

DATA ANALYSIS PORTFOLIO

PREPARED BY
EDIDIONG UDOH



Professional Background

My name is Edidiong Udoh, a recently certified data analyst, who has accomplished all the required missions and tasks for the certification. I am proficient in visualizations using tableau and excel, PowerPoint presentation and other Microsoft offices, I enjoy using my skills to contribute to the advancement of companies and businesses that seeks my services. I am smart, sound, insightful, resourceful, dynamic with a good sense of humor and communication skills. The link to my tableau page where I display my jobs done is on the appendix, along side other raw files of my work. My core values are Growth, Integrity, Puzzle solving and Services (GIPS). I concluded my certification program at Entrylevel, 2022. At the cause of the program, I earned badges of excellence for my performance and proficiency. I enjoy watching educational podcast on YouTube to increase my knowledge, I also enjoy travelling as it is educational.

Portfolio Outline



Professional Background	1
Table of Contents	2
Udemy Project Description	3
The Problem	4
Design	5
Findings	6-9
Analysis	10-11
Conclusion	12
Espado Project Description	13
The Problem	14
Data Design	15
Findings	16-18
Data Analysis	19-20
Conclusion	21
Appendix	22



Udemy Project Description

Udemy appointed one of its company's data analyst, which happens to be me, to analyze their course revenue using the dataset they provided.

The purpose of this task was to uncover where potential opportunities to increase business revenue lies. Since the dataset provided was on different courses/subject, it was sent on different spreadsheets.

The different spreadsheets were consolidated, cleaned up (for high quality data and high quality insights to be drawn), analyzed, after which visualization of insights found were made. The business problem in view was "How to increase revenue?". The analysis seek to understand the different performance of each course, ranging from number of subscribers, to prices of each course at different levels, which will help solve the business problem.

To arrive at an insight to solve the above mentioned problem, the following was done: consolidation of data, clean-up, organization (including sorting, creating of pivot tables, etc.), visualization.

Since it is a prescriptive analysis where previous data set are used to recommend future happenings, Regression analysis technique was used (Linear regression) to determine relationships between different parameters like Number of subscribers over prices, etc.

At the end of the analysis, it was observed that Web Development Subject had the highest Number of Subscribers, Web Development generated highest revenue, from the top 20 subscribed courses, 60% were free courses while 40% were paid.

In conclusion, based on the insight gotten, Web Development courses should be promoted more to **Increase Revenue** since it has more subscribers and has more cost. More emphasis should be led on the Intermediate Level, as it generated more revenue in time past, also, more paid courses and lesser free courses should be made available.

The Problem

Udemy seeks to know the different course revenue generated, to be able to uncover where the potential opportunities to increase revenue lies. The Manager of Udemy requested the project to be completed in 2 weeks. Different data on courses from different topics were provided for this analysis, including the Number of Subscribers, Prices, levels, course duration, date published, number of lecturers, course titles, subject, and other relevant data. The data was presented in a spreadsheet, according to the different subjects.

To achieve this, the following questions have to be answered:

- 1.What are the top 20 most Subscribed Courses?
- 2.Which Subject has the Highest Total Number and Average Number of Subscribers?
- 3.How does Average Cost of each Subject vary across the Different Levels?
- 4.How does the Subject Ratings Vary across Levels and Which Subject has the Most Rating and what Level?
- 5.Does Publication Date have an Impact on Number of Subscribers?

Data Design

The data based on the subjects, were presented on different spreadsheets with their corresponding information. The data was presented in a raw and unorganized state.

The Different Data collected were consolidated for proper rating of the data quality. It was observed that the data had some inconsistency, blank cells and duplicates, which reduced the quality of the data and would in turn affect the result. To avoid this, Clean up was carried out, which included removal of blanks, duplicates, modifications, etc. This Improved the data and made it more reliable and of high quality to produce good insights.

The Cleaned-up data was then fit in a Linear Regression Model to predict the relationships between various parameters, to determine the performance of each course.

The Visualization tool used in this project was Tableau Visualization and Google charts. These tools were used because it has hundreds of data import options available from csv files to Google Ads and works with dynamic data . This aids in proper organization and representation of relationships of different parameters.

