

**TASK**

**Exploratory Data Analysis on the Video Game Sales Data Set**

[](http://www.hyperiondev.com/portal/)

**Introduction**

The dataset that has been retrieved for this analysis can be found on Kaggle saved as “vgsales.csv” entailing video game sales made between the years 1980 and 2020. The data collected has video game sales separated by publishers of the game, the platforms the game has been distributed, the video game’s rank and sales made in North America, Europe, Japan, and other sales amounting to global sales.

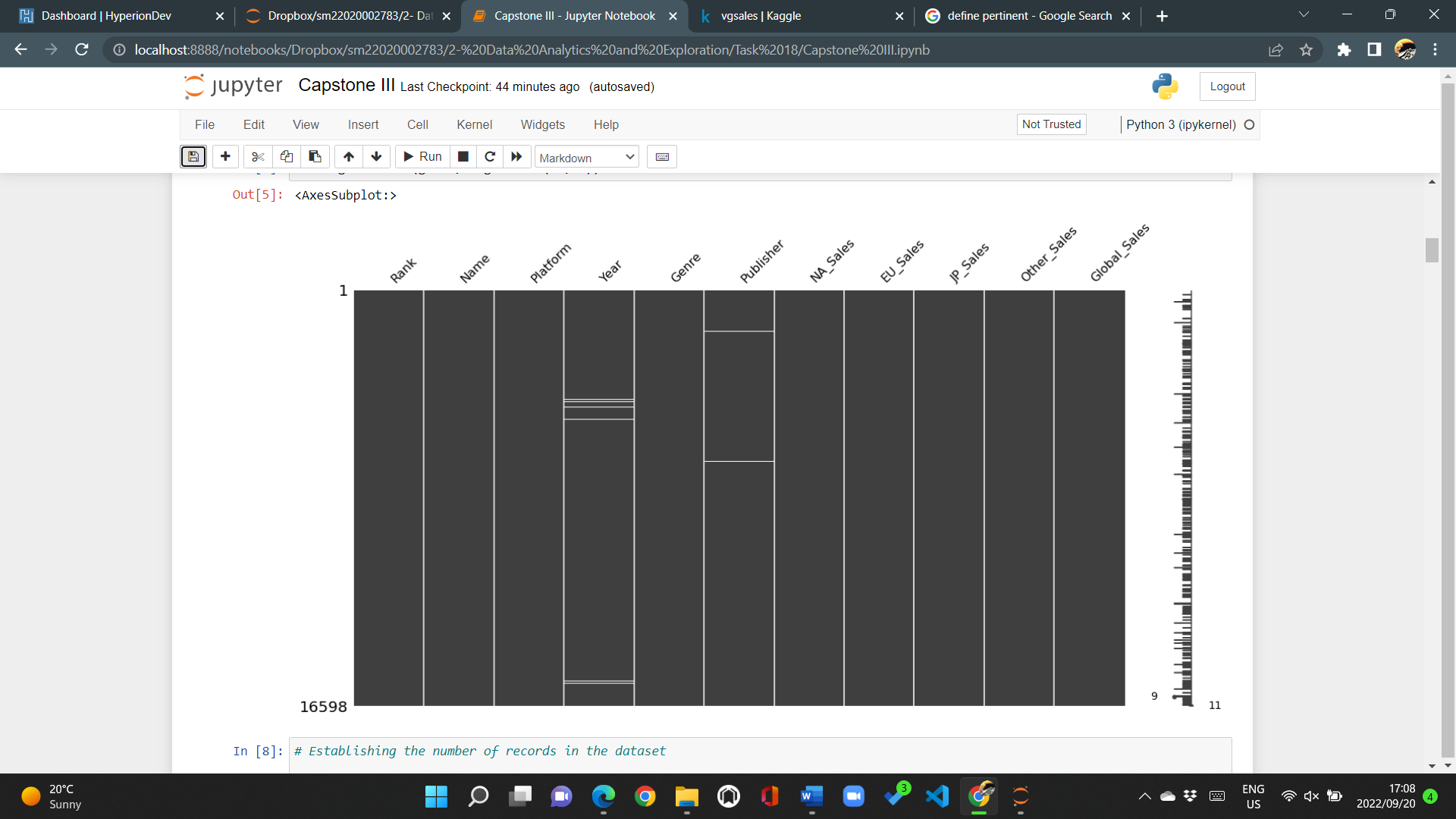
Not only separated by publisher, platform, rank and sales, the dataset includes separation by genre making for an easy analysis of sales and performance based on the game’s genre. Being a member of the gaming community myself, I found the dataset fitting for this exact capstone project. Having additional knowledge of the domain of the dataset, better inferences can be made which prompted my selection of this specific dataset.

**DATA CLEANING**

Before any analysis can be conducted, the dataset must be understood and cleaned. Cleaning the data involves a multitude of different processes including removing redundant columns, identifying repeating data and data that is not pertinent to the dataset and removing null values if any at all.

After carefully reviewing the dataset, the column titled “Rank” was deemed irrelevant. No clear correlation could be drawn from the rank as the ranking did not working in specific categories or groupings. The column functioned as a counter for the number of records essentially and therefore could not be used for analysis. The column titled “Other\_Sales” followed as the column could create more problems than statistical solutions. “Other\_Sales” appeared to be a very ambiguous column to include in the data as the world consists of 195 countries situated in seven continents. Referring to “Other” countries could not provide any use for statistical analysis as “other” may refer to only three or thirty other countries making it increasingly difficult to identify trends based on the country in question. The column has therefore been removed.

