**SCHOOL OF SCIENCE AND TECHNOLOGY**

**ASSIGNMENT FOR THE**

**BSC (HONS) IS; YEAR 2**

**BSC (HONS) IS (BUSINESS ANALYTICS); YEAR 2**

**BSC (HONS) IS (DATA ANALYTICS); YEAR 2**

**ACADEMIC SESSION AUGUST 2020**

**IST2034: ANALYTICS ENGINEERING**

**DEADLINE: Group Report - Week 13 (20 Nov, Fri, 5pm)**

**STUDENT NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ STUDENT ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**STUDENT NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ STUDENT ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**STUDENT NAME: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ STUDENT ID: \_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**INSTRUCTIONS TO CANDIDATES**

# 

# This assignment will contribute 30% to your final grade.

**IMPORTANT**

# The University requires students to adhere to submission deadlines for any form of assessment. Penalties are applied in relation to unauthorized late submission of work.

# Coursework submitted after the deadline but within 1 week will be accepted for a maximum mark of 40%.

# Work handed in following the extension of 1 week after the original deadline will be regarded as a non-submission and marked zero.

**Lecturer’s Remark** (Use additional sheet if required)

We …........................................... (Name) ………..……………….........std. ID received the assignment and read the comments ..............……………………...……… (Signature/date)

**Academic Honesty Acknowledgement**

“We ………............... (student name) verify that this paper contains entirely our own work. We have not consulted with any outside person or materials other than what was specified (an interviewee, for example) in the assignment or the syllabus requirements. Further, We have not copied or inadvertently copied ideas, sentences, or paragraphs from another student. We realize the penalties *(refer to page 16, 5.5, Appendix 2, page 44 of the student handbook diploma and undergraduate programme)* for any kind of copying or collaboration on any assignment.”

…………………………………………………………………............. (Student’s signature / Date)

**Report: 30% contribute to final**

Assessment Criteria:

* Research questions derived from preliminary data exploration : 5%
* Data validation, cleaning and manipulation to address the issues : 10%
* Output and discussion of the finding : 10%
* Programs with internal documentation : 5%

Evaluation Rubric:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Assessment criteria | Weight | 10-9 | 8-7 | 6-4 | 3-1 | 0 | Mark Awarded |
| Research questions derived from preliminary data exploration | 5% | Creating excellent research questions and support with preliminary data exploration techniques. | Creating good research questions and support with preliminary data exploration techniques. | Creating weak research questions and acceptable support preliminary data exploration techniques. | Creating weak research questions and minimum support with preliminary data exploration techniques. | No assignment submitted. |  |
| Data validation, cleaning and manipulation to address the issues | 10% | Excellent executed with no errors. | Well executed with few errors. | Poorly executed with many errors. | Very poorly executed and very difficult to read. | No assignment submitted. |  |
| Output and discussion of the finding | 10% | Addressed all of the research questions with additional components. Clearly illustrates the idea and well thought out the response. | Addressed all of the research questions. It is relatively detailed to illustrate the idea and relatively well thought out the response. | Addressed many of the research questions. Not detailed and poorly thought out the response. | Not address all research questions. No evidence of having given the assignment real thought or submitted late. | No assignment submitted. |  |
| Programs with internal documentation | 5% | Excellent executed with no errors. | Well executed with few errors. | Poorly executed with many errors. | Very poorly executed and very difficult to read. | No assignment submitted. |  |
| **Remark:**  **Total** | | | | | | | / 30% |

# IST2034 Analytics Engineering

# Movies Data Storytelling

In this assignment, work in a group of THREE members to explore and discover the patterns of movie audiences from the [MovieLens](https://movielens.org/) users who joined in year 2000. Dataset can be downloaded [HERE](http://www.grouplens.org/system/files/ml-1m.zip) and read the [README file](http://files.grouplens.org/papers/ml-1m-README.txt) before download.

You are required to read all the data files into SAS data format, follow by exploring the data to understand what each of the column and row represents. Next, making a basic visualization using plotting methods. Then, develop **TWO** research questions upon familiarize with the all datasets. In order to answer your research question(s), you will begin digging into the data by cleaning and manipulating. Finally, produce a group **REPORT** based on your research questions.

What to submit by WEEK 13 in a group?

* A report NOT more than 10 pages, use the format template as reference and ensure the report consists of
  + Introduction
    - Background about the research questions. State TWO research questions that your group have derived.

For example, is movie genres associate with the type of movie audience?