Findings

Table 1: Top 20 Most Subscribed Courses

S/N	Num_Subscribers	Course Title	Level	Course_status	Free Beginner Course	Course_Duration	Date_published	Subject
1	268923	Learn HTML5 Programming From Scratch	Beginner Level	Free	Yes	10.5	2/14/2013	Web Development
2	161029	Coding for Entrepreneurs Basic	Expert Level	Free	No	3.5	6/9/2013	Web Development
3	121584	The Web Developer Bootcamp	Beginner Level	Paid	No	43	11/2/2015	Web Development
4	120291	Build Your First Website in 1 Week with HTML5 and CSS3	All Levels	Free	Yes	3	4/8/2014	Web Development
5	114512	The Complete Web Developer Course 2.0	Beginner Level	Paid	No	30.5	3/8/2016	Web Development
6	101154	Free Beginner Electric Guitar Lessons	All Levels	Free	Yes	4.5	6/15/2012	Musical Instruments
7	98867	Web Design for Web Developers: Build Beautiful Websites!	All Levels	Free	Yes	3	4/13/2015	Web Development
8	84897	Learn Javascript & JQuery From Scratch	All Levels	Paid	No	2	10/10/2013	Web Development
9	83737	Practical PHP: Master the Basics and Code Dynamic Websites	Intermediate Level	Free	No	6.5	7/19/2014	Web Development
10	79612	JavaScript: Understanding the Weird Parts	All Levels	Paid	No	11.5	3/12/2015	Web Development
11	75499	PianoForAll - Incredible New Way To Learn Piano & Keyboard	Beginner Level	Paid	No	30	8/7/2014	Musical Instruments
12	73783	Angular 4 [formerly Angular 2] - The Complete Guide	Beginner Level	Paid	No	22	2/11/2016	Web Development
13	73110	Beginner Photoshop to HTML5 and CSS3	All Levels	Free	Yes	2	7/27/2012	Web Development
14	72932	Web Development By Doing: HTML / CSS From Scratch	All Levels	Free	Yes	1	9/25/2013	Web Development
15	70773	HTML and CSS for Beginners - Build a Website & Launch ONLINE	All Levels	Free	Yes	6	3/19/2015	Web Development
16	69186	Become a Web Developer from Scratch	All Levels	Paid	No	27.5	11/19/2011	Web Development
17	65576	Bitcoin or How I Learned to Stop Worrying and Love Crypto	All Levels	Free	Yes	8	4/20/2013	Business Finance
18	64128	Quickstart AngularJS	Beginner Level	Free	Yes	1.5	11/22/2014	Web Development
19	59639	Learn Responsive Web Development from Scratch	All Levels	Free	Yes	4.5	12/9/2013	Web Development
20	59361	Learn and Understand AngularJS	Beginner Level	Paid	No	7	9/24/2014	Web Development

Web Development subject has 80% of the 20 top subscribed courses

Figure 1: Sum Number of Subscribers By Subject

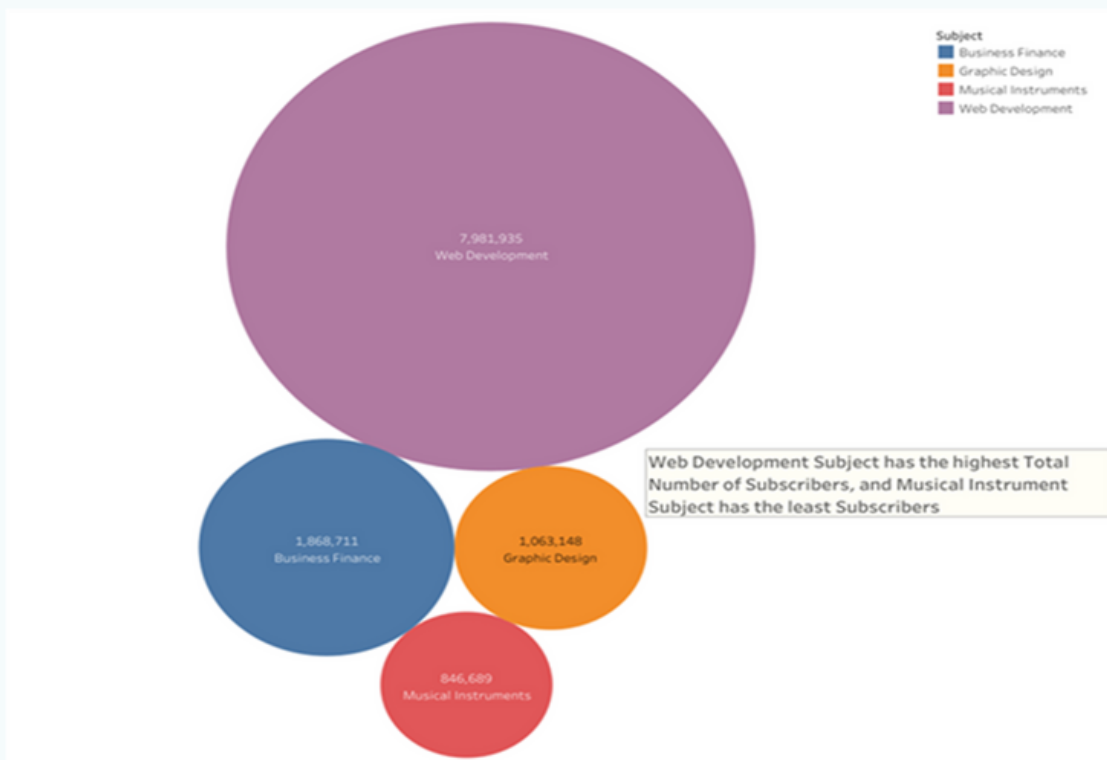
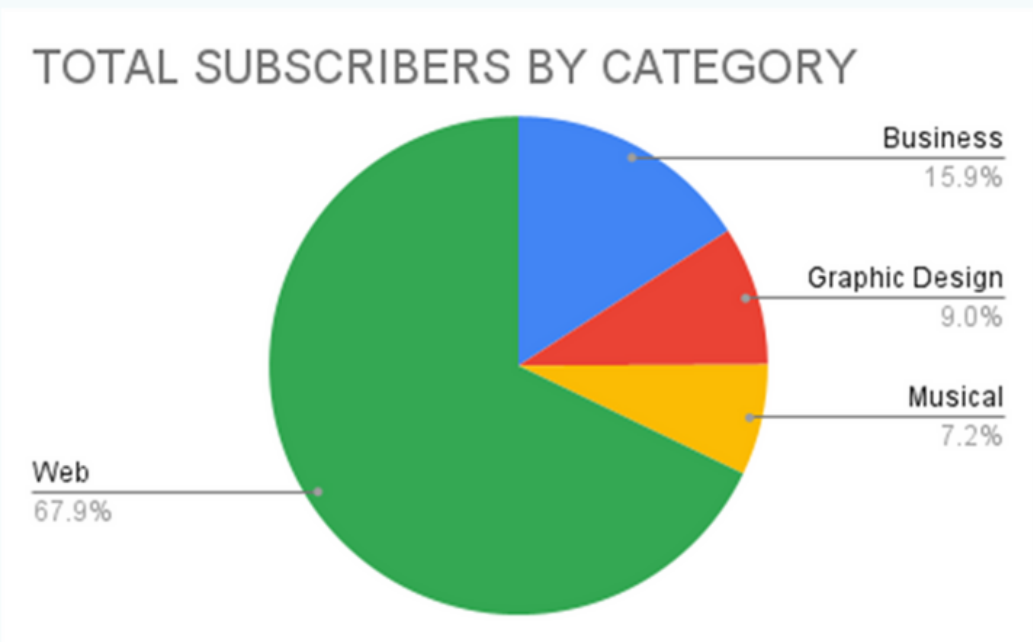


