MARKET RESEARCH REPORT

VIDEO GAME MARKETING RESEARCH PROJECT

GBUS -721

PROF: FATENA EL-MASRI

LAHARI TADEPALLI

Table of Contents: **page Numbers** **Abstract** Error! Bookmark not defined.

GBUS 721: Marketing Research
Final Project – Video Games Market Analysis

ABSTRACT

This Market research paper is an analysis performed framing the concepts of market analytics prior to launch a product. Our company is in the process of launching a video game and we are conducting an inhouse research analysis of the video games market. The primary research data is collected through a Qualtrics survey with around 150 respondents which included questions related to preferences on price, genre, game platform, device type, and other preferences. The secondary dataset of video games is taken from Kaggle which includes historic data of publishers, year of release, genre, demographic sales data with over 16,900 records. The geographical sales data for regions including North America, Europe, Japan, and Rest of the World is also analyzed to understand which region will offer higher return on investment. This data set was found in Kaggle created by Gregory Smith from Metacritic and extends some variables with additional observations by name VGChartz Video Games having around 6900 complete cases that were used in our secondary analysis. We have performed the descriptive and predictive analysis on the primary and secondary data to support our market establishment.

PRIMARY MARKET SURVEY

Our primary analysis includes a survey that was developed on Qualtrics software. Qualtrics helps us to Create the survey from scratch, having all kinds of input options to analyze and retrieve a specific solution for a given question. The methods of mandatory answering, loop and condition-based decisions, multiple and descriptive solutions can be achieved using Qualtrics. We can also retrieve the graphs and patterns with the In-built data analysis mode using Qualtrics based on the feedback of the customers. It also allows us to continuously make optimal changes and improvements for a proper survey.

In our primary analysis survey, we included 15 questions that basically focused on the market strategy and preferences of the customers interested in our game. We have also analyzed the factors that played a dominant role to launch our product with certain specifications and grouped the options possible. The responses recorded were around 150 that had some missing values which were unanswered. The data that we collected was mostly focusing with inputs received from age group 18-24 with 65% of total survey weightage. We also raised some interesting questions like level and mode of play to understand the gamers choices that we could implement on our product. The other factors that were given a range are scale of play time and other issues with agreeing or disagreement also helps us if certain factors which really matter to the customers.

Survey Method:

-> Specifying the information: Initially our aim was to understand the basic components that the gaming customers were interested in and then we devised the questions in a pattern of market segments. The main categorization of the factors specific to the domain of study were adopted in our survey.
->The type of survey method: We used Qualtrics for our survey development in which we included qualitative and range data to analyze the data insights that mainly influence our implementation and decisions of game development. This survey was available online and would be completed in less than 6minutes comprising 15 questions that were relevant to the ideology of our research.

->Determining the content of Individual questions: We made sure that the questions were grouped in sequence and the questions conveyed un-ambiguous meaning helping the customers to complete the survey without any confusion or hesitation. We could finally retrieve a dataset comprising 150 data points and varied outcomes.

->Designing the questions: The questions were well-structured and were designed to elicit recommendations from the respondents. We have also implemented the rank order scale and range for certain questionnaires. The decision and view based questions used semantic differential scale with 5 points making it easy for the responders to choose an outcome.

->Appropriate question flow: The importance was given to the qualitative responses and we tried to keep the sequence and applied multiple options wherever necessary. The survey was framed in a way that is easily understandable by avoiding the terminology used in gaming to certain extent.

->Pretesting before implementation: The pretesting of the survey was performed to observe the outcomes and it was a success as there were no looping or disorder patterns. We made sure to avoid implicit

dilemmas and biases-based questions that most of the respondents would not willingly answer. The final changes and alternatives were made before launching our survey and it was very reliable and accurate.

->Implementing the survey: The survey was implemented, and we received a good response from the customers. We also observed the average time was less than the expected time because the respondents felt it as an easy survey with basic thought processes. Finally, we could collect all the responses from Qualtrics into an excel file and summary statistics. The data we collected was pre-processed and used for analysis.

SECONDARY DATA COLLECTION

The secondary analysis is conducted on a dataset that was downloaded from Kaggle. We had fields like Name, Platform, Year of release, Genre, Publisher, Sales by regions and other criteria. We had 9 columns in total and this dataset helped us in understanding the trends of companies who were already performing in the gaming industry and their statistics. We explored the trends and insights that could help us in designing our output and taking decisions based on the success rate of our competitors. Using the secondary data set we could also analyze the sales of the publishers across the world and year wise analysis. We also considered the factors like critic and user having a certain score and count for a specific game they played. The data we collect can be viewed and downloaded here: (SID_TWR, 2019)

HYPOTHESIS

1.Understanding the Distribution of trends in the observations for Global sales with genre and platform.
2.Which factors Influence the game or product the most? Are critic scores and rating important?
3.The variations of points influence the brand or game strategy in the long run and revenue success.
4.Exploring the popular picks by genre, platform, and publisher sales.
5.Does publisher and rating affect the game's popularity and sales?

DATA EXPLORATION

1. Data Cleaning and Preprocessing:

Primary survey: In our primary survey, we had to clean our data according to the conjoint analysis that fit some of the values for quantitative representation. The data was converted into binary form which is 0 and 1 for independent variables including action, adventure genre, PC & laptop, smartphone platform, online, and offline mode-type. Price range as a dependent variable was divided into four levels from 0 as low (free), 1 as medium (up to $50), 2 as high ($50-$100), and 3 as very high (>$100). We have omitted the missing values and in-complete records from the survey. This data was processed in excel and the importance of each factor was outlined. The other challenge was segmenting the multiple choices as they were mainly having more importance and we made every value categorized.

Secondary Dataset: We have omitted the values that had missing data points by considering around 6900 entries with complete observations. The data set was selected only through a range from 1995

to 2016 based on our assumption of violating certain platforms that are not used or available as on today. We have selected and grouped the factors that we focused on and that would help to answer our hypothesis. We have segregated the year along with the publisher followed by the factors of interest and global sales. We ordered the genre and platform analysis by selecting the top 10 publishers' sales. Likewise, we also made the analysis with regions specific to sales data. The trends with review scores and count of users and critics were also discussed. We fitted K-means classification, principal component analysis and linear regression to obtain the accuracy and acceptance of the results. The secondary analysis was made using R and Python.

2. Primary Data Analysis

2.1. Conjoint Analysis

Conjoint analysis is used for determining the customers preferences on products and services. It is a statistical technique commonly implemented for promoting real world buying decisions by asking respondents the preference for one option for another. The outcome supports in identifying the choices consumers are willing to make depending upon Price thereby indicating the most important features. The below conjoint analysis is calculated based on the primary data collected from Qualtrics. The data was analyzed by using pivot tables. The survey data is collected from over 200 respondents where a video game user selects which level corresponds to their preference for that attribute and then these data points in total are collected and run through a conjoint analysis model.

The model output provides insight into each attribute and each level associated with survey or other data inputs through a value called a part worth. The objective of the part worth is to calculate the approximate average client preference for that attribute in the marketplace. The purpose of this approach is to help our business understand what our consumers are looking for on a video game and provide us with useful insight into how to include important features as per customer preferences and capture some of the market share.

The conjoint analysis provides excellent estimates of the importance of the features, especially about price. Considering Price as dependent variable and all other variables as independent variables the conjoint analysis offers clear understanding about the customers willingness to pay higher for features the most important features. The results can estimate the value of each level and the combinations that make up the optimal product.

