BYU Data and Modeling Case Study Competition

Purpose

The purpose of this case study competition is to provide an opportunity for students to use their skills in working with and modeling data in a more broad setting than is typically allowed in a classroom setting.

Instructions

Teams: Each team must consist of between 3 and 5 BYU Undergraduate students. Multiple teams can consult one with another in order to help guide, teach, or discuss, but the final product should be work created by each individual team.

Data and Methods: There will be data provided for you that will be sufficient to answer the question, but there are no limitations as to what data you can use. If you find other public data relevant to the problem you are trying to answer, you can include it in your analysis and your conclusions. You are not required to use any specific method. An advanced machine learning method will be weighted as equally as a useful figure as long as they are used correctly, are well justified, and effectively answer the problem you are given.

Presentation: A presentation will consist of a maximum of 5 slides and will take a maximum of 10 minutes. One possible (but not required) way to write 5 slides is

- Slide 1 states the problem you have been given
- Slide 2 describes the data you use and its appropriateness for answering the questions
- Slides 3 and 4 describe the modeling and/or summarizing efforts that were taken, along with useful figures or relevant model results
- Slide 5 includes a summary and conclusions that directly relate to the problem you have been given.

Each individual in the team must contribute to the presentation. Presentation skills will be evaluated as part of the competition.

Dates:
January 11th – Details about the data and the problem statement will be released to registered teams on and after this date
January 25th – Registration deadline
February 9th 6:00 PM to 8:00 PM – Data Contest Event, dinner will be provided at 6:00 PM for presenter and judges. Guests are welcome to attend the presentations that will begin at 7:00 PM.

Presentation Rubric:
10 points: Slide Quality. More points will be given for presentations where slides are neat, clear, and concise.
10 points: Presentation Quality. More points will be given for presentations where descriptions of methods and results are clear and effective and where presenters use good general presentation skills.
10 points: Data handling: More points will be given for presenters who can clearly describe, display, or explain their data along with describing strengths and weaknesses for how well the data they have collected is appropriate to answer the business problem given.
10 points: Methodology: More points will be given for presentations where the methods used effectively address the business problem.
10 points: Conclusions: More points will be given for presentations where the conclusions specifically address the business problem and are justified given the data and modeling work done.

Problem Statement

You and your team are employees of a brand new consulting firm, Anything Analytics. PEMDAS Gaming, a major video game publisher, has just contacted your consulting firm and has offered to provide your first major contract. Angela, a representative from PEMDAS Gaming, has provided a small report of what they expect from you:

Anything Analytics Team,

The last few years have been interesting for the video gaming industry. Many consider video games to be a luxury good, which means that sales react more to a struggling economy than do other goods. However, sales during the Covid-19 pandemic have actually been great despite many economic and financial challenges worldwide. Obviously, this could be directly related to how much time people were spending at home. In any case, PEMDAS Gaming has decided that we want to be prepared for future challenges. ***Our company wants to know how video game sales react to economic conditions that may affect consumer spending habits.*** We want to keep our personal sales data private, but there are a number of resources online for video game sales data from other publishers as well as for major economic indicators. We want you to provide us with a 10 minute presentation on the work you have done to gather, explore, and model the data that can help us answer this question.

Sincerely,
Angela, PEMDAS Gaming

Bobby is the first member of your team to get started on this project, but unfortunately he will not be able to make it to your presentation. He has found several useful data sources. First of all, there is a website that collects video game sales, https://www.vgchartz.com. This site has a variety of tables that contain current and total sales by region for a variety of video games. This data has actually been webscraped by other data enthusiasts and some of it is contained on Kaggle, the data set hosting site, and can be accessed using the link https://www.kaggle.com/datasets/gregorut/videogamesales. There is more data on vgchartz.com than in the Kaggle data set, but it may be sufficient for the problem you have been given. As with many Kaggle data sets, the video game sales data has some public analyses available, such as the one at https://datascience.fm/video-game-sales-analysis/. These may be helpful for understanding how to explore the data, but Bobby points out that the scope of the problem you have been given is very different than what is contained in these analyses.

As far as economic indicators, Bobby found a data warehouse of world development indicators through The World Bank, https://databank.worldbank.org/source/world-development-indicators. Bobby already downloaded some significant economic indicators for the U.S. and Japan, two of the most important video game markets, using this data warehouse. This can be found in the csv file https://richardson.byu.edu/data_contest/indicators.csv. Another variable that Bobby collected is called the consumer confidence index (CCI). This was not found in the World Bank data warehouse, but at the Organization for Economic Co-operation and Development (OECD). This data warehouse can be accessed using the link https://data.oecd.org/leadind/consumer-confidence-index-cci.htm. Again Bobby already downloaded CCI values for U.S. and Japan, which can be found at https://richardson.byu.edu/data_contest/CCI.csv.

Hopefully all the work Bobby did can help you get a head start. Your task is to explore and understand the data, combine it in useful ways, and evaluate the model in order to answer PEMDAS Gaming's main question about how video game sales are related to general economic conditions. It is important for the future of Anything Analytics that you are able to do well on this project.