

November 29, 2022

Movie Recommendation System

David Israel

Part A: Project Proposal for Business Executives.....	4
Letter of Transmittal	4
Project Recommendation	6
Problem Summary	6
Application Benefits	6
Application Description.....	6
Data Description	7
Objectives and Hypothesis.....	7
Methodology.....	7
Funding Requirements.....	8
Data Precautions.....	8
Developer's Expertise	8
Part B: Project Proposal	9
Problem Statement.....	9
Customer Summary	9
Existing System Analysis.....	10
Data.....	10
Project Methodology	11
Project Outcomes	12
Implementation Plan	13
Evaluation Plan.....	14
Resources and Costs	14
Timeline and Milestones.....	15
Part C: Application	16
Part D: Post-implementation Report	17
A Business (or Organization) Vision	17
Datasets	17
Data Product Code	18
Objective (or Hypothesis) Verification.....	21
Effective Visualization and Reporting	21
Accuracy Analysis	22
Application Testing	22
Application Files.....	23

User Guide 23

Summation of Learning Experience 24

Part A: Project Proposal for Business Executives

Movie Recommendation System

Letter of Transmittal

David Israel

9015 Blue Hawk Ln

Kernersville, AL 48596

Tim Turner

CEO

Movie Champ

895 River St

Kernersville, AL 48596

Dear Mr. Tim Turner,

With the attached estimated cost, you will find information on the benefits of acquiring a Movie Recommendation System for your business and my recommendations for a specific plan of action.

My analysis included information about the main problem that businesses like Movie Champ are facing in this new digital era. As a movie rental business, your company has been struggling with growing competition in the form of emerging online streaming platforms. In addition, your business may struggle with finding skilled staff with enough knowledge to create an optimal customer service experience.

How the Movie Recommendation System will benefit the business:

The Movie Recommendation System will provide your business with the necessary tools to optimize the customer service experience by providing a list of movie recommendations based on the movies of interest to every customer. This system will also support the staff who may lack the technical knowledge to assist customers when asking for recommendations. In addition, the provided recommendation can be later used in marketing campaigns that will also benefit your business.

Estimate of the total cost of the project:

\$7500.00

I am an experienced computer scientist, with experience working on projects with Automation and Machine Learning using different programming languages and frameworks. I will be pleased to show you that I can go above and beyond to bring your business back on track with new technologies.

If you have additional questions, feel free to contact me anytime at dvaca1@wgu.edu or (980) 989-0734. I look forward to speaking with you further on this project.

Sincerely,

David Israel

Project Recommendation

Problem Summary

- The Movie Recommendation System is a tool that provides movie recommendations based on the interest of every customer and based on the ratings provided by thousands of other customers. With an advanced algorithm, non-technical users and employees can easily access recommendations based on their favorite movies.
- Movie Champ is a movie rental business that has been facing growing competition in the form of online streaming platforms, and the lack of skilled employees with the knowledge to fully support the needs of the customers and create an optimal user experience. Therefore, a Movie Recommendation System is needed to bring Movie Champ back on track with today's technologies.
- The Movie Recommendation System will provide movie recommendations based on the user's interest and on the ratings that other customers provided. The recommendation will enable the employees to provide a better customer experience and for the marketing team to have a more focused approach when marketing to the customers.
- Items that will be delivered and achieved:
 - A fully functional Movie Recommendation System.
 - A user-friendly interface that customers and employees of Movie Champ can easily use.
 - A tool for the marketing team to take a more customer-focused approach in their marketing campaigns.

Application Benefits

- By using Machine Learning, the algorithm designed for the Movie Recommendation System will use the data to create a recommendation based on the customer's interest and on the rating thousands of other users provided, therefore making the movie recommendations more relevant and specific.

In addition, the System will have the capability to adapt when the customer's interest changes or when the ratings for the movies of interest change.

With a user-friendly interface, the System will be easy to learn for the employees, therefore quickly enabling them to make the customer experience better.

With an adaptive algorithm, the System can help the marketing team to focus their efforts on a more customer-focused approach, potentially reducing the costs of the marketing campaigns.
- Movie Champ will benefit from using the Movie Recommendation System by giving every customer a more personalized experience, therefore increasing the likelihood of returning customers.

The System will also provide the necessary tools that the employees need to assist the customer needs.

