Video Games Trend Analysis (Recommendation System)

BDM-2053 Project Report

Submitted to
Professor Dhruwal Shah

Submitted by Group 12
Auradee Castro
Bhumika Rajendra Babu
Miraj Sinya
Olivia Deguit
Roger Mais

Submitted on
April 20, 2023

TABLE OF CONTENTS

I. PROJECT SUMMARY

The client wanted to enter the video game industry to meet the growing demand by creating a recommendation system for video games based on user preferences. Additionally, they aimed to understand the market trends to make informed business decisions.

This project made use of a publicly accessible dataset from Kaggle that includes information about video games, such as sales (sold in units), genre, release year, platform, critic score, and user score. The dataset contains video games that were released over a 40-year period, from 1980 to 2020, which was verified for accuracy by cross-referencing it with VGChartz. The goal of this recommendation system is to provide video game recommendations to users who are looking to discover new games.

Deliverables to the client are as follows:

- **Recommendation Model** that provides suggestions for the video game titles to users based on the game title and platform *(optional)* as input [1]
- **Market Analysis Report** that presents an analysis of the Video Game Industry trend using available data up to the year 2020 [2]

II. PROJECT TEAM

The project team was composed of highly skilled and experienced professionals from various disciplines. Each member of the team brought their unique expertise to the project, contributing to its overall success.

Team members are listed below, along with their respective roles and contact information:

Name	Role	Email
Auradee Castro	Project Lead / Data Analyst	c0866821@mylambton.ca
Bhumika Rajendra Babu	Data Analyst	c0867081@mylambton.ca
Miraj Sinya	Business Analyst	c0863371@mylambton.ca
Olivia Deguit	Business Analyst	c0878491@mylambton.ca
Roger Mais	Data Engineer	c0863147@mylambton.ca

[1] Appendix E: Recommendation Model (Python notebook)
[2] Appendix D: Market Analysis report in Power BI

The project was completed successfully by the team using a standard project management methodology, where detailed planning and effective communication were essential factors contributing to the project's success.

III. PROJECT SCHEDULE

The project plan for the Video Games Trend Analysis (Recommendation System) entailed a 5-week timeline, starting from March 16, 2023, and concluding on April 20, 2023. The project was divided into several milestones that represented distinct stages in the project lifecycle, indicating both the planned schedule dates and the actual completion dates for each phase.

Project Milestone	Planned Completion	Actual Completion	Remarks
Project Details and Design	March 20, 2023	March 20, 2023	Completed on time
Data Acquisition	March 23, 2023	March 23, 2023	Completed on time
Data Analysis and Visualization	March 30, 2023	April 10, 2023	Completed behind planned schedule
Model Creation and Evaluation	April 6, 2023	April 17, 2023	Completed behind planned schedule
Project Presentation and Report	April 20, 2023	April 20, 2023	Completed on time

The reason for the modification of the actual completion dates for the "data analysis and visualization" and "model creation and evaluation" milestones was due to the project presentation deadline being pushed back by one week. This adjustment provided the team with more time to enhance the analysis and preprocessing of the data, which is the most vital phase of the project as it covers both the market analysis report and the pre-requisites of the recommendation model.

Moreover, extra time was allotted to the creation and evaluation of recommendation models as the team decided to build two models utilizing both supervised and unsupervised machine learning algorithms in order to properly evaluate their performance, and determine which model is more efficient for the users.

IV. PROJECT MILESTONES AND DELIVERABLES

The analysis on market trend in video games and the development of a recommendation system has been carried out successfully. The table below provides an overview of the deliverables for each milestone and their current status, which shows that the project has been completed to a high standard, and that the deliverables meet the requirements.

