Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

Artificial intelligence Lecture Orange Tool Practical No. 2

In this Practical Work, I have worked with the “Orange” tool to simulate machine learning on a dataset I found online with unsupervised and supervised learning and analyze my dataset afterwards.

My database is about mobile online games with their ratings and genres seperated showing the mobile game players rates on them.

Info about my dataset:

Author: Dhruvil Dave

Citation: @misc{dhruvil dave_2021,

```
        title={Top Games on Google Play Store},
        url={https://www.kaggle.com/dsv/2316698},
        DOI={10.34740/KAGGLE/DSV/2316698},
        publisher={Kaggle},
        author={Dhruvil Dave},
        year={2021}
}
```

Sources: https://play.google.com/store/apps/top/category/GAME

Licence: Database: Open Database, Contents: © Original Authors

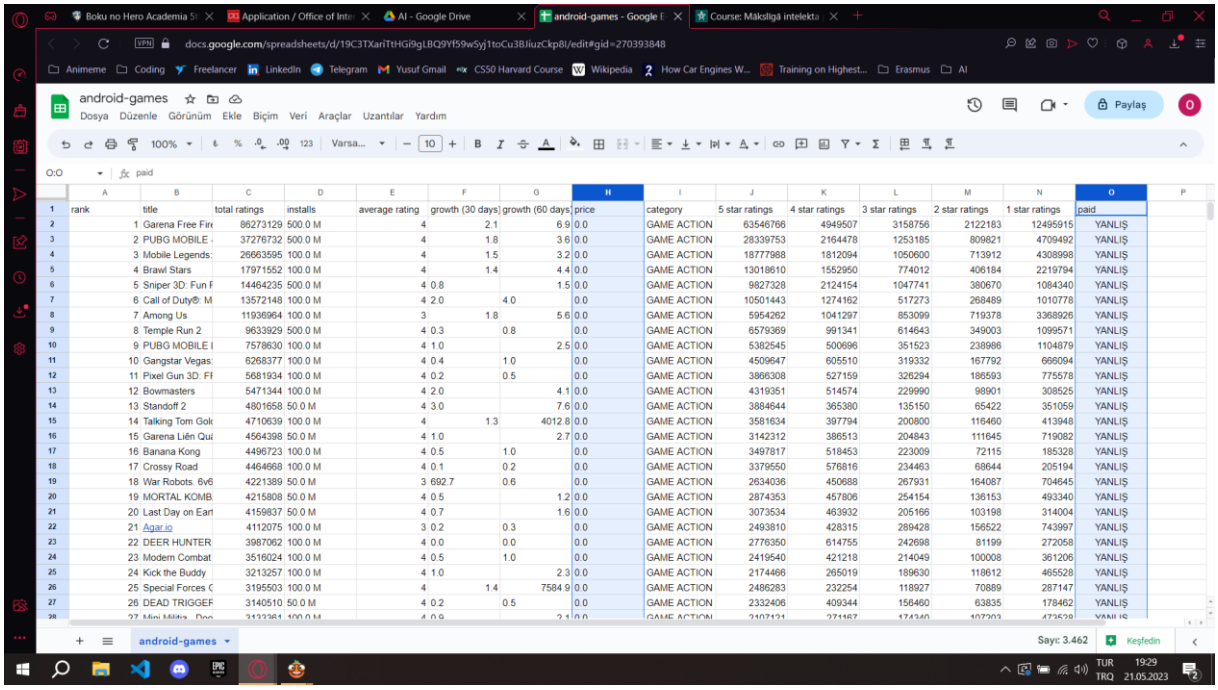
All the information above can be found from the link:
https://www.kaggle.com/datasets/dhruvildave/top-play-store-games

!!!Orange tool Workflow was not made as a section since diagrams’ change were given in between the sections below already.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

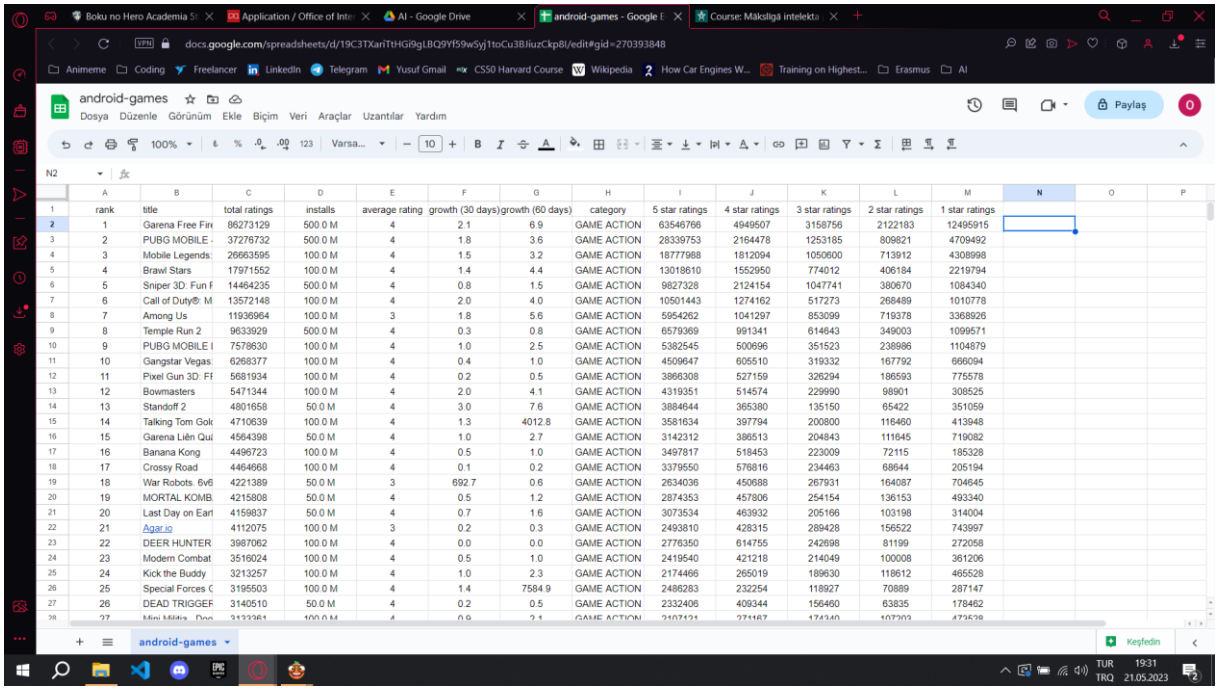
1) Pre Processing / Exploring Data Section:

First of all I started with Pre-Processing and Exploring of the data:



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rank	title	total ratings	installs	average rating	growth (30 days)	growth (60 days)	price	category	5 star ratings	4 star ratings	3 star ratings	2 star ratings	1 star ratings	paid	
2	1	Garena Free Fire	86273129	500.0 M	4	2.1	6.9	0.0	GAME ACTION	63546766	4949507	3158756	2122183	12495915	YANLIŞ	
3	2	PUBG MOBILE -	37276732	500.0 M	4	1.8	3.6	0.0	GAME ACTION	28339753	2164478	1253185	809821	4709492	YANLIŞ	
4	3	Mobile Legends:	26663595	100.0 M	4	1.5	3.2	0.0	GAME ACTION	18777988	1812094	1050600	713912	4308998	YANLIŞ	
5	4	Brawl Stars	17971552	100.0 M	4	1.4	4.4	0.0	GAME ACTION	13018610	1552950	774012	406184	2219794	YANLIŞ	
6	5	Sniper 3D: Fun F	14464235	500.0 M	4	0.8	1.5	0.0	GAME ACTION	9827328	2124154	1047741	380670	1084340	YANLIŞ	
7	6	Call of Duty®: M	13572148	100.0 M	4	2.0	4.0	0.0	GAME ACTION	10501443	1274162	517273	268489	1010778	YANLIŞ	
8	7	Among Us	11936964	100.0 M	3	1.8	5.6	0.0	GAME ACTION	5954262	1041297	853099	719378	3368926	YANLIŞ	
9	8	Temple Run 2	9633929	500.0 M	4	0.3	0.8	0.0	GAME ACTION	6579369	991341	614643	349003	1099571	YANLIŞ	
10	9	PUBG MOBILE I	7578630	100.0 M	4	1.0	2.5	0.0	GAME ACTION	5382545	500696	351523	238986	1104879	YANLIŞ	
11	10	Gangstar Vegas:	6268377	100.0 M	4	0.4	1.0	0.0	GAME ACTION	4509647	605510	319332	167792	666094	YANLIŞ	
12	11	Pixel Gun 3D: FF	5681934	100.0 M	4	0.2	0.5	0.0	GAME ACTION	3866308	527159	326294	186593	775578	YANLIŞ	
13	12	Bowmasters	5471344	100.0 M	4	2.0	4.1	0.0	GAME ACTION	4319351	514574	229990	98901	308525	YANLIŞ	
14	13	Standoff 2	4801658	50.0 M	4	3.0	7.6	0.0	GAME ACTION	3884644	365380	135150	65422	351059	YANLIŞ	
15	14	Talking Tom Gol	4710639	100.0 M	4	1.3	4012.8	0.0	GAME ACTION	3581634	397794	200800	116460	413948	YANLIŞ	
16	15	Garena Liên Quâ	4564398	50.0 M	4	1.0	2.7	0.0	GAME ACTION	3142312	386513	204843	111645	719082	YANLIŞ	
17	16	Banana Kong	4496723	100.0 M	4	0.5	1.0	0.0	GAME ACTION	3497817	518453	223009	72115	185328	YANLIŞ	
18	17	Crossy Road	4464668	100.0 M	4	0.1	0.2	0.0	GAME ACTION	3379550	576816	234463	68644	205194	YANLIŞ	
19	18	War Robots. 6v6	4221389	50.0 M	3	692.7	0.6	0.0	GAME ACTION	2634036	450688	267931	164087	704645	YANLIŞ	
20	19	MORTAL KOMB	4215808	50.0 M	4	0.5	1.2	0.0	GAME ACTION	2874353	457806	254154	136153	493340	YANLIŞ	
21	20	Last Day on Eart	4159837	50.0 M	4	0.7	1.6	0.0	GAME ACTION	3073534	463932	205166	103198	314004	YANLIŞ	
22	21	Agar.io	4112075	100.0 M	3	0.2	0.3	0.0	GAME ACTION	2493810	428315	289428	156522	743997	YANLIŞ	
23	22	DEER HUNTER	3987062	100.0 M	4	0.0	0.0	0.0	GAME ACTION	2776350	614755	242698	81199	272058	YANLIŞ	
24	23	Modern Combat	3516024	100.0 M	4	0.5	1.0	0.0	GAME ACTION	2419540	421218	214049	100008	361206	YANLIŞ	
25	24	Kick the Buddy	3213257	100.0 M	4	1.0	2.3	0.0	GAME ACTION	2174466	265019	189630	118612	465528	YANLIŞ	
26	25	Special Forces G	3195503	100.0 M	4	1.4	7584.9	0.0	GAME ACTION	2486283	232254	118927	70889	287147	YANLIŞ	
27	26	DEAD TRIGGER	3140510	50.0 M	4	0.2	0.5	0.0	GAME ACTION	2332406	409344	156460	63835	178462	YANLIŞ	