Application Description

- The Movie Recommendation system will do the following:
 - a. Analyze the movie data.
 - b. Find similar users who gave a high rating to the movie of interest.
 - c. Find movies highly rated by similar users.
 - d. Find the percentage of similar users who highly rated every found movie.
 - e. Find the percentage of all the users who highly rated every found movie.

- f. Analyze the difference between the percentages and create a similarity score.
- g. Return the movies with the highest similarity score.

Data Description

- The Raw Data is two data sets:
 - Movies data: 62,423 rows. 3 Columns: movieId, title, and genres.
 - Rating data: 25,000,095 rows. 4 Columns: userId, movieId, rating, and timestamp.
- Type and data structure:
 - Movies data: Type: Nominal data. Data Structure: Pandas DataFrame.
 - Rating data: Quantitative. Data Structure: Pandas DataFrame.
- Dependent and independent variables:
 - Dependent Variables:
 - Similarity Score
 - Independent Variables:
 - movieId
 - title
 - genres
 - userId
 - rating
- Anomalies and limitations:
 - Genres missing for some movies
 - Ratings are too low for some movies making it difficult to create a recommendation.
 - The year is included in the title.

Objectives and Hypothesis

- Desired outcomes of the project:
 - Movie Champ employees will be able to better support the customers.
 - The marketing team will reduce its budget and focus its efforts on a customer-focused campaign.
 - The number of returning customers will increase.
- Hypothesis:

The Movie Recommendation System will deliver custom accurate recommendations for most Movie Champ customers.
- Prediction accuracy:

The movie recommendations will be at least 80% accurate to what the customer's interests are.

Methodology

- The chosen methodology is Agile. This methodology was chosen because of the flexibility that provides and its adaptability to the needs of the business.
- Outline of the project methodology:
 - Requirements
 - Design
 - Development and Coding
 - Integration and Testing
 - Implementation and Deployment
 - Review

Funding Requirements

Description	Payment Frequency	Cost
Data set to transfer and clean	One time	\$500
System Development	One time	\$5000
AWS Hosting	1/year	\$1000
Employee Training	One time	\$1000
Support	One time	\$1000
Total One Time		\$8500
Total per Year		\$1000

Data Precautions

No sensitive or protected data will be used in this project.

All the data sets used in this project are public datasets and available for anyone, this is because the mentioned datasets do not have any personal information like names, addresses, SSNs, or medical records.

Developer's Expertise

- Developer's qualifications:
 - Bachelor's Degree in computer science at Western Governors University.
 - 2 years of experience working in automation for USTC Corp.
 - Experience in multiple programming languages and frameworks.
- Qualifications that are related to the needs of the project.
 - The focus of a computer science degree is data and how to analyze and use the data to provide better results. This gives me the capabilities and critical thinking to find the solutions to all the problems mentioned in previous sections.
 - While automation and machine learning are different, my focus while working at USTC was to automate the process of capturing data into the system and reduce the amount of time it took an employee to capture it. I can bring some of those methodologies to the table and help to better train non-technical employees.
 - My experience in multiple programming languages and frameworks gives me the adaptability that Movie Champ requires.

Part B: Project Proposal

Problem Statement

Movie Champ is a struggling business in a decaying industry that is in need of new technology that will bring the business back on track with modern business practices.

The business has recently seen a decline in sales, caused by two major factors that are related to each other:

- The recent pandemic forced many of Movie Champ's regular customers to stay at home.
- The surge of online streaming platforms (also propelled by the pandemic), has caused many potential customers to opt for those services instead.

Movie Champ lacks the following:

- A Data Product that can help the management make decisions and forecast the status of the business.
- A Data Product that can give the marketing team insight into what movie the customers want and when they would like to rent it.
- A Data Product that can help the front desk employees to provide customers with a wide range of suggestions for all their movie renting needs.
- A Data Product that can help the purchasing department determine what movies will be in demand, based on the customer's interests and season.

Customer Summary

- Movie Champ is a movie rental business located at 695 River St, Kernersville, AL 48596. Movie Champ has been in business since 1998 and has always been a family-owned business with 3 locations in the city.
- Why the proposed *application* will resolve the problem successfully:

The Movie Recommendation System will aid Movie Champ by bringing the business back on track with technological advancement.