Figure 2: Total Subscriptions by Category

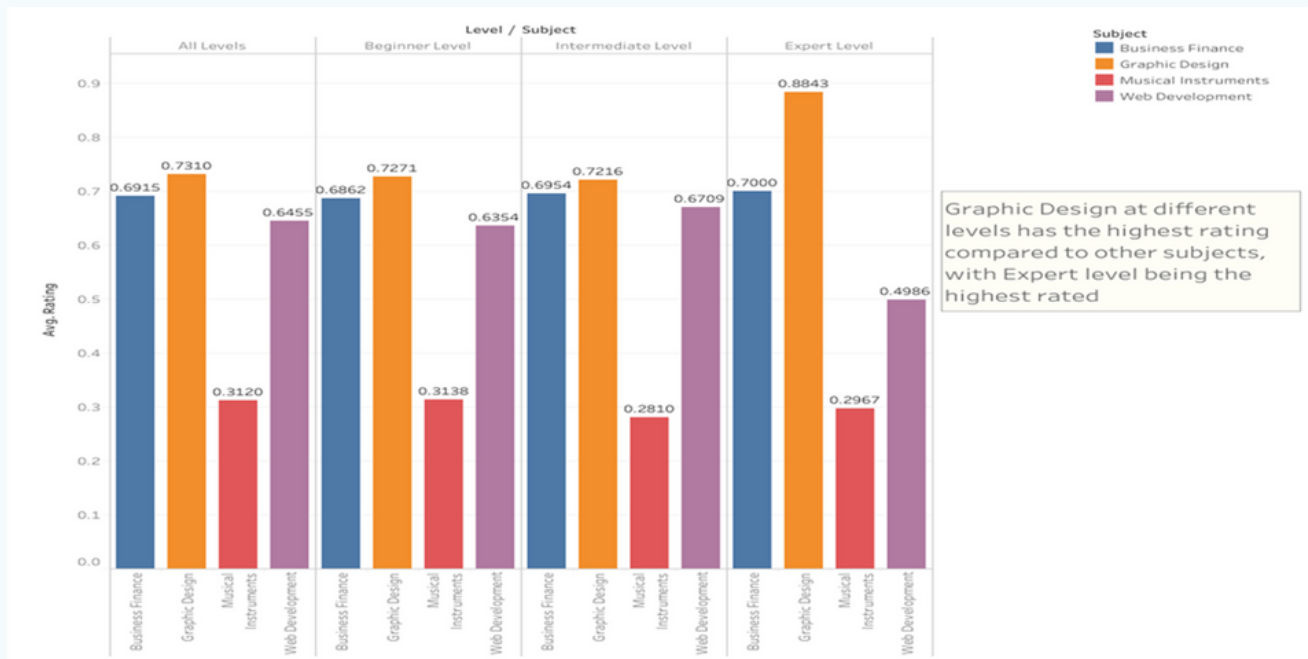


Web development has the highest number of subscribers both in the average and in category

Table 2: Average Subscribers by Subject

Subject	AVERAGE of Num_Subscribers
Business Finance	1569.0
Graphic Design	1766.0
Musical Instruments	1245.1
Web Development	6635.0
Grand Total	3199.3

Figure 3: Average Rating for Each Subject at Different Levels



At Different Levels (All Level, Beginner, Intermediate and Expert), Web Development has the highest Price, with Intermediate Level having the Highest Price. Also, there are more paid users than free users, even though in the top 20 subscribed courses, free users were more by 20%.

