October 25, 2025

Beno Recommendation Service

Alexandra Beno

Computer Science Capstone

C964 Task 2

Student ID 011095454

Table of Contents

Part A: Letter of Transmittal

October 16, 2025

Jeremy Elbertson

Beno Software Company

1225 Software Street, Chicago, IL

Dear Mr. Elbertson,

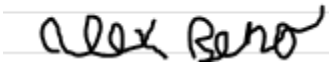
I am writing to you today in regards to the Beno Recommendation Service. As you are aware, the Beno Recommendation Service is a web-based program that is capable of recommending video games to users. However, it presently faces a key limitation in that it can only recommend the top five best-selling games. While they may be popular games, there is no guarantee that they will be of genres that the user enjoys. This has lead to user dissatisfaction, and has even caused certain users to stop using the service entirely in favor of competing products.

To address this issue, I propose the development and implementation of a new data-driven, AI-powered recommendation service. This new tool will employ a machine learning algorithm trained on the complete list of Steam Best-Selling Games, as of June 1st, 2025. Using the algorithm, the games will be grouped based on attributes like "FPS" or "turn-based." After prompting the user to enter a game they enjoy, the service will then be able to recommend a set of games from the same group of attributes. This will ensure that users are satisfied by the personalized results offered by the Beno Recommendation Service and are inclined to come back to it often for additional recommendations.

The development timeline of the project will be very short, going from programming to implementation in just a week. With five hours dedicated to planning and design, fourteen hours for development, and nine hours for documentation, there are an estimated 29 hours until completion of the project. Assuming a start date of this Saturday, October 18th, 2025, I expect completion of the project by October 25th, 2025. As the program will use a publicly available dataset and will be hosted online by a third party through Google Colab, there are no expected costs or ethical or legal concerns associated with the implementation of this project. With my background in Computer Science and previous history with dataset analysis, I am confident that this project will be completed successfully and without issues.

Thank you for taking the time to consider my proposal. I look forward to discussing this with you further.

Sincerely,



Alexandra Beno

Development Lead

Beno Software Company

Part B: Project Proposal Plan

Project Summary

The Beno Recommendation Service is a web-based program provided by the Beno Software Company that provides a list of video game recommendations to a user. However, it currently exhibits a key limitation in its design as it can only recommend the top five best-selling games at the time, with no regard for user preference. This has lead many users to feel disappointed with the Beno Recommendation Service and to look for alternative recommendation services. The Beno Software Company requires an intelligent, data-driven solution to enhance the Beno Recommendation Service and ensure that users continue to use it for video-game recommendations.

This proposal details a functional, AI-powered, data-driven solution which will enhance the Beno Recommendation Service with the ability to provide personalized recommendations for users based on preferred input game. This will come in the form of a Jupyter Notebook hosted on Google Colab, ensuring access to users through the web no matter their operating system. Users will be offered five games which are similar to the input game based on tags like genre and style of play, clustered through the k-means clustering algorithm. The program will also provide a variety of visualizations to help the user select the game that would best fit their needs, such as a bar graph for platforms the games are available on or a comparison of total reviews against the percentage of positive reviews. There will also be visualizations included, such as the silhouette coefficient analysis, to support data analysis and ensure that the k-means clustering algorithm is generating accurate results. Additionally, documentation will be produced, such as a User guide and Post-Implementation Report.

The implementation of this upgrade to the Beno Recommendation Service will help to ensure that users continue to use the service due to its personalized and accurate recommendations, resolving the program's current limitations. This should lead to increased user retention, and possibly even user growth as more potential customers are drawn to recommendations offered by the Beno Recommendation Service.

Data Summary

The dataset will be acquired from Kaggle, specifically the Best-Selling Steam Games of All Time list compiled by H. Bugra Eken here. The dataset is publicly available, and therefore there are no legal or ethical concerns involved in the use of the data. It features a list of the Best-Selling video games available on Steam as of June 1st, 2025. The dataset meets the needs of this project as it provides an excellent source of information on games available on Steam, and attributes of those games like genre and style of play. This is perfect for finding games with similar attributes to recommend to a user. This dataset specifically is strong because of the curated user defined tags section, meaning that sometimes extraneous tags added by users are ignored. Due to this, there are no expected outliers or data anomalies involved in the dataset, but it will still be examined for them regardless.

The dataset will not require a large degree of manipulation in order to be used in the project. The dataset already includes headers, and the user defined tags section to be used by this project has already been curated by the compiler of the dataset. In order for the user defined tags section to be used effectively by the Beno Software Recommendation Service, however, the tags will need to be split off into a matrix that can track multiple tags per game, rather than having all of a game's tags within the same variable. The

data will be stored by the program using a dataframe from the pandas library. The data will be viewed using a variety of python libraries to plot the data on graphs for better analysis, such as sklearn's t-SNE algorithm or matplotlib's pyplot. Future maintenance on the program will involve updating the program with a more recent dataset, but should otherwise require no changes.

Implementation

This project will implement the industry-standard SEMMA (Sample, Explore, Modify, Model, Assess) methodology in pursuit of completion of the project. The team will begin by obtaining the dataset from Kaggle that contains a Sample of the games available on Steam, including just those on the Best-Sellers list. The team will then Explore the data by analyzing data using plots and charts to determine if there are any issues or outliers within the dataset. If so, the dataset will be Modified and the issues or outliers are removed. Then the team will proceed to Model a machine-learning solution based off of the data, using the k-means clustering algorithm. The team will then Assess the results of the k-means clustering using silhouette coefficient analysis to ensure that the results offered by the new Beno Recommendation Service are reliable and make sense.

Timeline

Milestone or deliverable	**Project Dependencies**	**Resources**	**Start and End Date**	**Duration**
Proposal Acceptance	N/A	Development Engineer	10/16-10/18	2 days
Dataset Download	Proposal Acceptance	Kaggle	10/18-10/19	1 day
Dataset Exploration	Dataset Download	Spyder Pandas matplotlib	10/19-10/20	1 day
Dataset Modification	Dataset Exploration	Spyder Pandas matplotlib	10/20-10/21	1 day

Modelling	Dataset Modification	Spyder Pandas Sklearn	10/21-10/22	1 day
Assessment	Modeling	Spyder Pandas Sklearn AI Model	10/22-10/23	1 day
Cloud Implementation	Modeling	Google Colab	10/23-10/24	1 day
Documentation	Cloud Implementation & Assessment	Completed Project	10/24-10/25	1 day

Evaluation Plan

Throughout the development of the project, the Beno Recommendation Service will be subjected to a variety of unit tests to ensure that all of its pieces remain functional throughout its lifecycle. The dataset will be examined for outliers and anomalies, and the code and its functions will be repeatedly checked for errors and issues.

Meanwhile, silhouette coefficient analysis will be employed as a validation method to ensure that the k-means clustering algorithm employed by the Beno Recommendation Service is offering accurate results. The analysis will look at each point in a cluster, and determine whether it is close to its own cluster or closer to another cluster. This will allow for examination of the machine-learning model employed and determination of whether or not it is meeting the project requirements by accurately grouping data.