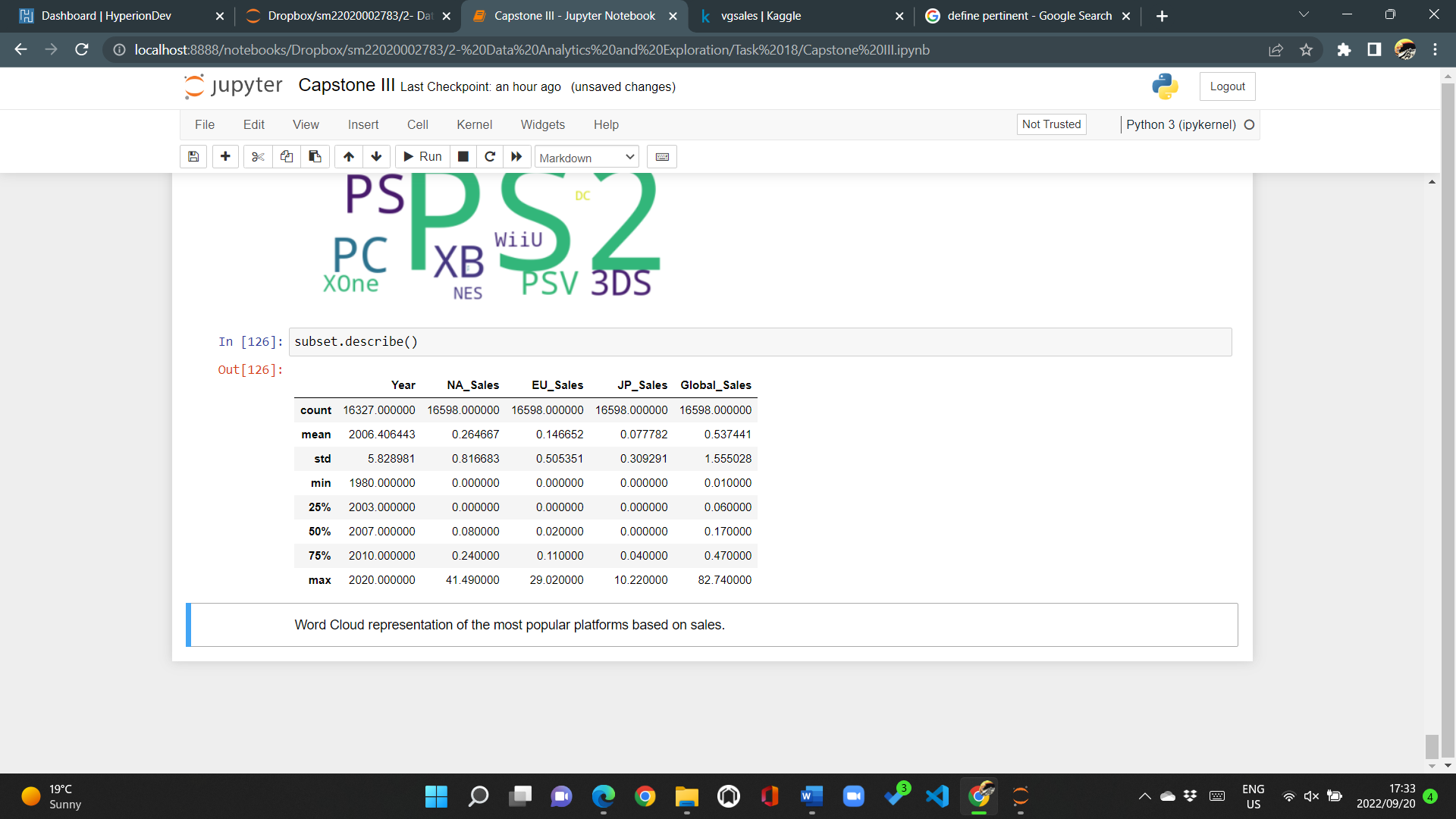
MISSING DATA



A visual above visualising null values present in the dataset. The dataset does not consist of many null values and therefore made for statistical analysis easier. Other records did have zero values as opposed to null values and therefore may function as substitutes for null values. Despite this possibility, no video game title recorded zero sales in total and therefore made it apparent that zeroes found in any of the columns may have not been a mistake and may be a true representation of the sales made across the world.

Enough data is present to continue with a statistical analysis. No additional work has been done on the dataset.

