

VIDEO GAMES METACRITIC SCORE

PREDICTION MODELING FOR METACRITIC SCORE



DATA SET

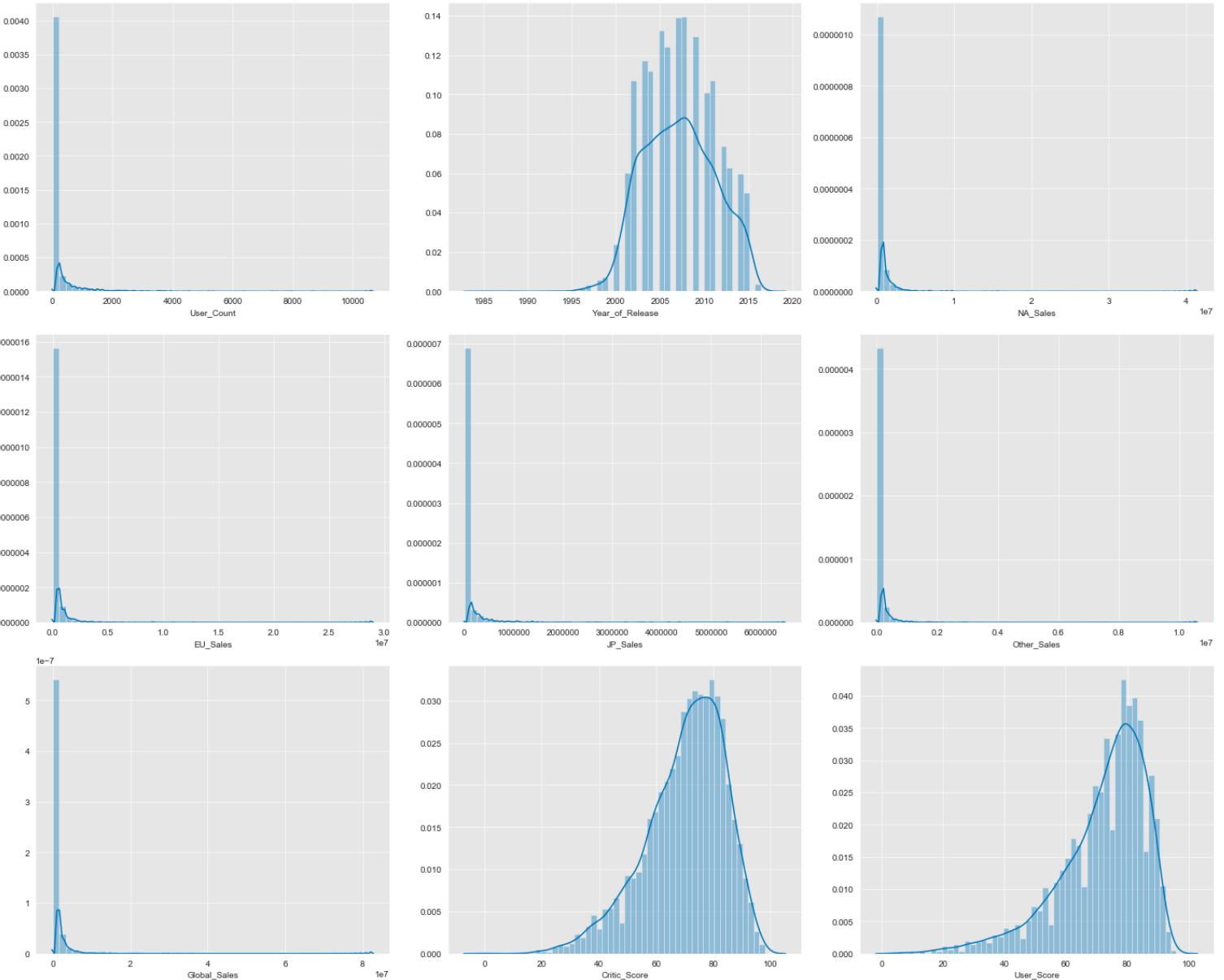
I use a data set from Kaggle and I also did some api calls for additional information on the Platforms for the games

