

Game Sales Prediction

Capstone project



Tristan Foo, Data Scientist

The Agenda

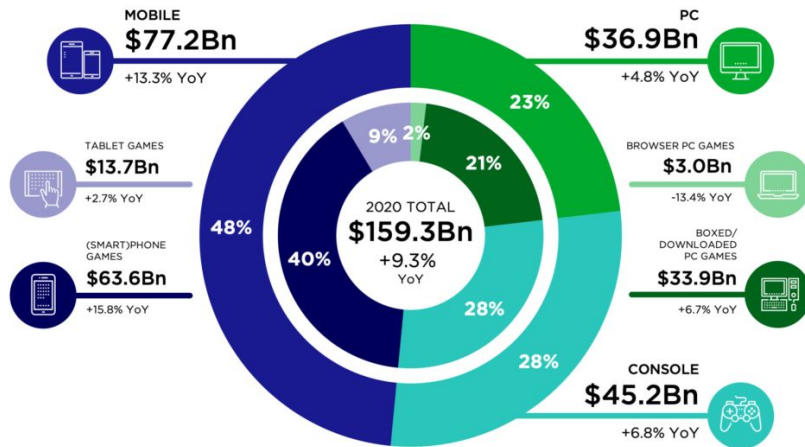
- The Domain Opportunity
 - Overview
 - Business Question vs Data Question
- Business and Data Process
- Delivery
- Summary : The Next Steps, and Conclusion

Domain Opportunity - Overview



2020 Global Games Market

Per Device & Segment With Year-on-Year Growth Rates



Source: ©Newzoo | 2020 Global Games Market Report | April Update
newzoo.com/globalgamesreport

The gaming industry is evaluated to generate \$159.3 Bn in 2020, out of which combined PC and consoles revenue has 52% of it (82.1 Bn)

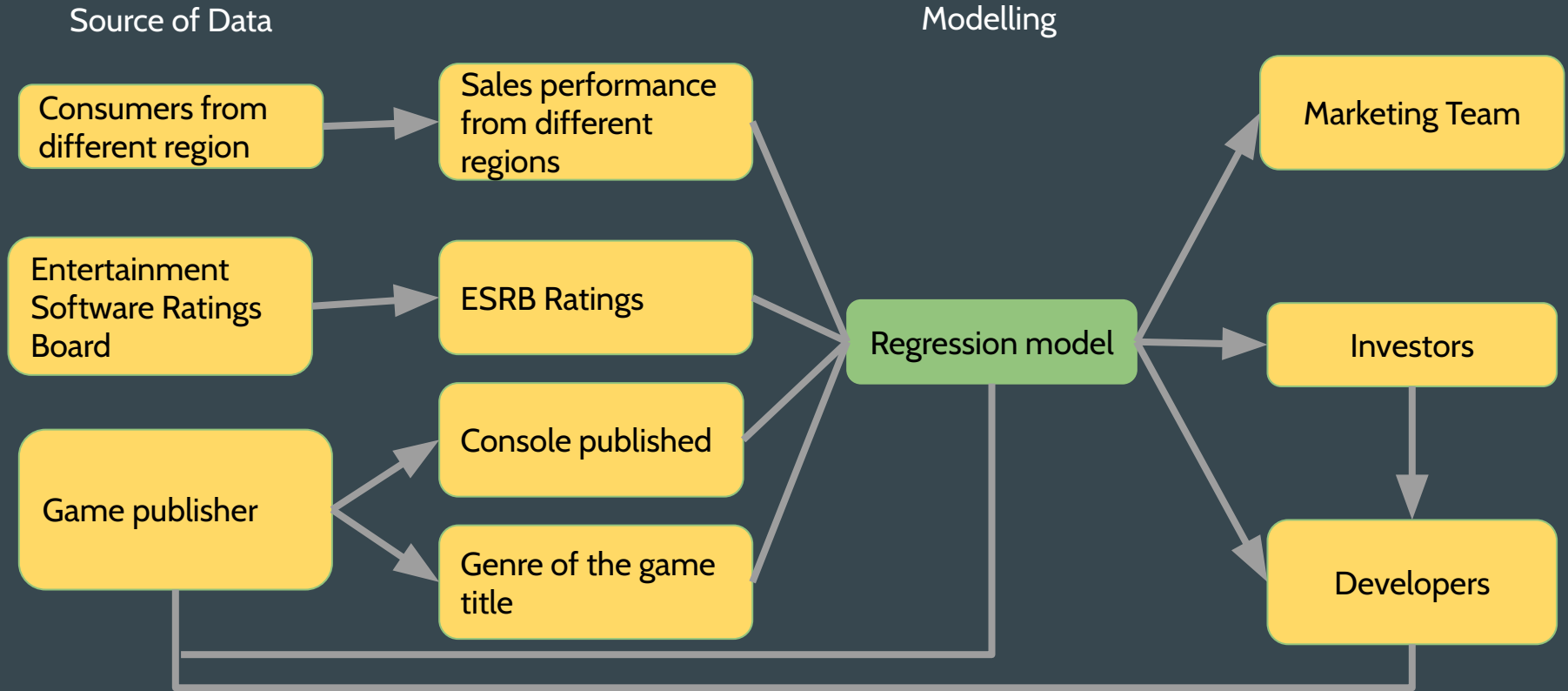
Console games projected to grow 6.8% grow year on year.