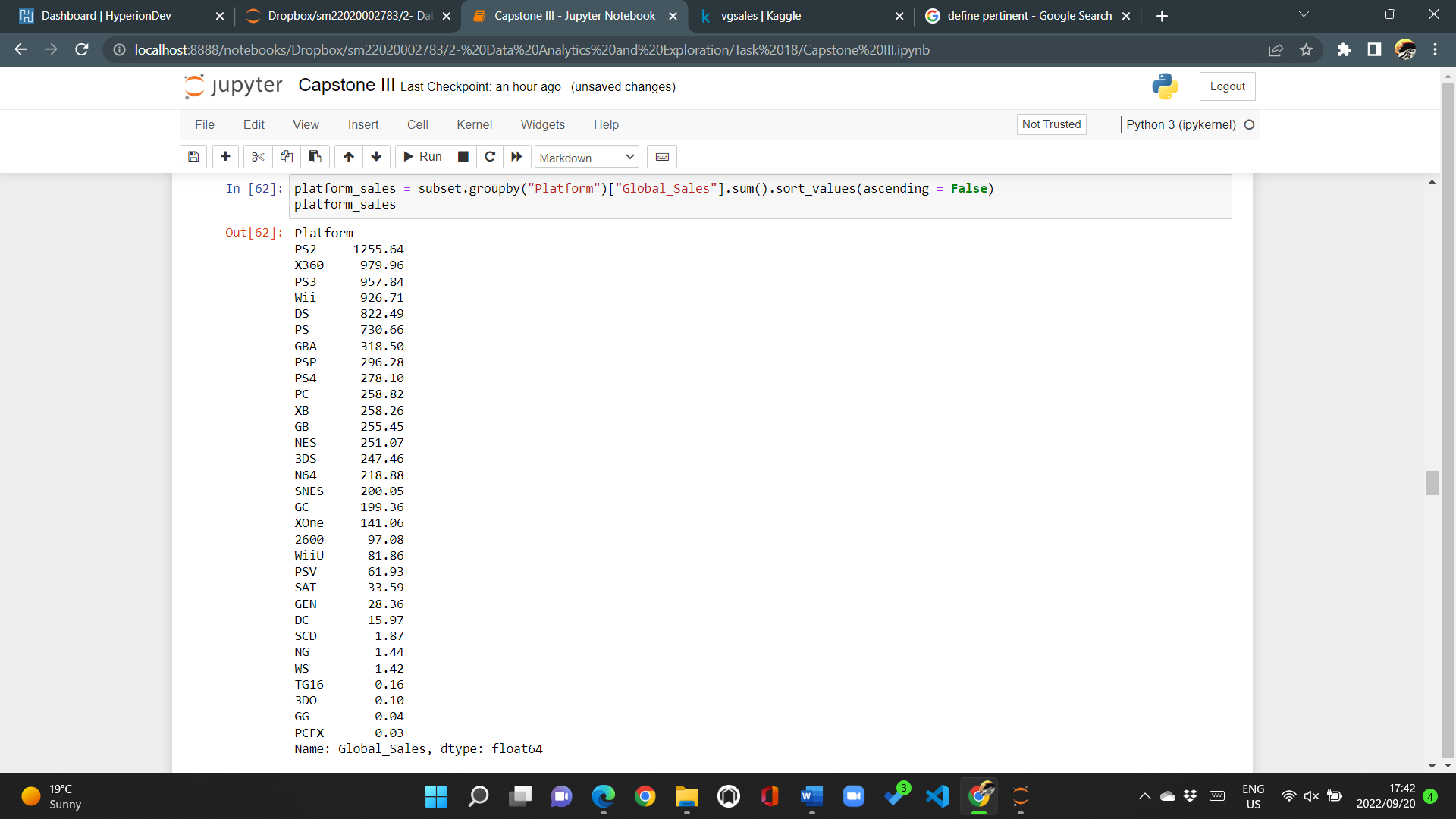
DATA STORIES AND VISUALIZATIONS



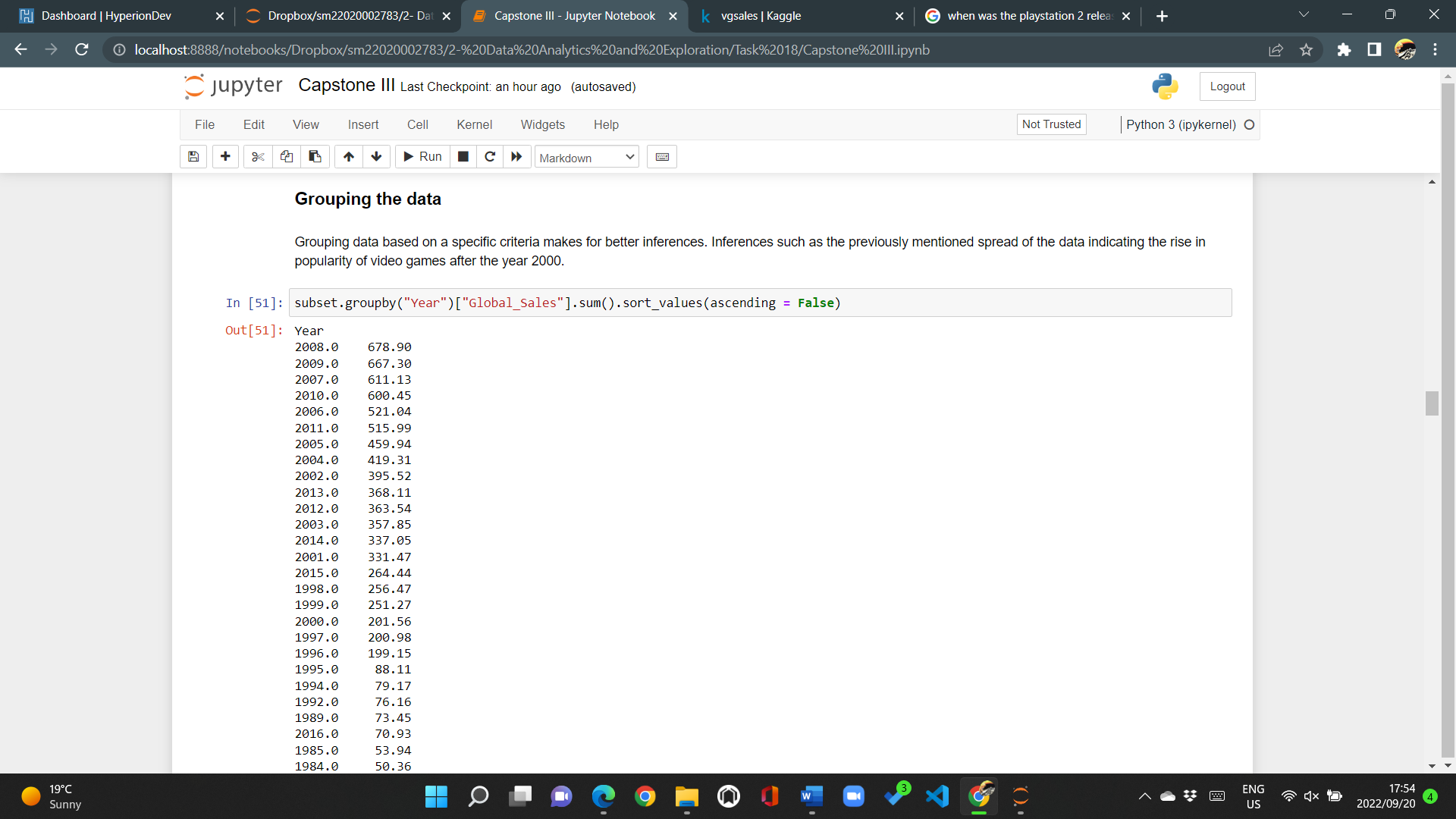
Beginning the statistical analysis began with a summarised statistical analysis of the columns in the dataset. Having a general understanding of the spread of the data, the mode in each column and the average of each column allows for a better understanding of what to look at when analysing the dataset.