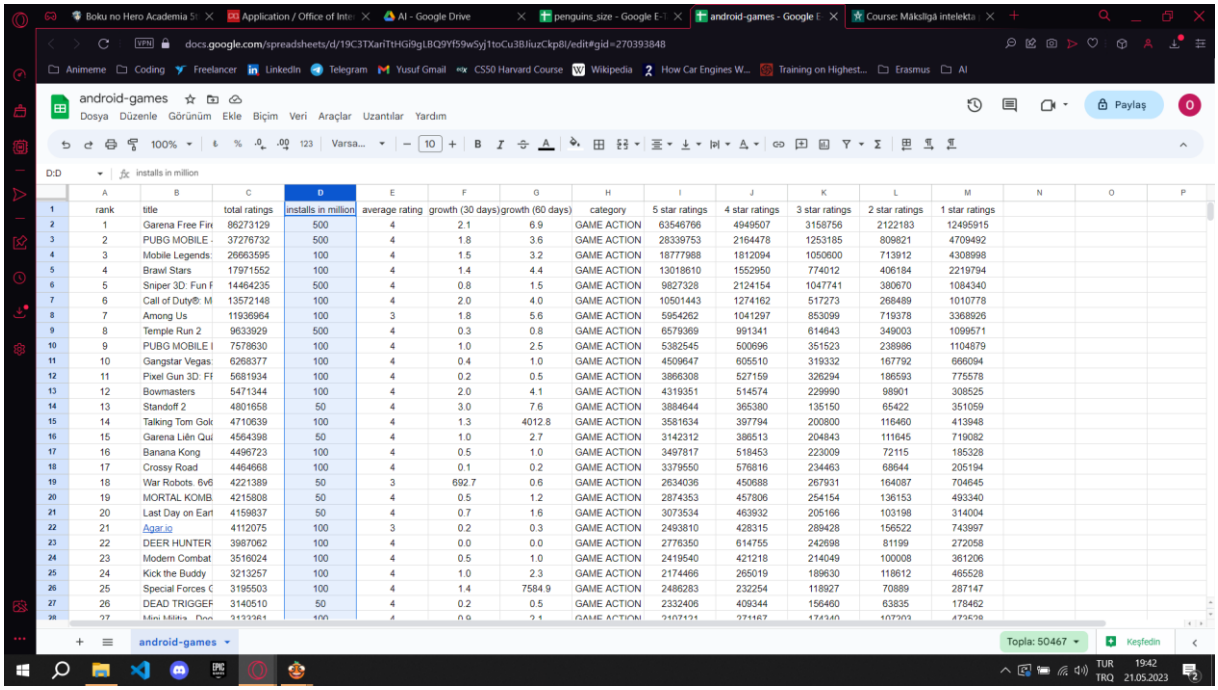
In here I decided to delete “price” and “paid” sections because firstly only 7 games out of 1700 were paid and these sections looked unnecessary to keep in the dataset.



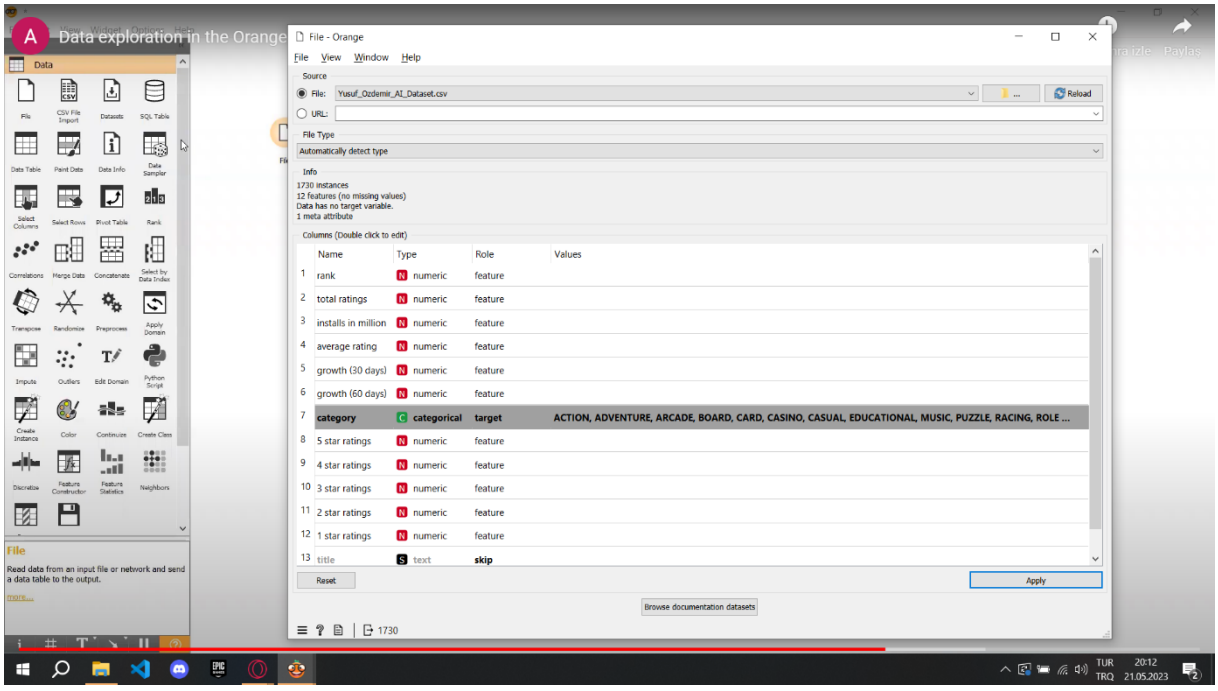
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	rank	title	total ratings	installs	average rating	growth (30 days)	growth (60 days)	category	5 star ratings	4 star ratings	3 star ratings	2 star ratings	1 star ratings			
2	1	Garena Free Fire	86273129	500.0 M	4	2.1	6.9	GAME ACTION	63546766	4949507	3158756	2122183	12495915			
3	2	PUBG MOBILE -	37276732	500.0 M	4	1.8	3.6	GAME ACTION	28339753	2164478	1253185	809821	4709492			
4	3	Mobile Legends:	26663595	100.0 M	4	1.5	3.2	GAME ACTION	18777988	1812094	1050600	713912	4308998			
5	4	Brawl Stars	17971552	100.0 M	4	1.4	4.4	GAME ACTION	13018610	1552950	774012	406184	2219794			
6	5	Sniper 3D: Fun F	14464235	500.0 M	4	0.8	1.5	GAME ACTION	9827328	2124154	1047741	380670	1084340			
7	6	Call of Duty®: M	13572148	100.0 M	4	2.0	4.0	GAME ACTION	10501443	1274162	517273	268489	1010778			
8	7	Among Us	11936964	100.0 M	3	1.8	5.6	GAME ACTION	5954262	1041297	853099	719378	3368926			
9	8	Temple Run 2	9633929	500.0 M	4	0.3	0.8	GAME ACTION	6579369	991341	614643	349003	1099571			
10	9	PUBG MOBILE I	7578630	100.0 M	4	1.0	2.5	GAME ACTION	5382545	500696	351523	238986	1104879			
11	10	Gangstar Vegas:	6268377	100.0 M	4	0.4	1.0	GAME ACTION	4509647	605510	319332	167792	666094			
12	11	Pixel Gun 3D: FF	5681934	100.0 M	4	0.2	0.5	GAME ACTION	3866308	527159	326294	186593	775578			
13	12	Bowmasters	5471344	100.0 M	4	2.0	4.1	GAME ACTION	4319351	514574	229990	98901	308525			
14	13	Standoff 2	4801658	50.0 M	4	3.0	7.6	GAME ACTION	3884644	365380	135150	65422	351059			
15	14	Talking Tom Gol	4710639	100.0 M	4	1.3	4012.8	GAME ACTION	3581634	397794	200800	116460	413948			
16	15	Garena Liên Quâ	4564398	50.0 M	4	1.0	2.7	GAME ACTION	3142312	386513	204843	111645	719082			
17	16	Banana Kong	4496723	100.0 M	4	0.5	1.0	GAME ACTION	3497817	518453	223009	72115	185328			
18	17	Crossy Road	4464668	100.0 M	4	0.1	0.2	GAME ACTION	3379550	576816	234463	68644	205194			
19	18	War Robots. 6v6	4221389	50.0 M	3	692.7	0.6	GAME ACTION	2634036	450688	267931	164087	704645			
20	19	MORTAL KOMB	4215808	50.0 M	4	0.5	1.2	GAME ACTION	2874353	457806	254154	136153	493340			
21	20	Last Day on Eart	4159837	50.0 M	4	0.7	1.6	GAME ACTION	3073534	463932	205166	103198	314004			
22	21	Agar.io	4112075	100.0 M	3	0.2	0.3	GAME ACTION	2493810	428315	289428	156522	743997			
23	22	DEER HUNTER	3987062	100.0 M	4	0.0	0.0	GAME ACTION	2776350	614755	242698	81199	272058			
24	23	Modern Combat	3516024	100.0 M	4	0.5	1.0	GAME ACTION	2419540	421218	214049	100008	361206			
25	24	Kick the Buddy	3213257	100.0 M	4	1.0	2.3	GAME ACTION	2174466	265019	189630	118612	465528			
26	25	Special Forces G	3195503	100.0 M	4	1.4	7584.9	GAME ACTION	2486283	232254	118927	70889	287147			
27	26	DEAD TRIGGER	3140510	50.0 M	4	0.2	0.5	GAME ACTION	2332406	409344	156460	63835	178462			

And after that I made the dataset more readable after deletion of the 2 sections.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

