DATA FASTING

Steam Store

Game Sales Report

College of Computing, Informatics and Media

DSC 651 - CS2306A

Prepared for :

Dr Zainura Idrus

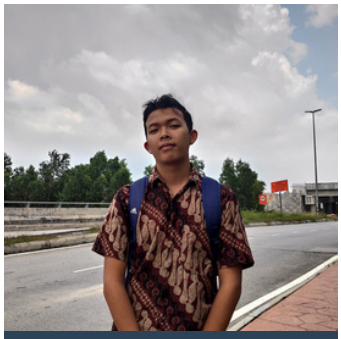
MEET OUR TEAM

Muhammad Akhmal Raziq Bin Arifin

2020819782

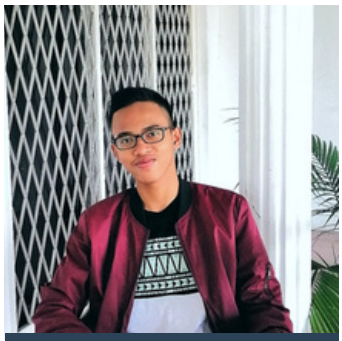
Muhammad Amirul Asyraf Bin Marzuki

2020898146



Mohamad Haniff Bin Abd Rahim

2020820016



Amirul Fahmi Bin Azman

2020601852

Mohamad Najib Bin Mohamad Aseri

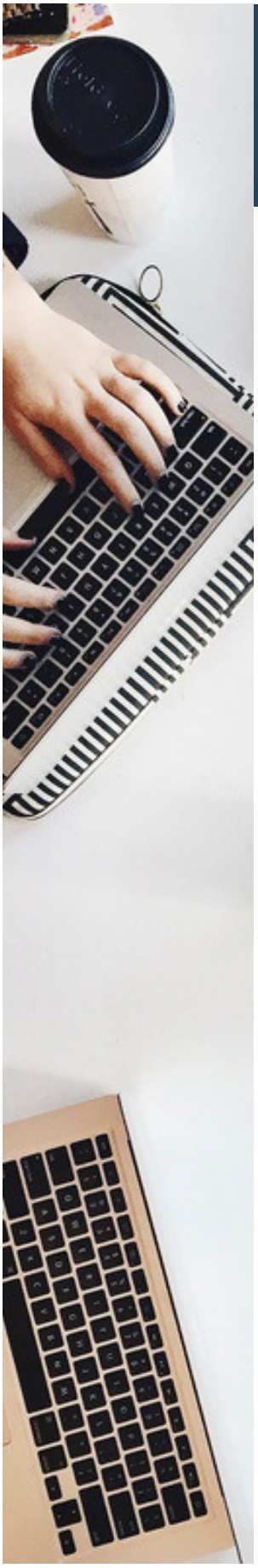
2020865578

We strive to deliver the best solutions for our stakeholders

CHAPTER 1

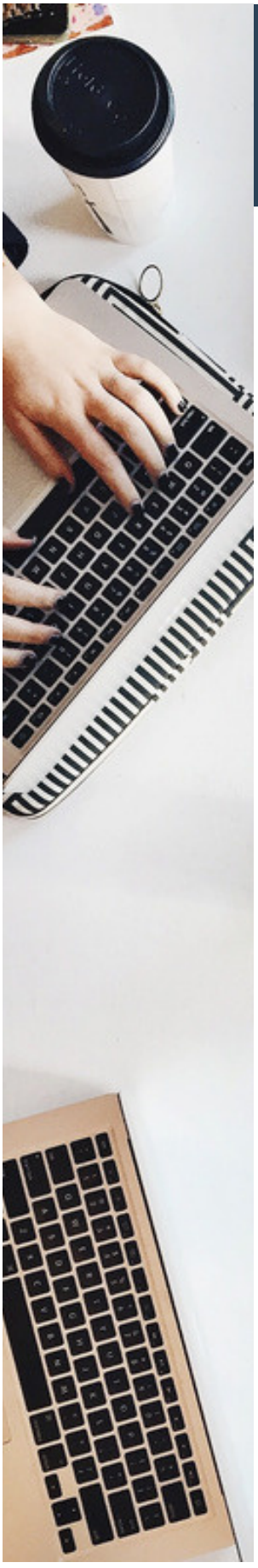
Planning

This chapter will discuss background of the study further.



INTRODUCTION

Steam is digital storefront, social networking platform, and gaming community all in one. It offers a user-friendly interface that allows users to browse and purchase games, organize their libraries, and interact with friends. It also provides a platform for developers to publish and promote their games.Steam aim is to offer gamers a one-stop-shop for purchasing and playing video games, as well as connecting with other players and developers.

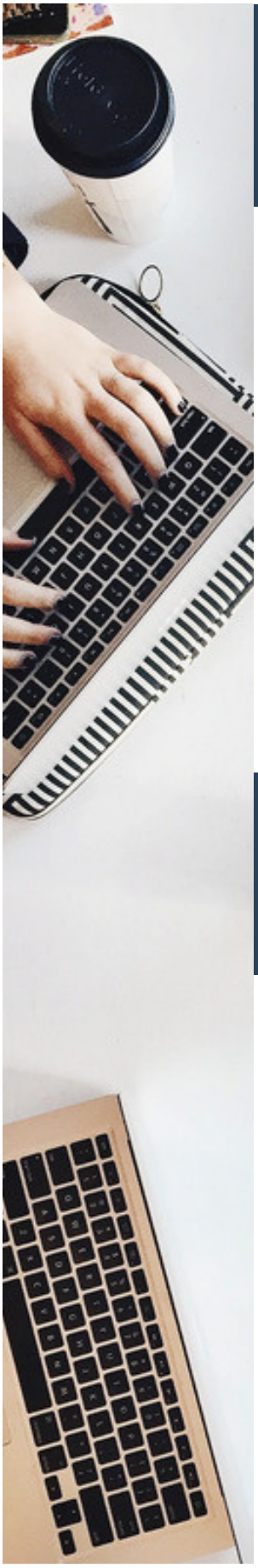


THE PROBLEMS

Gaming industry nowadays is rapidly evolving from times to times. New games being released regularly, and consumer preference always changing based on what are the newest trends.