Having completed the summary above, the average year under the “Year” column is calculated at 2006.4 highlighting that the data is skewed to the left. In other word, more data points lie to the right side of the data distribution. With the sales being recorded over a period of 40 years from 1980 and 2020, many questions arise. It is important to note that records for the years 2018 and 2019 are missing and very few sales were recorded in the years 2016, 2017 and 2020. However, the median of video game sales is 2007.

Many inferences can be made from the spread of the data but very few will be accurate. Gaming has always been a popular activity that has grown to greater heights attributed to the development of online gaming, gaming content in the form of YouTube videos and e-competitions increasing overall popularity. The sales of video games grew exponentially after the year 2000. In 1999, 251 million sales were recorded across the globe for video games sales with the sales reaching 331 million in the year 2001. The PlayStation 2, a Sony gaming platform, was released in the March of 2000 which would explain why the volume of sales increased by such a large factor.



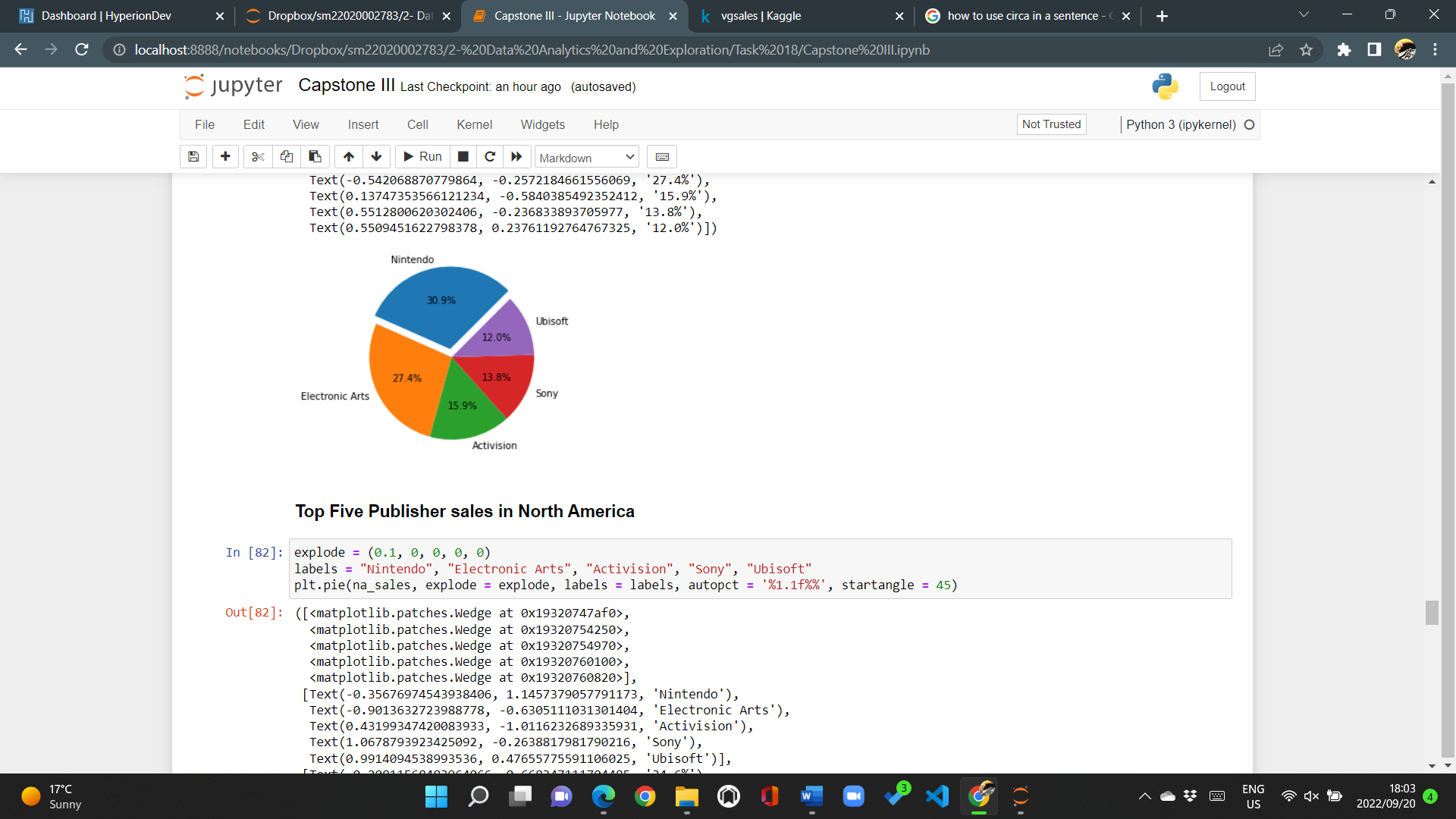
As seen above, the data is a grouping of global sales based on the platform. The top five above reveal that the PlayStation 2 (PS2) is the platform that has sold the greatest number of video games across the world. It is apparent that overall, PlayStation is considered the favoured platform to game on as there are two PlayStation platforms in the top five platforms across the world. Nintendo is a close second with two entries in the Nintendo Wii and Nintendo DS. Despite Xbox360 reaching second in global sales, it is the only platform of Microsoft that appears in the top ten platforms indicating that the despite Xbox not recording very many sales across all Xbox platforms, the Xbox360 is a popular gaming platform.



