**SALES INSIGHTS**

**A PROJECT REPORT**

***Submitted by***

**HARSHIL JAYANTBHARTHI GOSAI**

**180410107031**

***In fulfillment for the award of the degree of***

**BACHELOR OF ENGINEERING**

**in**

**Computer Engineering**

**Sardar Vallabhbhai Patel Institute of Technology, Vasad**

**Gujarat Technological University, Ahmedabad**

**April, 2022**

**Sardar Vallabhbhai Patel Institute of Technology**

**SVIT Road, Rajupura Village, Vasad, Anand, Gujarat 388306**

### CERTIFICATE

This is to certify that the project report submitted along with the project entitled **Sales Insights** has been carried out by **Harshil Jayantbharthi Gosai** under my guidance in fulfillment for the degree of Bachelor of Engineering in Computer Department, 8th Semester of Gujarat Technological University, Ahmedabad during the academic year 2021-2022.

Asst. Prof. Shrina Patel Dr. Neha Soni

Internal Guide Head of the Department

**Sardar Vallabhbhai Patel Institute of Technology**

**SVIT Road, Rajupura Village, Vasad, Anand, Gujarat 388306**

### DECLARATION

We hereby declare that the Internship / Project report submitted along with the Project entitled **Sales Insights** submitted in partial fulfillment for the degree of Bachelor of Engineering in Computer Department to Gujarat Technological University, Ahmedabad, is a bonofide record of original project work carried out by me Sardar Vallabhbhai Patel Institute of Technology, Vasad under the supervision of Asst Prof Shrina Patel and that no part of this report has been directly copied from any students’ reports or taken from any other source, without providing due reference.

|  |  |  |
| --- | --- | --- |
|  | Name of the Student | Sign of Student |
| 1 |  |  |
|  |  |  |

### ABSTRACT

This project will give you feel of how data analysis projects are executed in big companies. This project is based on a computer hardware business which is facing challenges in dynamically changing market. Here, sales director decides to invest in data analysis and would like to build PowerBI/Tableau Dashboard that can give you real time sales insights.

Data can help businesses to better understand their customers, improve their advertising campaigns, personalize their content and improvise their bottom lines. Raw data has a lot of potential, but you need data analysis tools and libraries to unlock the power to grow your business. Data Analysis is important in business to understand the problems faced by an organization and to explore data in meaningful way. Data Analysis organizes, interprets structures and present the data into useful information that provides content for the data.

In our case, the Sales Director is facing problems in reading the raw data i.e., excel file or csv file of transactions. So, he decided to invest in Data Analysis which will provide him proper visualization of his sales data along with the hidden insights. Building PowerBI/Tableau dashboard, along with two models:

* First, model will predict the games that has become popular/flop in the past years.
* Second, model will be providing important features that will be helpful for greater sales of new games.

This will help him to take decisions to grow his sales, manage and structure the items in his store. Also, he will able to take better decisions.

**List of Figures**

Figure 1 Gantt Chart.…………………………………………………………………………13

Figure 2 Timeline Chart……………………………………………………………………...13

Figure 3 Dashboard…………………………………………………………………………..19

Figure 4 Detailed Analysis…………………………………………………………………...20

Figure 5 Detailed Analysis – 2……………………………………………………………….20

Figure 6 Genre vs Critic Score……………………………………………………………….21

Figure 7 Developer vs Critic Score…………………………………………………………..21

Figure 8 Publisher vs Critic Score……………………………………………………………22

Figure 9 Platform vs Critic Score…………………………………………………………….22

Figure 10 Game count by Platform…………………………………………………………..23

Figure 11 Game count by Developer………………………………………………………...23

Figure 12 Game count by Publisher………………………………………………………….23

Figure 13 Game count by Genre……………………………………………………………..24

Figure 14 Hits………………………………………………………………………………...24

Figure 15 Flop………………………………………………………………………………..24

Figure 16 Outliers……………………………………………………………………………25

Figure 17 Sales by Month……………………………………………………………………25

Figure 18 Sales by Region…………………………………………………………………...26

Figure 19 Sales by Publisher…………………………………………………………………26

Figure 20 Sales by Platform………………………………………………………………….27

Figure 21 Sales by Year……………………………………………………………………...27

Figure 22 Highest sales Globally…………………………………………………………….28

Figure 23 Games released each Year………………………………………………………...28

Figure 24 Sales by Genre…………………………………………………………………….29

**List of Table**

Table 1.5.6 Roles & Responsibilities………………………………………………………..12

Table 1.6 Scheduling….……………………………………………………………………..12

Figure 3.2 Database Design………………………………………………………………….18

### Abbreviations

SVIT: Sardar Vallabhbhai Patel Institute of Technology

**Table of Contents**

Declaration………...…………………………………………………………………………..3

Abstract……………….…………………………………………………………………….....4

List of Figures...…….………………………………………………………………………....5

List of Tables…………….………………………………………………………………….....6

Abbreviations...……………...………………………………………………………………...7

Introduction………………….………………………………………………………………...9

System Analysis………………….…………………………………………………………..14

System Design…………………..……………………………………………………………18

Conclusion……………………...…………………………………………………………….30

References………………………..…………………………………………………………..32

Completion Certificate.…………..…………………………………………………………..33

1. **INTRODUCTION**
   1. **SUMMARY**

This project will give you feel of how data analysis projects are executed in big companies. This project is based on a computer hardware business which is facing challenges in dynamically changing market. Here, sales director decides to invest in data analysis and would like to build PowerBI/Tableau Dashboard that can give you real time sales insights.Data can help businesses to better understand their customers, improve their advertising campaigns, personalize their content and improvise their bottom lines. Raw data has a lot of potential, but you need data analysis tools and libraries to unlock the power to grow your business. Data Analysis is important in business to understand the problems faced by an organization and to explore data in meaningful way. Data Analysis organizes, interprets structures and present the data into useful information that provides content for the data.

* 1. **PURPOSE**

The main purpose of data analysis is to find meaning in data so that desired knowledge can be used to make informed decisions. Sales Insights is designed to support sales managers in improving the performance of their sales team activities. They are what helps to paint a clear picture of the world prospects live in, and helps sales leaders make clearer decisions about how and when to sell. Sales Insights allow sales leaders to guide their team towards more personalized and targeted outreach.

* 1. **OBJECTIVE**

The main objectives of this project are listed below:

* To create dashboard for providing better insights.
* To get important features to increase sales in future.
* To predict popularity of games released in past years.
  1. **TECHNOLOGY AND LITERATURE**

1. Technology

* Power Bi
* Tableau
* DAX Queries
* SQL
* Python and its libraries
* Excel

1. Literature

* **Predicting Global Video-Game Sales** – Alice Yufa, Jonathan L. Yu, Henry Chan, Paul D. Berger.
* **Sales Predictions on Video Games Using Machine Learning** – Bodduru Keerthana, Dr. K. Venkata Rao.
  1. **PLANNING**
     1. Approach

As per the requirements, I decided to develop my project with the Extreme Programming (XP) Approach. Extreme Programming is an agile development framework that aims to produce higher quality products, and higher quality of life for the development team. Also, it was best suited for this project as this rojects has following requirements:

* Dynamically changing requirements
* Risks
* Constant need of feedback
  + 1. Effort, Time and Cost Estimation
* Time Estimation

Predicted time: 2160 Hours

Required time: 2112 Hours

Time required to build dashboard: 264 Hours

Time required to build Model – 1: 216 Hours

Time required to build Model – 2: 264 Hours

Time for testing and taking feedbacks: 144 Hours

* Cost Estimation

1. Rate of Power Bi Developer: Rs 460 per hour

Total rate = 460 Rs x 264 Hours

Total rate = Rs 1,21,440

1. Rate of Data Analyst: Rs 480

Total rate = 480 Rs x 480 Hours

Total rate = Rs 2,30,400

1. Other rate (testing and etc.): Rs 900 per hour

Total rate = Rs 900 x 144 Hours

Total rate = Rs 1,29,600

Total Cost = Rs 1,21,440 + Rs 2,30,400 + Rs 1,29,600

**Total Cost = Rs 4,81,440**

* + 1. Roles and Responsibilities

|  |  |
| --- | --- |
| **Role** | **Responsibility** |
| Power BI Developer | Get data, Building Relationships, Creating Visualizations and Reports, Formatting/Designing Reports, Publishing Reports |
| Data Analyst | Gathering Data, studying data, Cleaning and Preprocessing data, EDA, preparing visualizations, Doing analysis, suggestions and conclusions, model building |