Milestone 1: Project Details and Design
Purpose: To prepare a document that outlines the project's specifics, approach, and scope by identifying the business problems and breaking them down into essential questions **Status:** Completed **Deliverables:** Project proposal containing the project's objectives, scope along with expected deliverables as well as project milestones and the estimated timeline [3]
Milestone 2: Data Acquisition
Purpose: To obtain a relevant dataset that aligns with the identified business problems and supports the project's objectives going forward **Status:** Completed **Deliverables:** Video Games Sales dataset sourced from Kaggle, which contains video games that were released between the years 1980 and 2020. The dataset is comprised of 16,719 records and 16 features. [4]
Milestone 3: Data Analysis and Visualization
Purpose: To conduct comprehensive data analysis and data preprocessing a.) to identify patterns, trends and relationships within data, which will provide support for business decision-making, and b.) to eliminate any missing values and convert the raw data into a more usable format **Status:** Completed **Deliverables:** - Descriptive statistics to understand dataset, and graphical representations of data distribution and correlations of all relevant features by leveraging Python's *matplotlib* and *seaborn* packages - Clean and preprocessed data as pre-requisite for model development

[3] Appendix A: Project Proposal
[4] Appendix C: Dataset Information

- Missing data-imputation on critic scores and user scores, and removal of missing values for other features by leveraging Python's *numpy* and *pandas* packages
- Conversion of categorical features to dummy indicators by leveraging Python's *pandas* package
- Transformation of categorical features to a standardized form by leveraging Python's *scikit-learn* package
- Records were reduced from 16,719 to 9,950 after data preprocessing

- ***Market analysis report*** of the overall sales based on the game's release year (1980-2020), and the top performing game genres and platforms using Power BI [5]

Milestone 4: Model Creation and Evaluation

Purpose: To develop recommendation model using the most appropriate algorithm, then assess its user efficiency

Status: Completed

Deliverables:

- ***Recommendation model*** using the nearest neighbor algorithm, which was designed to suggest the top ten video game titles based on user input, i.e., the game title and platform *(optional)*. To evaluate the effectiveness of the algorithm, it was assessed in terms of user efficiency of the results. [6]
- Program UI prototype to provide visual representation of how the model would work when integrated into an application, to ensure that it aligned with the client's expectations [7]

Milestone 5: Project Presentation and Report

Purpose: To present to the client the project's deliverables

Status: Completed

Deliverables: Project presentation and a final report highlighting important details and outcomes of the project, which involves the goals, scope, approach, and schedule of the project along with the expected deliverables [8][9]

[5] Appendix D: Market Analysis Report (Power BI)
[6] Appendix E: Recommendation Model (Python Notebook)
[7] Appendix F: Program UI Prototype
[8] Appendix G: Project Presentation
[9] Appendix H: Project Final Report

To sum up, the project team has successfully fulfilled all the documented project deliverables. The highlighted ones are the expected deliverables to the client, i.e., the Market Analysis Report and the Recommendation Model. This shows the project team's ability to meet the project's goals and objectives while ensuring that client's expectations were met.

V. PROJECT PERFORMANCE EVALUATION

The performance evaluation section of this report presents a thorough examination of the completed project, highlighting the significant factors that led to its overall success, as well as the difficulties encountered and the solutions implemented. This serves as a valuable tool for understanding the project's strengths and weaknesses, while offering actionable insights to enhance future project management practices and ensure continuous growth and development.

Key Achievements

- Successfully executed data loading and transformation using Python for data analysis
- Conducted a comprehensive Exploratory Data Analysis (EDA) on the data using Python
    - Descriptive statistics were generated to comprehend the data's range and distribution
    - Missing values were handled by imputing for critic and user scores, and removing them for other features
    - Data distribution for key features was visually represented using histogram, while scatter plot was used to show the correlation between critic and user scores
- Developed two models using different algorithms in Python; their user efficiency were evaluated successfully to determine the more suitable model for the recommendation system
- Utilized Power BI for video games market analysis, focusing on overall sales based on the game's release year (1980-2020), and the top performing game genres and platforms
- Proactively identified and addressed all potential risks and blockers in the project by performing preventive measures to avoid potential issues
- Fostered excellent team collaboration by maintaining open communication, sharing knowledge, and leveraging expertise to complete project more efficiently

Roadblocks and Remediation Plans

- A comprehensive research and analysis of online datasets were conducted to tackle the problem of insufficient data for video games analysis. The most appropriate algorithm was then selected to fit the available dataset.
- Challenge of choosing the most appropriate algorithm for the recommendation model was resolved by evaluating supervised and unsupervised models, which were compared based on their user efficiency to determine the most suitable algorithm for the recommendation model.