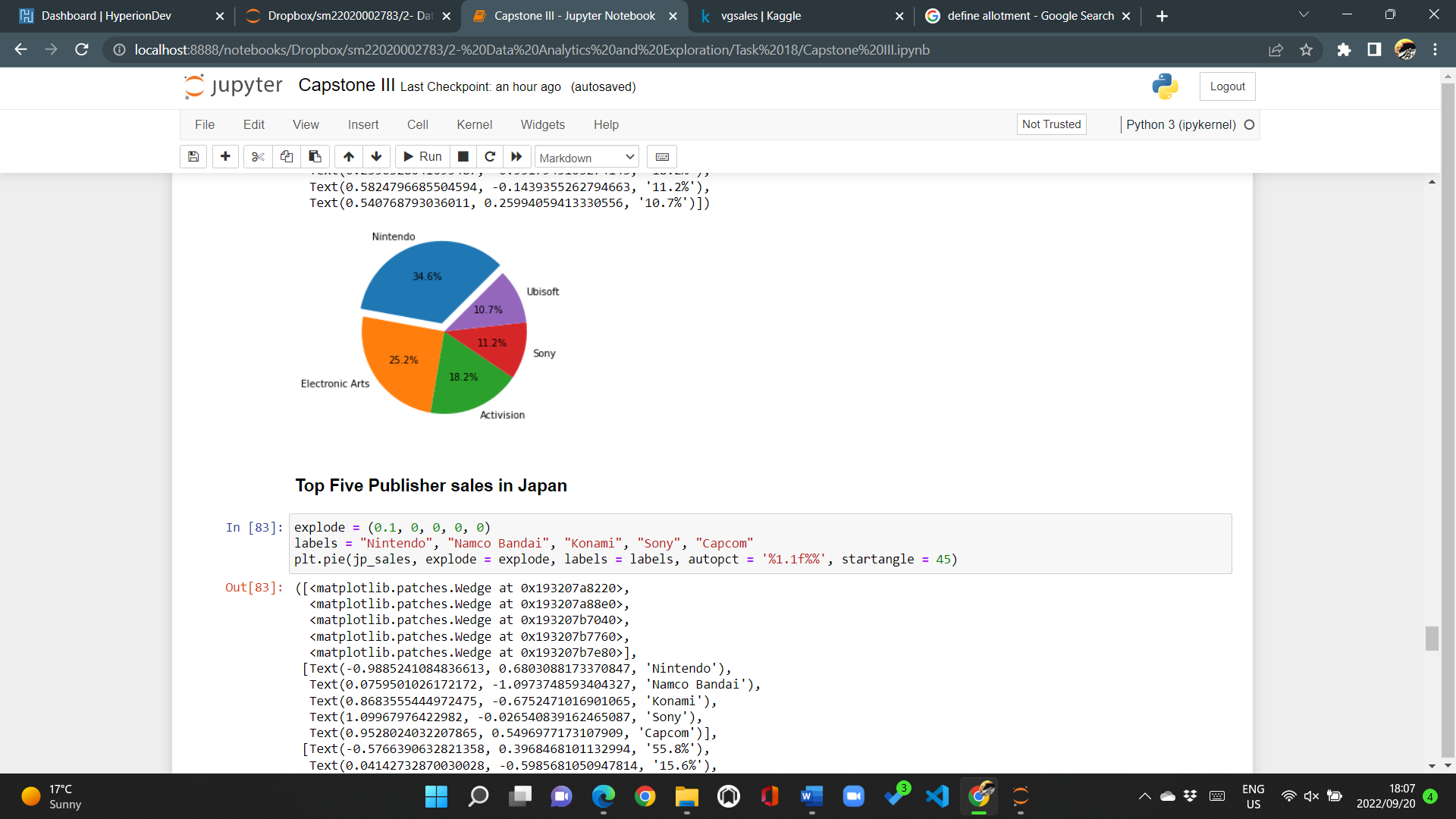
2008 is recognised as the year with the highest volume of sales with 2009, 2007 and 2010 following to make for four years of a many video game titles sold across the world. 2007-2010 being of the four most profitable years for gaming publishers may be attributed to many factors, namely, the release of the PlayStation 3 gaming platform. The platform released in November 2006 and became an extremely popular gaming console after that.

