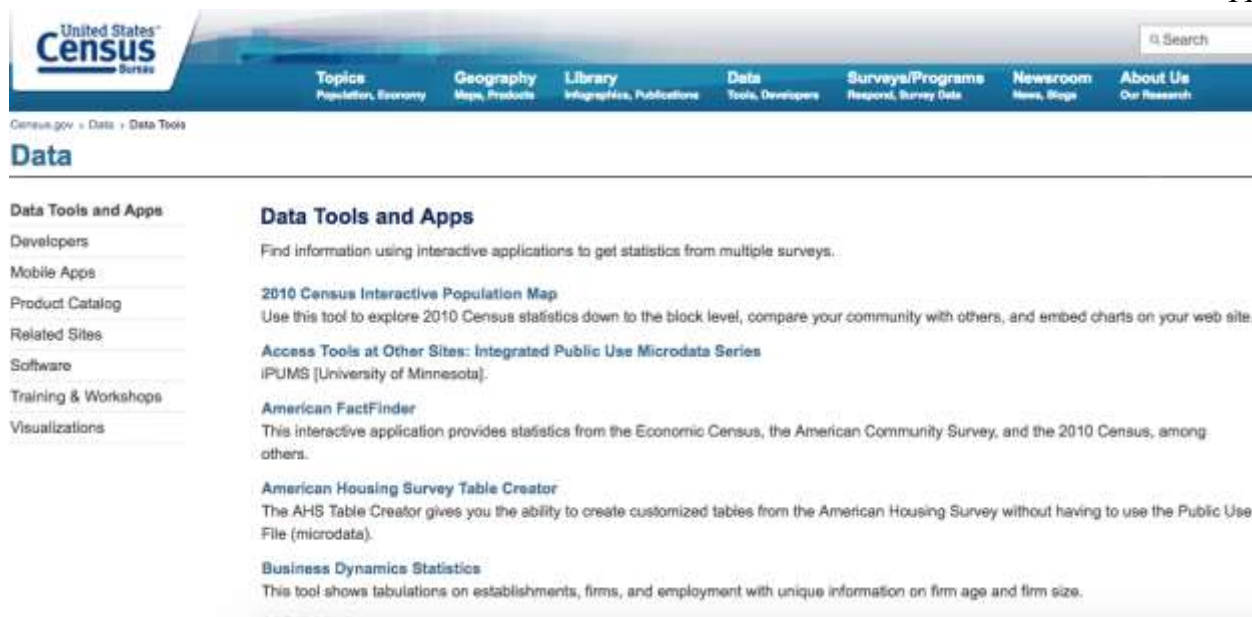


Figure 3: Different online tools that are available within the US Census Website

Aside from not being able to find a tool that can help clean and/or analyze the raw data that can be found on various websites and resources; the reports and information available on these sites tend to have the data either in raw form (meaning within an excel sheet or CSV file) or provides the user with a report that already has predefined conclusions from the author's analysis of the data. It is my opinion, that these two options do not allow the user to come to their own conclusions, by reviewing the information and trying to determine their own options on the data available to them.

2.2. Data Used

The statistics about graduation and dropout rates are made public conditional on public schools and the information that state or county deems necessary to either collect, or make available to the public. In addition, at the time of the project development, I was unable to locate one single location or database that aggregates this information in a readily available location. Setting aside the type of information that is being collected, the concept that you have to go to different and individuals locations, such as the county or state web pages to obtain information reflects the lack of data curation currently available for this specific topic. The information that is available, has already been compiled and grouped into specific categories to be downloaded. Therefore, searching through it (or in fact any further analysis) is not feasible at this time.

The data that is currently available in relation to dropout and graduation rates is not consistent among the different counties or states. For this purpose, I began to search for this information specifically for New Jersey, Maryland, and Virginia to see what type of information was available. When reviewing the different types of information available for the three states, I was able to notice the lack of consistency with the data that is collected among all three states. The information that was located for the states of Maryland, New Jersey, and Virginia, were located in their Department of Education website. More specifically, the dropout rates that are available for Maryland can be accessed by navigating directly to their Department of Education website, or by going to their *ReportCard.msde.maryland.gov* to view or download the data that was collected for the state [30]. For New Jersey, navigating directly to The New Jersey Department of Education website and searching for “dropout” will allow the user to obtain the dropout data

for their desired year or time frame [31]. Lastly, getting the dropout information for the state of Virginia, is very similar as the previous two states, as this information is also located within the Virginia Department of Education, and can easily be found by searching for “dropout” within the website [32]. These states did have the most basic information available, such as gender, race, dropout and graduation rates, however having limited variables to analyze, will by default limit the outcome or possible conclusions that are available for this research.

Among these three states, Virginia had the largest amount of information available, with the most amount of vast variables being incorporated. The data that was obtainable for all three states, were downloadable data sets with predefined variables or concepts for each dataset. The data available on the Virginia Department of Education site, incorporated the following type of information:

- 2008-2014 On-time graduation rate and cohort dropout rates
- 2010-2012 Annual dropout data
- 2006-2015 Annual HS Graduates and Completer's
- 2007-2015 Career & Technical Education Programs Graduates and Completer's
- 2006-2015 Test Data

The different factors that were incorporated in the majority of these datasets are below:

- School year

- Level code
- Division number and name
- School number and name
- Federal Race Code
- Gender
- Disability flag
- LEP flag
- Disadvantaged flag
- Dropout rate

Having these different types of variables available for analyzation can potentially help narrow down additional factors that can affect graduation and dropout rates. For this reason, the Virginia data was used in this project.

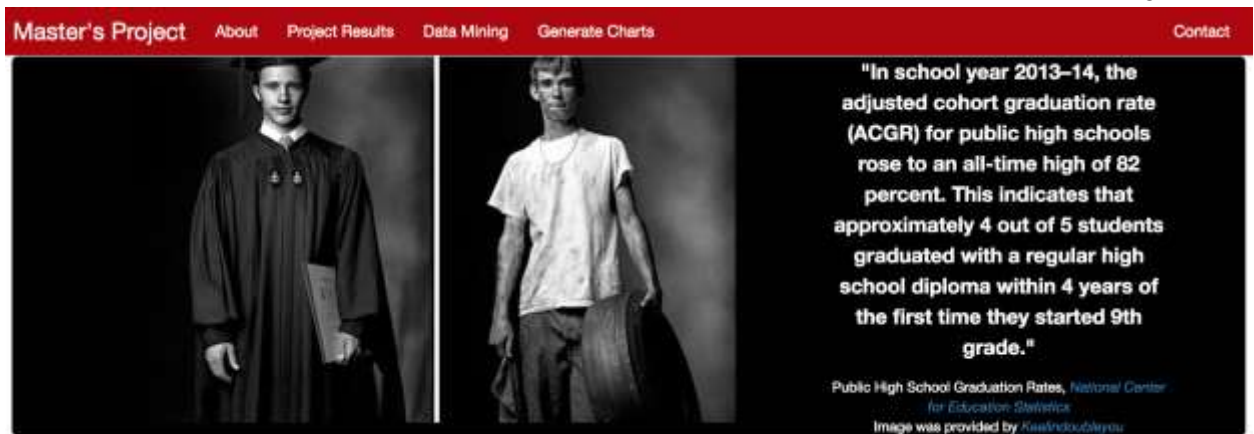
3. Application Design and Structure

3.1. Application Design

This website is broken up into six different pages that are easily accessed and viewable online.

Kindly see below for a clear explanation for each page.

- A. Master's Project Page: The Master's Project page is the home page for this website, which is introduced by an image of a young man, portrayed as a graduate, and as a labor intensive employee. It is my opinion, that this images reflects the very possibly paths some adolescents and young adults have ,when making the decision to complete their education. This page also provides a brief introductory section giving the user some background information on the foundation of the website, and discloses how this website can be used by the user. It provides a quick synopsis of the data and chart pages, and describes how they can be used by the user. The "Data Mining" and "Generate Charts" descriptions on this page, will also open a new tab and auto navigate the user to these pages, so they do not have to go to the navigation bar at the top for convenience.



Graduation and Dropout Statistics

Education is a very important factor that can help improve and enhance numerous lives. Within our society and communities, we have great number of young individuals who are dropping out of high school for various reasons. This unfortunate event tends to have a long lasting impact on their lives, lowering the amount of opportunities available to them, and possibly lowering the quality of life they can have in the future. In my opinion, having the statistics to how many individuals are dropping out of high school is not enough. We need to start focusing on other variables related to why individuals are dropping out of high school. With this information, we can begin the process of determining possible root causes of this unfavorable decision and if possible start providing guidance, help, or information to the affected students or families. This process can begin here.

This site allows you to upload information according to your needs, such as information specific to your state/city/county/school, clean (data mine) the information, and generate a graph to help summarize and correlate variables to help draw meaning conclusions. To begin, you may follow the steps below:

1. **Data Mining:** Using OpenRefine, you will be able to upload raw/messy data to the tool and conduct various data mining/cleaning techniques to make the data more useful and easier to use. For this website, the data must be exported as a CSV file from the tool, in order to generate graphs/charts section. After cleaning the data, you will download the updated CSV file and may

Figure 4: The Master's Project homepage within the web application.

B. About Page: This pages provides a brief introduction on the high school dropout data that I was able to locate during my research for this project. From reviewing different state pages and locating the data that is collected for multiple counties and public schools, every state and county collect different types of information and variables about the cohorts or the students that were enrolled in high school and their completion or dropout status'. Very few states/counties collect additional information such as economic status', disability flags, LEP flags, disadvantage flags, or any additional statistic that may be collected such as events that may have caused the student to dropout (i.e., incarceration/pregnancy). After reviewing the majority of the states in the Eastern Shore, I decided to go with data for

Virginia State, as it had the most variety of variables and information on high school graduates and dropouts.

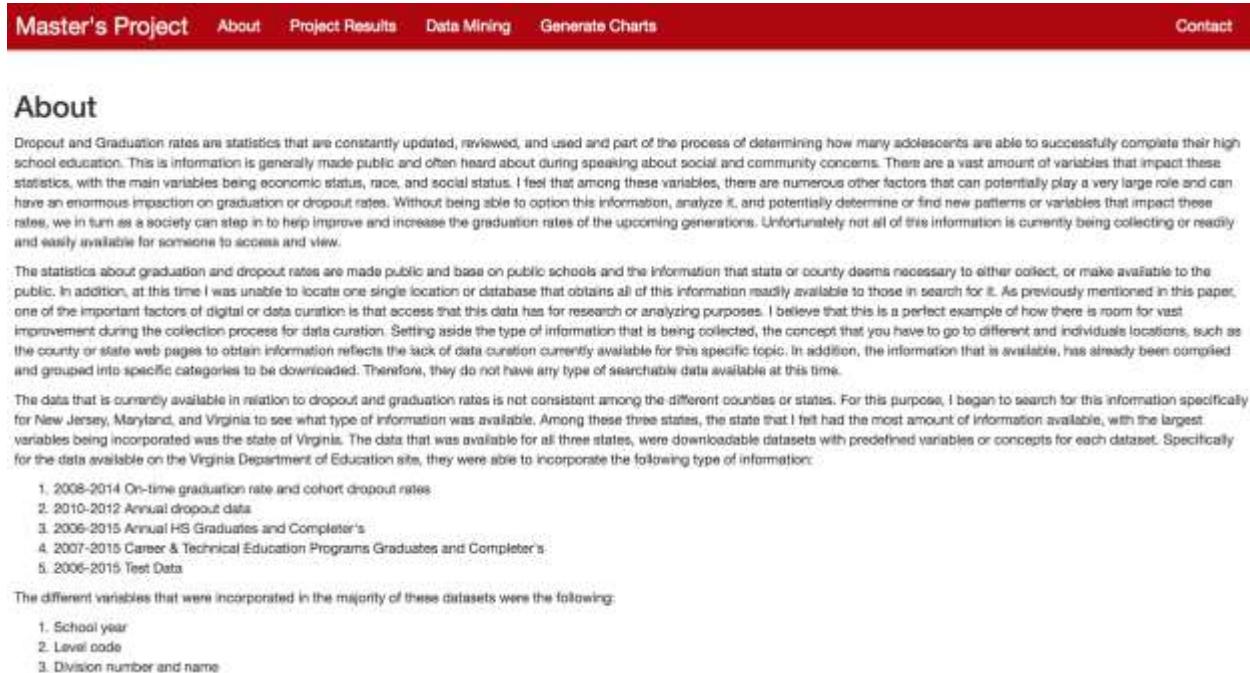


Figure 5: The About Page within the Web Application.

