C964: Computer Science Capstone – Video Game Sales Forecast

[Project Proposal for Business Executives 3](#_Toc142927076)

[Letter of Transmittal 3](#_Toc142927077)

[Project Recommendation 4](#_Toc142927078)

[Problem Summary 4](#_Toc142927079)

[Application Benefits 4](#_Toc142927080)

[Application Description 4](#_Toc142927081)

[Data Description 4](#_Toc142927082)

[Objectives and Hypothesis 4](#_Toc142927083)

[Methodology 4](#_Toc142927084)

[Funding Requirements 5](#_Toc142927085)

[Data Precautions 5](#_Toc142927086)

[Developer’s Expertise 5](#_Toc142927087)

[Project Proposal 6](#_Toc142927088)

[Problem Statement 6](#_Toc142927089)

[Customer Summary 6](#_Toc142927090)

[Existing System Analysis 6](#_Toc142927091)

[Data 7](#_Toc142927092)

[Project Methodology 7](#_Toc142927093)

[Project Outcomes 8](#_Toc142927094)

[Implementation Plan 8](#_Toc142927095)

[Evaluation Plan 9](#_Toc142927096)

[Resources and Costs 10](#_Toc142927097)

[Timeline and Milestones 11](#_Toc142927098)

[Post-implementation Report 12](#_Toc142927099)

[A Business (or Organization) Vision 12](#_Toc142927100)

[Datasets 12](#_Toc142927101)

[Data Product Code 12](#_Toc142927102)

[Objective (or Hypothesis) Verification 13](#_Toc142927103)

[Effective Visualization and Reporting 13](#_Toc142927104)

[Accuracy Analysis 14](#_Toc142927105)

[Application Testing 15](#_Toc142927106)

[Application Files 15](#_Toc142927107)

[User Guide 16](#_Toc142927108)

[Summation of Learning Experience 17](#_Toc142927109)

# Project Proposal for Business Executives

## Letter of Transmittal

8/14/2023

Jon Doe, CEO

Jon Doe Video Games

123 Main Street

Orlando, FL

Dear Mr. Jon Doe,

I hope this letter finds you in good health and high spirits. I am writing to bring your attention to a critical challenge our organization is facing and propose an innovative solution that can significantly enhance our strategic decision-making capabilities.

In the increasingly competitive video game industry, understanding and predicting sales by genre is paramount. Our current approach relies on manual analysis and traditional methods, leading to delayed and sometimes inaccurate forecasts. This issue hampers our ability to respond to market trends swiftly and optimize our marketing efforts.

I propose the deployment of a specialized data product that we have developed to predict future video game sales by genre. This application utilizes linear regression and machine learning techniques to analyze historical sales data and provide precise predictions. Customized to our unique needs, this tool will enable us to make informed decisions, aligning our strategies with real-time market dynamics.

The implementation of this data product offers invaluable advantages:

* **Real-Time Insights**: Instant access to sales forecasts will empower us to make timely decisions.
* **Strategic Alignment**: Targeted predictions by genre will enable us to allocate resources more effectively.
* **Competitive Edge**: Leveraging data-driven insights will position us ahead of competitors in identifying and capitalizing on emerging trends.

The estimated total investment for this project, including development, integration, training, and support, stands at $90,000. This investment is justified by the potential revenue growth and efficiency gains we expect to achieve.

With my extensive background in data science and hands-on experience in creating predictive models tailored to the video game industry, I am confident in my ability to lead this project to success. My team and I have successfully implemented similar solutions in the past, and we understand the unique dynamics of our market.

I would be delighted to discuss this proposal further with you and provide any additional information you may require. Thank you for considering this opportunity, and I look forward to your favorable response.

Sincerely,

Curtis

## Project Recommendation

### Problem Summary

The global video game industry is experiencing rapid change, driven by evolving consumer preferences and emerging trends. Our organization struggles to accurately predict sales by genre, leading to missed opportunities and suboptimal resource allocation. This application will solve the critical challenge of forecasting video game sales by genre, providing timely insights that align with our organization's strategic goals.