VI. PROJECT COSTS

The Video Games Trend Analysis (Recommendation System) project was assigned a budgeted cost of **CAD 31,400**. The cost was allocated based on project milestones, as shown in the table below, which compares actual costs with budgeted costs.

The budget for this project was not included in the initial project proposal. It was discussed during the planning stage after proposal submission. The allocated budget was finalized based on a detailed analysis of project requirements and was approved by the client before implementation.

Project Milestone	Budgeted Cost	Allocation %	Actual Cost	Utilization %
Project Details and Design	CAD 3,500	11.15%	CAD 3,350	10.93%
Data Acquisition	CAD 600	1.91%	CAD 600	1.96%
Data Analysis and Visualization	CAD 13,500	42.99%	CAD 13,100	42.74%
Model Creation and Evaluation	CAD 12,550	39.97%	CAD 12,350	40.29%
Project Presentation and Report	CAD 1,250	3.98%	CAD 1,250	4.08%
Total	**CAD 31,400**	**100%**	**CAD 30,650**	**100%**

Note: Budget utilization rate is at 97.61%. A positive budget variance of 2.39%

The actual overall expenses for the project totaled **CAD 30,650**, which means that the project not only achieved all its objectives and deliverables, but also completed within the allocated budget. [10]

VII. RECOMMENDATIONS

The Video Game Trend Analysis (Recommendation System) was a well-planned and effectively executed project that was designed to assist game enthusiasts in discovering new games. It also aims to provide quick insights on video gaming trends for businesses that are venturing into the gaming industry. Nevertheless, there were some recommendations and lessons learned from the project.

[10] Appendix I: Project Cost Breakdown

Recommendation 1:
To make predictions about future video game sales, it is recommended to use a comprehensive dataset that includes video game sales for each year and region. It is important to note that the current dataset is limited to the release year of each game. By obtaining a complete sales dataset, a predictive model can be created to forecast sales figures for future releases. This will provide valuable insights for developers and publishers in the gaming industry.

Recommendation 2:
Incorporating user data into video game recommendation system can lead to more personalized and accurate recommendations. This can improve user engagement, satisfaction, and retention. By analyzing user feedback, behavior, and demographics, recommendations can be tailored to the user's specific interests and playstyle.

Recommendation 3:
Using web scraping is recommended to collect data and fill in missing information. The model should include games played on smartphones, as they have a large player base, resulting in more precise recommendations for a wider audience. Furthermore, improving the dataset's quality and reducing the amount of missing data will increase the model's accuracy and effectiveness.

VIII. APPENDICES

Appendix	Item	Location
A	Project Proposal	File available on link
B	Project Report Status	File available on link
C	Dataset Information	Kaggle dataset link
D	Market Analysis Report (Power BI) - Regional Sales - Platform - Genre	Power BI Service link
E	Recommendation Model (Python Notebook) - Data analysis and data preprocessing - Model development and evaluation - Model implementation	Code available on link
F	Program UI Prototype	Images available on link
G	Project Presentation	File available on link
H	Final Project Report	File available on link
I	Project Cost Breakdown Project Details and Design: Expenses related to project planning, project management, and coordination, as well as the costs for personnel, travel, and project management software Data Acquisition: Costs for data collection and data preprocessing Data Analysis and Visualization: Expenses for data analysis software, data visualization tools, and personnel with expertise in data analysis and visualization	

	Model Creation and Evaluation: Expenses related to the creation of models, such as software or hardware, as well as personnel with expertise in modeling and evaluation Project Presentation and Report: Expenses for creating presentations, and travel expenses related to presenting the project to client	