Increased competition meant greater demand for new strategies to increase market share for console gaming

Domain - The Opportunity

- Business Question: Can we tell if which games would generate the most sales ?
- The industry is projected to hit 218 billion in 2024.
- Stakeholders:
 - Investors : VCs, crowdsourcing etc
 - Developers: Game studios
 - Game publishers
- Data Question: Based on the past data of consoles, genres and ESRB ratings, can we predict the sale number of a game title using various regression models ?

Business and data process

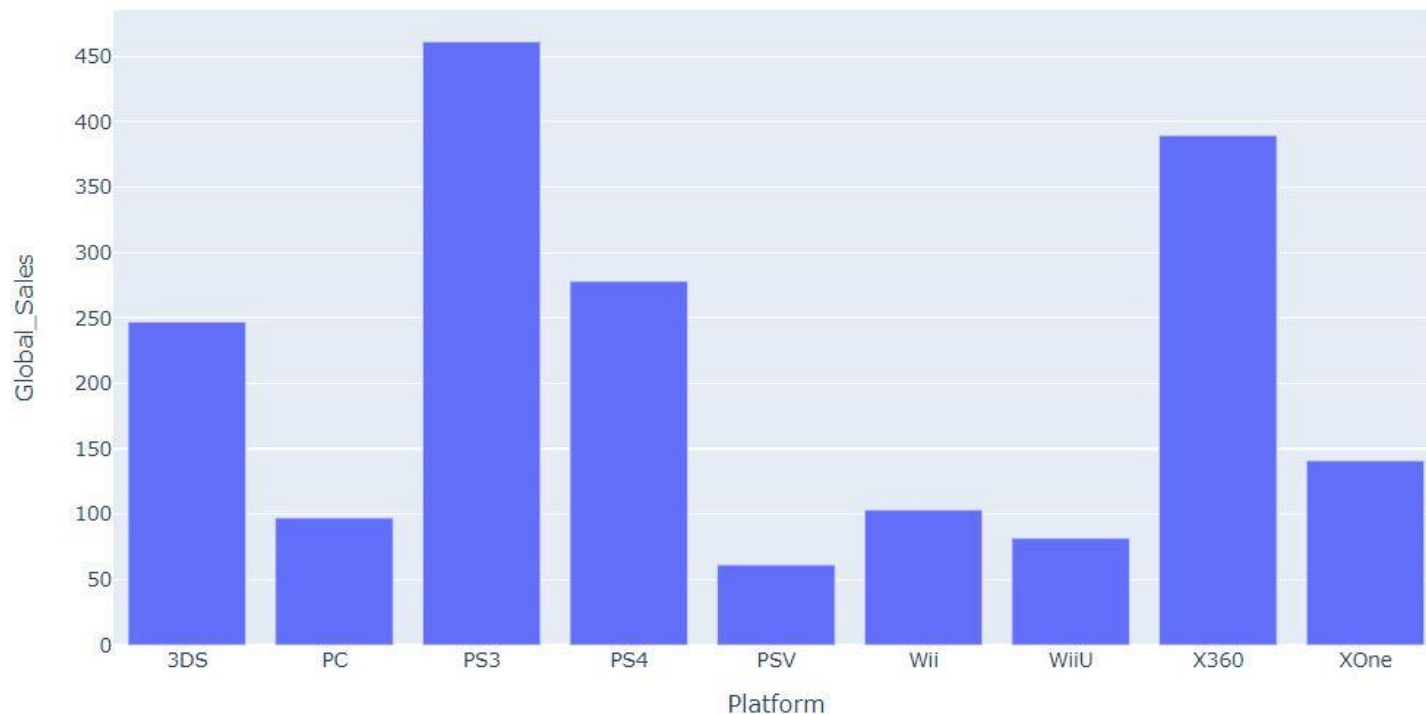


Data Process

- Physical copies sales performance of 16k+ game titles over the North
- America(NA), Europe (EU), Japan(JPN) and other regions, obtained from vgchartz.com
- Features include the consoles (Platform), Genre and the ESRB ratings of the title
- Sales numbers are recorded in millions (USD)
- Models uses : CatBoost Regressor, LGBM Regressor, XGB Regressor, RandomForest Regressor

Delivery - The Top Performers : Platform

Total sales of all time in the most important platforms (Millions)



Delivery - The Top Performers : Platform

	NA_Sales	JP_Sales	EU_Sales	Other_Sales	Global_Sales
Platform					
PS3	168.12	49.22	173.82	70.01	461.25
X360	227.85	3.41	125.01	33.17	389.63
PS4	96.79	14.28	123.70	43.36	278.07
3DS	78.55	97.29	58.50	12.61	247.04
XOne	83.18	0.34	45.61	11.92	141.01

PS 3 and XBOX 360 performed well in NA and EU as Sony and Microsoft have large presence there

While 3DS performed better in Japan as Nintendo is a popular local brand

Delivery - The Top Performers : Genre

Total sales by genre (Millions)