Name	Platform	Year_of_Release	Genre	Publisher	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	User_Score	Developer	Rating
Monster Hunter Stories	3DS	2017.0	Action	Capcom	0.00	0.00	0.28	0.00	0.28	79.0	8.2	Capcom	E10+
The Elder Scrolls V: Skyrim	PS4	2016.0	Role-Playing	Bethesda Softworks	0.48	0.59	0.04	0.21	1.33	81.0	7.1	Bethesda Game Studios	M
Rise of the Tomb Raider	PS4	2016.0	Adventure	Square Enix	0.23	0.53	0.04	0.14	0.94	88.0	8.1	Crystal Dynamics, Nixxes Software	M
MLB 16: The Show	PS4	2016.0	Action	Sony Computer Entertainment	0.72	0.00	0.00	0.17	0.89	85.0	7.3	SCEA San Diego Studios	E
Doom (2016)	XOne	2016.0	Shooter	Bethesda Softworks	0.44	0.25	0.00	0.07	0.76	87.0	8.4	id Software	M
The Elder Scrolls V: Skyrim	XOne	2016.0	Role-Playing	Bethesda Softworks	0.41	0.27	0.00	0.07	0.74	82.0	7.5	Bethesda Game Studios	M



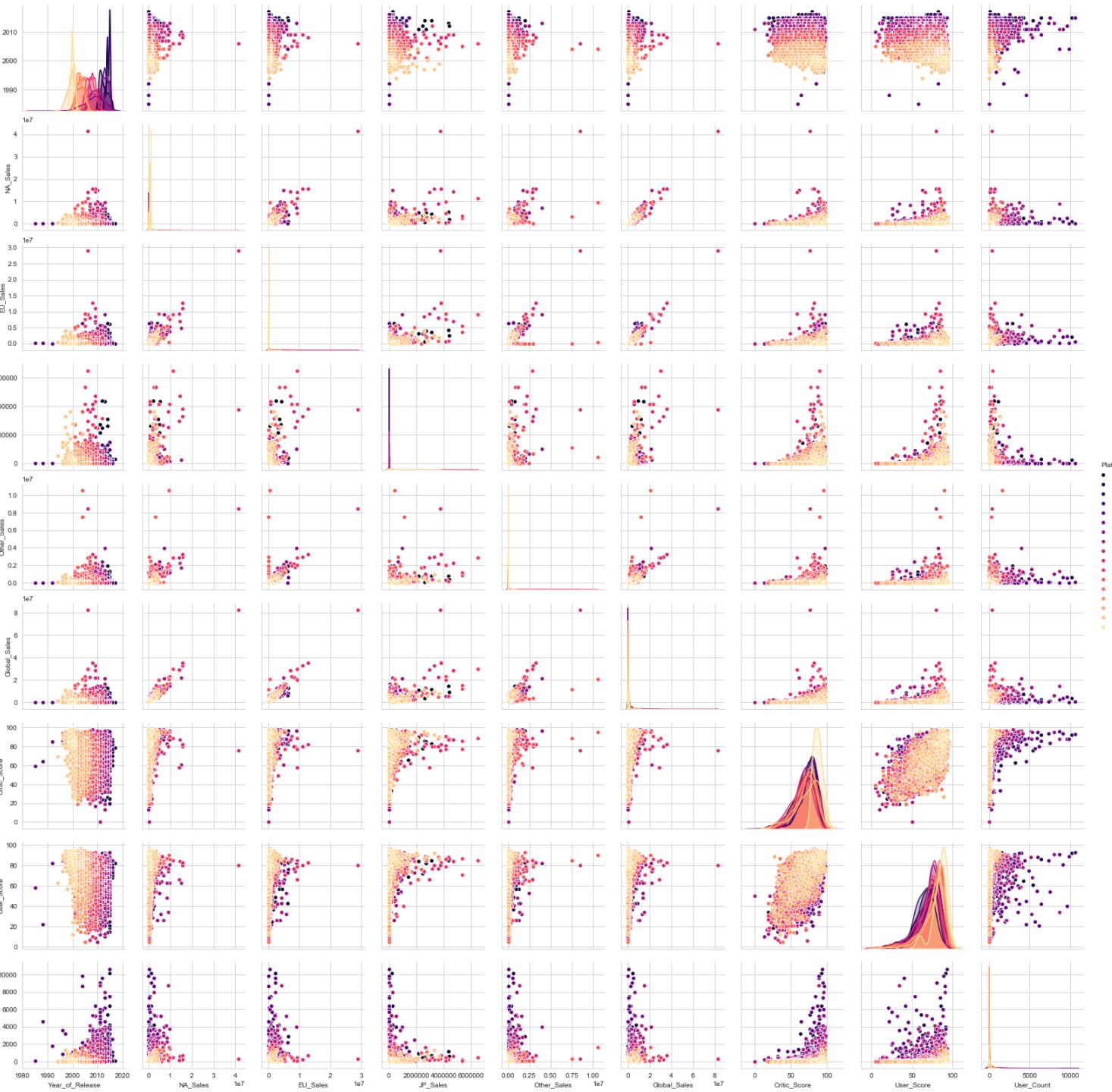
DISTRIBUTION

Plotted all numerical data to analyze the distribution of each category separately.



DISTRIBUTION PT2

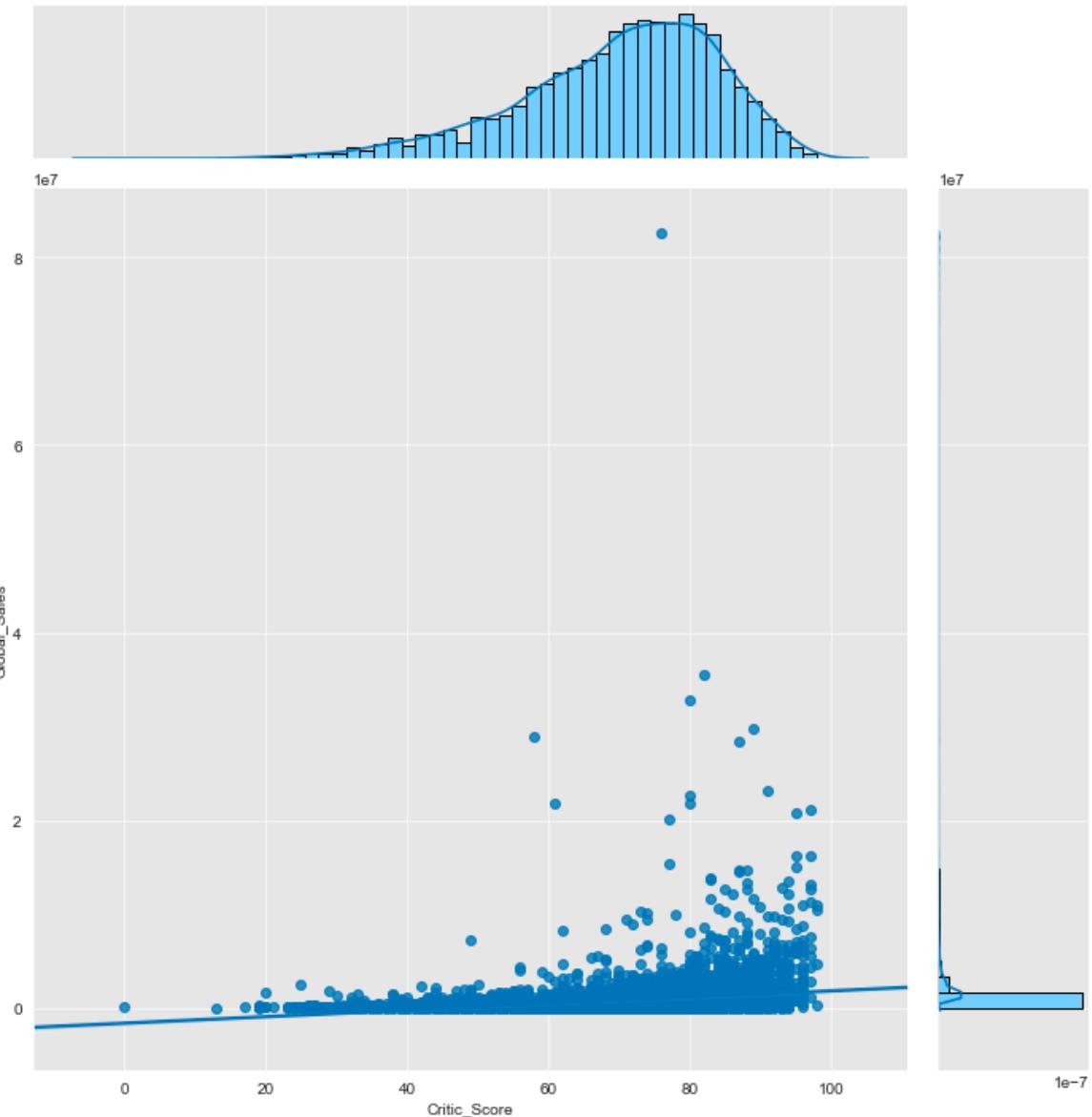
Plotted all numerical data to analyze the distribution of each category against the other categories. To see the relationships between categories with a color hue showing the different Platforms for the games. I then made dummy variables for non numerical values.