Table 1.5.3 Roles & Responsibilities

* 1. **SCHEDULING**

|  |  |  |
| --- | --- | --- |
| **Topic** | **Starting Date** | **Ending Date** |
| Project Start | 10-01-2022 | 10-01-2022 |
| Getting/Searching Data | 10-01-2022 | 20-01-2022 |
| Change in dataset | 23-01-2022 | 07-02-2022 |
| Power BI - tutorials | 10-02-2022 | 03-03-2022 |
| Building Report/visualization/insights | 04-03-2022 | 11-03-2022 |
| Testing-1 | 12-03-2022 | 12-03-2022 |
| Feedback-1 | 13-03-2022 | 13-03-2022 |
| Building Model-1 | 14-03-2022 | 22-03-2022 |
| Testing-2 | 23-03-2022 | 23-03-2022 |
| Feedback-2 | 24-03-2022 | 24-03-2022 |
| Building Model-2 | 25-03-2022 | 05-04-2022 |
| Testing-3 | 06-04-2022 | 06-04-2022 |
| Final Changes | 07-04-2022 | 07-04-2022 |
| Project Complete | 08-04-2022 | 08-04-2022 |

Table 1.6 Scheduling

* + 1. **Gantt Chart**

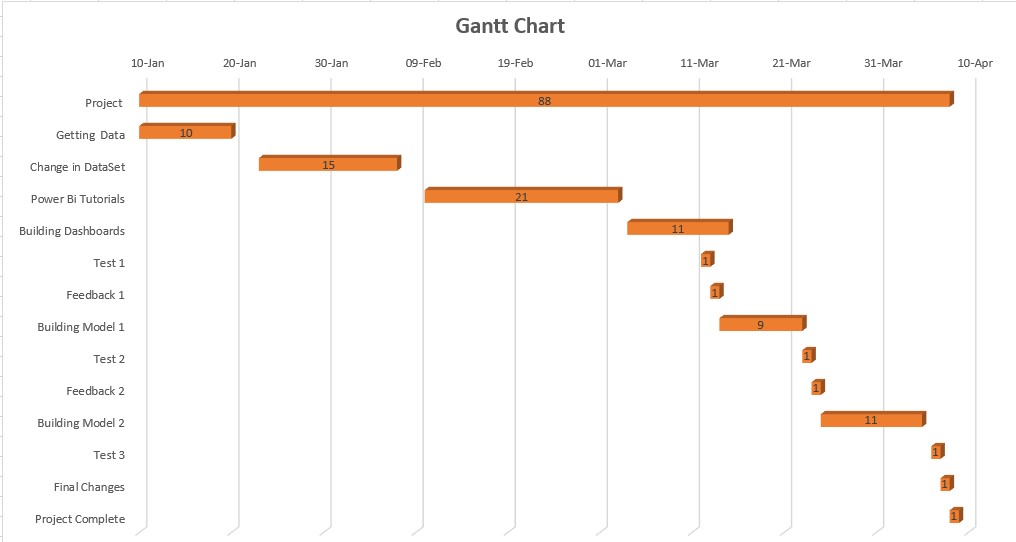
****

Figure 1 Gantt Chart

* + 1. **Timeline Chart**

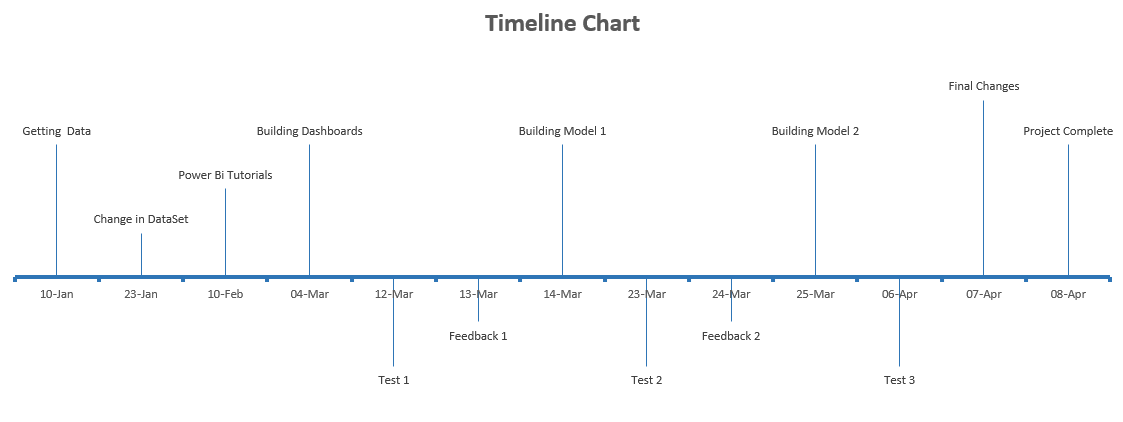
****

Figure 2 Timeline Chart

1. **SYSTEM ANALYSIS**
   1. **STUDY OF CURRENT SYSTEM**

Currently excel belong to Microsoft and are used for data processing and analysis. Excel, which is part of Office 365 and was released in 1985, is a program known by virtually everyone. Whether or not you are an expert in data analysis and whether or not you work with data, you have almost certainly used Microsoft Excel at some point. The program has the format of a spreadsheet that distributes data into rows and columns and has the ability to perform calculations and mathematical formulas easily and quickly. In addition, like Power BI, it is also used to transform data sets into visual information.

* 1. **PROBLEM AND WAEKNESS OF CURRENT SYSTEM**

Excel makes collaborative work difficult. Working together and simultaneously with other teammates is quite complicated in excel. If the file you are working on is not in the cloud, the only option is to share the file, but it does not allow simultaneous work. Limited amount of data. Excel is not designed to work with big data and has a more adjusted amount of data that will depend on the version we have of the program.

* 1. **REQUIREMENTS OF NEW SYSTEM**

The following list provides the minimum requirements to run Power BI Desktop:

* Windows 8.1 / Windows Server 2012 R2, or later
* .NET 4.6.2 or later
* Internet Explorer 11 or later
* Memory (RAM): At least 2 GB available, 4 GB or more recommended.
* Display: At least 1440x900 or 1600x900 (16:9) required. Lower resolutions such as 1024x768 or 1280x800 aren't supported, as certain controls (such as closing the startup screen) display beyond those resolutions.
* Windows display settings: If you set your display settings to change the size of text, apps, and other items to more than 100%, you may not be able to see certain dialogs that you must interact with to continue using Power BI Desktop. If you encounter this issue, check your display settings in Windows by going to Settings > System > Display, and use the slider to return display settings to 100%.
* CPU: 1 gigahertz (GHz) 64-bit (x64) processor or better recommended.
* WebView2, if not automatically installed with Power BI Desktop or uninstalled.
  1. **SYSTEM FEASIBILITY**
* Do I need to do thorough and comprehensive data analysis? If so, Power BI is the best option.
* Do I need to create reports in tabular format? In this case you should use Excel.
* Do I need my team to work collaboratively? Choose Power BI.
* Do I work with large amounts of data? Power BI will be more useful to you.
* Do I find it more necessary to create complex or visually attractive and interactive visualizations and graphics? In the first case, choose Excel. On the contrary, Power BI is more suitable.
* Do I work in Business Intelligence? Then Power BI is the ideal application for you.
  1. **FEATURES OF NEW SYSTEM**

To overcome the limitations of excel i decided to use power bi it has following advantages over excel:

* Superior connectivity
* Collaborative work
* Automatic update
* Designed for the big data
* Ideal for dashboard and KPIs
* Alerts
* Better visuals
  1. **TECHNIQUES of NEW SYSTEM**
* Power Bi
* Tableau
* Python
* SQL
  1. **SELECTION OF SOFTWARE AND ALGORITHMS**

1. Software

* Power Bi
* Tableau
* Google Collab
* Python

1. Algorithms

* Decision Tree
* Ridge
* Linear Regression
* Random Forest Regression
* KNN
* Random Forest Classifier
* Logistic Regression
* Lasso

1. Python Libraries

* Pandas
* Numpy
* Matplotlib
* Seaborn
* Calendar
* Datetime
* Series
* Ceil
* Sklearn
* Train Test split
* Grid Search
* Standard Scaler

1. **SYSTEM DESIGN**
   1. **SYSTEM DESIGN AND METHODOLOGY**

This project has below 3 phases:

1. Phase 1- Dashboard

This phase of the project includes the Power Bi dashboards. These dashboards will provide better insights to take better decisions related to the sales. This phase is the only phase which will be having user/ client interactions but that too with limitations i.e., on click data show only, no changes can be made by client.

