Video Games Sales Predictions

Created by: Marin Stoytchev
Date: Jan. 11, 2021

Executive Summary

- **Problem Statement**
  Can we predict which game based on Genre, ESRB Rating, Platform and Publisher will have high probability of being a high selling game in the next one-two years?

- **Recommendations Based on Data Analysis and Model Predictions**

  - Based on predictions of a model with 86% (TBD) accuracy, our recommendations for highly successful game are:
    - ✓ Genre – Shooter (TBD)
    - ✓ Platform – current generation PlayStation console (TBD)
    - ✓ ESRB Rating – Mature (TBD)
    - ✓ Publisher – Rockstar Games (TBD)

I. Process Outline

Six-Step Process

- **Problem Definition**
- **Data Collection**
- **Data Pre-processing/Cleaning**
- **Data Exploration and Visualization**
- **Data Modeling and Model Selection**
- **Data Interpretation and Recommendations Based on Model**

II. Problem Identification/Definition

Problem Definition

- **Problem**
  In the last year due to COVID-19 there is a noticeable increase in people staying at home and increase in the time spent playing video games. A small video games developer startup of excellent professionals wants to capitalize on this trend. However, due to limited resources, the team can develop only one game which can be released in the next year. Thus, it is of critical importance that the team focuses on developing a game which will have high probability to be top-selling game.

- **Question**
  Can we predict which combination of Genre, Platform, ESRB Rating and Publisher will result in a highly successful game?

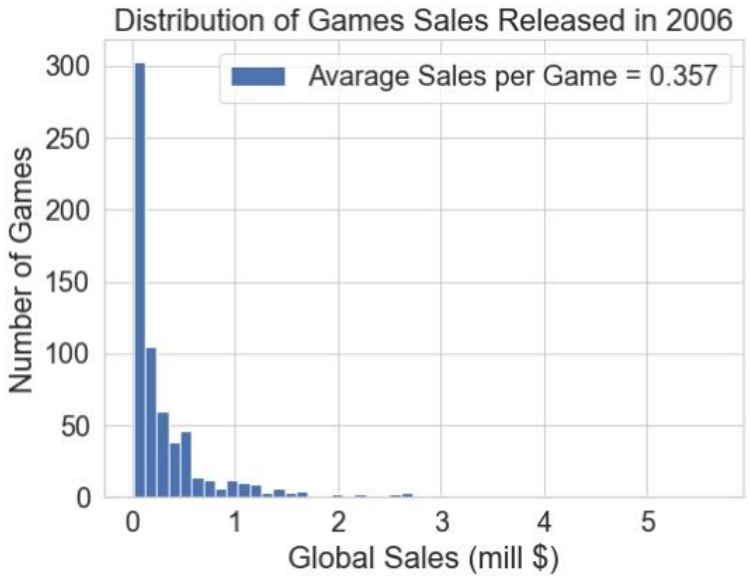
- **Risk**
  If our prediction is inaccurate, this could lead to the bankruptcy of the video game startup since they have the capacity of producing only one game in a period of one year and don't have additional resources to sustain themselves beyond that time period.

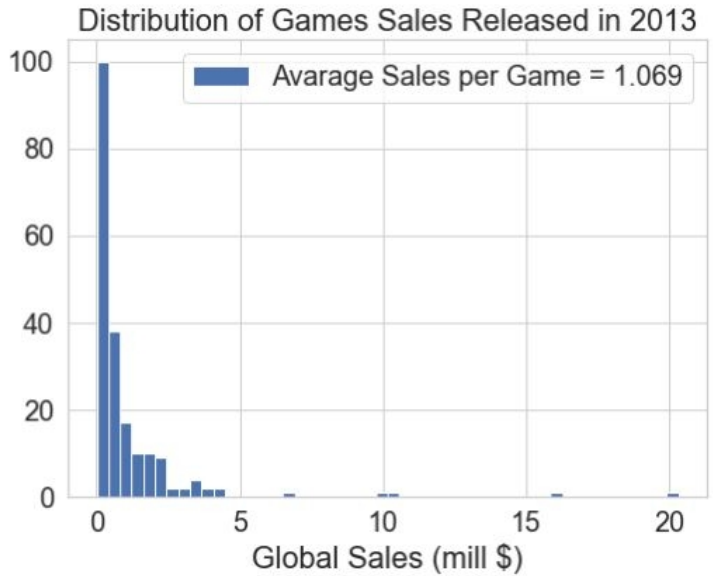
- **Task**
  Analyze available video games sales data and create a model which will accurately predict which video game will be successful in the next year.

