Report on final project:

Video game sales

1 Introduction

Over the last 40 years, the video game market has become increasingly important. Our work focused on the evolution of this market, in particular, the related sales figures through the globe. We chose this topic for two main reasons: It's a possible task our future boss might be interested in and we love video games so we were very eager to see the results. Our dataset was also analysed by [1] and as part of a solved Kaggle competition [2].

2 Data and methods

We base our analysis of the video game market on the data set found on Kaggle [3]. It considers video game titles with over 100.000 copies sold, released between the 1980s an 2016. The data was scraped from `vgchartz.com` in 2016, yet it contains four titles with later release date, which we deleted in the preprocessing step. It should be noted that this data does not include free-to-play titles and mobile games. To prepare the data for our analysis we removed games with missing sales figures or release date and fixed some formatting errors. As became apparent during our analysis, the video game market changed drastically in the 1990s. We therefore only considered data from 2000 up to 2016 for the regression. [4]
Our goal was to make predictions on the development of genres, platforms and overall video game sales for the year 2018. For that, we made use of linear and polynomial regression. We assumed a mean squared error loss function and determined the optimal parameters by minimization using the gradient descent method. If possible we compared our results to actual statistics from 2018.

3 Results

First, we had a look at the overall sales figures, the global as well as the regional data for North-America (NA), Europe (EU) and Japan (JP). After analysing the number of produced games for each year we further delved into the (relative) sales figures of the different genres and platforms. As our data is incomplete (as already stated in section 2) our predictions for 2018 are not representing the actual video game market. All the plots and results of the regression can be found in the according notebooks.

3.1 Regional sales

Our data are related to NA, EU and JP. all these regions follow the same trend and they fit a second-degree polynomial regression. NA has the highest number of

sales followed up by EU and at least JP. Looking at the regression we can see that for further years we will get negative revenues and that's impossible. One possible reason is the fact, as already pointed in the Data and Methods section, that we miss free-to-play and mobile games that cover a huge market share.

3.2 Number of games

From the observation on the number of products released concerning the different years, we noticed that the period with the biggest production of video games is from 2007 to 2011. Through a deeper analysis we saw that during those years, the platform that released more games is Nintendo DS (with 492 games released between 2007 and 2008, directly followed by Nintendo Wii (325 games in 2009) which the main genre was "Action". After this huge peak of releases of Nintendo, it fell and ended production, leaving the stage to a platform such as PS4, PC and Xbox One, with a trough of "Sport" games and a new peak of "Action" games, that had followed an oscillating trend from 2008 to 2014. The actual total of games released in 2018 is 974 [4] but our prediction through polynomial regression turns out to be 504. The reason could be that the data used contain information only on games with more than 100.000 sales (as we observed in section 2).

3.3 Regional relative sales per genre

For the analysis of the different genres, we took a look at the relative market share of the four most relevant genres, like in 3.2. The results for the prediction follow from the second-order polynomial model, except for the genre "Sports" in Japan, as the prediction value was negative. We observe, that in North-America, the EU and therefore globally the two most important genres are "Action" and "Shooter", whereas in Japan they are "Action" and "Role-Playing". Comparing this to reference values in [5] for North-America, we note that contrary to our prediction the most important genre is "Action" (27%), followed by "Shooter" (21%) and "Role-Playing" and "Sports" (each 11%). As stated in section 2, this data contains data on mobile games and is therefore not appropriate for a detailed discussion of our prediction values. However, it is also probable, that the chosen model doesn't fit the actual trend of the data.

3.4 Regional sales per platform

For the analysis of the different platforms, we grouped them by brands to get a simpler visualisation. In the predictions for 2018 for both linear and polynomial regression, we see that some values are negative. This means that we need to choose which regressions model to use for each brand as we cannot have negative sales. Comparing the prediction with the actual data for 2015, we see that the linear regressions seem to give a better prediction knowing that the sales have increased between 2015 and 2018. We can also compare the results with the value from Statista [6]. If we group the value from Statista in the four groups in common with our result (PC, Microsoft,

Nintendo and Sony) and discard the rest (like Mobile games), we obtain the following proportion: Sony lead with 28%, followed by PC and Nintendo with 25% and Microsoft is last with 22%. The prediction of the linear polynomial is close to this proportion: Sony is leading but Nintendo and Microsoft are close. The only problems with this comparison are that the prediction for PC is low. This can be explained by the fact that some games are missing from the data. It is also important to note that the value from Statista is based on a survey asking people what platforms they are using and note based on real sales.

4 Conclusion

Our predictions for the year 2018 suggest that a game belonging to the Action or Shooter genre, published for PlayStation is the most promising candidate. However, this was only a qualitative analysis as there are no errors taken into account. So we can't judge how well the chosen model fits the data respective how well our prediction fits the reality.

References

[1] Bartosz Trzaskowski. Video Games Sales 1980-2017. http://www.bmtfx.com/video-games-sales-1980-2017-2/?lang=en.

[2] Maksim Klimentyev. Video games - Predicting global sales. https://www.kaggle.com/maxkliment/video-games-predicting-global-sales.

[3] Rush Kirubi. Video Game Sales with Ratings. https://www.kaggle.com/rush4ratio/video-game-sales-with-ratings, 2016. [Online; accessed 28-November-2020].

[4] Sherif Saed. Published games in 2018. https://www.vg247.com/2018/12/18/every-game-released-pc-ps4-xbox-one-switch-2018, 2018.

[5] Entertainment Software Association. Genre breakdown of video game sales in the United States in 2018 . https://www.statista.com/statistics/189592/breakdown-of-us-video-game-sales-2009-by-genre/.

[6] Statista Research Department. United States: Which gaming platforms do you regularly use? . https://www.statista.com/statistics/561456/us-types-of-gaming-platform-used/.