Title (should be descriptive)

# Introduction

Introduce the data you found and describe the purpose of the assignment as it relates to this data. Make sure you motivate how the required analysis will solve the problem in the data. Be sure to cite any relevant sources. Your paper should contain the following sections: Introduction, Data Description, Methodology, Results and Discussion, Conclusions, and References. The length should be between four and six pages.

The last paragraph of the introduction should describe what is contained in the future sections. This should be something like this: “The future sections of this report describe the dataset, the methodology, results along with a discussion, and a conclusion. Section II contains a description of the dataset used for this analysis. The methodology for analysis is presented in section III. In section IV, I report and discuss the results. Finally, section V provides conclusions.”

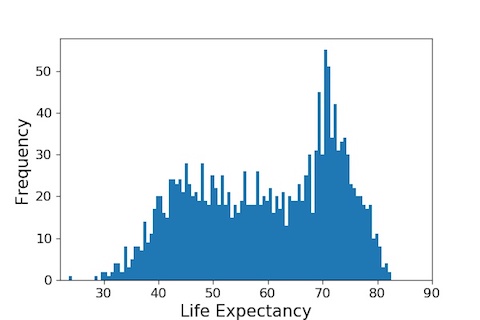
# Data Description

Describe the dataset you used in narrative form. Include a table that lists all attributes in the data along with types (nominal, numeric, ordinal, etc.) and example values. Refer to this table within the text. If the number of attributes is too large so that the table takes up more than a page, then only list the attributes you used in your analysis. If that’s too large, then put the table at the end of the report, as an appendix. The table should have the following formatting:

1. Table Title

| Attribute | Type | Example Value | Descripton |
| --- | --- | --- | --- |
| ID | Nominal (primary key) | 432987 | Record identifier |
| NAME | Nominal (string) | “John Smith” | Name of the client |
| AGE | Numeric (integer) | 24 | Reported age |
| INCOME | Ordinal (low, medium, high) | medium | Income level. Low is x<20k, medium is 20k<=x<80k, and high is x>=80k |
| EXP | Numeric (real) | 4.25 | Experience in years. |

Make sure to identify attributes you used for your analysis. Provide some descriptive statistics of these attributes (e.g. frequency distribution, mean, standard deviation, range, mode, etc.). These can be given as a table or using figures. For example, you can show a figure of the histogram of a variable. DO NOT just copy and paste screenshots from some software. Figures should be of high-quality, numbered, and include a caption like this:



1. Frequency distribution of life expectancy.

It may make sense to visualize the data as a whole. For example, if you’re analyzing network data, you can generate a figure of the network or at least a part of it.

# Methodology

In this section, you should present the steps you took to perform the analysis. It’s a good idea to include a flow chart of these steps. Be sure it’s detailed enough so that the reader could easily recreate your work. Make sure to cite appropriately. For example, if you mention a specific data mining method, be sure to cite the paper of the author that came up with this method.

# Results and Discussion

In this section, you will show and discuss the results of your analysis. This should include figures that visualize the results. These could be figures of models generated, graphs evaluating the performance of models, or plots showing the sensitivity of attributes to the target value. You should make sure to describe each result in detail and discuss the implications of the results.

# Conclusions

In this section, you should remind the reader what you have done throughout the paper (i.e., do a short summary), then describe the main takeaways of the paper.

# General Tips for Writing Data Science Reports

* Write as if you are the expert data scientist and the instructor is your client for whom you need to analyze the data.
* Remember that the purpose of data science is to find new knowledge in data. The whole report needs to be written around this purpose. The conclusions should be about new insights that come from the analysis of the data and how they could be applied.
* When choosing data to work on, think about the problem the analysis will solve in this data. Also, make sure you focus on finding up to date, real data sets. For example, choose a dataset on current crime data from city portals or recently gathered data from social networks. Using old, well-used datasets that are now primarily used for teaching purposes is not interesting. Find data about something that interests you.
* Make the title specific. Instead of using “Clustering on Data”. Write something that relates to the data and the problem around it: “Identifying Groups of Customers for Good Market Segmentation”.
* The introduction section should do several things:
  1. Begin by stating the problem, which in the case of data science will be based on the data.
  2. Motivate why your analysis work was useful for this data.
  3. Provide a short overview of what was done in the process and the general outcomes.
  4. Outline the rest of the paper. For example: “In section II, I provide an overview of the data. Then in section III, the analysis methodology is presented. Section IV, describes the results and discusses the analysis. Lastly, section V provides conclusions of the analysis.”
* DO NOT PASTE SCREENSHOTS! (that is, unless you are actually writing about what is going on in the computer’s screen, e.g. talking about graphical user interfaces – but that’s an exception). Only show what is needed to help the user understand your methods or results. Make sure everything is clearly legible (sufficiently large fonts, easily distinguished features on the graphs, etc.).
* Figures and tables should be used to help the reader understand the writing and you should refer to them in the text. They should be numbered and labeled per IEEE specifications.
* Make sure to provide references to things you mention in the text: data sources, software, algorithms, theorems, facts about the problem or data, etc. You need to attribute the source, otherwise, it’s plagiarism.
* Make sure to adhere to IEEE formatting guidelines – use the template for IEEE Transactions articles.
* Proofread your paper to make sure you avoid spelling and grammar mistakes and that the paper flows well. Get help on writing if necessary.
* Look up papers in IEEE Transactions journals for examples on how papers should be written. Look for journals with high impact scores that are also relevant to the field.
* Make sure to justify your methodology: why did you pick these particular algorithms? How did you go about finding the optimal parameters for the algorithms? Why did you preprocess the data in a particular way (e.g. normalized to mean of zero)?

# General IEEE Styling Guidelines

The following are guidelines from the IEEE template that you should keep in mind when working on your report. You should not have this section in your paper.

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.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.
* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

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.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*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 . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum **0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

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

## Figures and Tables

#### 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”.

##### References

This section is unnumbered and lists all the works you cited. This text is just for your information. Your report will only contain the list. The following is additional information from the IEEE template:

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 ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

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.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.