### Application Benefits

This application bridges the existing gap between data availability and actionable intelligence. By providing precise sales predictions by genre, it will:

* Enable proactive decision-making.
* Optimize marketing strategies.
* Enhance resource allocation.
* Increase overall profitability.

### Application Description

The data product employs linear regression and machine learning techniques, tailored to our specific industry needs. It leverages historical sales data to predict future trends, aligning with our organization's mission to stay ahead of market dynamics. The user-friendly interface ensures that even non-technical staff can harness its capabilities.

### Data Description

The application utilizes data, including Rank, Name, Platform, Year, Genre, Publisher, and Sales figures (in millions). Pros include comprehensive historical records and granularity by genre. Cons include potential inconsistencies in older data and the need for periodic updates. No significant anomalies have been detected in the dataset.

### Objectives and Hypothesis

* Objective: Achieve accurate, real-time sales predictions by genre to inform strategic decisions.
* Hypothesis: If we employ machine learning to analyze historical sales data by genre, then we can provide accurate forecasts that empower the organization to make informed decisions.

### Methodology

We will adopt the Agile methodology, which encourages flexibility and iterative development. This approach is ideal for this project because:

* Planning Phase: Determine specific requirements.
* Development Phase: Create, test, and refine the application.
* Deployment Phase: Implement and train users.
* Review Phase: Evaluate performance and make ongoing improvements.

### Funding Requirements

The total cost of this project is $90,000, including development, integration, training, and ongoing support.

### Data Precautions

While the data does not include sensitive or protected information, standard guidelines related to data integrity and confidentiality will be adhered to throughout the project's lifecycle.

### Developer’s Expertise

As the lead developer, my academic training in Computer Science, coupled with professional expertise in machine learning and predictive analytics, positions me uniquely to execute this project successfully. My past success in similar endeavors reinforces my ability to deliver a solution tailored to our organization's unique needs.

# Project Proposal

## Problem Statement

The current state of the video game industry is characterized by an immense and ever-growing diversity of genres, platforms, and regional markets. The dynamics of sales and consumer preferences are often fluid and require constant monitoring and analysis. The client's existing system of tracking video game sales by genre is manual, fragmented, and lacks the ability to provide real-time insights. This not only leads to delays in gathering essential market data but also contributes to a lack of coherent understanding of emerging trends, demands, and opportunities within various segments of the gaming market.

Moreover, the absence of predictive analytics in the existing tools is a significant hindrance to proactive decision-making. The industry's fast-paced nature necessitates predictive insights to anticipate market shifts, plan product launches, and tailor marketing strategies. The lack of a cohesive and automated system to track, analyze, and predict sales across various game genres has resulted in inefficiencies, increased labor costs, and missed opportunities for strategic alignment with market demands. Our proposed solution addresses these gaps by integrating data analytics, machine learning, and real-time reporting, fostering a more responsive and strategic approach to sales management.

## Customer Summary

The client represents a leading enterprise in the video game industry with a global presence, engaging in the development, marketing, and distribution of video games across various platforms and genres. Their vast portfolio and market outreach necessitate a sophisticated, data-driven approach to understanding the nuances of video game sales and consumer preferences. The existing methodologies have fallen short in delivering timely and actionable insights, causing fragmentation in strategic planning and execution. Our proposed application will cater precisely to the client's complex needs by offering an integrated, predictive analytics platform focused on sales by genre. This solution aims not only to automate and streamline the existing processes but also to enable the client to leverage advanced data analytics in aligning their products and strategies with emerging market trends. By resolving these critical challenges, the proposed application will significantly enhance the client's agility, efficiency, and competitive positioning in the ever-evolving video game industry.