The Movie Recommendation System will address the following problems:

- The System will provide the management team with insights into customers' preferences, therefore allowing the team to forecast patterns of customer spending and customer retention, based on customer satisfaction.
- The System will provide the marketing team with an insight into customer interests, therefore enabling the team to focus their manpower and efforts on sending targeted marketing to increase customer retention.
- The System will provide the front desk employees with the necessary tools to increase productivity and optimize the customer service relation by providing the customer with personalized movie suggestions.

- The system will provide the purchasing department with an insight into customer interests, therefore enabling the team to forecast customer interest and focus the purchasing of new movies in those interests.

Existing System Analysis

- What application or tool the client currently uses:

Movie Champ currently uses SQL Database to keep track of the rental movies and return dates.

- Shortcomings of this current technological environment:

The data is only used to keep track of sales and return dates. Movie Champ employees have no way to analyze the data rendering it useless for the purpose of business forecast and marketing strategies.

Data

- Descriptions of:
 - The Raw Data is two data sets:
 - Movies data: 62423 rows. 3 Columns: movieId, title, and genres.
 - Rating data: 25000095 rows. 4 Columns: userId, movieId, rating, and timestamp.
 - How data will be collected, processed, and managed throughout the application development life cycle: design, development, maintenance, or others:

The 2 data sets provided will be used to test and train the Movie Recommendation System.

Data collection and expansion will be the responsibility of the employees of Movie Champ, this will make the movie suggestions more accurate.

Data will be processed and managed in the following way:

- The movies.csv file will be imported as a Pandas DataFrame
- The ratings.csv file will be imported as a Pandas DataFrame
- Two new columns will be added to the movie's DataFrame: cleanTitle and cleanGenres. These columns will be added by removing all the special characters from the title and genre columns.
- A Matrix of Vectors will be created from the cleanTitle column and the cleanGenres column of the movie's DataFrame.
- Using the Matrix of vectors, a movie matching the user's searched title will be found by using the cosine_similarity function.
- Using the movie ID from the searched movie, a set of similar users that rated the searched movie 4 or higher will be found.

- Using the similar users' set, a set of movies rated 4 or higher by those users will be found.
- Using the found movies set, all the users who rated those movies 4 or higher will be found.
- Get the percentage of similar users that rated all the movies 4 or higher and narrow down the set to only users who liked more than 10 percent of the movies.
- Get the percentage of all users that rated all the movies 4 or higher and narrow down the set to only users who liked more than 10 percent of the movies.
- Using all the users set and the similar users set, a similarity score will be calculated by dividing the similar users' percentage, by all the users' percentages.
- Using the similarity score the final DataFrame will be sorted and the top 10 movies with the higher similarity score will be returned (Processed Data).

How data anomalies will be handled:

In the case where the movie ratings are not reliable enough to make a prediction, the System has a fallback function that uses the movie's genres to look for the 5 more similar movies by using the matrix of vectors `tfidf_genres`.

In the case where the movie ratings are not reliable enough to make a prediction and the searched movie is missing the genres information, the system has a fallback function that uses the movie's title to look for the 5 more similar movies by using the matrix of vectors `tfidf_title`.

Project Methodology

- Industry-standard methodology to be used to develop and deploy the application:

The Agile methodology will be used. Agile is a project management framework that will break the project down into several dynamic phases, known as sprints. The Agile framework is an iterative methodology. After every sprint, we will reflect and look back to see if there is anything that could be improved so we can adjust the strategy for the next sprint.

Agile is adaptable and since any project is prone to changes, especially when introducing a new system for a business, this is a huge benefit. With agile we will foster collaborative teamwork, looking for better ways to improve communication and working face to face.

Agile focuses on customer needs, therefore, is a perfect methodology for a system that has customer service in mind.

- Planned development of the application in each phase of the methodology:
 - Requirements:
 - Approval of this Project Proposal
 - Access to the Raw Data from Movie Champ servers
 - Design:

We will have a meeting to debate about desired UI/UX Design and backend functionalities.

In this meeting, a preview of the data sets will be presented to the development team, and a mockup of the system will be created.

The idea of merging the data sets and Movie Champ's database will be discussed.

- Development and Coding:

Translating the design ideas into code. This will be a very time-consuming task, so it will be necessary to present progress reports periodically.