Conjoint Analysis Results:

The below conjoint analysis results indicate the ideal profile for genre, platform, and mode type. We can see that after considering the intercepts, action genre, smartphone device platform, and offline mode of playing is given the highest importance by the customer. The results of the model identified the importance of attributes when deciding to purchase a video game. Below are customers' highest rated features. Video Game genre and platform are expected to have high importance of 44.9% and 36.5% respectively.

Part Worths			
Intercept		1.704979095	Ideal Profile
Genre	Adventure	-0.638814984	Action
	Action	0.106879513	
Platform	PC or Laptop	-0.286963132	Smartphone
	Smartphone	0.319156214	
Mode Type	Online	-0.237172178	Offline
	Offline	0.073013007	

Attribute Importance		
Attribute	Range	**Importance**
Genre	0.746	**44.9%**
Platform	0.606	**36.5%**
Mode Type	0.310	**18.7%**

Fig: This Table represents the importance of Genre, Platform, and mode for a video game from the primary survey collected. These factors show high importance.

The below video games Data part worth chart indicates the top features for which customers are willingly to pay high prices. Considering the results of conjoint analysis, our strategy would be to launch a video game based on the below features.

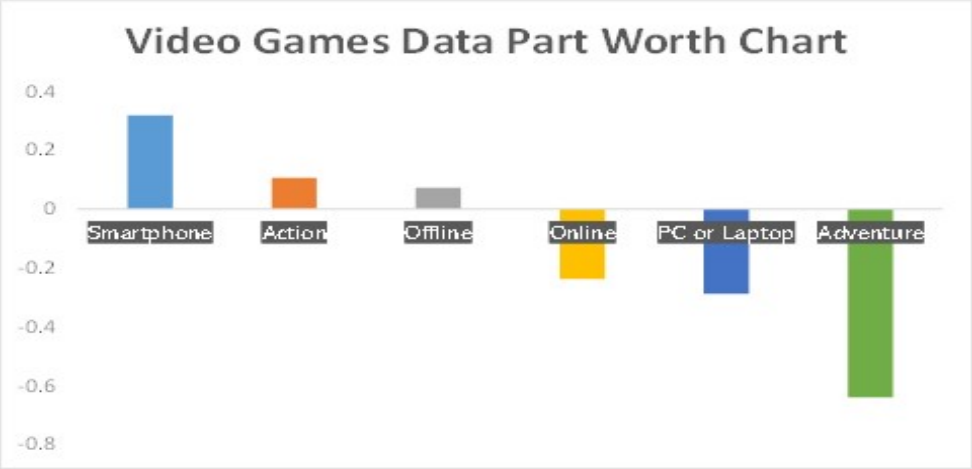

Fig: Part worth chart that depicts the feature contrast.

3. Secondary Data Analysis

3.1 K-Means Clustering:

K-Means Clustering in an unsupervised Non-linear algorithm that clusters data on group similarity by partitioning the observations on basis on the cluster number mentioned. The process includes the formation of 'K' clusters at random points by calculation the centroid value and assigning each data point to a cluster. This process is iterated until all the points find its closest cluster and are grouped accordingly. We performed the K-means clustering in R using the Cluster R and cluster packages. K-means clustering is a prototype based iterative approach on quantitative variables forming K-clusters of compete distance. It

performs closest center cluster basing on the centroid value of the observations. In our process, we observe that the User Count has a proper spread as there are favorites for every game. Whereas the critic score has less distribution as they only focus on top rated games and they are few in count compared to users. Clusters 1 and 2 have around 76% of point variability that means the split is tending towards a specific trend and is mostly explained by 2 clusters. We can observe the formation of 3 clusters with different proportions of shared data points and varied values. The factors of critics and users' count and score helps us in describing nearly 76% of the point variability. So, we can consider all the assumptions made using these factors. (Dhruv, n.d.)

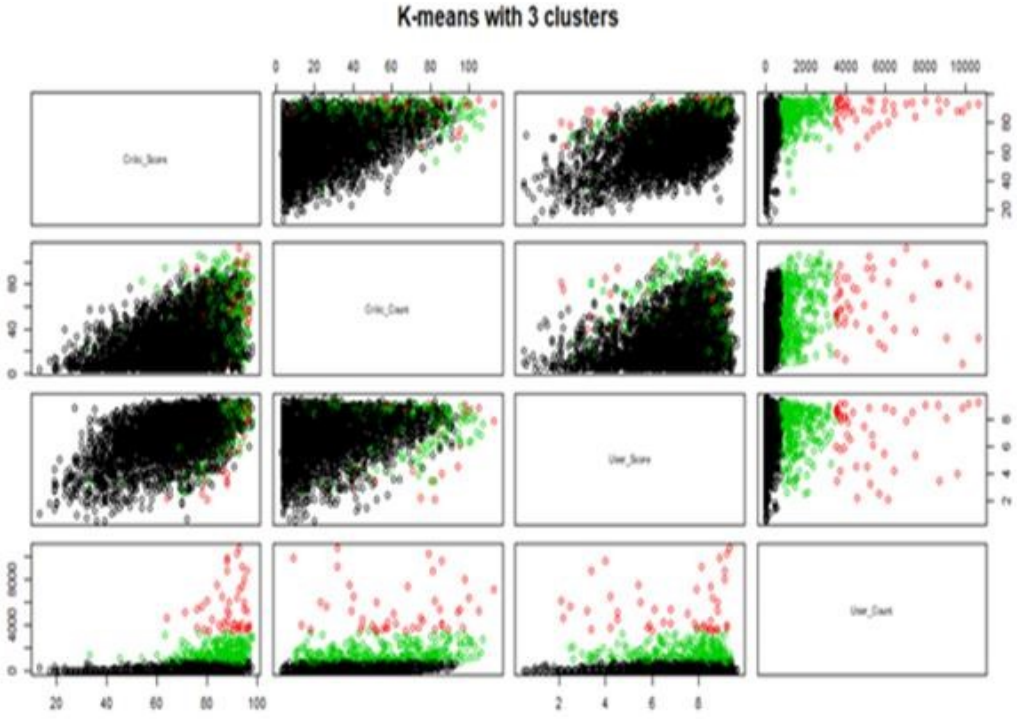

Fig: Formation and spill of 3 clusters for user and critic factors.

