College GPA Statistical Analysis

(COMP3125 Individual Project)

Devon Lapierre   
*Wentworth Institute of Technology School of Computing and Data Science*

*Abstract*—This electronic document is a “live” template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *\*CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract*. (*provide a short abstract*)

Keywords—example1, example2, example3, example 4, example 5 (provide 3-5 keywords)

# Introduction (*Heading 1*)

With college education becoming such an important factor in career success, it is essential to look in depth at the leading factors that contribute to the GPA of college students. It can be intriguing to analyze what categories of college students have a significant impact on their grades and overall measured success in their college education, since this measurement plays a key role in future life and career endeavors for most individuals. Some obvious categories that fit the general criteria for determining college GPA include high school GPA, major, study hours, attendance, and extracurricular activities, but how significant are these factors and others such as part-time jobs, library usage, sleep, and more in predicting a model for a given student’s GPA? A study done at the University of Alaska [1] determined that high school GPA is the most significant predictor of college GPA, but can other factors within the dataset be just as significant or more? One interesting variable to consider would be having a part time job while being a college student, and another could dive into the different major classifications and how they could play a role in determining GPA. With all of the different categories of potential significant explanatory variables, building a predictive model of their relationships with college GPA can itself be significant in looking at career success.

# Datasets

## Kaggle.com

The source of the dataset is Kaggle, a public site with free access to thousands of datasets. The dataset that analyzed in this report is licensed by Massachusetts Institute of Technology, giving it credibility and integrity. However, minimal information on its generation and its author is provided.

## Dataset Characteristics

The dataset [2] is a comma-separated values file just under 100 kB in size. The actual file contains 10 columns of nine explanatory variables and a response variable, with 2000 rows of data. The columns and their respective data types are as follows:

|  |  |
| --- | --- |
| **Column** | **Data Type** |
| Study Hours per Week | Hours |
| Attendance Rate | Grade Scale (0-100) |
| Major | Categorical |
| High School GPA | Grade Scale (0-4 |
| Extracurricular Activities | Integer Value |
| Part-Time Job | Binary (Yes/No) |
| Library Usage per Week | Hours |
| Online Coursework Engagement | Hours |
| Sleep Hours per Night | Hours |
| College GPA | Grade Scale (0-4) |

While no conversions were made, the data was cleaned to be easily interpreted as a data frame in the python programming language. Empty rows in the Major column were changed to “Undeclared”, as in Undeclared in the dataset rather than Undeclared majors. Empty rows in all other columns were dropped, as replacing them with 0 would inaccurately skew the data, and duplicates were dropped as well, resulting in 1473 rows of data. One additional column of Part-Time Job Indicator was created, with the binary data type of zeros and ones, correlating to No and Yes, respectively, in the Part-Time Job column.

# Methodology

In this part, you should give an introduction of the methods/model. First, what’s the method/model. What’s the assumption of this method/model. What’s the advantage/disadvantage of this method/model. Why did you choose it. What Python module or function do you apply to apply this method/model. Any optional input/extra work did you adjust to make the results better. If you have multiple methods, feel free to use subsection A., B. to separate them.

Example: Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

## Method A

Example: The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Method B

* Bulletin 1
* Bulletin 2.
* Bulletin 3

Identify applicable funding agency here. If none, delete this text box.

## Method C

Example: The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

An excellent style manual for science writers is [7].

# Results

In this section, present your findings using an appropriate method, such as equations, numerical summaries, or visualizations like charts and graphs. Clearly explain all results and provide guidance on how to interpret them. If any unexpected results arise, discuss possible reasons or contributing factors. To improve clarity and organization, consider using subsections (e.g., A, B) to separate different aspects of your results.

Example: After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Result A

Example: XXX

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Results B

Example: Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

## Results C

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
| Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

# Discussion

Every method/project has its shortage or weakness. Please discuss the unsatisfied results in your project. And discuss the feasible suggestions of future work to revise/improve your result.

Example: xxx

# Conclusion

In this part, you should summarize your project. What important results did you find for your topic and what’s the effect of this result on the real-world?

Example: xxx

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

Use the IEEE format for the citation. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...” Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

1. <https://ies.ed.gov/use-work/resource-library/report/descriptive-study/how-well-does-high-school-grade-point-average-predict-college-performance-student-urbanicity-and#:~:text=Summary,directly%20in%20college%2Dlevel%20courses>.
2. <https://www.kaggle.com/datasets/petermushemi/dataset-for-predicting-the-college-gpa-of-students?select=Academic.csv>

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**