Data Science Challenges

VIDEO GAME SALES
ANALYZE SALES DATA FROM MORE THAN 16,500 GAMES

Hany Elshafey

Data science challenges
Reveal data secrets

Short Story of Video Games

Data science challenges

Reveal data secrets

The Data

VIDEO GAME SALES ANALYZE SALES DATA FROM MORE THAN 16,500 GAMES.

THIS DATASET CONTAINS A LIST OF VIDEO GAMES WITH SALES GREATER THAN 100,000 COPIES. IT WAS GENERATED BY A SCRAPE OF VGCHARTZ.COM.

***FIELDS INCLUDE

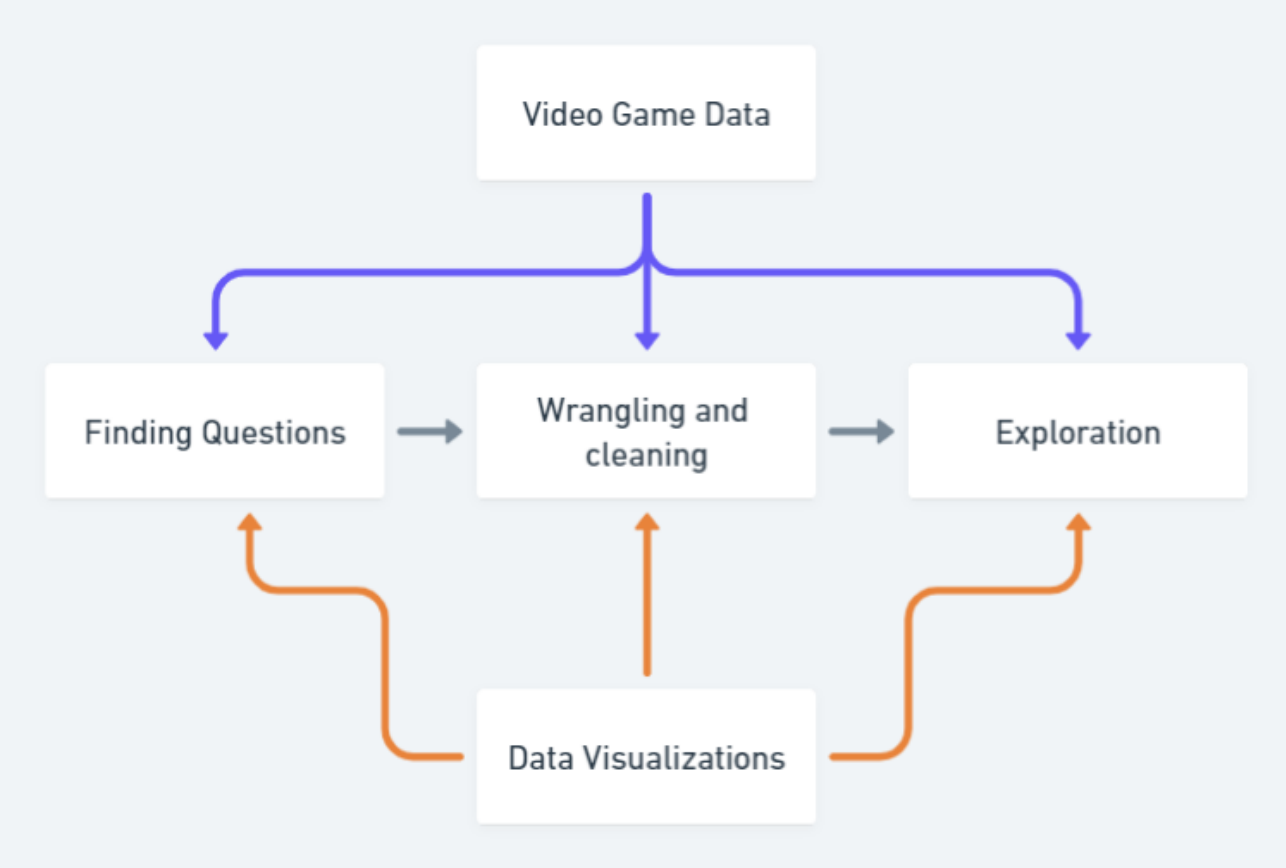
RANK - RANKING OF OVERALL SALES
NAME - THE GAMES NAME
PLATFORM - PLATFORM OF THE GAMES RELEASE (I.E. PC,PS4, ETC.)
YEAR - YEAR OF THE GAME'S RELEASE
GENRE - GENRE OF THE GAME
PUBLISHER - PUBLISHER OF THE GAME
NA_SALES - SALES IN NORTH AMERICA (IN MILLIONS)
EU_SALES - SALES IN EUROPE (IN MILLIONS)
JP_SALES - SALES IN JAPAN (IN MILLIONS)
OTHER_SALES - SALES IN THE REST OF THE WORLD (IN MILLIONS)
GLOBAL_SALES - TOTAL WORLDWIDE SALES.
THERE ARE 16,598 RECORDS. 2 RECORDS WERE DROPPED DUE TO INCOMPLETE INFORMATION.

Data science challenges
Reveal data secrets

Introduction

VIDEO GAMES HAVE BEEN AROUND FOR A VERY LONG TIME. THEY WERE FIRST INTRODUCED IN THE 1950S BUT DID NOT REACH THE MAINSTREAM PUBLIC TILL THE 1970S. THE POPULARITY OF VIDEO GAMES INCREASED IMMENSELY DURING THIS DECADE WITH THE INTRODUCTION OF GAMING ARCADES AND MANY HOME CONSOLES. SOON, VIDEO GAMES BECAME AVAILABLE ON HOME COMPUTERS. VIDEO GAMES ALSO REACHED MOBILE PHONES WHEN THEY LAUNCHED. THUS, VIDEO GAMES BRANCHED INTO 3 MAIN TYPES OF PLATFORMS: HOME COMPUTERS, GAMING CONSOLES, AND MOBILE PHONES. THIS ANALYSIS TRIES TO FIND THE TURNING POINT IN VIDEO GAME SALES.

Project FlowChart



Data science challenges
Reveal data secrets

Data Cleaning

Issues

Quality

1-271 value in year column is Nan and 58 value in Publisher column is Nan
2-year Dtype shoud be Date type
3-global sales column is not equal to the summation of (NA_Sales EU_Sales JP_Sales Other_Sales)

Tidiness

Columns NA_Sales EU_Sales JP_Sales Other_Sales was named in a confusing way. names should be clear

```
|:  1  df_vg.isna().sum()
executed in 28ms, finished 17:19:42 2021-08-03

[25]: Rank            0
      Name            0
      Platform        0
      Year            0
      Genre           0
      Publisher       0
      NA_Sales        0
      EU_Sales        0
      JP_Sales        0
      Other_Sales     0
      Global_Sales    0
      id              0
      dtype: int64
```

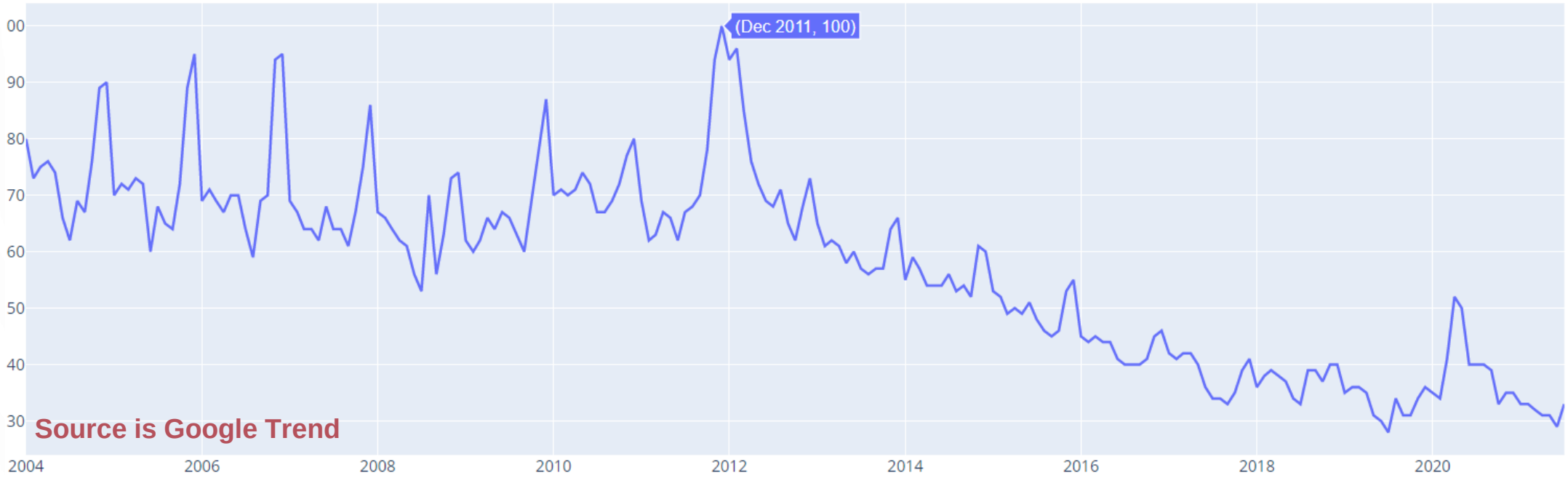
	Rank	Name	Platform	Year	Genre	Publisher	NorthAmerica_Sales	Europe_Sales	Japan_Sales	Other_Sales	Global_Sales	id
0	1	Wii Sports	Wii	2006	Sports	Nintendo	41.49	29.02	3.77	8.46	82.74	0
1	2	Super Mario Bros.	NES	1985	Platform	Nintendo	29.08	3.58	6.81	0.77	40.24	1
2	3	Mario Kart Wii	Wii	2008	Racing	Nintendo	15.85	12.88	3.79	3.31	35.83	2
3	4	Wii Sports Resort	Wii	2009	Sports	Nintendo	15.75	11.01	3.28	2.96	33.00	3
4	5	Pokemon Red/Pokemon Blue	GB	1996	Role-Playing	Nintendo	11.27	8.89	10.22	1.00	31.38	4
5	6	Tetris	GB	1989	Puzzle	Nintendo	23.20	2.26	4.22	0.58	30.26	5