- Integration and Testing:

Testing the System to ensure compatibility with the data sets, and the database obtained from Movie Champ and checking for any bugs.

- Implementation and Deployment:

The System will be deployed to the AWS cloud and made available to Movie Champ employees.

- Review:

Meeting with the development team to discuss progress and achievements. Evaluating if the goals and objectives were fulfilled.

Project Outcomes

- Deliverables:

- Project Proposal:

This Finished project proposal with all the requirements.

- Progress Report:

Multiple progress reports will be delivered to the development team. This will ensure communication and the timely process of change requests.

- Data Transfer:

This is one of the requirements, it needs to be done before the design meeting.

- The Movie Recommendation System:

Deliver the final version of the System after all the tests and integration with the Movie Champ system have been performed.

- A User's guide:

Provide a user guide and other training materials so Movie Champ's employees can prepare before the full integration of the new System is finished.

- Post-Implementation Report:

Final report explaining the changes performed and how the system has fulfilled the initial goals and objectives.

Implementation Plan

- Outline of how the project will be implemented:

- General strategy:

Once the Movie Recommendation System finishes going through the test and integration phases, the development team will hold a progress meeting to determine the best time to implement the new system in the business's computers.

Once the start date is set, the rollout will consist of two phases:

First, the rollout will include the Marketing, Management, and Purchasing departments. These three departments need the predictions for almost the same reason, so this rollout phase will ensure that all these departments are in constant communication and support each other.

The second phase of the rollout will include all the front desk employees. This will also allow for constant communication and support from every team member. Employees who have already finished their training can cover the front desk activities to ensure the continuity of business operations.

- Phases of the rollout:

Since the data will be used to generate movie recommendations, the normal operations of the business should not be affected by the rollout of the Movie Recommendation System.

- Marketing, Purchasing, and Management phase: The first ones to receive the system rollout will be the marketing team. This will serve as a test to evaluate the efficacy of the data that these 3 groups will be using as a guide for the decisions. This phase of the rollout will last 1 week.
- The front desk employees' phase: During this phase, we will use some of the employees from the marketing and purchasing departments to cover the normal operation of the front desk. Employees from the front desk will have time to learn about the new system and learn how to use the movie recommendations to better serve the customers.
- This phase of the rollout will last 1 week.

- Dependencies:

- A date for deployment must be agreed upon before starting the rollout.
- The continuity of business operations must be prioritized.

- Details for testing and distribution:

- Employees who are learning the new system will input several of their favorite movies so they can provide feedback on the recommendations.

- Once all the employees are trained, they will keep providing feedback when a customer asks for an odd movie. This will provide real-world feedback that is difficult to get in the testing phase.

Evaluation Plan

- Verification method to be used at each stage of development:

Feedback from the development team will be used as a verification method to test if the movie recommendations provided by the algorithm are accurate or not.

The verification will be assessed in three categories:

- The similarity of genres:

If the genres are similar, then it is more likely that the recommended movies are more relevant to the user.

- The similarity of the years the movies were released:

Although the year the movie was released does not affect the predictions in any way, there is a correlation between movies from closer years that have been recommended.

- The similarity of the title:

If a person liked the first movie from a series, then it is more likely that the sequels will be of interest to that person.

- Validation method to be used upon completion of the project:

The validation method at the end of the project will be the same format as before, but this time the test subjects will be the Movie Champ employees who will receive the training in the system.

Resources and Costs

The price change from the project proposal since the age of the computers from the stores was unknown at that time.

Resources	Payment Frequency	Cost
Old data set transfer	One time	\$250
Old data set clean and merge	One time	\$250
System Development	One time	\$5000
Updated PCs (3 pieces)	One time	\$1500
AWS Hosting	1/year	\$1000

Employees Training Group 1	One time	\$500
Employees Training Group 2	One time	\$500
Support	1/year	\$1000
Total One Time		\$10000
Total per Year		\$2000