C. Project Results Page: The Project Results pages show the data analysis and conclusions that I generated after data cleaning and formatting the information I was able to collect from the Virginia State Education website. While reviewing the data, I decided to separate the data into the available counties, and selected one county to review in order to simplify this process. From the finalized data, I was able to generate a graph disclosing the main points and variables from this data source. In addition, I also provided a brief explanation of my observations and conclusions of the data that was analyzed.

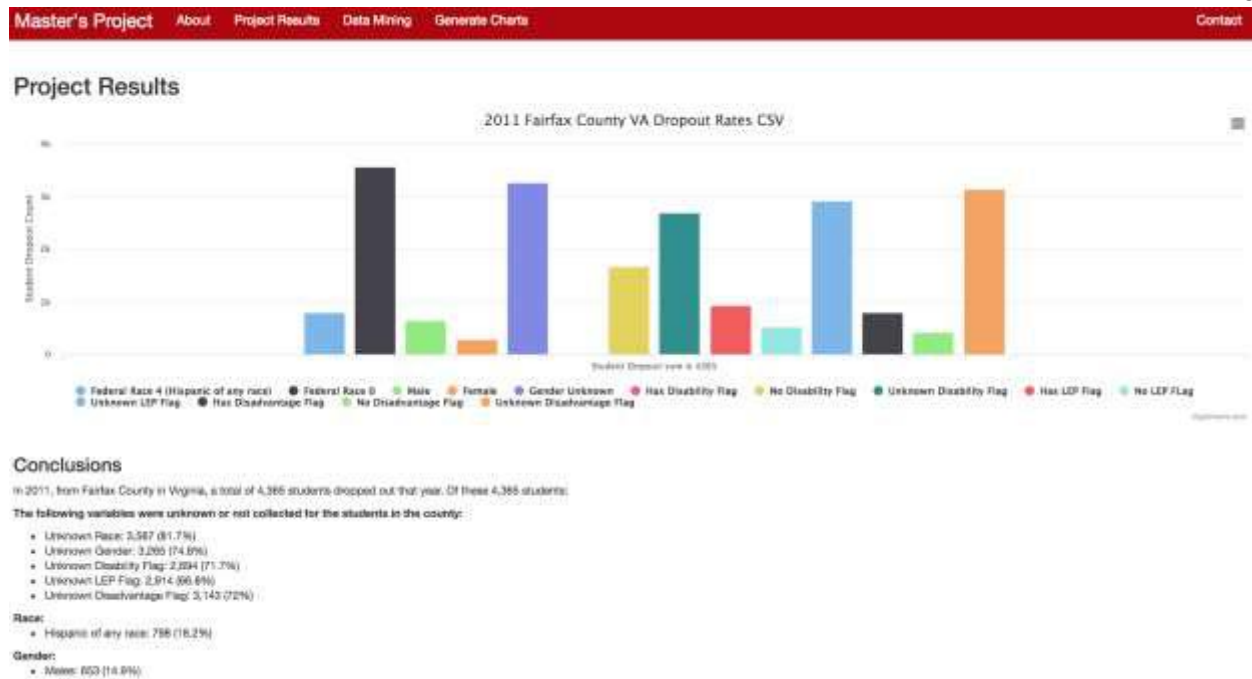


Figure 6: The Project Results page from the web application.

D. Data Mining Page: This page provides information to the user of how they can use OpenRefine, should they wish to clean, modify, transform, and export raw data, by using data mining techniques such as filtering, editing, and clustering the data to optimize its use. This tool works very well with multiple types of data formats and sizes. In the event that the user does want to use this tool, I have provided them with brief instructions on what they need to do to install the required software to use the tool. In addition, I was able to provide the user links to a tutorial and instruction videos to help explain and simplify this process.

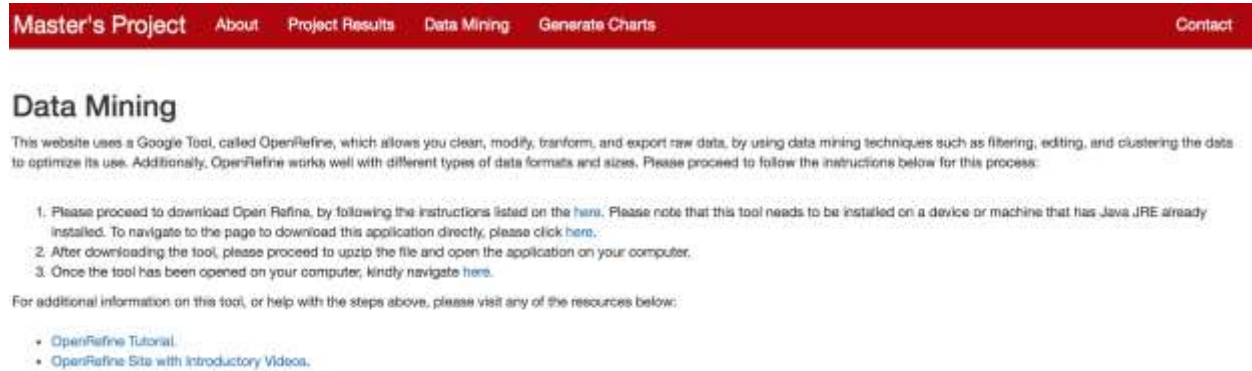


Figure 7: The Data Mining page from the web application.

- E. Generate Charts Page: The Generate Chart page on this website allows the user to upload their cvs file to generate a bar chart to help provide more useful information and draw conclusions form the uploaded data. In addition to uploading the data and generating the chart, the user is also able to manipulate the data be selecting which ledger variables to display on the chart, and finally has the option to print or download the finalized chart for their records.



Figure 8: Generate Graph page in web application.

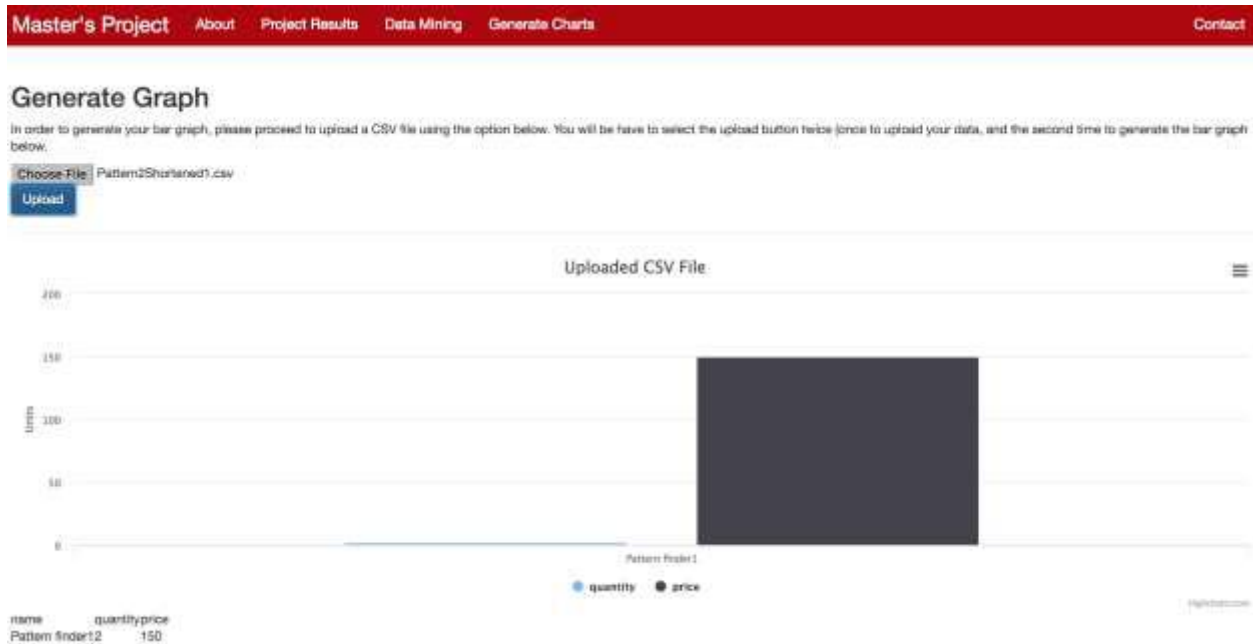


Figure 8.1: Generate Graph page in web application, with an external CSV file uploaded and chart generated.

- F. Contact Page: This page provide the option for the user to contact me with any feedback or requests for the website. Currently, completing the form from the website will pull from the current email account you have opened on your computer and generate the email format for the user. Then user, will then have to manually send the email to me, which goes directly to my personal Montclair State University account.

Master's Project [About](#) [Project Results](#) [Data Mining](#) [Generate Charts](#) [Contact](#)

Contact Me

If you have any questions, requests, or feedback in regards to this Master's Project, please feel free to email me using the form below.

Subject

Name

Email

Message

Figure 9: Contact page in web application.

3.2. Code Structure

This web application was created with the implementation of the specified platforms, API's, and resources listed below. The subsections below will cover Ruby on Rails, Bootstrap, OpenRefine, Highcharts, Github, Amazon S3, and Heroku will be explain how and why the mentioned softwares were used and implemented in the development of this web application.

3.2.1 Ruby on Rails

The Ruby on Rails platform was initially introduced in one of my web development classes. Upon further research of this platform, I learned that this was a very common and popular platform that is used for web development, both front and backend. I had very limited experience with web development, and wanted to take this opportunity to use this project to advance my knowledge with web development, and more specifically Ruby on Rails.

The foundation of this web application is created with Ruby on Rails, using Ruby version 2.3.0, and Rails version 5.0.0.1 [1]. This application was created utilizing HTML and Javascript as the coding language. To generate the required pages for this application, a combination utilizing the Ruby on Rails platform, as well as PowerShell command lines were used. During the development process of this application, I created a total of 2 versions of this project, in an attempt to determine what type of database was going to be used.

The common databases that Ruby on Rails works with are SQLite3, MySQL, and PostgreSQL [1]. “SQLite is an in-process library that implements a self-contained, serverless, zero-configuration, transactional SQL database engine and is an embedded SQL database engine that does not need to run on a specified server, and is very widely used with a large number of applications” [27]. Very similar to SQLite3, MySQL is “another open source SQL database management system, which is run on a database management system, such as MySQL Server, and is very useful for handling large amounts of data” [33]. Lastly, PostgreSQL is “an open

source object-relational database management system, which is can run on all major operating systems and is very diverse and powerful with different types of objects and data it can store and manage” [28]. The initial version of this application was created using SQLite3, as it was a much easier method to generate and implement changes within the application. After creating the foundation of the application, which allowed the user to upload a CSV file and store the contents to the database, I began to do research to determine what type of data mining techniques would be available with SQLite3. From my research, I was unable to find much information concerning the options SQLite3 had with data mining or data cleaning. Therefore, I began to expand my search to MySQL. I was able to find some information about a few data mining techniques that may be able to be implemented within MySQL through queries. In addition, I was also able to determine that MySQL is able to complete a few options, such as find and replace, removing/deleting characters, and concatenate or joining certain data rows/columns.