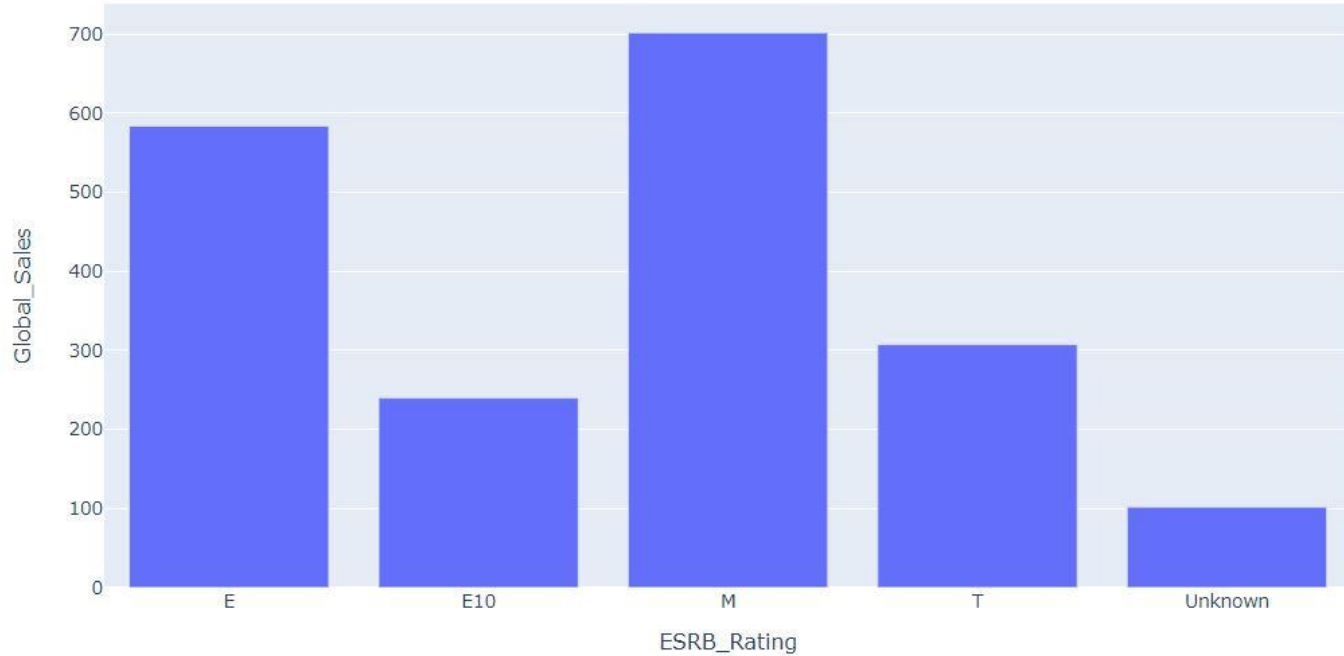
Delivery - The Top Performers : Genre

	NA_Sales	JP_Sales	EU_Sales	Other_Sales	Global_Sales
Genre					
Action	231.22	62.08	203.75	66.50	563.72
Shooter	184.70	11.45	142.60	46.73	385.68
Role-Playing	79.30	76.81	60.17	19.29	235.68
Sports	105.90	10.95	86.95	28.97	232.78
Misc	70.78	18.08	40.99	12.84	142.69

Interestingly
Western
markets prefer
Action games.

