The Zebra - Data Science Candidate Assignment

Overview

You are a data scientist for a top movie studio. After a series of box office flops, the producers at your studio are starting to question their strategy and need a new direction. You suggest a new approach - using data to determine what factors go into making a successful film.

Luckily, you have a dataset of over 5,000 films to mine for insights! The producers ask you to analyze the data and present a report detailing your findings and recommendations on revamping the studio's strategy.

Data

Attached is The Zebra movie dataset for use.

Data Dictionary (selected columns)

num_critic_reviews - number of movie reviews written by critics
num_user_reviews - number of movie reviews written by IMDB users
num_users_voted - number of IMDB users that rated the film
duration - the length of the film in minutes
actor_n_facebook_likes - the number of likes on the actor's Facebook page (we can assume this was measured before the film was released)
gross - the film's gross revenue
movie_score - the film's rating on a 1-10 scale by users that voted

Instructions

Objective: Demonstrate your approach to solving an analytical problem by working with a real dataset. Your goal is to showcase your data cleaning, visualization, and analysis skills.

Helpful Tips

- Given the limited time, prioritize a few key, actionable insights rather than aiming for a comprehensive analysis of all data. If you have additional ideas or steps you'd take with more time, feel free to include a brief outline in your submission.
- This assignment is purposely open-ended to see how you approach the problem - how should you quantify a successful film? Which columns from the dataset should you use? You decide! Please explain the rationale behind your decisions / assumptions where applicable.
- Building a model isn't necessary, but if you choose to include one, keep it simple with a baseline approach (e.g., avoid extensive parameter tuning or experimenting with multiple architectures).

Timeline: To respect your time, we recommend spending 4-6 hours on this project. **Please submit this take-home assignment within 5 business days of receiving it**. If you need more time, let us know.

Deliverables:

1. **Report or Presentation: Provide a concise report or presentation (in PDF, Google Doc, or Google Slides format) that include:**
   ○ Key findings, insights, and recommendations from your analysis, illustrated with relevant data and visuals.
   ○ A clear narrative explaining your thought process, assumptions, and any choices you made throughout the analysis.
   ○ Focus on storytelling and actionable insights over technical detail, and don't worry about perfect polish—our focus is on your approach.
2. **Code: Submit the code used for your analysis, either as a script with comments or as a Jupyter Notebook.**
   ○ Please use Python or R (no point-and-click tools like Excel or Tableau), and feel free to leverage any libraries.
   ○ If you use less common libraries, add comments to explain your code.

What We're Looking For

- **Practical Data Skills -** Are you effective at cleaning and transforming data? Are you able to identify and handle data quality issues? Do you incorporate domain knowledge into your analysis?
- **Programming Skills** - Is your code clean and readable? Do you show proficiency in your chosen language?
- **Communication** - Do you communicate your thought process clearly? Is your analysis structured in a way that is understandable to your audience?
- **Math / Stats Knowledge** - Are your modeling/analysis choices technically sound? Can you explain why you made certain assumptions or decisions?

Final Panel Interview

If you progress to the final panel interview, you will be asked to discuss your analysis with the "movie producers," AKA our hiring team. This is to assess how you approach a real data problem and communicate technical content to a general audience.

This will involve giving a 15 minute presentation of your findings and answering questions from the "producers". Please focus on storytelling and recommendations over technical details. We recommend turning your original submission into (simple) slides to facilitate presenting.