**Layton Industries**

**C964 Capstone Project**

Test preparation course value: Smallville ISD

Table of Contents

[Section A. 4](#_Toc78134163)

[Letter of transmittal 4](#_Toc78134164)

[Project Recommendation 5](#_Toc78134165)

[Problem summary 5](#_Toc78134166)

[Application benefits 5](#_Toc78134167)

[Application description 5](#_Toc78134168)

[Data description 6](#_Toc78134169)

[Objective and hypothesis 7](#_Toc78134170)

[Methodology 7](#_Toc78134171)

[Funding requirements 7](#_Toc78134172)

[Stakeholder impact 8](#_Toc78134173)

[Data precautions 8](#_Toc78134174)

[Developers’ expertise 8](#_Toc78134175)

[Section B. 8](#_Toc78134176)

[Problem Statement 8](#_Toc78134177)

[Customer Summary 9](#_Toc78134178)

[Existing System Analysis 9](#_Toc78134179)

[Data 10](#_Toc78134180)

[Project Methodology 10](#_Toc78134181)

[Project Outcomes 11](#_Toc78134182)

[Implementation Plan 11](#_Toc78134183)

[Evaluation Plan 11](#_Toc78134184)

[Resources and Costs 12](#_Toc78134185)

[Timeline and Milestones 13](#_Toc78134186)

[Project Purpose 13](#_Toc78134187)

[Datasets 14](#_Toc78134188)

[Data Product Code 15](#_Toc78134189)

[Hypothesis Verification 15](#_Toc78134190)

[Effective Visualizations and Reporting 16](#_Toc78134191)

[Accuracy Analysis 16](#_Toc78134192)

[Application Testing 17](#_Toc78134193)

[Application Files 17](#_Toc78134194)

[Users Guide 18](#_Toc78134195)

[Summation of Learning Experience 18](#_Toc78134196)

[Section E. 19](#_Toc78134197)

# Section A.

## Letter of transmittal

July 5, 2021

Jonathan Kent, Test Prep Initiative Team Lead, Smallville ISD

101 Hickory Lane

Smallville, Kansas 66605

Dear Mr. Kent,

The budget of school systems has always been at the forefront of every district. Deciding where best to use the taxpayer's funds for maximum efficiency is a critical process. Conveying the value of the projects to stakeholders has always been challenging. With an ever-changing environment and massive amounts of data to sort through, correlations become cloudy and difficult to ascertain.

What is needed is a way to illustrate the value of a project to stakeholders in a way that is easy to understand. The ability to use up-to-date data and metrics that correlate to the demographics of your district.

We at Layton Industries feel we have a tool that can benefit Smallville ISD and help you increase the efficiency of where your budget is spent. It offers a readout that is both graphical and quantitative. For the initial offering, we have prepared an application that could help you with your current test preparation course initiative.

Our application overview:

* Browser-based application using the Binder platform that takes in Smallville ISD test score data, as well as relevant demographic data. It then creates a visual as well as a numeric display of the projections for the selected demographic.

With his project, we hope to help alleviate a current challenge that your team is facing, as well as lay the groundwork for future partnerships. The application we are submitting can be expanded to other areas of data. Our team is currently expanding our knowledge in this growing field. As we progress the ability of this application to help you influence stakeholders in future initiatives will only grow as well.

This initial stage of the application would require a minimal investment in deference to the limited nature of the preliminary offering. With the open-source nature of the software, this is running on costs would be limited to the developer hours as well as projected time dedicated to tech support of the application. We project this to be $16,500.

We here at Layton Industries look forward to working with you,

Sincerely,

Greg Layton, CEO

## Project Recommendation

### Problem summary

The current issue facing project implementation in school districts is gaining funding for the projects. Taxpayers are generally reluctant to have their rates increased causing school districts to capitalize on the funds they already have secured. When new initiatives are proposed they have to be able to make their case to the board to gain funds. The workload needed to create successful presentations is substantial. Most districts do not have a dedicated implementation team, and the members are drawn from the current staff. The ability to pour through the data and project outcomes becomes a daunting task in addition to their normal workload. Often these projections are made too far ahead of time due to the time constraints of the team members. This data then struggles to convey an accurate measure to the board.