Costs

- Office Space: $300
    - Short-term office rental (1 week duration): $250
        - Workspace: $150
        - Utilities: $50
        - Internet (Spectrum): $50

    - Furniture: $50
        - Benocorp Desk: $20
        - Benocorp Chair: $30
- Hardware: $0
    - Dell Thinkpad Laptop (pre-existing): $0
    - Peripherals: $0
        - Dell Keyboard (pre-existing): $0
        - Logitech Mouse (pre-existing): $0
        - Dell Screen (pre-existing): $0
- Software: $0
    - Steam Best-Selling Games List (Kaggle Dataset): $0
    - Google Colab v2025-10-14 (host): $0
        - Jupyter Notebook (integrated IDE): $0
    - Spyder 6.1.0 (IDE): $0
    - Python 3.12 (language): $0
        - Python Libraries: $0
            - Pandas 2.2.3: $0
            - Sklearn 1.7.2: $0
            - Matplotlib 3.10.7: $0
- Labor: $1,015
    - Development Engineer: $1,015
        - Total Estimated Hours: 29
        - Hourly rate: $35/hr
- Ongoing costs: $0
    - Google Colab Maintenance: $0
    - Google Drive Storage: $0
    - Maintenance and Updates: $0

Total Project Cost: $1,315

Part C: Application

The following files are included in the application:

- AlexandraBenoCapstone.ipynb: This is the Jupyter Notebook file, intended to be uploaded to a Google Colab environment.
- bestSelling_games.csv: This file contains the dataset used by the program, from Kaggle.
- AlexandraBenoCapstone.pdf: This file, containing the User Guide, Post-Implementation Report, Project Proposal Plan, and Letter of Transmittal.

For instructions on how to use the application, please refer to the User Guide in Part D.

Part D: Post-implementation Report

Solution Summary

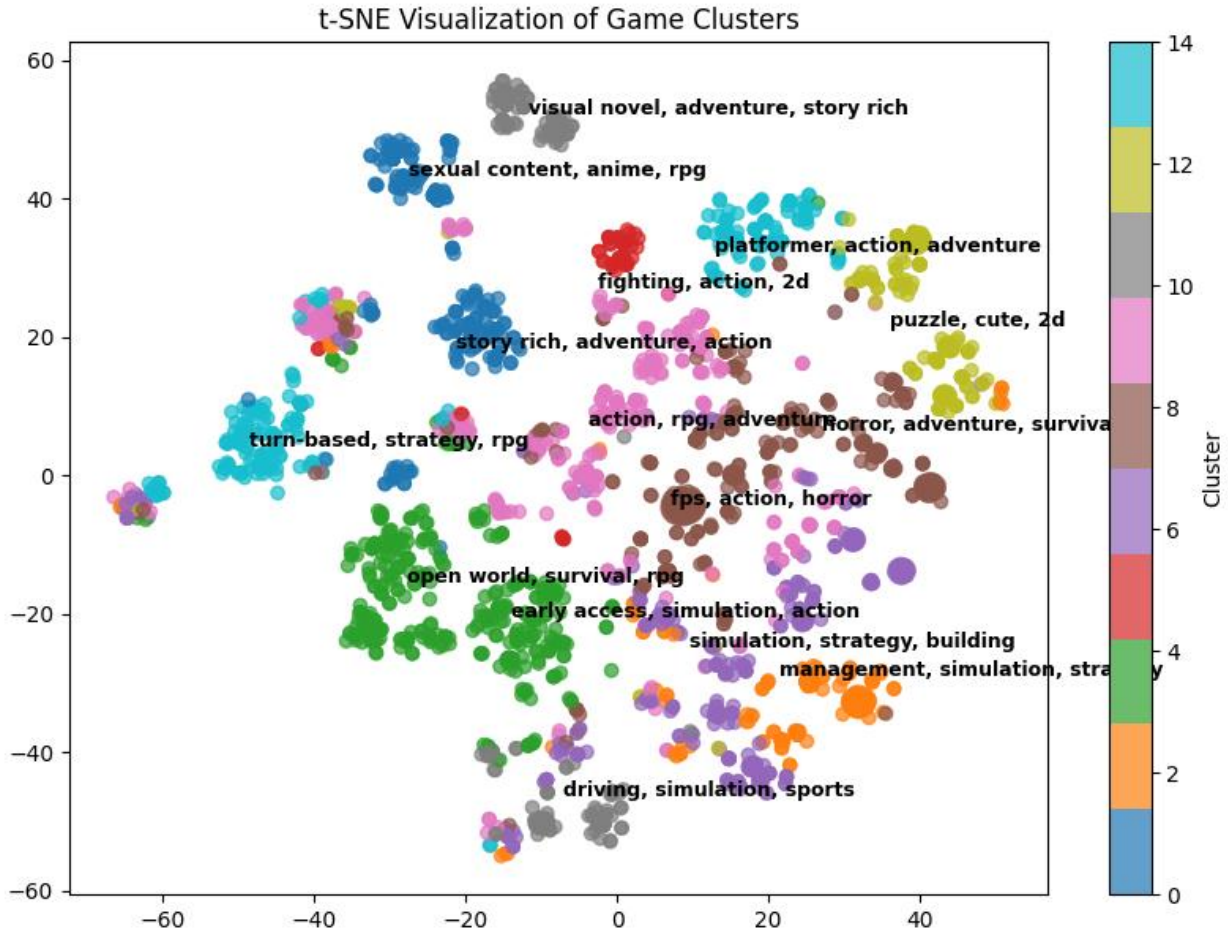
The Beno Recommendation Service recommends video games available on the Steam Platform to users. This project was created with the intent of resolve a key limitation in the programming of the original service which caused it to only recommend the five Best-Selling Steam games, with no regard for individual user preference. This left users feeling unrecognized and unfulfilled by the recommendations offered, leading to them pursuing suggestions from competing services. In order to improve the recommendations provided by the Beno Recommendation Service and restore user growth and satisfaction, a data-driven, AI-powered solution was implemented. By parsing the complete list of Steam Best-Selling Games as of June 1st, 2025, a k-means clustering machine learning algorithm is now used by the Beno Recommendation Service to provide personalized recommendations to users. Games are clustered into 15 different groups of similar games that the service then pulls from to recommend titles to the user based on an input game. This allows for each user to determine the best game for them to purchase based on games that they like, rather than just being given a generic list of popular games, directly addressing the preexisting limitation of the service. This solution has increased user satisfaction with the Beno Recommendation Service and lead to increased user growth since implementation.

Data Summary

The data for the Beno Recommendation Service is sourced from the Kaggle dataset "Best-Selling Steam Games of All Time" by H. Buğra Eken, available here. The dataset already features headers and 2,380 points of data. Each row of data describes a game that was on the Best-Sellers list on the Steam online gaming platform as of June 1st 2025. Each game in the dataset has a curated list of user defined tags, which represent details about the game such as genre, like "FPS," "Action," or "Tactical." The dataset is placed into a dataframe and read by the Beno Recommendation Service using the pandas library. Using the TfidVectorizer library from sklearn, the list of tags is converted into a matrix so that each tag is stored individually rather than as a group in a single cell. This matrix of tags is then used to create the clusters of games through the k-means algorithm detailed below. As the data is publicly available, there are no ethical or legal concerns present in the use of this dataset in the Beno Recommendation Service.

