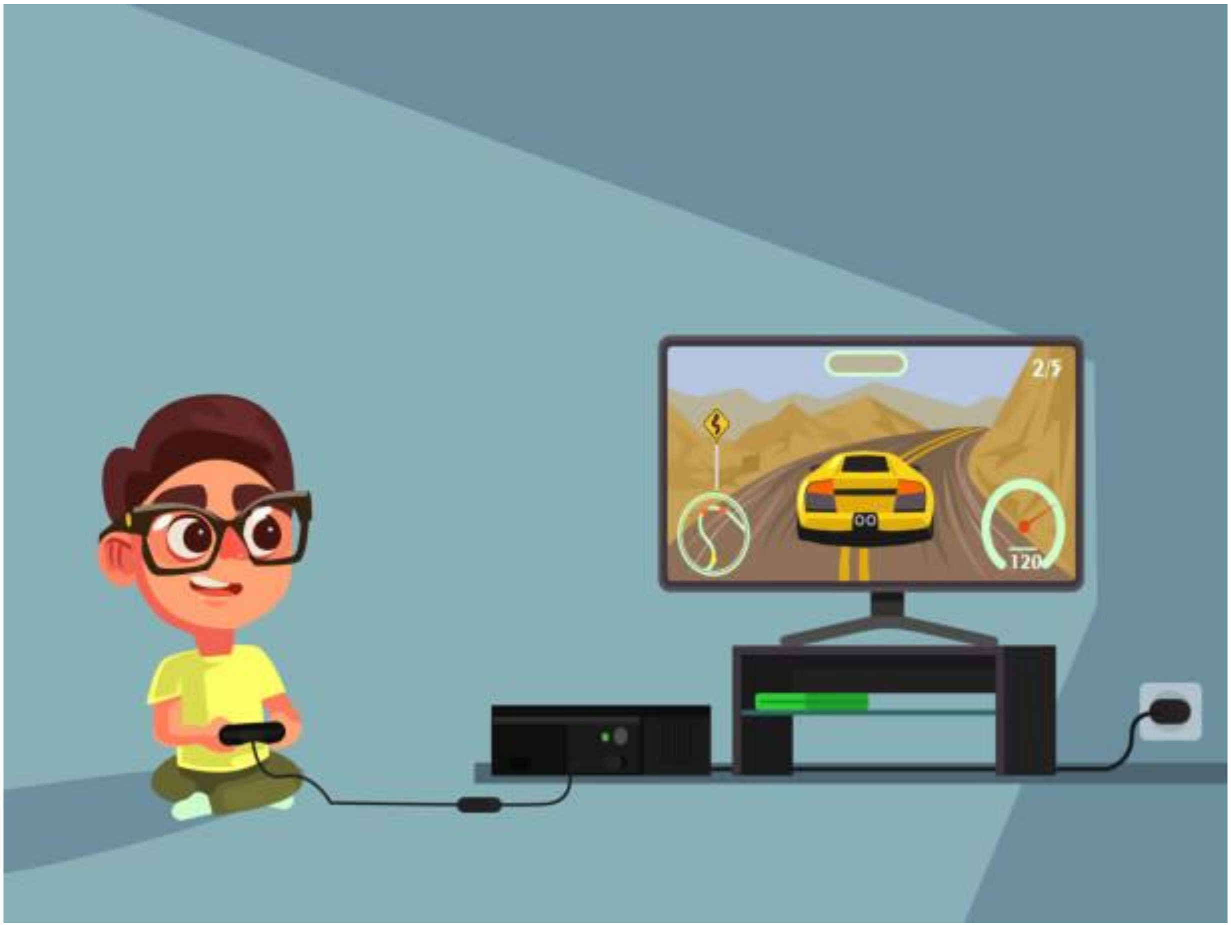


Video Game Sales

Eric Kemna | Cpts 315 | Dec 12 2022

.

Introduction

I grew up in the era with Atari gaming system. My first gaming console was the Nintendo when I was 6 years old. I know that it was only a 8 bit machine but I found it amazing. I witnessed the gaming world explode especially when online gaming arrived. It made the world smaller. I have met many people from all around the world and learned how they grew up. I would love to support my deepest passion and find a job as a data analyst for a gaming company.