Data Frame shape after Cleaning: `(16291 Row , 12 Column)`

Data science challenges
Reveal data secrets

DATA ANALYSIS

The world interest level in video games over time



Source is Google Trend

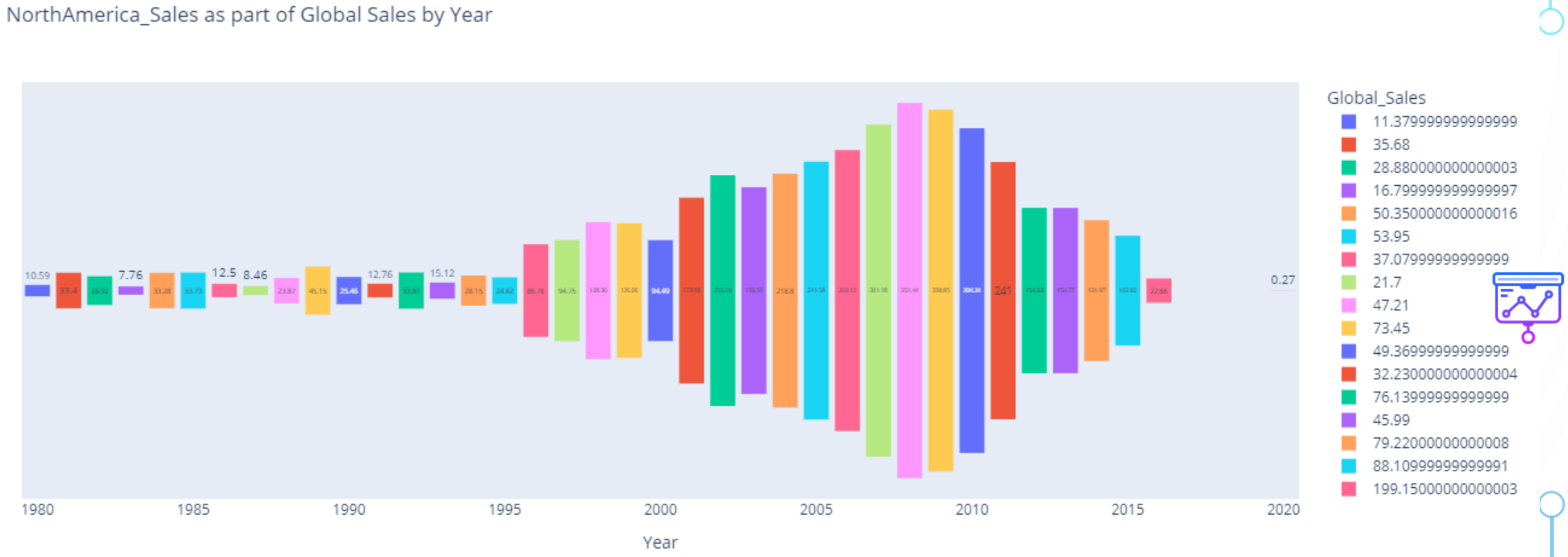
The Insights

The world interest level in video games started declining after a year 2011
From Now Our mission is to find the Causality

Data science challenges
Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year



The Insights

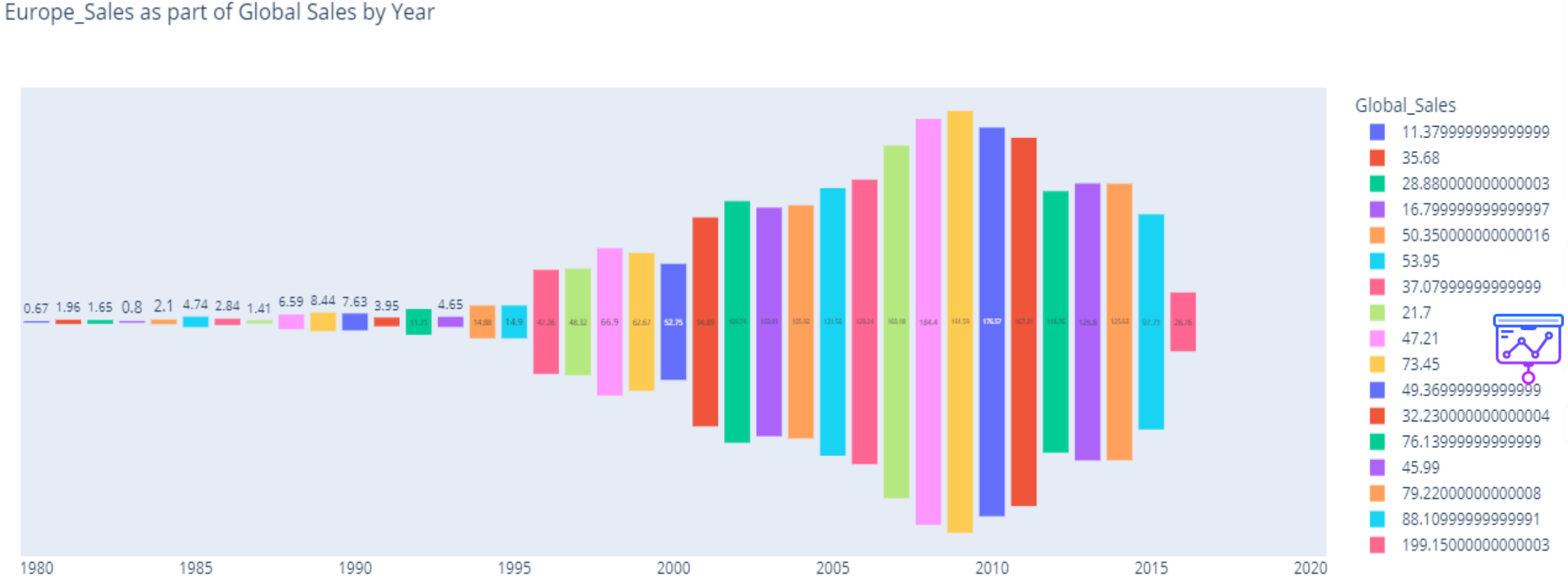
2008 is the best seller in North America with Value 351 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year