Therefore, in order to focus on three main aspects which are first, game genres which explores the most popular game genres on Steam. Second, user ratings and reviews which dives into the impact of user ratings and reviews on game sales. Lastly, game sales and promotion where is game sales and promotion giving a significant impacts to the game sales,

User need an one stop gaming platform to visualize those aspects above in more understandable form



AIMS

This project aims on developing open source project in form of data visualization to help users and game developers by sharing the overview on game sales, game rating and other interesting factors that can be utilize by them.

OBJECTIVE

1 To collect and analyze the data of Steam Store video game sales

2 To design interactive dashboard for visualize the data of Steam Store game sales

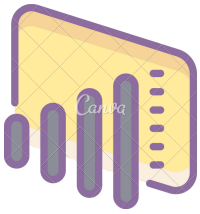
3 To obtain insight on the video game sales in Steam Store platform

SCOPE OF STUDY

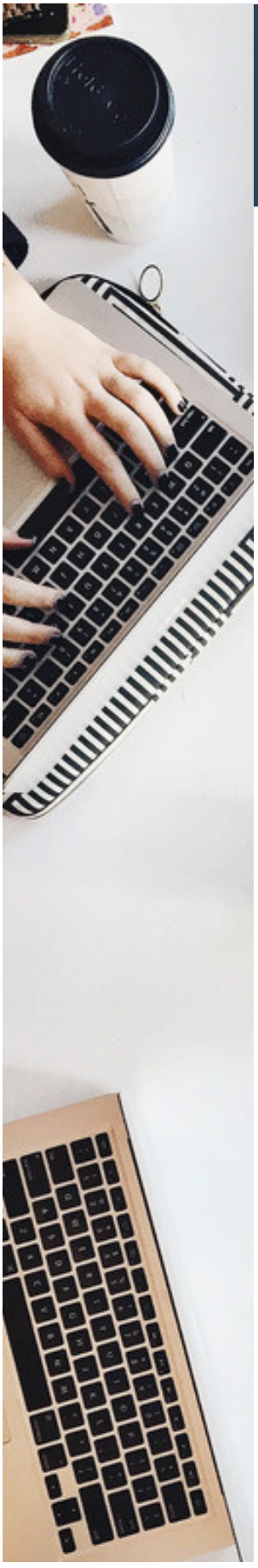
The data that will be used for this study is secondary data obtained from open source website Kaggle with the title of Steam Store Games.

The data contain video games releasing from 1997 until 2019 that was released on the Steam Store Platform and contain attributes such as video game price, categories, genre and estimation number of steam users that purchased the video game.

The software that will be used for developing interactive dashboard and visualizing the data is Microsoft Power BI Desktop.



This study will not covers up similar video game that were sold on different video game stores such as Epic Game Store and GOG.com store



SIGNIFICANT OF STUDY

The proposed project is expected to be beneficial toward user and video game developers. It will show the popular video games based on multiple factor for user to choose and play such as:

- reviews
- genres
- categories
- price

Also provide business strategy for video game developer such as what type of trending game that developers can produce for their next game

It will be convenient to create interactive dashboard that can be referred easily that by finding the information from multiple website or sources.

RESEARCH QUESTION

How do the price of games on the Steam Store impact the total of ownership?

How does game genre affect average playtime for specific genre, and are certain genres are less than others?

How do user ratings effect game sales success on the Steam Store?

Between free and paid games, which does player prefer to play?

Which platform has the most player?

CHAPTER 2

METHODOLOGY

In this chapter, the methods and techniques used during the study to achieve the expected outcome will be discussed.

SYSTEM DEVELOPMENT LIFE CYCLE

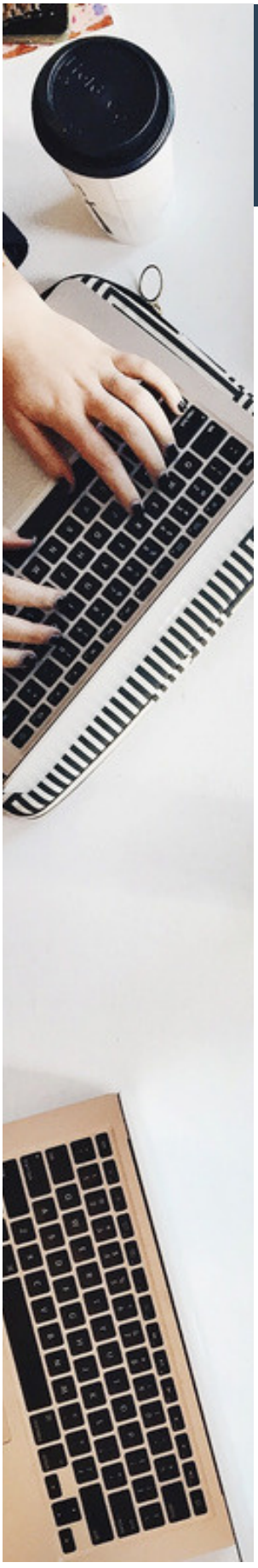
Plan what dataset to use

Collect dataset from Kaggle

Design the dashboard

Visualize the data using Power BI

Analyze and story-driven visualization.

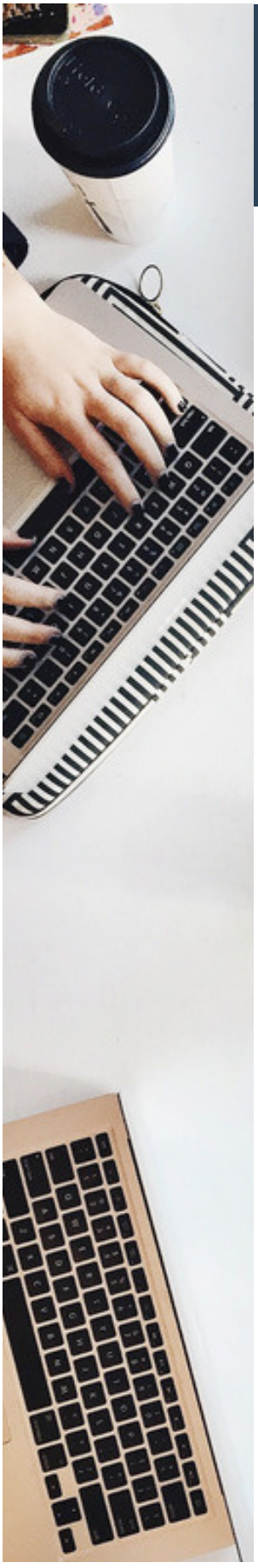


USER REQUIREMENTS

The dashboard should provide comprehensive insights into Steam Store game sales, including pricing impact analysis, genre-based playtime analysis, user rating and sales success, preference analysis, and platform player analysis.

