BMME128: Group Project

Problem Description

Sweek.com is one of the largest e-reading platforms in the world. It's a platform on which people can upload short stories and even entire books within minutes and publish them to a worldwide audience. Set up an account and have a look for yourself: sweek.com.

To serve its customers better, Sweek is trying to understand patterns of reading behaviour and what makes stories successful. To this end, they have approached you as soon-to-be data scientists to analyse their data and write a report of your analyses.

Sweek would like you to address two broad questions:

(1) Reading behaviour

Sweek are looking to position themselves as experts on consumer reading behaviour. To do so, they need to understand their users better. Below are a few example questions that the founders are wondering about. Use your newly acquired data skills to analyse them and come up with convincing insights.

You are also encouraged to go beyond these questions. Part of being a great data scientist is coming up with the right questions. Based on your own intuition and experience as a reader, come up with interesting ways to analyse reading behaviour and derive commercially relevant implications. The more novel and creative your insights, the better. Next to gaining expert status, Sweek is looking for positive, catchy findings that would do well in infographics for the press and social media.

What patterns in reading behaviour can you identify?
- Do people read more in the weekend or depending on the time of day?
- Do they only read certain genres or do they combine certain types of content?
- Can you distinguish types of readers based on their reading behaviour? Can you group readers by behaviour or the content they consume?
- How do types of readers relate to other observable characteristics (e.g., gender, language, …)?
- Can you find trending themes/topics?

What patterns of interaction behaviour can you identify?
- Do people interact more in the weekend or depending on the time of day?
- Do they only interact with certain genres?
- Can you distinguish types of readers based on their interactions? Can you group readers by their interaction behaviour?
- Do only certain people interact and what makes them interact with a story?
- How does interaction relate to reading behaviour? Does interaction enhance reading of the focal or other stories?
- Does interaction lead to more interaction (with the focal story or other stories)?

(2) Story success

Another important question for Sweek is if there are ways to anticipate the success of a story early on. Are there signs that show if a story will do well with readers? Is it the behaviour or the first readers, the number of chapters, the followers of the author on Sweek, any other metadata, etc.?

You are encouraged to come up with interesting questions that go beyond these initial suggestions. Is there a different angle from which to look at reading behaviour and story success? You are also encouraged to link Sweek's data with third party data, for example the popularity of certain genres, economic factors, etc.

Report Structure & Grading

At the end of this project, you need to submit your report (in PDF) on Canvas (25 April, 9am) and present your results to the founding team of Sweek. The presentations will be scheduled between 9am and 2pm on 25 April and slot assignments will be posted on Canvas. The report should not be longer than 15 pages (11pt font, 2.54cm margins, including everything; anything after the 15th page will not be graded) and the presentations should not be longer than 7 minutes to leave room for Q&A. Both the report and the presentation will be graded by the TA and your instructor. You can create your report using Jupyter notebooks, but you don't need to.

Please structure your report in the following way:
- Brief executive summary that concisely describes the main insights.
- General data engineering
- Question 1
  - Data engineering
  - Descriptive statistics, model selection, analysis, & validation
  - Business insights
- Question 2
  - Data engineering
  - Descriptive statistics, model selection, analysis, & validation
  - Business insights

The goal is to create a concise report that crafts a coherent story from beginning to end with important business insights for Sweek, while at the same time adequately documenting the underlying rigorous analyses. In particular, we are looking for reports that go beyond the listed questions above with creative analyses, new data, and surprising insights. For the presentation, focus in particular on conveying the most interesting business insights from your report.

Conditions & Data Notice

Please make sure you read these conditions carefully.

The data will be provided through GitHub. Each group will have access to its own private repository, which you need to use throughout the assignment to access the data and store your code, any additional data you may collect, and your report. **You are required to upload your analysis code to your group's GitHub repository before 9am on 25 April**. Failure to do so will result in a deduction of 1 full point from the final grade (out of 10).

During this group project, you will work in groups of 5. You can create teams by yourself on Canvas. Students who have not joined a team by Wednesday noon (12pm, 10 April 2019) will be assigned to a team at random.

For fairness reasons, late submissions will be penalised. For each day a report is late, its grade will be reduced by 1 full point (out of 10). Extensions can only be granted before the deadline and only in exceptional circumstances.

The provided data is completely anonymised. By participating in this group project, you agree not to share this data, parts of it, or any results derived from it with anyone who is not enrolled in BMME128 2019. In particular, you agree not to make this data available publicly. This data is solely for the purpose of this group assignment and must be deleted once the course has ended. You obtain no rights to the data or the results of your analyses.