Our application will allow for easier visualizations to be created, as well as project outcomes in real-time as data is entered. Using the historical data as models, this browser-based application can project changes to test scores based on the demographics of the individual school. This application will allow for updated information to be added to the database as long as the data remains in the current format.

We will not currently support any methods of data transformation to automatically convert future data to the current format if it were to change. The application runs in a virtual environment separate from any operating system requirements. As such there will be no separate support for mobile, Mac OS, or Windows. The application will be supported through the major browsers Chrome, Safari, Firefox, and Edge.

### Application benefits

Our application streamlines the data analysis portion of the project by automating some tasks for the users. After loading the data source into the application, the program will allow the user to select the demographic of the school in question and see the project changes to the score if a test preparation course was offered and completed. There are already several graphs included in this application that will also automatically adjust to the data set used (it will adjust as the data is updated). The projection model will expedite a task that would historically be labor-intensive.

With this, the implementation team can focus on the presentation itself instead of data analysis and visualization.

### Application description

The application will be created using the current Python programming language (Python 3) as well as the traditional Markup language used to display non-code-based information. Python is largely accepted for its data analytics and ease of use. It is these reasons that have made it the preferred choice for the scientific community. The application will be deployed in a Jupyter Notebook, a locally hosted server that acts as a virtual machine to allow for the running of code. This application would allow Smallville ISD to automatically feed in the testing data of the district and see the projected change to student scores. This would allow for immediate calculation of the projected changes with the input of new test data, creating visuals as well as reports of the data. The application also allows narrowing the fields applied to the projection. As with all larger districts, the demographics of each school can vary widely. Our application can accept the demographics of a school and adjust the projection of the score change based on its specific population's makeup.

### Data description

The data will be stored as a .csv file (comma-separated value) list that is formatted with commas to facilitate the automatic loading into the application. An example of the data format is below. The columns hold data on the student's demographic as well as their test scores in three tested fields.

To prevent any bias the race/ethnicity data will be altered to provide only grouping and not specify the groups directly. Though not used by the application in its first iteration the student's lunch type and gender will be included in the data as well. For security, all names will be removed from the data set as well. The complete data spread including min, max, standard deviation, and mean will be displayed in the application using the full data set.

The projection model uses the test scores across all three areas as well as a calculated cumulative score as the dependent variable, and the completion of the test course as the independent variable. The projection only compares students of similar demographics when making its projection, i.e. only students whose highest parental level of education matches are used to compare test scores with regard to completing a test course or not.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| gender | race/ethnicity | parental level of education | Lunch | test preparation course | math score | reading score | writing score |
| female | group B | bachelor's degree | standard | none | 72 | 72 | 74 |
| female | group C | some college | standard | completed | 69 | 90 | 88 |
| female | group B | master's degree | standard | none | 90 | 95 | 93 |
| male | group A | associate's degree | free/reduced | none | 47 | 57 | 44 |
| male | group C | some college | standard | none | 76 | 78 | 75 |
| female | group B | associate's degree | standard | none | 71 | 83 | 78 |
| female | group B | some college | standard | completed | 88 | 95 | 92 |
| male | group B | some college | free/reduced | none | 40 | 43 | 39 |
| male | group D | high school | free/reduced | completed | 64 | 64 | 67 |

### Objective and hypothesis

This application is designed to expedite data analysis and visualization for the implementation team as well as provide more concise and accurate projections.

The hypothesis is that students whose parents have a higher level of education would have a smaller amount of change to their final test scores (dependent value) based on whether they completed a test prep course (independent value). The value of this increase would be represented by the increase in scores over similar students who did not take a test prep course.

For accuracy, only students with similar access outside of a test prep class to mentorship will be compared. This is represented by the use of a filter to compare only students whose parents have a similar level of education.

### Methodology

This project will be created using the Agile methodology, in particular the Scrum variation of Agile. The ability to iterate on the application through feedback is paramount to the success of this project. An overview of this workflow is presented below.

* The requirements are generally defined as the focus is more on iterable products than paperwork and rigidly defined documents
* The development will consist of a series of events that will be repeated until the product has reached the appropriate level of acceptance from the client.
  + During each of these cycles, “Sprints”, the development team will work towards that sprint's goals. At the end of each sprint, a product will be presented to the customer for feedback.
  + The feedback from this will be used to direct the focus of the next Sprint.
  + Continuous testing will be possible during each stage of development through these Sprints using black, white, and grey-box testing.
* After the final iteration meets acceptance from the client delivery of the application will begin. This consists of turning over the application and access to the hosted environment that runs the server.