1. Phase 2 – Model-1 (Predicting Popularity of Games released in 2016)

This phase will show two types of outputs in table format, first, the games that has become popular in 2016 and second, games that were flop in the year 2016.

1. Phase 3 – Model-3 (Stating important features for better sales of games in future)

This phase is the last phase and will provide the features that are important from sales point of view to the sales director of the firm.

* 1. **DATABASE DESIGN**

|  |  |
| --- | --- |
| **Feature** | **Data Type** |
| Name | object |
| Genre | object |
| Platform | object |
| Publisher | object |
| Developer | object |
| Vgchartz\_score | float 64 |
| Critic\_score | float 64 |
| User\_score | float 64 |
| Total\_shipped | float 64 |
| Total\_sales | float 64 |
| NA\_sales | float 64 |
| PAL\_sales | float 64 |
| JP\_sales | float 64 |
| Other\_sales | float 64 |

Table 3.2 Database Design

* 1. **INPUT/OUTPUT AND INTERFACE DESIGN**

Screenshots:

1. Dashboard

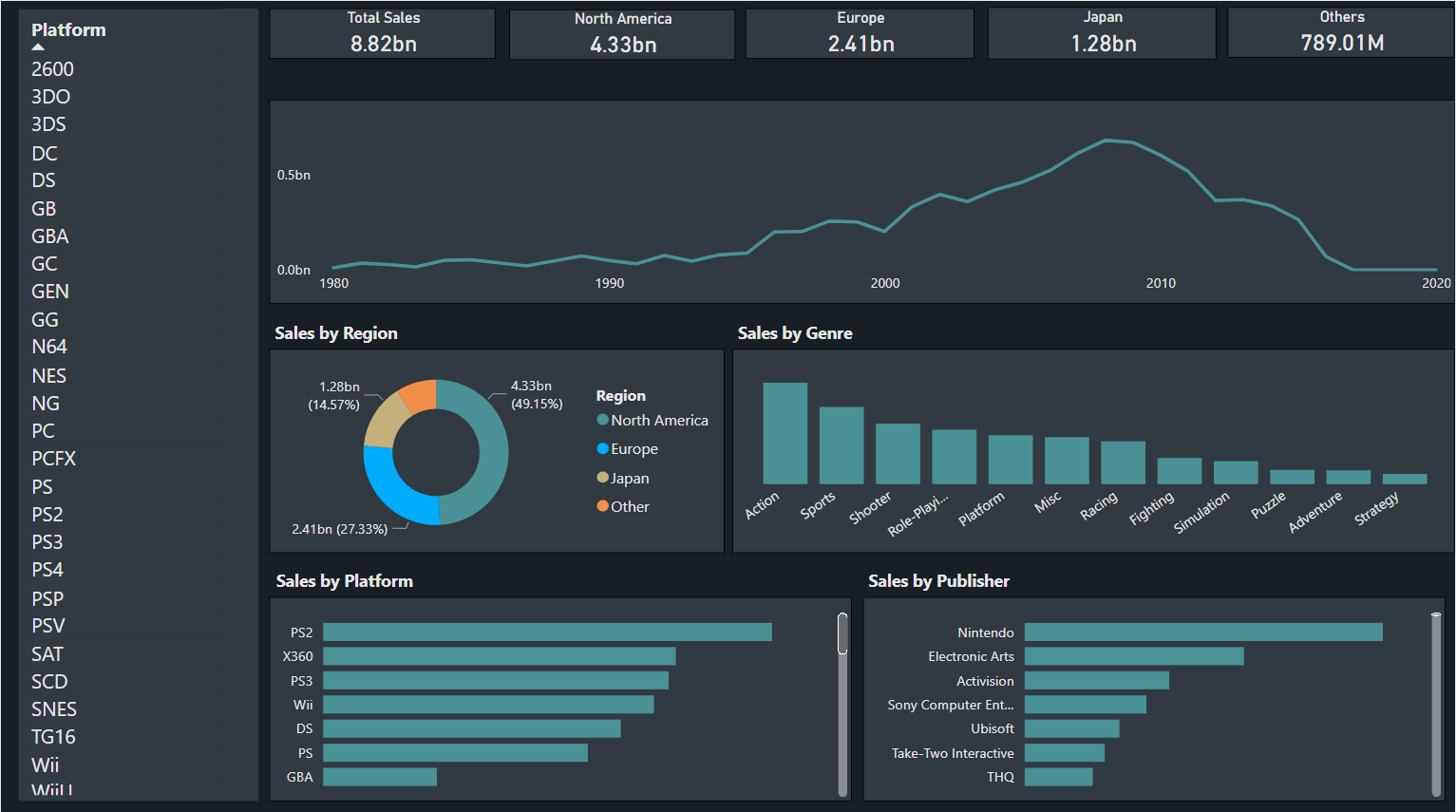


Figure 3 Dashboard

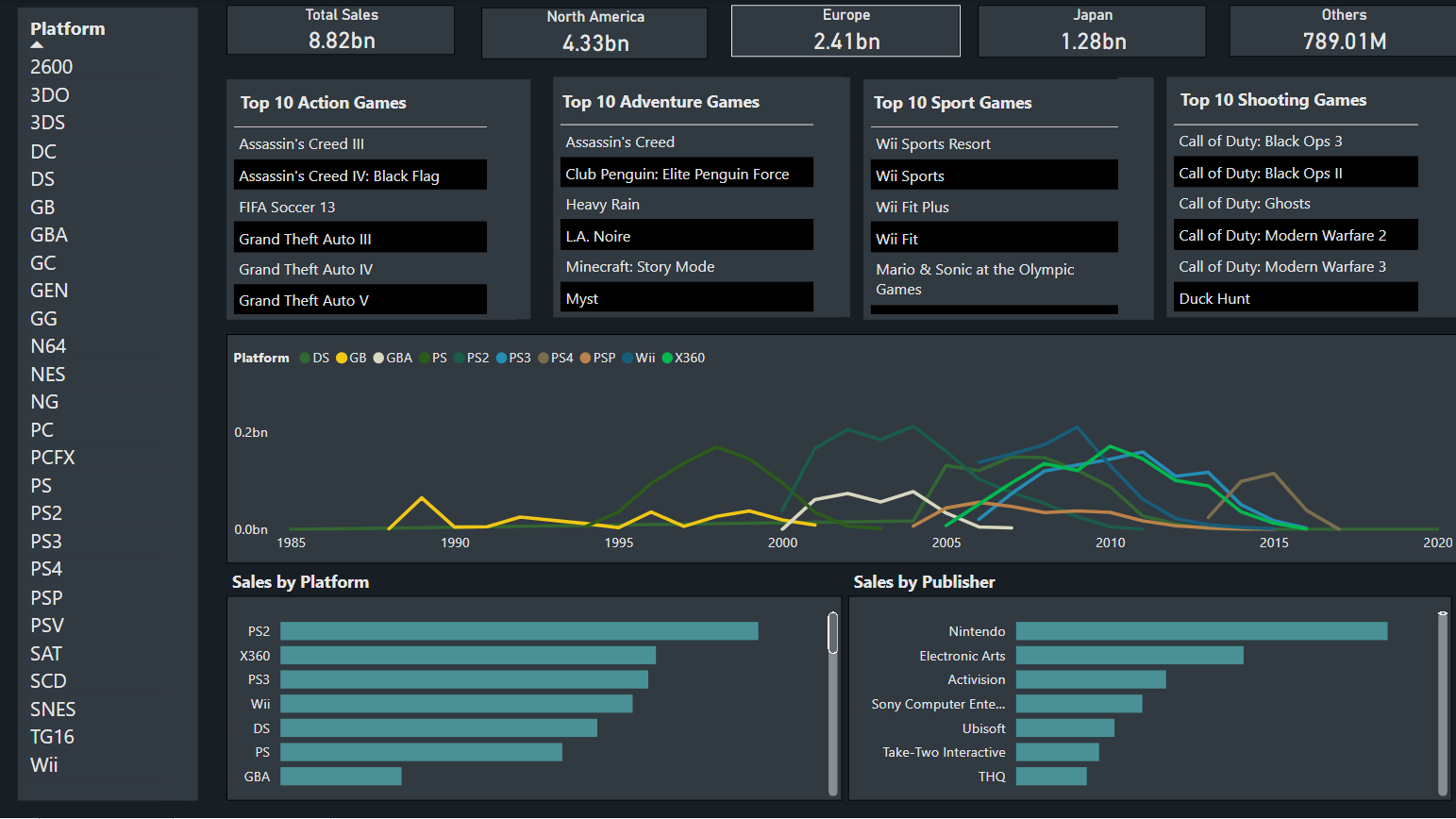


Figure 4 Detailed Analysis - 1

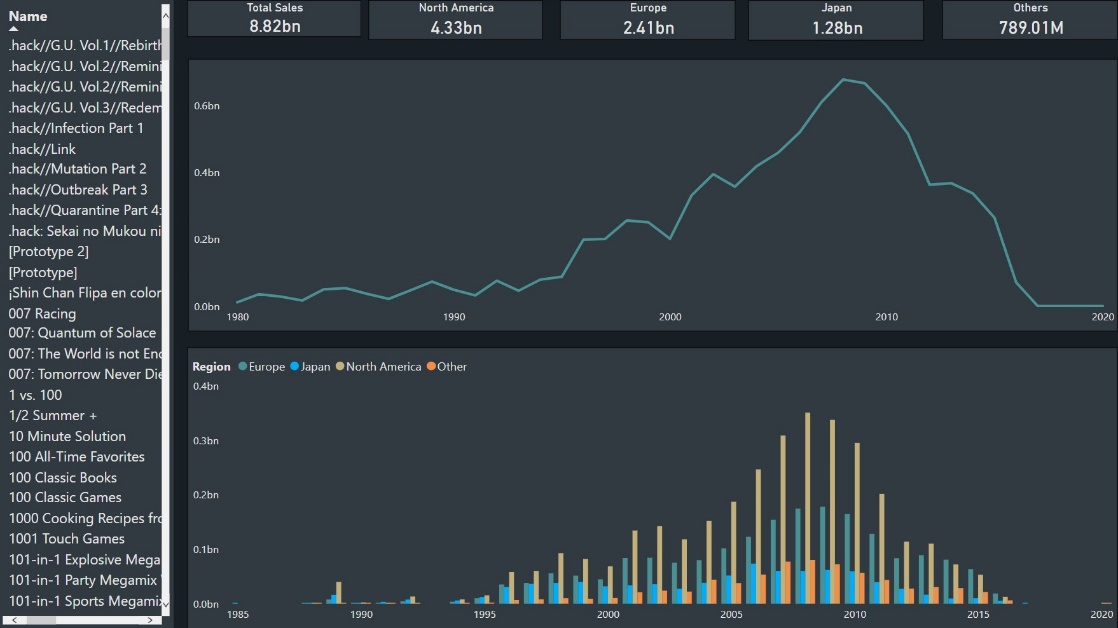


Figure 5 Detailed Analysis - 2

1. Model -1

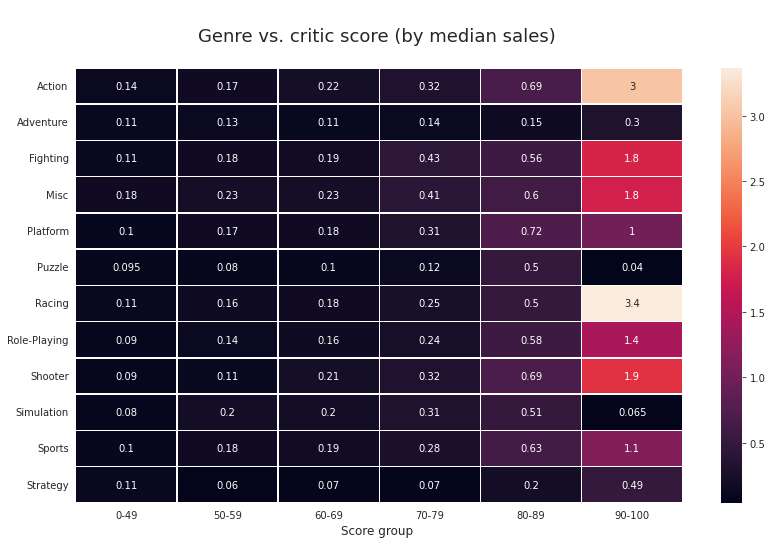


Figure 6 Genre vs Critic Score

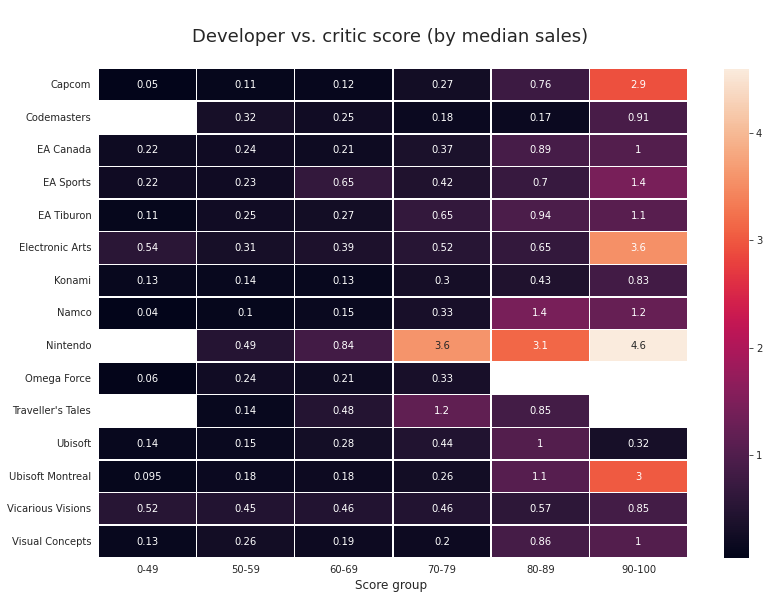


Figure 7 Developer vs Critic Score

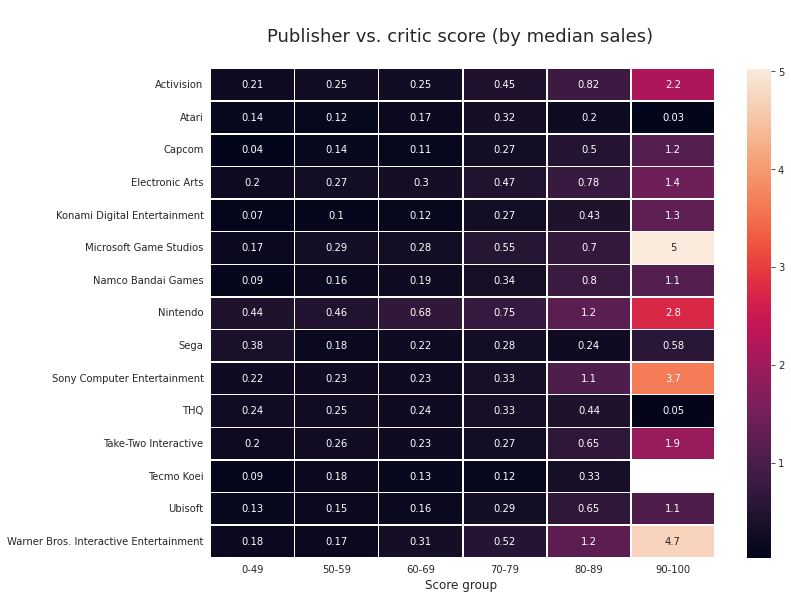


Figure 8 Publisher vs Critic Score

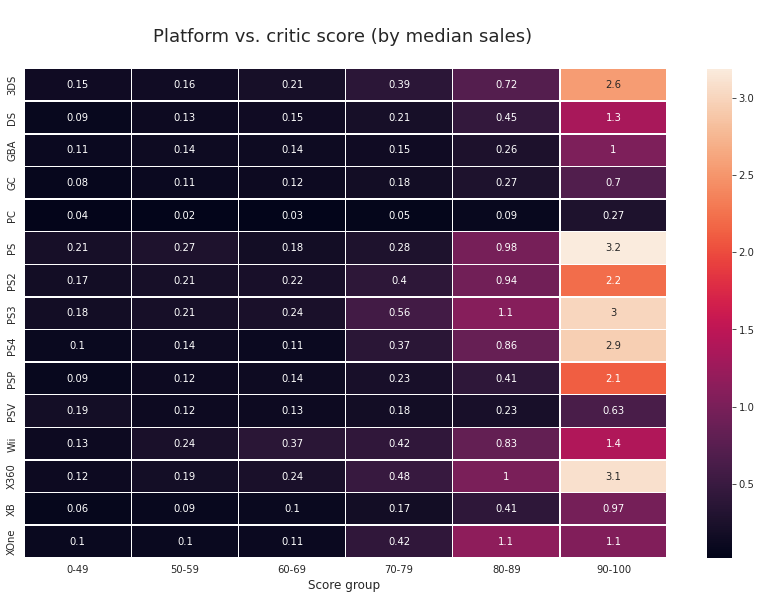


Figure 9 Platform vs Critic Score

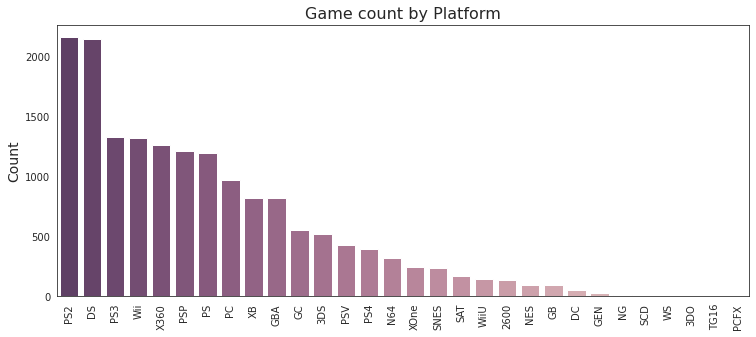


Figure 10 Game count by Platform

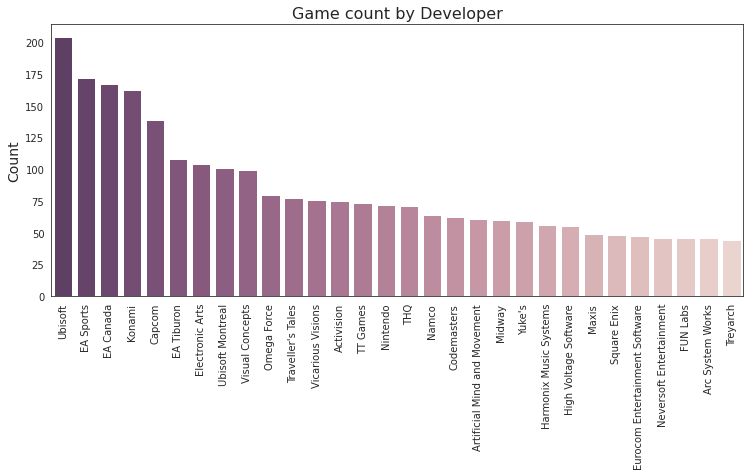


Figure 11 Game count by Developer

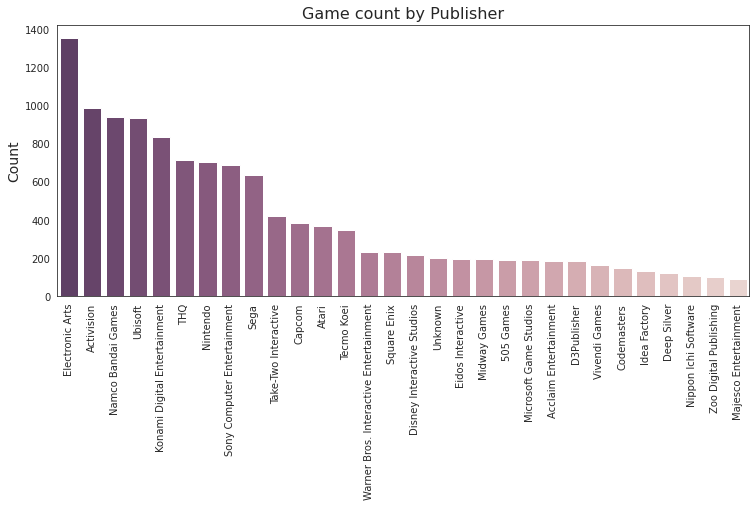
****

Figure 12 Game count by Publisher

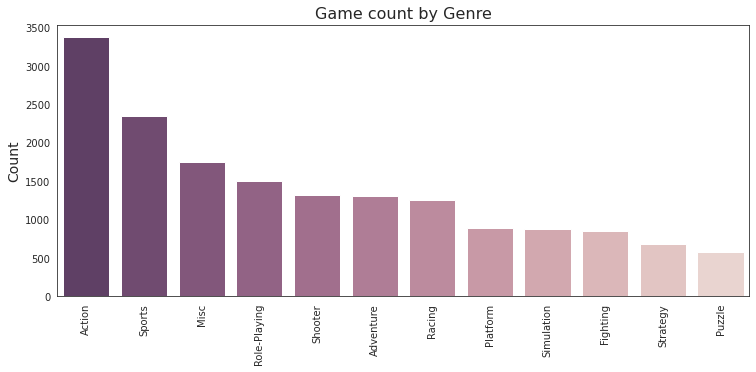
****

Figure 13 Game count by Genre

