The datasets we are going to investigate are based on the videogames sales, hardware sales, and ratings in Japan.

https://www.perfectly-nintendo.com/japanese-sales-media-create-famitsu-dengeki/

Scrape the weekly sales data of Media create and famitsu from 2015 to 2023 in the above website. Build models that answer the following question. (Additional details available below)

"Exploring Popular Game Genres in Japan: Implications for Game Companies"

Missing data, data only from 2015, but it can be fixed.. Please looks for plausible models so we can discuss how to apply them. Please add market conditions predictors. This could be market share for that game overall, proportion of games with the same genre/publisher, characteristics such time of entry etc.

Games survival times, and sales over the period, as a function of characteristics of the games, genre, publisher, time of release, market conditions.

**********NO CHAT GPT**********

In this project, you will apply statistical methods and concepts from class to a real-world decision making situation. You will identify a data-driven decision scenario, source appropriate data, analyze the data, and derive actionable insights from your analyses in any industry/sector of your choice. Your final deliverables include a project report (10-12 pages) and an end-of-semester team presentation. Your project report must include ALL of the following sections:

(1) Executive Summary (less than 1 page): Brief summary of problem, data, analysis, and key findings.

(2) Problem Definition & Significance: Who is the target client for this project, and what business problem are you trying to address for this client? Why is this an interesting or important problem? Present some brief industry statistics or background research to make a strong case for your target problem.

(3) Prior Literature: How have others tried to address this problem and with what outcomes? Note that without this "domain knowledge", your report will be sub-par.

(4) Data Source/Preparation: Where did you source your data? What variables did it contain and how were they measured? Which variables (DV, IV) did you select for your analysis and why? How did you clean the data (if applicable)?

(5) Variable choice: What are your predictors of interest. Explain the rationale for these predictors.

(6) Descriptive Analysis & Data Visualizations: What patterns/trends do you see in your data? What do you infer from these trends?

(7) Models: How did you statistically model your data? Present at most 3-4 models, but you must carefully choose your models. Which model is best and why?

(8) Quality Checks: How do you know that your analysis is trustworthy? Test of assumptions? Robustness checks?

(9) Recommendations: What recommendations do you have for your client, based on your analysis? Your recommendations must be actionable, i.e., things that managers can act upon (e.g., "target more white customers because they spend more" is not actionable because racial discrimination is illegal in the USA).

(10) References.

(11) Appendix: R Code. This section can go beyond the 10-12 page limit.

Note that your report is intended for an executive audience. These are people who have no background or interest in statistical methods or R. Hence, please tone down technical details and write in a language intended for an average educated person. Absolutely, no cut-and-paste from R in the body of the report (appendix excluded). You can summarize relevant numbers from your R output in user-friendly tables by stripping out unnecessary details and/or use data visualizations. You can include your models in a compact manner in the report. All of your R code and outputs should be included in the Appendix section, following your 10-12 page report.

Sections 2 (Problem definition) and 9 (Insights) are the most important sections in this project. About "actionable" recommendations in Section 9, note that saying that house prices increases with CPI or that a certain disease is more prevalent in women than in men are not actionable recommendations, since there's nothing that a manager or a doctor can do about CPI or gender. However adding a bathroom or combining a drug treatment with therapy are actionable recommendations. You must think about this issue in Sections 3, 4, and 5 (while choosing your research problem, data, and hypothesis); Section 9 is too late to address this very important issue.

Please note that classification problems are NOT acceptable for this project. Even though there are many public data sets on classification problems, I want you to demonstrate what you learned from THIS class, not from your data mining class.

You project will be judged based on the criteria listed above, as well as by the quality of your writing. This project is a test of your writing skills, as much as it is a test of your analytic and problem-solving skills. You have only 10-12 pages to address all of the above sections. Hence, you must be ultra-compact and organized in your use of space, while making sure that you present all important information without missing out critical details. Use tables and graphics intelligently to organize information in a compact manner. But just throwing in some graphics and not talking about it in the text is also a bad idea. Please use the best writer in your team to write/supervise/coordinate/organize the project write-up. It is a terrible idea to assign different students to write different sections, in which case, some sections will be well-written and others will be poorly-

written, and overall, you will have an incoherent report that does not flow from section to section. Because this is a team project and not an individual project, the entire team will be penalized for poor writing by some of your team members.

Lastly, note that your professor will compare your code with similar projects on the Internet, run a plagiarism check, and ask you to explain your work. Feel free to browse such code and learn from them, but your work must be sufficiently original, practical, and useful. Downloading an online project, making minor modifications, and submitting as your team project is NOT acceptable, and will get the entire project team a zero grade for the entire project.