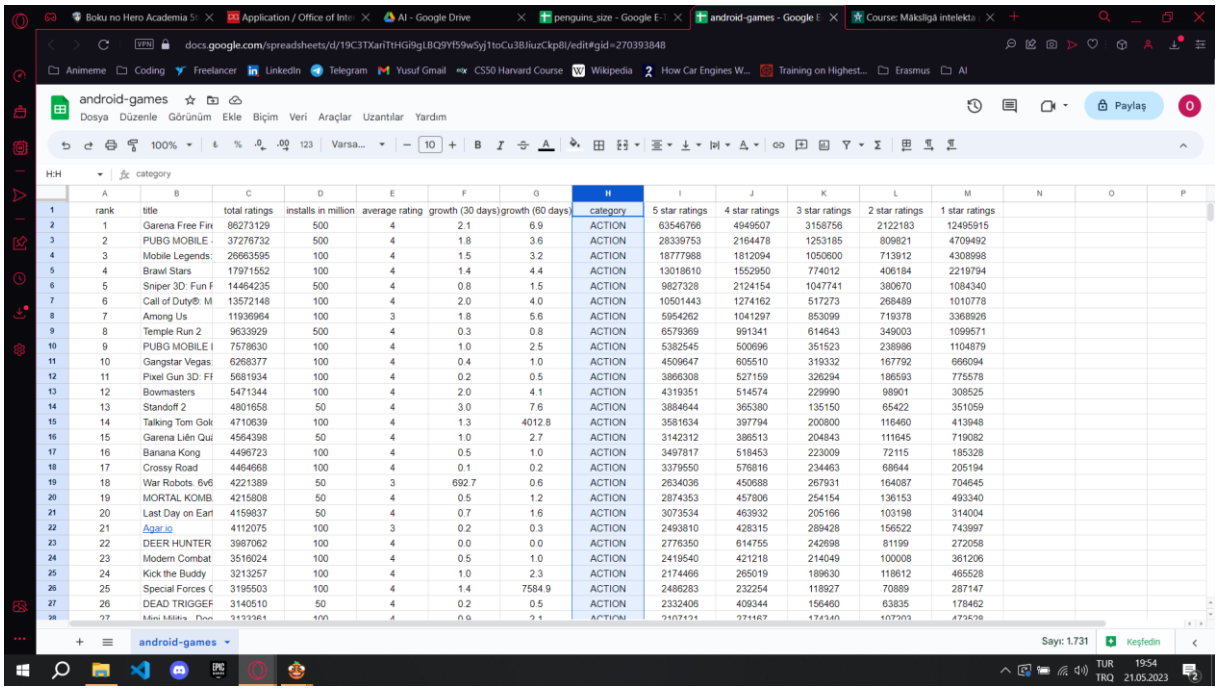


Then I deleted the “M” symbol which stands for million from the data and wrote it to the title so that machine learning can work easily on my data as numerical values.

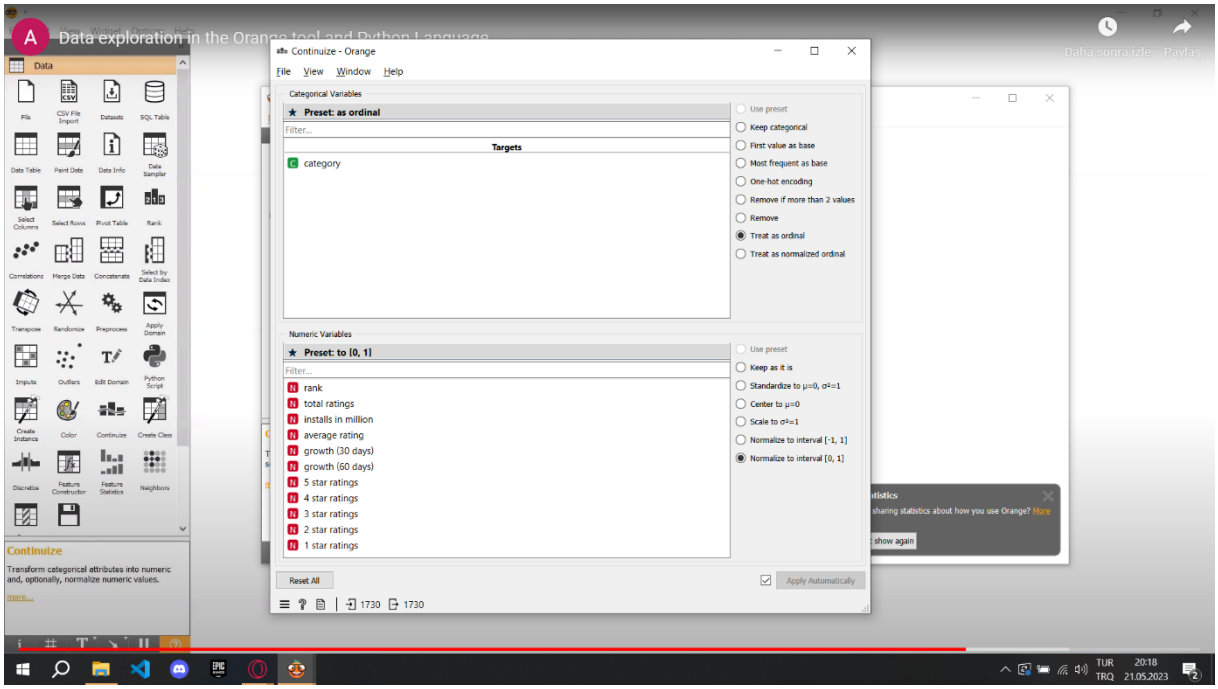


Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

I've put the role of target on my "category" section because in my dataset the way to put them in sections would be from genres of games. All other numerical values were made to be a feature except title which I felt wasn't needed in categorising and in a dataset of 1700 at that 😊

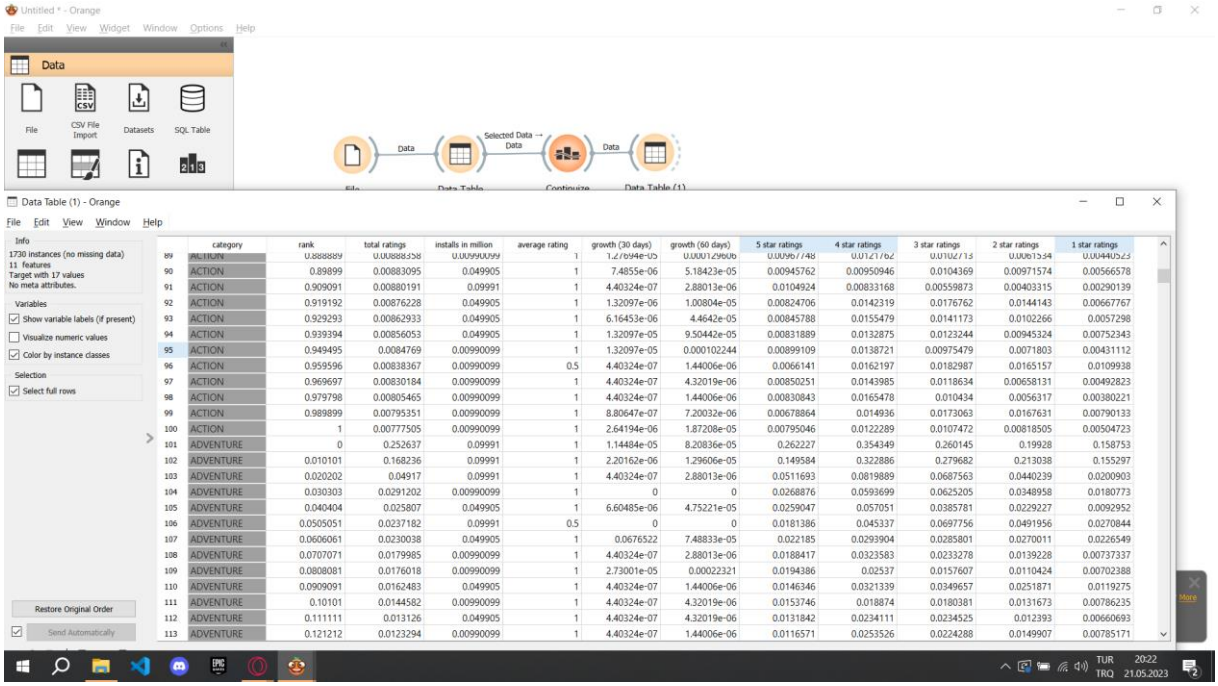


Afterwards deleted the Word "GAME" under category section because it was unnecessarily long.

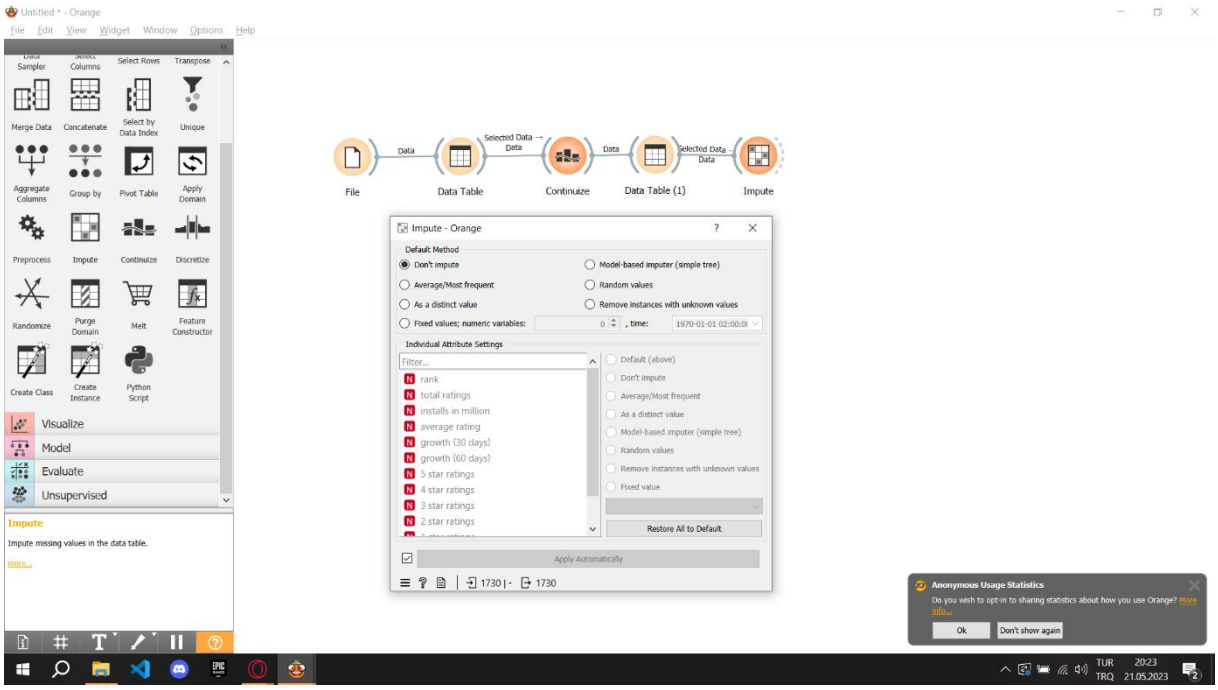


Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

Later I opened constinuize in my diagram and put it as “ordinal” and interval for numerical values as [0,1].