In my transition of migrating the database from SQLite3 to MySQL, I came across a Google tool called OpenRefine, previously known as Google Refine [7]. OpenRefine is an open source tool that has numerous of data mining/data cleaning tools already embedded and ready to use. This is a web based application, with a requirement of downloading the tool on to the machine in use, in order to utilize the tools online. After learning about this free online tool, I made the decision to use this tool within my web application, as it offered more options to clean the user's data. Using this tool, allowed me to revert back to SQLite3, as the database function for the application was no longer going to be heavily used.

3.2.2. Bootstrap

Bootstrap is an HTML, CSS, and JS framework that is an open source code which normalizes and auto adjusts web application layouts and pages, to properly display on variables screens (cell phones, tablets, laptops) with different dimensions [2]. This framework ensures that content layout remains intact and is easily readable and useable by the user. Bootstrap was also introduced through my web development course, by a student who completed a presentation of the benefits and use for this code.

The implementation of Bootstrap into the HTML code, is done by either downloading the provided CSS, Javascript, or source code and adding it to the code foundation, or directly adding the Bootstrap html to the code. The method that was used for this web application was to directly add the html scripts and code to the layouts section of this application. In this project, it was primary used to determine the layout and stylesheets of the web application, such as the navigation bar, jumbotron, text layout, and contact form style. Figure 13 below shows the required script that needs to be added to the layouts page in the code, in order to be able to utilize Bootstrap's layouts and functionalities.

```

<title>Mastersproject</title>
<%= csrf_meta_tags %>

<%= stylesheet_link_tag 'application', media: 'all', 'data-turbolinks-track': 'reload' %>
<%= javascript_include_tag 'application', 'data-turbolinks-track': 'reload' %>

<!-- Latest compiled and minified CSS -->
<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" integrity="sha384-
BVYiISIFeK1dGmJRAKycuHAHRg320mUcww7on3RYdg4Va+PmSTsz/K68vbdEjh4u" crossorigin="anonymous">

<!-- Optional theme -->
<link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap-theme.min.css" integrity="sha384-
rHyON1iRsVXV4nD0JutlnGaslCJuC7uwjduW9SVrLvRYooPp2bWYmgJQIXwL/Sp" crossorigin="anonymous">

<!-- Latest compiled and minified JavaScript -->
<script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js" integrity="sha384-
Tc5Iqib027qvyjSMfHjOMaLkfuwVxZxUPnCJA7L2mCWNIpG9mGCD8wGNICPD7Txa" crossorigin="anonymous"></script>

```

Figure 10: Bootstrap script requirements for implementation within web application.

Once the required scripts have been added to the code, the implementation of Bootstrap is added to the code through HTML formatting, in addition to using the templates provided within the Bootstrap website, for specific layouts or formatting of the web application. Figure 11 below shows how the layouts of the web application will be called upon in the main application.html layout page.

```

<!DOCTYPE html>
<html>
  <head>
    <%= render 'layouts/bootstrap' %>
    <%= stylesheet_link_tag 'application', media: 'all', 'data-turbolinks-track' => true%>
    <%= javascript_include_tag 'application', 'all', 'data-turbolinks-track' => true%>
    <%= csrf_meta_tags %>
  </head>
  <body>
    <!--<div class="jumbotron"> -->
    <!-- <h2>Montclair State University</h2>
    </div> //-->
    <%= render 'layouts/navbar' %>
    <div class="container-fluid">
      <%= yield %>
      <!-- <%= link_to "about", root_path %>!-->

      <%= render 'layouts/footer' %>
    </div>
  </body>
</html>

```

Figure 11: This image reflects how the layouts of the web application will be called upon in the main application.html layout page.

3.2.3. OpenRefine

OpenRefine previously known as Google Refine, is a free desktop application, that is run on a web server, which allows the user to clean “dirty” or “messy” data [7]. In order to utilize this tool, the user must have Java JRE installed on the machine where this tool is going to be downloaded [7]. This tool requires the user to download the OpenRefine application, open the application on the machine in use, and lastly select the option to open the browser in order to use this tool through my web application.

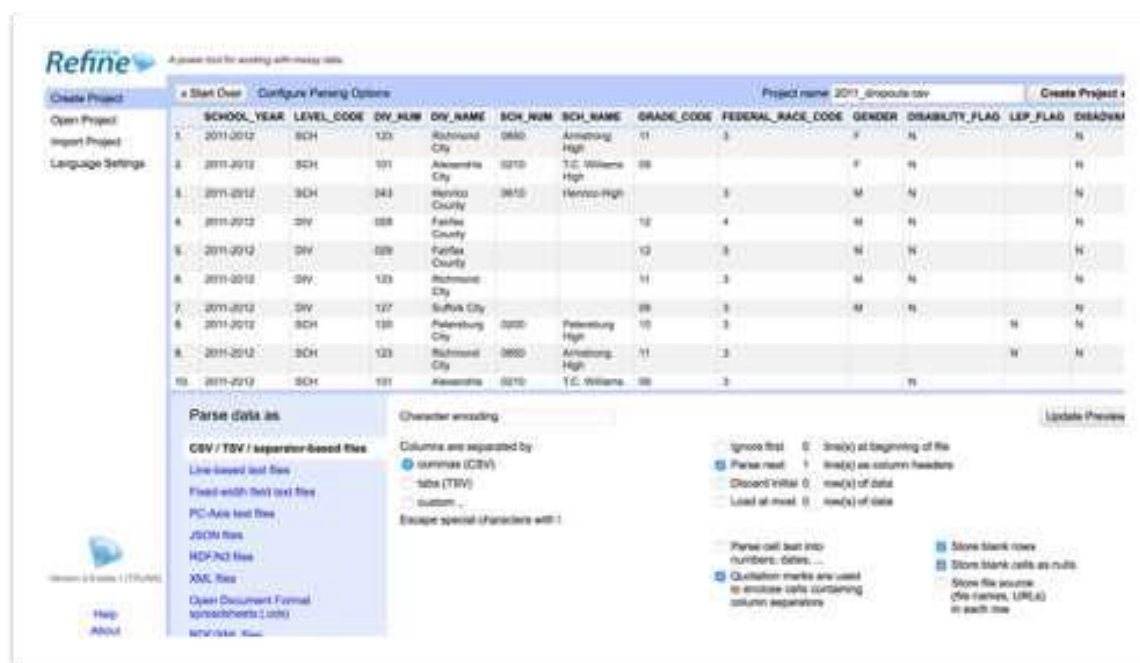
Once the browser is open, the user is prompted to upload the data that needs to be cleaned. Through this tool, the user can upload multiple data sources, such as TSV, CSV, *SV, Excel, JSON, XML, RDF as XML, and Google Data documents (Figure 12 below).



Figure 12: OpenRefine upload file option.

After uploading the selected file, the user is able to make a few quick changes to the uploaded data, by changing the project name and selecting some quick changes for the entire dataset, before moving to edit the data (Figure 13 below), such as:

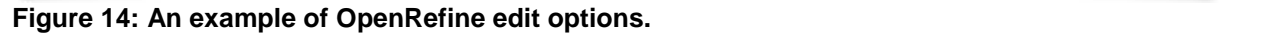
- Selecting how you would like the data parsed as (CSV, TSV, JSON files, XML file, etc)
- Character encoding
- How the columns should be separated (commas, tables, or customer options)



Figure

13: OpenRefine initial uploaded file changes/edits.

OpenRefine has multiple options available to clean and organize the data. The following options below were used to clean the data source that was selected for this project [7]. Figure 12 below shows an example of how some of these edit options can be accessed through this tool.



After completing the data mining/data cleaning for the uploaded data, you are able to save and export this data as TSV, CSV, Excel, or an HTML table. This option allows you to restart the process and continue working with your data if necessary.

3.2.4. Highcharts

Highcharts is an open sourced charting API that allows the development of flexible and interactive charts [3]. The implementations of this API was applied to the application by directly adding the javascript to the code and page of choice. In addition, Highcharts also allows the customization of the layout/style, and multiple different types of charts/graphs options based on the data and the desired outcome. Figure 15 below, reflects the standard code that was used to generate the Project Results chart using the Highchart API.

```

$(document).ready(function() {
    var options = {
        chart: {
            renderTo: 'container',
            type: 'column'
        },
        title: {
            text: '2011 Fairfax County VA Dropout Rates'
        },
        xAxis: {
            categories: []
        },
        yAxis: {
            title: {
                text: 'Student Dropout Count'
            }
        },
        series: []
    };

    $.get('data.csv', function(data) {
        // Split the lines
        var lines = data.split('\n');
        $.each(lines, function(lineNo, line) {
            var items = line.split(',');

            // header line contains categories
            if (lineNo == 0) {
                $.each(items, function(itemNo, item) {
                    if (itemNo > 0) options.xAxis.categories.push(item);
                });
            }

            // the rest of the lines contain data with their name in the first position
            else {
                var series = {
                    data: []
                };
                $.each(items, function(itemNo, item) {
                    if (itemNo == 0) {
                        series.name = item;
                    } else {
                        series.data.push(parseFloat(item));
                    }
                });
                options.series.push(series);
            }
        });

        //add suffix to title
        options.title.text = options.title.text + ' CSV';
    });
});

```


Figure 15: This image reflects the standard code that was used to generate the Project Results chart using the Highchart API.

Currently this application is only using a bar graph, for both the Project Results page, and the Generate Charts page, as it is a very simple method of reading and understanding the results from this project. In addition, the bar graph is able to pull data from a CSV file that is either uploaded to the application or included in the code. When uploading the CSV in the Generate Chart page, the application converts the first row in the CSV as the headers for the X axis, and the items in the first column are converted to the interactive variables that can be included or excluded within the generated chart. Lastly, after finalizing the chart on the web application, the user has the option to either print or download (as a PNG, JPEG, PDF, SVG file) the chart for record keeping purposes [3]. Please reference Figure 16 below.

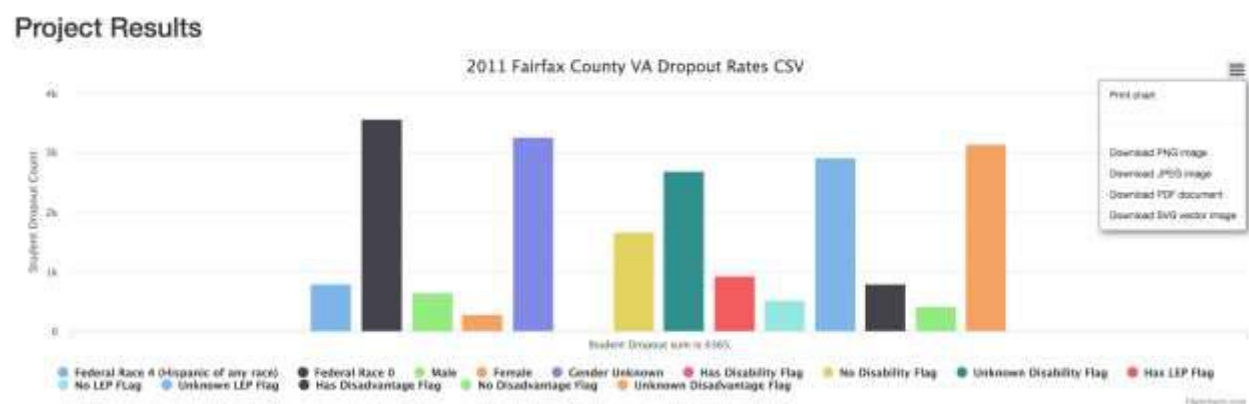


Figure 16: Highcharts view

3.2.5. Github, Amazon S3, and Heroku

In order to make this web application available publicly online, I used Github, Amazon S3, and Heroku. Github is a web based online repository service that allows you to store (private or publicly) the applications source code and its progress. This web based repository allows you to upload save points for the code, which can be used in the event that you need to revert to the last saved point in your development [5]. A Github account and new repository must be created in order to save the work onto this online repository. The following PowerShell commands must be run in the order listed below, within the folder of the selected project on the machine, in order to save this project to Github.