Timeline and Milestones

Activity	Start date	End date
Present the project proposal to the Business Executives	12/05/2022	12/07/2022
Present the project proposal to the IT professional leadership group	12/08/2022	12/09/2022
Have a kickoff design meeting with the development team	12/12/2022	12/16/2022
Transfer the data from Movie Champ servers	12/19/2022	12/23/2022
Start development of the system	12/26/2022	01/20/2023
Present a progress report to the development team	01/23/2023	01/27/2023
Finish development of the system	01/30/2023	03/10/2023
Present a progress report to the development team	03/13/2023	03/17/2023
Start testing and integration	03/20/2023	03/31/2023
Have a meeting to schedule implementation	04/03/2023	04/07/2023
Start implementation with group 1	04/24/2023	04/28/2023
Start implementation with group 2	05/01/2023	05/05/2023
Test and review	05/08/2023	05/19/2023

Part C: Application

List of submitted files:

- “Movie Recommendation Engine.ipynb”
- “movies.csv”
- “ratings.csv”

Part D: Post-implementation Report

A Business Vision

- Problem Description:

Movie Champ is a movie rental business that was struggling with the growing competition in the form of online streaming platforms. Movie Champ used to lack employees with the right skills and knowledge to provide optimal customer support. The marketing team of Movie Champ used to lack a customer focus approach to their strategies.

- How the Movie Recommendation System solved the problem:

The Movie Recommendation System has provided Movie Champ employees with the necessary tools to compensate for the lack of knowledge and they are now offering better customer support.

The marketing team from Movie Champ has used the movie recommendations of every user as the foundation for all their marketing campaigns.

The purchasing department has used the movie recommendation to determine which movies will be more in demand.

- How the user can use the Movie Recommendation System to solve the problem:

Example: John is a customer of Movie Champ for 10 years. John is interested in old movies from the 60s. Tom is an employee of Movie Champ. Tom is assisting John, but Tom does not know much about movies from the 60s. Tom uses the Movie Recommendation System to recommend a movie to John. The marketing team uses John's recommendations and sends an email to John offering him a coupon if he decides to rent the sequel of the movie that he rented.

Datasets

- Raw and processed data:
 - Raw Data:
 - Movies data: 62423 rows. 3 Columns: movieId, title, and genres.
 - Rating data: 25000095 rows. 4 Columns: userId, movieId, rating, and timestamp.
 - Processed Data:
 - Recommended movies: 10 rows. 3 Columns: similairtyScore, title, genres.

- Example of the raw and processed data:

```
In [3]: #Previewing the movies data
#Descriptive method
movies.head()
```

Out[3]:

	movielid	title	genres
0	1	Toy Story (1995)	Adventure Animation Children Comedy Fantasy
1	2	Jumanji (1995)	Adventure Children Fantasy
2	3	Grumpier Old Men (1995)	Comedy Romance
3	4	Waiting to Exhale (1995)	Comedy Drama Romance
4	5	Father of the Bride Part II (1995)	Comedy


```
In [4]: #Previewing the ratings data
#Descriptive method
ratings.head()
```

Out[4]:

	userid	movielid	rating	timestamp
0	1	296	5.0	1147880044
1	1	306	3.5	1147868817
2	1	307	5.0	1147868828
3	1	665	5.0	1147878820
4	1	899	3.5	1147868510

Figure 1 Raw Data

Movie Title

'Searched movie: Avengers Age of Ultron 2015'

''

'Recommendations:'

	score	title	genres
25058	43.782728	Avengers: Age of Ultron (2015)	Action Adventure Sci-Fi
25073	27.025799	X-Men: Apocalypse (2016)	Action Adventure Fantasy Sci-Fi
20513	25.491917	Thor: The Dark World (2013)	Action Adventure Fantasy IMAX
21454	24.751835	The Amazing Spider-Man 2 (2014)	Action Sci-Fi IMAX
53867	23.601189	Ant-Man and the Wasp (2018)	Action Adventure Comedy Fantasy Sci-Fi
20018	22.895555	Wolverine, The (2013)	Action Adventure Fantasy Sci-Fi
25061	22.802299	Ant-Man (2015)	Action Adventure Sci-Fi
25071	21.473613	Captain America: Civil War (2016)	Action Sci-Fi Thriller
19678	21.189794	Iron Man 3 (2013)	Action Sci-Fi Thriller IMAX
27980	20.332400	Furious 7 (2015)	Action Crime Thriller

Figure 2 Processed Data

- Access to any datasets used:
 - The following files are included:
 - “movies.csv”
 - “ratings.csv”