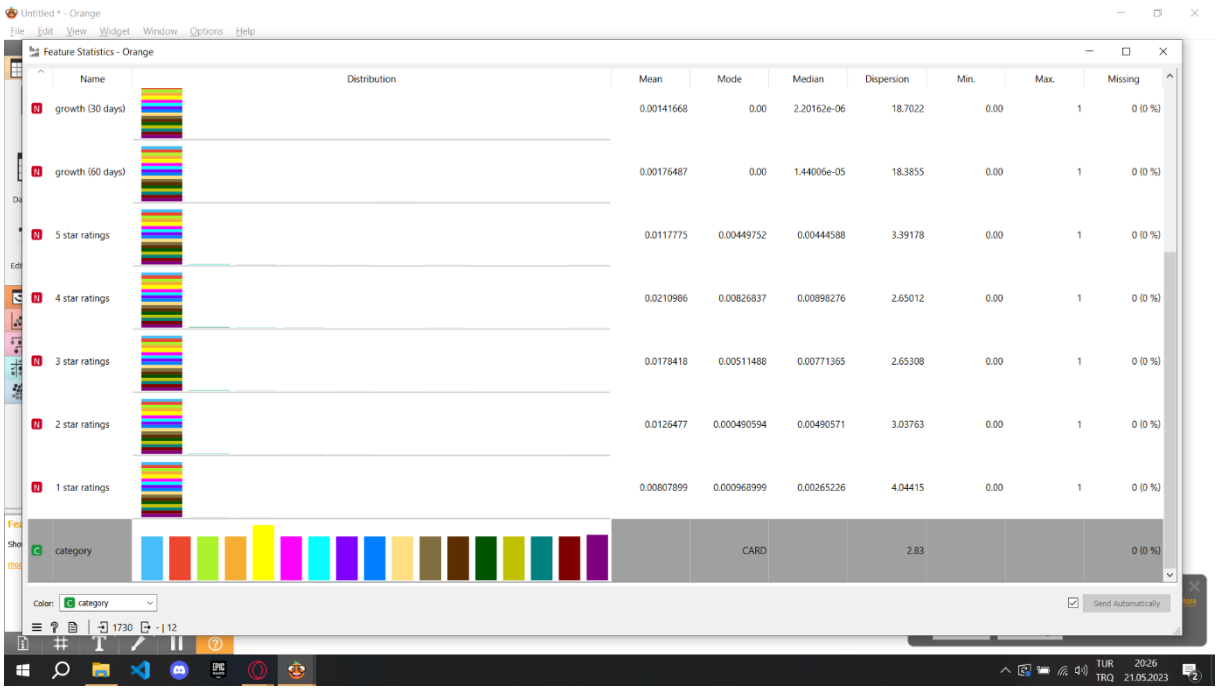


This is the result after the change in numerical and categorical values.

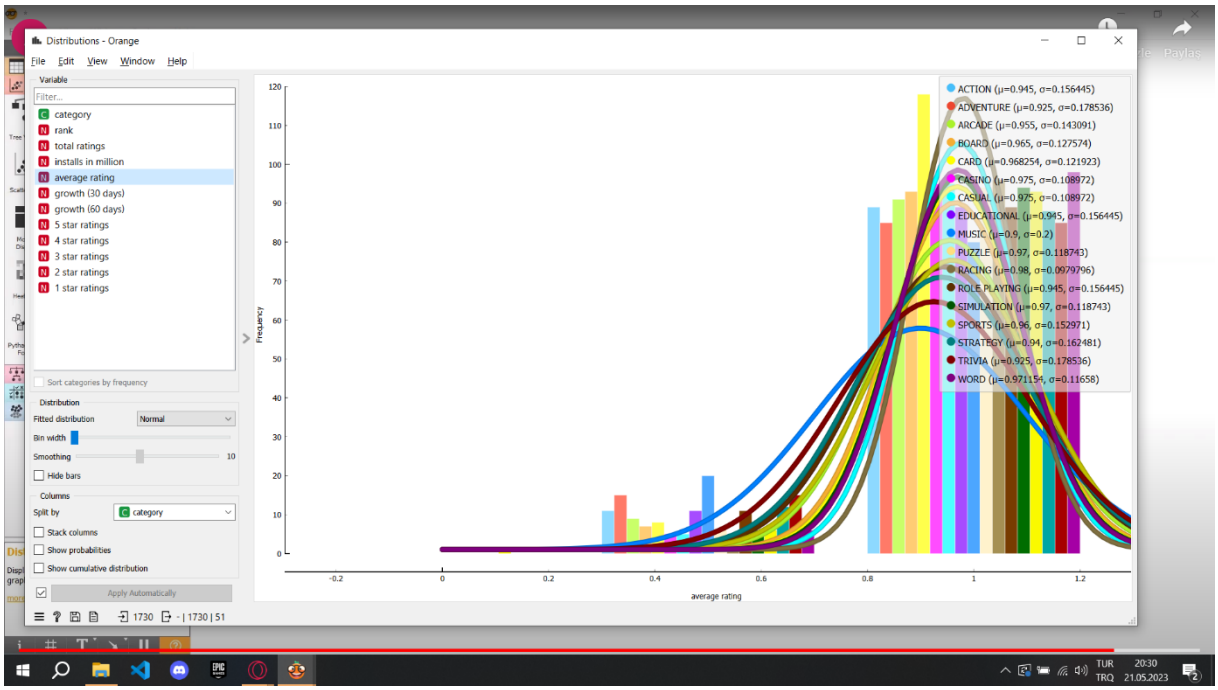


I have added impute function in my diagram but didn’t impute my dataset since I did it at the start for convenience.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

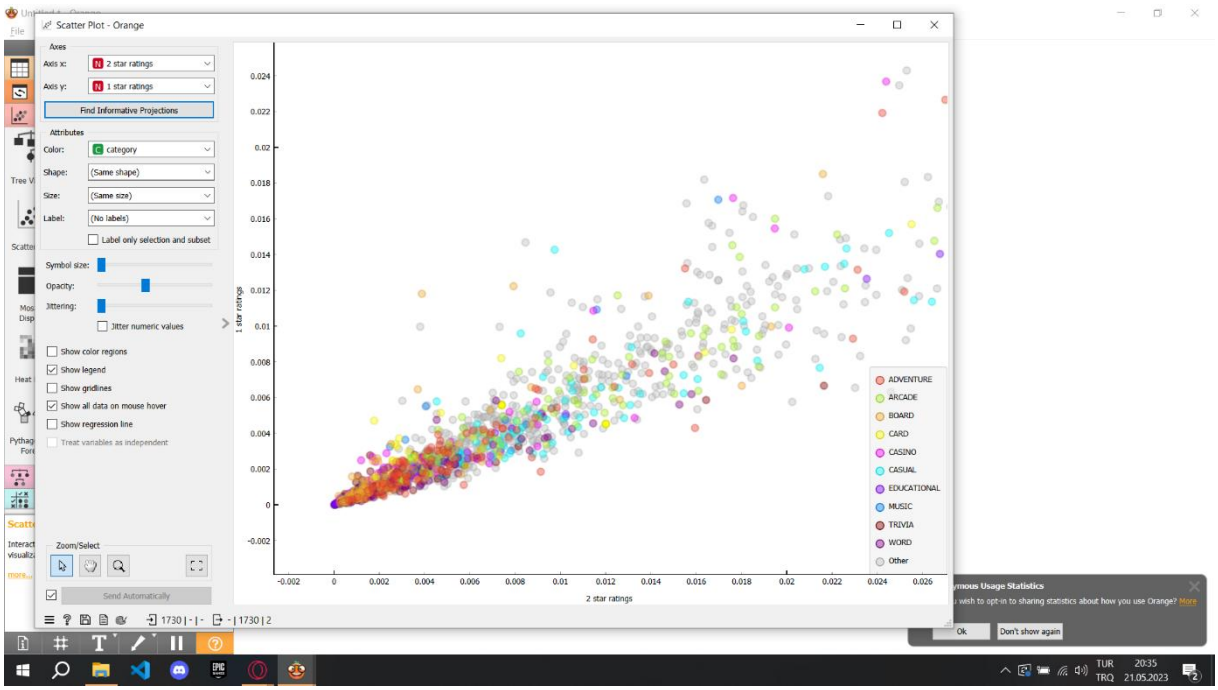


I chose the target for the color section. This showed my feature statistics.

