Elizabeth Earl

DSC680-T301

(Project 2) Milestone 1 – Proposal and Data Selection

Topic:

For my second project I wanted to dive into what makes something great; I want to focus specifically on toys. Not all toys are equal and some new toys flop while others become major and rare hits overnight but what sets these two apart?

Business Problem:

In a growing industry of new technology, it is hard to keep up with constant toys hitting shelves. Children (and adults alike) are now more commonly found behind a screen and managing a way to get them away can be very difficult. Toys like Squishmallows which are essentially just a stuffed egg have become a new craze. How can a toy that does nothing competing against technology?

The aim of my project is to discover which toys are more likely to be successful in a competitive market, when put against technological advances. Some potential questions I will consider:

- Are people more likely to purchase toys when they are part of a collection?
- Does adding words like ‘rare’ or ‘collectible’ make a toy more sought out for?
- Because technology is everywhere, are people more likely to buy toys to catch a break from it?
- How long do toys typically last in top selling spots?

Datasets:

The dataset I'm using can be found here: Amazon eCommerce | Kaggle The dataset includes top toys sold on Amazon's website. I believe Amazon is a reliable central point to base top toys on as most parents (or anyone buying toys) resort to online services like Amazon to figure out what a child would like to be gifted.

I may potentially use similar datasets such as top video games sold. This dataset can be found: Video Game Sales | Kaggle

The reason for this dataset is this will give a sort of insight on technological toys (aka video games) alongside 'regular' toys.

Methods:

For this project I will primarily focus on sentiment modelling; I will run polarity on my data set. I will run various graph types to get a visual understanding of how top toys/games are stacking up against each other. There will be a lot of relying on sales revenue to correlate with which toys are the most popular.

Ethical Considerations:

It will be hard to figure out the consumers of certain toys as the target audience is not always the consumer. For this reason, I will not assume anything about the age bracket of any toys being purchased or a factor in why a toy/game will be successful.

Challenges/Issues:

A potential challenge I see with my project is being able to figure out what makes a toy great. I am going to use assumptions as to what makes certain toys top toys but the assumption should be obvious. For example, if a toy like Squishmallows is the top toy I will assume it is a top toy because it is a type of stuffed animal (which correlated to safety). I will not go as far as to assume (in this example) that the reason it is a top toy is because of its egg like shaped because in such assumption it would be a stretch.

References:

A lot of the datasets are from Kaggle which will be used heavily to discover more datasets. I will continue to use such site as means of researching new datasets if needed.