A new system often entails innovative technologies and capabilities that allow for users a better gaming experience. The PlayStation 3 implemented a feature that allowed individuals to play gaming titles online and with many other individuals anywhere across the globe, provided a stable internet connection is available. Popular gaming titles such as Grand Theft Auto saw release dates between 2007 and 2013, which has continued to grow.

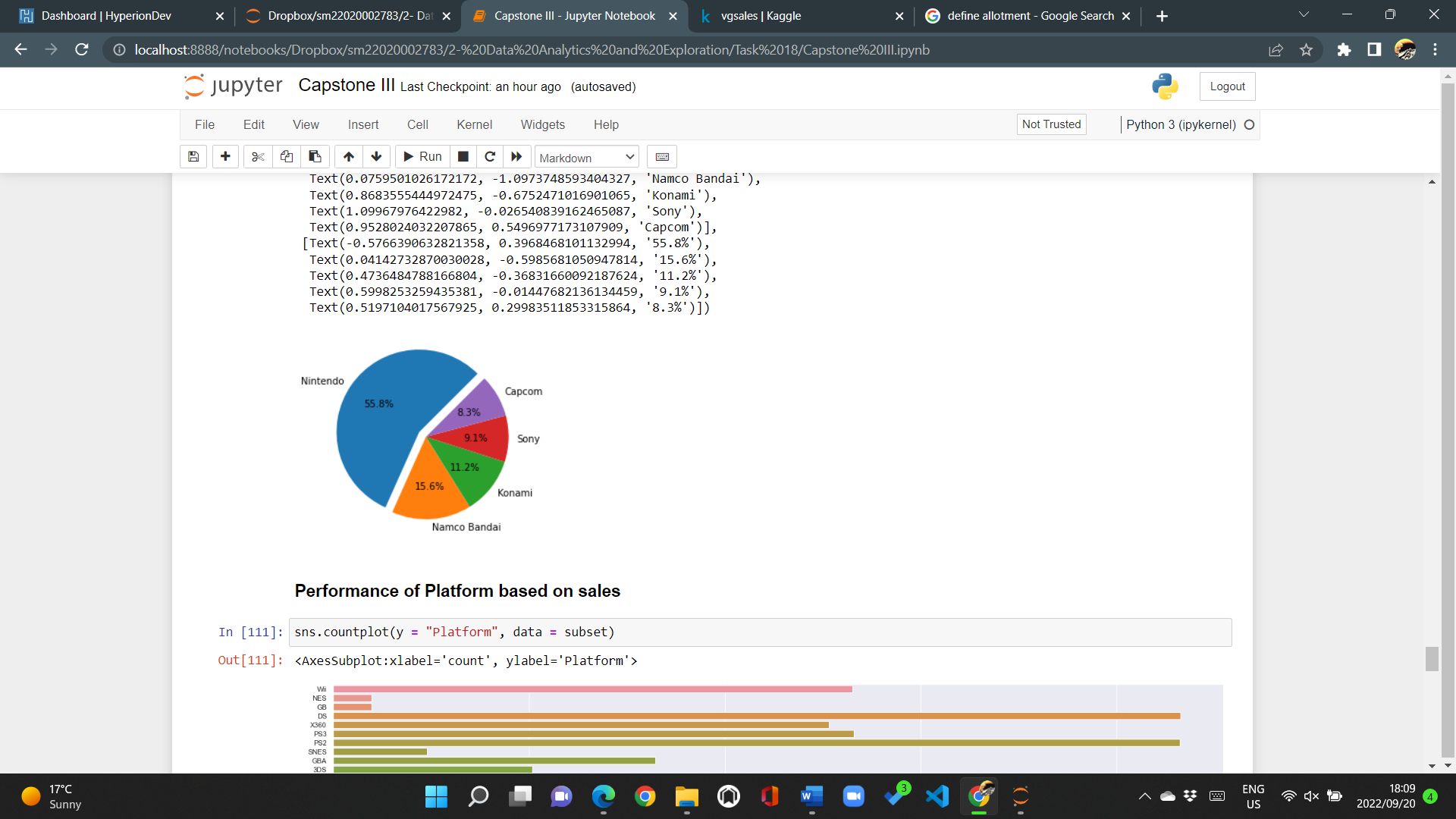
Nintendo offered a different format of gaming by introducing light VR technology that required the user to simulate the movement happening the game such as in tennis, basketball, or netball in the Wii Sports franchise, providing the gaming community with a feature that made for interactivity even better.



A visual representation of the spread of sales based on the publisher in the European market. The graph indicates that between 1980 and 2020, Nintendo dominated the video gaming market having made sales that amount to nearly 31% of all video game sales made in Europe.