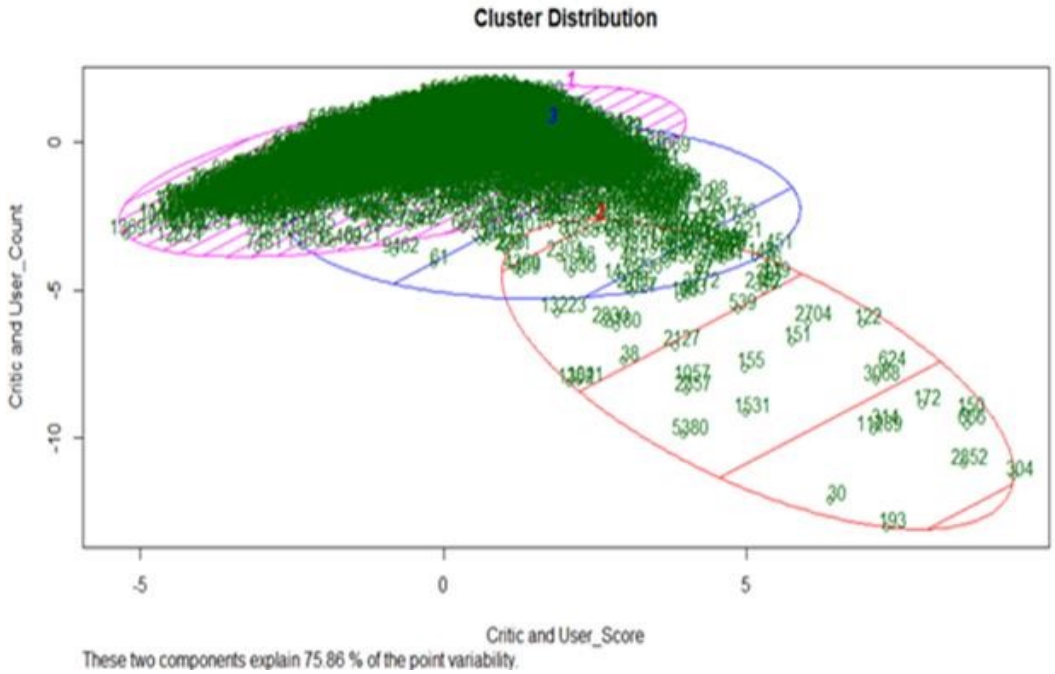

Fig: Explanation of point variability (76%) using score and count variables.

3.2 Principal Component Analysis (PCA):

Principal component Analysis (PCA) is a technique that is used to emphasize variation and develop strong patterns to explore and visualize the vital components in the dataset. In the PCA model, we observe that the critic score values a lot compared to the user score I.e., In most played games are the ones that have a higher critic score which attracts the gamers. The other surprising trend is that the critic count being too less than the user count is considered as an important factor while the sales show low priority for a user.

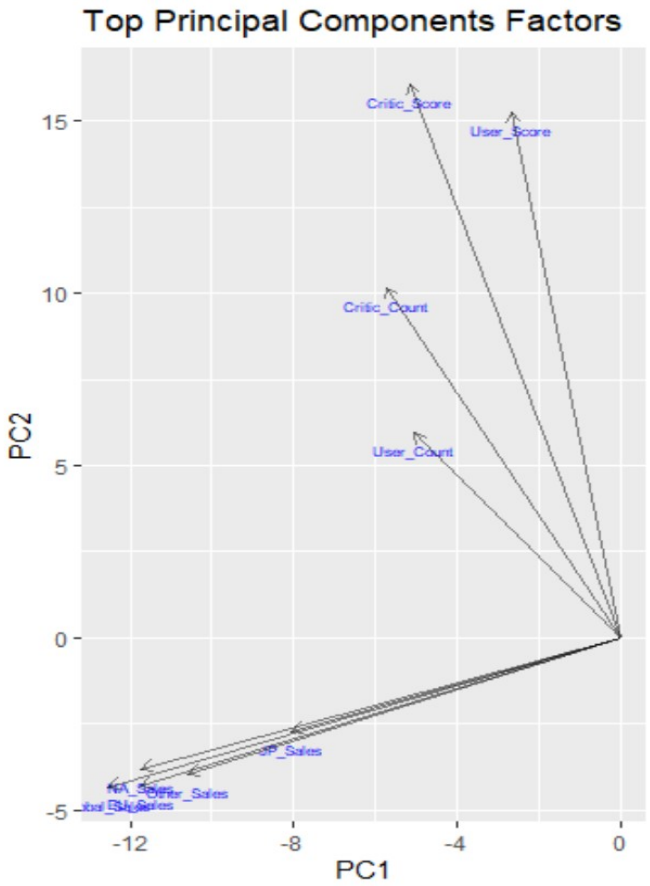

Fig: PCA importance for critic over user and sales.

3.3 Linear Regression:

Linear Regression is a statistical approach for modelling the relationship between a dependent variable with a given set of independent variables. The feature scaling is done by using the Standard scalar ,The data is divided into Training and Test set.

We have obtained a Training score of 0.92 and Test score of 0.86. The scores are very close which indicates that we avoided over fitting of the data.

The line equation is used to predict the dependent variable.

```
lm.fit(X_train, y_train)
lm_score_train = lm.score(X_train, y_train)
print('Training score: ',lm_score_train)
lm_score_test = lm.score(X_test, y_test)
print('Testing score: ',lm_score_test)
print('Equation: ', str(lm.intercept_ ) + ' + ' + str(lm.coef_))
```

```
Training score:  0.92128344225782
Testing score:  0.8640390780704887
Equation:  -0.0005291873063433261 + [1.97631038]
```

This plot indicates the prediction of global sales.

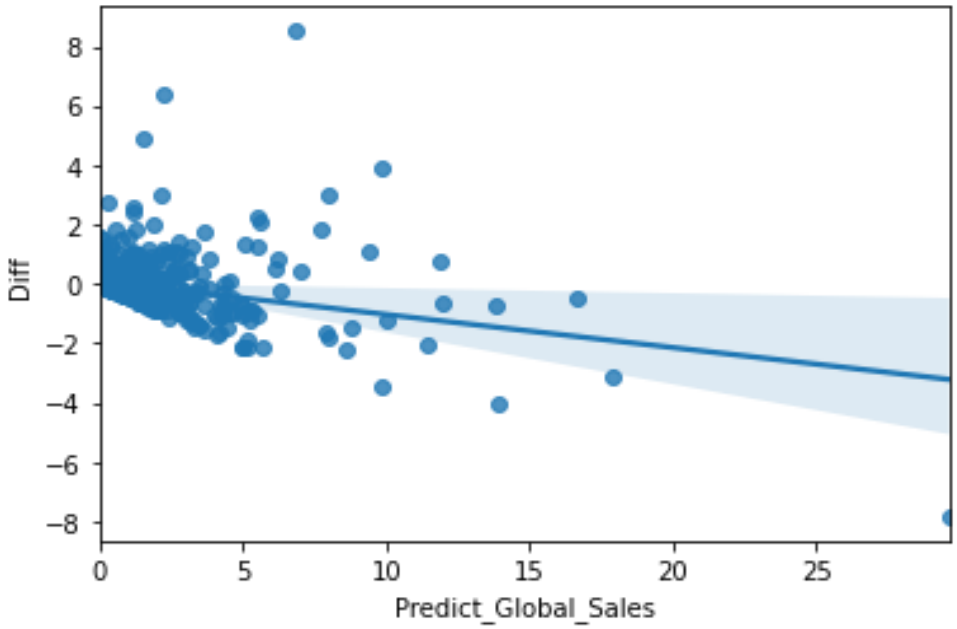

Fig: Observing the linear fit across Global sales and spread.