1. Run 'Git status': This lists all the files that have had any changes made since the last time the code was saved to Github.
2. Run "Git add .": This command will prompt Github to add all the files to the current directory.
3. Run 'Git commit -m "your message here"': This option allows the user to commit the changes to Github, and provide a customized message for this update, in order to reflect the state of the update within the repository.
4. Run 'Git push': This command pushes all of the changes to Github.

After the commands above have been successfully run, the project within Github will display the updated files with the specified message next that was provided. Kindly see Figure 17 below, which provides an example of what the uploaded project would look like within Github.

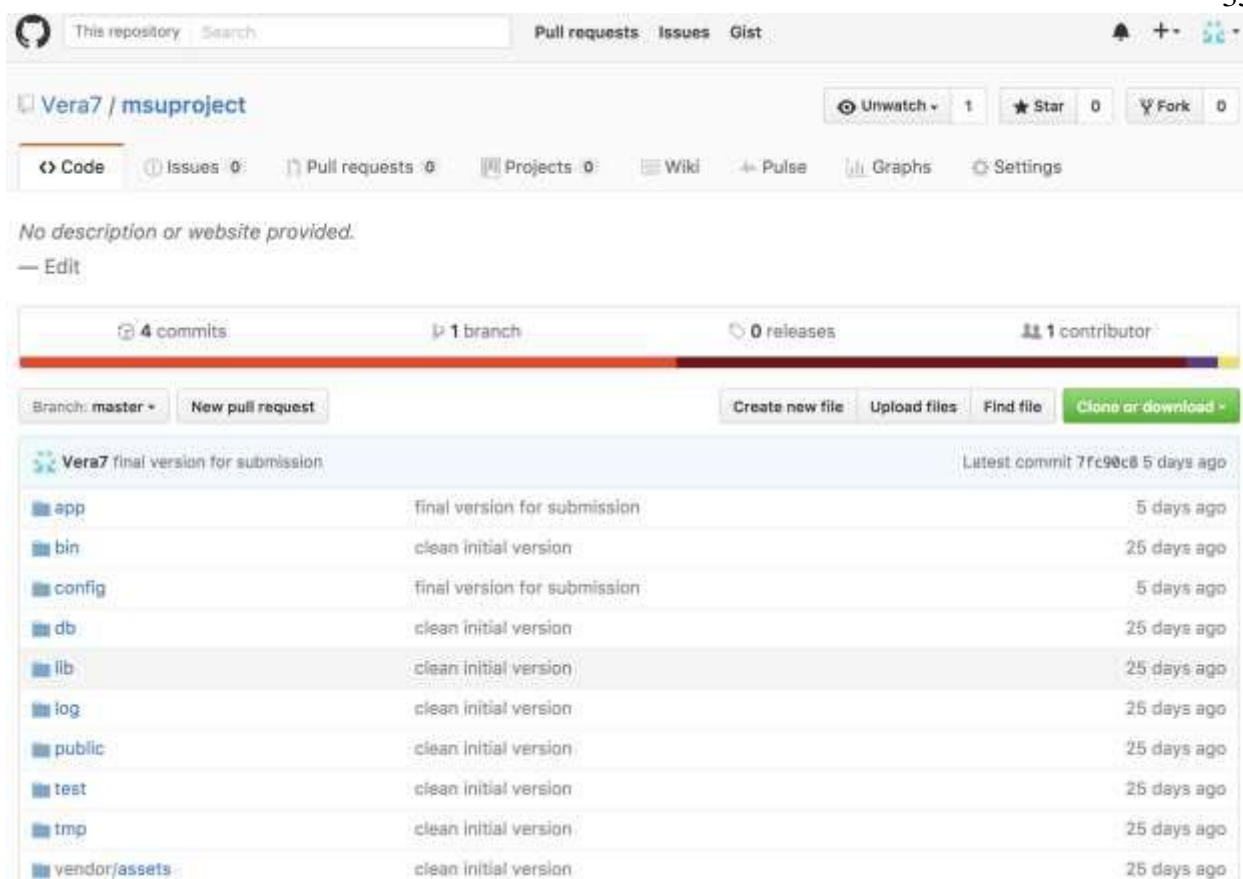


Figure 17: Github repository updated code with comments and dates.

Amazon S3 is a cloud storage device that I am using to host any necessary documentation or images that will need to be accessed by the web application [6]. At this time, I am only hosting the image that is displayed on the Master's Project page. I was able to use this service by downloading and saving the image/file that I wanted to upload to Amazon S3. After setting up the account with Amazon S3, a new bucket will have to be created for the specified project or purpose, and then the file/image can be uploaded to this specified bucket. Amazon S3 provides a link to the uploaded file that is used within the code to call the specified image, in order to be

used within the web application. Kindly see Figure 18 below, which provides an example the uploaded files to Amazon S3, and the link that is provided for the file.

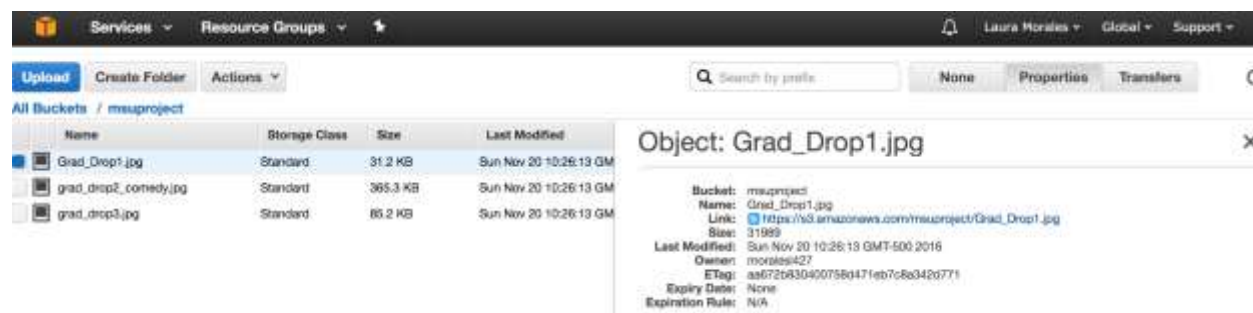


Figure 18: Amazon S3 uploaded file and file properties.

Lastly, I am using Heroku, which is a cloud platform that allows users run web applications, to make the web application available online. Heroku is able to run the web application by pulling the source code of the application from Github repository and deploying the web application online [4]. In order to complete this process, Heroku has to have already been deployed for the project in question. Once this is completed, the user can run the following PowerShell command “*Git push heroku master*”, from the project folder on the machine in use, to push the changes in Github to Heroku [4]. In addition, you may login to Heroku to view what has been deployed and if necessary, the user does have the option to revert back to previous deployments as well.

4. Project Results

4.1. Field Information

The data that was selected to complete this project, was collected from the Virginia Department of Education, under Annual Dropout Data. This data was for the year 2011 and the provided information that was collected for all the public schools in the state of Virginia. The variables below were the original columns options for the CSV being used. Please note that the descriptions for the categories listed below, were collected from the Virginia Department of Education website [32].

- “School year: States what school year that is related to the collected data.
- Level code: This column reflects whether the data is related to state level data, division level data (school districts), or school level data.
- Division number: This column contains the three digit state number that reports the student’s data to the Virginia Department of Education, disclosing that one of the following options are true about the student [32]:
 - The student lives in that division (school district)
 - If the student attends the school through open enrollment
 - If the tuition is waived for the student

- If the student is served at a school or if the student has a disability, which caused the division to be responsible for providing free education to the student.
- Div name: State the division (school district's) name
- School number: This is the four digit state number for the specific school within the division.
- School name: This is the specific name of the school within the division.
- Grade code: This states what grade the student is in:
 - 09= 9th grade
 - 10= 10th grade
 - 11=11th grade
 - 12= 12th grade
 - If the cell is blank, it implies that the grade code was not collected.
- Federal Race Code: This column identifies the students racial identification, based on how they identify themselves. These are the options below:
 - 0=unspecified
 - 1=American Indian/Alaska Native
 - 2=Asian
 - 3=Black or African/American
 - 4=Hispanic of any race
 - 5=White
 - 6=Native Hawaiian/Other Pacific Islander

- 99=Two or more races, non-Hispanic
- Gender: This field states if the students are male, female, or if the cell is blank, this means that this data was not collected.
- Disability flag: This field will specify Yes/No/cell left blank (no data collected), depicting if the student has any of the following:
 - Intellectual disability
 - Hearing impairment (including deafness, speech, or language impairment)
 - Visual impairment
 - Emotional disturbance
 - Orthopedic impairment
 - Autism
 - Traumatic brain injury
 - Development delay
 - Other health
 - Impairments
 - Learning disability
- LEP flag: This field will state Yes/No/cell left blank (no data collected) if any of the following are true:
 - Students that were not born in the United States or whose native languages are languages other than English
 - Who are a Native American or Alaska Native

- Who come from an environment where languages other than English have a significant impact on their level of language proficiency
- Who are migratory, whose native languages are languages other than English, and who come from an environment where languages other than English are dominant
- Whose difficulties in speaking, reading, writing, or understanding the English language may be sufficient to deny individuals:
 - the ability to meet the state's proficient level of achievement on state assessments described in section 1111(b)(3)
 - the ability to successfully achieve in classrooms where the language of instruction is English
 - the opportunity to participate fully in society
- Disadvantaged flag: This category reflects Yes/No/cell left blank (no data collected) identifying a student who is economically disadvantaged if the meeting one of the following below:
 - Eligible for free/reduced meals
 - Receives TANF
 - The student is eligible for Medicaid
 - The student is identified as either a migrant or is experiencing homelessness
- Dropout rate: This number reflects the amount of student who dropped out of high school and did not re-enroll back into high school” [32].

4.2. Data Processing

The original dataset that was downloaded, had 7,340 rows of data for the state of Virginia, for the year 2011 (please see Figure 19 below for an example of the raw data). To help simplify the data in order to facilitate the process of reviewing and analyzing the data, I decided to clean up and narrow down the data to one school division. The division I selected was the Fairfax County division. As I cleaned the data, I was able to remove the columns below, as the data for these columns were either not necessary because I narrowed down the data based on division, or the data was not collected.

- School year
- Level code
- Division number
- Division name
- School number
- School name

Once this was completed, I was able to consolidate multiple columns and rows to end up with a dataset that I was able to use within the web application. Please see Figure 9 below for the finalized CSV file for Fairfax County.

After reviewing the finalized data, I was able to generate a bar chart that reflected the data, based off of the finalized CSV file that was uploaded to the code. My expectations coming into this project, were related to the general stereotype of dropouts, and that a very large percentage

of the data would support that the majority of dropouts are of minority races and within the lower economic status; however the conclusions of this set of data provided slightly different results than I anticipated. In 2011, from Fairfax County in Virginia, a total of 4,365 students dropped out that year. Of these 4,365 students:

- The following variables were unknown or not collected for the students in the county:
 - Unknown Race: 3,567 (81.7%)
 - Unknown Gender: 3,265 (74.8%)
 - Unknown Disability Flag: 2,694 (71.7%)
 - Unknown LEP Flag: 2,914 (66.8%)
 - Unknown Disadvantage Flag: 3,143 (72%)
- Race:
 - Hispanic of any race: 798 (18.2%)
- Gender
 - Males: 653 (14.9%)
 - Females: 279 (6%)
- Disability Flag
 - Has Disability flag: 0
 - No Disability flag: 1671 (38.2%)
- LEP Flag:
 - Has LEP flag: 932 (21.3%)
 - No LEP flag: 519 (11.8%)

- Disadvantage Flag:
 - Has Disadvantage flag: 801 (18.3%)
 - No Disadvantage flag: 421 (9.6%)

This data supports that about 73% of the total data that is collected, was incomplete, missing, not reported, or not collected. Therefore, I am only able to review and analyze about 27% of the data collected for Fairfax County. From this 27%, the majority (50% or more) of the students who dropped out were Hispanic males, with no disability flags (meaning that they did not have any recognized or known disabilities), had LEP flags (reflecting that English may have not been their first/primary/or dominant language, possibly impacting their ability to properly communicate), and lastly that they did have a disadvantage flag (generally proposing that they were part of the lower income class) [32]. Overall, the 27% of data that had variables that can be analyzed did support my general understanding of those students that were more prone to dropout of high school. However, I do feel that if a larger part of the data had been more complete, it could either hold a stronger confirmation of this conclusion, or possibly provide a broader perspective of the diversity of students who may be prone to dropping out and any additional support variables.

