Video Games Project

Client is a small games developer who is planning next release and want to use Data to decide what kind of game to make.

They have 2 datasets with historical information on top selling games ***OF ALL TIME*** and want to understand what TYPES of games sell a lot of copies. In particular they are looking for the direction to take their company in.

Day 1 (Thursday)

Initial thoughts:

They want me to use the Data which is historical which raises some questions:

How old is it? Are there any repetitions? Have trends moved on? Can you determine that from the data? Can you compare games that are on both files and determine if there is a trend upwards or downwards towards that type of game? Is there a price point which determines popularity? Can it be linked to reviews? Are reviews on games fairly consistent? Do people base their purchase in a review? Do any products defy reviews? Can I build a model that at least identifies the relationship between the different variables that might determine or indicate volume of sales? Is there a relationship between how good a game is and sales or is it something else? Does it help if there have been series predecessors regardless of the review, essentially are you buying a franchise? Is there room at a later date to include a dashboard for the future that shows the impact of different variables on a video games sales model? Can it later be predictive or what else does it require? How much cleaning does it require? Do I need a few cleaned scripts in order to analyse different things particularly comparing historical information? Seeing as they are small gaming company, are they looking to develop the most sophisticated game or the one with greater potential as realistically they don’t have the resource for scalability yet to grow too fast? Can I determine or predict potential for a series based on first release? Product maturity lifecycle is the ideal model but it comes with inherent risks.

Tasks:

1. Create a private repo in GitHub and load initial stuff that will eventually include files such as: raw data, cleaning scripts, clean data, data analysis, picture files with graphs and other visuals, a readme file, further considerations, a markdown presentation and a PowerPoint or pdf presentation.
2. Go over my CodeClan texts on dplyr and cleaning to make sure I am asking the right questions of the data. Think about how I want it to look in order to manipulate it best. I am particularly thinking of pivot long for example but there are others: mutate being another.
3. Determine existing data variables and what new ones I need (mutate).
4. Go over the business question again and think about what kind of graphs I would like to create
5. Mock up some wire framed potential graphs and see if they answer business questions
6. Start noting the risks to the client regarding growth and building a top selling product.