* + Main body of the report should have sections:
    - Data exploration, cleaning and/or manipulation
      * How you handling missing data, data validation, creating secondary variables, binning or group variables in order to answer the research questions.
      * Output of your finding with labels and discussion: support with graphs and/or tables.
  + Conclusion
  + Appendix with SAS, R and/or Python programs
* Source code of SAS, R and/or Python

Assessment Criteria:

* Submit report on 20 Nov 2020 (Friday) before 5pm on eLearn by the group leader. Group members have responsibility to ensure the submission is done by their group leader.
* Mark distribution
  + Report will contribute 30% to your final grade.
    - Research questions : 5%
    - Data exploration, cleaning and manipulation : 10%
    - Output and discussion : 10%
    - Programs : 5%

\*Group registration must be done before 18th September 2020, 8am: Group leader please register [HERE](https://forms.gle/GWvb1AfGKj6yh5BG7)

# Create Your Own Movies Data Storytelling Title

Student Name Student Name Student Name  
Student ID Student ID Student ID

**Abstract** (The abstract should not exceed 300 words. It should briefly summarize the essence of the paper and address the following areas without using specific subsection titles.):Objective:Briefly state the problem or issue addressed, in language accessible to a general audience. Method:Briefly summarize data validation, cleaning and manipulation to address the problem. Results:Provide a brief summary of the results and findings. Conclusions: Give brief concluding remarks on your outcomes. SAS/R/Python Programming Impact:Comment on the programming aspect of the work presented in the paper and analysis of the plans. Detailed discussion of these aspects should be provided in the main body of the paper.

(Note that the organization of the body of the paper is at the authors’ discretion; the only required sections are Introduction, Methods and Procedures, Results, Conclusion, and References. Acknowledgements and Appendices are encouraged but optional.)

INTRODUCTION[[1]](#footnote-1)

T

HIS document is a template for Microsoft *Word* versions 6.0 or later.

To insert images in *Word,* position the cursor at the insertion point and either use Insert | Picture | From File or copy the image to the Windows clipboard and then Edit | Paste Special | Picture (with “float over text” unchecked).

Please observe the assignment page limits.

MAIN BODY (GIVE APPROPRIATE TOPIC)

## Problem Statement

If you want to submit your file with one column electronically, please do the following:

--First, click on the View menu and choose Print Layout.

--Second, place your cursor in the first paragraph. Go to the Format menu, choose Columns, choose one column Layout, and choose “apply to whole document” from the dropdown menu.

--Third, click and drag the right margin bar to just over 4 inches in width.

The graphics will stay in the “second” column, but you can drag them to the first column. Make the graphic wider to push out any text that may try to fill in next to the graphic.

## Submission

Submit a softcopy of this paper with cover page and rubric, and SAS, R and/or Python source code with (relevant extension), clean data sets, zip all files, named using this format <your group members’ last name> to elearn submission link.

HELPFUL HINTS

## Figures and Tables

Large figures and tables may span both columns. Place figure captions below the figures; place table titles above the tables. If your figure has two parts, include the labels “(a)” and “(b)” as part of the artwork. Please verify that the figures and tables you mention in the text actually exist. Use the abbreviation “Fig.” even at the beginning of a sentence. Do not abbreviate “Table.” Tables are numbered with Roman numerals.



Fig. 1. Magnetization as a function of applied field. Note that “Fig.” is abbreviated. There is a period after the figure number, followed by two spaces. It is good practice to explain the significance of the figure in the caption.

Figure axis labels are often a source of confusion. Use words rather than symbols. As an example, write the quantity “Magnetization,” or “Magnetization *M*,” not just “*M*.” Put units in parentheses. Do not label axes only with units. As in Fig. 1, for example, write “Magnetization (A/m)” or “Magnetization (Am−1),” not just “A/m.” Do not label axes with a ratio of quantities and units. For example, write “Temperature (K),” not “Temperature/K.”

Multipliers can be especially confusing. Write “Magnetization (kA/m)” or “Magnetization (103 A/m).” Do not write “Magnetization (A/m) × 1000” because the reader would not know whether the top axis label in Fig. 1 meant 16000 A/m or 0.016 A/m. Figure labels should be legible, approximately 8 to 12 point type.

TABLE I

Units for Magnetic Properties

|  |  |  |
| --- | --- | --- |
| Symbol | Quantity | Conversion from Gaussian and  CGS EMU to SI a |
| Φ | magnetic flux | 1 Mx → 10−8 Wb = 10−8 V·s |
| *B* | magnetic flux density,  magnetic induction | 1 G → 10−4 T = 10−4 Wb/m2 |
| *H* | magnetic field strength | 1 Oe → 103/(4π) A/m |
| *m* | magnetic moment | 1 erg/G = 1 emu  → 10−3 A·m2 = 10−3 J/T |
| *M* | magnetization | 1 erg/(G·cm3) = 1 emu/cm3  → 103 A/m |
| 4π*M* | magnetization | 1 G → 103/(4π) A/m |
| σ | specific magnetization | 1 erg/(G·g) = 1 emu/g → 1 A·m2/kg |
| *j* | magnetic dipole  moment | 1 erg/G = 1 emu  → 4π × 10−10 Wb·m |
| *J* | magnetic polarization | 1 erg/(G·cm3) = 1 emu/cm3  → 4π × 10−4 T |
| χ*,* κ | susceptibility | 1 → 4π |
| χρ | mass susceptibility | 1 cm3/g → 4π × 10−3 m3/kg |
| μ | permeability | 1 → 4π × 10−7 H/m  = 4π × 10−7 Wb/(A·m) |
| μr | relative permeability | μ → μr |
| *w, W* | energy density | 1 erg/cm3 → 10−1 J/m3 |
| *N, D* | demagnetizing factor | 1 → 1/(4π) |

Vertical lines are optional in tables. Statements that serve as captions for the entire table do not need footnote letters.

aGaussian units are the same as cgs emu for magnetostatics; Mx = maxwell, G = gauss, Oe = oersted; Wb = weber, V = volt, s = second, T = tesla, m = meter, A = ampere, J = joule, kg = kilogram, H = henry.