Figure 4: Average Cost per Subject at each Level

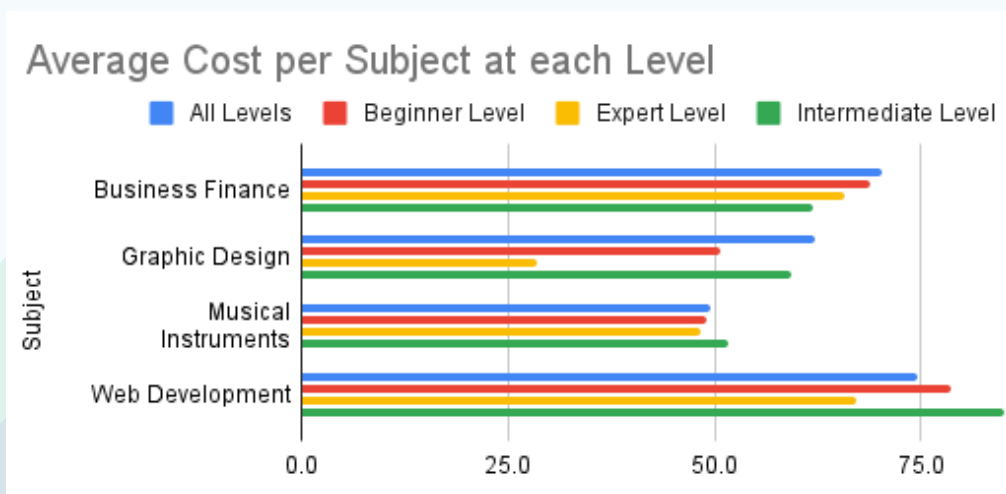
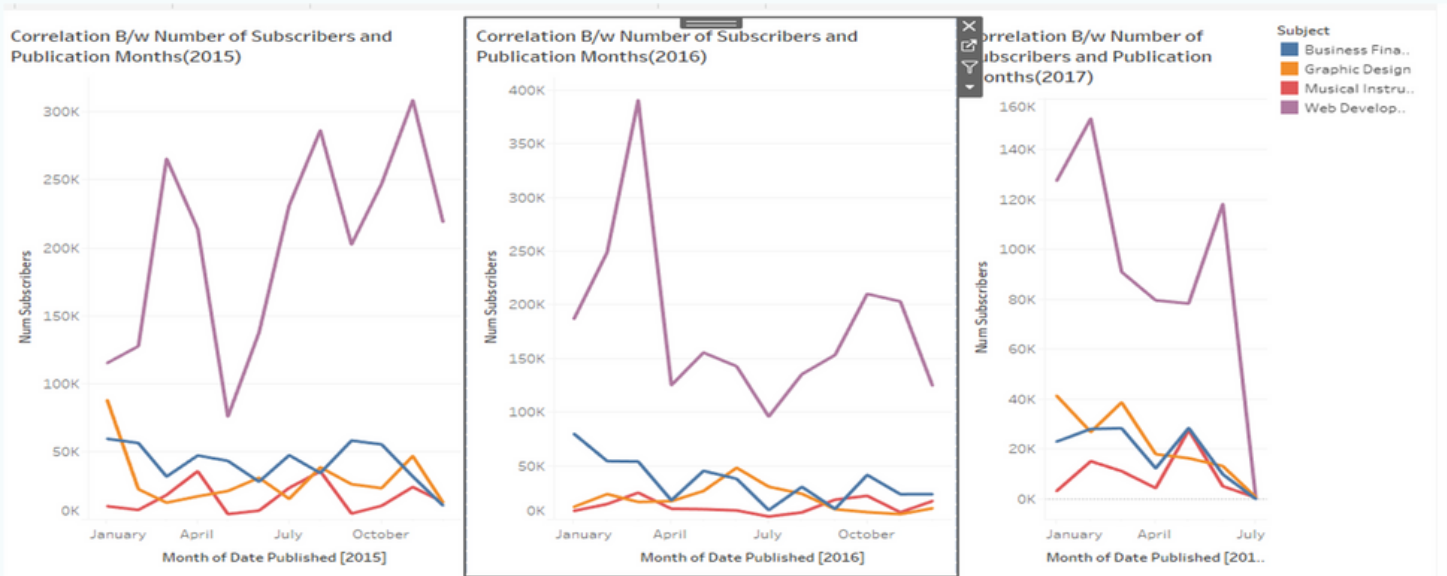


Table 3: Free/Paid Users

Course_Type	COUNTA of Course_Type
Free	311
Paid	3365
Grand Total	3676

Figure 6: Correlation Between Number Of Subscribers And Publication Months (2015-2017)



More Subscribers at Month March and November and Great Decline between April-July, and December in the last 3 years

Analysis

The Five-Why Analysis:

How to Increase Business Revenue?

1. Why have there been no increase in the business revenue?

Ans: Because there are no enough paid course subscribers.

2. Why are there no enough paid course subscribers?

Ans: Because there hasn't been enough awareness made for the courses

3. Why hasn't there been enough awareness made for the courses?

Ans: Because the campaign personnel don't really know which course to focus more on for a better result.

4. Why are the campaign personnel not aware of which course to focus more on for a better result?


Ans: Because they are not aware of the response level and behavior of previous and present subscribers towards the courses

5. Why are they not aware of the response level and behavior of previous and present subscribers towards the courses?

Ans: Because they do not track and understand the different performance of each course

The Root Cause: Not Understanding the Different Performance of Each Course

To be able to uncover the potential opportunities in Increasing Revenue, The Performance of each course must be fully understood and structured out.



1. Number of Subscribers: Web Development Subject has the highest Number of Subscribers, and Musical Instrument Subject has the least Subscribers.

2. Average Number of Subscriber: Web Development Subject has the highest Average Number of Subscribers, and Musical Instrument Subject has the least Subscribers.

3. Average Rating per Subject for each Level: Graphic Design at different levels has the highest rating compared to other subjects, with Expert level being the highest rated.

4. Average Cost per Subject for each Level: At Different Levels (All Level, Beginner, Intermediate and Expert), Web Development has the highest Price, with Intermediate Level having the Highest Price. Also, there are more paid users than free users, even though in the top 20 subscribed courses, free users were more by 20%.

5. Correlation between Number of Subscribers and Date Published (between 2015-2017): More Subscriber at Month March and November and Great Decline between April-July, and December in the 3 years . Publication date has influence on number of subscribers.

Conclusion

85%, 10% and 5% of the top 20 most subscribed courses are from Web Development, Musical Instrument and Business Finance Subjects respectively. Irrespective of this, 60% were free courses and only 40% were paid.