Machine Learning

The machine learning model included in the Beno Recommendation System is the k-means clustering algorithm. K-means clustering is used to categorize items by similarity into a number of groups equal to k. In this project, it is used to group the games from the Steam Best Sellers list according to the similarity of their user defined tags. The following is an example output produced by the program using an algorithm which visualizes the clusters of games in a 2D space. The closer games are together, the more similar they are, while the further they are apart, the more different. Different colors represent the different groups of similar games. The clusters are represented by a label that shows the top three user defined tags for each cluster.

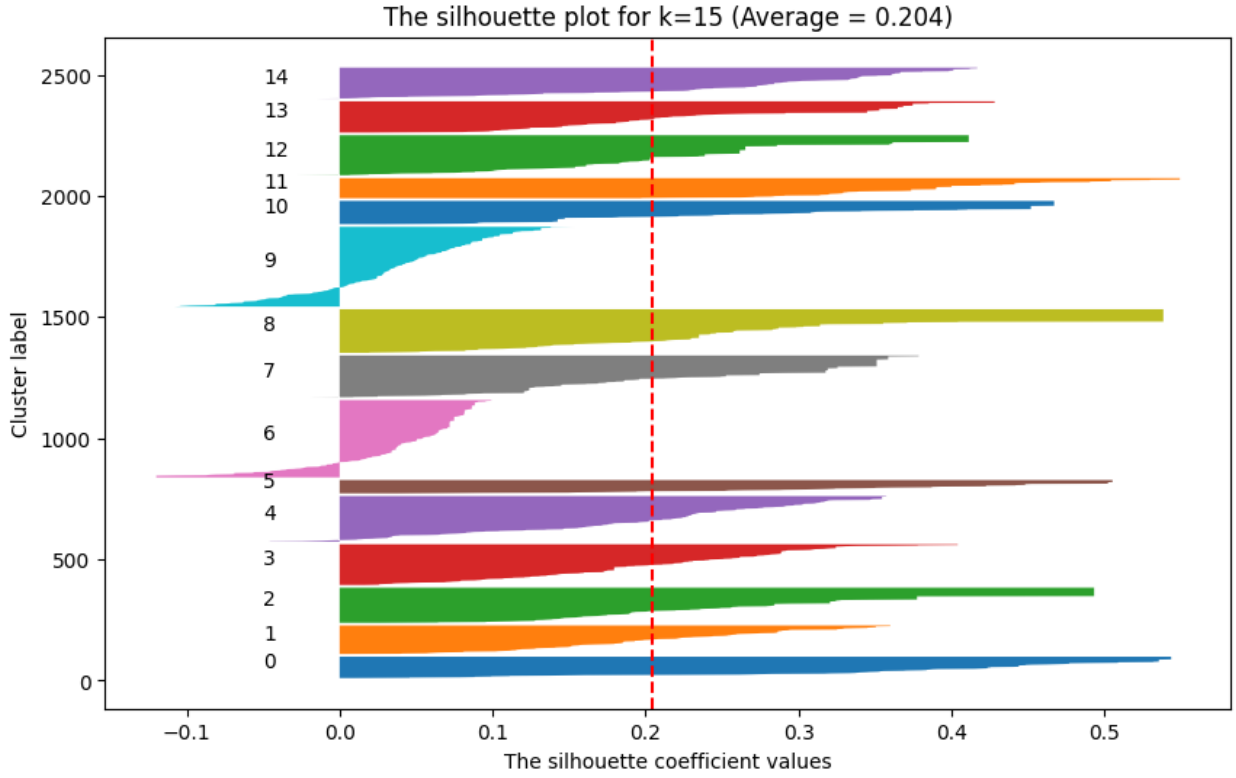


In the project, the k-means clustering algorithm was implemented through the sklearn KMeans library. k is set to a value of 15 to ensure accurate grouping of games. To start with, 15 games are selected at random. The remaining games are then divided based on their tags and placed into 15 clusters based on the similarities they share with the initially selected and already placed games. Then, the clusters are labeled with colors and with the top three tags totaled from the games in their group. When the user inputs a game, its tags are then compared to the others within its cluster, and the most similar five games are selected to display to the user.

The k-means clustering algorithm was chosen because the dataset of Steam Best-Selling Games does not have labeled results, and therefore required an unsupervised learning solution which can predict based on information like tags. The k-means clustering algorithm was perfect for this purpose, as it is easily able to sort through vast datasets of categorical data and sort them into different clusters. The sklearn implementation of the k-means clustering algorithm was also chosen because it is easily able to interact with other sklearn and python libraries to produce a variety of plots and graphs. The value 15 for k was chosen based on the application of the “elbow method” to the dataset, which is the process of plotting the variation within the dataset for different numbers of clusters and then choosing the point where it begins to level off. The accuracy of the k-means clustering algorithm was tested through silhouette analysis, as detailed below.