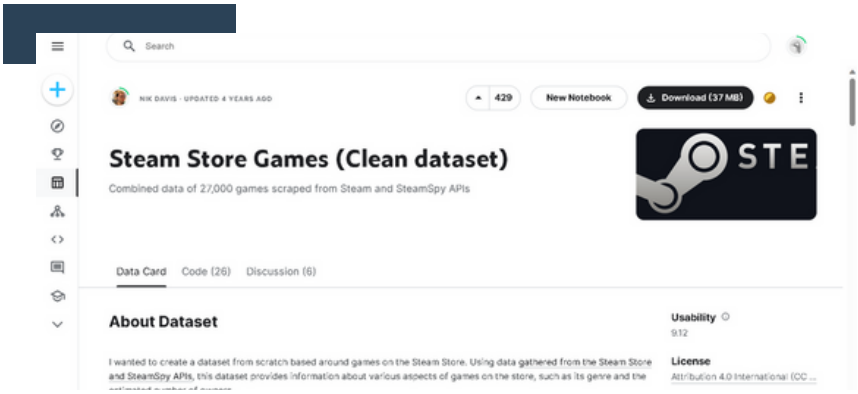
It should offer various visualizations such as line charts, bar charts, and pie charts to present the data in a clear and straightforward manner, enabling users and developers to make data-driven decisions.

The dashboard should deliver a user-friendly experience, ensuring that users can easily access and navigate the information, enabling them to receive insightful information
and make informed decisions based on the analysis of Steam Store game sales.



DATA COLLECTION

The dataset offers valuable insights into various aspects of Steam Store games, providing a rich resource for analysis and research purposes. It encompasses a wide range of attributes that can be explored to understand trends, preferences, and characteristics of games available on the platform.



The dataset contains information about games on the Steam Store, with 27,075 instances and 18 attributes. It was sourced from the Kaggle website and can be accessed at

Steam Store Games Dataset (Click Here to View)

DATA DESCRIPTION

NO.	ATTRIBUTES	DESCRIPTION
1.	appid	Application id. Unique identifier for each title
2.	name	Title of the game.
3.	release_date	Release date of the game.
4.	english	Language support: 1 if it's in English.
5.	developer	Name of developers. Semicolon delimited if there are multiple developers.
6.	publisher	Name of publishers. Semicolon delimited if there are multiple publishers.
7.	platforms	Type of supported platforms for the game.
8.	required_age	Minimum required age according to PEGI UK standards. Data with 0 are unrated or unsupplied.
9.	categories	Type of categories in the game.
10.	genres	Type of genres in the game.
11.	steamspy_tags	Type of genres in the game but based on community voted.
12.	achievements	Number of in-games achievements.
13.	positive_ratings	Number of positive ratings.
14.	negative_ratings	Number of negative ratings.
15.	average_playtime	Average user playtime (how many hours had been spent to play the game).
16.	median_playtime	Median user playtime (how many hours had been spent to play the game).
17.	owners	Range number of owners that install the game.
18.	price	Price of the game in pounds sterling.

DATA DICTIONARY

NO.	ATTRIBUTES	TYPE	DATA FORMAT
1.	appid	Integer	-
2.	name	Text	-
3.	release_date	Date	YYYY-MM-DD
4.	english	Integer	-
5.	developer	Text	-
6.	publisher	Text	-
7.	platforms	Text	-
8.	required_age	Integer	-
9.	categories	Text	-
10.	genres	Text	-
11.	steamspy_tags	Text	-
12.	achievements	Integer	-
13.	positive_ratings	Integer	-
14.	negative_ratings	Integer	-
15.	average_playtime	Integer	-
16.	median_playtime	Integer	-
17.	owners	Integer	-
18.	price	Double	-

DATA PREPARATION

The platform column is converted to binary representation (1 for available, 0 for not available) for simplified analysis.

The "required_age" column is transformed into nominal categories (e.g., "parental_advisor," "universal," "Adult") for easier grouping and analysis.

Low-frequency categories in certain columns are removed to streamline the dataset and reduce complexity.

Unnecessary columns such as "steamspy_tags" and "median_playtime" are eliminated to simplify the data.

The "price" column is transformed into ordinal categories (e.g., "free," "low," "medium," "high") for easier comparison and analysis.

The "median_playtime" column is deleted, and the "average_playtime" column is used instead. This change allows for a more representative measure of playtime, avoiding potential biases introduced by outliers.

Video Game with "date of released" outside of year 2018 are removed.

