CODING WITH R & SHINY



VIDEO GAME SALES

Course Instructor: Prof. Guillaume Coqueret

Report for Shiny App

By:

BANERJEE Hiya

emlyon business school

2021-222

Table of Contents

About the Dataset

Video Games Sales Dataset

This dataset is sourced from Kaggle and contains a list of videogames that were released between the years 1980 and 2016.

Through this dataset, the objective is to understand the sales trend of the dataset over the duration period.

It is to further find out if there were –

- significant shifts in sales pattern in the different decades,
- to understand the reasons behind this shift and if there was a company playing monopoly over the videogame market;
- And to also understand the change in the preferences of the videogame geeks out there.

The raw dataset contains the following variables:

a) Rank
b) Name
c) Platform
d) Genre
e) Publisher
f) Year of release
g) Sales over different regions (North America, Europe, Japan, Other, Global)

Forming 'Tidy Data'

The main libraries used to clean this dataset is **tidyverse**.

First we clean the dataset so that it satifies the conditions of the tidy format. This involved the following –

a) Renaming the columns so that they have short names in small letters with no spaces using the function **mutate()**.

b) Checking the data type of the columns using the function **sapply(data, typeof)** and setting them to the correct datatype.

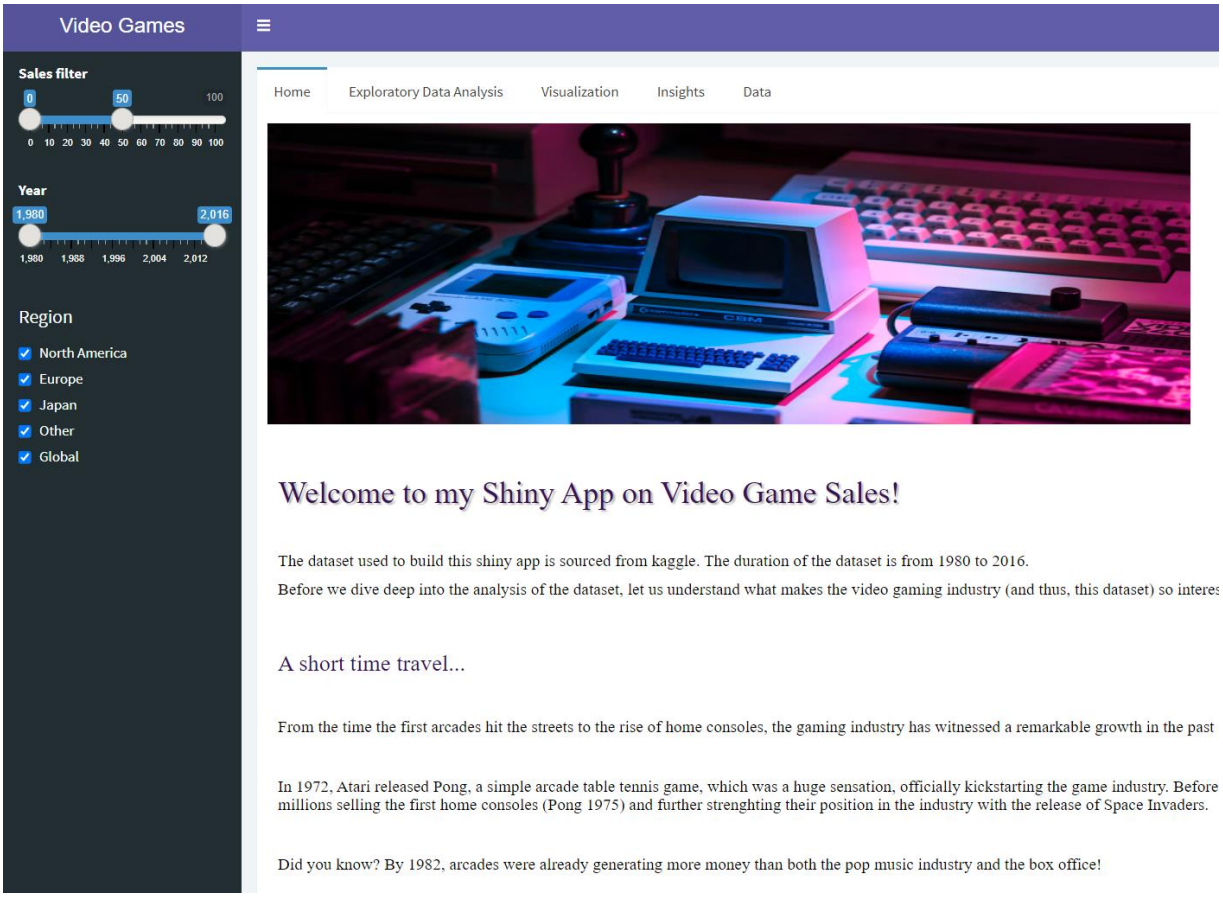
c) Removing duplicate values and checking for Nan values.

   a. In this dataset, the 'year' column had missing values for the years, 2017 and 2020, which were removed.
   b. On observing the last few rows of the dataset, we also observe that there are games with zero sales values for all regions. However, these rows were retained based on the assumption that these games were not released in the regions and hence record a sales value of zero.

d) Applying the function **pivot_longer** to the dataset to ensure all rows are observations and all columns are variables.