While Asian
markets like
Japan prefer
Role-playing
games

Delivery - The Top Performers : ESRB Ratings

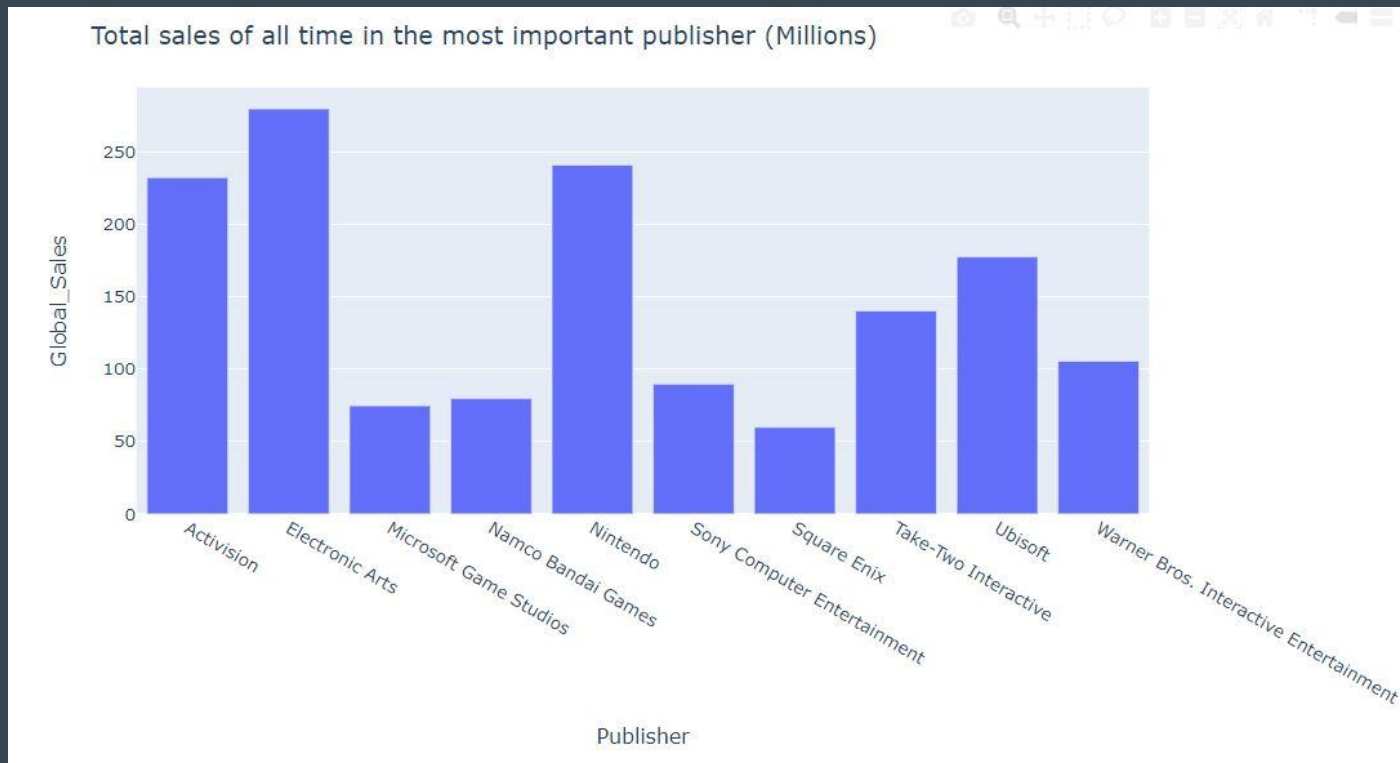


Delivery - The Top Performers : ESRB Ratings

	NA_Sales	JP_Sales	EU_Sales	Other_Sales	Global_Sales
ESRB_Rating					
E	253.70	60.18	209.04	60.47	583.65
E10	108.72	15.76	88.61	26.75	239.83
M	319.71	27.86	265.33	88.75	701.60
T	132.23	43.10	97.43	34.31	307.35
Unknown	0.29	93.61	6.65	1.04	101.78

‘Unknown’ rated games sell higher in Japan because Japan has its own regulatory rating systems

Delivery - The Top Performers : Publisher