Web Development Subject has the highest Number of Subscribers.

Web Development generated highest revenue, with highest cost at Intermediate level, followed by Business Finance at All Level. Majority of the users were paid users.

Graphics Design has the highest rating at the different levels, with Expert Level having the highest rating.

There is a relationship between Number of Subscribers and Date of Publication.

Recommendation

Web Development courses should be Promoted more to Increase Revenue. More emphasis should be led on the Intermediate Level, as it generated more revenue in time past.

Availability of Free courses should be minimal. More paid courses and lesser free courses.

More Course Publications should be done on Month February, March and October for more Subscribers.

Even though Web Development has the highest price/cost and number of Subscribers, it has the second least rating at the different levels. This calls for concern.



Espado Gaming World Project Description

A video game selling company has been out of market for some years now and wants to know how to get back to market and generate revenue.

This report is to help Espado Gaming World understand their sales performance/track data in the past years in order to know their targeted customers to make more sales and in turn generate revenue. The data received from the company was of high quality, hence needless of clean-up. "How to get back to business and generate revenue?" was the problem to be resolved. To resolve this, some questions concerning the business performance in the past years (2013-2020 sales), has to be answered like: total revenue generated from 2013-2020, State that generated the highest revenue and Genre sales generally and in specific location (genre preference among locations of sale).

This was a descriptive/prescriptive analysis. To solve the above stated problem, the following were carried out: Exploratory analysis (including sorting and organizing the data using Pivot table), Visualization (Using Google Charts and tableau).

Since it is a prescriptive analysis where previous data set are used to recommend future happenings, Regression analysis technique was used (Linear regression) to determine relationships between different parameters like Genre sales over locations, etc.

At the end of the analysis, it was observed that revenue was generated most in 2013 of which It declined as the Year went by, and no sales was made 2018-2019. North America (NA) had the highest sales in 2013-2015. Genre "Action" had the highest Sales and there are genre preference among locations of sale.

In conclusion, based on the findings, video game genre with the highest sales should be produced more and made available at the locations in which there are preferred, also, the marketing campaign should be based on targeted audience (preferred genre at preferred location).

The Problem

Espado seeks to know the sales performance of different video games in order to proffer possible ways of getting back to business and generating business revenue. The Manager requested the project to be completed in one week. The information needed for this project is the past years sales tracking record, including information like the game names, genre, years, the different location sales records, the global sales records, etc. The data was presented in a spreadsheet.

To achieve this, the past sales records were analyzed (2013-2020 sales), to answer the following questions:

1. What is the total number of sales between 2013- 2020
 2. What are the total revenue generated from 2013-2020?
 3. Which state generated the highest revenue?
 4. Which Genre sales more globally and at Different Location?
- B. Does the different sales state have genre preference?

Note: All prices are in Million dollars

Data Design

An exploratory analysis was carried out on the data and it was observed that the data was of high quality, it was free of blanks, inconsistency, duplication, etc. Clean up was not required. Since the data was of high quality, reliable and high quality insights were produced.

The Data was then fit in a Linear Regression Model to predict the Relationships between various parameters, to determine the Performance/ Track of Sales.

The Visualization tool used in this project was Tableau, visualizing using Bar charts, Column Charts, Line graphs, and the likes of other. These tools were used because it has hundreds of data import options available from CSV files to Google Ads and works with dynamic data. This aids in proper organization and representation of relationships of different parameters.

Findings

Table 1: Number Of Sales Between 2013–2020

Years	Count of Global Sales
2013	546
2014	582
2015	614
2016	344
2017	3
2020	1
Grand Total	2090

2014 had the highest number of sales, followed by 2013. No sales were made b/w 2018/2019.

2013 generated the highest revenue

Table 2: Total Revenue Generated Between 2013–2020

Years	Sum of Global_Sales
2013	368.11
2014	337.05
2015	264.44
2016	70.93
2017	0.05
2020	0.29
Grand Total	1040.87