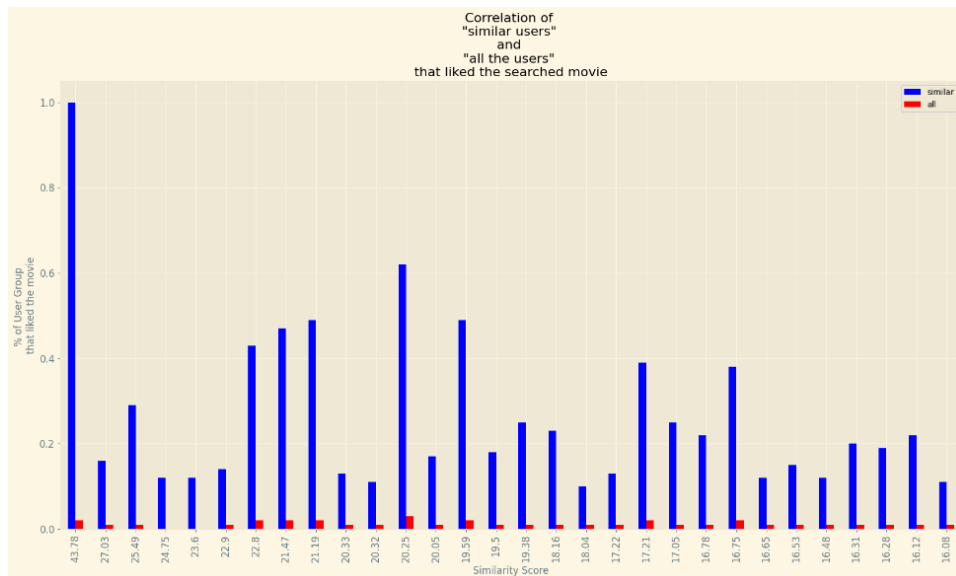
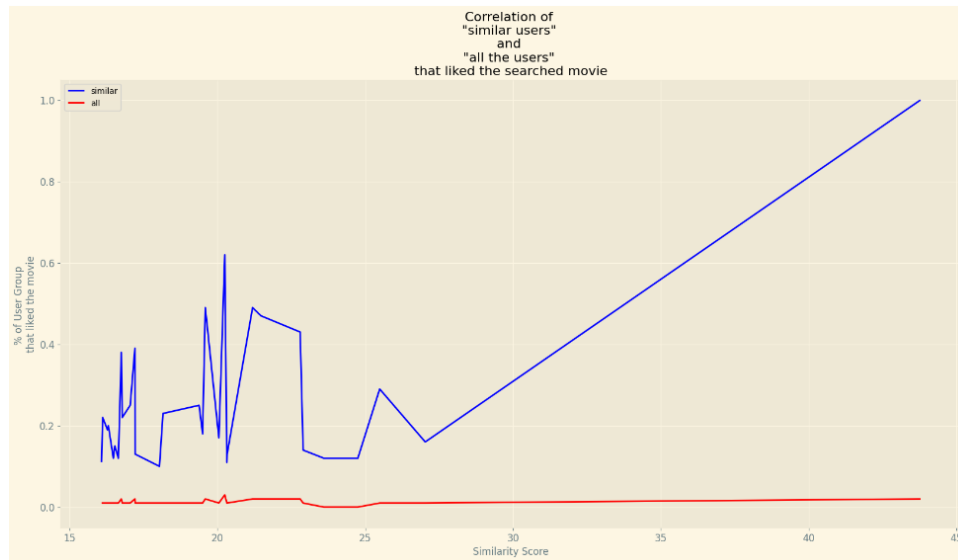
Data Product Code

- Review the code’s functionality used to perform the analysis and develop your application.
 - Processing raw data:
 - The movies.csv file was imported as a Pandas DataFrame
 - The ratings.csv file was imported as a Pandas DataFrame
 - Two new columns were added to the movie's DataFrame: cleanTitle and cleanGenres. These were added by removing all the special characters from the title and genre columns.
 - A Matrix of Vectors was created from the cleanTitle column and the cleanGenres column from the movie's DataFrame.
 - Using the Matrix of vectors, a movie matching the user’s searched title was found.
 - Using the movie id from the searched movie, a set of similar users that highly rated the searched movie was found.
 - Using the similar users' set, a set of movies highly rated by those users was found.
 - Using the found movies set, all the users who highly rated those movies were found.
 - Using all the users' sets and the similar users' sets, a similarity score was calculated.