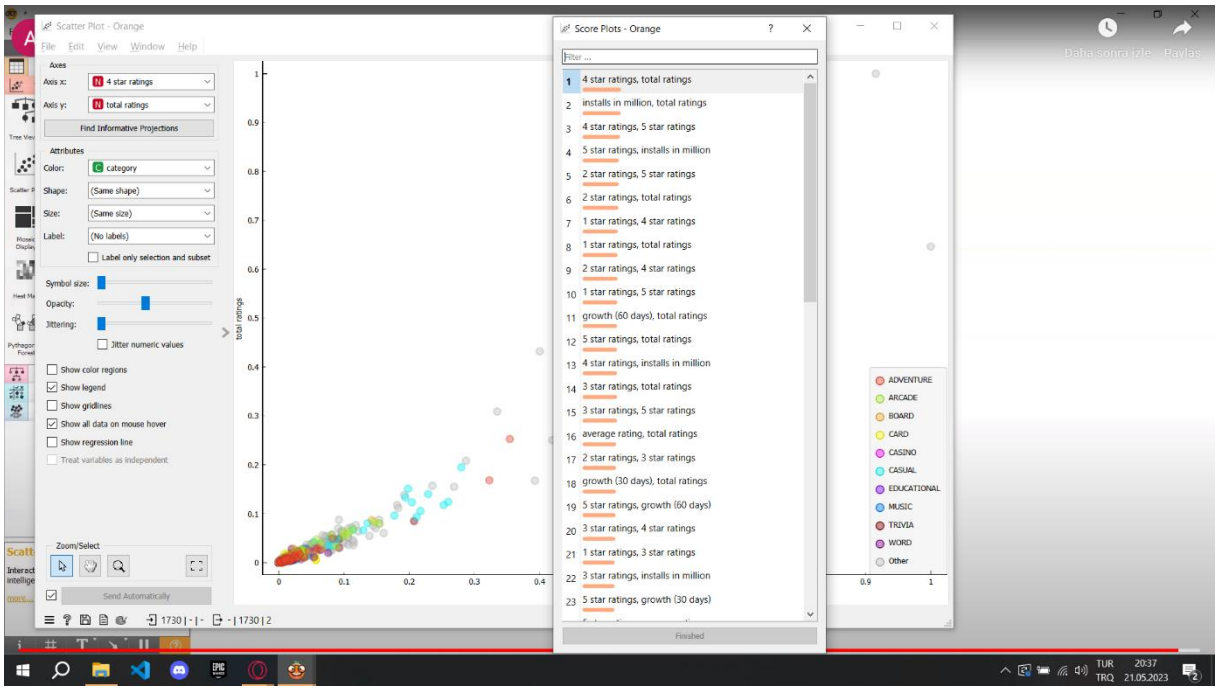


Later on I checked my distributions. E.g: Average Rating… It was done in normal distribution.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

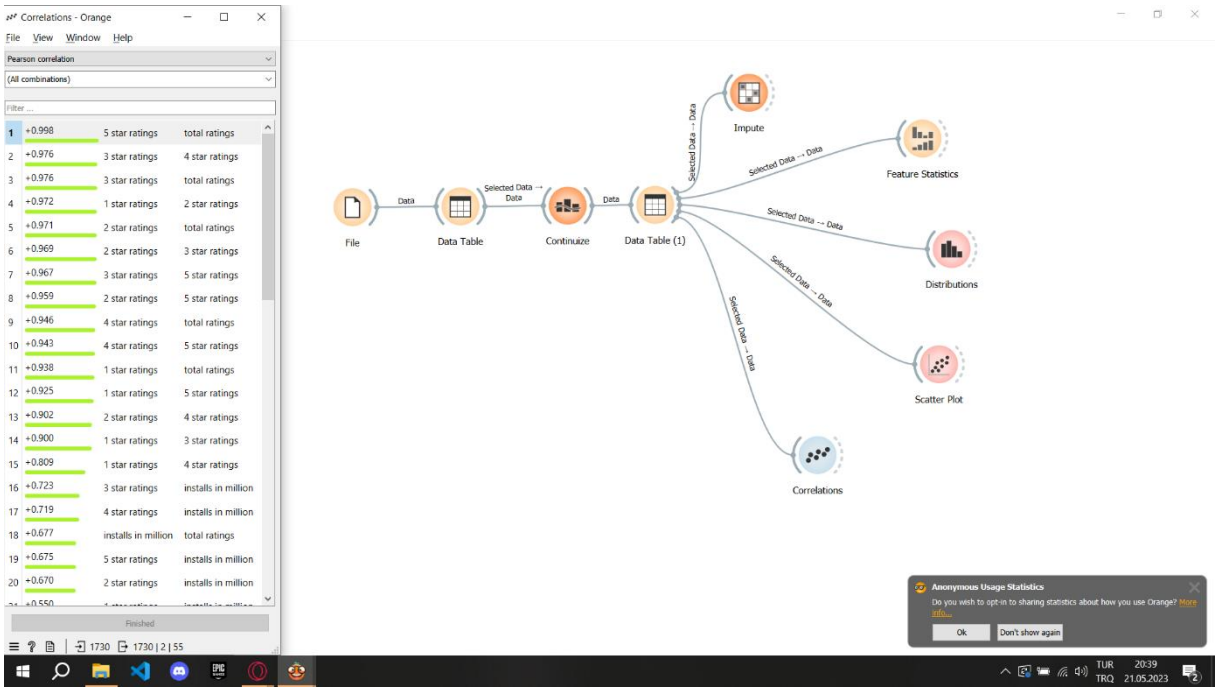


Here's the projections in scatter plot I received after inputting 1 and 2 star ratings.



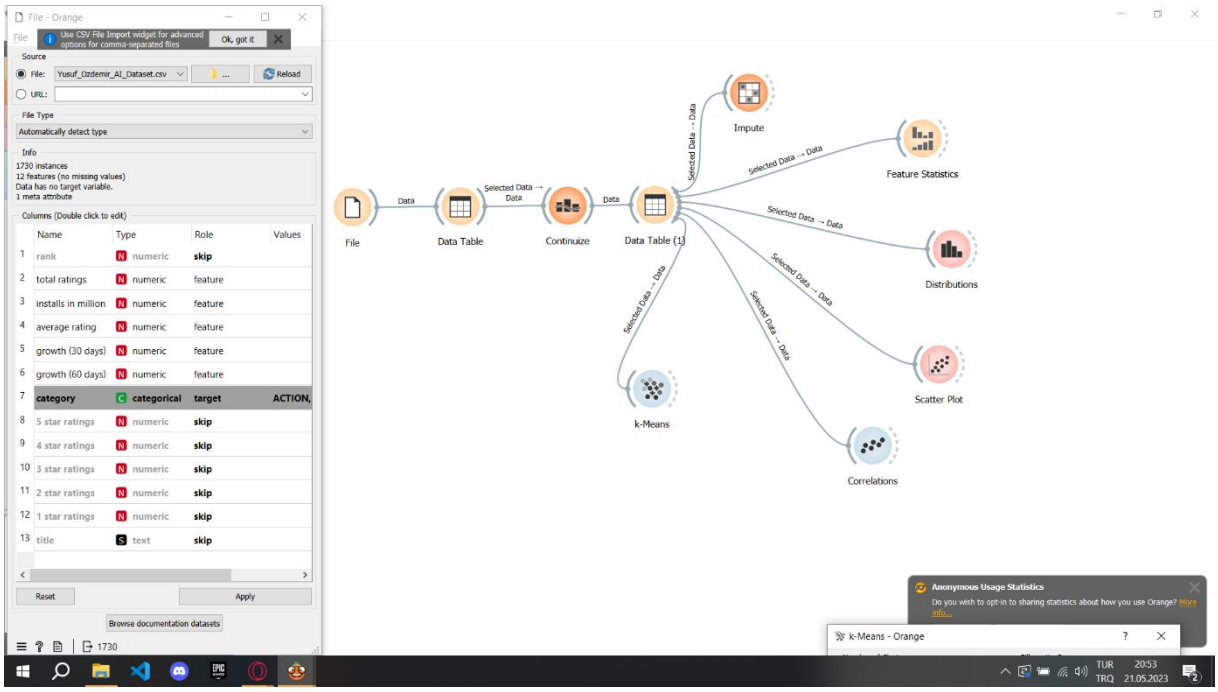
And here's the informative results of the same scatter plot.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git



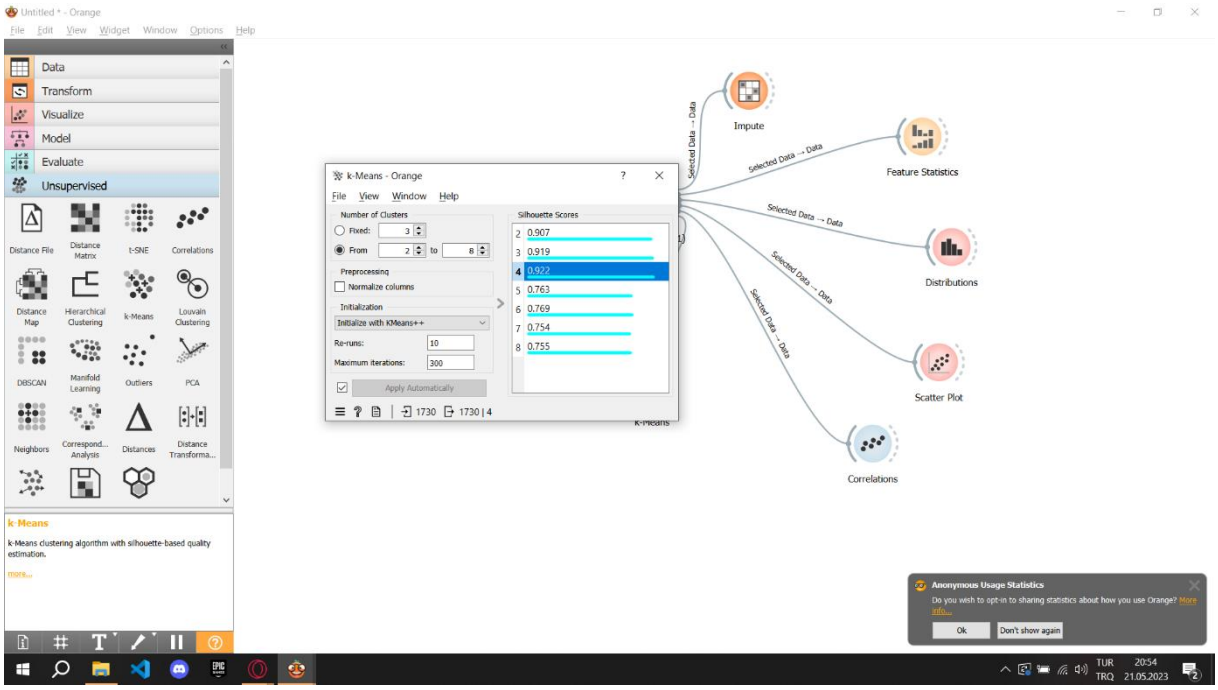
This is the updated look of my diagram at the moment with correlations shown on the left side.

2) Unsupervised Learning



I have minimized my objects before attempting the k-Means algorithm. Around half were left as feature and the other half was skipped.