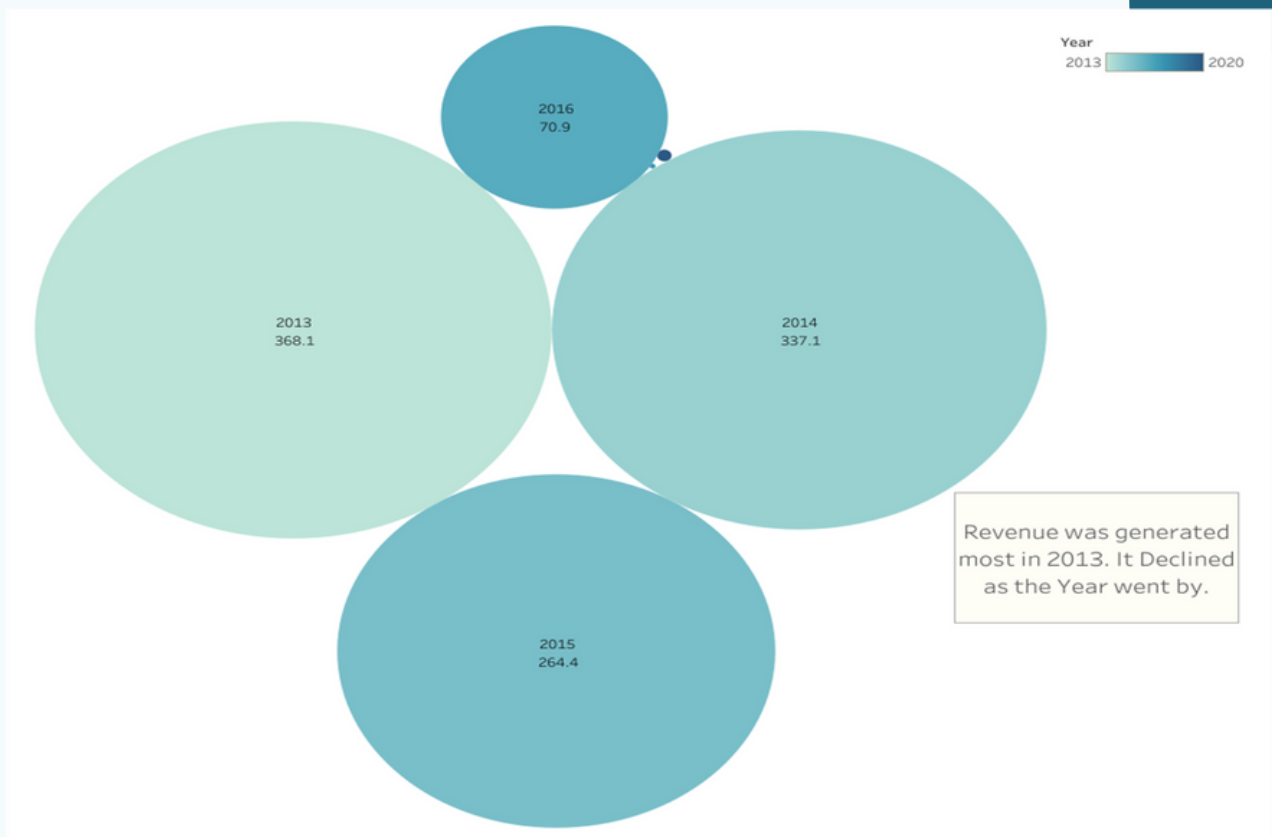
Note: All prices are in Million dollars

Table 3: Total Revenue Generated Between 2013–2020 vs Location sales

	Values			
Row Labels	Sum of NA_Sales	Sum of EU_Sales	Sum of JP_Sales	Sum of Other_Sales
2013	154.77	125.8	47.59	39.82
2014	131.97	125.65	39.46	40.02
2015	102.82	97.71	33.72	30.01
2016	22.66	26.76	13.7	7.75
2017	0	0	0.05	0
2020	0.27	0	0	0.02