DATA PREPARATION

Before Preprocessing

appid	name	release_date	english	developer	publisher	platforms	required_age	categories	genres	steamspy_	achieveme	positive_rat	neg
10	Counter-Strike	11/1/2000	1	Valve	Valve	windows;mac;lin	0	Multi-player;Onlir	Action	Action;FPS	0	124534	
20	Team Fortress Classic	4/1/1999	1	Valve	Valve	windows;mac;lin	0	Multi-player;Onlir	Action	Action;FPS	0	3318	
30	Day of Defeat	5/1/2003	1	Valve	Valve	windows;mac;lin	0	Multi-player;Valv	Action	FPS;World	0	3416	
40	Deathmatch Classic	6/1/2001	1	Valve	Valve	windows;mac;lin	0	Multi-player;Onlir	Action	Action;FPS	0	1273	
50	Half-Life: Opposing Fc	11/1/1999	1	Gearbox Softwa	Valve	windows;mac;lin	0	Single-player;Mult	Action	FPS;Action	0	5250	
60	Ricochet	11/1/2000	1	Valve	Valve	windows;mac;lin	0	Multi-player;Onlir	Action	Action;FPS	0	2758	
70	Half-Life	11/8/1998	1	Valve	Valve	windows;mac;lin	0	Single-player;Mult	Action	FPS;Classic	0	27755	
80	Counter-Strike: Condi	3/1/2004	1	Valve	Valve	windows;mac;lin	0	Single-player;Mult	Action	Action;FPS	0	12120	
130	Half-Life: Blue Shift	6/1/2001	1	Gearbox Softwa	Valve	windows;mac;lin	0	Single-player	Action	FPS;Action	0	3822	
220	Half-Life 2	11/16/2004	1	Valve	Valve	windows;mac;lin	0	Single-player;Stea	Action	FPS;Action	33	67902	
240	Counter-Strike: Sourc	11/1/2004	1	Valve	Valve	windows;mac;lin	0	Multi-player;Cros	Action	Action;FPS	147	76640	
280	Half-Life: Source	6/1/2004	1	Valve	Valve	windows;mac;lin	0	Single-player	Action	FPS;Action	0	3767	
300	Day of Defeat: Source	7/12/2010	1	Valve	Valve	windows;mac;lin	0	Multi-player;Cros	Action	FPS;World	54	10489	

After Preprocessing

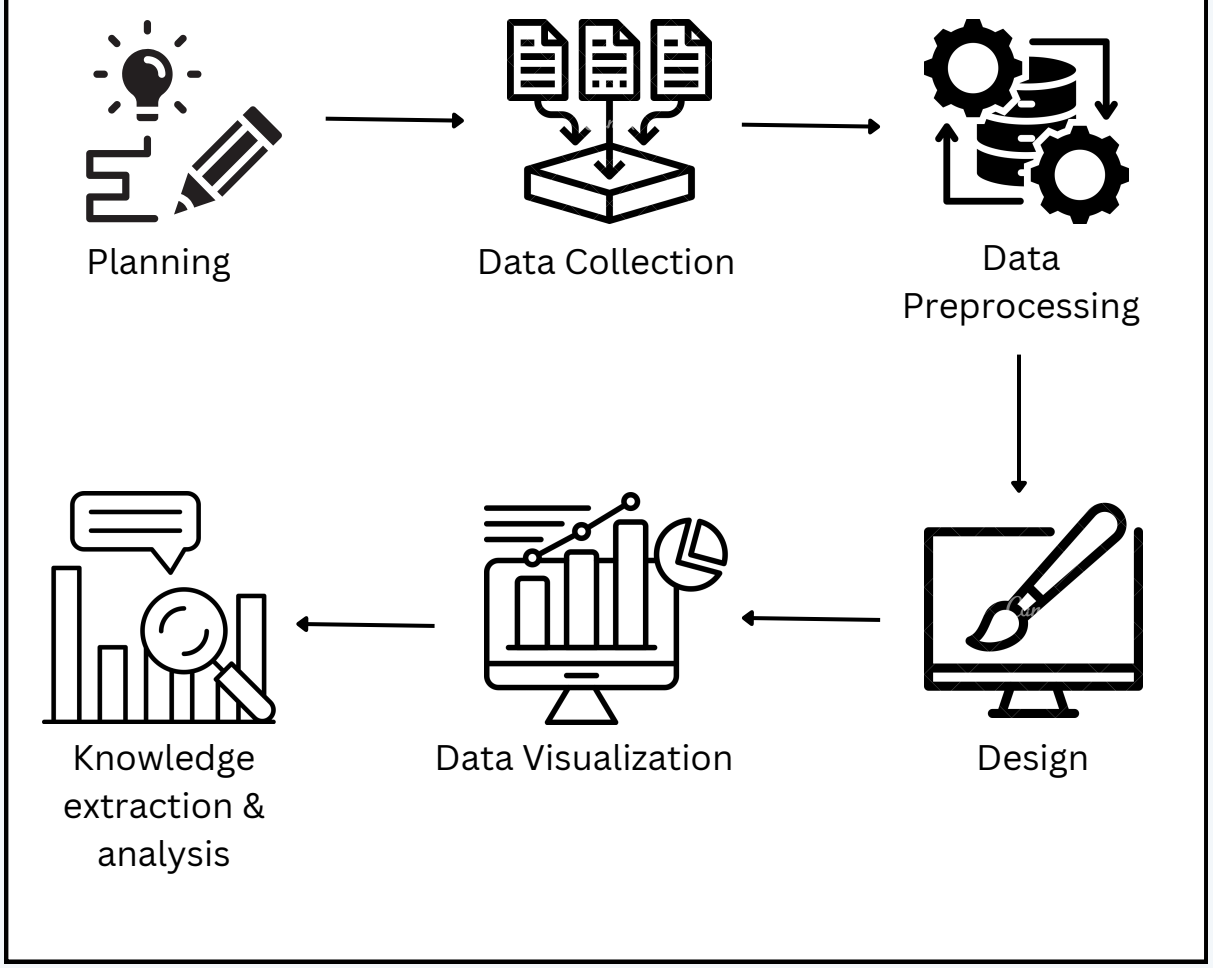
appid	name	release_date	english	developer	publisher	platforms_	platform_	platform_	required_age	categories	genres	sub_genres	achieve
684210	7	1/26/2018	1	Zac	Zac	1	0	0	0	Single-player	Adventure	Free to Play	
938520	21	9/28/2018	1	Softwaves Mini	Softwaves Dist	1	0	0	0	Single-player	Adventure	Casual	
913850	103	12/21/2018	1	Dystopia Interac	Dystopia Inter	1	1	1	0	Single-player	Adventure	Casual	
696860	900	11/30/2018	1	Barnaque	Barnaque	1	1	1	0	Single-player	Action	Indie	
876840	999	6/29/2018	1	Tydecon Games	Tydecon Game	1	0	0	0	Single-player	Casual	Indie	1
942050	2048	10/31/2018	1	Lucas Alexander	Lucas Alexand	1	0	1	0	Single-player	Casual	Indie	
987850	2084	12/13/2018	1	Feardemic	Feardemic	1	0	1	0	Single-player	Action	Indie	
952250	Sep-99	10/10/2018	1	98DEMAKE	98DEMAKE	1	1	1	0	Single-player	Indie	Indie	
866510	!AnyWay!	6/6/2018	1	GQG DEVELOPEF	GQG PUBLISHE	1	0	1	0	Single-player	Adventure	N/A	49
870990	!LABrpgUP!	6/13/2018	1	UPandQ	UPandQ	1	0	1	0	Single-player	Adventure	Casual	20
782100	#CuteSnake	3/15/2018	1	Sunrise9	Sunrise9	1	0	1	0	Single-player	Casual	Indie	1
834180	#CuteSnake 2	4/20/2018	1	Sunrise9	Sunrise9	1	0	1	0	Single-player	Casual	Indie	1
841860	#Have A Stick	4/26/2018	1	Red Dolphin Gan	VT Publishing	1	0	1	0	Single-player	Casual	Strategy	50
779420	#WarGames	3/14/2018	1	Eko	Eko	1	1	1	0	Single-player	Adventure	Indie	