Validation

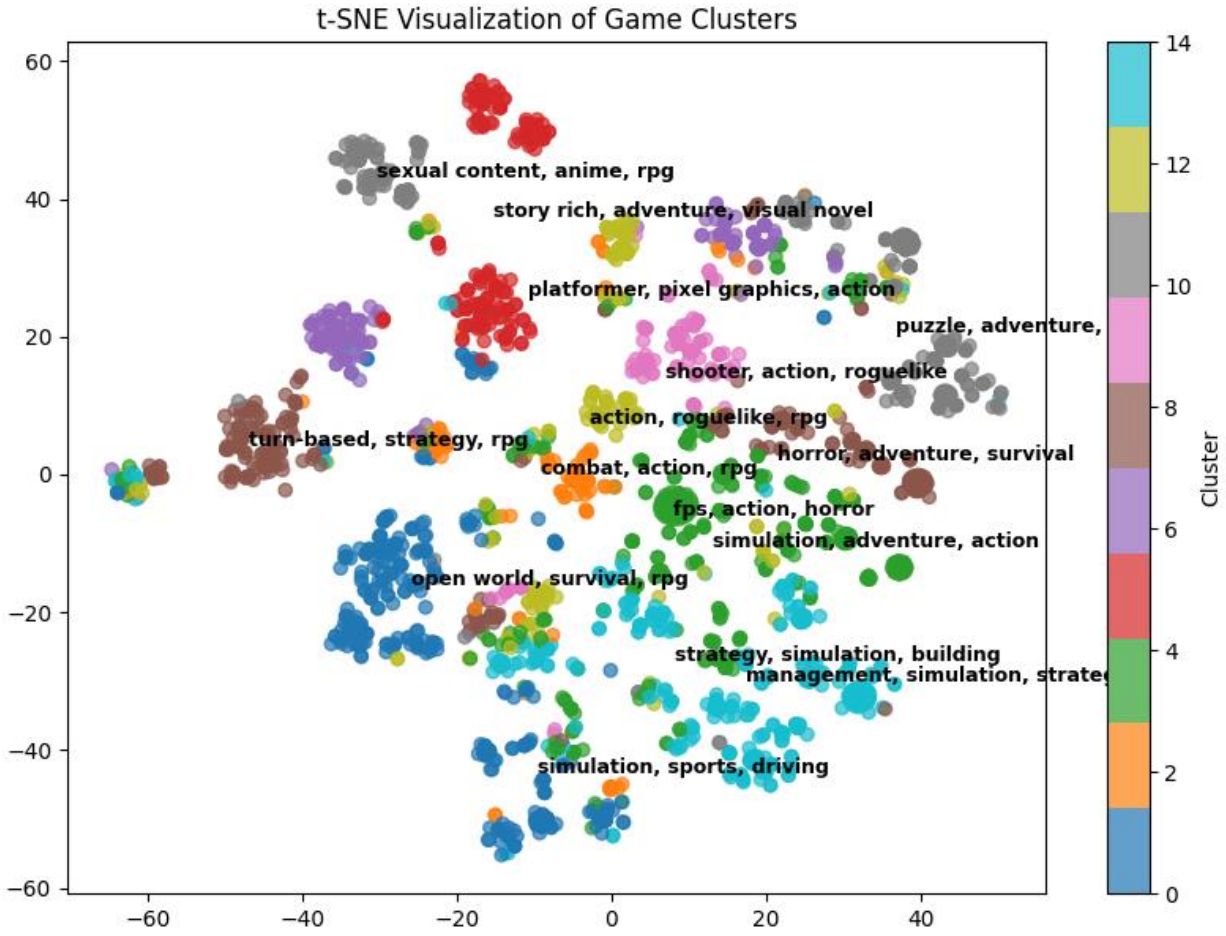
The k-means clustering algorithm is an unsupervised learning model, meaning it is focused on discovering relationships between unlabeled data. In order to measure its accuracy, a silhouette coefficient analysis was employed. This measures the similarity of each point to the other points in its cluster as well as those of other clusters. Each point is assigned a score, ranging from -1 to 1, with higher numbers indicating that the point is closer to the center of its own cluster, while a lower score indicates it is closer to other clusters than its own. Below is an example output where the program produced an average score of 0.204, indicating that the games were accurately grouped enough to ensure that the Beno Recommendation Service functions properly:



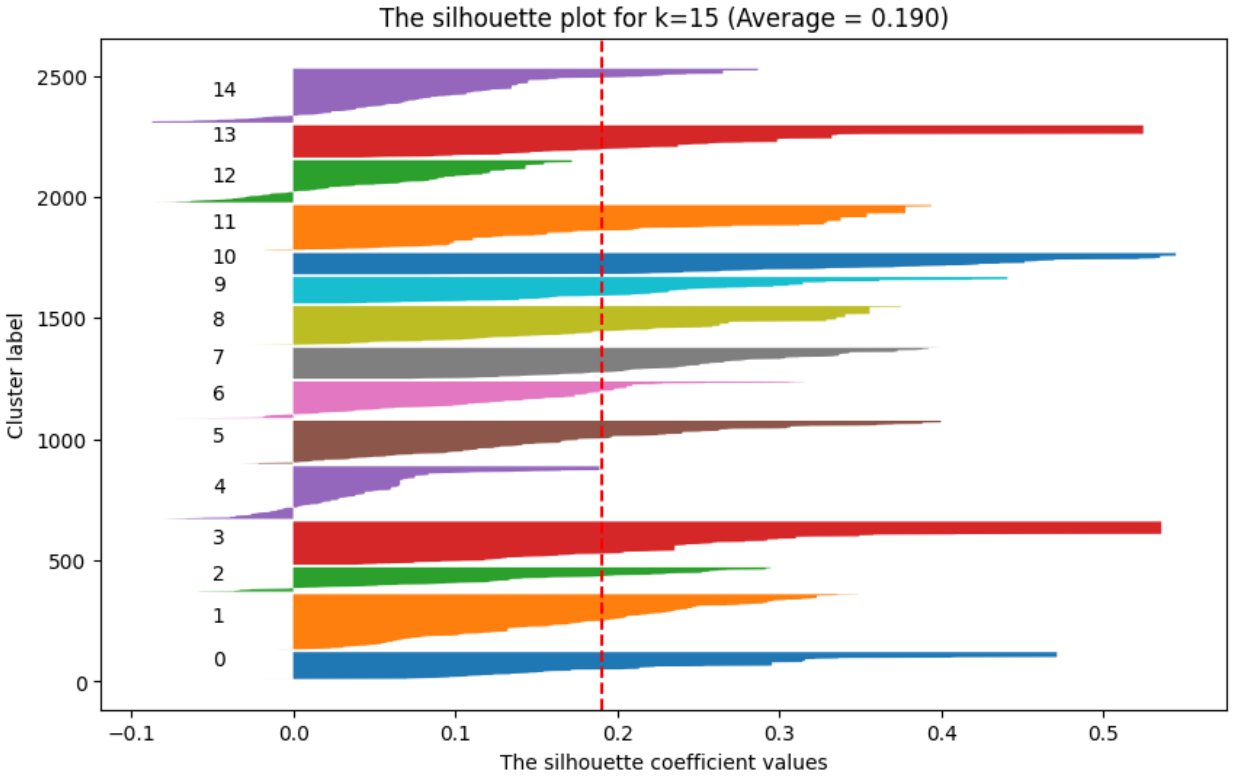
Visualizations

The Beno Recommendation Service is capable of providing four different visualizations:

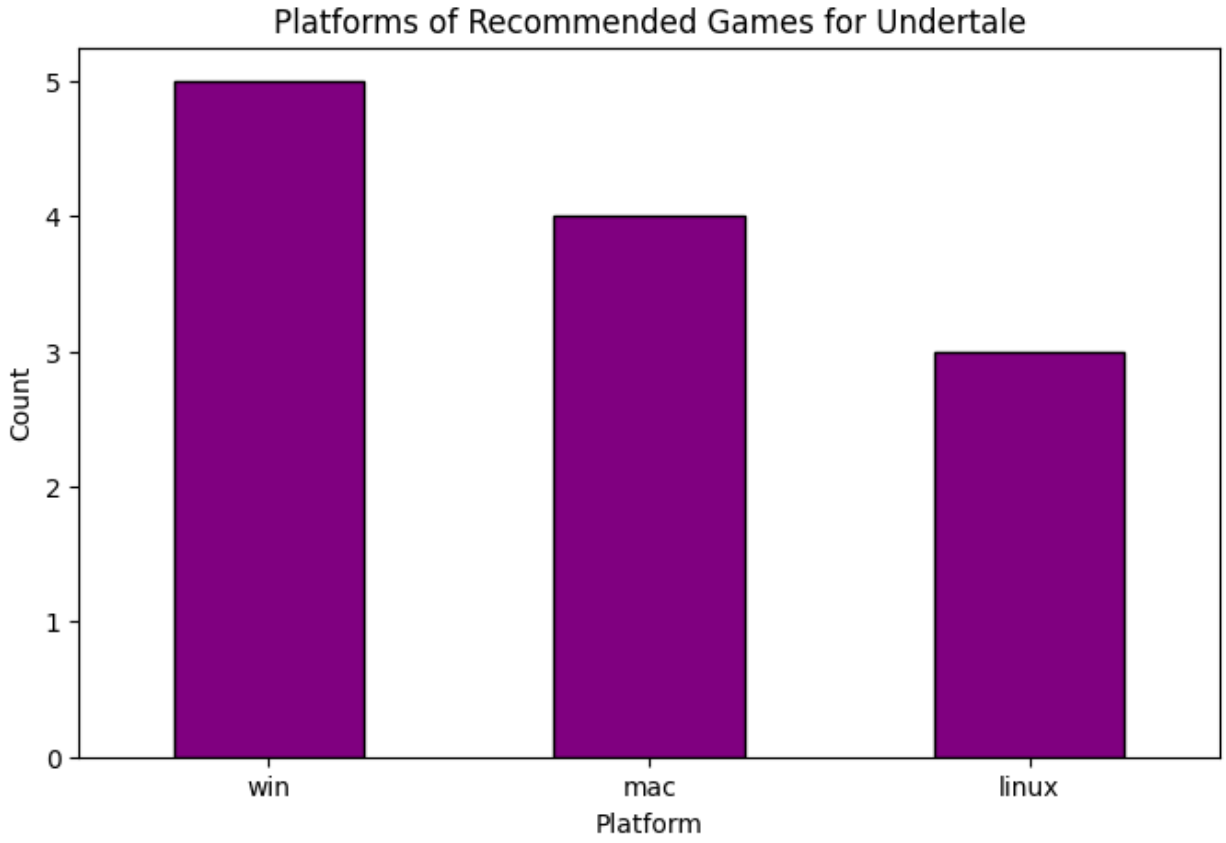
1. By clicking “Data Analysis”, the user can access a visualization powered by the t-distributed stochastic neighbor embedding algorithm that displays the clusters of games in a 2D space as assigned by the Beno Recommendation Service:



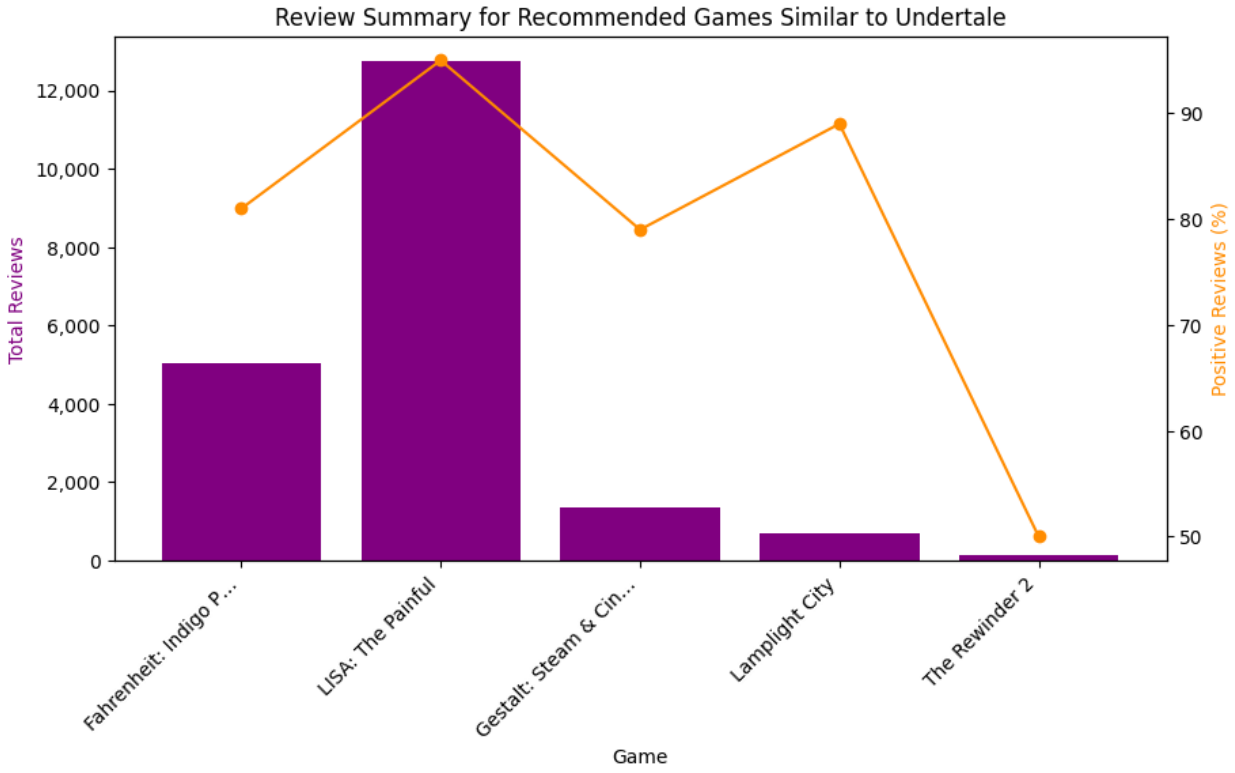
2. By clicking “Data Analysis”, the user can access a visualization powered by silhouette coefficient analysis that displays the accuracy of the Beno Recommendation Service in assigning games to clusters of similar games:



3. By clicking “Recommendations” and inputting a game title, like “Undertale” (without quotes), and then clicking “Confirm”, the user can access a bar graph displaying the platforms that 5 games similar to the input title are available on:



4. By clicking “Recommendations” and inputting a game title, like “Undertale” (without quotes), and then clicking “Confirm”, the user can access a summary of the reviews for games similar to the input title, including the total number of reviews and the percentage of the reviews that are positive:

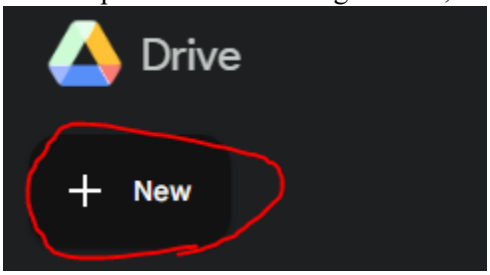


User Guide

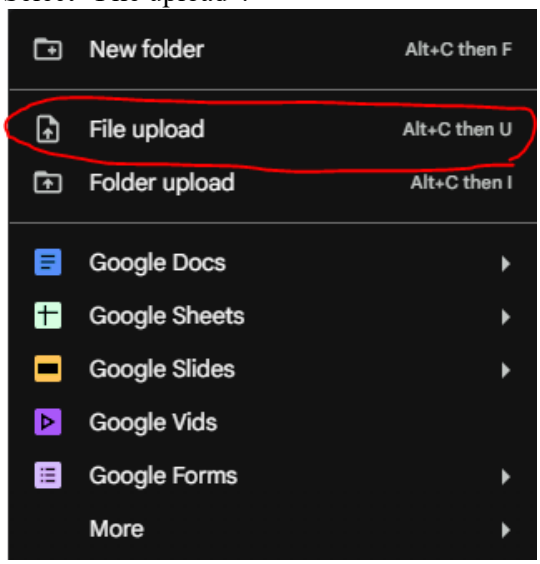
Welcome to the Beno Software Company’s exclusive Beno Recommendation Service! This is a lightweight, AI-driven, cloud-based program designed to recommend best-selling games on Steam based on game titles that a user inputs.