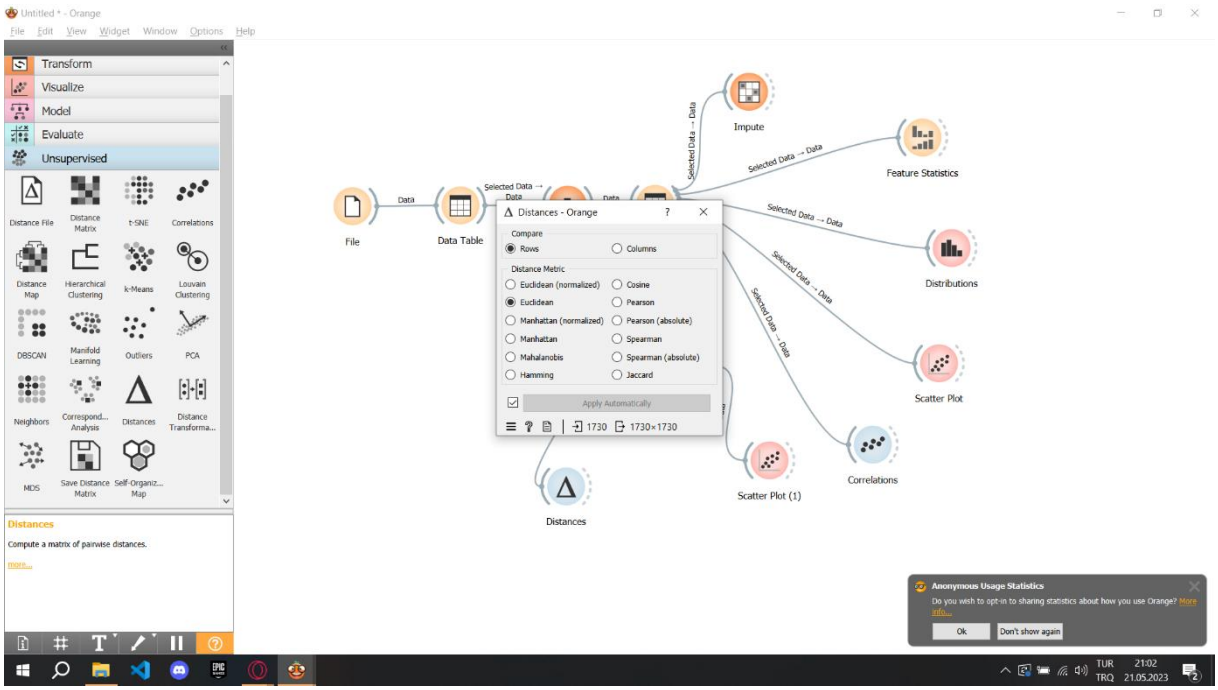
Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git



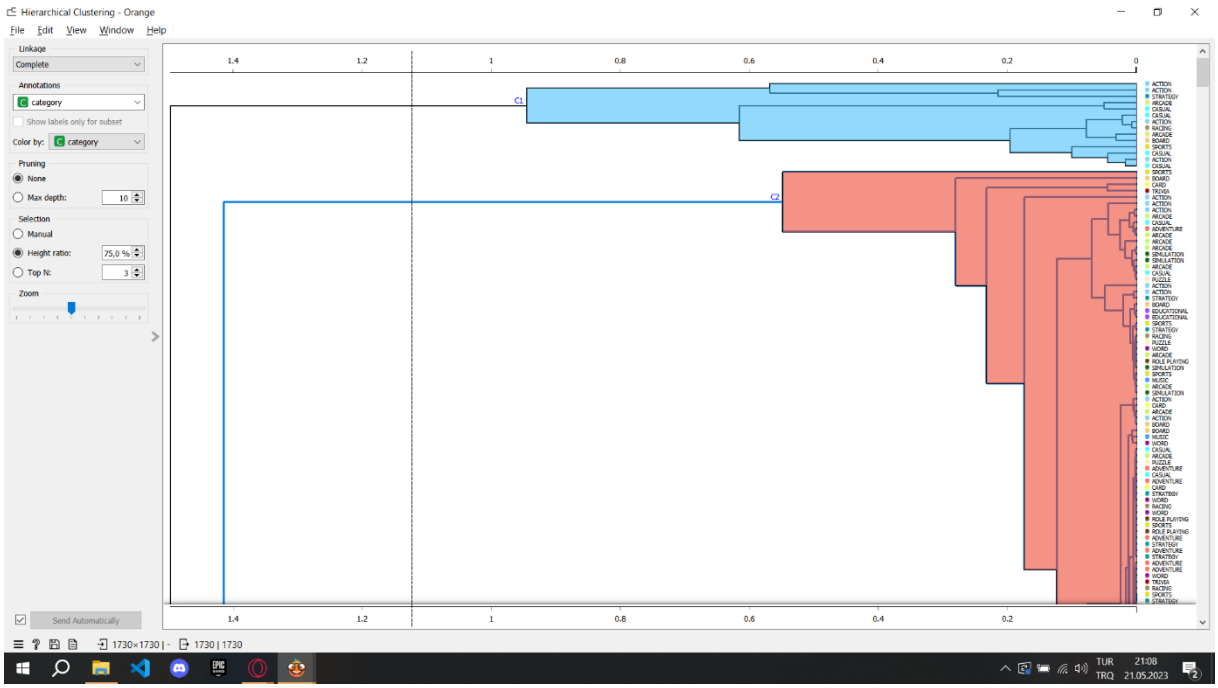
Since I Already Had Normalized Value I Didnt Tick the Option And Got my Silhouette Scores. For me 4 clusters was the sweet spot it seems.

Here's scatter plot of my k-Means algorithm. The reason it looks so undecipherable is because of the amount of objects crammed up in each other.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git



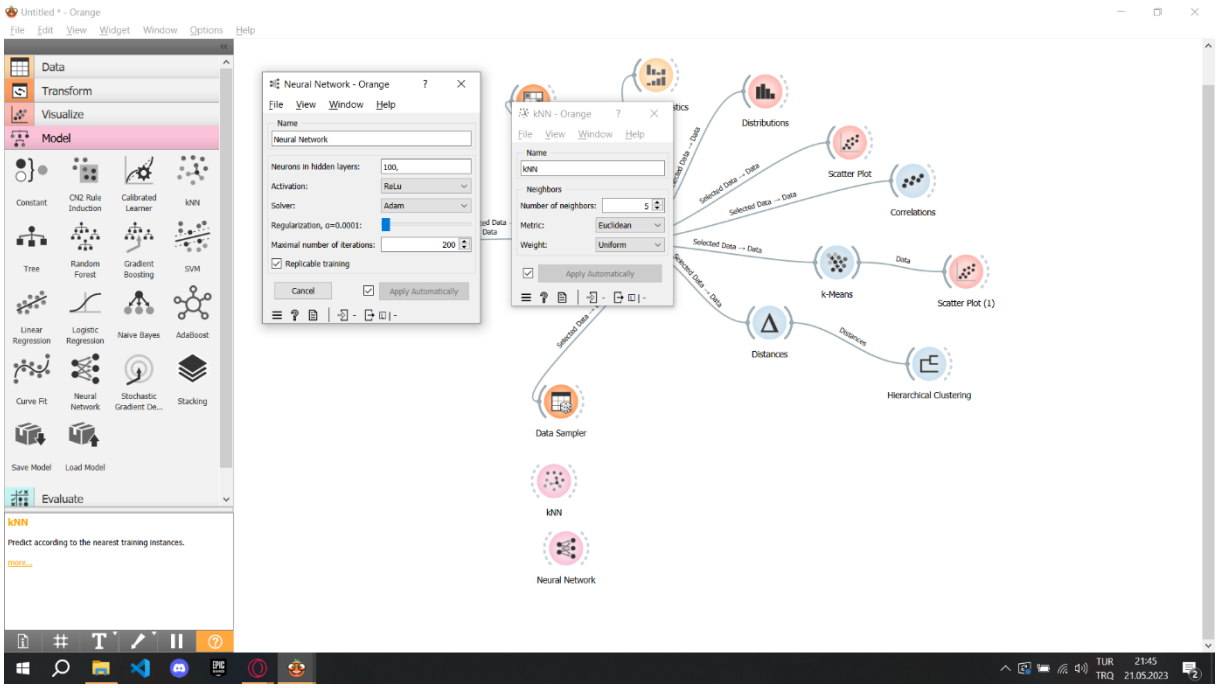
I added the distance function to my diagram here and didn't tick normalize as usual, chose euclidean method for it.



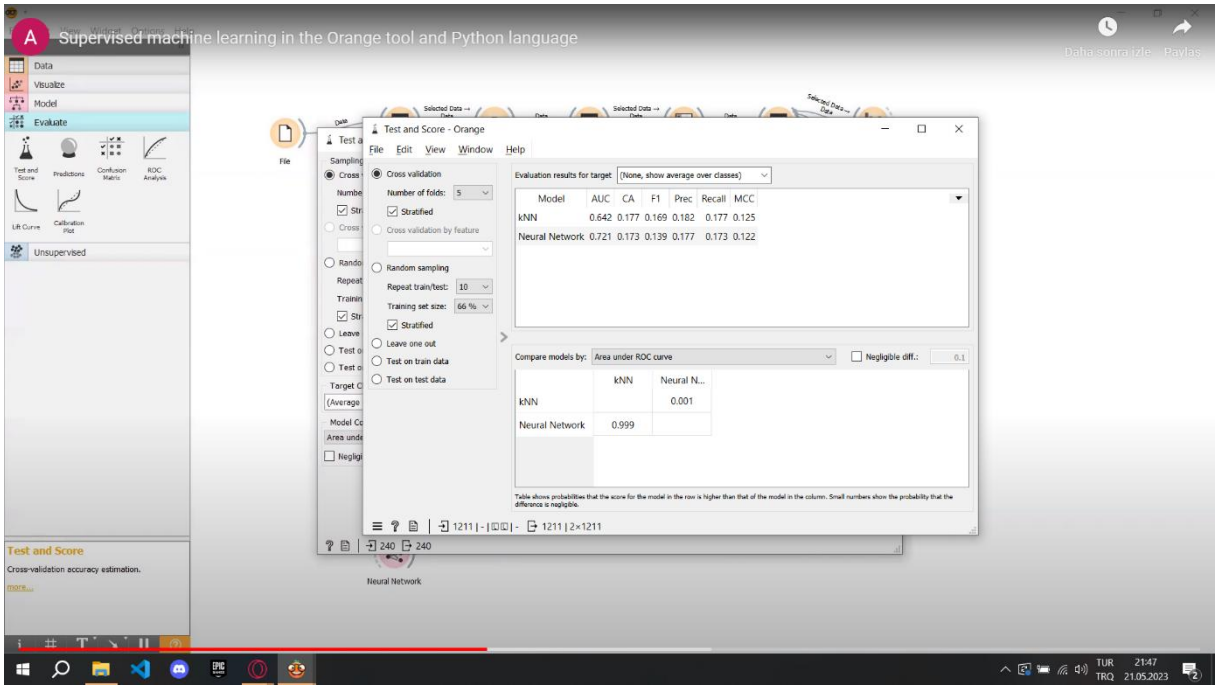
In my case after adding the hierarchical clustering, clustering data properly was nearly impossible so I stopped trying to find the cut-point after a while.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