VISUALIZING BY CRITIC SCORE VS SALES

Fig. 1 Shows the relationship between Metacritic score and Global sales of the game.

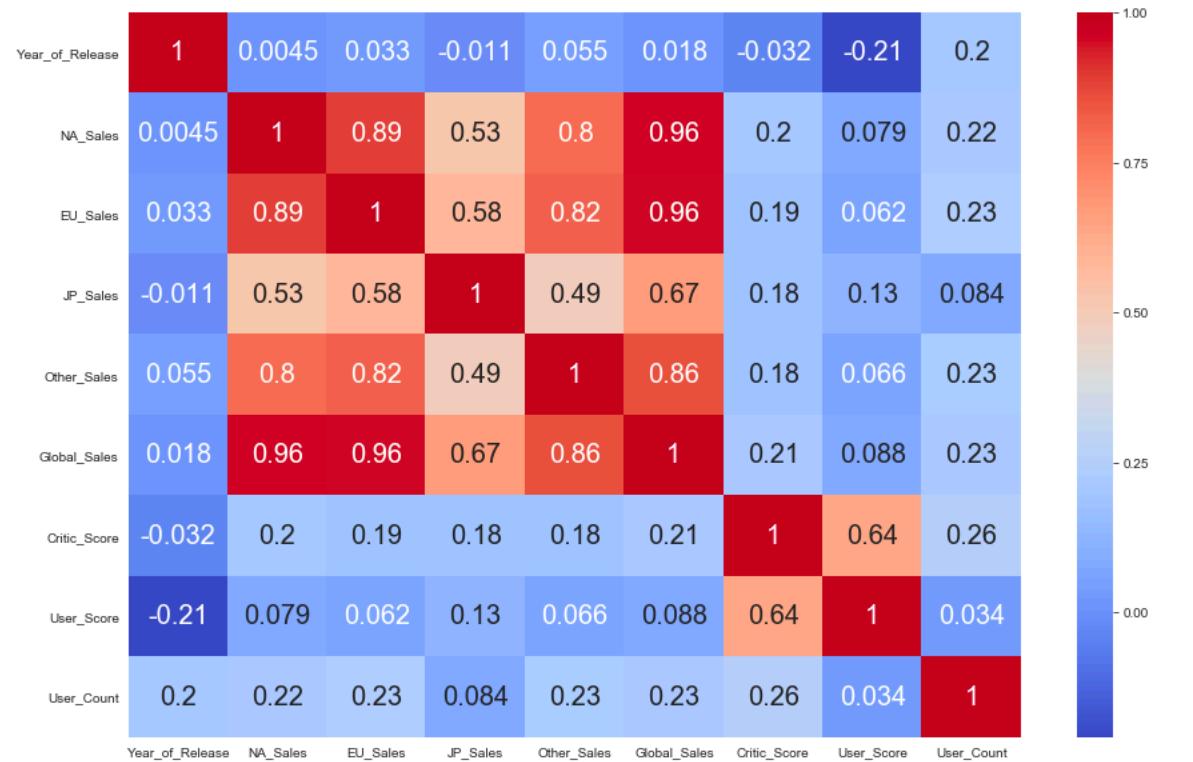
Based on this graph you can say there seems to be a slight correlation.