DATA VISUALIZATION

Figure 1:

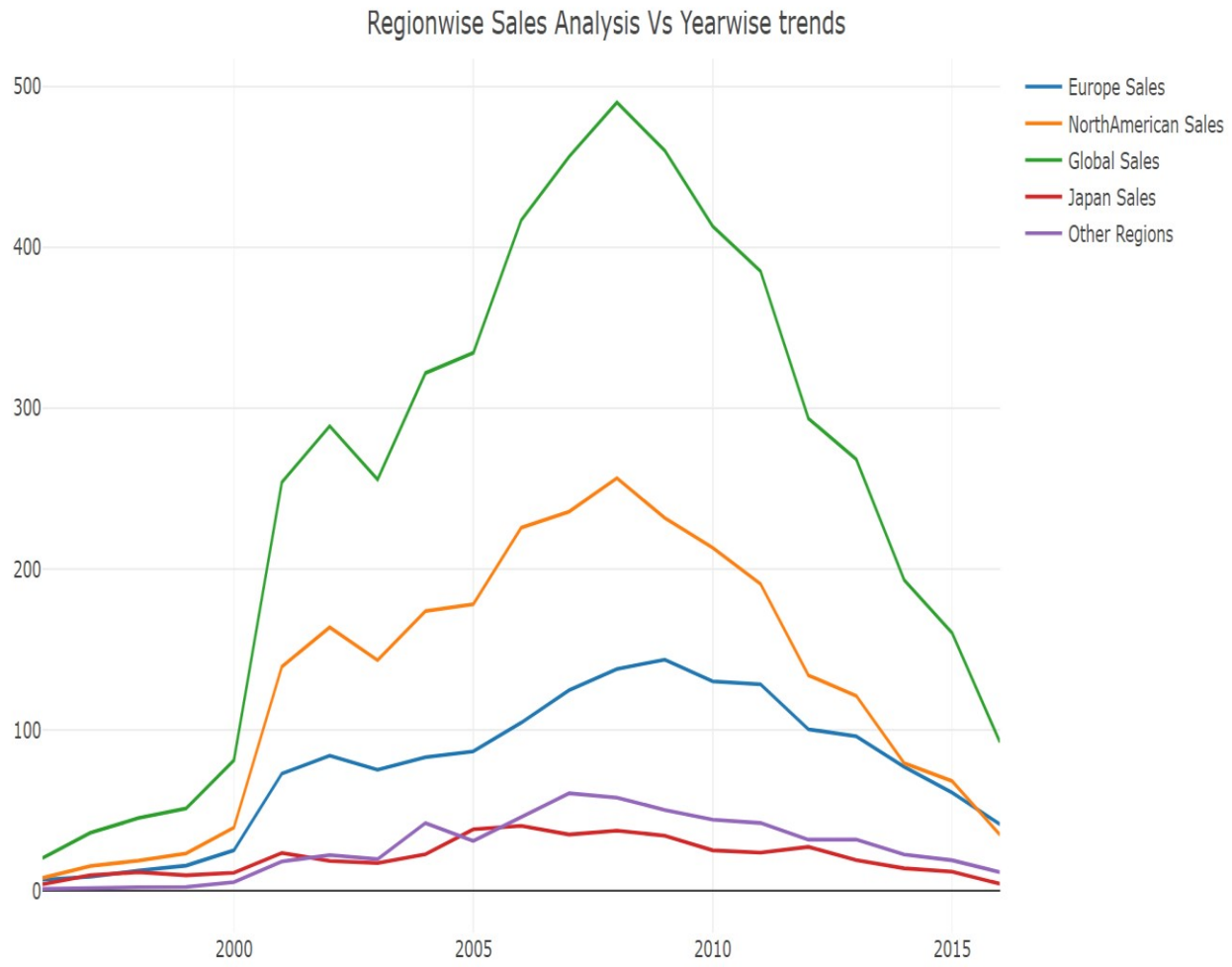

Fig: In this we have considered Region wise sales analysis Vs Year wise Trends. In between the years 2005 and 2010 the Global sales were very high. When we compare the Sales of the countries like Japan, North America ,Europe ,Other countries, the sales of North America are high, and the Japan and Other countries have the same trend in the sales .

Figure 2:

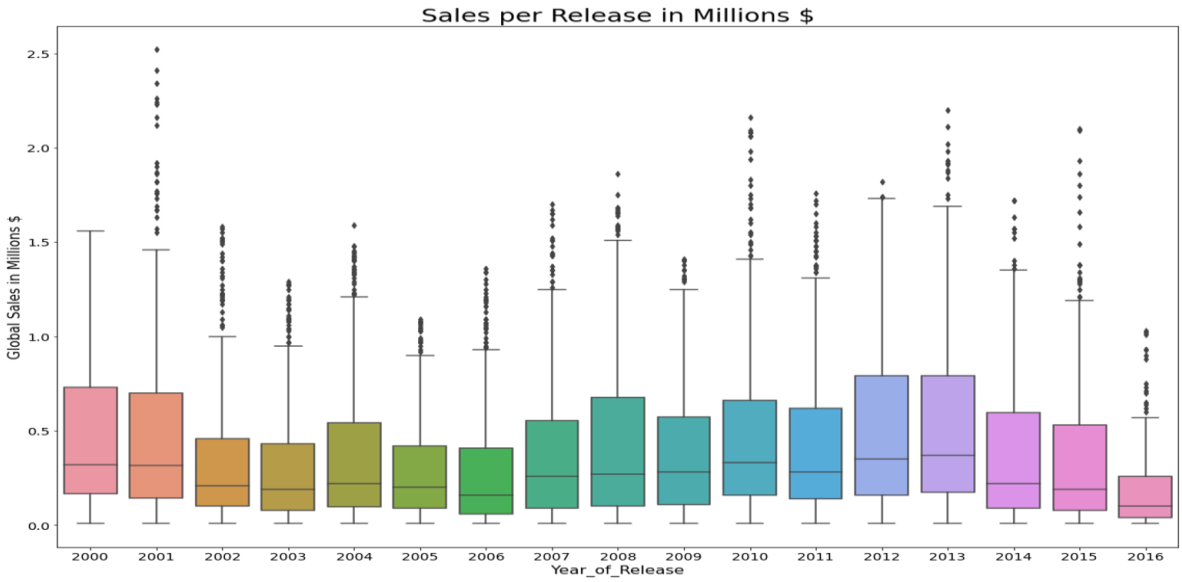

Fig: This Box plot shows the trends of sales in millions for Global sales from 2000 to 2016. The interesting insight that we observed is that there was a peak obtained after 2015 (5 years) while the highest was in 2001.

Figure 3:

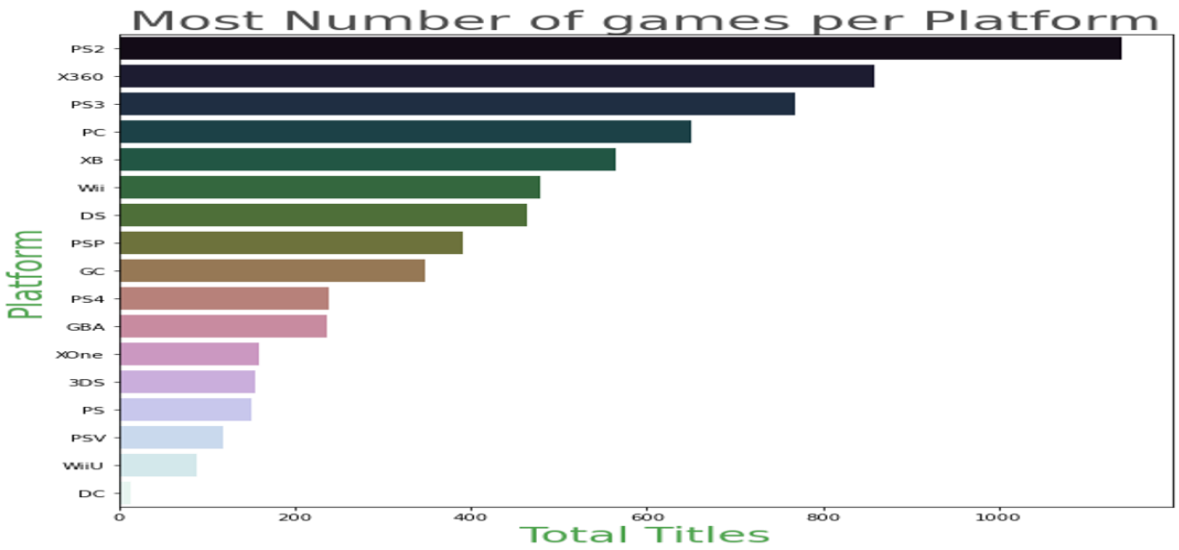

Fig: In this we want to observe the usage of the most popular platform for gaming. We found that gamers are mostly interested in The PS2 and second preference is given to the X360. The DC ,WiiU ,PSV are less preferred by the gamers.