North America (NA) generated highest revenue

Figure 1: Sum Revenue Generated Between 2013–2020



Note: All prices are in Million dollars

Figure 2: Location sales over years

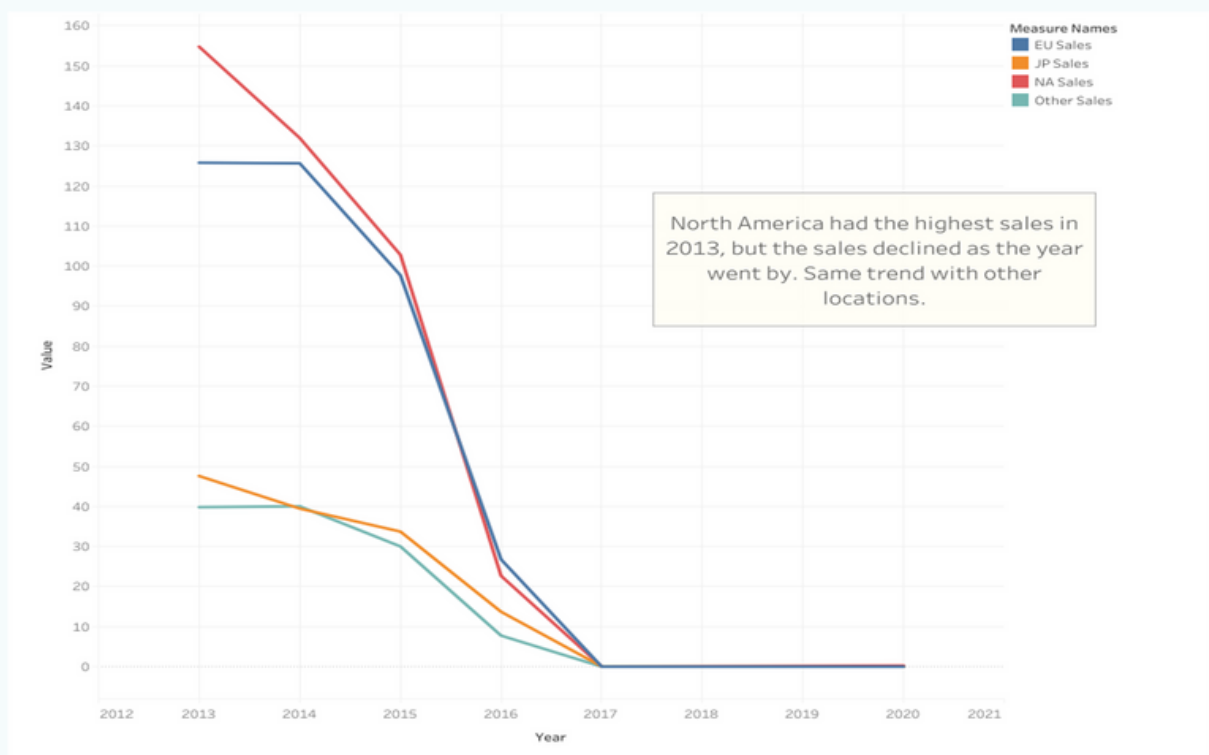


Figure 3: Global Sales vs. Genre

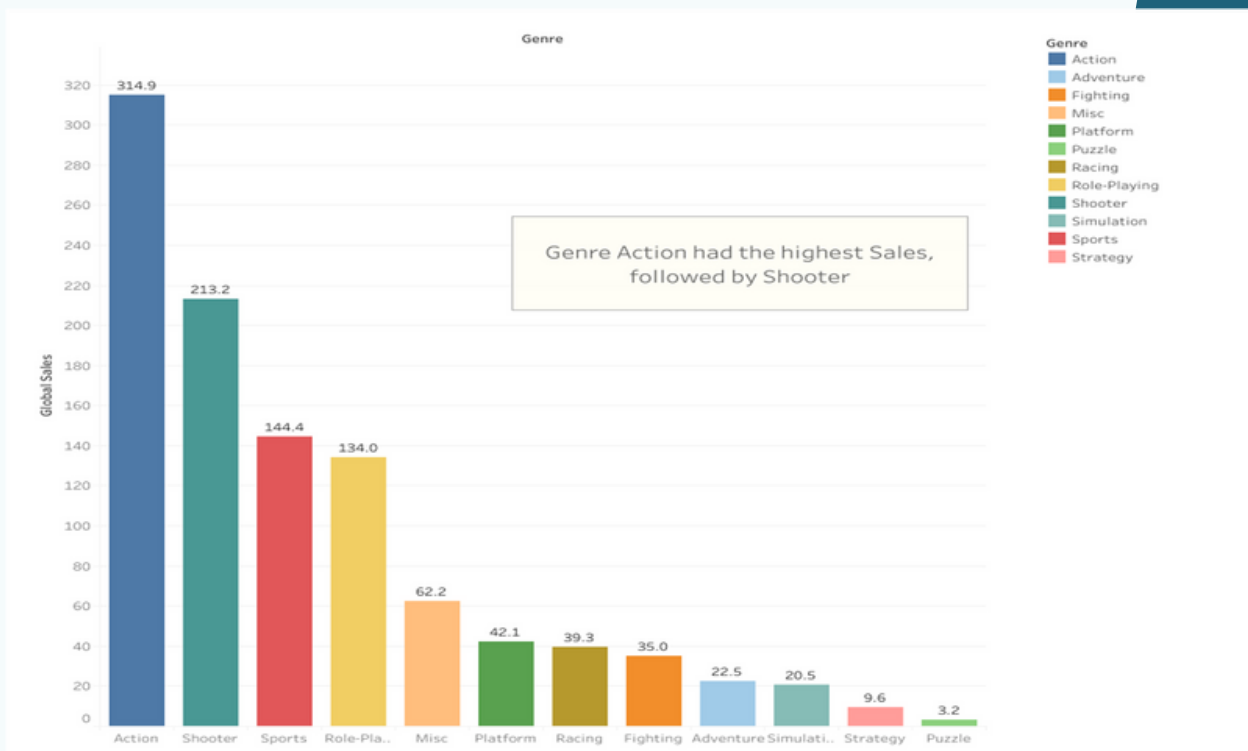
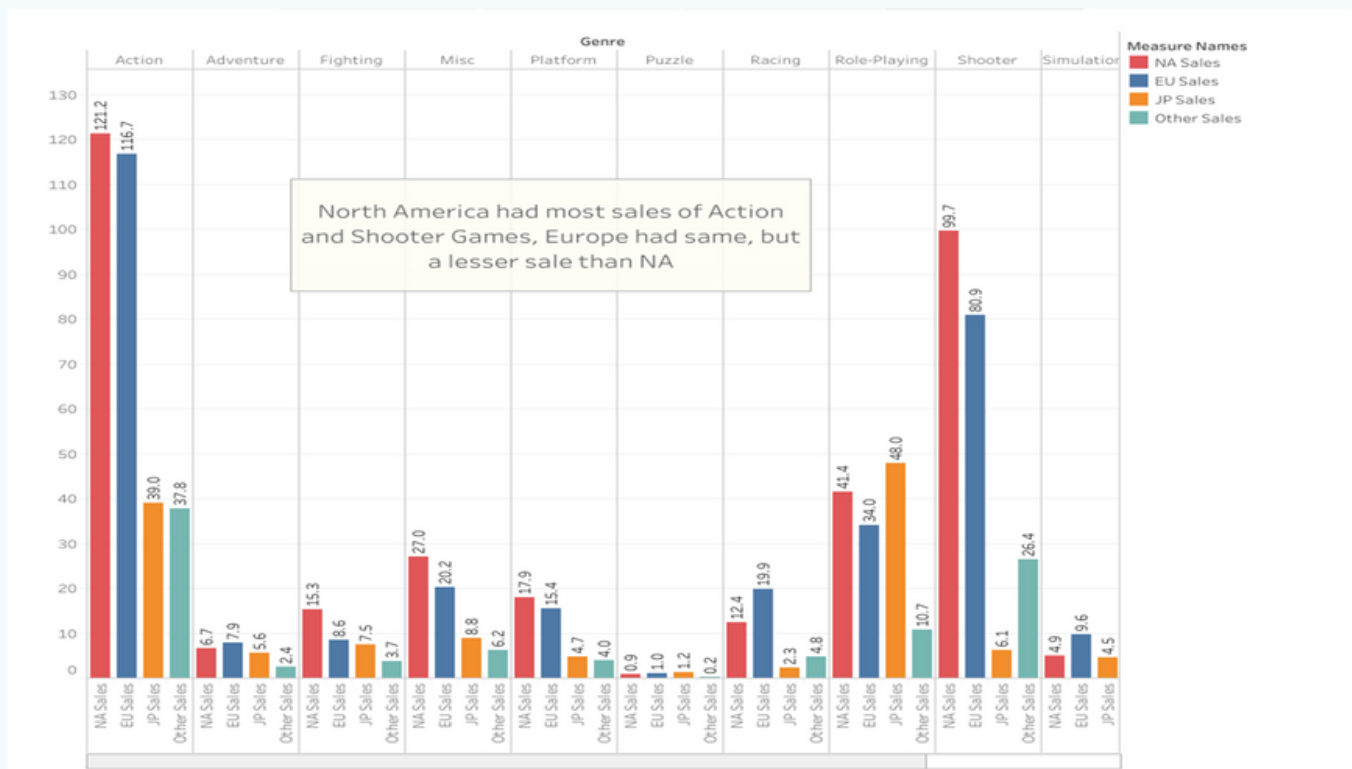