The Insights

2009 is the best seller in Europe with Value 191.5 millions

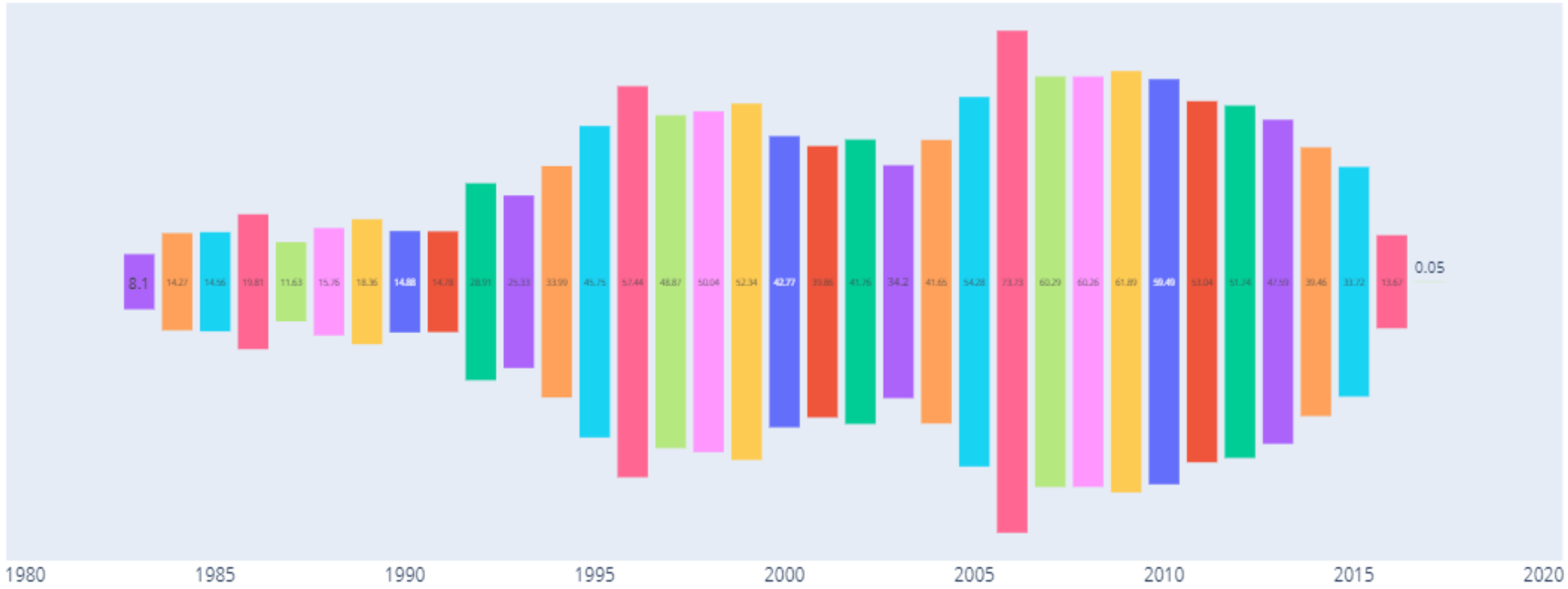
Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year

Japan_Sales as part of Global Sales by Year



The Insights

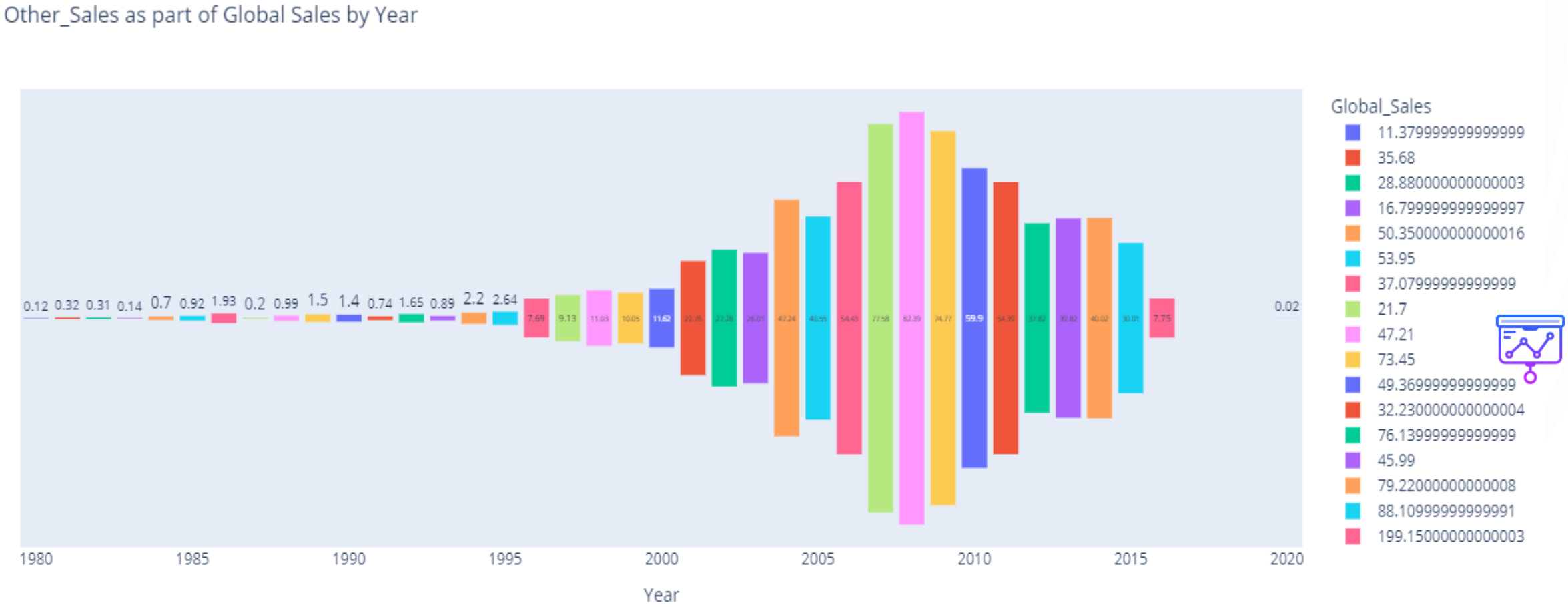
2006 is the best seller in Japan with Value 73.3 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year



The Insights

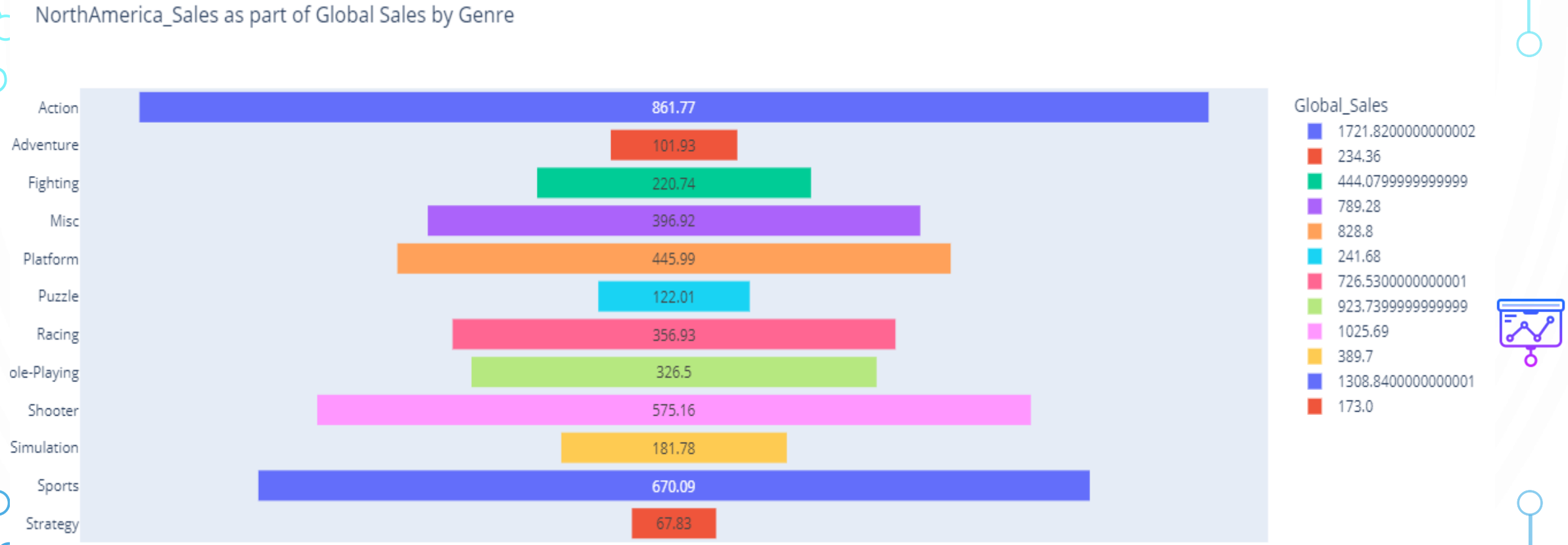
2008 is the best seller for other sales with Value 82.4 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Genre



The Insights

Action Games is the best seller in North America with Value 861.77 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Genre