3) Supervised Learning

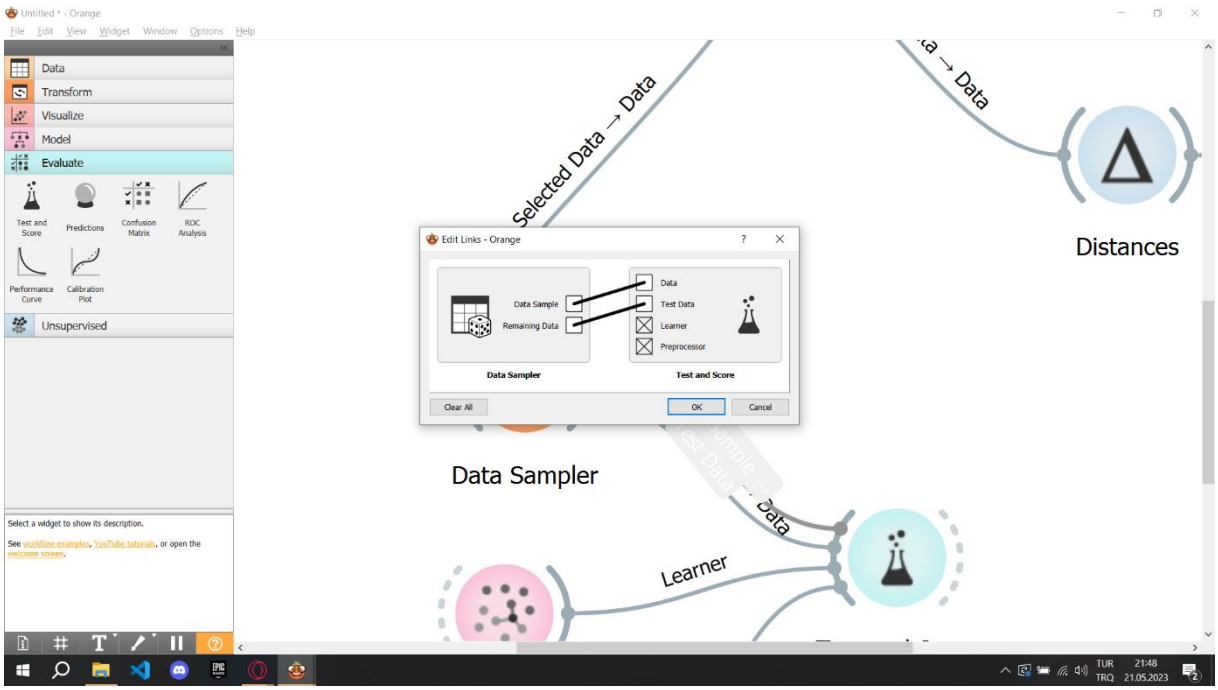


Added my KNN and Neural Network sections to my diagram here. Their specifications can be seen in the Picture above.

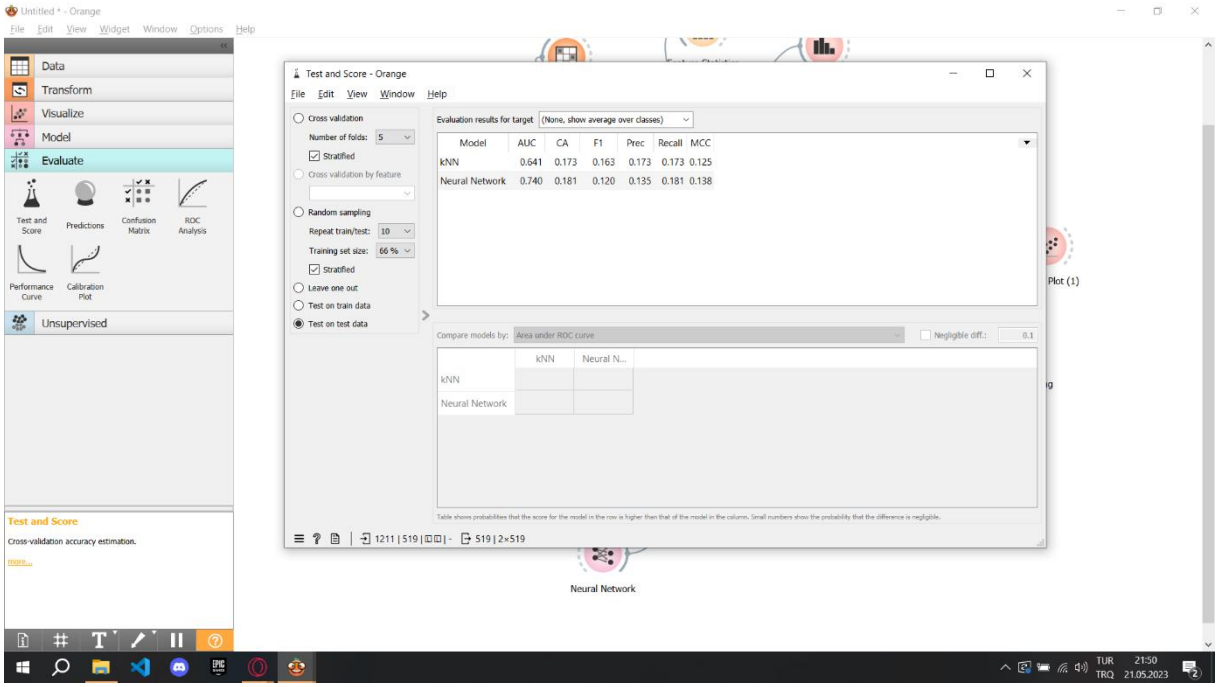


After adding my test and score section, I checked the results of Knn and Neural network. For me neural network did a better job. I used cross validation for the testing.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

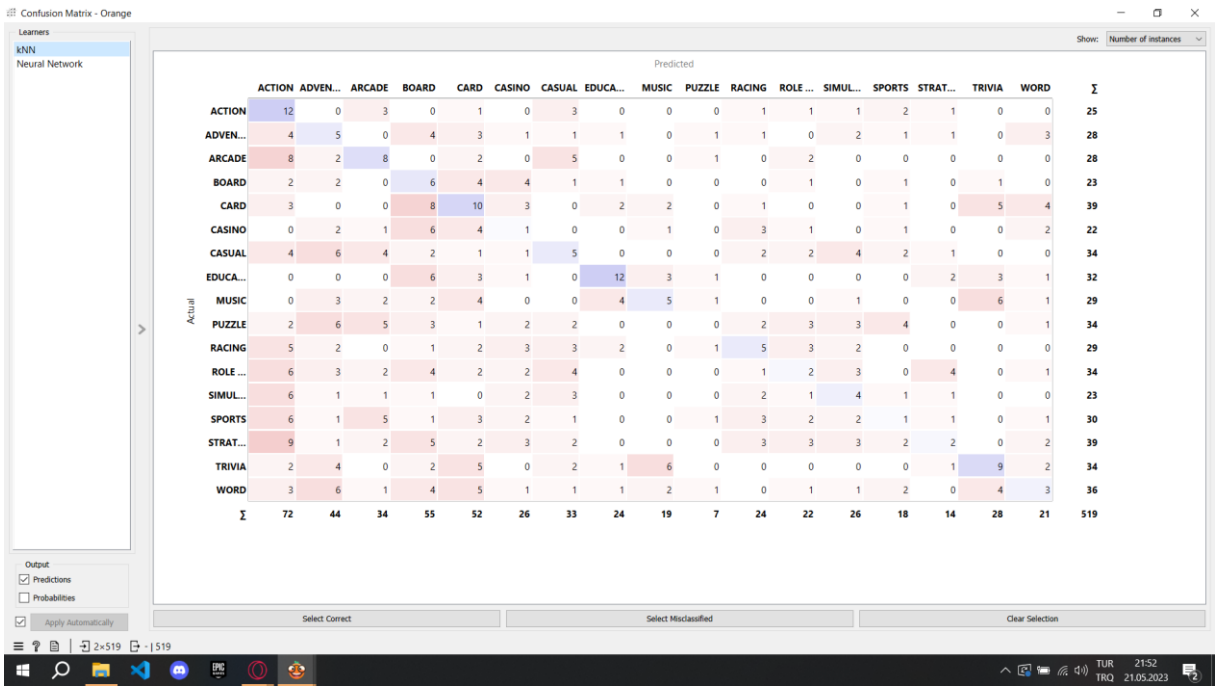


Changed the linking between knn neural network and test data in this part.

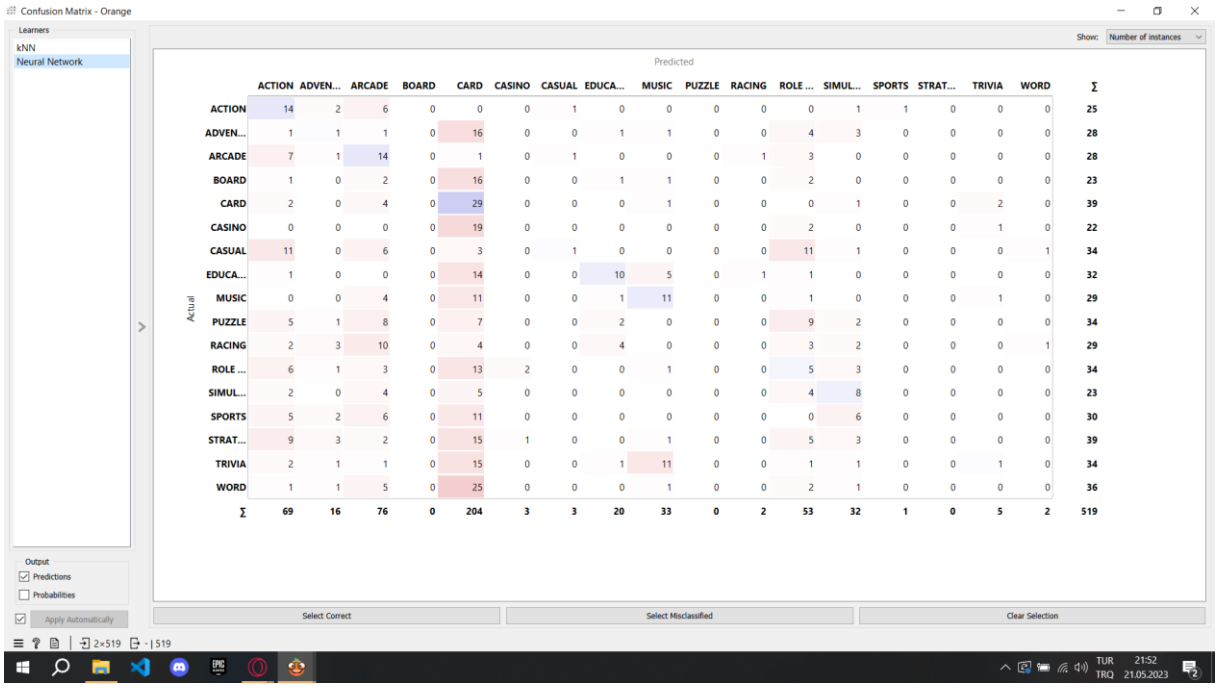


And this is the result after the change in linkage. Neural network still wins.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