The same appears to be true for video game sales recorded in North America with Nintendo once again dominating in sales, Electronic Arts claiming second place and Activision claiming the third position.

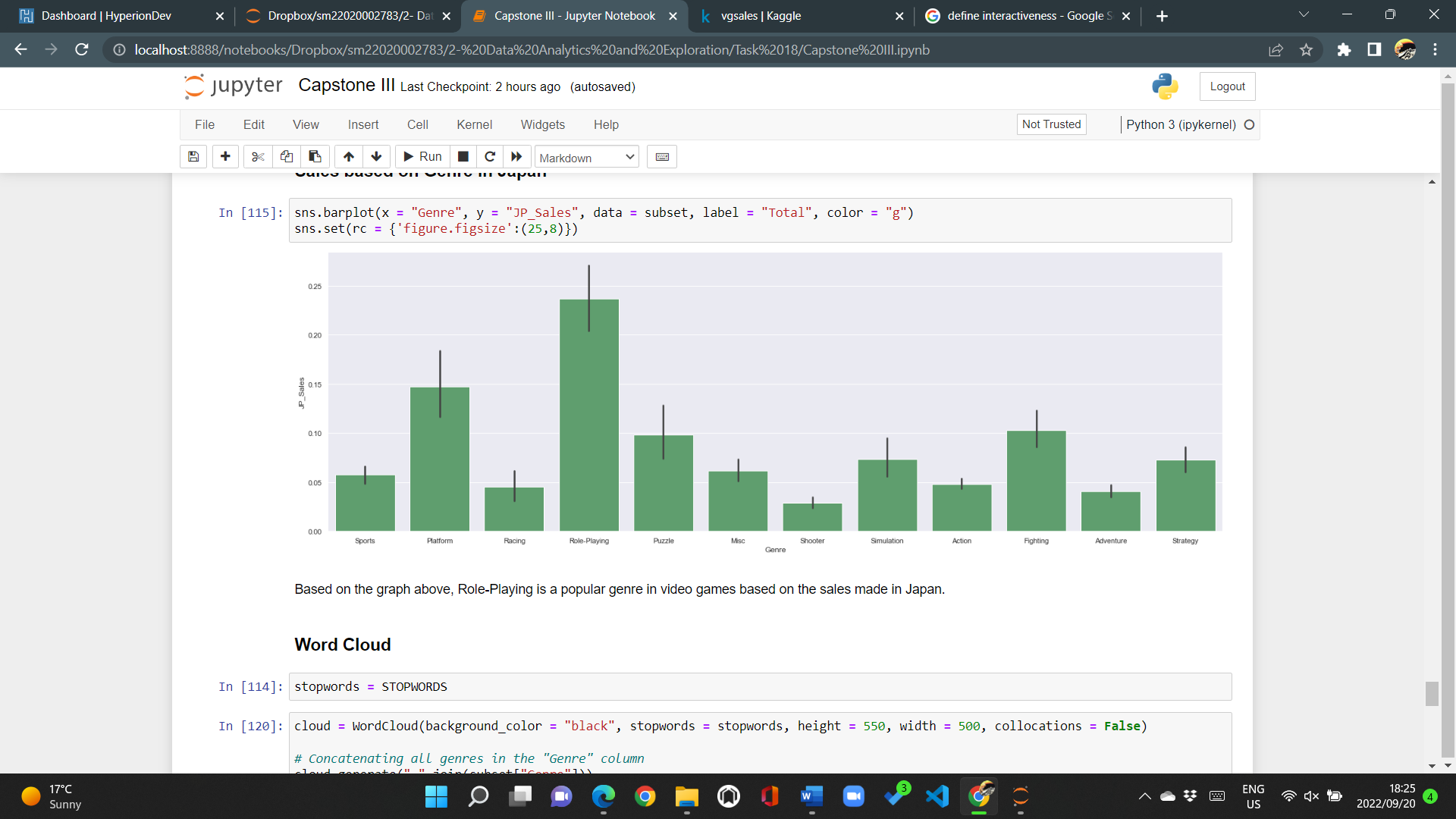


Sales in Japan, however, indicate that despite Nintendo dominating the market once again, favourite publishers in Europe and North America, in the form of Electronic Arts and Activision, are not favoured in the Japanese gaming market. All, except one in Sony, are Japanese based companies hinting that the population of the gaming community situated in Japan favours local development as opposed to gaming titles created outside of Japan.

