Project Proposal

Question/need:

What is the framing question of your analysis, or the purpose of the model/system you plan to build?

Which best five Publishers had received the copies sales overall?

Which best five games had received the most copies sales overall?

Which best five genres had received the copies sales overall?

Can you predict Which genre will receive the most copies sales next year?

Who benefits from exploring this question or building this model/system?

Will benefit Publishers to know which genre will receive the most copies than can publish more games from this genre

Data Description:

What dataset(s) do you plan to use, and how will you obtain the data?

Video games sales from Kaggle link:
https://www.kaggle.com/gregorut/videogamesales

Fields include

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year of the game's release
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales.

What is an individual sample/unit of analysis in this project? What characteristics/features do you expect to work with?

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year of the game's release
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales.

If modeling, what will you predict as your target?

Genre Which will receive the most copies sales next year.

Tools:

How do you intend to meet the tools requirement of the project?

Python:-

- Pyplot
- NumPy
- Matplotlib
- Pandas

Are you planning in advance to need or use additional tools beyond those required?

Mostly in the future I may use tools other than the above to accomplish all project objectives

MVP Goal:

What would a minimum viable product (MVP) look like for this project?

Which Genre has received the most copies sales in Japan overall?

- Explore the features
- Drop any unnecessary features
- Fill null value in ['Year'] column
- Check on ['Genre'] column include
- We found 12 different genres include ['Genre'] column
- Rename ['Genre'] column to number
- Create module to base on genre columns
- Accuracy = 20.2%