## References

Number citations consecutively in square brackets [1]. The sentence punctuation follows the brackets [2]. Multiple references [2], [3] are each numbered with separate brackets [1]–[3]. When citing a section in a book, please give the relevant page numbers [2]. In sentences, 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] shows ... .” Please do not use automatic endnotes in *Word*, rather, type the reference list at the end of the paper using the “References” style.

Number footnotes separately in superscripts (Insert | Footnote).[[2]](#footnote-2) Place the actual footnote at the bottom of the column in which it is cited; do not put footnotes in the reference list (endnotes). Use letters for table footnotes (see Table I).

Please note that the references at the end of this document are in the preferred referencing style. Give all authors’ names; do not use “*et al*.” unless there are six authors or more. Use a space after authors’ initials. Papers that have not been published should be cited as “unpublished” [4]. Papers that have been accepted for publication, but not yet specified for an issue should be cited as “to be published” [5]. Papers that have been submitted for publication should be cited as “submitted for publication” [6]. Please give affiliations and addresses for private communications [7].

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 [8].

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have already been defined in the abstract. Abbreviations such as IEEE, SI, ac, and dc do not have to be defined. Abbreviations that incorporate periods should not have spaces: write “C.N.R.S.,” not “C. N. R. S.” Do not use abbreviations in the title unless they are unavoidable.

## Equations

Number equations consecutively with equation numbers in parentheses flush with the right margin, as in (1). First use the equation editor to create the equation. Then select the “Equation” markup style. Press the tab key and write the equation number in parentheses. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Use parentheses to avoid ambiguities in denominators. Punctuate equations when they are part of a sentence, as in

Equation here (1)

Be sure that the symbols in your equation have been defined before the equation appears or immediately following. Italicize symbols (*T* might refer to temperature, but T is the unit tesla). Refer to “(1),” not “Eq. (1)” or “equation (1),” except at the beginning of a sentence: “Equation (1) is ... .”

## Other Recommendations

Use one space after periods and colons. Hyphenate complex modifiers: “zero-field-cooled magnetization.” Avoid dangling participles, such as, “Using (1), the potential was calculated.” [It is not clear who or what used (1).] Write instead, “The potential was calculated by using (1),” or “Using (1), we calculated the potential.”

Use a zero before decimal points: “0.25,” not “.25.” Use “cm3,” not “cc.” Indicate sample dimensions as “0.1 cm × 0.2 cm,” not “0.1 × 0.2 cm2.” The abbreviation for “seconds” is “s,” not “sec.” Do not mix complete spellings and abbreviations of units: use “Wb/m2” or “webers per square meter,” not “webers/m2.” When expressing a range of values, write “7 to 9” or “7-9,” not “7~9.”

A parenthetical statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.) In American English, periods and commas are within quotation marks, like “this period.” Other punctuation is “outside”! Avoid contractions; for example, write “do not” instead of “don’t.” The serial comma is preferred: “A, B, and C” instead of “A, B and C.”

If you wish, you may write in the first person singular or plural and use the active voice (“I observed that ...” or “We observed that ...” instead of “It was observed that ...”). Remember to check spelling.

RESULT

Discuss your result and findings to answer both research questions.

CONCLUSION

Please include a brief summary of your work in the conclusion section. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. Consider elaborating on the translational importance of the work or suggest applications and extensions.

REFERENCES

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. E. H. Miller, “A note on reflector arrays (Periodical style—Accepted for publication),” *IEEE Trans. Antennas Propagat.*, to be published.
3. J. Wang, “Fundamentals of erbium-doped fiber amplifiers arrays (Periodical style—Submitted for publication),” *IEEE J. Quantum Electron.*, submitted for publication.
4. C. J. Kaufman, Rocky Mountain Research Lab., Boulder, CO, private communication, May 1995.
5. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces (Translation Journals style),” *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].
6. M. Young, *The Techincal Writers Handbook.* Mill Valley, CA: University Science, 1989.
7. J. U. Duncombe, “Infrared navigation—Part I: An assessment of feasibility (Periodical style),” *IEEE Trans. Electron Devices*, vol. ED-11, pp. 34–39, Jan. 1959.
8. S. Chen, B. Mulgrew, and P. M. Grant, “A clustering technique for digital communications channel equalization using radial basis

APPENDIX

Include segments of your SAS, R and/or Python source code here.

1. [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)