The Insights

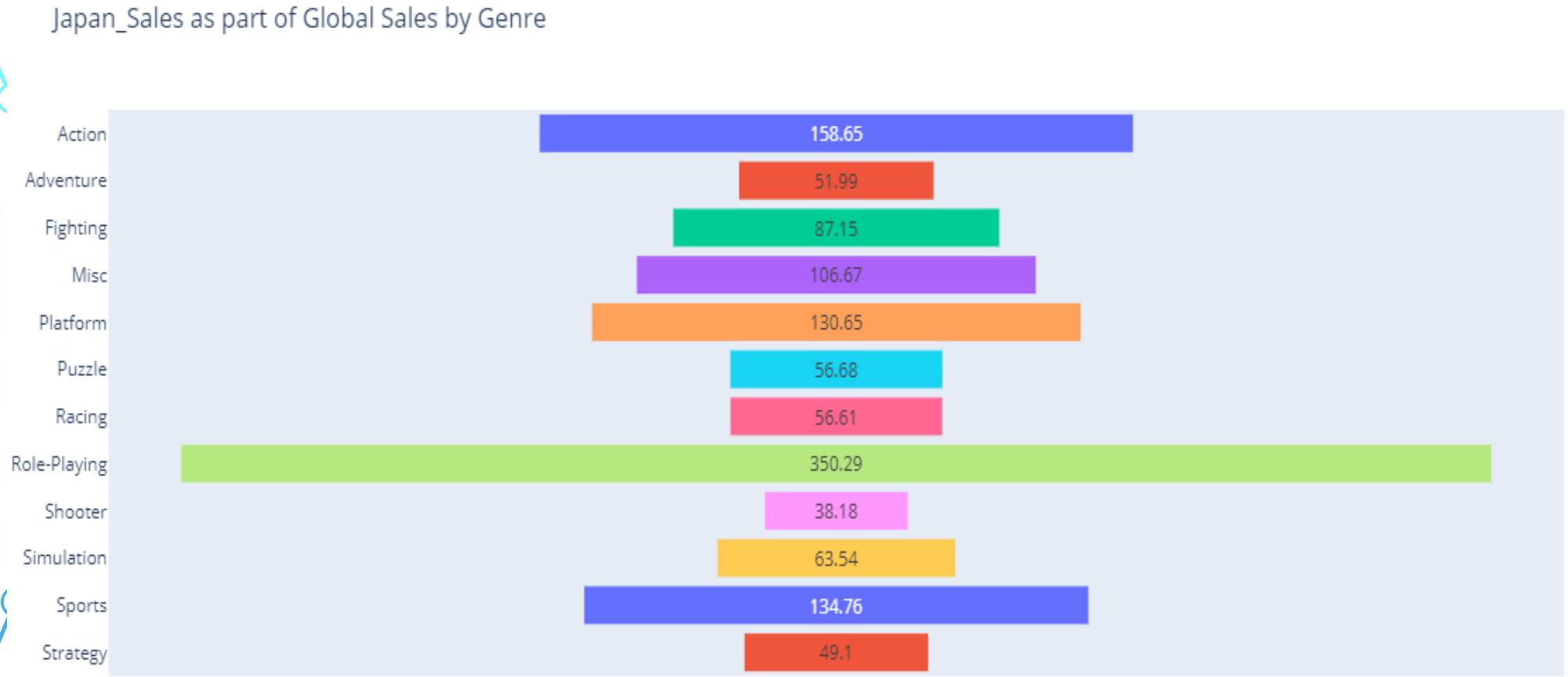
Action Games is the best seller in Europe with Value 516.48 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Genre



The Insights

Now there is return point in Japan as Role-Playing Games is the best seller in Japan with Value 350.29 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Genre

The Insights

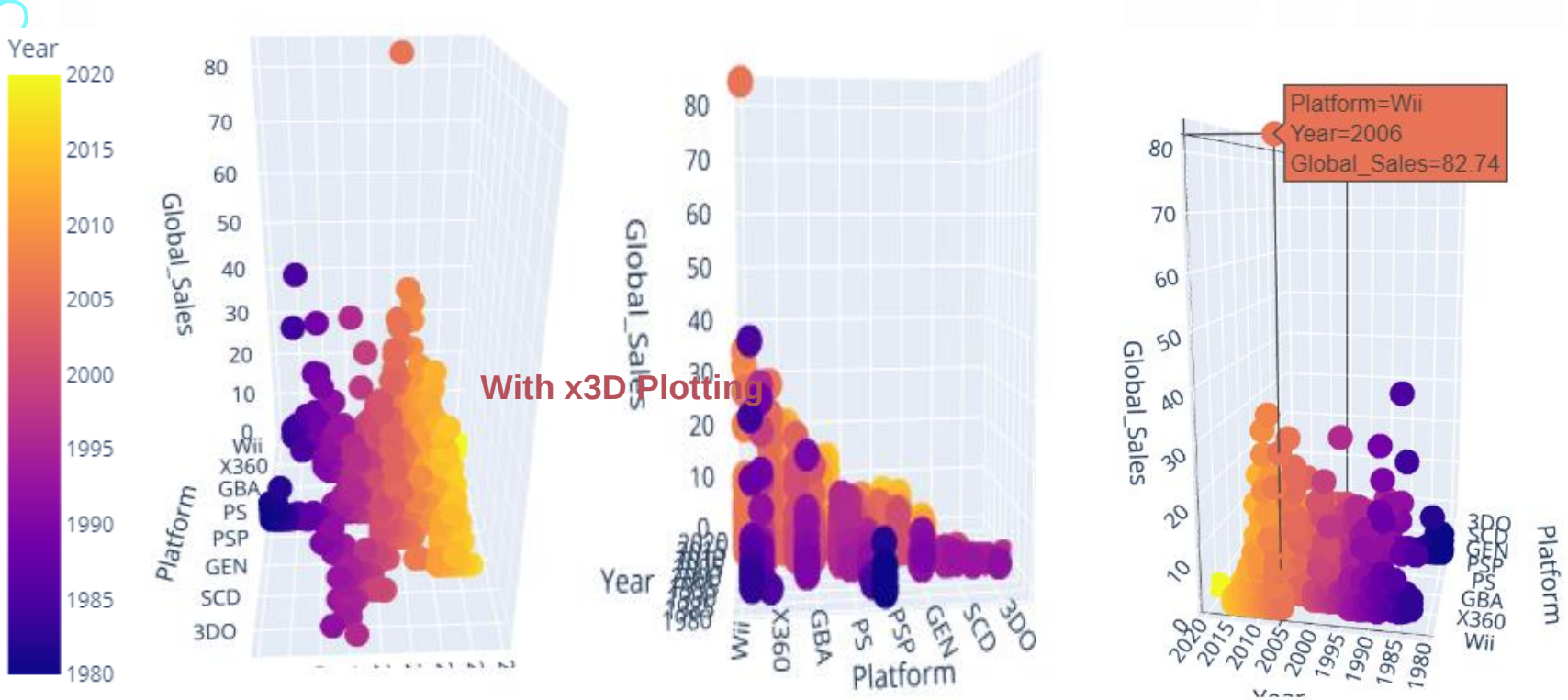
Actions Games is the best seller for other sales with Value 184.92millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Data Exploration With x3D Plotting



With x3D Plotting

The Insights

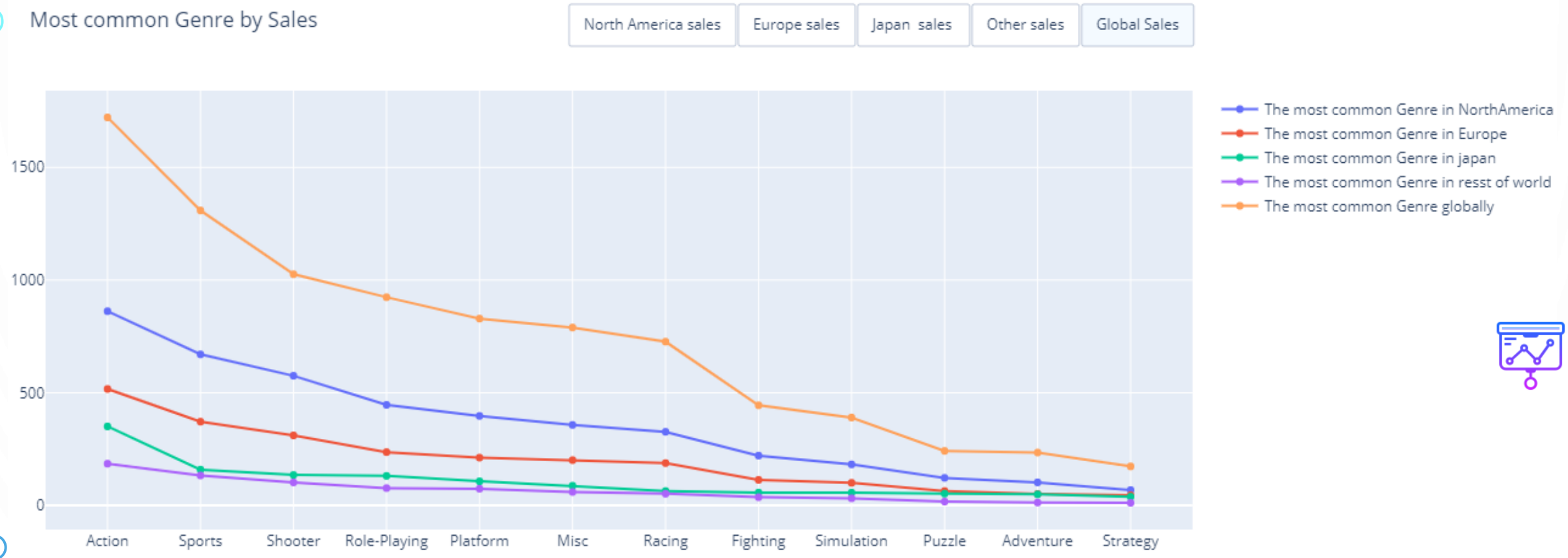
Wii Platform is the best seller in 2006 with Value 82.7 millions