## Existing System Analysis

The client's current technological environment consists of various disparate systems and tools, employed to analyze, and forecast video game sales across different regions and genres. While these systems provide basic statistical insights, they are plagued with shortcomings such as a lack of real-time analysis, limited predictive capabilities, and manual dependencies, all of which hinder their ability to adapt to the rapidly changing dynamics of the video game market. These applications often operate in silos, causing fragmentation and inconsistency in the data flow, which ultimately leads to delayed or misguided decisions. The absence of an integrated platform that can provide a comprehensive view of the market, analyze historical trends, and predict future sales based on genre-specific behaviors has created a significant void in the client's strategic planning. The proposed solution aims to replace this patchwork of tools with a unified and intelligent system that will not only overcome the current limitations but also enable the client to pioneer data-driven decision-making within their industry.

## Data

The raw data set that will be utilized for this project comprises several key attributes: Rank, Name, Platform, Year, Genre, Publisher, and Sales figures segregated by region (North America, Europe, Japan, and Other). The data is compiled from various sources and includes historical sales information across different platforms and genres.

Data will be collected both from available historical records and real-time feeds from various distribution channels, online platforms, and market research firms. This multi-sourced approach ensures a comprehensive view of the market. The processing phase will include cleaning, normalization, and transformation of the data to ensure it is fit for analysis. Anomalies such as outliers or incomplete data will be handled systematically through imputation or exclusion, based on the nature of the anomaly. This processed data will be managed through a robust data warehouse system, designed to support both the development lifecycle and ongoing maintenance of the application.

Data anomalies, if present, can skew the results and lead to inaccurate predictions. The system will implement advanced techniques to detect and handle anomalies such as outlier detection algorithms and data imputation methods. Incomplete data will be addressed through either exclusion or statistical imputation, depending on the nature and extent of the incompleteness.

## Project Methodology

For the development and deployment of this application, we will employ the Agile development methodology. Agile's iterative approach ensures flexibility, collaboration, and adaptability, making it particularly suitable for the ever-changing dynamics of the video game industry.

Phases of Agile Methodology:

**1. Discovery Phase:**

Needs Analysis: Understand the specific requirements and constraints.

Feasibility Study: Assess the project's viability, including technical and financial aspects.

**2. Planning Phase:**

Sprint Planning: Divide the development process into smaller sprints, with specific goals and timelines.

Resource Allocation: Identify and allocate necessary human and technological resources.

**3. Development Phase:**

Design: Create wireframes and design the architecture of the application.

Coding and Integration: Develop the code and integrate various components.

Testing Within Sprints: Conduct continuous testing to ensure functionality and fix any emerging issues promptly.

**4. Testing Phase:**

User Acceptance Testing (UAT): Validate that the application meets user expectations.

Security and Compliance Testing: Ensure that all security measures and compliances are met.

**5. Deployment Phase:**

Staging: Deploy the application in a staging environment to mimic the production environment.

Production Deployment: Roll out the application in the production environment.

**6. Maintenance and Iterative Improvement Phase:**

Ongoing Support: Provide ongoing support and maintenance.

Iterative Improvements: Continuously collect feedback and make iterative improvements.

The Agile methodology aligns well with the client's need for a responsive and adaptable solution. By embracing changes and promoting collaboration between cross-functional teams, Agile ensures that the project stays aligned with the evolving business priorities of the client.

## Project Outcomes

The finished application will be a comprehensive solution designed to analyze and visualize video game sales by genre. Featuring a user-friendly graphical user interface, it will offer interactive data visualization tools, real-time analytics, customizable reporting, and seamless integration with existing systems. A comprehensive user guide, video tutorials, and hands-on workshops will ensure the client's team fully understands the application's capabilities. Ongoing technical support, regular updates, and a detailed maintenance schedule will form the support and maintenance plan, ensuring robustness and efficiency.

In addition to technical aspects, the project outcomes include stringent security measures to ensure data integrity and compliance with industry standards. A well-planned promotion strategy will aim to maximize user adoption within the organization, while regular performance analysis and user satisfaction surveys will assess the application's impact and identify areas for improvement. The outcomes collectively aim to provide a tailored solution that aligns with the client's business goals and offers substantial value.

