CA Rough Work

Name: Martin Casey

X-Number: X00119025

Date: Wait till last minute

Overview

The CA in question is designed to take datasets from various sources and create visualisations in Tableau to make aspects of the data more visually noticeable. The purpose of creating visualisations out of the datasets is to make the data more readable to the average user, so that they can easily intercept the data and identify any correlations with the data.

For this CA the author is trying to create visualisations based on video game sales, to make it easier to intercept information’s about games sales, across different regions. The author is hoping to find significant disparities between some aspects of sales across the three different regions of North America, Europe and Japan. The aspects of the video games the author is keeping in mind relate to factors such as genre, platform, year of release, etc.

Background

The author came up with the idea to use video game sales data base the basis for the dataset they would turn into a visualisation, after hearing the requirements of the CA from the lecturer. Once the lecturer stated how much detail would be required, the author immediately though of video game sales, as based on their past interest they found there would be a lot of organisational data to work with when dealing with the sales of video games. Data that ranges from the game developers, publishers, up to the platforms and regions the games were released in.

After some time acquiring the dataset the author found that some prior analysis had been done on the subject in the past. Notably by a former student at the institute that the author studies in.

Datasets

The author got their video games dataset from the online dataset site known as Kaggle.

Seven Stages

Acquire

For the acquire phase the author acquired a dataset from Kaggle based on video games sales. The dataset was for the most part suitable but some values in different columns were missing, so it is likely that the ca work will require the use of a second dataset to meet the required work load.

Parse

In the parse phase the author began the basics of refining the data for use in tableau. For starters the columns that displayed the name of the video games was removed, as there were not many repeating values or noticeable trends in the data to indicate it would be data visualisation in tableau.

For now, all rows are being kept in the table, even the empty ones, as they will need to be examined in more detail at a later stage.

All rows where the value in the year column is “NA” are removed. This is because the year column needs to ensure that its values are in the year format so that calculations can be performed on them. Also, it would be very difficult to look up the exact years of release for the various video games.

All instances of “tbd” are removed from the user score column, as the values in the column would need to be in a number format. It would be too difficult to try and find the various user scores that were missing, especially since there is no clear source as to where the scores were taken from.

Filter

All columns that are associated with scores in the tables are now removed. This is because it has been deciding that the visualisation will mainly focus on the sales for the video games, as it would be easier to correlate game sales, with a games genre.

The other sales column has now been removed as it did not provide any good information as to the source of its sales, while the other sales columns clearly define their sales based on regions, and a combination of regions for the global sales.

The developer column has been removed as it is they are not nearly as relevant from a data analysis stage for examining the trends in video game sales, unlike the publishers.

Mine

All the ratings columns were removed from the table as they contained for too many white spaces, which would cause major issues for tableau when it comes to reading and converting the data.

References

<https://public.tableau.com/en-us/s/resources?qt-overview_resources=1#qt-overview_resources>

2/11/2017