### Funding requirements

This initial stage of the application would require a minimal investment in deference to the limited nature of the preliminary offering. With the open-source nature of the software this is running on, costs would be limited to the developer hours as well as projected time dedicated to tech support of the application.

|  |  |  |
| --- | --- | --- |
| Labor type | Hours allotted | Cost |
| Developer/Programmer | 100(@$125) | $12,500 |
| Tech Support | 25(@$40) | $1,000 |
| Data Anyalist | 30(@100) | $3,000 |
| Total | 180 | $16,500 |

### Stakeholder impact

This application affects several different stakeholders, both those who have invested in the school district monetarily as well as the students and their parents who can benefit from the enhanced efficacy of future projects.

The school board will have a new method to help decide the best place to allocate funds most effectively. The implementation team will have a valuable tool to help reduce their preparation work as well as create accurate visuals. The taxpayers will be able to see the projected benefits of proposed projects and can be assured that their money is being spent wisely. Students can benefit from more effective projects that help produce greater results furthering their education.

### Data precautions

Student data is protected (FERPA) and as such must be handled correctly. The school district will be responsible for preparing the data for outside use, such as with our application. At no time will Layton Industries have access to the raw data files. The remaining data contains sensitive data and that will be protected by limiting access to the application. This will be implemented by users having to authenticate to the application using a secure Username and Password.

### Developers’ expertise

Our team consists of A junior-level programmer, with 1 year of experience in application development using Python. A senior developer with several successful projects under their belt and current CompTIA Project+ certification. Also, a data analyst comes to us from a university research lab where they worked on several published papers.

The general design and interface will be designed by the senior developer and coded by the junior programmer. The analyst will be consulted when the visualizations are designed as well as to ensure data validity with regards to the projection model. Tech support will be provided by our in-house team and available throughout the agreed time for this project.

# Section B.

## Problem Statement

The current issue facing project implementation in school districts is gaining funding for the projects. Taxpayers are generally reluctant to have their rates increased causing school districts to capitalize on the funds they already have secured. When new initiatives are proposed they have to be able to make their case to the board to gain funds. Implementation teams are tasked with presenting accurate and easily understood data products including visualizations. Accurate and timely presentation of expense to value data is critical for the project to gain approval.

The current approach for this is for the implementation team to create visuals and analyze all pertinent data. This is time-consuming and often done too far ahead of time for the data to have the same impact. This way also does not allow for any adjustment to the projections. If a stakeholder wanted to see the benefits to another demographic than the one the implementation team created, they would be at a disadvantage to produce one.

Our application will allow for easier visualizations to be created, as well as project outcomes in real-time as data is entered. Using the historical data as models, this browser-based application can project changes to test scores based on the demographics of the individual schools. Using easy to create .csv files the database can be updated and all functionalities will be retained. Graphics and projections will accurately reflect the updated data set as soon as the kernel is running. The projection model has a dropdown function that allows for real-time adjustment to the demographic used to predicts scores. As this is a browser-based solution no special hardware or downloads are required. A benefit when presenting offsite where the available equipment may not be known beforehand.

Using Python as its underlying base and taking advantage of the Pandas, Numpy, and Pyplot libraries we will be able to use data analysis standards and methods to accurately derive useful information from the data set. To further the value, we will make use of the Seaborn and ipywidgets libraries to increase the visualizations available with our product. This will increase the available graphic types available from the basic plots to more advanced visuals where suitable.

## Customer Summary

This application is designed primarily to be used by the members of the implementation team. They would be able to use it to help create a presentation more easily with higher confidence in the data they want to present. However, the application could also benefit others who wish to see the correlation of data or the projections. Such as a board member who wanted to see the data projections for a specific campus demographic. The majority of the application is running in the background so the impact of inexperienced users can be minimized. The drop-down box used for the selection for the projection model prevents invalid data from being entered.