My motivation for this project is to understand gaming industry from a business side. One day, I could maybe create the next big game. George Romero created a Call of Duty World at War game mode called “zombies”. He wanted to create a different experience then just a Player vs Player (PVP) shooter. Romero knew from their data analytics that the player base did not like PVP and preferred the game style. By creating this style, it would include those people in an online game that would fit their preferences. A fortunate outcome of the creation would be the popularity in the gaming world. The primary reason I went back to school for data analytics is because of my love of numbers and games.

Based on sales data in the dataset, I classified whether a game belongs to sports or a another less popular genre. The primary challenges for this project are accurately completing my analysis of the data in this report, learning the overall process of data organization and how to display it. My goal was to gain a better understanding analysis data process. I started my approach by wanting to answer the four questions below:

1.) What type of video game should I make and what genre type of genre?
2.) How should I group my classifications in genres?
3.) What classifier tells optimizes for sales?
4.) Is there a trend from top video games?

I knew that these questions would be hard to answer. I reviewed notes from class and did an internet search for good practices for classifiers. I learned more about Sklearn Python Library. I also found Seaborn and learned more about Pandas Library. As I went through this and studied the dataset, my challenges changed, and my main questions were not relevant. However, I chose to compare the four classifiers discussed in class

DATA MINING

Cleaning Data

The data set was not cleaned and most of the sales data cells were empty.The data set was in a .cvs file with all numerical sales data fields values are units sold into the millions.

The region sales are

- Na_Sales: North America

- JP_Sales:Japan
- Pal_Sales: Which include China, European Countries (except France).
- Other Sales: Is all countries that was not listed above

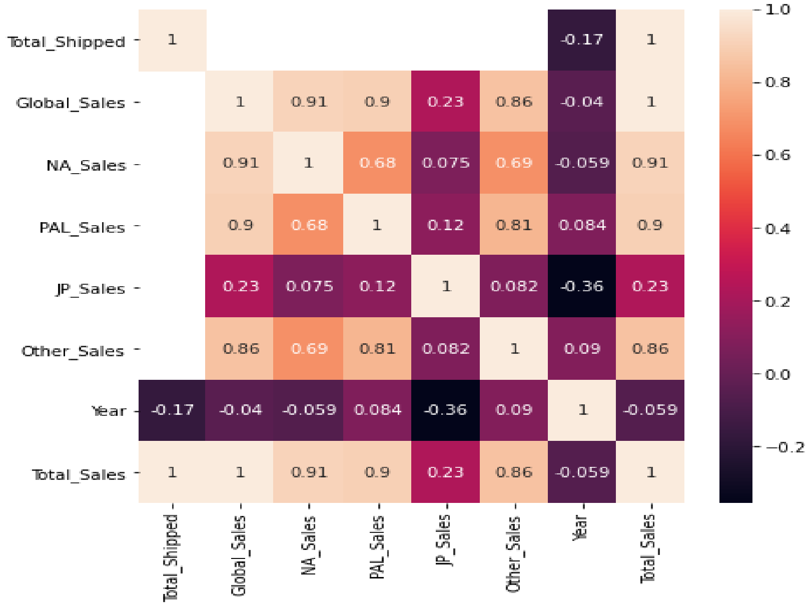
I created global sales which should be equal to the sum of all region sales which adds Na_Sales + JP_Sales + Pal_Sales + Other_Sales.

Name and platform cannot be null as they are key components. I deleted the data from 2019 because there was not a full year and the data stopped on 12-4-2019

Task Details

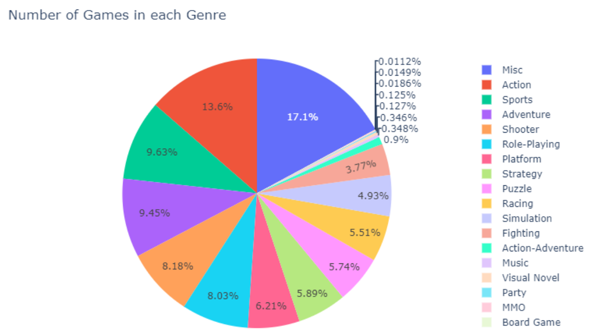
My first task after cleaning was to analyze the dataset using heatmap from the Seaborn Library.