CORRELATION

Shows correlation between each numerical features against all others

	Year_of_Release	NA_Sales	EU_Sales	JP_Sales	Other_Sales	Global_Sales	Critic_Score	User_Score	User_Count
Year_of_Release	1.000000	0.004540	0.033129	-0.010530	0.055233	0.018044	-0.031577	-0.208178	0.199946
NA_Sales	0.004540	1.000000	0.890921	0.528168	0.800006	0.963701	0.202548	0.079322	0.219646
EU_Sales	0.033129	0.890921	1.000000	0.583327	0.817553	0.959271	0.188492	0.061604	0.234283
JP_Sales	-0.010530	0.528168	0.583327	1.000000	0.492099	0.673409	0.178593	0.126806	0.083786
Other_Sales	0.055233	0.800006	0.817553	0.492099	1.000000	0.859098	0.183351	0.066456	0.226771
Global_Sales	0.018044	0.963701	0.959271	0.673409	0.859098	1.000000	0.212058	0.087559	0.225325
Critic_Score	-0.031577	0.202548	0.188492	0.178593	0.183351	0.212058	1.000000	0.636489	0.259023
User_Score	-0.208178	0.079322	0.061604	0.126806	0.066456	0.087559	0.636489	1.000000	0.033650
User_Count	0.199946	0.219646	0.234283	0.083786	0.226771	0.225325	0.259023	0.033650	1.000000



CLASSIFICATION FOR CRITIC'S SCORE

I classified three categories for ranges of Metacritic scores:

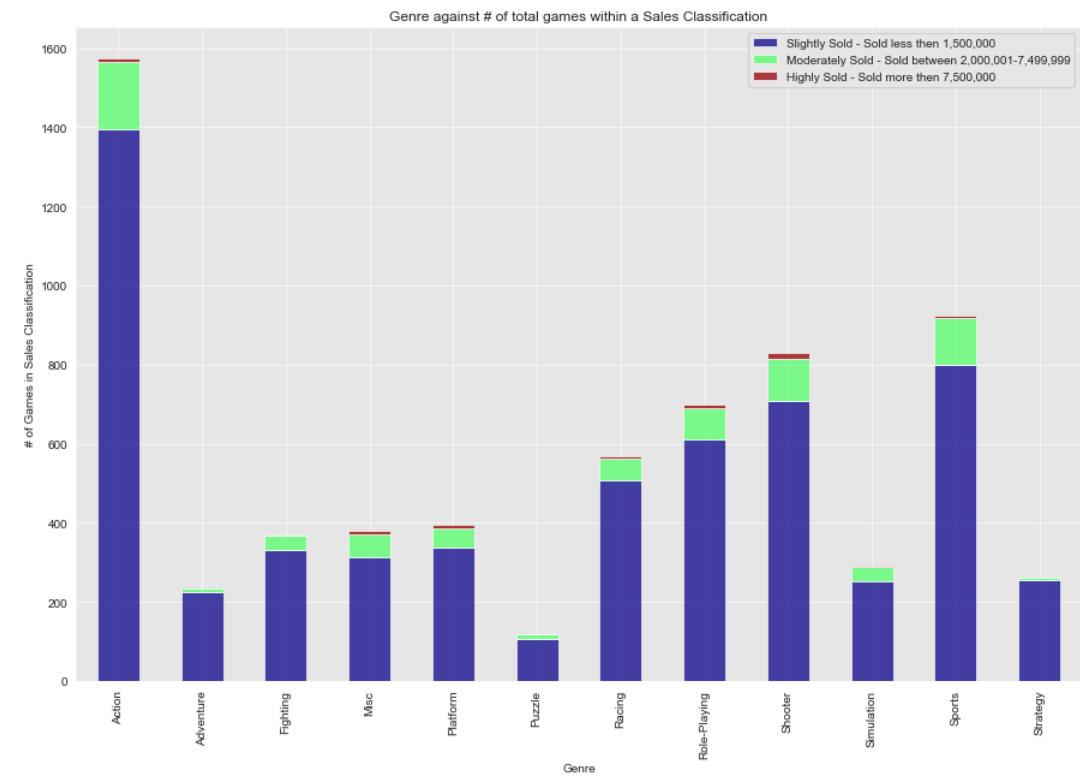
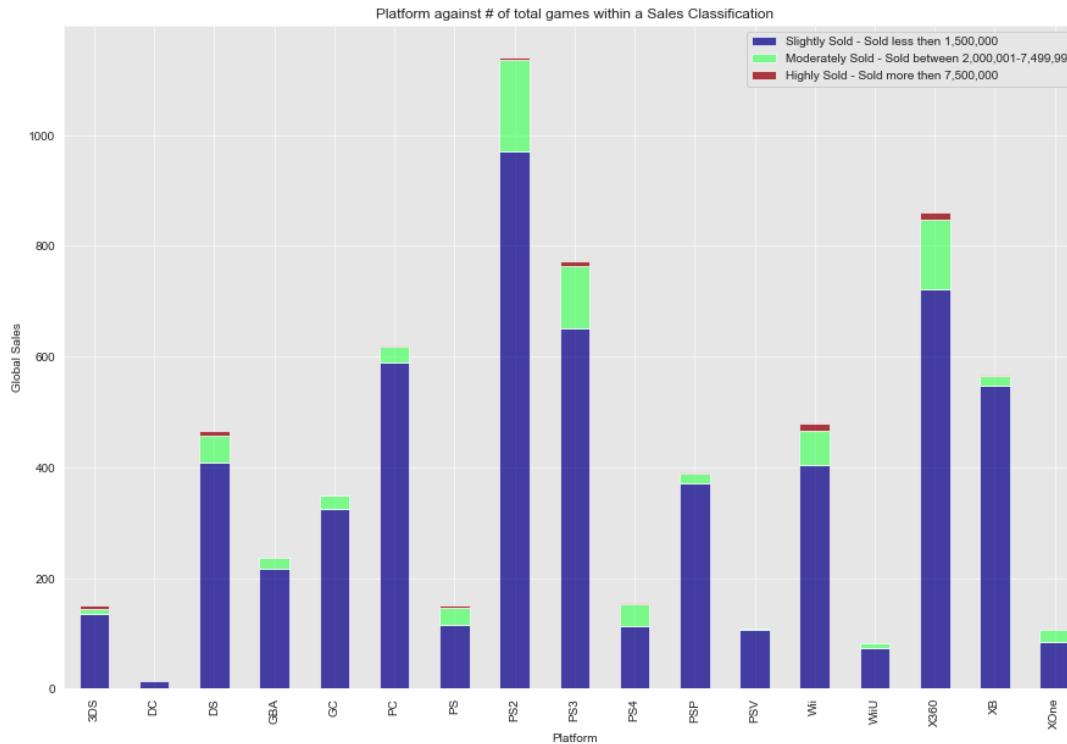
- 3 if the game had a score higher then 7.5
- 2 if the game had a score between 5.1-7.4
- 1 if the game had a score lower then 5.0