## Existing System Analysis

The solution we are offering runs in a browser with no needed downloads or special software. As such the only requirements would be functioning networks to connect to the servers. Smallville ISD currently has a dedicated fiber network connecting all campuses with a mesh wireless network also available. After configuring the firewall to allow traffic to the server the only obstacle with the existing system would be internet or other network outages.

## Data

The data we are going to use will be from Smallville ISD’s databases. The agreed format for this project includes the columns available as well as the order they will be parsed to our server. All sensitive data will be excluded from the data sent to Layton Industries. This data will be in the form of a .csv text file. It will include pertinent data needed for the application to predict score changes, as well as data that could be used at a later date if the current project succeeds and Smallville ISD wishes to alter the application scope.

The data set contains columns with information on the student, including:

* Gender
* Race/Ethnicity (hidden into associated groups)
* Parental Level of Education
* Lunch type
* Whether they have completed a test preparation course or not
* Scores in Math, Reading, and Writing

To prevent any data skew due to outlier’s a minimum of 1000 data entries will be required before the projection model is accurate. A sample set of data will be used to present the initial application product to Smallville ISD.

This data was collected from:

<https://www.kaggle.com/spscientist/students-performance-in-exams>

## Project Methodology

The project will be developed using an Agile development methodology. The testing and feedback loops will best serve the customer to ensure needs will be met by the application. The iterative nature will allow for features and changes to be incorporated during development.

**Requirement’s phase:**

Working with the implementation team lead we will outline the application requirements. We will need a clear image of what it is the customer wants the final product to look like, items to include, and ensure that both parties are clear on the deliverable application to be created.

**Development phase:**

Using the requirements gather we will begin the development of the application using an iterative approach. At defined points, we will conclude sprints and present the current product for review to the client. Using the feedback gained we will steer the next sprint towards creating the desired application.

Testing will be conducted during the development phase. It will be integrated with development and will use multiple testing methods. Black box testing to ensure proper data projections using defined inputs, as well as full systems testing. Unit testing will not be required as each cell acts as its program.

**Delivery:**

Once development has concluded the application will be delivered to the client for final review and acceptance. Final instruction of application will be available, and this will begin the defined tech support window.

## Project Outcomes

The project deliverables consist of the budget document, with a breakdown of all the projected costs and their sources. A sample dataset will also be included so that the application can be demonstrated to the client.

The product deliverables consist of the application itself. It will consist of a Jupyter notebook containing the developed projection model as well as custom-tailored data visualizations. This will be hosted on servers for the time being. After the initial period with included tech support has ended, the server access and maintenance will be turned over to Smallville ISD.

## Implementation Plan

The first stage of the product will consist of a rough framework that will demonstrate the core functionality of the application. This stage will use the sample data previously mentioned to illustrate the data visualizations and validate the projection model's predictions. After the presentation of this iteration to the client we will take their feedback into development and cycle through development and testing phases. At the end of each, we will present the new iteration to the client for feedback. Once the application reaches acceptance, we will then incorporate the actual data provided from Smallville ISD.

This second stage will begin the tech support of the project and ensure that the application functions to agreed standards. All authorized users of Smallville ISD (authentication file provided by client) will be given access to the notebook. This is a browser-based application so there is no need for specific hardware rollouts. The application will update on every instance as the client updates the data set that the information is pulled from.

## Evaluation Plan

This product is to be evaluated using sample data before live deployment. Known variables will be compared with outputs from the application to ensure data validity and accuracy in a black box testing environment. The application demonstrations will allow the end-users to run the application on their own devices to ensure functionality on the final systems. Application iterations will be presented to the client periodically for feedback and to ensure that acceptance goals are met. Once the user data set has been loaded into the program the same black box data comparisons will be run again to ensure accuracy across the new data set. The projection model will be verified by our data analyst using known data points and adjusted if any variances are encountered. The first stage with sample data will be completed once the client has reached acceptance of the product.

The second stage will load the client's data set into the application and begin our tech support stage of the project. This will be completed after the time frame agreed upon by both parties. Graphic functionality will be tested to ensure proper output and correct information is displayed they will be compared against requirements collected in the planning stages.