In the below example, a simple heat map is plotted with a linear regression. The closer to 1 the parameters are shows correlation of each variable.



I used the statistical learning data mining technique, to evaluate the data points and how correlated they are to each other. It is common to set the linear regression to one and the closer to one the variables are equates to correlation of those variables.

.



After review of number of games in a genre and with some limitations of the data due to null values, I reset my questions. I wanted to test if a game belonged to the sports genre or belonged to a different genre such as music, party, or visual novel.

TECHNICAL APPROACH

To test if a game belonged to the sports genre or a less popular genre, I had to break my data into a test and training dataset. The size split was test 0.33 and training 0.67. I chose implementation of Gaussian Bayers in Sklearn Libraries. It makes predictions about unknown classes using Bayers theory of probability. As I got more familiar with Sklearn, I used the decision tree to find my answers.

Decision trees are the most important elements of a Random Forest. They are capable of fitting complex data sets while allowing the user to see how a decision was made. I used Sklearn's model selection module in order to provide various functions to cross-validate my model, tune the estimator's hyperparameters, produce validation and learning curves. I used two classifiers to test and determine which had better results. a better result

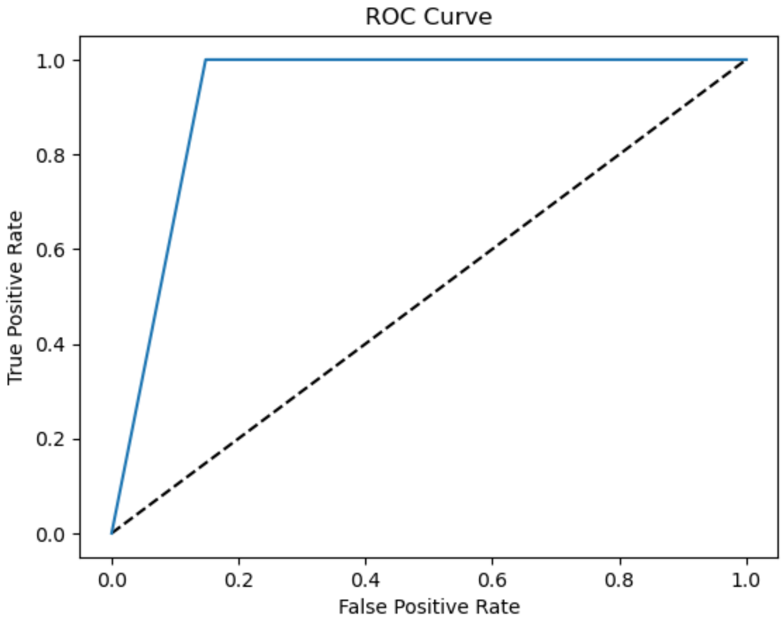
My chosen classifiers were as follows:

- GaussianNB
- confusion_matrix
- model_selection
- model_selection.KFold(n_splits=10)
- DecisionTreeClassifier()

.

Result

The result of GaussianNB Classifer had an accuracy score of 92%. This means that 92% of the time the genera was set to the correct classification.



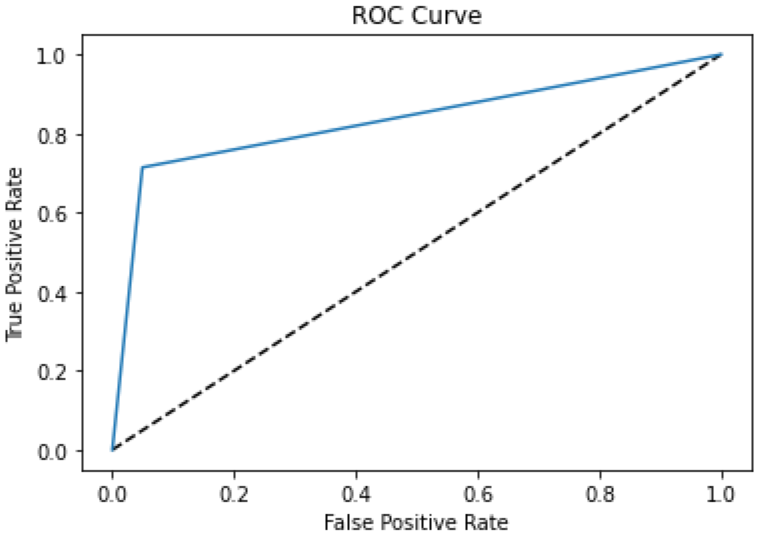
```
The accuracy is: 92.04545454545455 %
          precision    recall  f1-score   support

       0       0.95      0.96      0.96        81
       1       0.50      0.43      0.46         7

   accuracy                           0.92        88
  macro avg       0.73      0.70      0.71        88
weighted avg       0.92      0.92      0.92        88
```