Initial exploratory analysis reveals there are - 31 platforms, 12 genres and 566 publishers. The summary statistics of the data does not provide a lot of accurate insight due to the presence of videogames for which sales value is zero.

The Shiny App



The Shiny App made with the video games sales dataset features many functionalities as observed from the home page above. It allows users to explore the different variables present in the dataset to get better insights about the sales trend in the industry.

For instance, it makes it possible for a user to find out the effect of just one genre on the total sales of a particular region. This is just one of the many other things that the app makes possible.

The benefit of using the app is efficiency. With the app, the user can adjust the parameters to obtain exactly what they're looking instead of working on huge CSV files!

The customized interface is just a bonus!

Let us now explore in detail the workings of each component of the app.

The Sidebar

The Sidebar features the following widgets:

1. Sales Filter: Using the slider, the user can decide the range of sales for which they want to see the video games. The slides helps the user to set their own upper limit and lower limit.

2. Year Filter: Works in a similar way to the sales filter. With this, the user can set a particular range to view the video games released only during the selected period.

3. Region Checkbox: The checkbox allows users to further narrow down their research to only viewing the video games that were released in a particular region, or in the desired regions. Multiple regions can be selected at the same time.

All the three sidebar widgets combined provide efficient ways for users to find their desired data. If a user wants to view only the video games released in Europe that sold above a particular sales figure between 1990 and 1995, they can now do it with the help of the app.

The Dashboard

The dashboard features 5 panels, namely – Home, Exploratory Data Analysis, Visualization, Insights and Data.

1. **Home**

   The home panel, also the home page, tells the user about what makes the video games industry so interesting. It outlines a short history of the evolution of the industry over the years.

2. **Exploratory Data Analysis (EDA)**

   The EDA panel provides information such as the size of the dataset, the presence/absence of missing values, how the data was converted into a tidy format and the major differences between the raw data and the tidy data.

3. Visualization

The visualization panel is further sub-divided into 5 tabs featuring plots. All the plots are interactive customizable using the sidebar widgets.

Sales

It features two plots mapping –

i. total sales over the duration

ii. average sales over the duration

Genre

It features three plots mapping –

i. genre vs sales

ii. genre vs sales and region (can also be achieved using widgets)

iii. distribution of genres most sold over the years

Platform

It features a pie plot and distribution plot and allows the user to pick the particular genres that they would like to view the result for. This gives insight into which were the popular platforms for a particular genre.

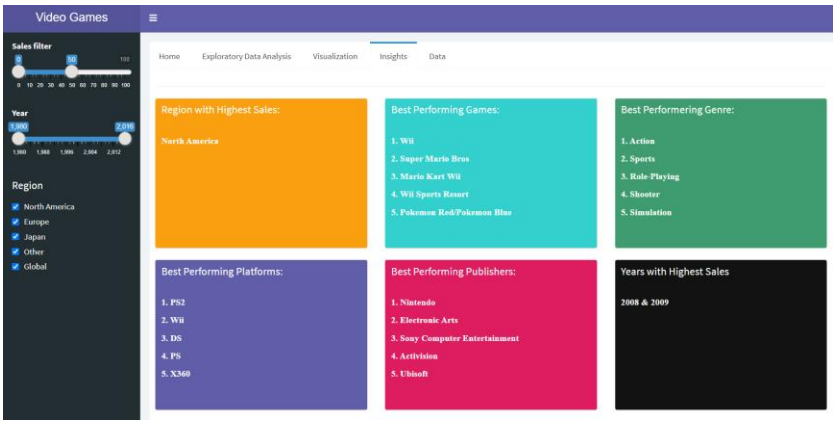
Publisher

Similar to platform, it gives insight into which were the popular publishers for a particular genre.

Games

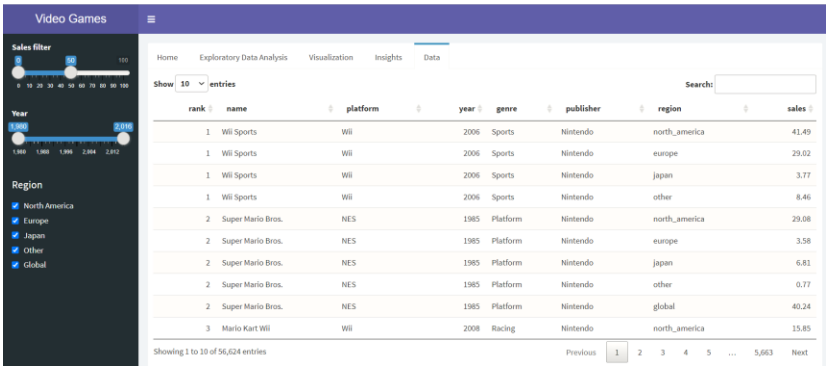
This features a plot displaying the top 50 bestselling games. Using the genre widget, it can be customized to view the top 50 bestselling games for a particular time period.

4. Insights



The insights panel summarizes the best performers of the dataset providing a quick reference to the user, to make swift comparisons.

5. Data



As the name suggests, this tab provides the user with the tidy data table containing all the fields and observations. It can be customized with the sidebar widgets and the user can choose the desired number of entries they'd like to view. It also provides a search bar to search for specific observations.