Analytics Report Critique

Formative and Summative Assessments
BEM2031 - 2020/21 T2 - Instructor: Jesse Michael Fagan

There are two components to consider 1) the outline for the critique is expected two weeks before the full report but may be sent to me earlier, and 2) the full critique of the report along with your additional analyses.

File	Description
hr_analytics.zip	Archive file with a PDF of the report and the data.
kaggle_hr_analytics.csv	A CSV file with the data for the reports. Description of the columns can be found below.
don-t-know-why-employees-leave-read-this.pdf don-t-know-why-employees-leave-read-this.html	This is a personal reproduction of the Kaggle Kernel found here. This report was a "Report of the Week" in 2017. The creator was awarded $500 for this distinction. But there are many flaws in this analysis. Use the attached PDF or HTML files, not the online version. The online version is missing some things. The PDF doesn't have the interactive chart, so the HTML is recommended.

Outline for Critique

Length: 300-500 words
Weight: formative
Deadline: 2021-03-12T15:00:00 GMT
Feedback: written response

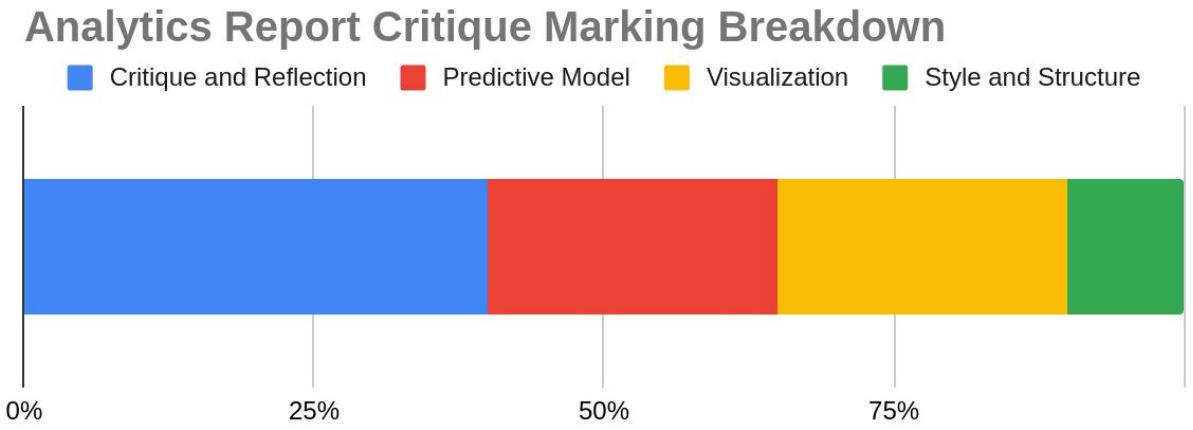
Due two weeks prior to the deadline for the report critique. I am looking to see that you have thought through the assignment and what you plan to write, analyze, or create in response. This can be delivered as an outline with bullet points, or with complete thoughts in well-reasoned paragraphs.

Analytics Report Critique

Word Count: 3,000 words
Weight: 70%
Deadline: 2021-03-26T15:00:00 GMT

This assessment is divided into three primary tasks. See the marking table at the end for more details.

1. You must critique the given report. See the points of critique below.
2. You must provide at *least one* new predictive model of your own. Assess the fit of the model, justify your approach, and interpret the findings.
3. You must create at least one new visualization of your own. You must justify your visualization approach, and provide an interpretation of what you learn from it.