It is my opinion that having this data more readily available, and easier to read and understand, it can be used to target those individuals that fall into the categories mentioned above (Hispanics, with LEP and Disadvantage flags) and provide additional support or assistance to help prevent them from dropping out. Being able to determine who may be at risk of dropping

out and providing that additional support, or assistance can help continue to lower the high

school

dropout

rate in

within

the

nation,

and

more

specific

ally

within

SCHOOL YEAR	LEVEL CODE	DIV NUM	DIV NAME	SCH NUM	SCH NAME	GRADE CODE	FEDERAL RACE CODE	GENDER	DISABILITY FLAG	LEP FLAG	DISADVANTAGED FLAG	DROPOUT CNT
2011-2012	SDH	100	Richmond City	888	Armstrong High	11	3 F	N			N	10
2011-2012	SDH	101	Alameda City	210	T.C. Williams High	8	3 F	N			N	18
2011-2012	SDH	40	Henrico County	610	Henrico High		3 M	N			N	15
2011-2012	DEV	20	Fairfax County			12	4 M	N			N	19
2011-2012	DEV	20	Fairfax County			12	5 M	N			N	10
2011-2012	DEV	120	Richmond City			11	3 M	N			N	15
2011-2012	DEV	121	Stafford City			8	3 M	N			N	15
2011-2012	SDH	120	Petersburg City	200	Petersburg High	10	3			N	N	10
2011-2012	SDH	120	Richmond City	888	Armstrong High	11	3			N	N	20
2011-2012	SDH	101	Alameda City	210	T.C. Williams High	8	3	N				11
2011-2012	SDH	20	Fairfax County	190	Woodson Adult High	10	4 M					10
2011-2012	SDH	20	Fairfax County	190	Woodson Adult High	10		N		Y	N	10
2011-2012	SDH	40	Henrico County	610	Henrico High	12		N			N	11
2011-2012	SDH	20	Fairfax County	190	Woodson Adult High	12		M			N	15
2011-2012	SDH	20	Fairfax County	190	Woodson Adult High	12		M	N			25
2011-2012	SDH	120	Petersburg City	200	Petersburg High		3	N		N	N	18
2011-2012	SDH	121	Portsmouth City	240	L.C. Meekins High		3	N			Y	15
2011-2012	SDH	120	Richmond City	730	John Marshall High		3	N			N	11
2011-2012	SDH	20	Fairfax County	190	Woodson Adult High		4 M				N	15
2011-2012	SDH	00	Franklin County	0311	Franklin County High		5 F				N	11
2011-2012	SDH	120	Richmond City	888	Armstrong High		3 F				N	14
2011-2012	SDH	121	Portsmouth City	240	L.C. Meekins High		3 M	N				18
2011-2012	SDH	120	Richmond City	730	John Marshall High		3 M	N				10
2011-2012	SDH	40	Henrico County	190	Highland Springs High		5	N			N	11
2011-2012	SDH	40	Henrico County	190	Highland Springs High		M	N			N	17
2011-2012	SDH	101	Alameda City	210	T.C. Williams High		M	N			N	15
2011-2012	SDH	120	Richmond City	888	Armstrong High		F	N			N	12
2011-2012	SDH	120	Richmond City	888	Armstrong High		F	N			N	12

this county.

Dropout Sum	Student Dropout sum is 4365
Fed Race 4 (Hispanic of any race)	798
Fed Race 0 (Unknown)	3567
Male	653
Female	279
Gender Unknown	3265
Has Disability Flag	0
No Disability Flag	1671
Unknown Disability Flag	2694
Has LEP Flag	932
No LEP Flag	519
Unknown LEP Flag	2914
Has Disadvantage Flag	801
No Disadvantage Flag	421
Unknown Disadvantage Flag	3143

Figure 19: Raw CVS data (above)

Figure 20: CSV finalized data after data mining/data cleaning (left)

5. Implementation

The main goal of the master's project was to provide a more clear and supportive conclusion of some of the main factors that can impact the high school dropout rate. The About and Project Results pages are geared specifically towards my master's project and my conclusions that were generated from reviewing the finalized data. The user is able to manipulate the finalized chart, to display the desired data based on the options within the ledger.

Should the user want to clean or generate a chart based on their own data, they may navigate to the homepage, Master's Project page, which provides a general explanation related to high school dropouts and instructions on how the user can begin this process, using the same methods from my project.

To use OpenRefine to clean raw data, the user can navigate to the Data Mining page and follow the instructions to download and operate OpenRefine tool. One requirement that is needed to use this tool, is to have Java JRE already installed on the machine of use, prior to downloading the OpenRefine tool. Once opening tool, the user will proceed to upload the desired data, use the tools within the OpenRefine to clean the data, and then proceed to download the cleaned copy of data for their records. Should the user want to generate a bar chart from this data, they will be able to navigate to the Generate Chart page, upload the CSV file (will be required to select the upload button twice at this time) and the bar will be created based off of the data that is uploaded. Once the bar chart is generated, the user can manipulate the chart by unselecting/selecting the variables listed within the ledger in the chart. Lastly, the user will be

able to print or download the chart to keep for their record. Kindly see Figure 21 below for a screenshot of how the user can view these options within the chart in the web application.

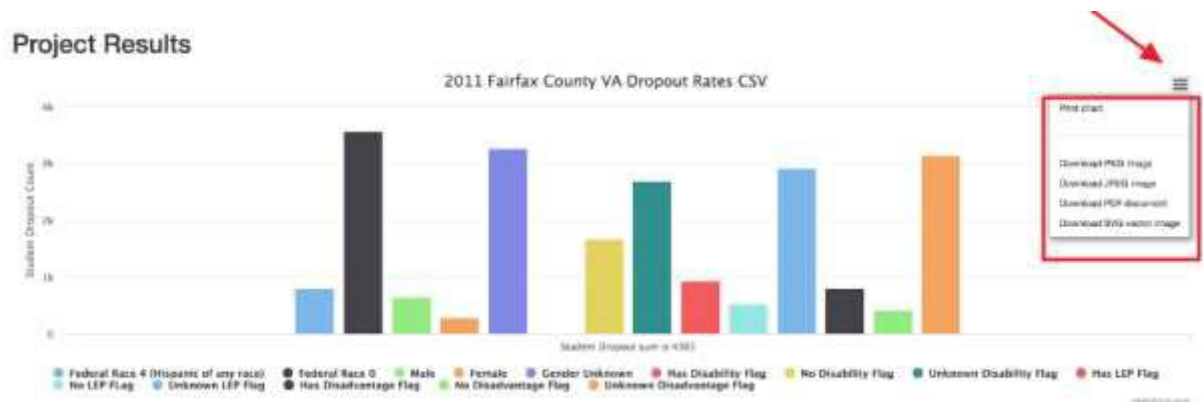


Figure 21: Highchart example of how a user can select the options to print or download the chart.

6. Future Advancements

This web application has the potential to become more innovative and advanced with the data mining options, as well as with the methods and future customizations to create charts or graphs. Below are some concepts that I would like to implement into this web application to improve its functionality for users.

6.1. Chart Options and Enhancements

Currently, this web application is only able to generate bar charts with the CSV file that is uploaded by the user. The charts have a predefined y axis labeled “Units” and the x axis is pulled from the first row of the CSV file which is considered the headers. Lastly, the variables listed in the ledger for the chart are pulled from the first column within the uploaded file. In the future, I would like to implement the following improvements specifically for generating charts within this application:

- Creating options to select variables: At this time, the user does not have the ability to dictate what variables are going to be used for either the x axis or y axis. I would like to develop a method which will generate two or three dropdown menus from the uploaded CSV file, listing the variables located within the first row of the file. The user should then be able to select the desired variables and then select the option to generate the chart with the specified selected variables.
- Generating different types of charts or graphs: Currently, this web application only allows the user to generate a bar chart from the uploaded data. I would like to provide the ability to create different types of charts/graphs, such as pie charts, area charts, and line charts, to display different types of information. The ideal implementation would be to have a button for each chart/graph option, and after the user selects their variables, they would be able to select the type of chart/graph they would want generated.

- Fixing the bug within the Generate Charts page: Currently, this is a bug within this page, which forces the user to select the “upload” button twice, in order to generate the chart and display the data below the chart. I would like to work on addressing this bug, by either potentially using parallel computing to have both sets of data displayed at the same time, or limiting the need to show the data below the graph, and only generate the chart with one push of the button.
- Processing large data: Currently, for large sets of data (around 500 rows of data or more), there is a lag with the current process that is being used to read the uploaded data and generate the chart. These large sets of data can cause the web application to run very slowly, or possibly hang when trying to process this information. I have done some preliminary research on how to address this issue, and there are a few methods that can be used to help increase the performance in these situations, that I would like to implement in the future.

Implementing these changes will help provide more customization to the way the data can be analyzed and presented with having multiple types of chart options.

6.2. Website Design and Layout Improvements

The current layout and foundation of the web application is very basic and straight to the point. It is easily understandable and users are able to navigate through the web application without much complexity. I would like to make the following improvements for this web application, to make the interface more appealing and possibly more user friendly.

- Text layout: The current layout and text format is very basic and generally. I would like to make some changes to the layout of the web application, to make it more appealing to the eye, such as changing the font type, and possibly reformatting the text layout of the application as well.
- Contact Page: The current Contact page is very simple, and once the user inputs their information and selects submit, it auto creates an email that is sent from the user's mailbox. Currently, the format of the generated email is not the cleanest and has a few extra character's within the body, that should not be there. I would like to improve this section by either cleaning up the way the email template is displayed, or by getting the from auto emailed without the need to generate it in the user's mailbox first.
- Generate Charts Page: When the user uploads the selected CSV file to this page, the data that has been uploaded is displayed below the chart, without much formatting and layout. The lack of table layout is due to how the application is reading and generating the code. I want to work on improving either how the data is being read so that way it

does not display on the page, or improve the table layout results to have a much nicer visual of the data that was uploaded.

The improvements listed above, will help provide a cleaner and more professional experience with the user of this web application.

7. Conclusion

This Master's Project has challenged me and pushed me to learn and create a web application with platform's, API's, and additional resources that I was not previously familiar with. The research that was conducted to learn more about the variables that can affect high school dropout rates, provided confirmation of some already known facts, such as the majority of high school dropouts are minorities within the lower economic class. However, more importantly disclosed that a large amount of the data that is being collected is not complete which can potentially impact our current belief and understanding of some of the variables that can help determine students that are at higher risk for dropping out. My overall goal is to use this information that can be drawn from this web application as an educational tool for professionals, to either potentially help discover new factors and locate at risk students, or help educate students of the current trends, and the available assistance and options, to help ensure that they do complete their high school education.