Data science challenges

Reveal data secrets

DATA ANALYSIS

Most common Genre by sales



The Insights

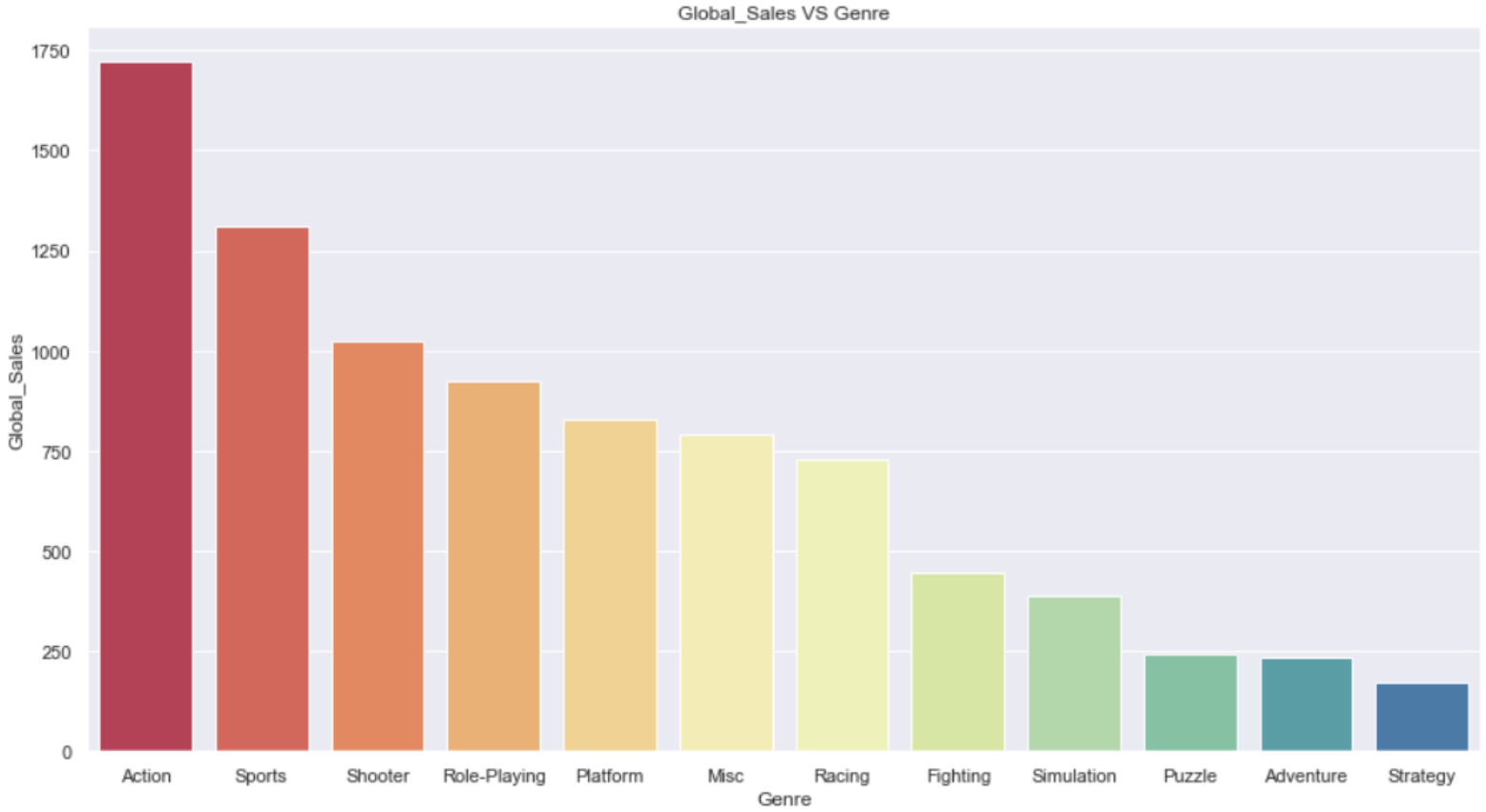
Most common Genre by sales is Action , Sports and Shooter

Data science challenges

Reveal data secrets

DATA ANALYSIS

Which genre game has sold the most in a single year?



The Insights

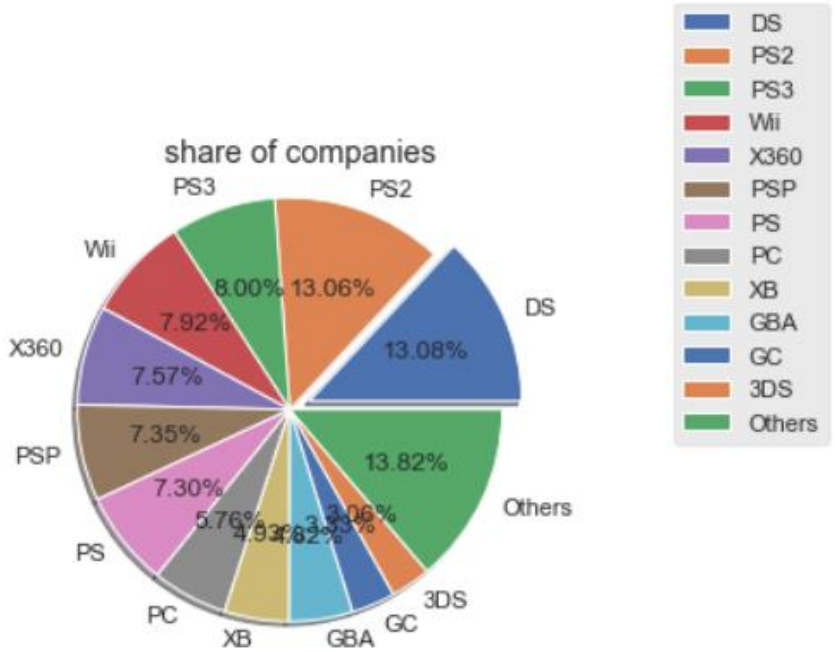
Action , Sports and Shooter are always The Global's best-selling.

Data science challenges

Reveal data secrets

DATA ANALYSIS

The Platforms percentage



The Insights

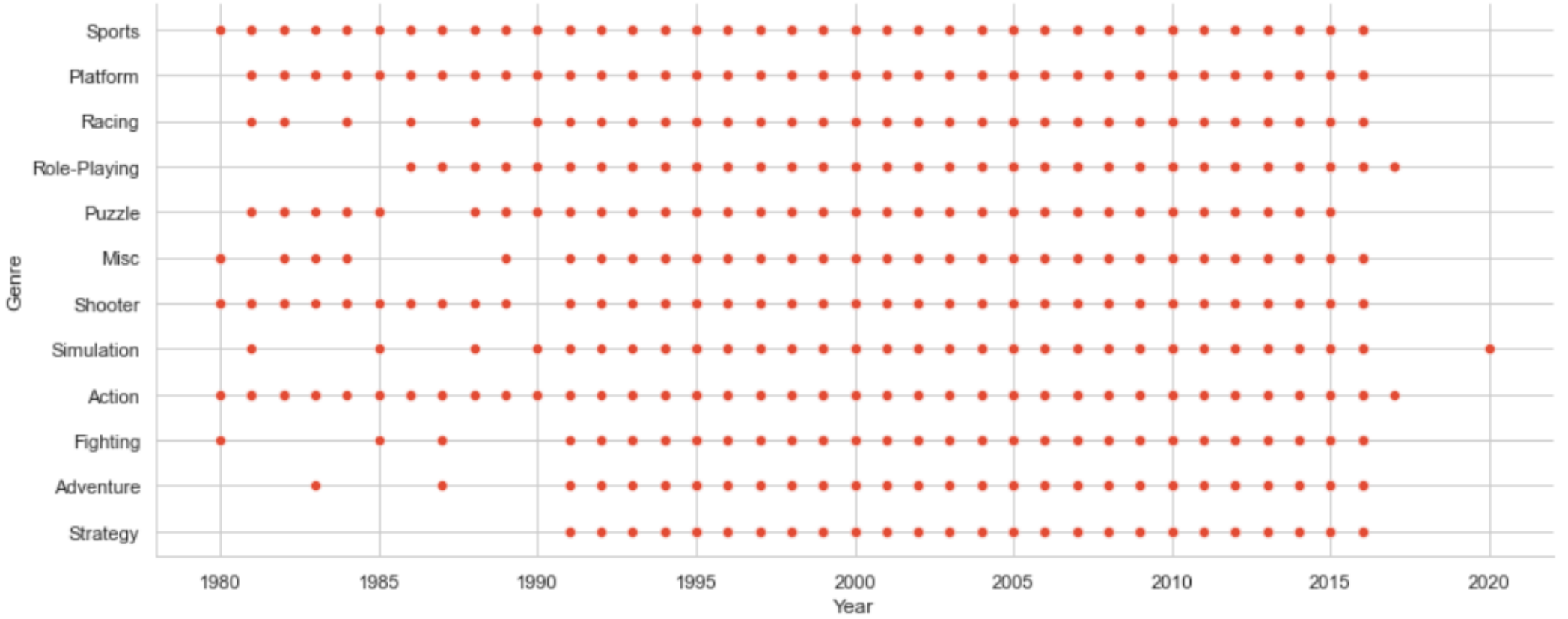
PS2 , PS3 and Wii are always highly percentage

Data science challenges

Reveal data secrets

DATA ANALYSIS

What year did the games start and when did they stop, depending on the genre?