## 

## Implementation Plan

**General Strategy**: The implementation of the video game sales by genre application will follow a carefully planned strategy that emphasizes collaboration, efficiency, and risk management. The deployment will be executed in multiple phases to allow for controlled, effective integration with existing systems.

**Phases of the Rollout:**

1. **Pilot Phase**: This phase involves the deployment of the application within a confined environment. It will be tested by a select group of users who represent the target audience, and their feedback will be used to refine the system.

2. **Integration Phase**: The application will be integrated with existing databases and tools. This phase will involve close collaboration with the client's IT team to ensure seamless integration.

3. **Deployment Phase**: The full-scale deployment of the application to all intended users. Training and support materials will be provided to ease the transition.

**Dependencies**: The project relies on specific milestones and collaboration with various stakeholders. Dependencies include the completion of the previous project phases, availability of the client's IT resources, and timely feedback and approval from key decision-makers.

**Details for Testing and Distribution**:

**Testing**: Rigorous testing will be conducted at every stage of the rollout. This includes unit testing, system testing, integration testing, and user acceptance testing.

**Distribution**: The application will be distributed through secure channels that comply with organizational and industry standards. Regular updates and patches will be provided to ensure optimal performance.

**User Training & Support**: Training sessions, webinars, and written guides will be provided to users at different levels within the organization to ensure a smooth transition to the new system. A dedicated support team will be available to handle any issues that arise during or after the rollout.

**Monitoring & Evaluation**: Continuous monitoring and evaluation will be implemented to ensure the system is meeting its objectives. Metrics will be established to assess user engagement, system performance, and the overall success of the implementation.

By following this detailed implementation plan, the project will provide a seamless and efficient transition to a new, powerful tool for analyzing video game sales by genre. The plan's emphasis on collaboration, phased rollout, and continuous evaluation ensures that the application is aligned with the client's needs and provides value from day one.

## 

## Evaluation Plan

The evaluation of the video game sales by genre application encompasses a multifaceted approach that includes both verification and validation methods. During the development stages, verification will be conducted to align the project with client needs through reviews, inspections, and testing at each phase, from requirement analysis to integration. Design documents, code, and integrated components will be meticulously verified to meet specifications. In the post-development stage, validation methods like User Acceptance Testing (UAT), performance, security, and compliance testing will ensure that the application satisfies user needs, adheres to security standards, and complies with legal regulations.

Furthermore, the plan emphasizes ongoing monitoring and quality assurance. Continuous checks on system performance, feedback loops with users, adherence to industry standards, and a robust reporting mechanism will ensure that the application consistently meets project goals and allows for continuous improvement. Thorough documentation, regular reporting to stakeholders, and a final comprehensive evaluation report will summarize the evaluation process, affirming the alignment of the product with the client's technical and business objectives.

## Resources and Costs

The comprehensive budget for the development, implementation, and maintenance of the video game sales by genre application is estimated at $90,000. Below is a detailed breakdown of the costs:

1. Hardware Costs: $10,000

* Workstations, servers, and other necessary hardware components for development, testing, and deployment.

2. Software Costs: $8,000

* Licenses for development tools, database management systems, and other required software.

3. Development and Design Costs: $45,000

* Includes salaries for developers, designers, testers, and project managers.
* Additional costs for freelance specialists if required.

4. Deployment Costs: $5,000

* Costs for deploying the application on a cloud-based platform or dedicated server, including hosting fees.

5. Maintenance Costs: $7,000

* Ongoing support and updates to ensure the application remains functional and secure.
* Includes bug fixes, updates, and necessary adjustments based on user feedback.

6. Testing Costs: $5,000

* Includes costs for various testing phases, such as unit testing, integration testing, user acceptance testing, and performance testing.

7. Project Management and Other Overheads: $10,000

* Includes administrative costs, project management software, communication tools, documentation, and other miscellaneous expenses.

This carefully estimated budget aligns with industry standards and ensures an efficient allocation of resources to create a robust and effective solution for the client.

## Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Start Date | End Date | Duration |
| Requirement Analysis | 09/01/2023 | 09/15/2023 | 15 days |
| Design Phase | 09/16/2023 | 10/15/2023 | 30 days |
| Development Phase | 10/16/2023 | 12/15/2023 | 60 days |
| Testing Phase | 12/16/2023 | 01/15/2024 | 30 days |
| Deployment | 01/16/2024 | 01/31/2024 | 15 days |

# Post-implementation Report

Create a post-implementation as outlined below. Provide sufficient detail so that a reader knowledgeable in computer science but unfamiliar with your project can understand what you have accomplished. Using examples and visualizations (including screenshots) beyond the three required is highly recommended. **Write everything in the past tense.**

## A Business (or Organization) Vision

The organization faced a problem of efficiently tracking video game sales by genre. The existing system was outdated and lacked real-time analytical capabilities, making it difficult to identify trends and make informed business decisions. We developed a data-driven application that allows users to monitor, analyze, and visualize video game sales by genre in real-time. By leveraging machine learning algorithms, the application provided predictive insights and facilitated decision-making. Users could use our application to filter sales data by genre, view trends through interactive charts, and generate predictive sales forecasts. An example use case was the marketing team utilizing this application to plan promotional campaigns based on past sales performance and future predictions.

## Datasets

The raw data for our application consisted of intricate collections of sales figures, dates, genres, platforms, and regional information, each representing different aspects of video game sales. This raw data was meticulously processed, mainly involving cleaning to remove outliers, incomplete records, and normalization of inconsistent entries. Relatively little processing was needed due to the well-structured nature of the raw data, but the processing that was performed was essential to make the data accessible to our machine learning algorithms. The processed data was then formatted into a structure that was optimal for our predictive modeling. Access to the raw dataset can be obtained from [Kaggle](https://www.kaggle.com/datasets/gregorut/videogamesales).

## Data Product Code

The code began by loading and processing a raw dataset comprising historical video game sales data. This process included cleaning to remove missing or irrelevant entries and normalization to bring data into a consistent format. The preparation of this data was essential for feeding it into predictive models, making it accessible to our algorithms.

Following the data processing, descriptive statistical methods were employed to understand the underlying trends and patterns. Various visualizations were generated to offer insights into the data, such as the impact of genres on sales. These initial explorations allowed us to home in on the most important features and provided an essential context for the non-descriptive analysis. In the non-descriptive phase, the predictive analytic method that was applied was Linear Regression. This was chosen based on the complexity and nature of the sales prediction task.

These models were trained on a subset of the data and then tested using separate validation sets, ensuring unbiased performance assessment. Hyperparameter tuning was performed to optimize each model, and cross-validation techniques were applied to verify their robustness.

The combination of descriptive and non-descriptive methods was instrumental in creating a comprehensive analysis. The descriptive methods informed the choice and tuning of predictive models by revealing underlying data characteristics, while the non-descriptive methods were tailored to predict the sales figures effectively. The iterative nature of this approach allowed for continuous refinement, leading to a more accurate and reliable application.

All code, datasets, visualizations, and related resources for this project have been compiled and are accessible through the Jupyter Notebook file. The dataset used can be found on [Kaggle](https://www.kaggle.com/datasets/gregorut/videogamesales).

## Objective (or Hypothesis) Verification

The core objective of this project was to develop an analytical model that could predict video game sales specifically by genre. This necessitated a deep understanding of how various genres perform in different markets, platforms, and over time. The hypothesis driving this project was that genre plays a significant role in the sales of a video game and that an accurately tailored predictive model could forecast sales based on genre-related factors.

The objective was met with substantial success. By focusing on the genre as a main variable, the project team was able to isolate and understand the unique characteristics that drive sales within various gaming genres. Machine learning algorithms tailored to handle categorical data were deployed, allowing for nuanced insights into how different genres interact with other variables such as region and platform.