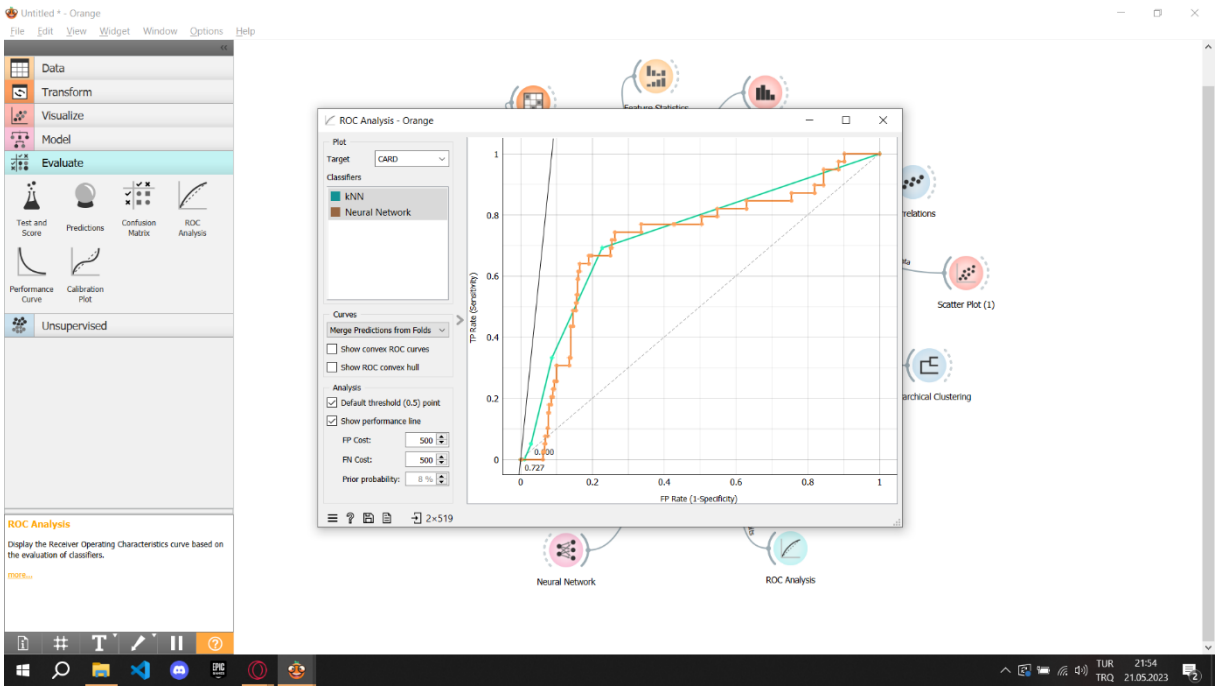


My KNN confusion matrix with its predictions.

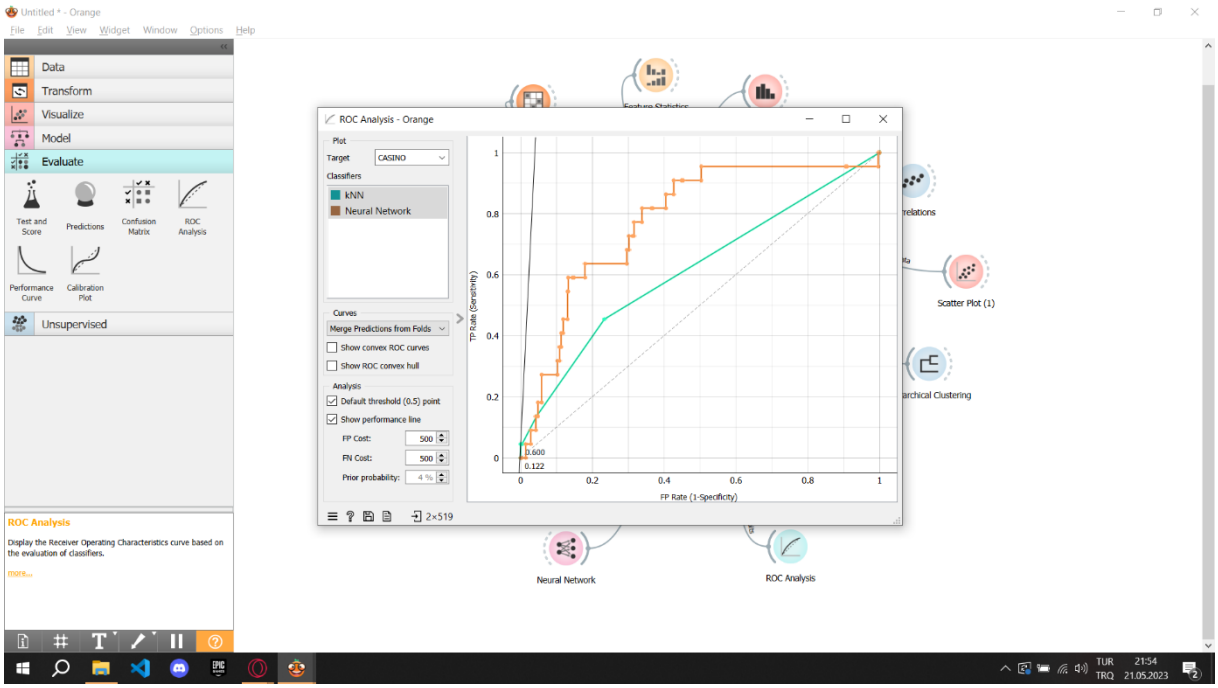


My Neural Network confusion matrix with its predictions. It is easily seen this matrix is far more accurate compared to the Picture before.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

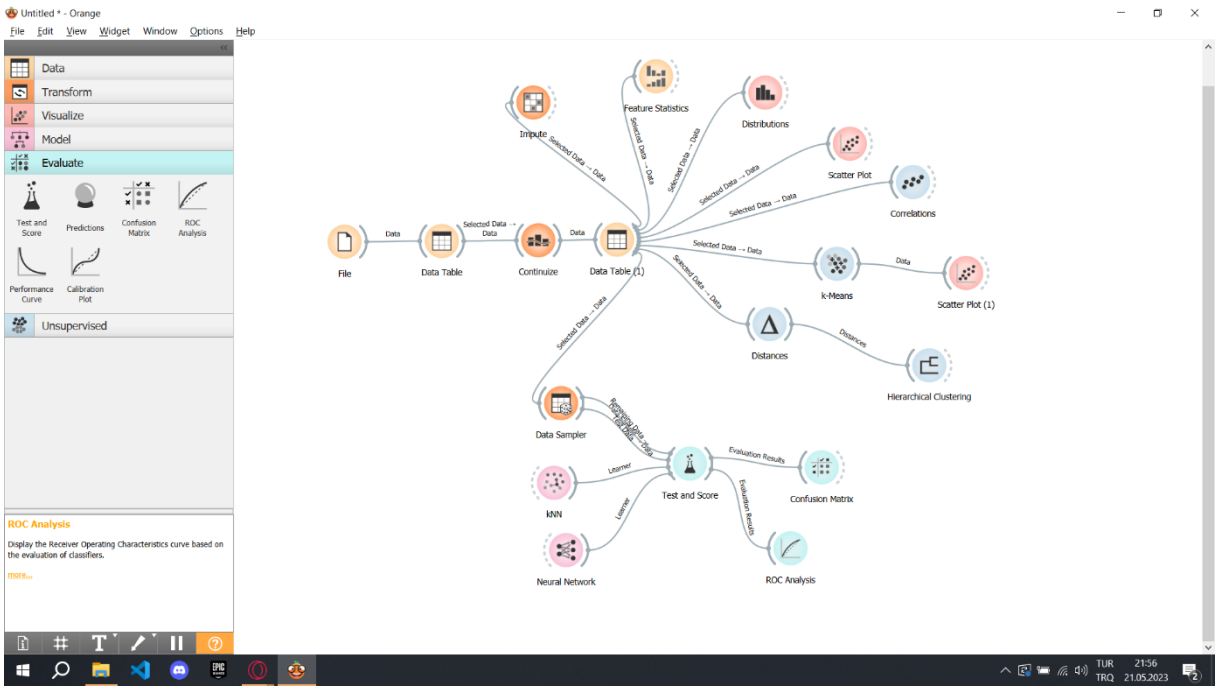


Here I added the ROC analysis to my diagram and chose a random genre. In this instance it was CARD games.



In this instance it was CASINO genre of games. It is easily noticable how in both Roc analysis' Neural Network did a better job.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git



And to finalize everything, this is how my finished diagram looks like in the end.

!!!It is important to add I have tried tests of different values of measuring in each function in this Project and decided not to add a screenshot for every single change so I just played around with them myself and experimented. Otherwise my document would be ridiculously long…

4) Conclusion:

In this pracitcal work, I have learned how to work with a dataset and manipulate them first to make it usable for machine learning.

Then I have downloaded and used the orange tool which made working with a dataset incredibly easy. It is a drag and drop tool which was very beneficial for my learning experience since I understand better with pictures.

After finishing up with cleaning the dataset and starting unsupervised and supervised machine learning I realized how useful it could be and already had some ideas myself such as using the data from my fitness watch to better track and predict, maybe even create diet plans and/or workout plans using the data.

I also saw differences between my data and my professors' in which our results were quite different in the sense that where Neural Network algorithm worked for me, KNN worked for her and such.

Yusuf Ozdemir 211ADB094
https://github.com/JobMonarch/AIYusuf.git

There is likely so much to AI and this is only the tip of the iceberg but I thank my professor for creating such an understandable Project and inspring me on AI even more 😊