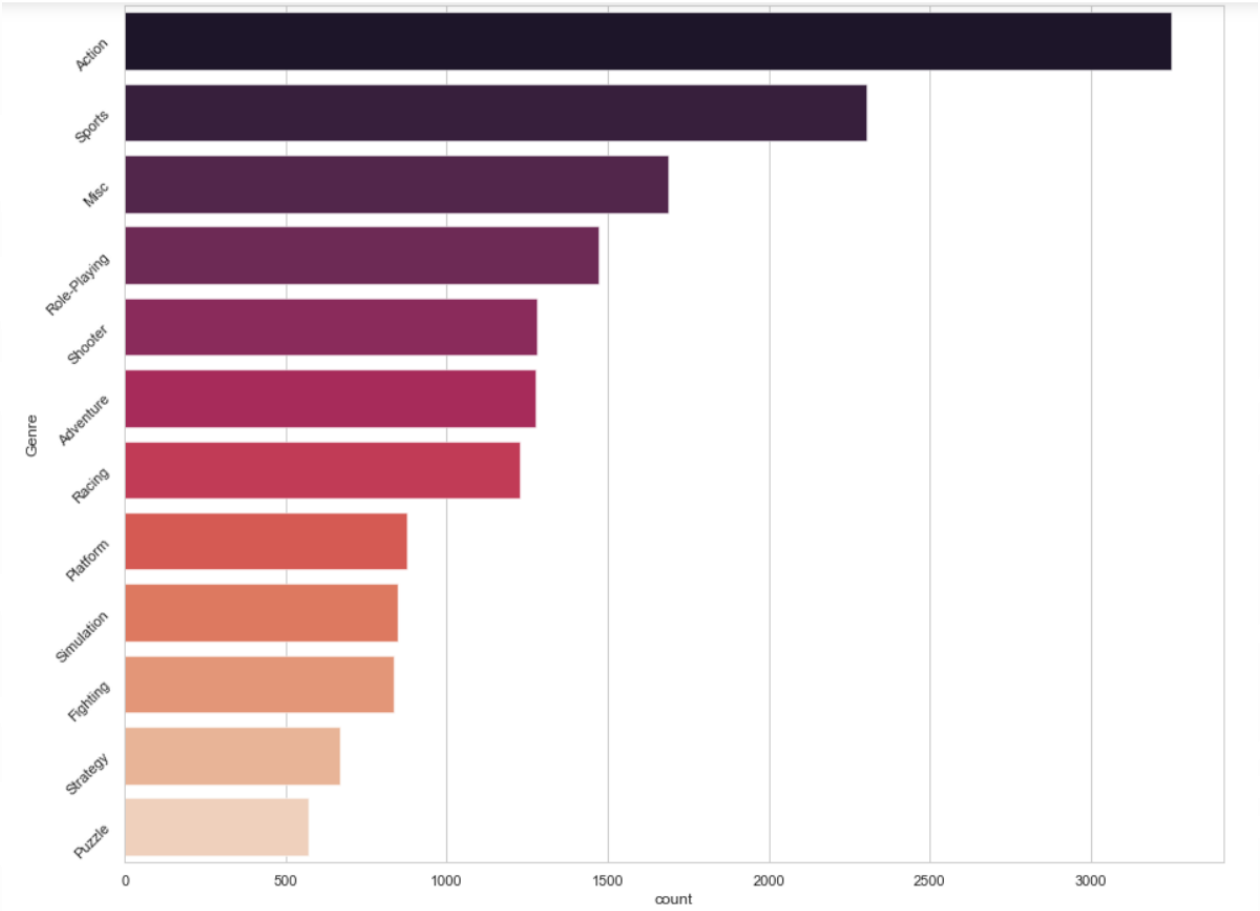
The Insights

- Sports Games started from 1980 to 2016 then stoped. .
- Simulation games stoped with Sports games and appear again in 2020. .
- strategy games started from 1991 to 2016 .

Data science challenges
Reveal data secrets

DATA ANALYSIS

What is the most popular game ?



The Insights

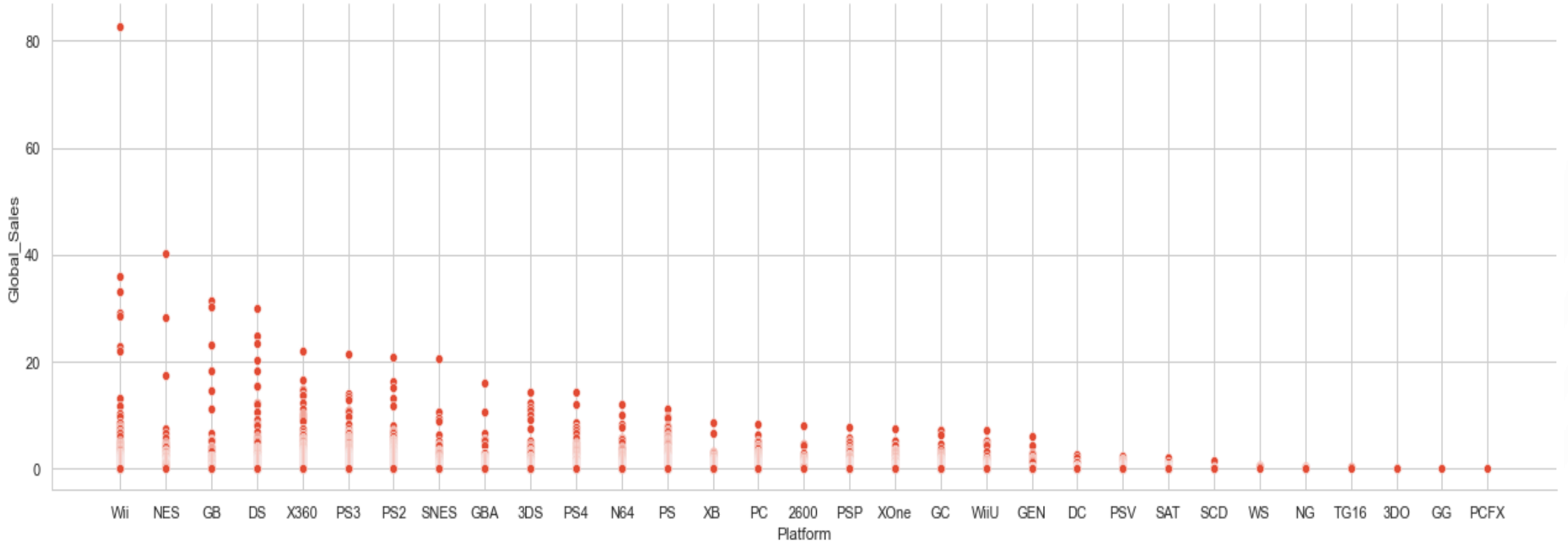
- Action and Sports Games are the most popular than others ..

Data science challenges

Reveal data secrets

DATA ANALYSIS

Which platform with the highest price individual game globally?