## Effective Visualization and Reporting

1. Data Exploration:

During this phase, a comprehensive understanding of the dataset was sought. Visualizations were used to get a quick insight into the relationships between variables. The scatter plot helped in exploring how sales varied across different genre, pinpointing the concentration of data points, and identifying potential outliers.

2. Data Analysis:

At this stage, the visualizations played an essential role in diving deeper into the data:

* **Bar Plot**: A bar plot was employed to analyze the sales by genre. This allowed us to identify which genres were leading in sales and where most of the market was concentrated. For example, a bar plot might reveal that action and sports genres were consistently high-selling categories.
* **Line Plot**: Line plots were used to visualize the trend of sales across different time periods by genre. This helped in identifying seasonal patterns or growth trends in sales, aiding in the selection of appropriate non-descriptive predictive models.
* **Scatter Plot**: Scatter plots assisted in understanding the relationships between different variables, like sales by genre versus user ratings. This helped to assess whether these factors could be influential in the prediction model.

3. Data Summary:

Summary statistics and visualizations were generated to collate findings:

* **Bar Plot**: Summarized the comparative sales figures across genres, reinforcing the key areas of focus.
* **Line Plot**: Provided a summarized view of temporal trends within genres.
* **Scatter Plot**: Summarized relationships between different variables, influencing the selection of features for the non-descriptive model.

4. Analysis Application of Three Visualizations:

* **Bar Plot**: Helped in feature selection by identifying the most prominent genres, which served as vital inputs to the non-descriptive prediction models.
* **Line Plot**: Guided the selection of time-series or regression models by highlighting trends over time, tailoring the non-descriptive methods to the inherent temporal characteristics of the data.
* **Scatter Plot**: Informed decisions about potential data transformations or normalizations needed, and aided in identifying suitable non-linear relationships that might require specific non-descriptive methods.

## Accuracy Analysis

The Mean Squared Error (MSE) was the primary metric used to evaluate the accuracy of the non-descriptive model in predicting video game sales by genre. MSE is calculated by taking the mean of the squared differences between the predicted and actual values. It's a standard metric in regression problems that quantifies how well the predicted values align with the true values. A lower MSE signifies better alignment between the model's predictions and the actual data.

In this specific project, the MSE value was 4.275239370828089. This value indicates the quality of the model's predictions. Though the interpretation of the MSE depends on the context and the range of the sales figures, this value does provide a concrete measure of how much the predicted values deviate from the actual values on average. Depending on the scale of the sales figures, this MSE may signify that there was a reasonably good fit between the predicted sales by genre and the actual sales.

Suppose the model was predicting sales for the action genre in different regions. The actual sales might be [100, 150, 200] (in thousands of units), and the model's predictions might be [105, 140, 190].

Calculating the squared differences:

* (105-100)2 = 25
* (140-150)2 = 100
* (190-200)2 = 100

Mean of these squared differences: (25 + 100 + 100) / 3 = 75

In this specific example, the error would be 75, which is part of the data contributing to the overall MSE of 4.275239370828089. Depending on the rest of the data, this overall MSE would suggest a certain level of accuracy in the model's predictions.

## Application Testing

The application was rigorously tested through various stages of development to ensure that it met the requirements and performed efficiently. Testing methods employed include:

* **Unit Testing**: Individual components of the code were tested in isolation to ensure that they worked as intended.
* **Integration Testing**: Different parts of the system were combined and tested as a group to validate the interactions between various components.
* **Validation Testing**: The final product was tested against the user requirements to confirm that the application performed as it was designed.

The application's testing results were carefully analyzed and led to several key improvements:

* **Model Tuning**: The initial performance of the model was not optimal, and the testing phase identified areas for improvement.
* **Bug Fixes**: Some bugs were identified during integration testing, related to data preprocessing and visualization. They were fixed promptly, resulting in a more stable application.
* **User Interface Adjustments**: Usability testing provided feedback on the user interface (UI), leading to some enhancements to improve the user experience.