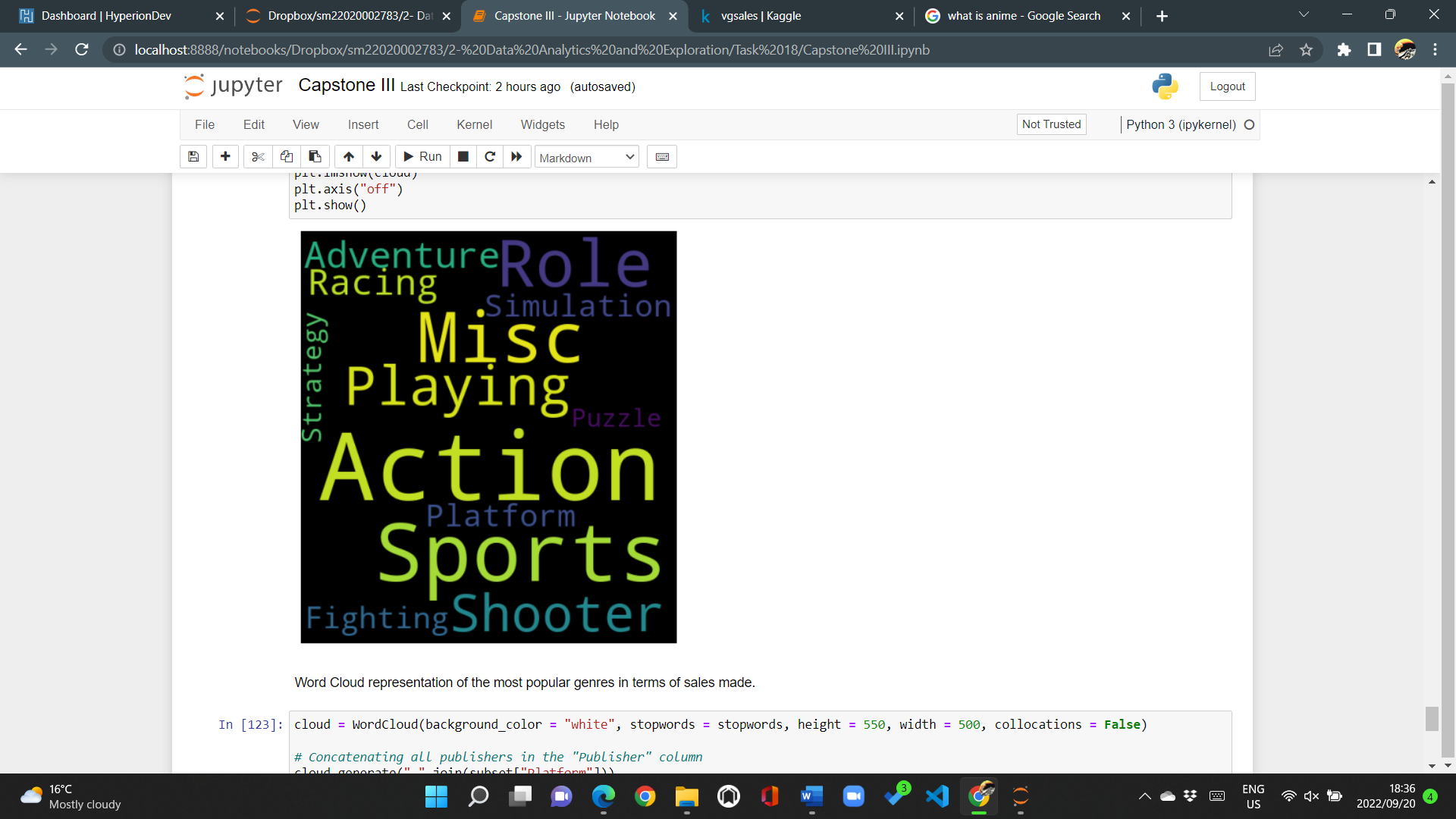
**Why Nintendo dominates**

As previously mentioned above, Nintendo has dominated sales in all the above countries indicating the level of appeal Nintendo holds to the gaming community. Nintendo, being a Japanese based company, dominates the sales volume in Japan. Though not common, the sales recorded do make sense as the Japanese market very clearly leans towards Japanese-origin gaming titles. Nintendo’s domination in video game sales exists across the world and is not only limited to Japan. The only logical explanation to this finding is Nintendo’s uniqueness in what the company offered to the gaming community.

Nintendo provided a new way of gaming by increasing the level of interactivity of the gaming world. Nintendo provided users with the ability to simulate game movement in real life in popular titles such as Wii Sports and Wii Sports Resort. As opposed to publishers such as Electronic Arts and Activision, players were restricted to game movement that only required the movement of few sets of fingers to interact with the game. Nintendo offered a platform that would allow for the distribution of gaming titles that require the player/user to make use of the full body. Nintendo could be considered a light alternative to exercise as competitive gaming became exciting and elevated the heartrate of the user while playing. In short, Nintendo offered a spin to the conventional way of gaming which many individuals have very clearly enjoyed. Nintendo’s dominance is attributed to the innovation of gaming that the company has introduced to the world.



The graph above reveals video game sales made based on the genre in Japan. “Role-playing” is the best-selling genre in the Japanese gaming market as visualised above. In Europe and North America, Action is considered the best-selling genre yet in Japan, Role-playing is a favourite. Without any additional context, few inferences as to why the Japanese market prefers role-playing as opposed to action like the other countries can be made. What may attribute the popularity in role-playing video games may be a result of Anime. A popular Japanese genre that is hand drawn and animated, often turned into comics, series, and movies too. Anime is immensely popular, not only in Japan, but across the world. Role-playing video games are the closest genre of games that simulate the ideas and core principles of anime therefore suggesting why the volume of video game sales for role-playing video games is higher than the other genres in Japan.



A screenshot of a computer

Description automatically generated

The above Word Cloud visualisations function as representations for the level of popularity in video game genre and gaming platform. As revealed above, Action is the favourite in global sales. The PlayStation 2 and Nintendo Wii and DS gaming platforms are the most popular being the most exclaimed of the platforms recorded in the dataset.

**THIS REPORT WAS WRITTEN BY: Simphiwe Motsoasele**