The Insights

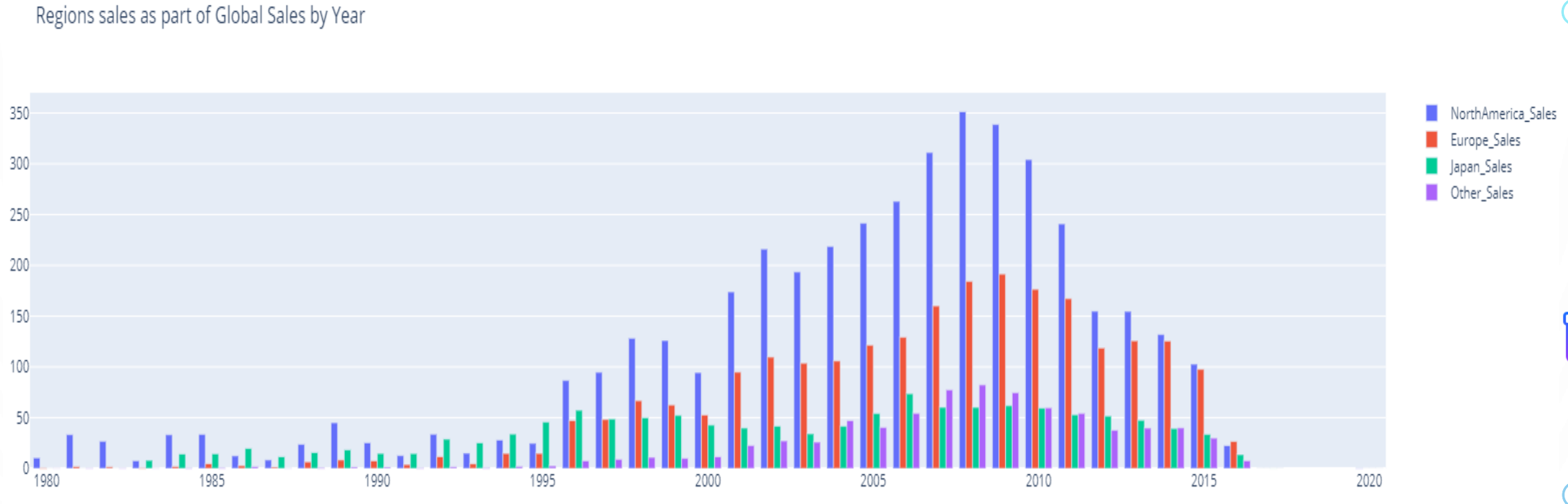
- WII Sports get the highest price individual game globally with sports genre ..

Data science challenges

Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year



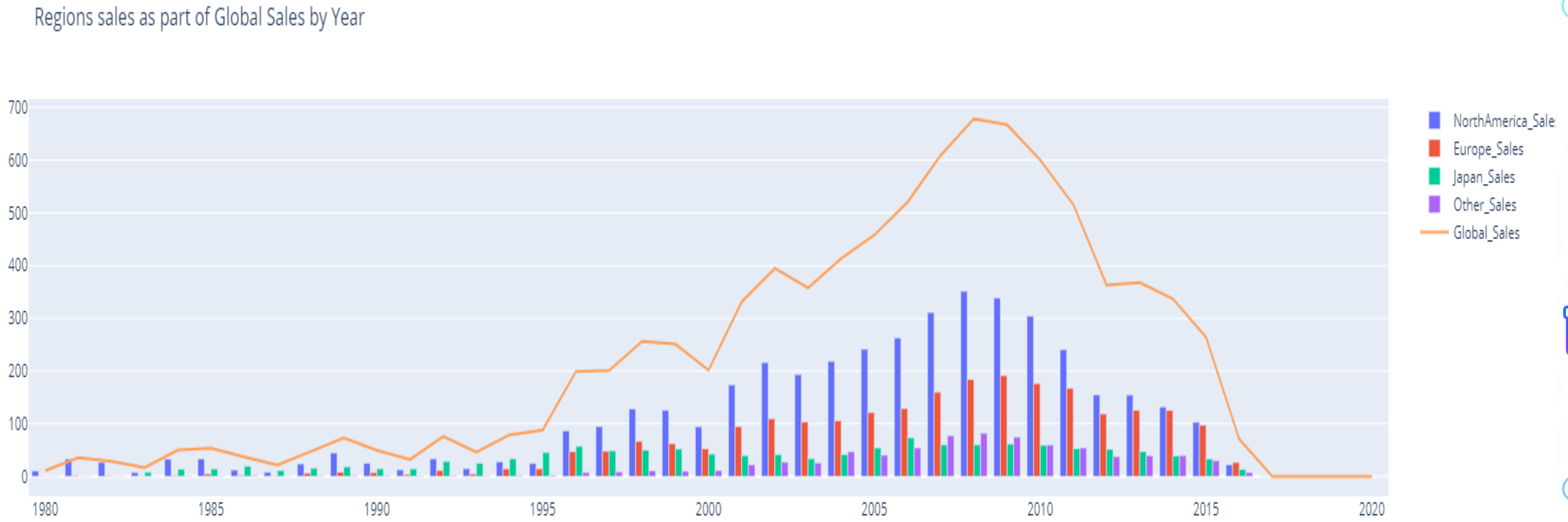
The Insights

- North America is always been the best seller Before Europe.

Data science challenges
Reveal data secrets

DATA ANALYSIS

Regions sales as part of Global Sales by Year



The Insights

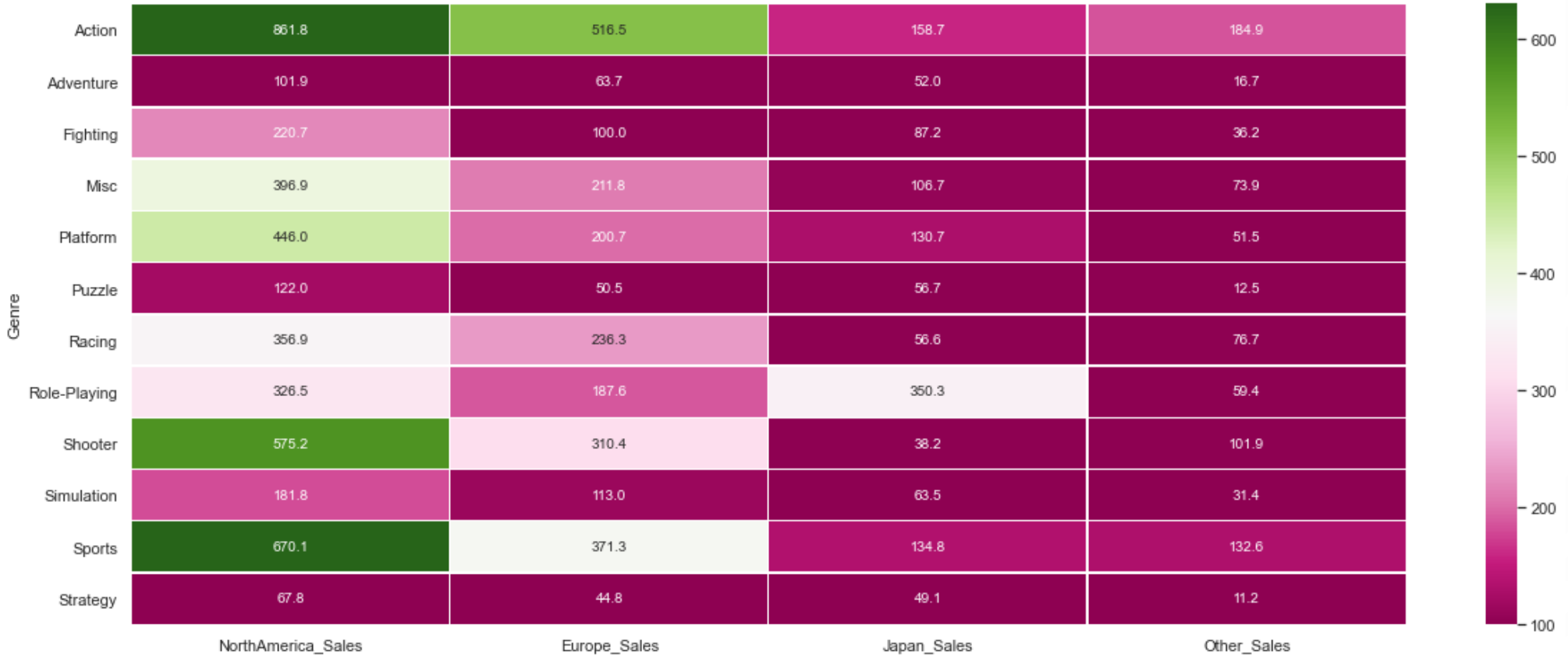
- 2008 is the best-selling year.Globally

Data science challenges

Reveal data secrets

DATA ANALYSIS

Which platform with the highest price individual game globally?