Figure 4: Different Genre Sales at Different Locations



Analysis

The Five-Why Analysis:

How to Generate Revenue at Espado Gaming World?

1 Why has there been no revenue generated in the past few years?

Ans: Because there haven't been sales.

2. Why haven't there been sales?

Ans: Because the marketing campaign results came out poor.

3. Why are the marketing campaign results poor?

Ans: Because they didn't have a targeted audience

4. Why don't they have a targeted audience?

Ans: Because the sales department doesn't organize the sales tracks of video games sold, thereby giving the marketing department no information to work with.

5 Why doesn't the sales department organize their sales tracks of video games sold?

Ans: Because they do not understand the Sales Performance or tracking data, in respect to genre or location.

The Root Cause: Not understanding and organizing Sales Performance or tracking data, in respect to genre or location.

To Generate Revenue, Sales Performance or tracking data, in respect to genre or location must be fully understood and organized.

The Insights drawn are:

- Total Number of Sales Between 2013- 2020: 2014 had the highest number of sales, followed by 2013. No sales were made b/w 2018/2019.
- Total Revenue Generated between 2013-2020: Revenue was generated most in 2013. It declined as the Year went by. In 2018 and 2019, no sales was made. In 2020, only 1 sale was made.
- Location Sales over Years (2013-2020): North America (NA) had the highest sales in 2013-2015, but the sales declined as the year went by. Same trend with other locations.
- Global Sales vs. Genre: Genre "Action" had the highest Sales, followed by "Adventure"; "Puzzle" had the least sales.
- Different Genre Sales at Different Locations: North America (NA) had most sales of Action and Shooter Games, Europe (EU) had same, but a lesser sale than NA. Puzzle had the least sales, with Japan (JP) having the highest sales.

Conclusion

Total Number of Sales Between 2013-2020 are as follows: 546, 582, 614, 344, 3, 1 respectively.

Revenue generated at year 2013, 2014, 2015, 2016, 2017, 2018, 2019 and 2020 are \$368.1M, \$337.1M, \$264.4M, \$70.9M, \$0.05M, \$0, \$0, and \$0.29M respectively.

North America (NA) generated the highest revenue for the past 8 years, except for 2016, where Europe (EU) generated revenue slightly higher than NA.

Genre "Action" of video games sold most globally, followed by genre "Shooter". Genre "Puzzle" had the least sales globally.

From the findings, there is preference of video game genre at the different states. North America (NA) and Europe (EU) sold more of genre "Action" and "Shooter" than any other genre, while Japan (JP) sold more of "Role Playing" genre.

Recommendation

Video game genre with the highest sales should be produced more and made available at the appropriate places i.e the locations in which there are preferred.

The marketing campaign should be based on targeted audience. This means, for instance, North America sold more of Action games, Campaigns done in NA should be more of Actions games, even though other genres could be campaigned for too.

Also, to improve number of sales / revenue, more marketing strategies could be developed, like Promo, etc.

Appendix

Google Sheets Data Set for Udemy Project

[\[https://docs.google.com/spreadsheets/d/1vC3c9bTcwbtR3_hFcnZ8jliHMLEjBLA-Bbevi68Ey2w/edit?usp=sharing\]](https://docs.google.com/spreadsheets/d/1vC3c9bTcwbtR3_hFcnZ8jliHMLEjBLA-Bbevi68Ey2w/edit?usp=sharing)

Tableau Data Visualization For Udemy Project

[\[https://public.tableau.com/views/UdemyTechProject/General-Dashboard?:language=en-US&:display_count=n&:origin=viz_share_link\]](https://public.tableau.com/views/UdemyTechProject/General-Dashboard?:language=en-US&:display_count=n&:origin=viz_share_link)

Google Sheets Data Set for [chosen] Project

[\[https://docs.google.com/spreadsheets/d/1iXw2vNBTtrNS1SPYeRu_rjZr8TUI9v1Jhj1_v7pJVd8/edit?usp=sharing\]](https://docs.google.com/spreadsheets/d/1iXw2vNBTtrNS1SPYeRu_rjZr8TUI9v1Jhj1_v7pJVd8/edit?usp=sharing)

Tableau Data Visualization For [chosen] Project

[\[https://public.tableau.com/views/EspadoVideoGameSales/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link\]](https://public.tableau.com/views/EspadoVideoGameSales/Dashboard1?:language=en-US&:display_count=n&:origin=viz_share_link)