- The top 10 movies with the higher similarity score were retrieved (Processed Data).
- Descriptive methods and visualizations:

Correlation of how the percentage of similar users who like a specific movie and the percentage of all the users who like the same movie helps makes the recommendation more accurate.

If the difference between the percentages is greater then the similarity score will be greater, making that movie a more relevant recommendation.



- Non-descriptive method:

Prediction of movies that the user will like based on the similarity score.

Movie Title

'Searched movie: Avengers Age of Ultron 2015'

..

'Recommendations:'

	score	title	genres
25058	43.782728	Avengers: Age of Ultron (2015)	Action Adventure Sci-Fi
25073	27.025799	X-Men: Apocalypse (2016)	Action Adventure Fantasy Sci-Fi
20513	25.491917	Thor: The Dark World (2013)	Action Adventure Fantasy IMAX
21454	24.751835	The Amazing Spider-Man 2 (2014)	Action Sci-Fi IMAX
53867	23.601189	Ant-Man and the Wasp (2018)	Action Adventure Comedy Fantasy Sci-Fi
20018	22.895555	Wolverine, The (2013)	Action Adventure Fantasy Sci-Fi
25061	22.802299	Ant-Man (2015)	Action Adventure Sci-Fi
25071	21.473613	Captain America: Civil War (2016)	Action Sci-Fi Thriller
19678	21.189794	Iron Man 3 (2013)	Action Sci-Fi Thriller IMAX
27980	20.332400	Furious 7 (2015)	Action Crime Thriller

- For the non-descriptive portion, describe:

- What analytic methods were applied?

The correlation between the movies that all the users liked and the movies that only the similar users liked was used to calculate the similarity score. This process gave a higher similarity score to the movies that are more relevant to the user's input and not only the ones with the highest ratings. Therefore, this predicted the movies that the end user would like the most.

- Why these methods are appropriate for the project:

Because only considering genres or title similarity would make all the predictions less relevant for the end user.

- How they were trained and tested.

A tool to get the predictions based on genres or title similarities was created for testing, in every instance, the movies with the higher similarity score assigned by the algorithm were more relevant.

- How the data analysis supported the choosing and improving of your descriptive and non-descriptive methods?

By analyzing the data, we could determine the relationship between the percentage of similar users who liked a set of movies and all the users who liked the same set of movies, helping make the recommendations more accurate.

- Sources necessary for developing the project:
 - <https://drive.google.com/drive/folders/1dZ96cKZMeuMJt8Jm4SCL36ER7eLg9jmN?usp=sharing>
 - The shared folder includes three files:
 - movies.csv
 - ratings.csv
 - Movie Recommendation Engine.ipynb

Hypothesis Verification

- Project's hypothesis:
The Movie Recommendation System will deliver custom accurate recommendations for most Movie Champ customers.
- Why the hypothesis was met:

With the exception of some movies with very few ratings, most Movie Champ customers have had an accurate movie recommendation.

Effective Visualization and Reporting

- Describe how the descriptive method and visualizations supported your non-descriptive method development process. Items discussed should include:
 - Data exploration:

Previewing both data sets helped me determine the correlation between the two and the order in which each part of the data was going to be accessed.
 - Data analysis:

Analyzing the data was the most important step to train the model. The training mostly was accessing different parts of the data sets, each step depending on the results from the previous step.
 - Data summary:

Based on the preview of the data and the analysis, a result containing the movie predictions was obtained.
 - Analysis application of three visualizations:
 - The tables help me preview the data sets and determine the course of action for data analysis.
 - The line chart helped me see the correlation between similar users that liked a set of movies and all the users that liked the same set of movies. In most cases when one value was going up, the other one also was going up. The lines never overlapped each other.
 - The bar chart helped me to observe the correlation between similar users that liked a set of movies and all the users that liked the same set of movies. In all the

cases the bigger the difference between the two values of the bars the bigger the similarity score would be.

Accuracy Analysis

- Metric used to assess your model:

The feedback from Movie Champ employees was used as a metric to assess the accuracy of the Movie Recommendation System. During the deployment and training phases, each employee was assigned a list of movies to search and provide feedback if the movie recommendations were accurate or not.