The Insights

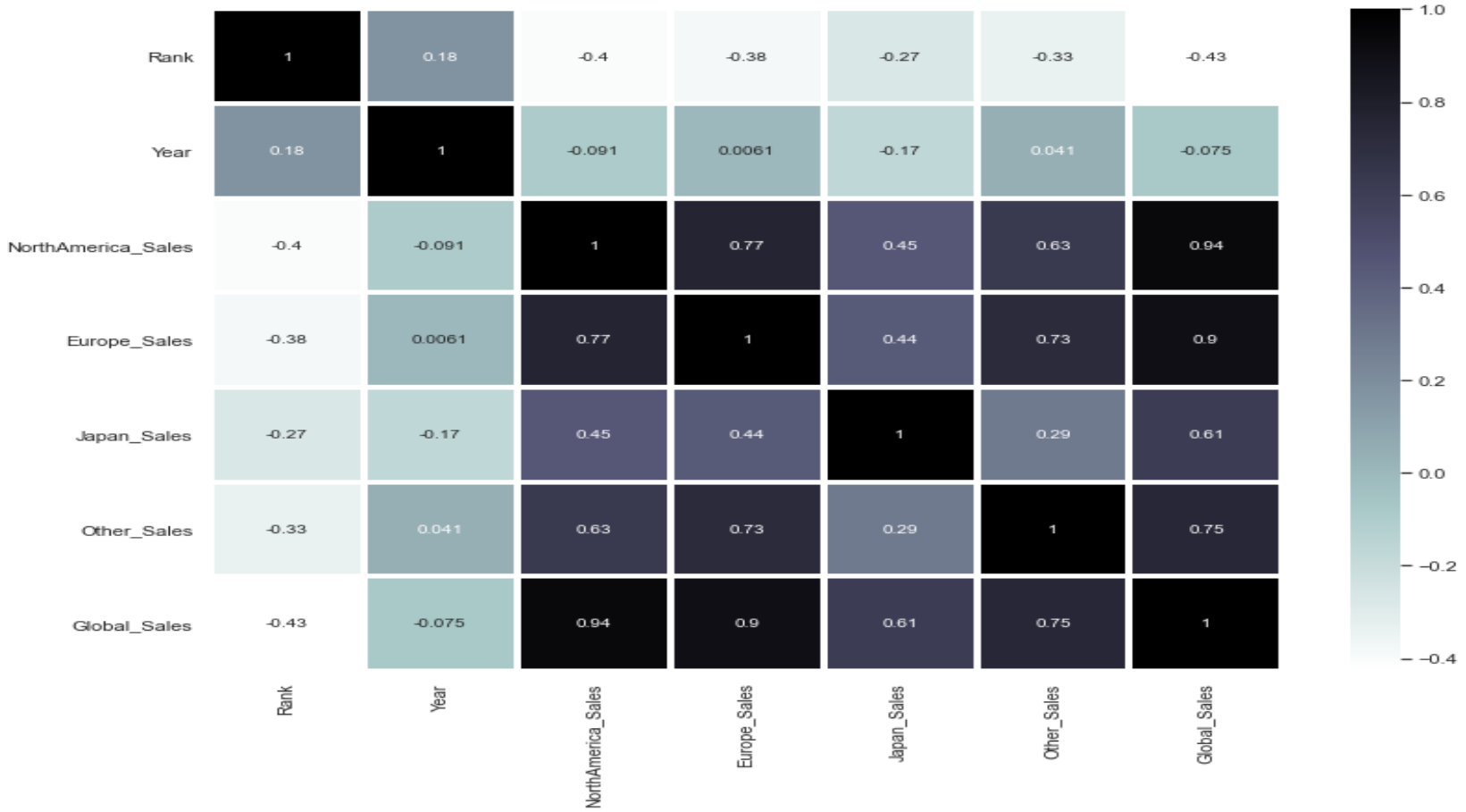
- North America is the most popular region for video games without a competitor ..

Data science challenges

Reveal data secrets

DATA ANALYSIS

The biggest influencer in the world price ?



The Insights

- North America and Europe is The biggest influencer in the world price without a competitor .
- There is a good correlation Between NorthAmerica_sales , Europe_slaes And Global_sales

Data science challenges

Reveal data secrets

Finally, we finished our main mission

The reason for the decline in game sales after 2008 is the emergence of the smartphone

Data science challenges

Reveal data secrets

Machine Learning

Linear Regression Model

- I used NorthAmerica_sales , Europe_slaes with 'Platform', 'Genre', 'Publisher' to predict Global_sales Because of the correlation between them

	precision	recall	f1-score	support
0.0	0.98	0.98	0.98	4011
1.0	0.89	0.89	0.89	877
accuracy			0.96	4888
macro avg	0.93	0.93	0.93	4888
weighted avg	0.96	0.96	0.96	4888

The Accuracy

executed in 27ms, finished 17:20:32 2021-08-03

```
Linear Regression Accuracy in the training data : 96.634049891285 %
Linear Regression Accuracy in the test data : 96.14793318822102 %
```

Data science challenges

Reveal data secrets

Machine Learning

Then I tried another Algorithms like

GradientBoostingRegressor Model:

executed in 1.04s, finished 18:04:59 2021-08-03

```
GradientBoostingRegressor Accuracy in the training data : 98.40411906463221 %%!!
GradientBoostingRegressor Accuracy in the test data : 94.24531253598228 %
```

DecisionTreeRegressor Model:

executed in 73ms, finished 18:04:59 2021-08-03

```
DecisionTree Accuracy in the training data : 99.83845899573934 %
DecisionTree Accuracy in the test data : 92.6551175983054 %
```

RandomForestRegressor Model

```
RandomForest Accuracy in the training data : 98.09633717549644 %
RandomForest Accuracy in the test data : 96.07290616702541 %
```

SVR Model

```
SVR Accuracy in the training data : 57.108483863643514 %
SVR Accuracy in the test data : 67.01127829545256 %
```

Data science challenges

Reveal data secrets

Machine Learning

Then I tried another Algorithms like

GradientBoostingRegressor Model:

executed in 1.04s, finished 18:04:59 2021-08-03

```
GradientBoostingRegressor Accuracy in the training data : 98.40411906463221 %%!!
GradientBoostingRegressor Accuracy in the test data : 94.24531253598228 %
```

DecisionTreeRegressor Model:

executed in 73ms, finished 18:04:59 2021-08-03

```
DecisionTree Accuracy in the training data : 99.83845899573934 %
DecisionTree Accuracy in the test data : 92.6551175983054 %
```

RandomForestRegressor Model

```
RandomForest Accuracy in the training data : 98.09633717549644 %
RandomForest Accuracy in the test data : 96.07290616702541 %
```

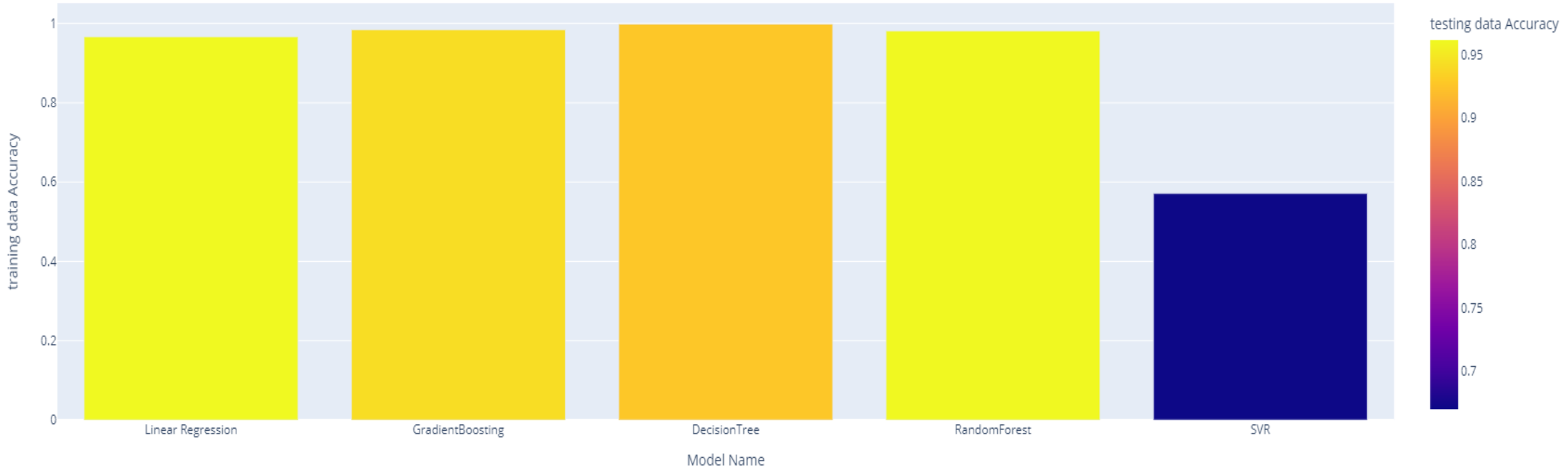
SVR Model

```
SVR Accuracy in the training data : 57.108483863643514 %
SVR Accuracy in the test data : 67.01127829545256 %
```

Data science challenges

Reveal data secrets

Machine Learning



The Insights

The least accurate is the output from the SVR algorithm

Data science challenges

Reveal data secrets

I THANK DR : DOAA AND ANYONE PUSH TO SUCCESS

ALL THANKS FOR ALL OF YOU

Hany Elshafey