References

- [1] "Imagine What You Could Build If You Learned Ruby on Rails...." Ruby on Rails. N.p., n.d. Web. 09 Sept. 2016.
- [2] Mark Otto, Jacob Thornton, and Bootstrap Contributors. "Bootstrap · The World's Most Popular Mobile-first and Responsive Front-end Framework." Bootstrap · The World's Most Popular Mobile-first and Responsive Front-end Framework. N.p., n.d. Web. 09 Sept. 2016.
- [3] "Interactive JavaScript Charts for Your Webpage | Highcharts." Interactive JavaScript Charts for Your Webpage | Highcharts. N.p., n.d. Web. 12 Oct. 2016.
- [4] "Cloud Application Platform | Heroku." Cloud Application Platform. N.p., n.d. Web. 15 Oct. 2016.
- [5] "Build Software Better, Together." GitHub. N.p., n.d. Web. 20 Sept. 2016.
- [6] "Amazon Simple Storage Service (S3) - Cloud Storage." Amazon Web Services, Inc. N.p., n.d. Web. 15 Oct. 2016.
- [7] "OpenRefine." OpenRefine. N.p., n.d. Web. 01 Nov. 2016.
- [8] OpenRefine. "OpenRefine/OpenRefine." GitHub. N.p., n.d. Web. 15 Nov. 2016.
- [9] "Code." Google Code Archive - Long-term Storage for Google Code Project Hosting. N.p., n.d. Web. 15 Nov. 2016.
- [10] "OpenRefine Tutorial." OpenRefine Tutorial - Enipedia. N.p., n.d. Web. 20 Nov. 2016.
- [11] "The Fastest Growing Open Source Database." Open Source Database, Enterprise Database | MariaDB. N.p., n.d. Web. 27 Oct. 2016.

[12] "HTML." W3Schools Online Web Tutorials. N.p., n.d. Web. 12 Oct. 2016.

[13] "The Condition of Education - Elementary and Secondary Education - Student Effort, Persistence and Progress - Public High School Graduation Rates - Indicator May (2016)." The Condition of Education - Elementary and Secondary Education - Student Effort, Persistence and Progress - Public High School Graduation Rates - Indicator May (2016). N.p., n.d. Web. 01 Dec. 2016.

[14] Kealindoubleyou. "Kealindoubleyou." Stercus Accidit. N.p., 2013. Web. 01 Dec. 2016.

[15] "Online Courses - Anytime, Anywhere | Udemy." Udemy. N.p., n.d. Web. 15 Sept. 2016.

[16] Education, Virginia Department of. "Data for Researchers & Developers." VDOE:: Statistics and Reports; Data for Researchers & Developers. N.p., n.d. Web. Jan. 2016.

[17] "School Characteristics Related to High School Dropout Rates." Remedial and Special Education. N.p., n.d. Web. Jan. 2016.

[18] "Digest of Education Statistics-Digest of Education Statistics - Home." Digest of Education Statistics-Digest of Education Statistics - Home. N.p., n.d. Web. 01 Feb. 2016.

[19] "Fast Facts." Fast Facts. N.p., n.d. Web. Jan.-Feb. 2016.

[20] "Drop Out Central." Drop Out Central. N.p., n.d. Web. 01 Dec. 2016.

[21] "FRONTLINE." PBS. PBS, n.d. Web. 01 Dec. 2016.

[22] Freeman, Andrew. "High School Dropout Facts: The Numbers Don't Lie." TakePart. N.p., n.d. Web. 01 Dec. 2016.

[23] "Parental Education - Child Trends." Child Trends. N.p., n.d. Web. 10 Dec. 2016.

[24] "2015 Graduation Rates." 2015 Graduation Rates. N.p., n.d. Web. 10 Dec. 2016.

- [25] "Maryland High School Graduation Rate Hits Record High." Maryland High School Graduation Rate Hits Record High. N.p., n.d. Web. 10 Dec. 2016.
- [26] Bureau, US Census. "Data Tools and Apps." US Census Bureau. N.p., n.d. Web. 10 Dec. 2016.
- [27] "About SQLite." About SQLite. N.p., n.d. Web. 10 Dec. 2016.
- [28] "About." PostgreSQL: About. N.p., n.d. Web. 10 Dec. 2016.
- [29] "Find, Install, and Publish RubyGems." RubyGems.org | Your Community Gem Host. N.p., n.d. Web. 11 Dec. 2016.
- [30] "2016 Maryland Report Card." 2016 Maryland Report Card. N.p., n.d. Web. 11 Dec. 2016.
- [31] "New Jersey Department of Education." New Jersey Department of Education. N.p., n.d. Web. 10 Dec. 2016.
- [32] Education, Virginia Department of. "Virginia Department of Education." VDOE :: Virginia Department of Education Home. N.p., n.d. Web. 10 Dec. 2016.
- [33] "MySQL." MySQL. N.p., n.d. Web. 27 Oct. 2016.

Appendix I - Creation of Web Application

This web application can be created and run on all major operating systems and internet browsers. In order to create this web application with the downloaded source code, kindly proceed to follow the instructions below:

1. After downloading the zipped source code, please proceed to unzip the folder and save its contents to a specified location.
2. Once the source code has been saved, kindly proceed to open this code within a source code editor of your choice (ie Brackets, Sublime Text 2, or anything of your preference).
3. To confirm that the source code can be run through your localhost, please launch the command prompt and complete the following:
 - a. In this window, please navigate to the project folder that was saved on the machine.
 - b. In a second tab, also navigate to the project folder and start the rails server by running the “*rails s*” command. To confirm that the rails server is running properly, you should see something similar to the image below.


```

Lauras-MacBook-Pro:msupproject Vera$ ls
Gemfile      Rakefile      config        lib           test
Gemfile.lock app            config.ru     log           tmp
README.md    bin           db            public        vendor
Lauras-MacBook-Pro:msupproject Vera$ rails s
=> Booting Puma
=> Rails 5.0.0.1 application starting in development on http://localhost:3000
=> Run `rails server -h` for more startup options
Puma starting in single mode...
* Version 3.6.0 (ruby 2.3.0-p0), codename: Sleepy Sunday Serenity
* Min threads: 5, max threads: 5
* Environment: development
* Listening on tcp://localhost:3000
Use Ctrl-C to stop

```

- c. After the rails server has been initiated, you may proceed to open a internet browser tab and navigate to “<http://localhost:3000>”. Here you should see the main page for the web application.
4. In order to create this make this web application available publicly, accounts and access to Github and Heroku will need to be created and accessible for this process.
5. Github: To save this source code to Github, you will need to do the following:
 - a. Ensure that the computer in use has an SSH code associated with Github.
 - b. Login to Github and create a new repository.
 - c. In order to link this project to the new Github repository, please open the command prompt and navigate to the project folder. Once in the project folder, please run “Git init”. Then copy the
 - d. In order to finalize connecting the application to the new repository, you will need to copy and run the command line that is provided within Github that looks similar to this “*git remote add origin* <https://github.com/Vera7/msupproject.git>”. For any

questions about this process, you can navigate to Github.com where they have available user guides for this initial process.

- e. Once the new repository has been linked with the project, you may proceed to save the code and progress onto Github by doing the following:
 - i. Run 'Git status': This lists all the files that have had any changes made since the last time the code was saved to Github.
 - ii. Run "Git add .": This command will prompt Github to add all the files to the current directory.
 - iii. Run 'Git commit -m "your message here"': This option allows the user to commit the changes to Github, and provide a customized message for this update, in order to reflect the state of the update within the repository.
 - iv. Run 'Git push': This command pushes all of the changes to Github.
6. Heroku: In order to make the web application available online, Heroku needs to pull the saved code from Github. Kindly complete the following to get this setup:
 - a. Kindly make sure to initially deploy Heroku by:
 - i. Opening the command prompt, navigating the project folder location and entering "*heroku login*". Here you will be prompted to enter your credentials for your Heroku account. In order to proceed, please make sure that you have an SSH public key already generated for the machine in use, in order to push the code to Heroku. For more information on setting the SSH Key up, kindly navigate to Heroku.com for a user guide to assist with this process.

- ii. In the command prompt, please run "*heroku create*" to create the new project within Heroku.
 - iii. Kindly run "*git push heroku master*" to deploy the code within Github to Heroku.
 - b. After the project deployment has been successful, you may proceed to view the web application online by running "heroku open" in the command prompt, or by visiting your Heroku account and obtaining the URL for the project that was just deployed.
7. Using OpenRefine: In order to use this web based tool within the Web Application, the user must have Java JRE installed on the machine where this tool is going to be downloaded [7]. This tool requires the user to download the OpenRefine application, open the application on the machine in use, and lastly select the option to open the browser in order to use this tool through my web application. These instructions are provided within the Web Application "Data Mining" page.

Appendix II - Source Code

```

custom.css.scss — msuproject

1  .jumbotron {
2    position: relative;
3    width: 100%;
4    height: 100%;
5    background-size: cover;
6    overflow: hidden;
7
8    padding-top: 0px;
9    padding-bottom: 0px;
10   margin-bottom: 0px;
11   margin-top: -20px;
12   background-color: #010101;
13   text-color: #ffffff;
14   text-align: center;
15   .container-fluid {
16     margin-left: -10px;
17     padding-left: -10px;
18     position: relative;
19     color: #ffffff;
20
21   }
22 }
23
24
25
26 .navbar-default {
27   .container-fluid {
28     background-color: #000100;
29     .navbar-nav > li > a {
30       color: #ffffff;
31       font-size: 16px;
32     }
33     .navbar-brand {
34       color: white;
35       font-size: 24px;
36     }
37   }
38 }
39
40
41 html {
42   position: relative;
43   min-height: 100%;
44 }
45
46 body {
47   padding-top: 0px;
48   padding-bottom: 50px;
49   font-family: "Helvetica Neue",Helvetica,Arial,sans-serif;
50   font-size: 14px;
51   line-height: 1.42857143;

```

```

50 font-size: 14px;
51 line-height: 1.42857143;
52 }
53
54
55 .footer {
56 position: absolute;
57 padding-top: 10px;
58 bottom: 0;
59 width: 100%;
60 height: 40px;
61 text-align: center;
62 overflow: auto;
63 }
64
65
66

```

_bootstrap.html.erb -- msuproject

```

1 <title>Mastersproject</title>
2 <%= csrf_meta_tags %>
3
4 <%= stylesheet_link_tag 'application', media: 'all', 'data-turbolinks-track': 'reload' %>
5 <%= javascript_include_tag 'application', 'data-turbolinks-track': 'reload' %>
6
7 <!-- Latest compiled and minified CSS -->
8 <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap.min.css" integrity="sha384-
  BVYi15iFeK1dGnJRAkycuHAHRg32DmUcw7on3RYdg4Va+PmSTz/K68vbdEjh4u" crossorigin="anonymous">
9
10 <!-- Optional theme -->
11 <link rel="stylesheet" href="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/css/bootstrap-theme.min.css" integrity="sha384-
  rhty0N11RsVXV4n0JutlnGasLC3u7u7jduW9SVrLvRYooPp26WYgmgJQIXwL/Sp" crossorigin="anonymous">
12
13 <!-- Latest compiled and minified JavaScript -->
14 <script src="https://maxcdn.bootstrapcdn.com/bootstrap/3.3.7/js/bootstrap.min.js" integrity="sha384-
  Tc5IQ1b827qvyjSMHj0MaLkfuWVxZxUPnC3A712NtG9nGCD8wGNIcPD7Txa" crossorigin="anonymous"></script>
15

```

_footer.html.erb -- msuproject

```

1 <footer class="footer">
2   <div class="container">
3     <p class="text-muted">By Laura Morales</p>
4   </div>
5 </footer>
6
7
8