Mark	Fail	3rd	2:2	2:1	1st	Weighting
Marking Criteria	0-39	40-49	50-59	60-69	>70	% of total mark
CRITIQUE & REFLECTION	Completely ignores glaring issues, or utterly fails to explain any of the reasoning	Mis-identifies strengths and weaknesses. Poorly expresses reasoning for the selected aspects.	Broadly identifies strengths and weaknesses, or misses important problems with the report	Clearly states the different strengths and problems of the report, but misses some issues, or does not fully justify choices	Displays excellent and detailed understanding of the strengths and weaknesses of each element of the report and can explain their reasoning clearly	40%
MODELING / ANALYSIS	Analytic approach is wrong, incorrectly applied, or incorrectly explained	Analytic approach is wrong, misunderstanding of core aspects of the method used. implications are not correctly understood or interpreted	Analytic approach is inappropriate, understanding of the method is not demonstrated, and implications are not correctly understood or interpreted	Appropriate analytic approach, but some elements of the method are not explained, and important implications are missed	Analytic approach is appropriate, correctly explained in detail, and the implications are well understood	25%
VISUALIZATION	Visualization is completely unsuited to the data, very poorly executed, and it's impossible to understand what it's trying to say.	Visualization is unusable for interpretation. Not appropriate for the data, execution is poor, and it's unclear what it's purpose is	Visualization is not appropriate for the data, execution is poor, and it's unclear what it's purpose is	Visualization is appropriate, but is not well executed, the store is present but requires additional explanation	Visualization use is appropriate, clean and clearly explained, it stands on its own, and has a quick, clear meaning	25%
STYLE, STRUCTURE, & KEY COMPONENTS	Poor structure and grammar which is hard to follow or understand; incorrectly formatted, with no references or poorly chosen ones. Complete misunderstanding of the purpose of different components of the report	There is difficulty reading the paper. Many errors in grammar, vocabulary & referencing. Poor use of references. Poorly developed understanding of the elements of the report	A reasonable structure and use of language. Some errors in grammar, vocabulary & referencing. Limited choice of references. Poorly developed understanding of the elements of the report	Good structure and grammar, which is easy to follow & understand. Few instances of typos or formatting errors. Evidence of further research and wider reading. Complete overview, but misses elements of the purpose and intent of elements	Clearly structured and lucidly expressed. Only minor errors in language, grammar or referencing. Evidence of further research. Excellent range and quality of references to support analysis. Excellent overview of the content and mission of the report	10%
						Total 100%

Context:

Consider you lead an analytics team at an international corporation. There have been a number of notable departures of your top talent recently - some of the best contributors have quit their jobs and left for new companies. The top management is concerned that something is systematically wrong with their retention policies. You have been collecting data on job satisfaction, performance, and other metrics for years and merged this information with other data on employee work loads and other information in order to determine what is leading to employees to leave the company.

The report found in **don-t-know-why-employees-leave-read-this.html** is the report that was delivered to you by a member of your data science team. It is heavy on analytics and visualization, but very light on interpretation and explanation since it was intended to be a discussion piece for a series of meetings that will happen soon. You are to read the report and come to the meeting prepared to ask a number of questions about the analysis, ask for changes, and show some of your own results for comparison.

The data science team was able to produce a dataset, **kaggle_hr_analytics.csv**, consisting of the following features from 15,000 employees.

- Employee satisfaction level, based on survey data (satisfaction_level)
- Last evaluation, supervisor rated performance evaluation (last_evaluation)
- Number of projects employee worked on (number_project)
- Average monthly hours (average_montly_hours)
- Time spent at the company in number of years (time_spend_company)
- Whether they have had a work accident (1 = yes, 0 = no)
- Whether they have had a promotion in the last 5 years (1 = yes, 0 = no)
- Department, text data based on the different departments
- Salary, are they highly paid, medium paid, or low paid.
- Whether the employee has left (1 = yes, 0 = no)

How to get started

1. Read the report, try and understand it as much as you can. Take notes. Determine where the report is probably lacking or potentially made some mistakes. Ask me questions when you get a chance.
2. Play with the data. Load the RStudio cloud workspace I created for the summative assessment. Look at the data and try to filter it, create some of the charts, maybe. I don't know.
3. Read the assignment brief, especially the list of questions I prompt you to think about. Consider the business problem - why are people quitting their job? And how well does the data, analysis, etc. address this problem?