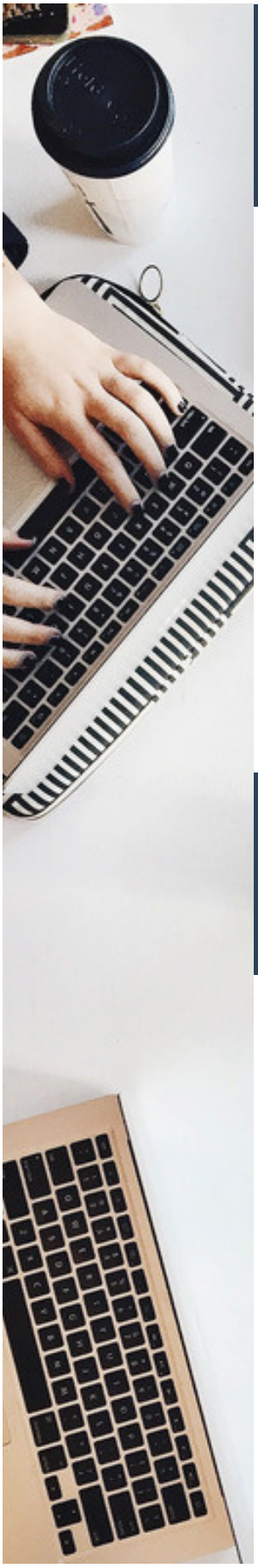
CHAPTER 3

DESIGN &DEVELOPMENT

This chapter involves the creation and implementation of structured sections that outline the process and progress of a project.

SYSTEM ARCHITECTURE





HARDWARE

- Processor: 1.6 GHz or higher
- RAM: Minimum 4 GB (8 GB recommended)
- Disk Space: At least 10 GB free
- Display Resolution: Minimum 1024x768

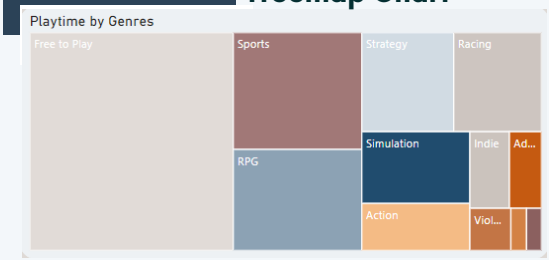
SOFTWARE

- Operating System: Windows 10
- Excel : Data Cleaning
- Power BI Desktop: Install for creating reports and visualizations

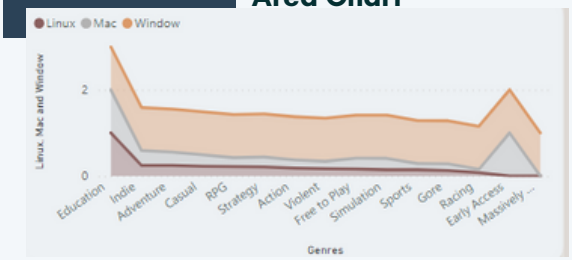
GRAPH DESIGN

A treemap chart is a visual representation that displays hierarchical data using nested rectangles, with each rectangle's size indicating a specific metric. The larger the rectangle, the higher the cumulative playtime for games within that genre, allowing for a quick visual comparison of playtime distribution among different genres.

Treemap Chart



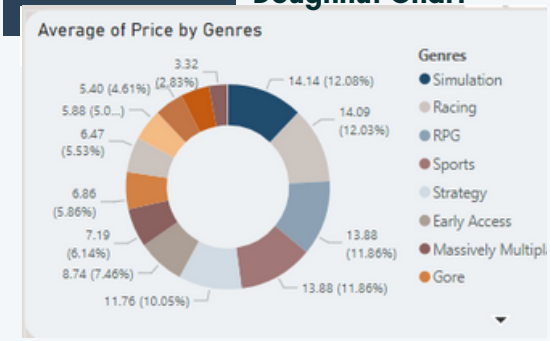
Area Chart



The area chart depicts the distribution of game genres across different operating system (OS) platforms. It visually highlights the relative popularity or prevalence of genres within each OS platform, enabling quick insights into the genre preferences of users on different operating systems.

The donut chart showcases the distribution of average prices across different game genres, providing insights into pricing trends. It visualizes the varying average prices associated with each genre, allowing for easy comparison and identification of genres that tend to have higher or lower average prices.

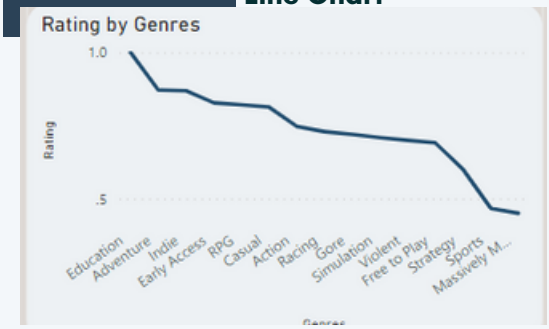
Doughnut Chart



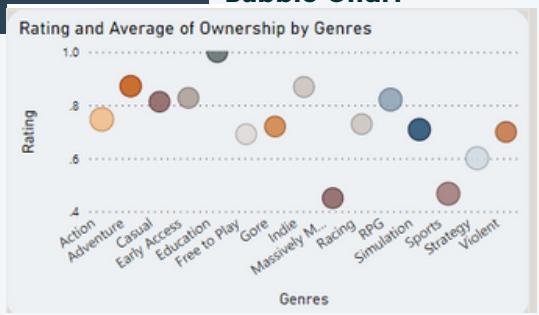
GRAPH DESIGN

The line chart shows how game ratings vary across different genres, providing a visual comparison of which genres tend to have higher or lower ratings. It enables the identification of trends and patterns, indicating which genres tend to have higher or lower ratings and how they compare to one another.