```

_navbar.html.erb -- msupproject

```

1 <nav class="navbar navbar-default">
2   <div class="container-fluid">
3     <!-- Brand and toggle get grouped for better mobile display -->
4     <div class="navbar-header">
5       <button type="button" class="navbar-toggle collapsed" data-toggle="collapse" data-target="#bs-example-navbar-collapse-1" aria-expanded="false">
6         <span class="sr-only">Toggle navigation</span>
7         <span class="icon-bar"></span>
8         <span class="icon-bar"></span>
9         <span class="icon-bar"></span>
10      </button>
11      <a class="navbar-brand" href="/pages/index">Master's Project</a>
12    </div>
13
14    <!-- Collect the nav links, forms, and other content for toggling -->
15    <div class="collapse navbar-collapse" id="bs-example-navbar-collapse-1">
16      <ul class="nav navbar-nav">
17        <li><a href="/pages/about">About</a> <span class="sr-only">(current)</span></li>
18        <li><a href="/pages/projectResults">Project Results</a></li>
19        <li><a href="/pages/data">Data Mining</a></li>
20        <li><a href="/pages/chart">Generate Charts</a></li>
21      </ul>
22      <ul>
23      </ul>
24      <ul class="nav navbar-nav navbar-right">
25        <li><a href="/pages/contact">Contact</a></li>
26      </ul>
27    </div>
28  </div><!-- /.navbar-collapse -->
29 </div><!-- /.container-fluid -->
30 </nav>

```

application.html.erb -- msupproject

```

1 <!DOCTYPE html>
2 <html>
3   <head>
4     <%= render 'layouts/bootstrap' %>
5     <%= stylesheet_link_tag 'application', media: 'all', 'data-turbolinks-track' => true %>
6     <%= javascript_include_tag 'application', 'all', 'data-turbolinks-track' => true %>
7     <%= csrf_meta_tags %>
8   </head>
9
10
11   <body>
12     <!--<div class="jumbotron">-->
13     <!-- <h2>Montclair State University</h2>
14     </div> -->
15     <%= render 'layouts/navbar' %>
16     <div class="container-fluid">
17       <%= yield %>
18       <!-- <%= link_to "about", root_path %> -->
19
20
21     <%= render 'layouts/footer' %>
22   </div>
23 </body>
24 </html>
25

```

```

1 <!DOCTYPE html>
2 <html>
3 <head>
4   <h2>About</h2>
5
6 </head>
7 <body>
8
9   <p>
10     Dropout and Graduation rates are statistics that are constantly updated, reviewed, and used as part of the process of determining how
    many adolescents are able to successfully complete their high school education. This information is generally made public and often
    heard about during speaking about social and community concerns. There are a vast amount of variables that impact these statistics, with
    the main variables being economic status, race, and social status. I feel that among these variables, there are numerous other factors
    that can potentially play a very large role and can have an enormous impact on graduation or dropout rates. Without being able to
    obtain this information, analyze it, and potentially determine or find new patterns or variables that impact these rates, we in turn as a
    society can step in to help improve and increase the graduation rates of the upcoming generations. Unfortunately not all of this
    information is currently being collected or readily and easily available for someone to access and view.
11   </p>
12
13   <p>
14     The statistics about graduation and dropout rates are made public and based on public schools and the information that state or county
    deems necessary to either collect, or make available to the public. In addition, at this time I was unable to locate one single location
    or database that obtains all of this information readily available to those in search for it. As previously mentioned in this paper, one
    of the important factors of digital or data curation is that access that this data has for research or analyzing purposes. I believe that
    this is a perfect example of how there is room for vast improvement during the collection process for data curation. Setting aside the
    type of information that is being collected, the concept that you have to go to different and individuals locations, such as the county or
    state web pages to obtain information reflects the lack of data curation currently available for this specific topic. In addition, the
    information that is available, has already been compiled and grouped into specific categories to be downloaded. Therefore, they do not
    have any type of searchable data available at this time.
17   </p>
18
19   <p>
20     The data that is currently available in relation to dropout and graduation rates is not consistent among the different counties or states.
    For this purpose, I began to search for this information specifically for New Jersey, Maryland, and Virginia to see what type of
    information was available. Among these three states, the state that I felt had the most amount of information available, with the largest
    variables being incorporated was the state of Virginia. The data that was available for all three states, were downloadable datasets with
    predefined variables or concepts for each dataset. Specifically for the data available on the Virginia Department of Education site, they
    were able to incorporate the following type of information:
22   </p>
23
24   <p>
25     <ol>
26       <li>2008-2014 On-time graduation rate and cohort dropout rates</li>
27       <li>2010-2012 Annual dropout data</li>
28       <li>2006-2015 Annual HS Graduates and Completer's</li>
29       <li>2007-2015 Career & Technical Education Programs Graduates and Completer's</li>
30       <li>2006-2015 Test Data</li>
31     </ol>
    </p>

```

```

32
33
34
35     The different variables that were incorporated in the majority of these datasets were the following:
36
37
38
39
40     <ol>
41         <li>School year</li>
42         <li>Level code</li>
43         <li>Division number and name</li>
44         <li>School number and name </li>
45         <li>Federal Race Code</li>
46         <li>Gender</li>
47         <li>Disability flag</li>
48         <li>LEP flag</li>
49         <li>Disadvantaged flag</li>
50         <li>Cohort count</li>
51         <li>Diploma rate</li>
52         <li>Dropout rate</li>
53     </ol>
54
55
56
57
58
59     Having these different types of variables available for analyzation can potentially help narrow down other important factors that can
60     affect graduation rates and dropout rates. Another factor that I noticed when reviewing the different types of information for the three
61     states, is the lack of consistency among all three states. All three states did have the most basic information available, such as gender,
62     race, dropout and graduation rates, however having limited variables to analyze, will by default limit the outcome or possible conclusions
63     that are available for this research.
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```



```

1 <!DOCTYPE html>
2 <html>
3 <head>
4 <head>
5 <h2>Generate Graph</h2>
6 </head>
30
31 </head>
32
33 <body>
34 <script src="https://code.highcharts.com/highcharts.js"></script>
35 <script src="https://code.highcharts.com/modules/data.js"></script>
36 <script src="https://code.highcharts.com/modules/exporting.js"></script>
37
38 <script type="text/javascript" src="http://ajax.googleapis.com/ajax/libs/jquery/1.8.3/jquery.min.js"></script>
39 <script type="text/javascript">
40   $(function () {
41     $("#upload").bind("click", function () {
42
43
44
45       Highcharts.chart('container', {
46         data: {
47           table: 'CSV'
48         },
49         chart: {
50           type: 'column'
51         },
52         title: {
53           text: 'Uploaded CSV File'
54         },
55         yAxis: {
56           allowDecimals: false,
57           title: {
58             text: 'Units'
59           }
60         },
61         tooltip: {
62           formatter: function () {
63             return '<b>' + this.series.name + '</b><br>' +
64               this.point.y + ' ' + this.point.name.toLowerCase();
65           }
66         }
67       });
68
69       var regex = /^[a-zA-Z0-9\s\_\.~:;]+(csv|txt)$/; // this will accept csv and txt files
70       if (regex.test($("#fileUpload").val().toLowerCase())) {
71         if (typeof FileReader != "undefined") {
72           var reader = new FileReader();
73           reader.onload = function (e) {
74             var table = $("#table");

```

```

73     reader.onload = function (e) {
74         var table = $('<table/>');
75         var rows = e.target.result.split("\n");
76         for (var i = 0; i < rows.length; i++) {
77             var row = $('<tr/>');
78             var cells = rows[i].split(",");
79             for (var j = 0; j < cells.length; j++) {
80                 var cell = $('<td/>');
81                 cell.html(cells[j]);
82                 row.append(cell);
83             }
84             table.append(row);
85         }
86         $('#CSV').html('');
87         $('#CSV').append(table);
88     }
89     reader.readAsText($('#fileUpload')[0].files[0]);
90 } else {
91     alert("This browser does not support HTML5.");
92 }
93 } else {
94     alert("Please upload a valid CSV file. If this CSV file has been edited previously, kindly attempt re-downloading it again
    and try uploading it again.");
95 }
96
97
98
99     });
100
101 </script>
102
103 <!--Attempt to generate multiple graphs-->
104
105 <p>In order to generate your bar graph, please proceed to upload a CSV file using the option below. You will be have to select the upload
    button twice (once to upload your data, and the second time to generate the bar graph below. </p>
106
107 <input type="file" id="fileUpload" />
108 <input type="button" id="upload" class="btn btn-primary" value="Upload" />
109 <br />
110
111 <div id="container" style="min-width: 310px; height: 400px; margin: 0 auto">
112 </div>
113
114 <div class="table-responsive" id="CSV">
115     <table class="table" id="datatable">
116     </table>
117 </div>
118
119 <!-- <div id="container" style="min-width: 310px; height: 400px; margin: 0 auto"></div> //If this is kept here, the data that is uploaded
    will print above the chart, and the graph will be generated below the printed data

```

```

contact.html.erb — msuproject

contact.html.erb
30 <div class="form-group">
31   <label for="Textarea">Message</label>
32   <textarea name="Message" class="form-control" id="Textarea" rows="3"></textarea>
33 </div>
34
35 <br/>
36 <button type="submit" class="btn btn-primary">Submit</button>
37
38 <input type="hidden" name="body" />
39
40 </form>
41
42
43 <!--
44 <form method="GET" action="mailto:Morales1427@gmail.com" enctype="text/plain">
45 </form>
46 -->
47
48 <script>
49   var form = document.getElementsByTagName('form')[0];
50   form.addEventListener('submit',contact,false);
51   function contact(e) {
52     // Prevent Default Form Submission
53     e.preventDefault();
54
55     var target = e.target || e.srcElement;
56     var i = 0;
57     var message = '';
58
59     // Loop Through All Input Fields
60     for(i = 0; i < target.length; ++i) {
61       // Check to make sure it's a value
62       if(target[i].type != 'text' && target[i].type != 'textarea') {
63         // Skip to next input it doesn't match the set rules
64         continue;
65       }
66
67       // Add Input Name and values followed by a line break-- not exactly working as expected... need to research
68       message += target[i].name + " : " + target[i].value + "\r\n";
69     }
70
71     // Modify the hidden body input field that is required for the mailto: scheme... need to research
72     target.elements["body"].value = message;
73
74     // Submit the form
75     this.submit();
76   }
77 </script>
78 </body>
79 </html>

```

```

data.csv — msuproject

contact.html.erb
1 DropoutSum,Student Dropout sum is 4365
2 Federal Race 4 (Hispanic of any race),798
3 Federal Race 0,3367
4 Male,653
5 Female,279
6 Gender Unknown,3265
7 Has Disability Flag,0
8 No Disability Flag,1671
9 Unknown Disability Flag,2694
10 Has LEP Flag,932
11 No LEP Flag,519
12 Unknown LEP Flag,2914
13 Has Disadvantage Flag,801
14 No Disadvantage Flag,421
15 Unknown Disadvantage Flag,3143

```

```

data.html.erb -- msuproject

contact.html.erb
1 <!DOCTYPE html>
2 <html>
3   <h2>Data Mining</h2>
4
5 </head>
6 <body>
7
8
9   <p>
10    This website uses a Google Tool, called OpenRefine, which allows you clean, modify, transform, and export raw data, by using data mining
    techniques such as filtering, editing, and clustering the data to optimize its use. Additionally, OpenRefine works well with different types
    of data formats and sizes.
11
12    Please proceed to follow the instructions below for this process:
13
14    <br></br>
15
16    <ol>
17      <li>Please proceed to download Open Refine, by following the instructions listed on the <a href="https://github.com/OpenRefine/OpenRefine/wiki/Installation-Instructions" target="_blank">here</a>. Please note that this tool needs to be installed on a
    device or machine that has Java JRE already installed. To navigate to the page to download this application directly, please click <a href="
    https://code.google.com/archive/p/google-refine/downloads" target="_blank">here</a>.</li>
18      <li>After downloading the tool, please proceed to unzip the file and open the application on your computer.</li>
19      <li>Once the tool has been opened on your computer, kindly navigate <a href="http://127.0.0.1:3333/" target="_blank">here</a>.</li>
20    </ol>
21
22    For additional information on this tool, or help with the steps above, please visit any of the resources below:
23
24    <br></br>
25
26    <ul>
27      <li><a href="http://enipedia.tudelft.nl/wiki/Google_Refine_Tutorial" target="_blank">OpenRefine Tutorial</a>.</li>
28      <li><a href="http://openrefine.org/index.html" target="_blank">OpenRefine Site with Introductory Videos</a>.</li>
29    </ul>
30
31    </p>
32
33 </body>
34 </html>
35
36
37
38

```