Your Predictive Model and Visualization

You have a lot of freedom regarding your choice of visualization or predictive model. You can use code directly from the code sharing document. Whatever choice you make you should do three things to contextualize and explain your decisions and results:

1. Say what you're going to and why decided to do it. What motivated your choice of visual or model?
2. Say what you did, and how. What challenges did you face and how did you solve them? Challenges could be conceptual or technical.
3. Say what you learned and what value it brings to the analysis. How do the results alter your perspective on the original business problem?

Additional Details

This data was part of a Kaggle competition to predict who would leave and who would stay based on the data. The original Kaggle report by Yassine Ghouzam can be found here. You are encouraged to explore the comments on Kaggle or other reports and analytics that have used this dataset, although the original dataset has been removed from Kaggle.

Refer to the *Proposal Review Guide* in Appendix A in your *Data Science for Business* book for a good outline of what to look for and what to critique in the report. There is another sample report and critique in Appendix B as well.

The document should be prepared in word processing program (e.g. Microsoft Word or Google Docs) and submitted as a PDF or DOCX format. Copy and paste all tables and visualizations into the document. You are not required to include the code you wrote to produce the analyses or visualizations. If you do choose to include the code, add it to the end of the document in an appendix. The visualizations, tables, and code do not count towards the word count.

You do not have a required quota for academic references. They are not expected to add academic references to your report. Support your claims using a sound analysis of the data provided. If you do reference material outside the reports or data described here, you are expected to cite that material. Use APA style for your references if you include them.

A good way to structure the report is to consider the CRISP-DM - the CRoss Industry Standard Process for Data Mining.

Structuring Your Report

Your goal is to be critical of the report that was given to you by another team member. You need to consider ways in which this analysis could be improved and provide your own interpretation of the situation before this report is shared with company leadership. The questions listed in each heading are suggestions. You do not need to answer each one, and your report can explore other questions not listed here.

- **Business Understanding**
    - What is your understanding of the goals of this project? Is the data analysis suitable and the data used going to be to be to help guide decision making? What are the costs and benefits of this analysis? Who do you think will be harmed or benefited by this analysis?
- **Data Understanding.**
    - Is the data appropriate? What don't we know from the data that would be helpful when understanding the results? What data should they have included that was missing?
    - How effective are the visualizations at building the narrative of the report? How could they be improved? What visualizations are missing that could help?
- **Data Preparation**
    - Did the report appropriately explore all the different ways in which the data may

be corrupted? What were additional cleaning steps they could have considered? Should they have reshaped the data in any way? Do you trust the data? What would make you trust or distrust the data?

- **Modeling**
    - Were the analytics choices here appropriate? Did they apply them correctly? You don't need to know the specifics of the code, but more about the general approach (e.g. was a decision tree a good choice, or is there another analysis that would have been better used?)
    - How were the models evaluated? How do you know that they fit the data appropriately? What approaches did they use to avoid overfitting? Do we know if these models will work on unknown data in the future? What metrics could they have used to assess the quality of the model?
    - What were the important variables in the models and how do you know they are important? Do we know how these variable impact the outcome? How could they have measured that impact?
- **Evaluation**
    - Should this project move on to deployment? Do you think that the process here, the data used, the analytics and visualizations produced helped solve the problem? What changes to the process would you make to continue this assessment in the future? What was missing from the data or analytic process that you feel should be included?
- **Deployment**
    - What action would you recommend considering the results of the report and your own analyses? What makes you think these actions would be feasible and effective? Note - you don't have to be an expert in human resource management (HRM), you don't have to justify your decisions based on a deep understanding of HRM, *you should justify it based on what the data and analytics suggest*. Even if you *do* have a good idea of HRM, I expect you to justify any actions using analytics based on this data.
    - How would you deploy this analysis in an ongoing process? What are the needs of a descriptive, predictive, or prescriptive deployment? What would you need to do to implement an ongoing prescriptive dashboard system? What consequences do you expect from these analyses?