Also had to feature engineer to get a more accurate predictions

(0.0, '3DO')
(0.0, '505 Games')
(0.0, 'BAM! Entertainment')
(0.0, 'Home Entertainment Suppliers')
(0.0, 'Ignition Entertainment')
(0.0, 'Jaleco')
(0.0, 'Kalypso Media')
(0.0, 'Kemco')
(0.0, 'Lighthouse Interactive')
(0.0, 'MTV Games')
(0.0, 'Majesco Entertainment')
(0.0, 'Nihon Falcom Corporation')
(0.0, 'Scholastic Inc.')
(0.0, 'Square')
(0.0, 'Sting')
(0.0, 'Success')
(0.0, 'Tru Blu Entertainment')
(0.0, 'Universal Interactive')
(0.0, 'Vivendi Games')
(2.495366766927972e-06, 'SquareSoft')
(2.826891156192845e-06, 'PopCap Games')
(3.103632207517172e-06, 'Spike')
(4.368634599476728e-06, 'NewKidCo')
(4.827269785531538e-06, 'Agatsuma Entertainment')
(5.379463323170304e-06, 'Ubisoft')
(5.631814749786764e-06, 'bitComposer Games')
(8.000220531431693e-06, 'Midas Interactive Entertainment')
(8.301802182086178e-06, 'RedOctane')
(1.2336356118580708e-05, 'Mercury Games')
(1.5178568712680928e-05, 'Nintendo')
(1.5294966415824597e-05, 'Interplay')
(1.6110111689289514e-05, 'Slightly Mad Studios')
(1.6631369824254693e-05, 'D3Publisher')
(1.778931583930655e-05, 'DTP Entertainment')
(2.2615129249542536e-05, 'Hudson Soft')
(2.74318090435295e-05, 'Mastertronic')

Features with no significance I removed them

VISUALIZING CLASSIFIED CRITIC SCORES – AGAINST PLATFORMS & GENRES



MODELS

- I ran my data through DummyClassifier ,KNN , Logistic Regression, Random Forest, Support Vector Classification, Decision Tree and Bernoulli NB models. And it seemed KNN worked the best.

	Model	Accuracy Score	Cross Validation Accuracy Score	Precision Score	Recall Score	F1 Score
0	Baseline	0.407321	0.409534	0.415020	0.407321	0.410516
1	K_nearest Neighbor = 5	0.743769	0.628279	0.745829	0.743769	0.738570
2	Logistic Regression	0.682243	0.691396	0.686127	0.682243	0.675784
3	Random Forest	0.678349	0.650507	0.675365	0.678349	0.672418
4	Support Vector Classification	0.684579	0.556383	0.689607	0.684579	0.677887
5	Decision Tree	0.637850	0.562559	0.640579	0.637850	0.638023
6	BernoulliNB	0.582555	0.556383	0.565494	0.582555	0.566475