- Assessing the accuracy of your non-descriptive method:

The accuracy of the prediction is at least 85%, this is based on the recommended movies obtained during testing. Almost all the time the movies were closely related, even when old movies with fewer ratings were selected.

- Example demonstrating the non-descriptive method and discussing the accuracy:

The searched term was: “Lion King The 1994.”

The first movie is always the search term, which indicates how accurately the System looks for a movie fitting the description. Any other result has very similar genres, and years and most of them were created by the same company.

Movie Title

Lion King The 1994

'Searched movie: Lion King The 1994'

..

'Recommendations:'

score	title	genres
359 6.122578	Lion King, The (1994)	Adventure Animation Children Drama Musical IMAX
1818 3.824691	Mulan (1998)	Adventure Animation Children Comedy Drama Musical
999 3.789861	Cinderella (1950)	Animation Children Fantasy Musical Romance
587 3.713947	Beauty and the Beast (1991)	Animation Children Fantasy Musical Romance IMAX
1992 3.584819	Little Mermaid, The (1989)	Animation Children Comedy Musical Romance
580 3.503659	Aladdin (1992)	Adventure Animation Children Comedy Musical
586 3.445692	Snow White and the Seven Dwarfs (1937)	Animation Children Drama Fantasy Musical
588 3.440762	Pinocchio (1940)	Animation Children Fantasy Musical
1989 3.385253	Jungle Book, The (1967)	Animation Children Comedy Musical
313 3.339588	Santa Clause, The (1994)	Comedy Drama Fantasy

Application Testing

- How the application was tested:

Many terms were imputed to determine if the recommendations were accurate or not.

- how the testing results were used to improve the application:

Some of the results at the beginning were not very accurate, this led me to include the TfidfVectorizer function, so the search would look at sets of 2 words instead of just word by word. This increased the accuracy.

Application Files

- Hierarchical list of files and libraries required to execute your application through a Windows 10 machine:
 - Files:
 - \movies.csv
 - \ratings.csv
 - \Movie Recommendation Engine.ipynb
 - Libraries
 - Python3
 - Pandas
 - Ipywidgets
 - Matplotlib
 - Sklearn
 - IPython
- How the files are organized in the submission:

The files are in the Google Drive Folder named C964. All three files are in the root of the folder.

User Guide

Download Project Files

1. Go to <https://drive.google.com/drive/folders/13YqxRV2ZOWWA9We5YCrJ-TaOMWTynwkP?usp=sharing> and download all the project files.

Instructions to install Python 3

2. Go to <https://www.python.org/ftp/python/3.11.0/python-3.11.0-amd64.exe> on your preferred web browser and download the Python3 installer
3. Go to your downloads folder and double-click on the file named python-3.11.0-amd64.exe
4. Follow the instructions on the installer.

Instructions to install Jupyter Notebook

5. Open the Command Prompt application.
6. Type the following commands, press ENTER after each command, and wait for each command to finish before typing another one:
 - pip install notebook
 - pip install pandas
 - pip install numpy
 - pip install ipywidgets

- pip install matplotlib
 - pip install ipython
 - pip install -U scikit-learn
- 7. Run the command: jupyter notebook
- 8. A new web browser window will open.
- 9. In the web browser click on the Downloads Folder.
- 10. Within the Downloads folder look to the folder that was downloaded in step 1 and click on it.
- 11. Click on the file Movie Recommendation Engine.ipynb
- 12. A new tab will open.
- 13. Go to the "Kernel" menu and "restart and run all."
- 14. Wait for the process to finish.
- 15. Scroll all the way down.
- 16. In the Movie Title Search Box type a movie title.
- 17. When the menu populates select one of the movies.
- 18. The movie recommendations will display at the bottom of the page along with the charts.

Summation of Learning Experience

- How your previous experience readied you for this project:

Ever since I got a promotion at work last year, I have been dealing with more of the corporate world, and this project is a clear example of how it works. My academic experience has greatly helped me to improve my communication and critical thinking skills, and those skills have been essential for this project.
- Additional learning or resources needed to complete this project:

I always try to find examples of anything I'm about to do on YouTube, and this project was no different, the visualization of my ideas is what help me to complete this project.
- How this experience contributed to your concept of lifelong learning:

By expanding my understanding of what computer science is and making me realize how the possibilities to keep going forward are endless with this degree.