Line Chart



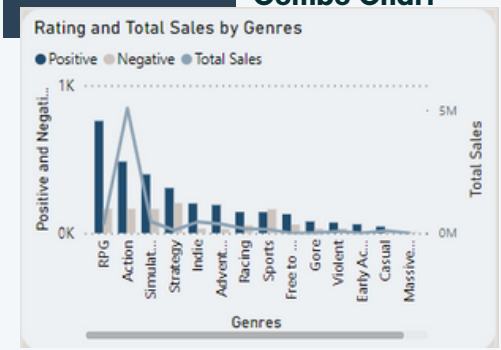
Bubble Chart



The bubble chart visually depicts the relationship between game genres, their ratings, and average ownership, allowing for quick identification of popular genres based on bubble size and position.

The combo chart combines multiple types of charts, such as bar and line chart. It provides a comprehensive visual display, enabling the comparison and analysis of these metrics for different genres in a single chart.

Combo Chart



GRAPH DESIGN

Slicers are interactive filters that allow users to select specific game names, enabling focused analysis and exploration of data related to those specific games

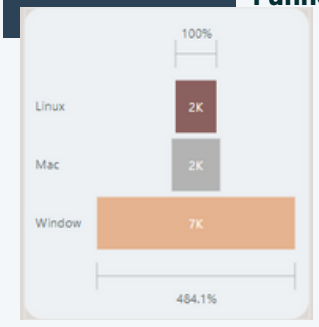
Slicer

Column Chart

The column chart displays how the average ownership of games varies across different price ranges, providing a visual representation of the distribution of ownership within each price category.

The funnel chart shows the number of games for each operating system platform (Windows, Mac, and Linux), providing a simplified representation of the distribution of games across the platforms.

Funnel Chart



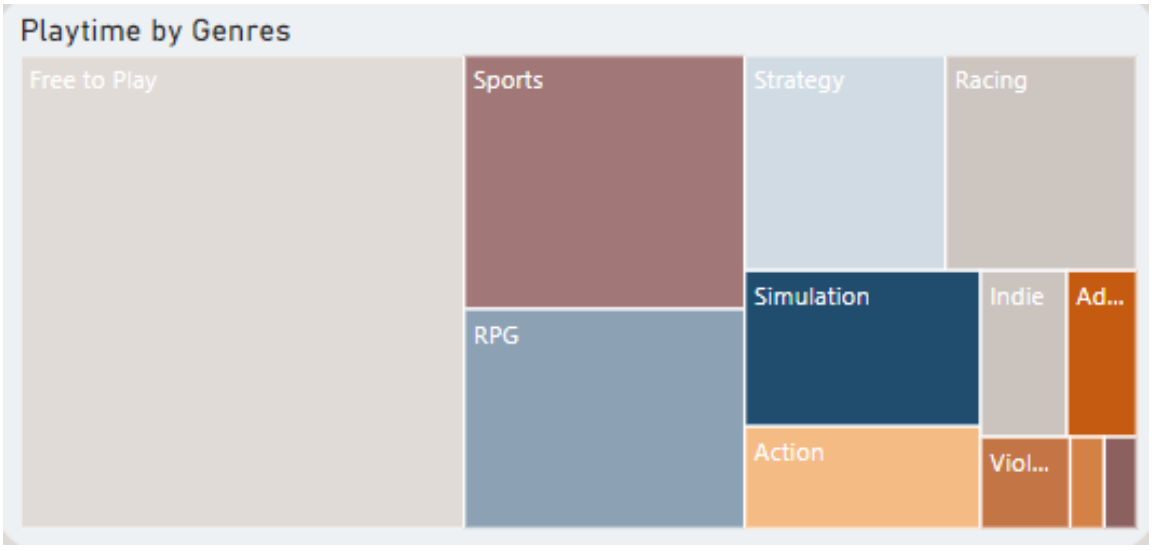
CHAPTER 4

ANALYSIS & FINDINGS

This chapter involves the results and analysis of the project

FINDING & ANALYSIS

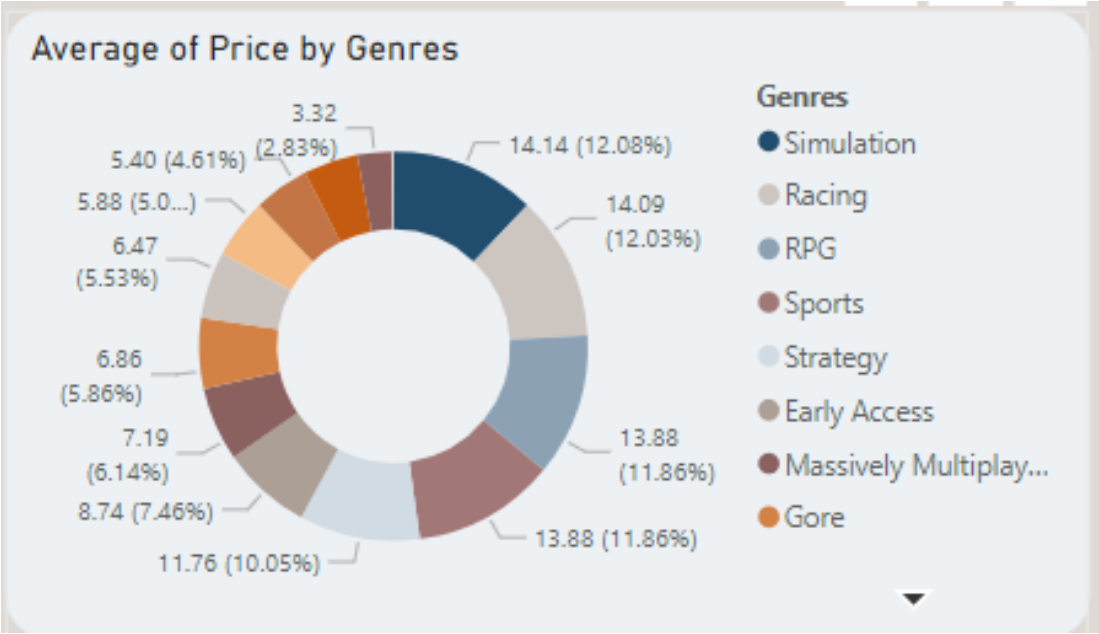
This treemap chart visualizes the playtime of games based on the game's genre. This will assist game developer to determine which genre that are most played by user and will be the targeted genre for their future games. The chart shows that free to play game has the most playtime. Hence, game developer can develop more free games for user and generate income using other platform such as advertising or in game purchases. Not just that, sports and RPG games also have high playtime by users. Hence, developer also can focus on those genres in their future games.