Delivery - The Top Performers : Publisher

	NA_Sales	JP_Sales	EU_Sales	Other_Sales	Global_Sales
Publisher					
Electronic Arts	116.49	3.40	122.31	37.48	279.74
Nintendo	88.63	73.02	64.36	14.94	240.91
Activision	118.75	3.42	82.21	27.72	232.26
Ubisoft	81.16	2.43	71.95	22.00	177.53
Take-Two Interactive	74.46	2.29	45.54	17.92	140.20

Nintendo, which had presence across the 3 markets had stable performance

Western publishers like EA and Activision only performed well in their domestic markets

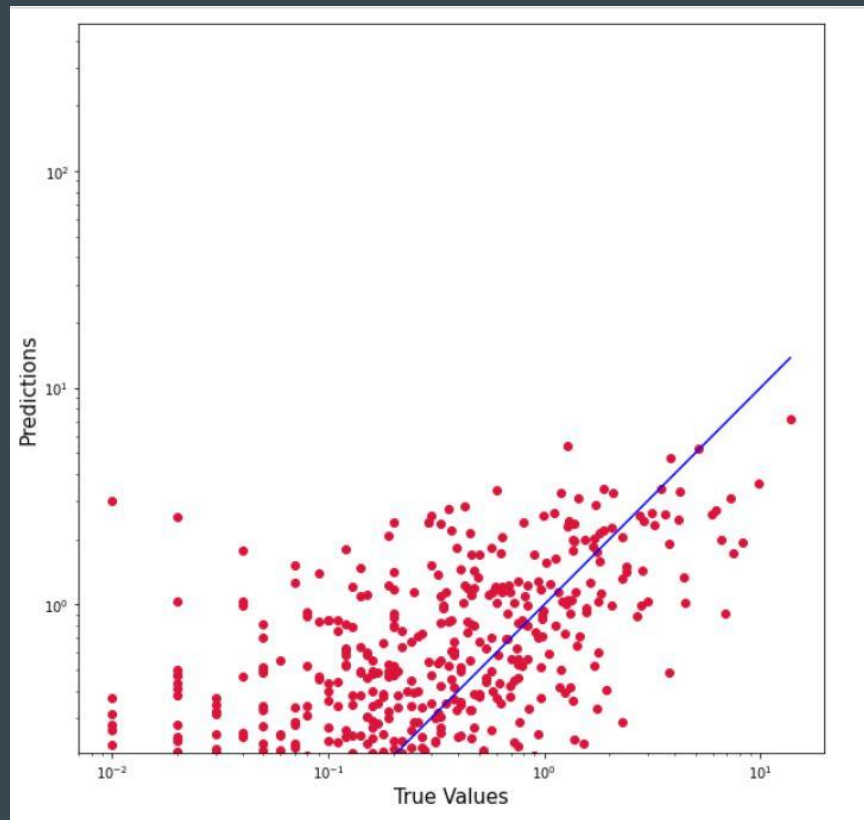
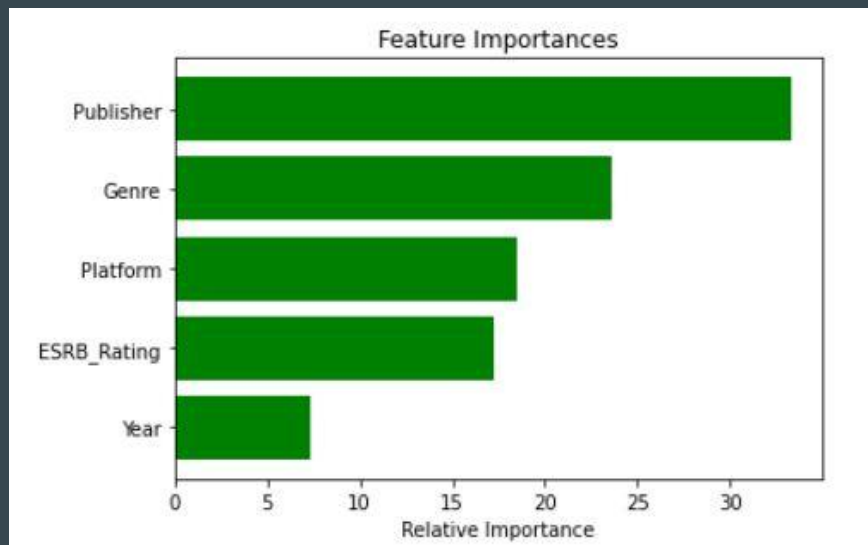
Delivery - Comparing models