In some areas of the application, no modifications were necessary. For instance, certain parts of the data processing pipeline and the implementation of the bar plot, line plot, and scatter plot were found to be robust and effective through the testing phases. They met the expected standards, and no further adjustment was required.

## Application Files

The project is structured with a main folder that contains the primary Jupyter notebook (`C964\_SIM2.ipynb`), housing the entire code. Additionally, there are subfolders:

1. Data Folder: Inside the data folder, you'll find the raw data file (`raw\_data.csv`) and the processed data file (`processed\_data.csv`).

2. Images Folder: This folder contains various visualizations used in the analysis, like bar plots, line plots, and scatter plots.

3. README File (`README.md`): A markdown file that provides a general overview of the project, instructions, and other relevant information.

4. Requirements File (`requirements.txt`): A text file containing the list of required Python libraries necessary to run the code.

To execute the application, the following libraries will need to be installed using a package manager like pip within the Jupyter Notebook environment or through the command line:

* pandas
* numpy
* matplotlib
* seaborn
* scikit-learn

The submission will be packaged in a ZIP file containing the following structure:

* `C964\_SIM2.ipynb`: This Jupyter notebook file contains all the code, including data preprocessing, analysis, and predictive modeling.
* `C964\_SIM2\_Task2.pdf`: A PDF document that includes a detailed write-up of the project, summarizing the methods, results, and conclusions.

## User Guide

1. Installing Anaconda:

* Visit the [Anaconda download page](https://www.anaconda.com/products/individual).
* Download Anaconda for Windows.
* Open the downloaded file and follow the on-screen instructions to install Anaconda, which includes Python and Jupyter Notebook.

2. System Requirements:

* Windows 10 Operating System.
* Stable internet connection.

3. Launching Jupyter Notebook:

* Open Anaconda Navigator from the Start Menu.
* Click on 'Launch' under Jupyter Notebook. This will open Jupyter in your default web browser.

4. Downloading the Application:

* Navigate to the application's repository URL.
* Click on the 'Download ZIP' button.
* Extract the ZIP file to a preferred location on your computer.

5. Opening the Application in Jupyter:

* In the Jupyter Notebook interface, navigate to the directory where you extracted the application's code.
* Open the Notebook file (e.g., `C964\_SIM2.ipynb`).

6. Installing Required Libraries:

* In the first cell of the Notebook, type and execute the following command:

!pip install pandas matplotlib scikit-learn numpy

7. Loading the Dataset:

* + Ensure that the dataset file (e.g., `vgsales.csv`) is placed in the same directory as the Notebook.

8. Executing the Application:

* + Go to the 'Cell' menu and select 'Run All.' This will execute the entire Notebook.

9. Using the Application:

* The application will process the data and display the predicted sales for the selected genre, along with visualizations (bar plot, line plot, scatter plot).
* At the bottom of the application, there will be a drop-down menu where you can select a specific genre to view the predicted sales.

## Summation of Learning Experience

My academic journey towards a bachelor's degree in computer science has laid a robust foundation for the skills and knowledge required to undertake the video game sales prediction by genre project. Through various courses, I have been well-equipped to handle the technical complexities of this project. The hands-on experience gained from class assignments and individual projects prepared me to deal with real-world data, apply appropriate analytic methods, and create meaningful visualizations.

While my academic background provided the core skills, there were areas where additional learning was necessary. The specificity of predicting video game sales by genre required me to research and understand the gaming industry's dynamics and trends. Resources such as online tutorials, technical forums, and industry reports played a crucial role in supplementing my existing knowledge. Utilizing platforms like Stack Overflow and GitHub allowed me to engage with community-driven insights and best practices.

This project experience has been a significant steppingstone in my lifelong learning journey. It has demonstrated that while formal education lays the groundwork, continuous learning is essential to adapt to new challenges and evolving technologies. The synthesis of theoretical knowledge with practical application has underscored the importance of adaptability and curiosity in driving my professional growth. It's not merely about mastering a set curriculum but about fostering an inquisitive mindset that seeks to learn, innovate, and contribute beyond the classroom walls.