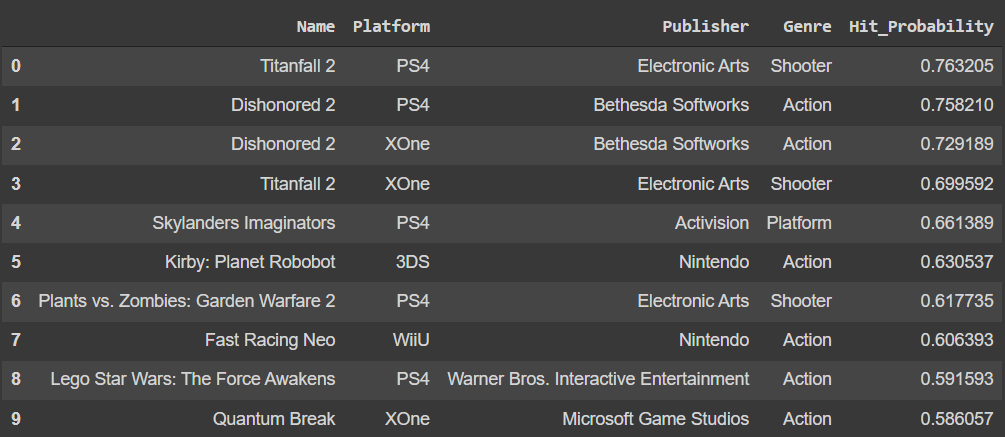


Figure 14 HIts



Figure 15 Flop

1. Model -2

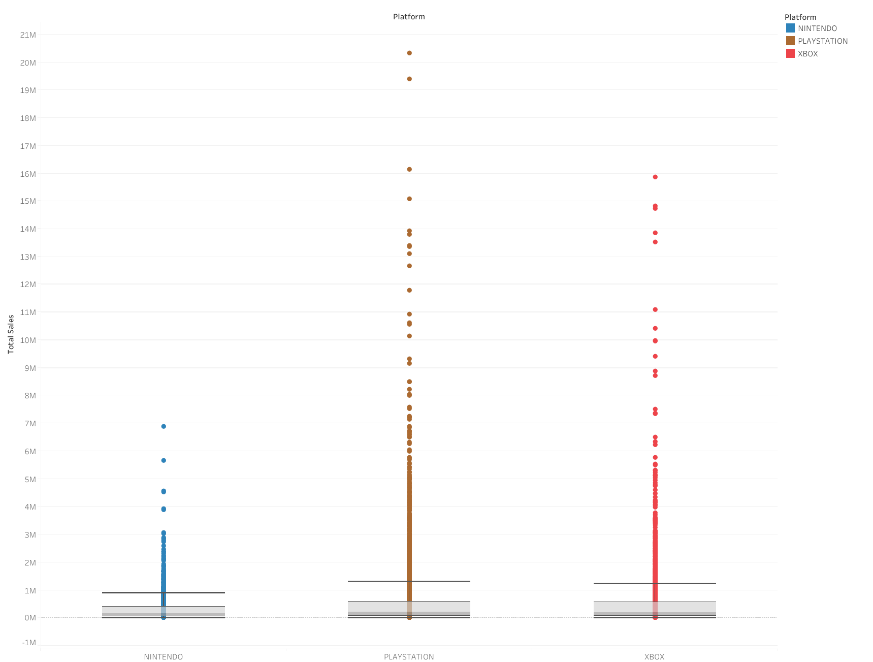


Figure 16 Outliers

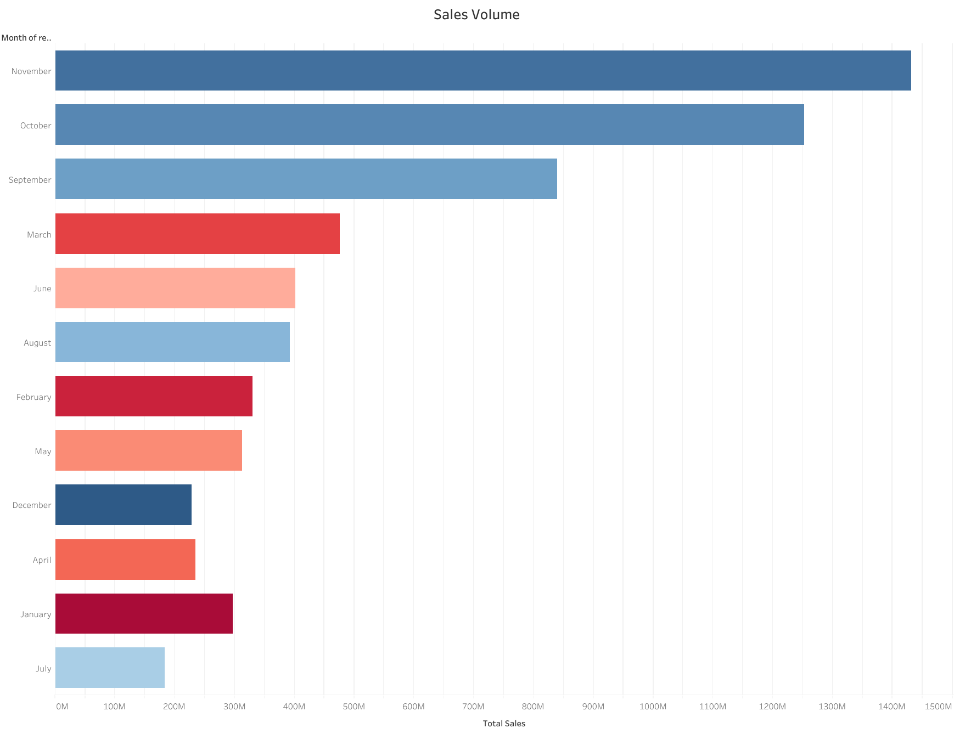


Figure 17 Sales by Month

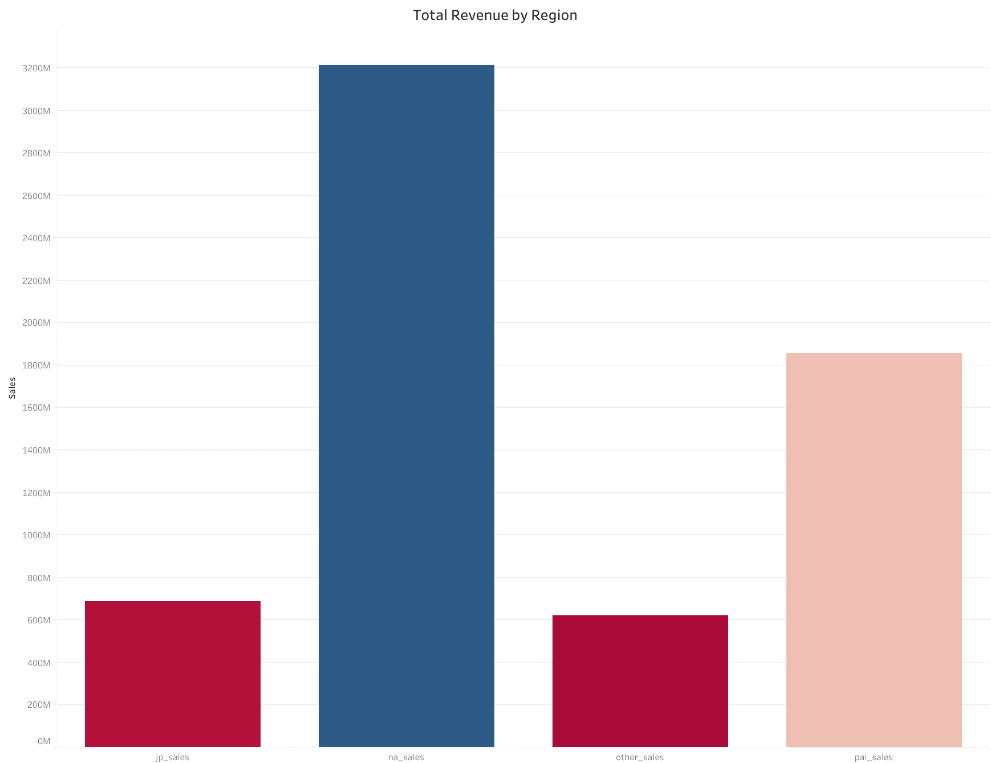


Figure 18 Sales by Region

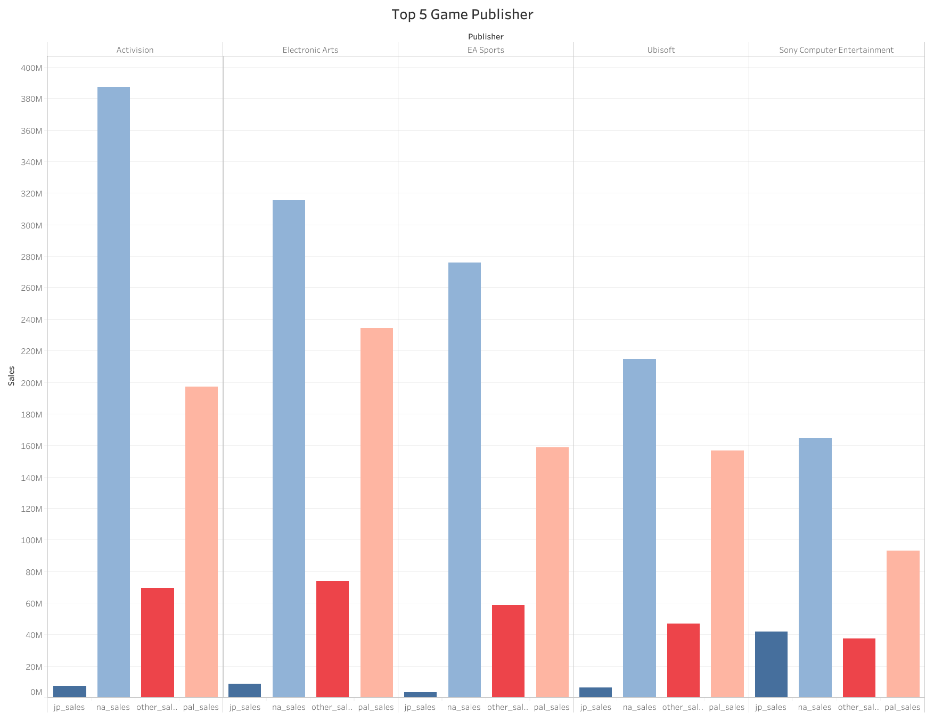


Figure 19 Sales by Publisher

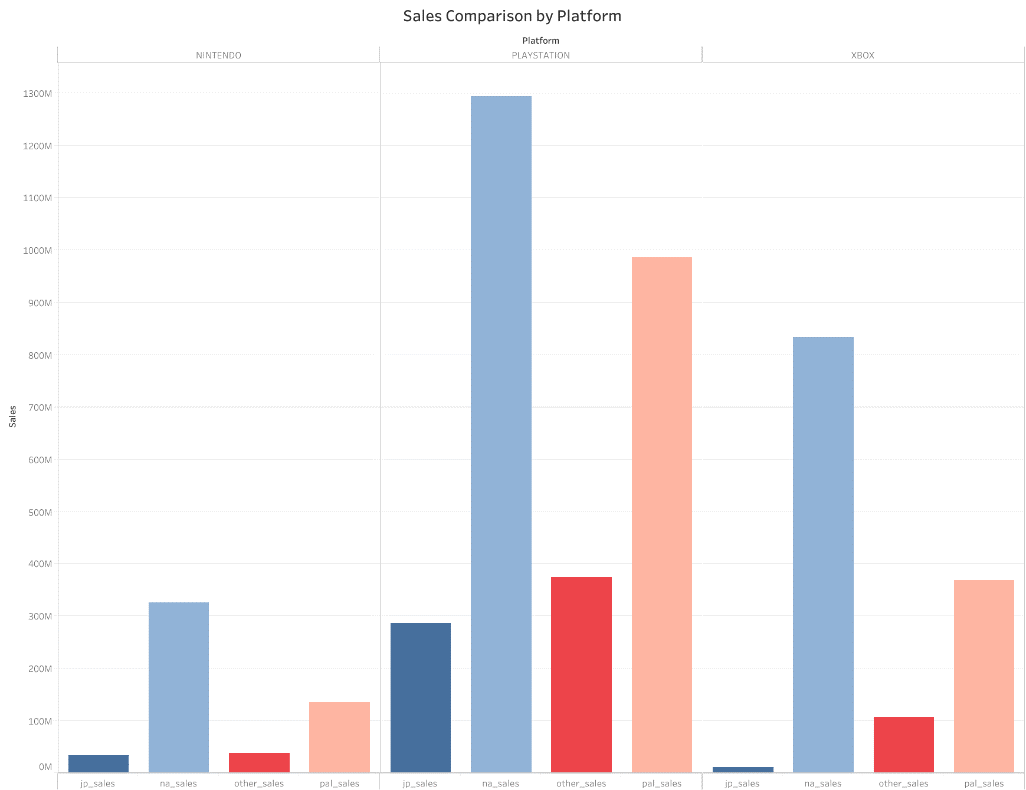


Figure 20 Sales by Platform

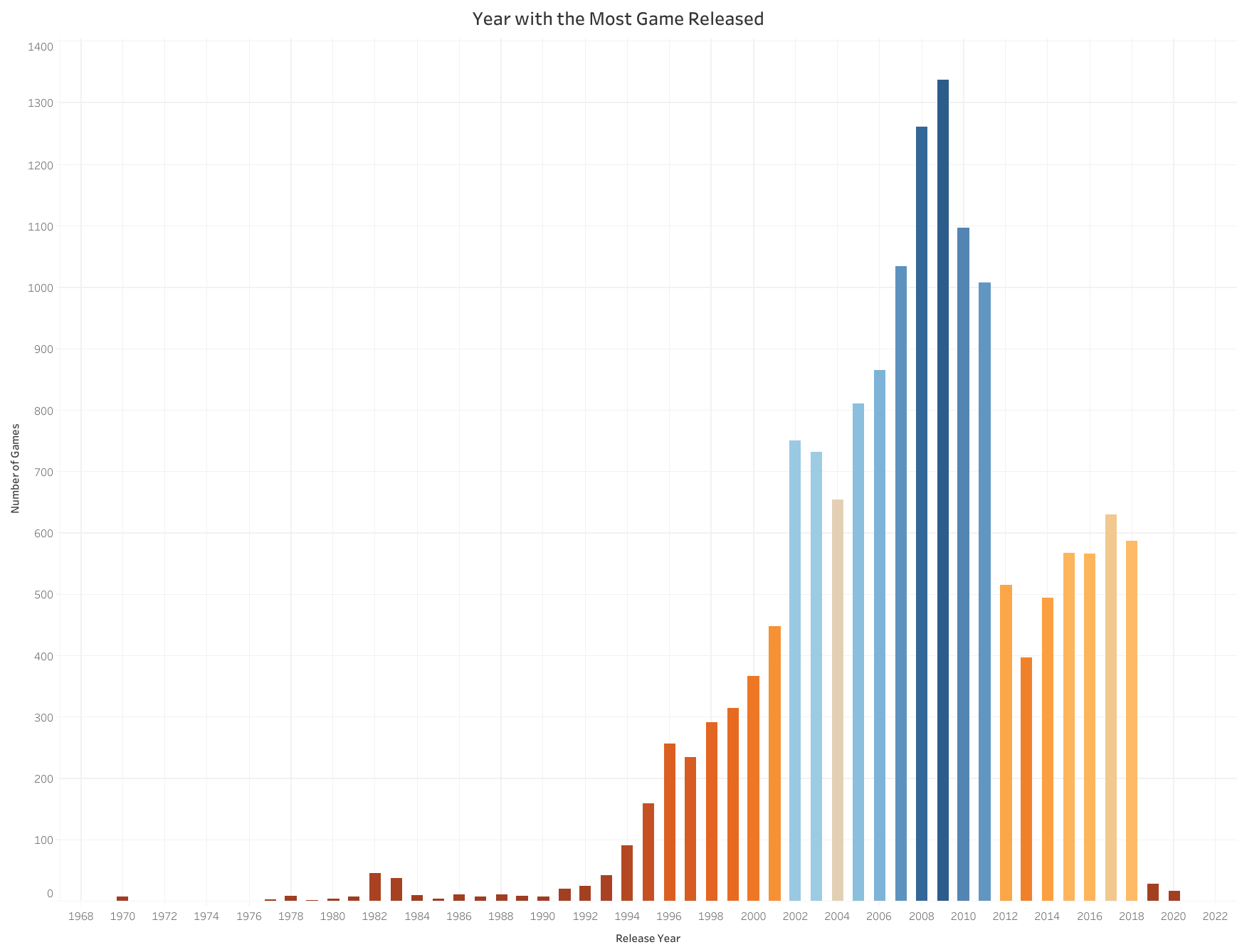


Figure 21 Sales by Year

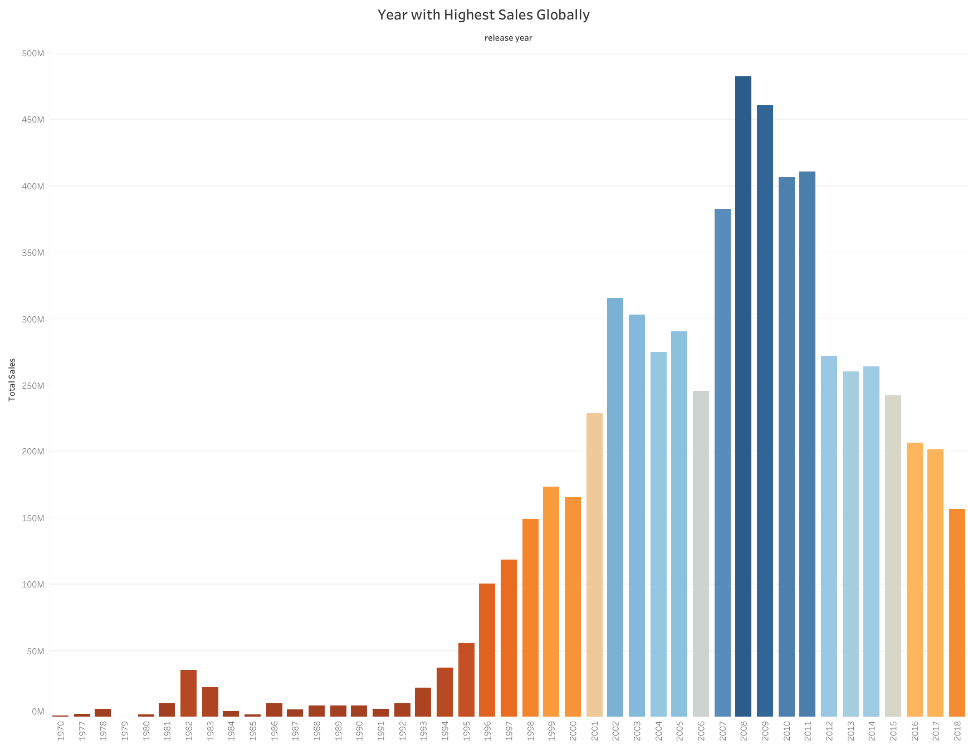


Figure 22 Highest sales Globally

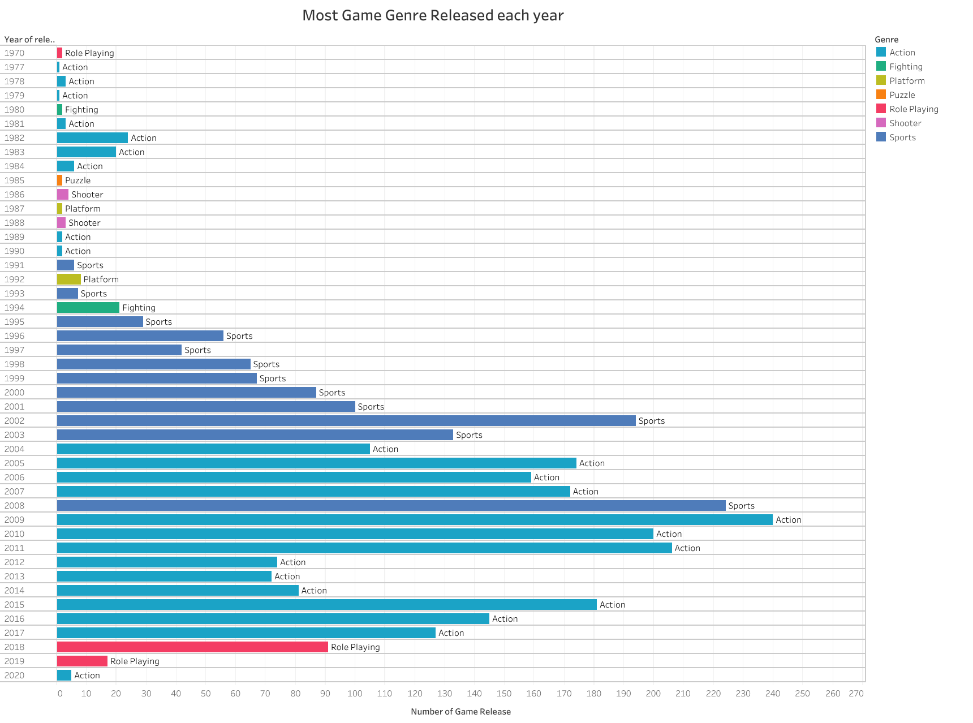


Figure 23 Games released each year

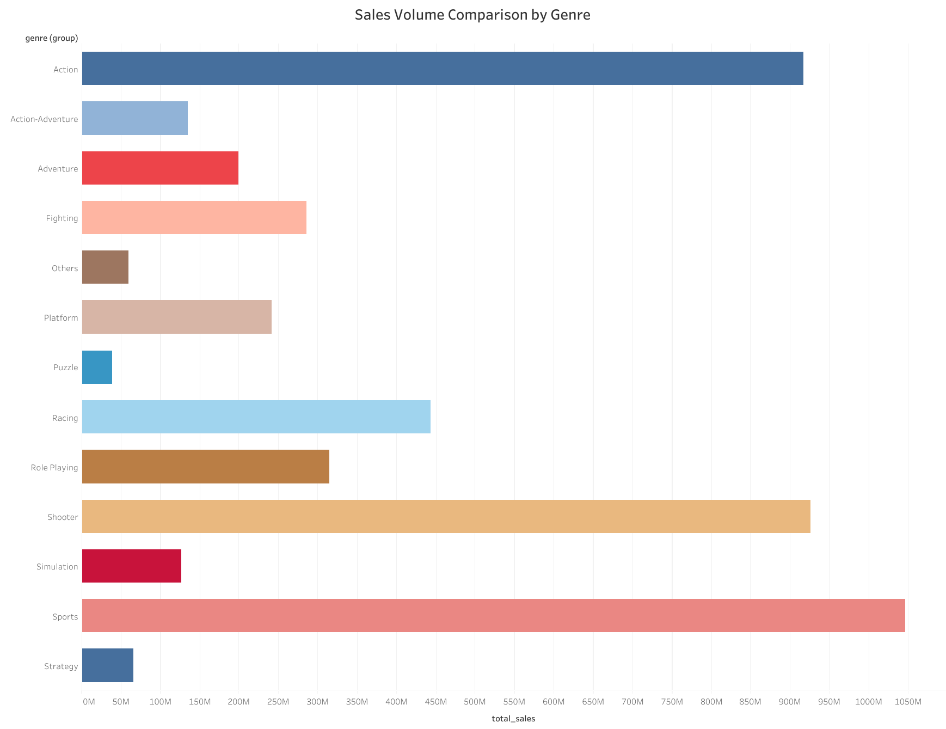


Figure 24 Sales by Genre

1. **CONCLUSIONS**

The important features that affect overall total sales are:

* **Critic score**
* **Release\_of\_month (November)**
* **Platform (PlayStation)**