## Resources and Costs

* Programming Environment: All used software is open source and free to use. The computers needed to run the application as well develop it will be counted among the necessary for business hardware already in use by both Layton Industries and Smallville ISD. As such the total costs for the programming environment is:
  + $0.00
* Environment Costs: there are no special hardware requirements as the application is browser-based. All used hardware is considered items already deemed necessary for business and in use by both Layton Industries and Smallville ISD. As such the total costs for the Hardware environment is:
  + $0.00
* Human Resources Requirements: The entirety of the funds for this project reside with the labor needed to create the application and support it after deployment. The costs are broken down by category below.

|  |  |  |
| --- | --- | --- |
| Labor type | Hours allotted | Cost |
| Developer/Programmer | 100(@$125) | $12,500 |
| Tech Support | 25(@$40) | $1,000 |
| Data Anyalist | 30(@$100) | $3,000 |
| Total | 180 | $16,500 |

## Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Start Date | End Date | Acceptance Criteria |
|  | Stage 1 |  |  |
| Requirements Gathering | 1/4/2021 | 1/18/2021 | Requirements agreed upon |
| Initial application (demo) | 1/18/2021 | 2/8/2021 | Initial application demo presented to the client |
| Second Iteration | 2/8/2021 | 3/1/2021 | Feedback collected and the next sprint scheduled |
| Data Visualizations (added) | 3/1/2021 | 3/22/2021 | Feedback collected and the next sprint scheduled |
| Final Iteration | 3/22/2021 | 4/12/2021 | Customer acceptance of the product |
|  | Stage 2 |  |  |
| Tech support for the application  Provided by Layton Industries | 4/12/2021 | 4/30/2021 | Support time frame completed |

Source: Layton Industries, Smallville ISD projected timeline

**Section D.**

## Project Purpose

The issue that was facing Smallville ISD’s project implementation team was the difficulty in creating successful presentations for gaining funding for projects. Smallville ISD focuses on capitalizing on the funds they already have secured. When the new test preparation course initiative was proposed the Implementation team had to be able to make their case to the board to gain the needed funds. The workload needed to create their successful presentation was historically substantial. The implementation team was made up of members drawn from the current staff. They had limited ability to pour through the data and project in addition to their normal workload.

We at Layton Industries created an application that allowed for easier visualizations to be created, as well as project outcomes in real-time as data is entered. It used the historical data as models, to project changes to test scores based on the demographics of the individual school. The application allowed for updated information to be added to the database to create the most up-to-date visualizations and projections possible.

This was a browser-based application to facilitate its use across the largest number of devices with minimal hardware requirements and no added software installation. Using a Jupyter notebook with Python coded data projections and visualizations this tool allowed the implementation team the ability to create a presentation in less time and with better results than they had been able to previously. The application can read in updated data and automatically adjust itself to the new data and updates the projection model as well. The application created 3 main visuals that demonstrated the various demographics represented in Smallville ISD’s test data. In addition to providing a projection model that showed the change in test scores of students with regards to the completion of a test preparation course separated by their demographics. This allowed the Implementation team to better illustrate the effectiveness of their proposed project to the school board.

## Datasets

The data was collected and formatted by Smallville ISD due to regulations regarding student data (FERPA). It was formatted as a .csv and the structure already in use by the school district was adapted by the programmers to ease coding requirements. The below is a sample of data as it was received by Layton Industries.

"gender","race/ethnicity","parental level of education","lunch","test preparation course","math score","reading score","writing score"

"female","group B","bachelor's degree","standard","none","72","72","74"

"female","group C","some college","standard","completed","69","90","88"

"female","group B","master's degree","standard","none","90","95","93"

"male","group A","associate's degree","free/reduced","none","47","57","44"

"male","group C","some college","standard","none","76","78","75"

"female","group B","associate's degree","standard","none","71","83","78"

"female","group B","some college","standard","completed","88","95","92"

"male","group B","some college","free/reduced","none","40","43","39"

The application was written in such a way as to need no further cleaning of the data. It was passed as shown straight to into the application and computations were performed on required columns as needed. The data as it appears inside the application is shown below.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Gender | race/ethnicity | parental level of education | Lunch | test preparation course | math score | reading score | writing score |
| Female | group B | bachelor's degree | standard | none | 72 | 72 | 74 |
| Female | group C | some college | standard | completed | 69 | 90 | 88 |
| Female | group B | master's degree | standard | none | 90 | 95 | 93 |
| Male | group A | associate's degree | free/reduced | none | 47 | 57 | 44 |
| Male | group C | some college | standard | none | 76 | 78 | 75 |
| Female | group B | associate's degree | standard | none | 71 | 83 | 78 |
| Female | group B | some college | standard | completed | 88 | 95 | 92 |
| Male | group B | some college | free/reduced | none | 40 | 43 | 39 |
| Male | group D | high school | free/reduced | completed | 64 | 64 | 67 |