In addition, the chart uses different colors of rectangle boxes to indicate different genres. This will help viewer differentiate the genres

FINDING & ANALYSIS

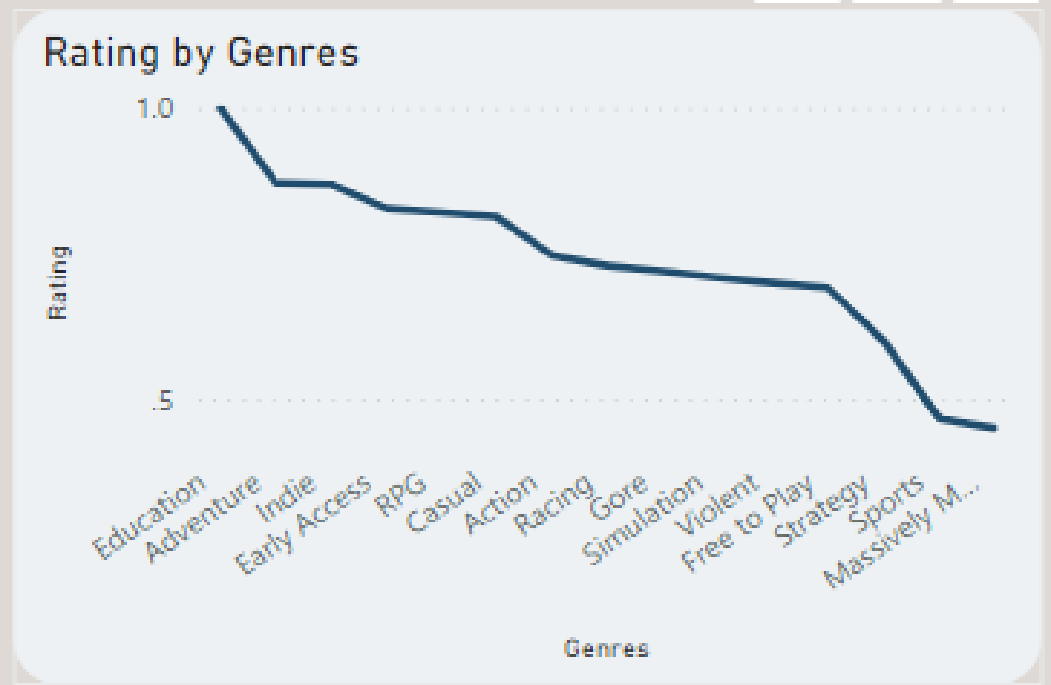
This doughnut chart visualizes average price of games in each genres. This visualization can assists gamers to choose what game to play based on their budgets. This also can gives insight to customer if they have targeted particular game to buy, the average price can gives early information and gives customer time to prepare the budget. For example, if there is a new sports game will be released, customer can prepare budget according to the average price of sports game in the doughnut chart.



In addition, the chart uses colors same with the treemap chart for the genres so that viewer did not confuse with the colors.

FINDING & ANALYSIS

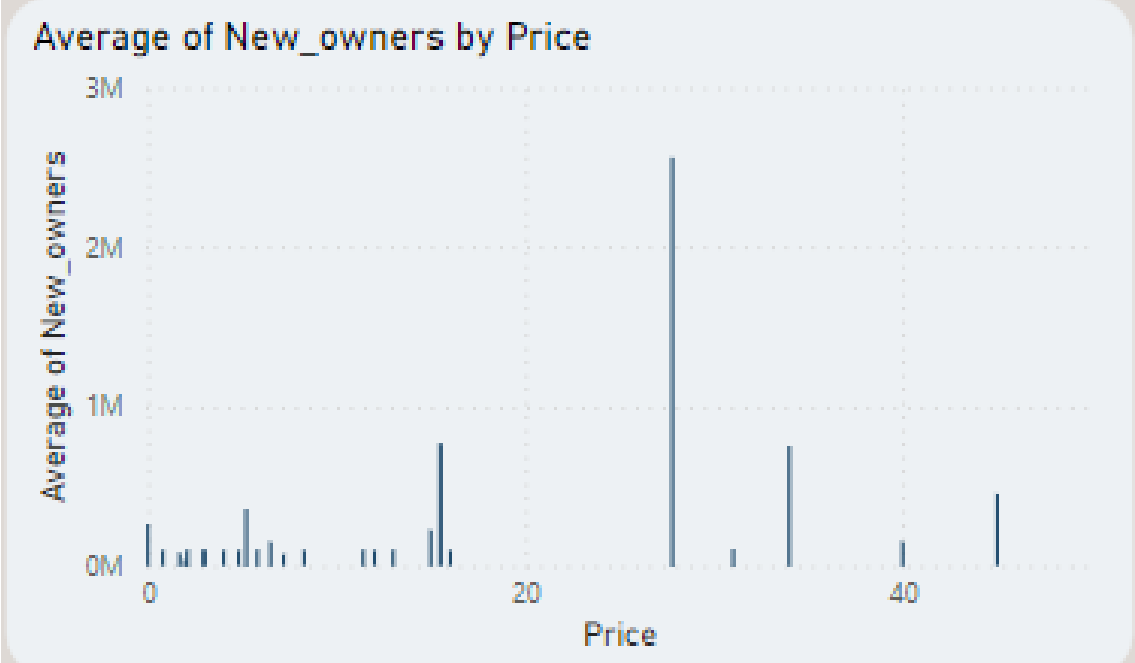
This line chart visualize the rating of the game based on the video games genre. The chart show the trend of the rating among the genres. Based on the graph, education reach the highest rating by 1.0 and massively multiplayer reach the lowest rating by 0.45. Hence, the education genres receive great feedback from the gamers which the games is good, or satisfying. Customer can create a perception in which games can reach trend in the future avoiding flop and churn for a new games development.