```

index.html.erb — msuproject
contact.html.erb
1 <!DOCTYPE html>
2 <html>
3
4 <head>
5 </head>
6 <body>
7   <div class="jumbotron">
8     <div class="container-fluid">
9
10      <div class="row">
11        <div class="col-xs-12 col-sm-6 col-md-8">
12          <img src=""><%= image_tag "https://s3.amazonaws.com/msuproject/Grad_Drop1.jpg" %></img>
13        </div>
14        <div class="col-xs-6 col-md-4">
15          <p><strong>In school year 2013-14, the adjusted cohort graduation rate (ACGR) for public high schools rose to an all-time
16            high of 82 percent. This indicates that approximately 4 out of 5 students graduated with a regular high school diploma within
17            4 years of the first time they started 9th grade.</strong></p>
18
19          <footer>Public High School Graduation Rates, </a> <cite title="Source Title"><a href="http://nces.ed.
20            gov/programs/coe/indicator_coi.asp" target="_blank"> National Center for Education Statistics</a> </cite></footer>
21
22          <footer>Image was provided by </a> <cite title="Source Title"><a href="http://kealindoubleyou.tumblr.
23            com/post/58748356868/jewist-detroit-born-photographer-mark-laita" target="_blank">Kealindoubleyou</a> </cite></footer>
24        </div>
25      </div>
26    </div>
27  </div>
28  </div>
29  <h2>Graduation and Dropout Statistics</h2>
30
31  <p>
32    Education is a very important factor that can help improve and enhance numerous lives. Within our society and communities, we have great
33    number of young individuals who are dropping out of high school for various reasons. This unfortunate event tends to have a long lasting
34    impact on their lives, lowering the amount of opportunities available to them, and possibly lowering the quality of life they can have in
35    the future. In my opinion, having the statistics to how many individuals are dropping out of high school is not enough. We need to start
36    focusing on other variables related to why individuals are dropping out of high school. With this information, we can begin the process of
37    determining possible root causes of this unfavorable decision and if possible start providing guidance, help, or information to the
38    affected students or families. This process can begin here.
39  </p>
40  <p>
41    This site allows you to upload information according to your needs, such as information specific to your state/city/county/school, clean (
42    data mine) the information, and generate a graph to help summarize and correlate variables to help draw meaning conclusions. To begin, you
43    may follow the steps below:
44  </p>
45  <ol>
46    <li><a href="/pages/about" target="_blank"><strong>Data Mining</strong></a>: Using OpenRefine, you will be able to upload raw/messy
47    data to the tool and conduct various data mining/cleaning techniques to make the data more useful and easier to use. For this website,
48    the data must be exported as a CSV file from the tool, in order to generate graphs/charts section. After cleaning the data, you will
49    download the updated CSV file and may proceed to generate the graph.</li>
50    <li><a href="/pages/chart" target="_blank"><strong>Generate Chart</strong></a>: Upload the desired CSV file to generate graph, which may
51    be manipulated and downloaded for your records. </li>
52  </ol>
53  </p>
54  </body>
55  </html>

```

```

projectResults.html.erb — msuproject

contact.html.erb

1
2 <!DOCTYPE html>
3 <html>
4 <head>
5 </head>
6
7 <h2>Project Results</h2>
8
9
10 <meta http-equiv="Content-Type" content="text/html; charset=utf-8">
11
12 <script src="http://ajax.googleapis.com/ajax/libs/jquery/1.9.1/jquery.min.js"></script>
13 <script src="http://code.highcharts.com/highcharts.js"></script>
14 <script src="http://code.highcharts.com/modules/drilldown.js"></script>
15 <script src="https://code.highcharts.com/modules/exporting.js"></script>
16
17
18 <!-- Local copy of jQuery js file is
19 <script src="js/jquery.min.js"></script>
20 </head>
21 -->
22
23 <!-- Local copy of Highcharts js file is
24 <script src="js/highcharts.js"></script>
25 </head>
26 -->
27
28
29 <!-- JavaScript to initialize the chart on document ready -->
30
31 <script type="text/javascript">
32
33 $(document).ready(function() {
34
35     var options = {
36         chart: {
37             renderTo: 'container',
38             type: 'column'
39         },
40         title: {
41             text: '2011 Fairfax County VA Dropout Rates'
42         },
43         xAxis: {
44             categories: []
45         },
46         yAxis: {
47             title: {
48                 text: 'Student Dropout Count'
49             }
50         },
51         series: []

```

```

projectResults.html.erb — msuproject
contact.html.erb
52     });
53
54     $.get('data.csv', function(data) {
55         // Split the lines
56         var lines = data.split('\n');
57         $.each(lines, function(lineNo, line) {
58             var items = line.split(',');
59
60             // header line contains categories
61             if (lineNo == 0) {
62                 $.each(items, function(itemNo, item) {
63                     if (itemNo > 0) options.xAxis.categories.push(item);
64                 });
65             }
66
67             // the rest of the lines contain data with their name in the first position
68             else {
69                 var series = {
70                     data: []
71                 };
72                 $.each(items, function(itemNo, item) {
73                     if (itemNo == 0) {
74                         series.name = item;
75                     } else {
76                         series.data.push(parseFloat(item));
77                     }
78                 });
79
80                 options.series.push(series);
81             }
82         });
83
84         //add suffix to title
85         options.title.text = options.title.text + ' CSV';
86
87         var chart = new Highcharts.Chart(options);
88     });
89
90     });
91
92
93
94
95
96 </script>
97
98 </head>
99
100 <body>
101
102     <div id="container" style="width:100%; height: 400px; margin: 0 auto"></div>

```



```

100 <body>
101
102 <div id="container" style="width:100%; height: 400px; margin: 0 auto"></div>
103 <br>
104 <h3>Conclusions</h3>
105
106 <p>
107 In 2011, from Fairfax County in Virginia, a total of 4,365 students dropped out that year. Of these 4,365 students:
108 </p>
109 <p>
110 <strong>The following variables were unknown or not collected for the students in the county: </strong>
111
112 <ul>
113 <li>Unknown Race: 3,567 (81.7%)</li>
114 <li>Unknown Gender: 3,265 (74.8%)</li>
115 <li>Unknown Disability Flag: 2,694 (71.7%)</li>
116 <li>Unknown LEP Flag: 2,914 (66.8%)</li>
117 <li>Unknown Disadvantage Flag: 3,143 (72%)</li>
118 </ul>
119
120 <strong>Race: </strong>
121 <ul>
122 <li>Hispanic of any race: 798 (18.2%)</li>
123 </ul>
124 <strong>Gender: </strong>
125 <ul>
126 <li>Males: 653 (14.9%)</li>
127 <li>Females: 279 (6%)</li>
128 </ul>
129 <strong>Disability Flag: </strong>
130 <ul>
131 <li>Has Disability flag: 0</li>
132 <li>No Disability flag: 1671 (38.2%)</li>
133 </ul>
134 <strong>LEP Flag: </strong>
135 <ul>
136 <li>Has LEP flag: 932 (21.3%)</li>
137 <li>No LEP flag: 519 (11.8%)</li>
138 </ul>
139 <strong>Disadvantage Flag: </strong>
140 <ul>
141 <li>Has Disadvantage flag: 881 (18.3%)</li>
142 <li>No Disadvantage flag: 421 (9.6%)</li>
143 </ul>
144 </p>
145 <p>
146 This data supports that about 73% of the total data that is collected, was incomplete, missing, not reported, or not collected. Therefore, I
    am only able to review and analyze about 27% of the data collected for Fairfax County. From this 27%, the majority (50% or more) of the
    students who dropped out were Hispanic males, with no disability flags (meaning that they did not have any recognized or known disabilities),
    had LEP flags (reflecting that English may have not been their first/primary/or dominant language, possibly impacting their ability to
    properly communicate), and lastly that they did have a disadvantage flag (generally proposing that they were part of the lower income class).
    Overall, the 27% of data that had variables that can be analyzed did support my general understanding of those students that were more prone
    to dropout of high school. However, I do feel that if a larger part of the data had been more complete, it could either hold a stronger
    confirmation of this conclusion, or possibly provide a broader perspective of the diversity of students who may be prone to dropping out and
    any additional support variables.
147 </p>
148 <p>
149
150 It is my opinion that having this data more readily available, and easier to read and understand, it can be used to target those individuals
    that fall into the categories mentioned above (Hispanics, with LEP and Disadvantage flags) and provide additional support or assistance to
    help prevent them from dropping out. Being able to determine who may be at risk of dropping out and providing that additional support, or
    assistance can help continue to lower the high school dropout rate in within the nation, and more specifically within this county.
151
152 </p>
153
154 </body>
155 </html>

```



```

# routes.rb — msuproject

contact.html.erb

1 Rails.application.routes.draw do
2   # For details on the DSL available within this file, see http://guides.rubyonrails.org/routing.html
3
4   root to: "pages#index"
5
6   get 'pages/about'
7
8   get 'pages/contact'
9
10  get 'pages/data'
11
12  get 'pages/chart'
13
14  get 'pages/index'
15
16  get 'pages/projectResults'
17
18  end
19
20

```

```

Gemfile — msuproject

contact.html.erb

1 source 'https://rubygems.org'
2
3 ruby '2.3.0'
4
5 # Bundle edge Rails instead: gem 'rails', github: 'rails/rails'
6 gem 'rails', '~> 5.0.0', '>= 5.0.0.1'
7 # Use sqlite3 as the database for Active Record
8 gem 'sqlite3', group: :development
9 # Use Puma as the app server
10 gem 'puma', '~> 3.0'
11 # Use SCSS for stylesheets
12 gem 'sass-rails', '~> 5.0'
13 # Use Uglifier as compressor for JavaScript assets
14 gem 'uglifier', '~> 1.3.0'
15 # Use CoffeeScript for .coffee assets and views
16 gem 'coffee-rails', '~> 4.2'
17 # See https://github.com/rails/execjs#readme for more supported runtimes
18 # gem 'therubyracer', platforms: :ruby
19
20 # Use jquery as the JavaScript library
21 gem 'jquery-rails'
22 # Turbolinks makes navigating your web application faster. Read more: https://github.com/turbolinks/turbolinks
23 gem 'turbolinks', '~> 5'
24 # Build JSON APIs with ease. Read more: https://github.com/rails/jbuilder
25 gem 'jbuilder', '~> 2.5'
26 # Use Redis adapter to run Action Cable in production
27 # gem 'redis', '~> 3.0'
28 # Use ActiveModel has_secure_password
29 # gem 'bcrypt', '~> 3.1.7'
30
31 # Use Capistrano for deployment
32 # gem 'capistrano-rails', group: :development
33
34 group :development, :test do
35   # Call 'byebug' anywhere in the code to stop execution and get a debugger console
36   gem 'byebug', platform: :mri
37 end
38
39 group :development do
40   # Access an IRB console on exception pages or by using <%= console %> anywhere in the code.
41   gem 'web-console'
42   gem 'listen', '~> 3.0.5'
43   # Spring speeds up development by keeping your application running in the background. Read more: https://github.com/rails/spring
44   gem 'spring'
45   gem 'spring-watcher-listen', '~> 2.0.0'
46 end
47
48 # Windows does not include zoneinfo files, so bundle the tzinfo-data gem
49 gem 'tzinfo-data', platforms: [:mingw, :mswin, :x64_mingw, :jruby]
50

```

```
50
51 #highcharts
52 gem "highcharts-rails"
53
54 gem "rails_12factor", group: :production
55 gem "pg", group: :production
56
57 #carrierwave gems for incorporating AmazonS3
58 gem "carrierwave"
59 gem "fog-aws"
60
61 #making passwords private so they are not viewed on Git
62 gem "figaro"
63
```