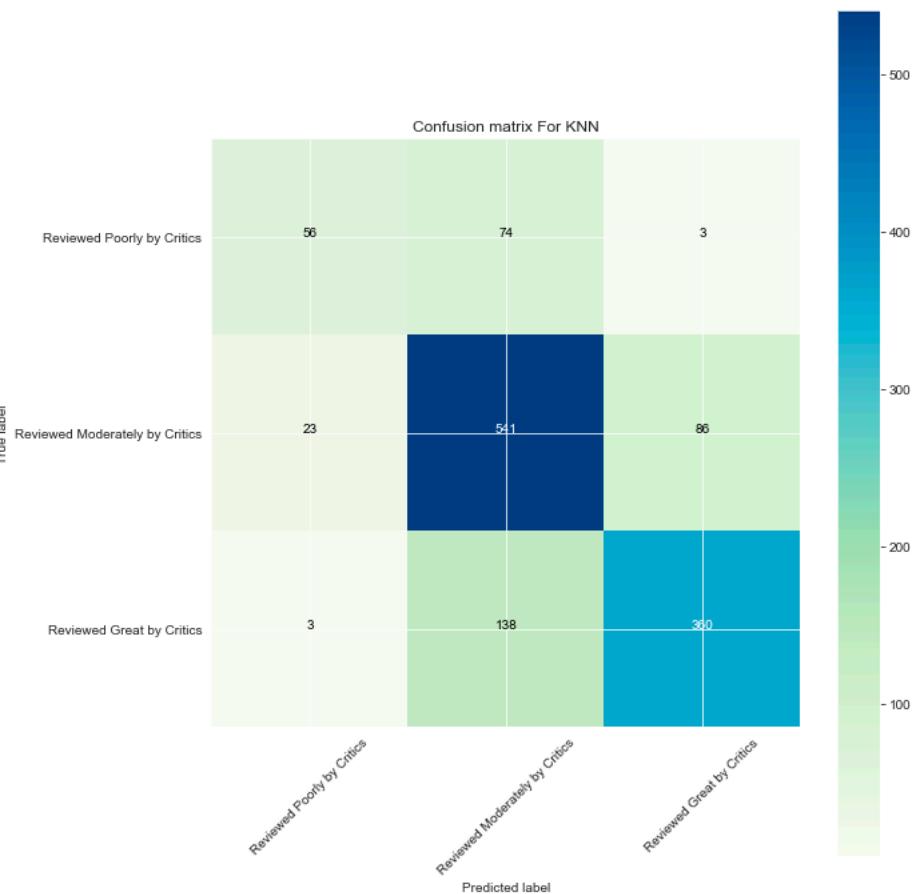


FINAL CONCLUSION

K_nearest Neighbor = 5 Regression train accuracy is 0.6439545758183033
K_nearest Neighbor = 5 Regression test accuracy is 0.7437694704049844
K_nearest Neighbor = 5 Prediction accuracy is: 0.7437694704049844
K_nearest Neighbor = 5 Mean cross validation is: 0.6282787768110097
K_nearest Neighbor = 5 Precision score is: 0.7458290585870638
K_nearest Neighbor = 5 Prediction recall score is: 0.7437694704049844
K_nearest Neighbor = 5 Prediction f1 score is: 0.7385703525214701

Manually selected a sampling of Actual vs Predicted Values from KNN and calculated Accuracy

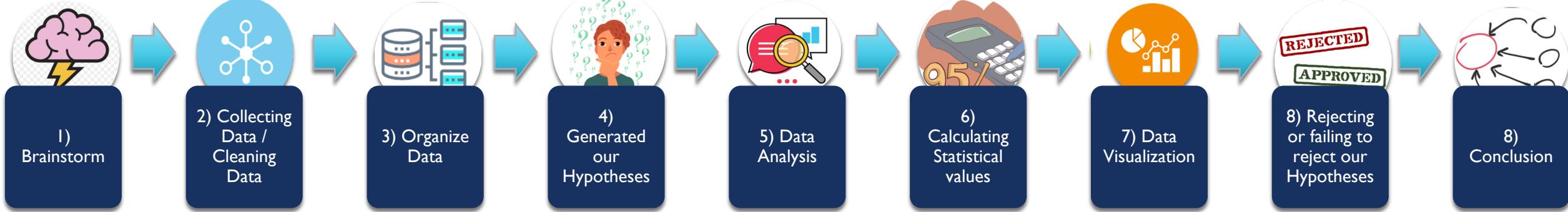
True: [2 2 3 2 1 3 1 3 2 2 3 3 3 2 3 1 1 2 2 3 3 2 3 2 2]
Pred: [2 2 3 2 2 3 1 3 2 2 2 3 3 3 2 3 1 3 2 2 2 3 2 2 2 2]
Accuracy from 1st 25 values is 80.0%



APPENDIX

Other useful information

WORKFLOW :



LIMITATIONS

Time - more time to clean more data, for more parameters to filter through
Collecting data - to collect data from more sources to collect more game data
Running more statistical tests/models
Not having a partner...