III. Data Analysis and Model Results

Typical Distribution of Video Games Sales

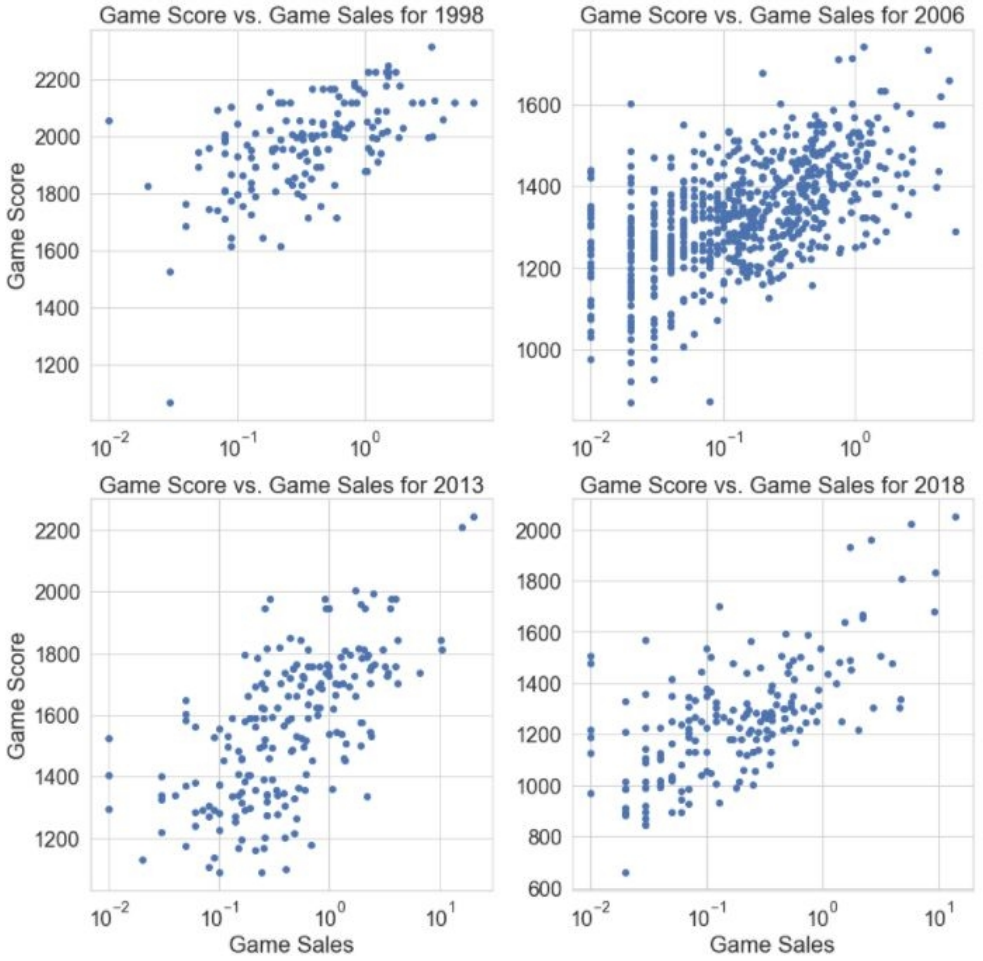
- Typical distributions of the sales of video games since their release is shown for two different years in our data
- The distributions are characterized by:
  - ✔ Very small number of games with large sales since their release date
  - ✔ Large number of games with extremely small sales numbers
- That' why it is critical to be able to predict what makes the best sellers