figure 4:

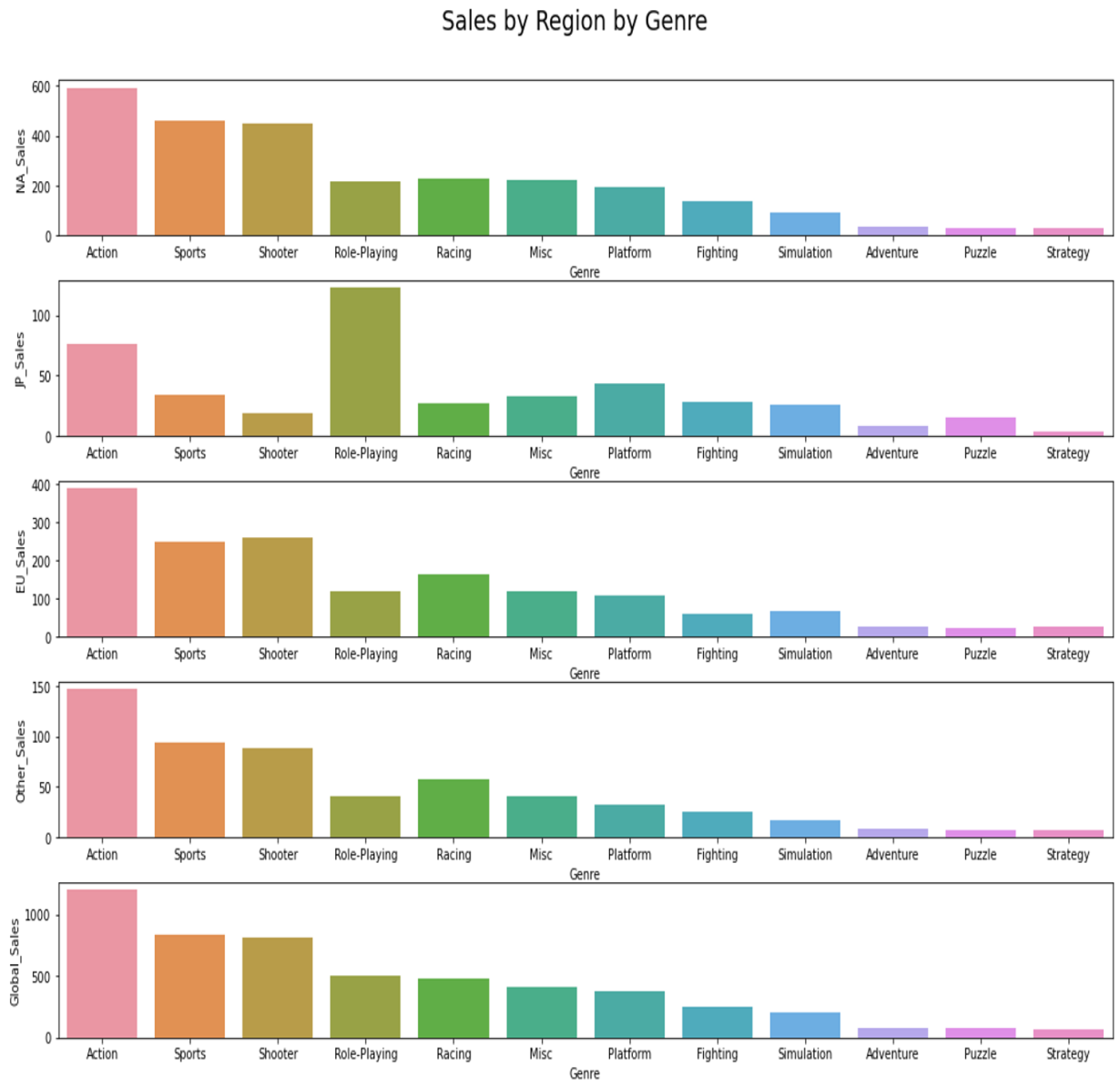

Fig: In this Visualization we have considered the sales by Region by Genre .

In North America, the action games had great demand whereas the Adventure, puzzle, Strategy had less demand. The sports and shooter games are equally preferred by the people. Role play and Racing are in the same trend of preference.

Considering the Japan sales, the role play had high demand. The action is the second most preferred genre in japan. Adventure ,puzzle and strategy are the least popular genres in the country.
In Other countries the action is mostly preferred, and the sports and shooter had the same preference. Adventure ,puzzle, strategy had low sales compared to fighting ,simulation.

The global sales are very high compared to other countries like Japan, north America. The action had a huge demand, sports and shooter genres had the same preference by the gamers. Adventure , puzzle, strategy had low sales.

Figure 5:

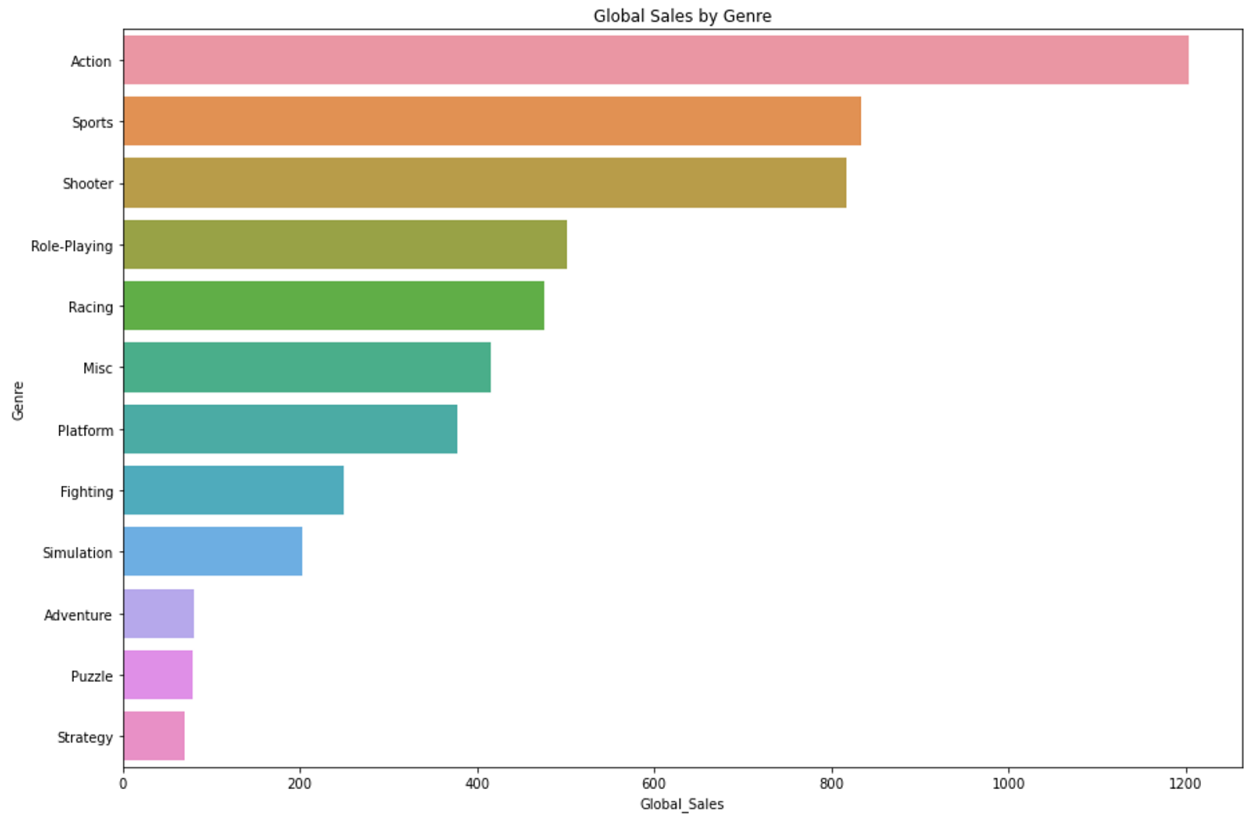

Fig: As we have seen that compared to other countries the global sales are very high. So, In this we want to visualize the global sales .The action and sports genres are mostly interested by the gamers. The sports and shooter games are equally preferred by the gamers. The role play, racing had 450 audience. The adventure , puzzle, strategy had less sales.

Tableau Visualizations

Figure 6:

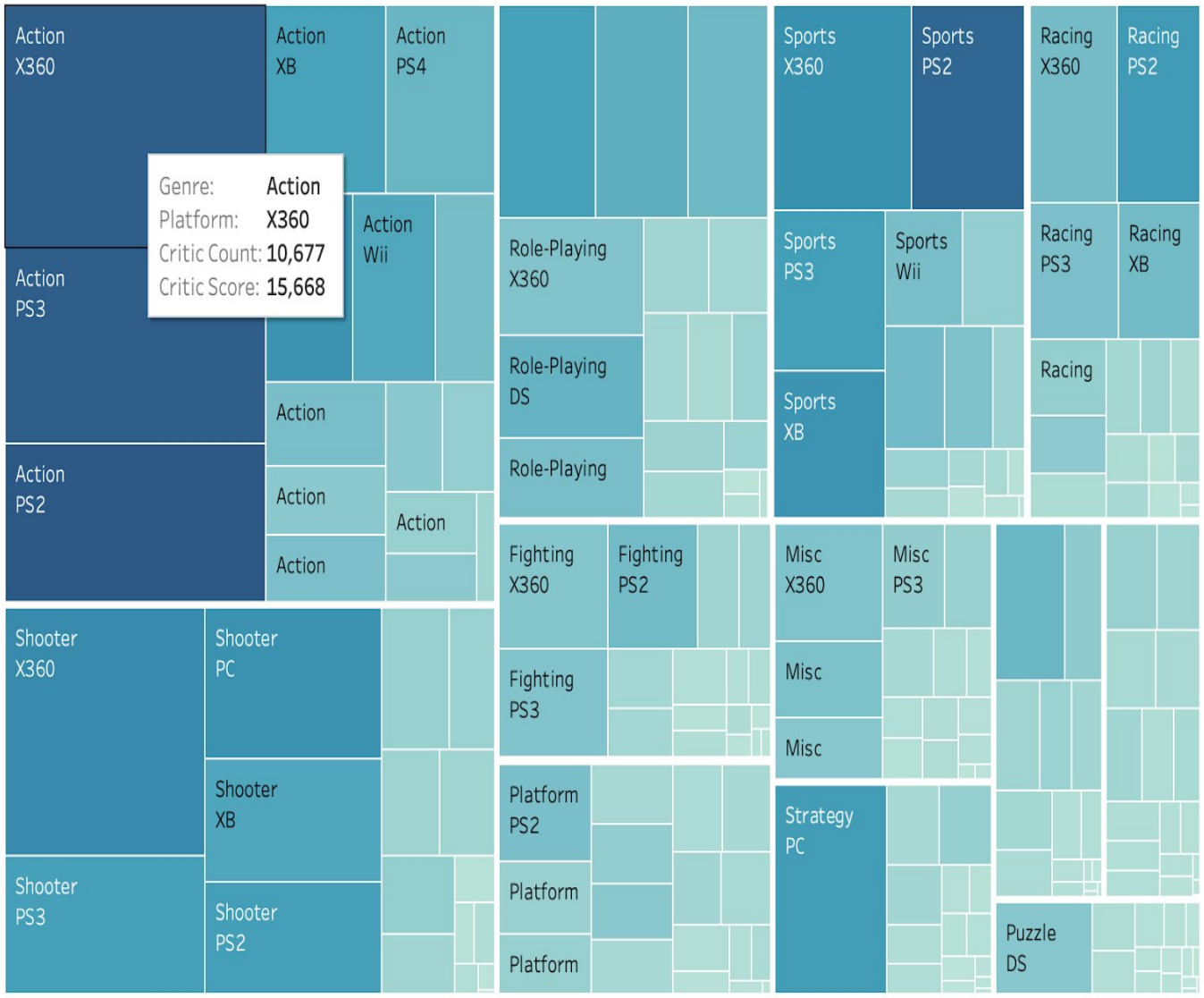

Fig: Critics are the people who review all the platforms .In this we have given to the Critic score and critic count . So, the maximum number of the critic score is given to Action on X360,Ps3,ps2.

Figure 7:

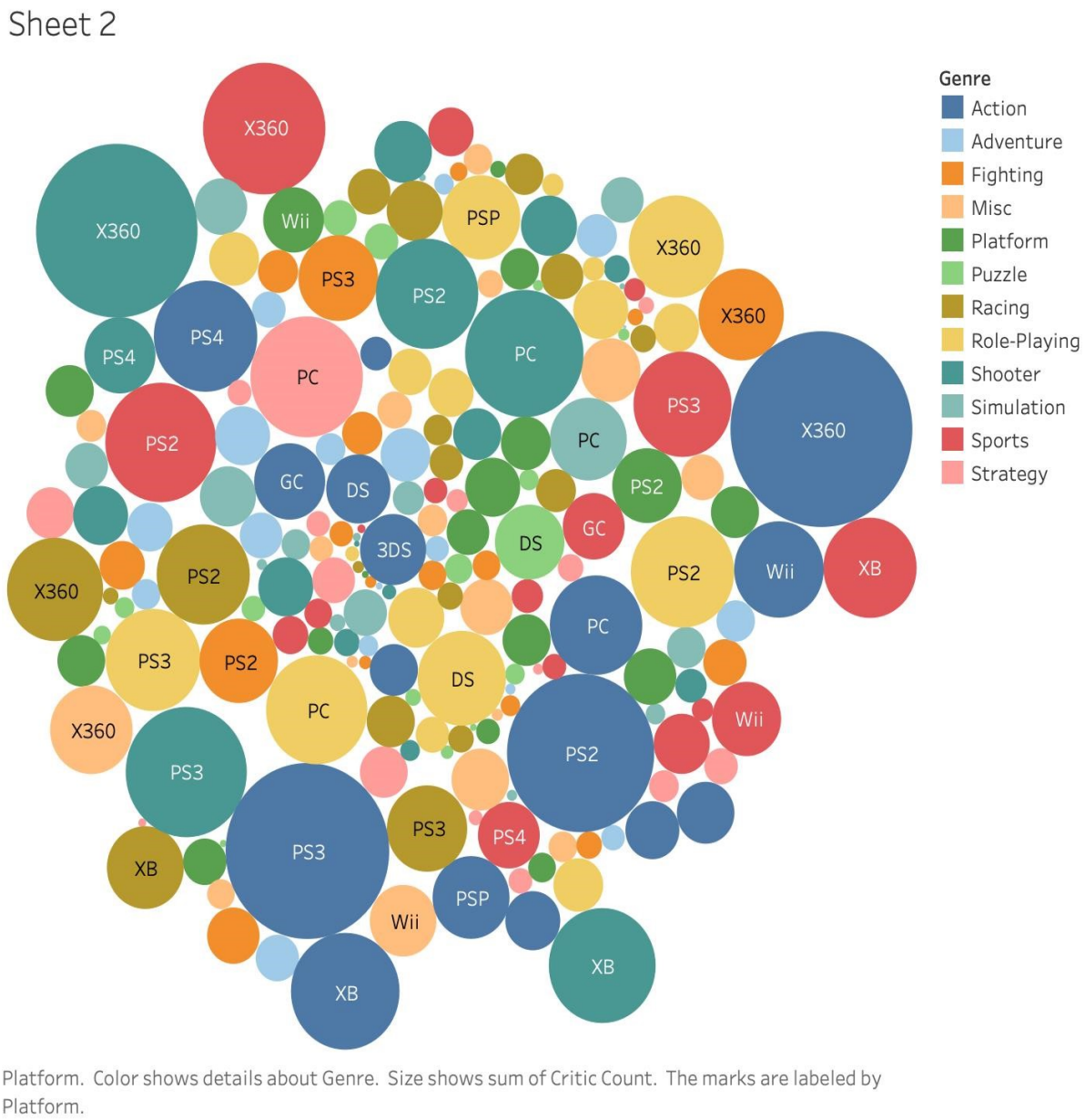

Fig: The Bubble chart representing critic count over various platform and genre. The wide varieties of platforms and Genre helps us in analyzing how critics perform their reviews and the reason why they get the audience base. It is really interesting how a critic chooses him platform for same game in specific genre.

Figure 8:

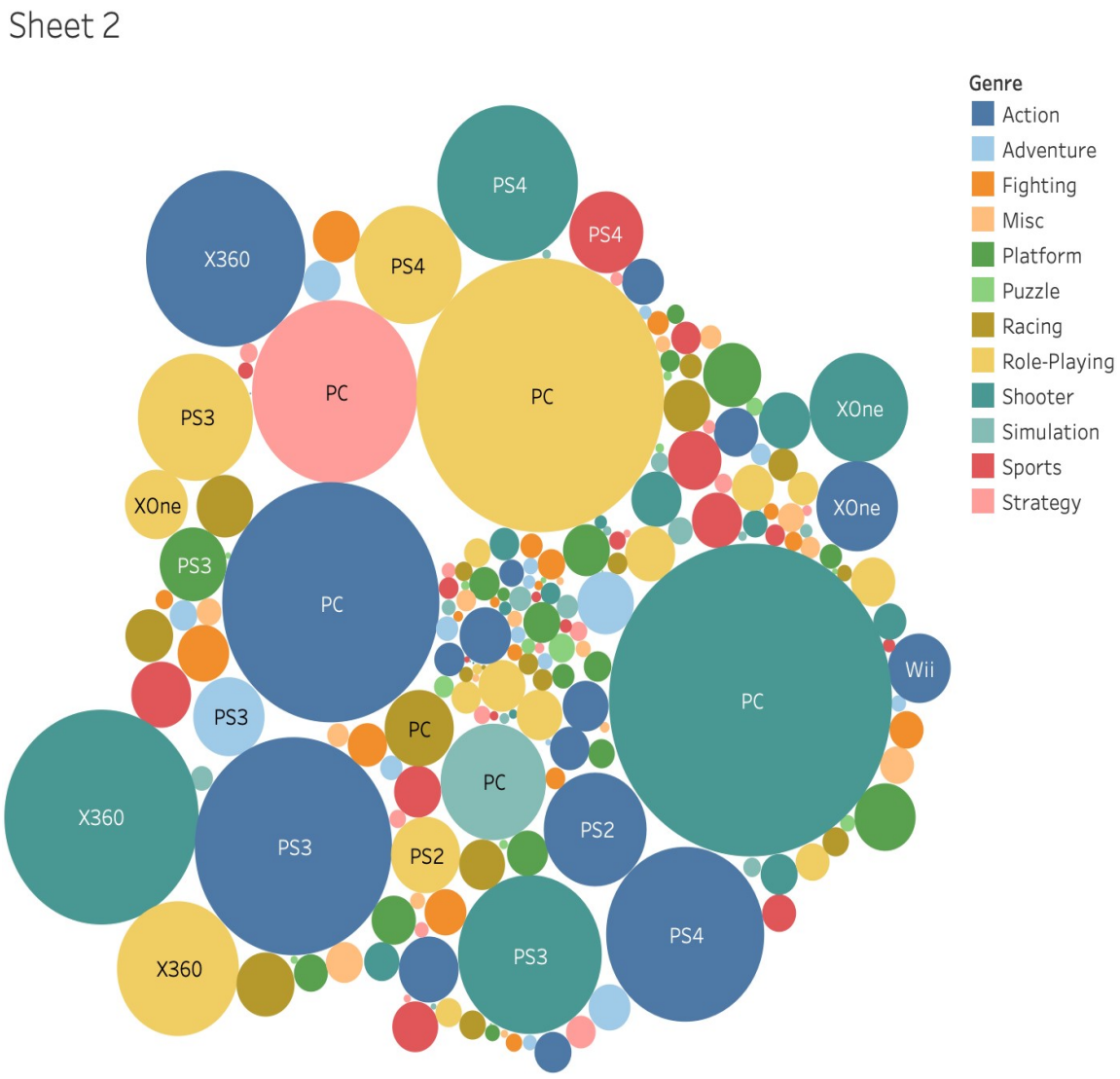

Fig: The importance of critic score over user score is explained in this bubble chart. We understand that the critic count is vital for the game and platform played as it is distributed randomly across all the platforms and genres unlike the user count that is spread only on certain platforms and genres. This chart also gives us some of the interesting split of genres across various platforms by game play.