Instructions on how to use the program are as follows:

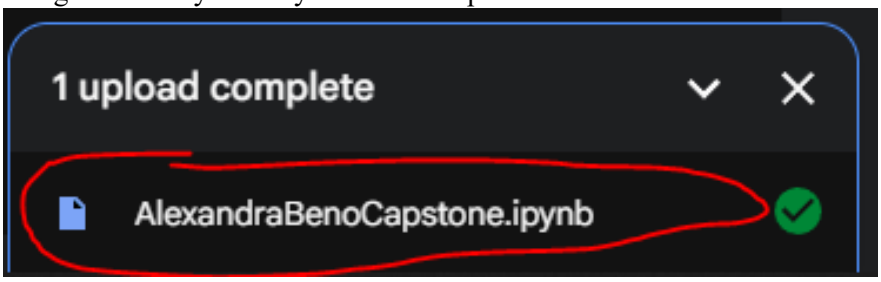
1. Unzip all included files.
2. Navigate to drive.google.com
3. In the top left corner of Google Drive, click the “New” button:



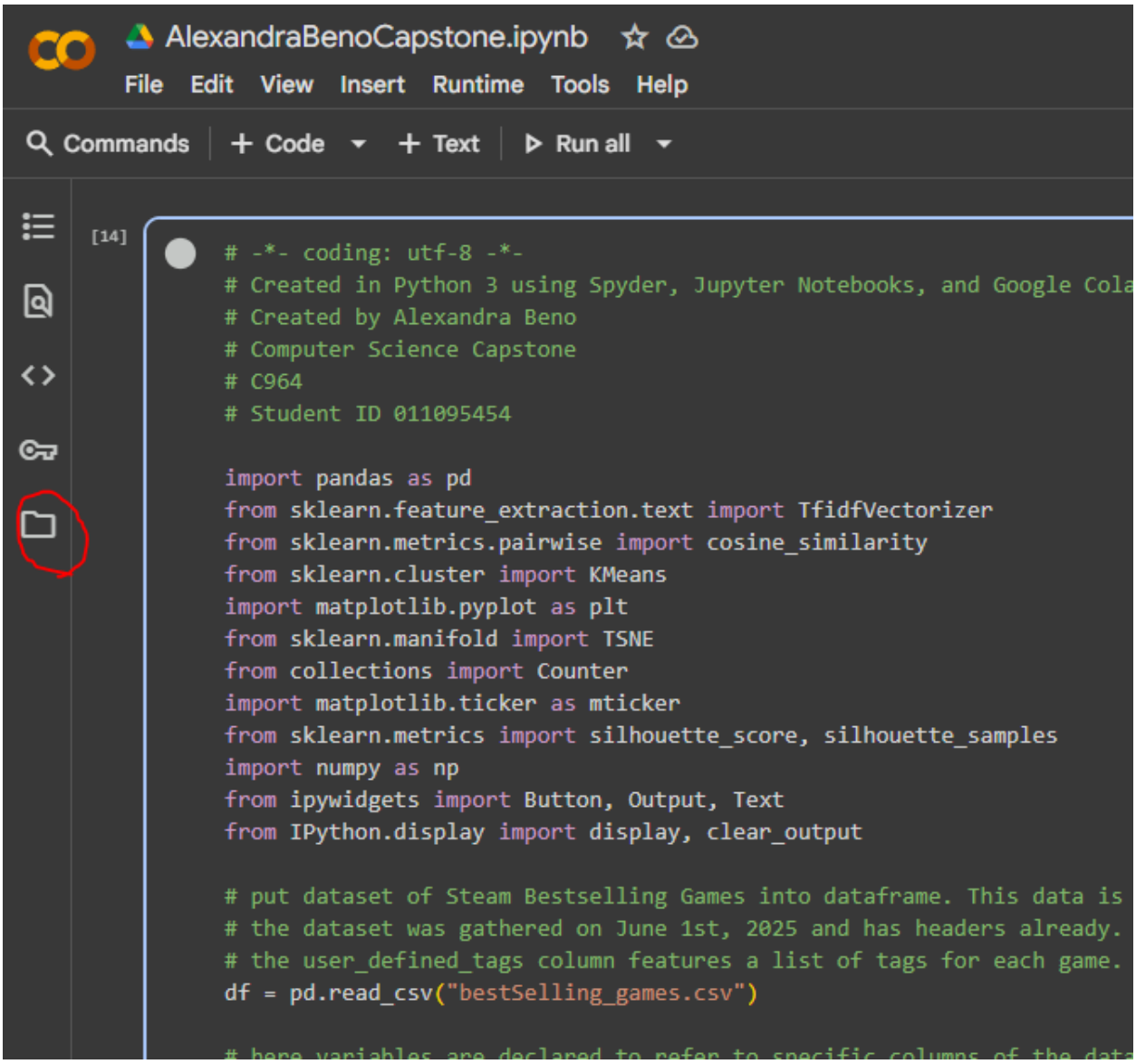
4. Select “File upload”:



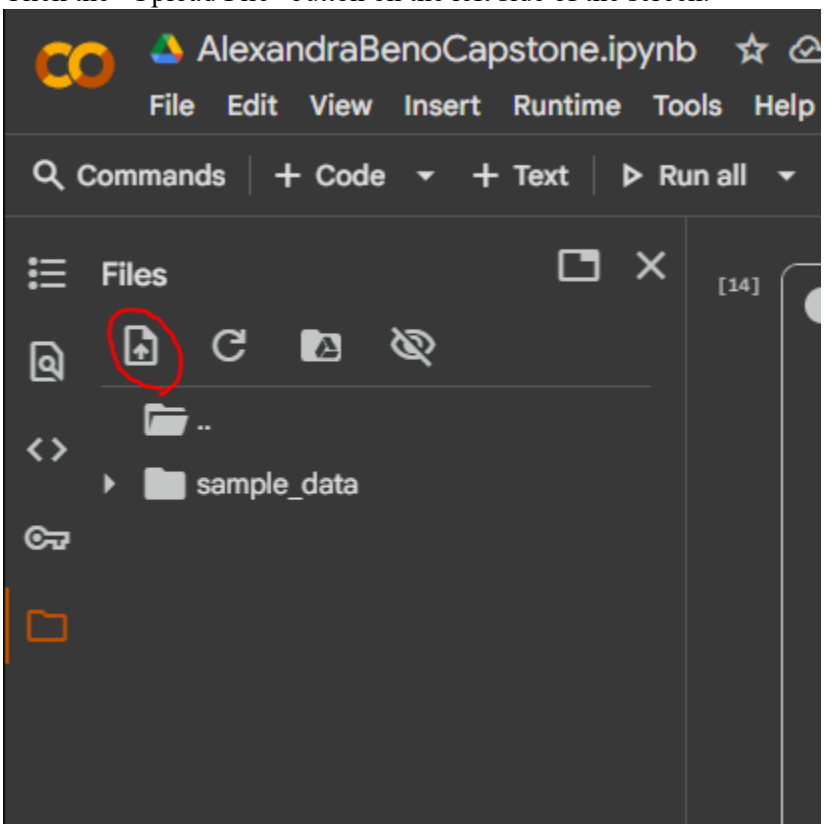
5. Navigate to and upload the AlexandraBenoCapstone.ipynb file.
6. Once it has been uploaded, click on the uploaded file in the lower right corner of your screen, or navigate to it in your “My Drive” and open it:



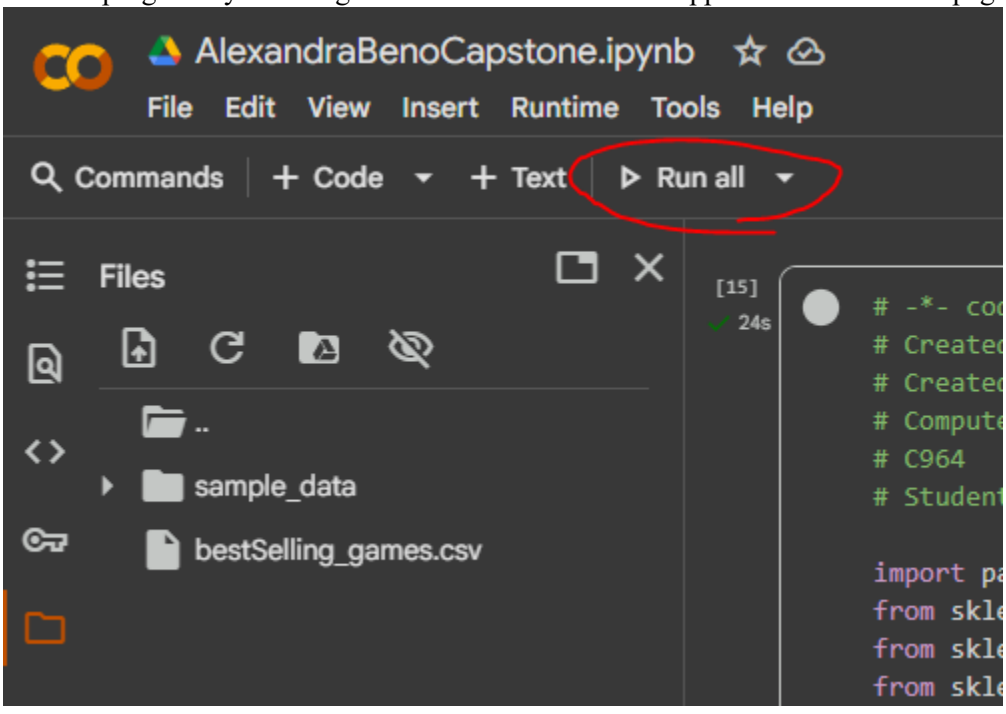
7. Click the “Files” folder button on the left side of the screen:



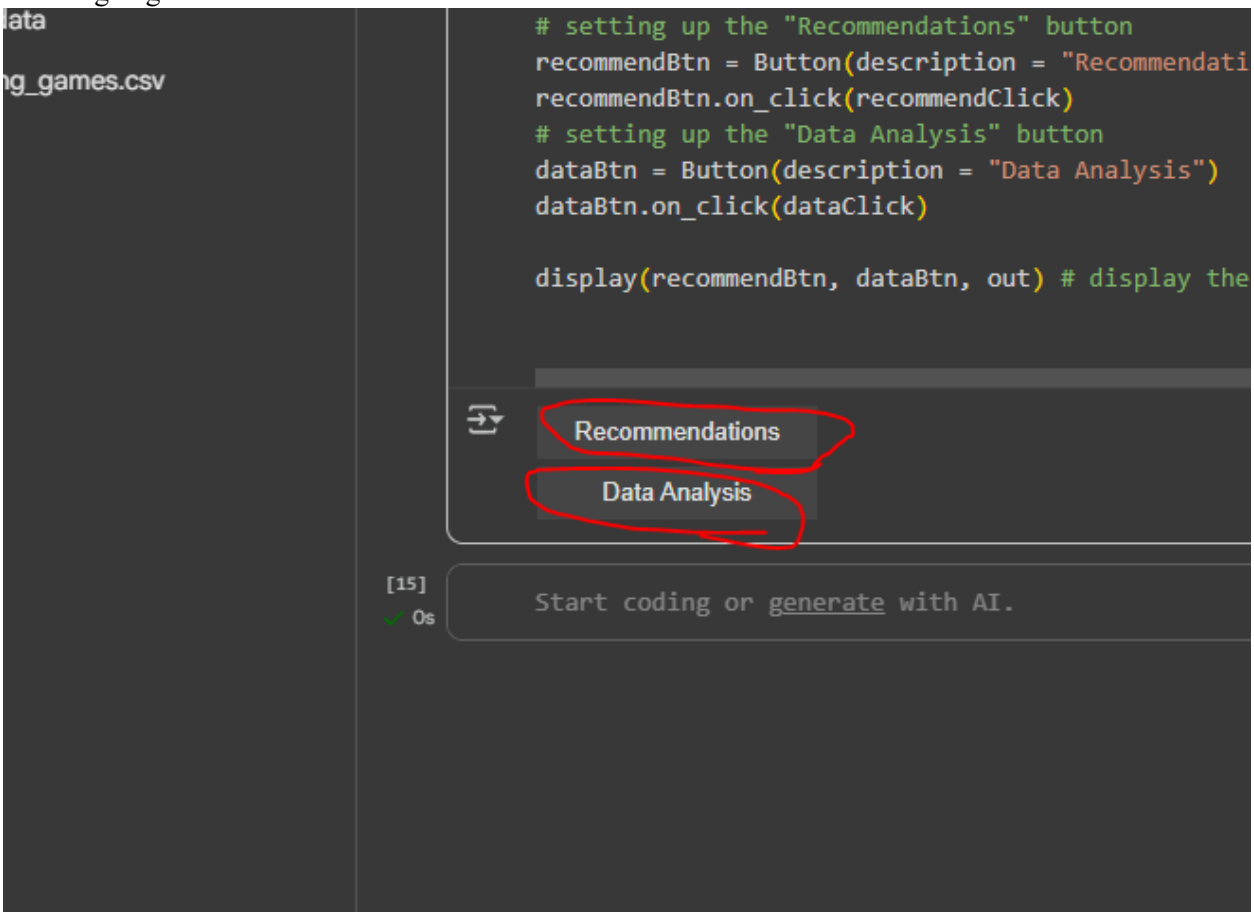
8. Click the “Upload File” button on the left side of the screen:



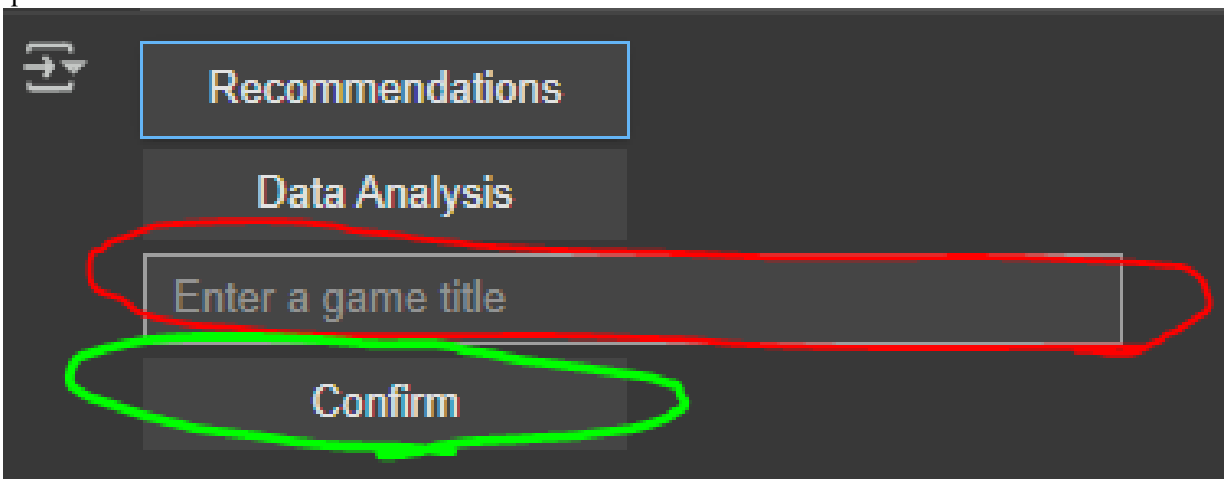
9. Navigate to and upload the bestSelling_games.csv file.
10. Run the program by selecting the “Run all” button in the upper left corner of the page:



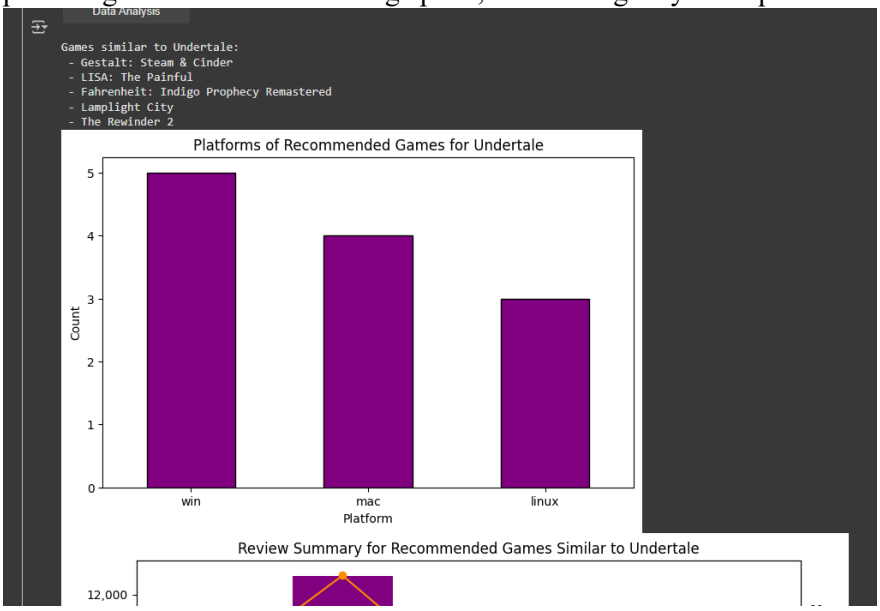
11. Scroll down and select either “Recommendations” for game recommendations or “Data Analysis” for analysis on how the data has been grouped by the Beno Recommendation Service’s Machine Learning Algorithm:



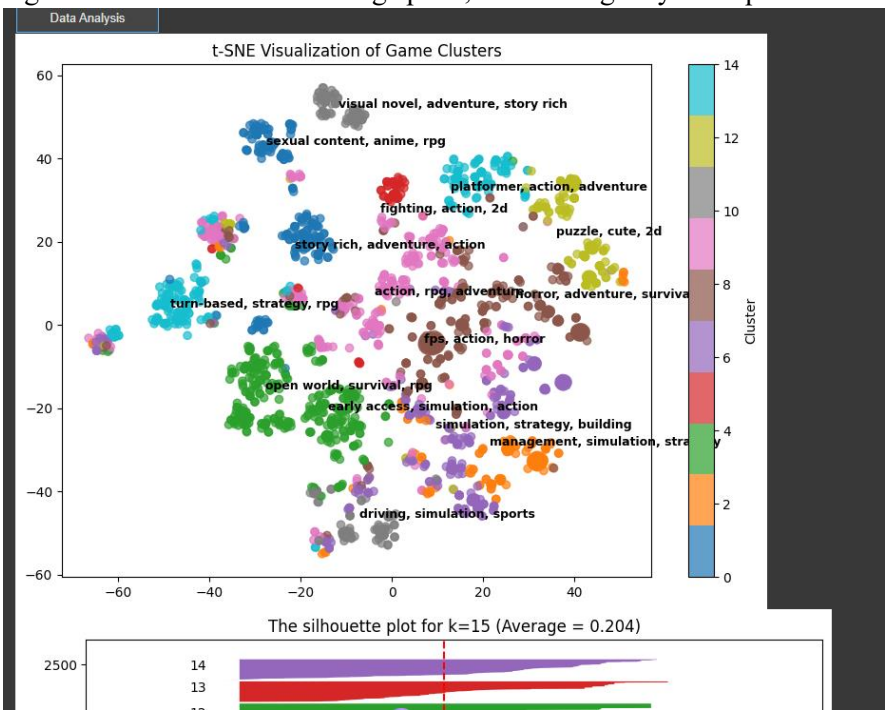
12. If “Recommendations” is selected, input a game title in the text field circled in red, then select the “Confirm” button circled in green. Example game titles include “Undertale” or “Terraria” without quotes:



13. A list of 5 similar games to the input game will be produced, as well as graphs illustrating the platforms of the recommended games as well as their total reviews and positive review percentage. There are several large plots, so scrolling may be required:



14. If "Data Analysis" is selected, two plots are produced that describe how the data was grouped by the Beno Recommendation Service's k-means Machine Learning algorithm, including a t-SNE visualization of how the data was grouped and a silhouette plot to gauge the accuracy of the algorithm. There are several large plots, so scrolling may be required:



References

No references were used.