## Data Product Code

We created the projection model using the historical data. Each student's test data was grouped with similar students with regard to demographics. Then they were further separated into whether a test preparation course was taken or not. These scores created the basis for the model.

The averages of the test scores for the student who meets the selected demographics who did not take a test preparation course was used to establish a base score. When compared with students' test score averages of the same demographic but who had completed a test preparation course we could see the projected change to test scores.

To create a clearer picture a cumulative score system was created. This score was the result of taking the three tested area's scores and combining them to average out each student's total scores. These were weighted the same as no test had a higher value for this projection model. The individual test scores are still kept allowing for visualizations of any trends regarding test prep course effectiveness with regard to the subject. This was not the focus of this project, so no conclusions were drawn at that time.

The change in scores for students of a particular demographic, their parental level of education for this project, when compared to whether the student has taken a test preparation course or not is represented with this application. This data was presented to the user with a table including the test scores of those who had and had not taken a test preparation course. This included each subject test as well as the cumulative score. To further add value a visualization of this data was created to show projected change to test scores. The application's demographic selection was facilitated by a drop-down menu system to prevent users from searching for data that is not stored in the dataset.

## Hypothesis Verification

The stated hypothesis was those students whose parental level education was of a higher level would receive less benefit (positive score change) from a test preparation course than students with a lower level of parental education. At this time our model has confirmed this hypothesis, showing an overall decrease in a positive change of test scores comparing students with a lower parental level of education vs higher levels. Students with the lowest level of parental education showed the greatest increase in test scores. Students with the highest level of parental education showed the smallest growth in scores but did still benefit from taking a test preparation course.

Further runs of the application with more datasets would help confirm the validity of this hypothesis but we feel confident with the data we have analyzed that it will trend in the same direction as our results.

## Effective Visualizations and Reporting

The first visualization was created separating the average scores of students in each of the three tested subjects. It was further broken down with the gender of the students and whether they completed a test preparation course or not. Looking at this shows that there was no major difference between the genders with regard to the change in test scores. Both genders reported similar changes due to the completion of the test prep course. As such gender was disregarded through the rest of the model and the student data was only separated by parental level of education. We can see that overall that math has the lowest average scores of the three subjects.

The second visualization was used to help clarify the breakdown of the parental level of educations segmentation. The only marginally greater than the rest was the percentage of students with parents holding a Bachelors's degree. As a whole, parents with a completed degree of post-secondary education make up the majority of the demographic. This graphic allows for a better understanding of how much of the student population would be affected by the addition of the test preparation course. This was created by separating each student's entry into a parental level of education and calculating their sum. These were then compared to the overall total of students in the dataset to get the percentage makeup of each level.

The third visualization is a box graph. Though a less common graph type this was selected to help illustrate the score ranges of each parental level of education. The boxes represent the scores that make up the 25th to 75th percentile of each level. The line inside the box is the median of all scores in that demographic. Each line extends to include the remaining scores inside a defined range. The dots seen on some levels represent outliers that could not be grouped. We chose to include these as a trend emerged with more low outliers appearing with lower levels of education. The data shows that the spread seems to grow as levels are decreased, students have a larger gap between scores. Noticeably the “Some High School” level has the largest overall spread, covering the lowest group of scores in the 0 to 25th percentile and 75th to 100th percentile reaching even beyond that of the “Masters Degree”.

## Accuracy Analysis

The representative data tracks with projections when compared to the historical data. This was a specific application created for a narrow selection of data. When expanded to include other criteria for demographics the accuracy of the application could become skewed. The projection modeling as it stands is only focused on the differences in scores based on parental level of education. In that regard, it has been shown to match up to expected outputs with no observable deviation. The projection shows an increase in a cumulative score for students in the largest demographic of “Bachelors Degree” of 5 points and that follows the historical data found in the dataset.