<u>Visuals Using R : Radar Charts</u> Figure 9:

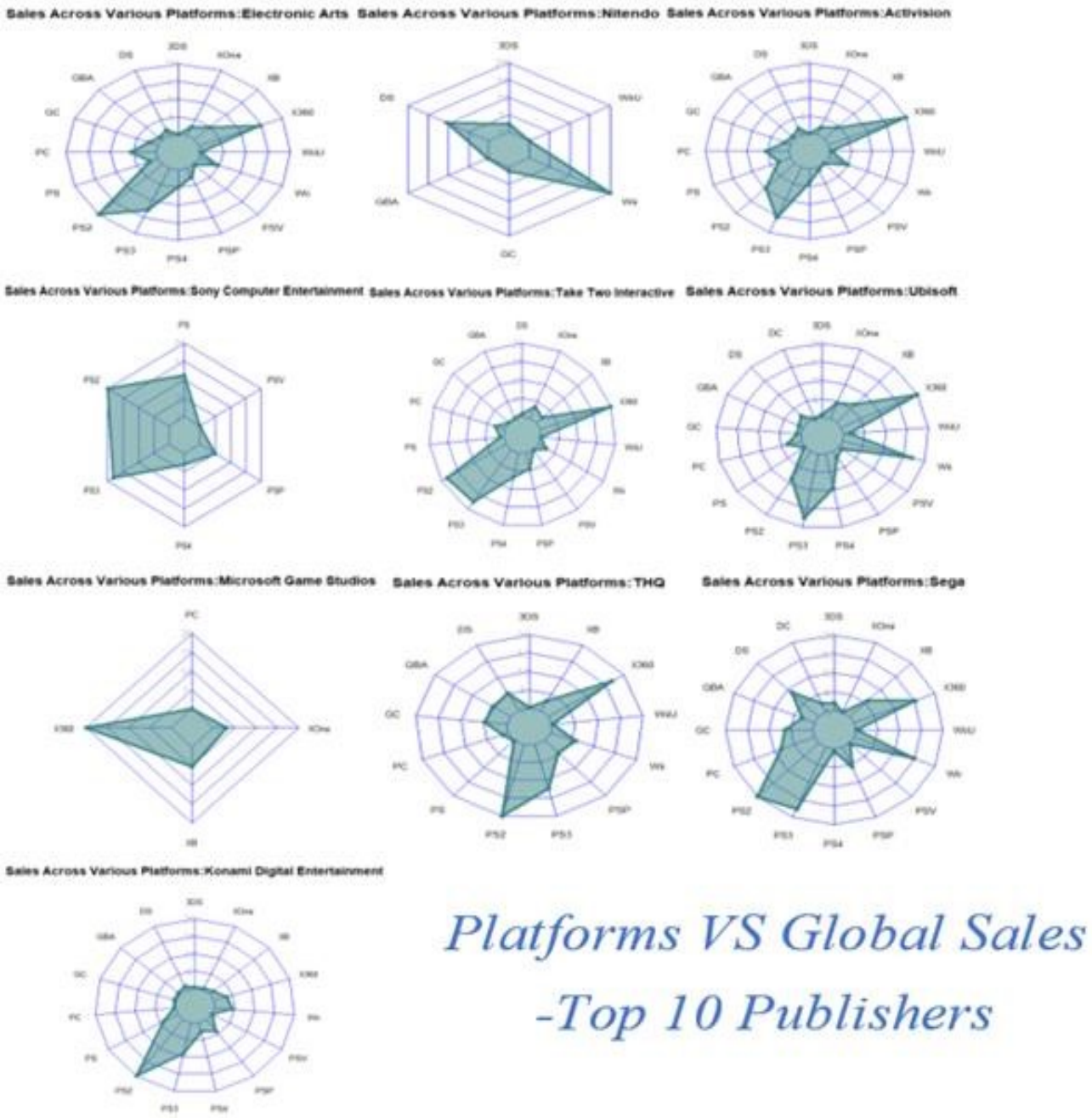

Fig: Thi figure illustrates the platform-oriented spread for the Top 10 publishers considering the Global Sales.

Figure 10:

Fig: The Radar charts that depict the information for the Top 10 Publishers considering their Global Sales. The Radar graph has huge significance in determining the spread of the publisher's interest and scope in the platform released and the Genre of the Game. The area covered by the graph determines the higher scope of importance and classification. These charts are performed in R and we have taken the features like Genre and Platform to analyze the results of the Publishers Global sales.

MARKET STRATEGY AND KEY TAKEAWAYS

We have been able to identify the key characteristics that will attract consumers to purchase our video game. Also, from the secondary dataset we identified the genre and demographic sales which is expected

to offer a potential opening in the existing video games market. Based on the Conjoint Analysis, we have determined that the consumers give more importance for video games which can be played offline on smartphones with action genres. By developing a video game that meets these specifications, we believe we will have a product that can compete in the upper mid to high tier price range. Moreover, after analyzing the demographics and sales data we would be targeting genres which have highest sales in a particular region. Additionally, a strong marketing campaign will offer us better competitive advantage to address the existing competition.

●The conjoint analysis concludes that users prefer video games which can be played offline on smartphones with action genres. Hence, our company would be targeting to launch video games with this specification.
●It is seen that for the last 5 years only few trending games have crossed 100k copy sales and the low revenue games have come into play.
●As per stats N.A and Europe share 75% of market share by revenue.
●Ages 18-24 play on a weekly basis and are interested in Graphics, Game story and Nature of play.
●PlayStation, Pc and Smartphone are popular platforms. The trends are PSP and Xbox for Sports, Shooting and Action.
●EA has top-most releases and Wii has top revenue.
●Action and Adventure sports have been the most picked Genres.
Observed : Grand Theft Auto, Super Mario, Football 2016,Assassin and PubG Mobile.

<u>The 4P's in Market Research:</u>

The Marketing Mix model identifies the 4p's as a set of marketing objectives in the target market. We would describe how the product, place, price, and promotion influence our video game.

1. Product: The above analysis clearly mentions that we would mostly be interested in developing a game either in action or sports genre as it is the favorite pick by many responders. We would also do certain specifications to our product like mode independent, platform independent and frequent updates to attract more customers. The product with high graphics, game story and multiplayer would surely attract a gamer.
2. Place: Analysis of the sales data gave us insight that publishing a game in Europe or North America would be more beneficial. It could also be beneficial if we extend it globally after its phase 1. A role play action game would also be an option when we look to publish it in japan as they are interested in those genres.
3. Price: The survey that we collected had people who were mostly playing games that were below $50 /month and which also offered some free subscriptions or updates. We would implement this on our game initially so that it could be easily purchased within the budget of the customer and allow certain free subscription for a period. The other factor would be game sharing up to 3 people can be made accessible initially as it would interest others in buying our package.
4. Promotion: It is understood that user and critic reviews play a major role in promoting the game. We could record live streaming of top gamers and developers to interact with beginners and help

them develop our game play. Apart from these, the promotion on several gaming websites, giving the players trial versions would also be a great idea.

LIMITATIONS AND FUTURE WORK

The conjoint analysis is limited towards 150 respondents and the insights for factor importance like Genre, Platform and Mode type. However, an increase in survey records can give more accurate outcomes. Due to the massiveness of the secondary dataset we have a lot of data points which help us in understanding the insights but it's challenging as different games have different publishers and there is a new one every day. The sample made was skewed and had good point variability, but it was not for the complete data as there were a lot of missing values. The other limitation was we only had data specific to certain age groups and that could help to analyze only their priorities and it would lead to an availability bias. This could be handled if we focus on collecting data points from varied personnel.

The available data was from 1995 to 2016. So, it is also an availability bias as we don't have the trending data from 2016 and the data before 2000's is not really useful as we don't use those platforms anymore. The Future work would be implementing the game over various platforms across several linked genres. It would also be functional if we develop a game that is liked and played by all age groups. The factors like mode and platform independence would give the game more advantage. Later, we can also develop an interface or a specific platform for our game play as trending simulators and devices are highly purchased.

References

Dhruv. (n.d.). *K Means Clustering In R Programming*. Retrieved from Geeks for Geeks: https://www.geeksforgeeks.org/k-means-clustering-in-r-programming/

SID_TWR. (2019, MAY 20). *VIDEO GAMES SALES DATASET*. Retrieved from KAGGLE: https://www.kaggle.com/sidtwr/videogames-sales-dataset