In addition, the chart brought a future perspective of the current trend to enhanced knowledge understanding of the gaming researcher.

FINDING & ANALYSIS

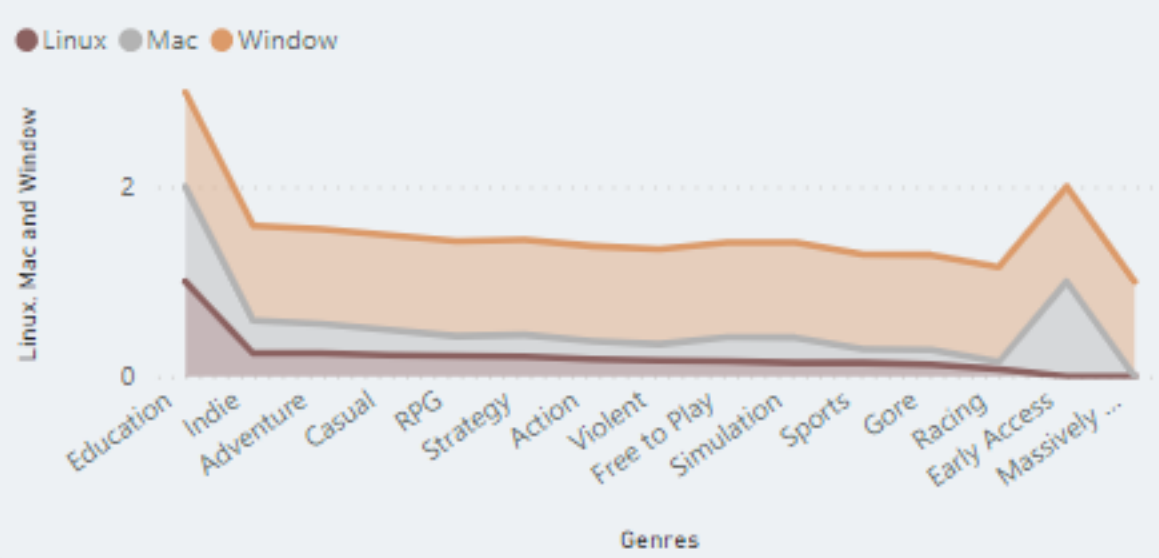
This columnar chart visualize the average ownership by it price. The customer can visualize the data changes of the ownership by the price from free to the maximum price of 49.99. They can determine the optimal price range for the game to buys. Customer with a budget price, can create an optional games to choose. Thus, the distribution of the chart help the customer to determine the perfect get a go game to fit their budget. As for the developer, they could cluster the best range of price to their game that competent the customer budget value while maintain the interest.



In addition, the chart show different data that may spike or drop base on the distribution. Larger data is perfect to stimulate the cluster trend.

FINDING & ANALYSIS

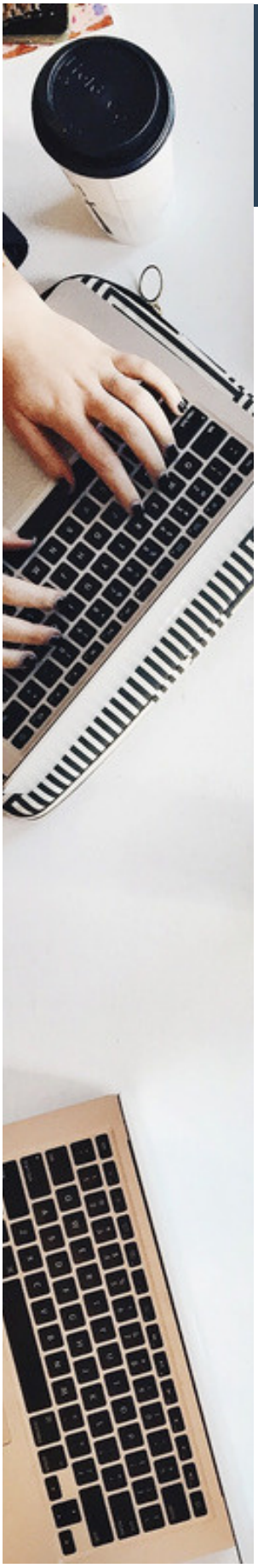
This area chart visualize the average platform of the video game been made. Based on the chart, windows have a linear line sitting at the average of 1.0. It tell that every games used windows as their main platform. Other example, in genre action, linux show 0.18, and mac show 0.19 on average. Both platform lack the compatibility of the software usage toward the game genre. Certain game required high-end hardware specification, in which window software is the suitable for it. Hence, the customer can insight the best platform for their games. This insight can manage customer development cycle so that they have a targeted platform. A goals in mind, can create a well revised budget of the upcoming platform launch.



CHAPTER 5

CONCLUSION

This chapter involves the results and analysis of the project



CONCLUSION

In conclusion, this study utilizes data visualization to delve into the world of Steam Store game sales. By analyzing various factors such as pricing impact, genre-based playtime, user ratings, sales success, preference analysis, and platform player analysis, the dashboard provides valuable insights for game developers and stakeholders. Through this project, developers gain a better understanding of the types of games that are likely to succeed on the Steam Store, considering elements like genre, category, and user demographics.

The data visualization capabilities offered by the project, particularly using tools like Microsoft Power BI, prove to be an invaluable resource for developers in gaining insights and enhancing their game applications before their release.

DATA FASTING

Thank You

College of Computing, Informatics and Media

DSC 651 - CS2306A

Presentation Slide (Click Here)

Dashboard (Click Here)