Model	Training R sq	Testing R sq	RMSE
CatBoost MAE = 0.393 MSE= 0.719	0.554	0.404	0.832
LGBM MAE = 0.484 MSE= 1.071	0.232	0.079	1.035
XGB MAE = 0.50 MSE= 1.222	0.263	-0.0517	1.106
RF Regressor MAE = 0.500 MSE= 1.222	0.263	-0.0618	1.111

Under XGB and RF Regressor, there is no constant term, hence the negative R-square

After comparing the models, we can conclude that CatBoost is the best model for the dataset (Highest R sq with lowest RMSE)

Delivery



Summary - The Next Steps: Data

Accuracy of the model can be further enhanced by :

- Having access to the most updated and comprehensive data and use it
- Collect other data to interact with the model (Monthly Active User, Online downloads, User play time etc)
- Collection of data can be done better now with people having access to online servers and shops

Summary - The Next Steps: Business

Design and Development	Marketing
Know your TA	International : Can adopt similar marketing strategies to NA and EU
If TA is in niche market, work closely with dominant publishers in the target market	Niche (Japan): Adopt a more tailored strategy and adapt to local tastes accordingly
Introduce microtransaction (Use with caution)	Moving sales to online
Cross Platform playing	Early Access

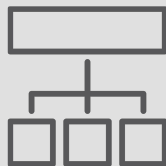
Summary



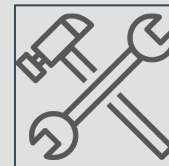
Can we tell what kind of games would generate the most sales?



Can we predict game title sales?

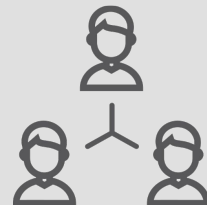
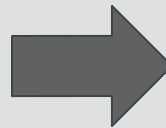


Segment into different markets



Identify the right features for feature engineering

Build and evaluate model



Provide insights to stakeholders and implement strategies

Appendix

- Dataset :

- <https://data.world/julienf/video-games-global-sales-in-volume-1983-2017>

- Reference :

- <https://machinelearningmastery.com/gradient-boosting-with-scikit-learn-xgboost-lightgbm-and-catboost/>
 - <https://www.analyticsvidhya.com/blog/2017/06/which-algorithm-takes-the-crown-light-gbm-vs-xgboost/>
 - <https://www.udemy.com/course/catboost-vs-xgboost-a-gentle-introduction/learn/lecture/17756724#overview>
 - <https://medium.com/analytics-vidhya/regression-and-performance-metrics-accuracy-precision-rmse-and-what-not-223348cfcafe>
 - <https://newzoo.com/insights/articles/newzoo-games-market-numbers-revenues-and-audience-2020-2023/>
 - <https://catboost.ai/>

THANK YOU