Through this model, we are able to predict the total sales with RMSE score of 660,000, what this means is that our model's prediction is on average off by $660,000. Given the better performance of the model, it is clear that we achieved the objective of our project successfully.

The ultimate outcome will be, with the details of the total sales prediction for the sales volume, I will use the data to convince stakeholders of the company to set an amount of budget for any new game development, focus on building games based on the PlayStation and also allocating more marketing strategies and ensuring that the critics give a good score for our games.

Also, the sales director will be able to figure out the past insights easily like which games were popular and which were flop.

* 1. **RECOMMENDATIONS**

Additional outside research found that video game downloadable contents and microtransactions to drive $4 billion in sales alone. Therefore, much consideration should be made to increase content with new updates and additional features in the game. With the genre of Action and Sports games being popular, perhaps we can brainstorm and focus on developing games that are based on those genres which will be of most interest to most gamers now. One of the ideas could be to create a game whereby it has to deal with stock market trading, business creation be it legal or illegal sort of business and enable players get to buy luxurious ranging from cars to even purchasing a soccer team which could be a linkup business opportunity with other sports games. Additional consideration for future game development is that we can create a game which players get to play on their main tv game console and at the same time have an mobile game app which is linked. So, in this way, we are able to capture the upcoming trend of mobile gaming and ride the wave on the potential revenues that possible to be explode in  
the years to come.

1. **REFERENCES**
2. Predicting Global Video-Game Sales – Alice Yufa, Jonathan L. Yu, Henry Chan, Paul D. Berger.
3. Sales Predictions on Video Games Using Machine Learning – Bodduru Keerthana, Dr. K. Venkata Rao.
4. Microsoft - <https://docs.microsoft.com/en-us/dax>
5. Payscale - <https://www.payscale.com>
6. Tutotrialspoint - [Python - Data Science Tutorial (tutorialspoint.com)](https://www.tutorialspoint.com/python_data_science/index.htm)
7. Google Colab - [Google Colab](https://colab.research.google.com/?utm_source=scs-index)
8. Model 1 - <https://colab.research.google.com/drive/1Fh2T_cJYL3ZdApiXbY3dqSf8wAXYfiXN?usp=sharing>
9. Model 2 - <https://colab.research.google.com/drive/1CgyOukD64uUHtkAXECDl794QMs1Kmm9t?usp=sharing>

<https://colab.research.google.com/drive/1pE7dHNNr2E72JurfPAIk-ZMjKb9uABgX?usp=sharing>

1. Power Bi - <https://powerbi.microsoft.com/en-au>