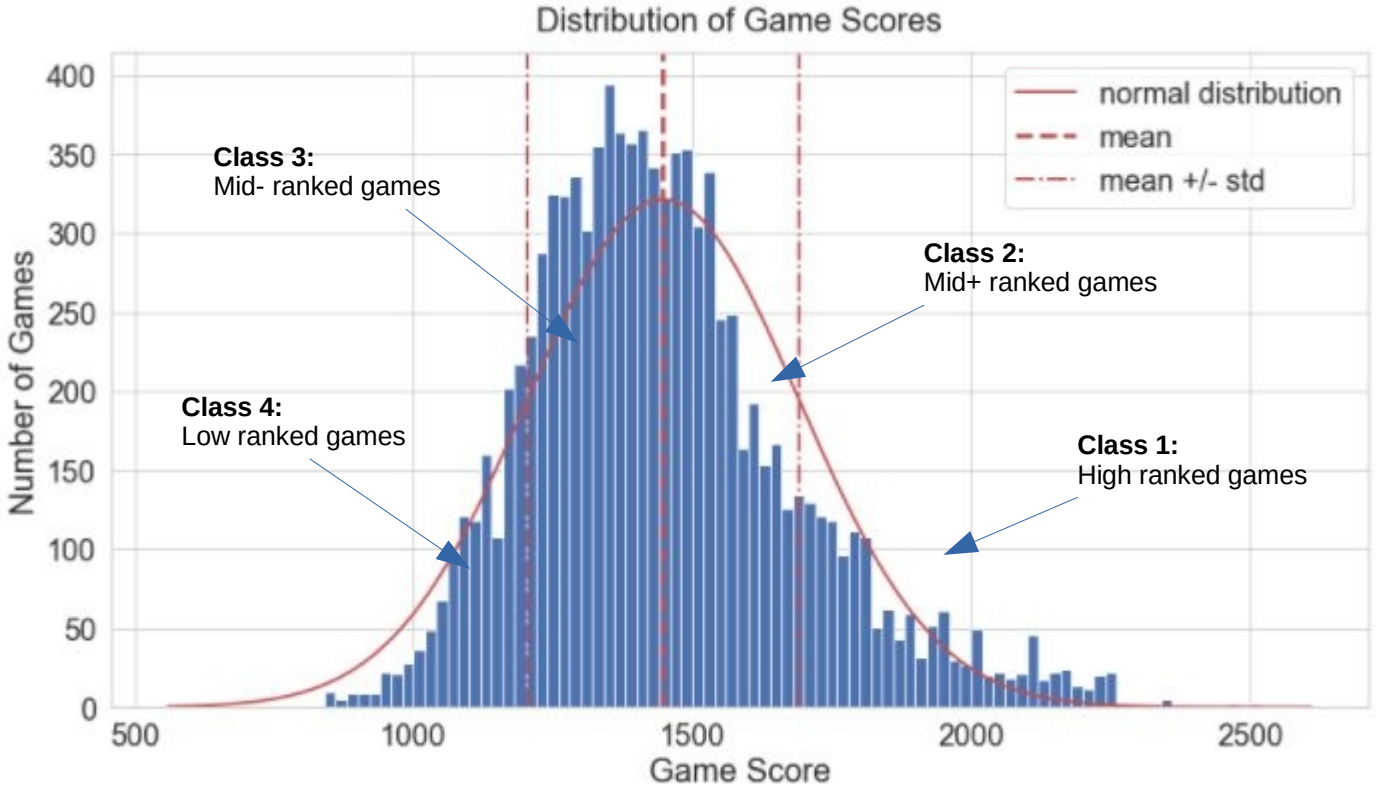
Feature Engineering for Game Classification

- Based on the success of different categories in Genre, ESRB Rating, Platform, and Publisher, from the available data we have created a Game Score feature which is highly correlated with game sales and allows for reliable separation of games in different classes



Feature Engineering for Game Classification (cont.)

- Based on the distribution of game scores we can reliably separate all game in the data set into four different classes as shown below



Model Predictions

- **We have built Random Forest Classification model**
- **After training and optimization (TBD), the model predicts with an average accuracy of 86% (TBD) the different classes of the test data set**

Confusion Matrix

```
[[280  33   3   0]
 [ 20 492  75   1]
 [  2  51 721  34]
 [  1   1  46 190]]
```

Classification Report

	precision	recall	f1-score	support
1	0.92	0.89	0.90	316
2	0.85	0.84	0.84	588
3	0.85	0.89	0.87	808
4	0.84	0.80	0.82	238
accuracy			0.86	1950
macro avg	0.87	0.85	0.86	1950
weighted avg	0.86	0.86	0.86	1950

IV. Summary

Summary

- **TBD**