The visualizations help illustrate the overall makeup and the score variances. With these, we feel the application is complete in its accuracy to project change in test scores. With further expansion of the dataset, the application should improve its accuracy further tracking with the score trends.

## Application Testing

During the initial programming, several different tests were employed. Unit tests were run to verify that proper data from the .csv file that represents the database were loaded into the application. Acceptance tests were run to see the most effective visualizations for the data we wished to represent.

Known data points were loaded into the application and calculated to ensure that the basic functions required to run the application were functioning as designed. White box testing was used to verify expected outcomes from each visualization and the projection model itself. During this testing, a strange error has become apparent. At times and unpredictably the notebook will fail to recognize the code to create the widgets used for the projection model. It is a failure to correctly import the ipywidgets library and has been unable to be completely removed. The workaround is to merely restart the kernel in the notebook, but as the library import is handled through Jupyter notebooks itself there is little our team can do to improve the situation further.

The projection model was limited to inputs from a drop-down menu to in part limit the ability for the application to crash due to unknown variables being passed to the model. For this reason, the projection required minimal functionality testing to ensure its operation.

During each sprint's conclusion, the application was presented to the client for acceptance testing and feedback. Each iteration allowed us to capitalize on the client's needs and refine the finished product to meet their goals.

## Application Files

All the files for this application have been submitted along with this document. To ease use this will be loaded into Binder. Binder sets up a virtual environment with all the necessary libraries. This allows for a more seamless experience and ease of use.

Included in the linked repository:

* The Jupyter notebook in its raw form with all code,
* README document with instructions and a copy of the Users guide
* A required .yml file that has environment details required for the Binder setup
* The dataset used in this application

## Users Guide

1. Ensure you have a computer and an up-to-date browser. (Google Chrome recommended)

2. Navigate to the following webpage by either clicking the link or copying and pasting it into the address bar of your web browser: <https://github.com/GLayton-TX/C964_capstone>

3. Scroll down to view the readme. On it is a badge labeled “Launch Binder”. The link will launch the Binder environment with the Jupyter Notebook that contains the application. Please click on the “Launch Binder” badge.

4. Allow for the notebook to load as this is a VM environment with multiple dependencies it may take a few moments to complete.

5. The application is run as an interactive notebook. Even though the notebook will prepopulate with some data and visualizations, please click the “Kernel” tab, from the drop-down select “Restart & Run All”. On the pop up confirm “Restart and Run All Cells”. This will initialize the Projection model.

6. At the bottom of the notebook the projection model contains drop-down boxes with the choices for “Average Parental Education Level”, you can select any of the reported levels to see the effects of a test preparation course on that demographics score. There is a numerical table display of the scores as well as a visual representation.

7. this environment is separate from the repository hosted on GitHub. As such any changes made here are nonpermanent. Feel free to explore any code and change variables if desired. If you encounter any issues or need to restart the application. Close the tab hosting the binder page and relaunch from step 2.

## Summation of Learning Experience

The numerous courses throughout the degree program helped prepare me for this capstone project. We have completed several projects that helped me determine the best approach to creating an application. I have been exposed to many programming languages and was able to evaluate them to determine the best one to use for this. We have studied data structures and the best way to maintain our data integrity, which helped in sorting through datasets to determine the most appropriate for this use. The project management courses allowed me to see how the project could be created and implemented.

As I had no experience with creating an application that focused on data analysis and visual representations, I first went through the study materials I could find on the internet, including some tutorials from YouTube. I sought out my course instructor to help me understand the nature of the capstone and better direct my focus and energy to its completion. In addition, I relied on my friends and family for testing the application. They have varied technical experience and their feedback helped uncover potential pitfalls that would occur.

This capstone helped me improve my understanding of my tendencies and how to ensure that I can be more productive with future projects. The knowledge gained on effective searching methods as well as credible sources is possibly the most valuable. With the varied and vast amount of programming languages, methods, applications, and methodologies attempting to memorize them all would be impossible. A firm understanding of how to find the knowledge needed would be key to future success.

# Section E.

Reference

Seshapanpu, J (2018) Students Performance in Exams, Marks secured by the students in various subjects. Kaggle.

https://www.kaggle.com/spscientist/students-performance-in-exams