The accuracy came out at 92%.

The Decision Tree classifier had an Auc accuracy rate of 83%.

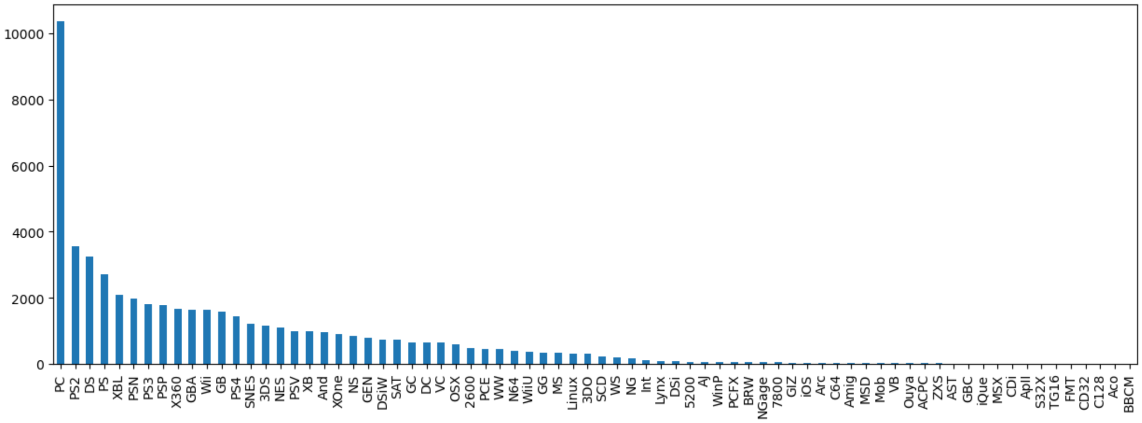


Because the Auc Accuracry rate has a lower rate at 83% than the Navie Baynes classification, I would use the Naïve Baynes for model since it has a higher accuracy rate at 92%

Answer questions

What type of video game should I make and what type of genre?
For testing the genre, the market is evenly split between the top ten genres. As long as the focus is not on the bottom, and stay in the top ten genres the analysis is accurate.

Is there a trend from top video games?
No, there is much of a trend. Sales seem to be balanced across the genre categories. Based on the way the market is shifting, analysis may need to be completed for a Computer based gaming as it is a major share of the makart market.



.

Lessons learned

I learned a lot about Sklearn. I had a job opportunity with a banking job last year. I told them I was not strong enough because I needed to learn how to do some data analysis from Sklearn. I wished that I knew about Sklearn a year ago. I feel a lot more confident about it. I learned a lot more about programming and Jupyter. I have had some major issues with figuring out why my environments were wrong. For a person that has never done a research-based report, I thought I did a decent job. Overall, I need to learn a lot more and it gave me a ton of experience. I hope that I can remember down the road.

Reference page

Sharma. (2021) Gaussian Naïve Bayes Implementation in pythom Sklearn [Online]. Available: Gaussian Naive Bayes Implementation in Python Sklearn - MLK - Machine Learning Knowledge

Garfieldliang (2015) Video Games Analysis [Online]. Available: video games analysis | Kaggle

Radovanovic (2022) An introduction Guide to Machine Learning [Online]. Available: Sklearn - An Introduction Guide to Machine Learning - AlgoTrading101 Blog