**WGU C964**

**Computer Science Capstone**

**Project Documentation**

**-------------------------------------------------------------**

**Joshua Thompson**

**Student ID # 000992212**

**9/23/2023**

# Table of Contents

|  |  |
| --- | --- |
| **A. Project Proposal for Business Executives** | |
| Letter of Transmittal | pg. 4 |
| Project Recommendation | pg. 5 |
| Problem Summary | pg. 5 |
| Application Benefits | pg. 5 |
| Application Description | pg. 5 |
| Data Description | pg. 5 |
| Objectives and Hypothesis | pg. 5 |
| Methodology | pg. 5 |
| Funding Requirements | pg. 6 |
| Data Precautions | pg. 6 |
| Developer’s Expertise | pg. 6 |
| **B. Project Proposal** | |
| Problem Statement | pg. 7 |
| Customer Summary | pg. 7 |
| Existing System Analysis | pg. 7 |
| Data | pg. 7 |
| Project Methodology | pg. 8 |
| Project Outcomes | pg. 8 |
| Implementation Plan | pg. 8 |
| Evaluation Plan | pg. 8 |
| Resources and Costs | pg. 9 |
| Timelines and Milestones | pg. 10 |
| **C. Application – pg. 10** | |
| **D. Post-Implementation Report** | |
| Business Vision | pg. 11 |
| Datasets | pg. 11 |
| Data Product Code | pg. 11 |
| Objective Verification | pg. 12 |
| Effective Visualization and Reporting | pg. 12 |
| Accuracy Analysis | pg. 13 |
| Application Testing | pg. 13 |
| Application Files | pg. 13 |
| User Guide | pg. 14 |
| Summation of Learning Experience | pg. 14 |

# 

# Part A: Project Proposal for Business Executives

Joshua Thompson

Code Conjurers Development Solutions

893 Church Street

Rockledge, Florida, 32955

jtho840@wgu.edu

407-910-3283

09/23/2023

William Harrington

Timeless Pages Booksellers

4551 Roosevelt Blvd.

Jacksonville, Florida, 32210

Dear Mr. Harrington,

I am writing today in response to your request for development of Timeless Pages’ new online storefront. For the last 50 years, Timeless Pages has been one of the premier bookstores in the state of Florida, and I have been a happy customer of your brick and mortar store since I was a child. With your expansion into the digital marketplace, Code Conjurers Development Solutions is proud to be able to offer our assistance in making your online presence the strongest it can be.

To that end, we are here to offer development of a machine learning algorithm that will enhance your online business by using a customer’s purchase history to find the books in your inventory that they are most likely to be interested in, and then recommend them to the customer either as they go to check out or as a marketing email at regular intervals.

The obvious benefit to this solution is the enhancement to your marketing strategy. By targeting advertisements to your customers’ interests, they are much more likely to engage with those ads, which will in turn lead to more sales and increased revenue. It will also strengthen your reputation as the most prolific bookstore in the state, as it will expose your customers to the depths of your vast inventory and show them books they did not know you had, and books they did not know they wanted.

As a graduate of Western Governors University’s prestigious computer science program, I am equipped with cutting edge professional skills to develop this solution for you. I have spent the last five years developing software for the federal government, and my resume and references speak for themselves. My attached budget requirements will outline the cost to develop the solution (with a total estimated budget of 146,719.00), and the increased revenue that this solution will generate for Timeless Pages will make that budget even more appealing. Thank you for this opportunity, please reach out when you wish to collaborate to implement this solution. I look forward to hearing from you soon.

Sincerely,

Joshua Thompson

Code Conjurers Development Solutions

## Project Recommendation

### Problem Summary

The goal of this project is to create a machine learning algorithm that will use a customer’s purchase history to determine which books in Timeless Pages’ inventory they will be most likely to purchase, so they can be used for targeted advertising to that customer. With Timeless Pages’ new online storefront allowing for expansion into the digital marketplace, targeted advertising will help to replace the sensation of browsing the stacks in a brick and mortar bookstore, leading to more purchases in the online bookstore and the reputation of Timeless Pages as the place where any book can be found.

### Application Benefits

By implementing our solution, our preliminary research has shown that Timeless Pages will be able to increase engagement with their online listings by 20 percent within the first three months of deployment. Of these additional engagements, we believe that a further 50 percent of those engagements will lead to at least one additional sale. With additional market research, we will be able to refine our metrics to increase these results even further.

### Application Description

The application uses item-based collaborative filtering to create a similarity score between a customer and each book in Timeless Pages’ inventory. Then, the items with the highest similarity score will be provided as output for use in targeted advertising. The program does this by using a series of tags based on relevant information from each book in Timeless Pages’ inventory, such as the author, genre, subject, characters, ect. The tags from the books in the customer’s recently purchased history are compared to the tags of each book in the inventory, and based on the number of common elements, a similarity score is assigned to each item. Further details will be provided further down in this document.

### Data Description

The raw data for this solution comes entirely from within Timeless Pages’ database, consisting of qualitative data from the books in the inventory and the most recently purchased books from any given customer. The primary advantage of this is that we do not need to gather data from outside sources, which reduces overhead and allows for greater security, as no third parties risk coming into contact with proprietary information about the business. Another advantage to this data structure is that the data for each book in the inventory only needs to be prepared once and then stored in the database for when the algorithm is used, greatly improving performance. The only variable that needs to be considered are the books purchased by a customer, as the algorithm compares to every book in the inventory when run.

### Objectives and Hypothesis

Our objective is to identify the books that they are most likely to be convinced to buy and enjoy and push them as recommendations, thereby increasing sales and strengthening customer satisfaction.

### Methodology

We will use the Agile methodology to develop this solution. We prefer to use Agile for all of our software development projects due to its emphasis on flexibility from being able to adjust sprint goals as other sprints are completed and the emphasis on quality that results from being able to test our work as we create it. A tenative timeline of the initial development process is listed in Section B.

### Funding Requirements

Because Code Conjurers Development Solutions is an independent firm, we already have much of the required hardware and software for development in house. The cost of our services includes the use of our physical and digital assets for development, as well as the cost of labor. An itemized budget can be found in Section B of this document on page ##.

### Data Precautions

The data that we will be using for this solution will come from Timeless Pages’ own database, and as such, no additional security measures are needed to work with this data during this project. We will work with the security standards already in place for the online storefront and its data, including access restriction to anybody without a need-to-know, encryption protocols for transmitting customer data, and adherance to any additional security measures that Timeless Pages’ IT department already has in place for its use.

### Developer’s Expertise

As the owner and operator of Code Conjurers Development Solutions, I am a highly independent worker who’s sole focus will be on producing value for you, my client. I earned my Bachelor’s Degree in Computer Science from Western Governors University and founded Code Conjurers as a way to provide my services to a variety of clients while maintaining a consistent development process and style. My primary client has been the Department of Homeland Security, developing numerous data tracking programs and systems that use machine learning processes to process data and provide insights into employment processes for my clients.

# Part B: Project Proposal

The project proposal should target your client’s technically savvy IT (Information Technology) professional leadership. Use appropriate industry jargon and sufficient technical details to describe the proposed project and its application. Remember, you’re establishing the technical context for your project and what it will accomplish for the client. Typically, this section is 8 – 10 pages. **Write everything in the future tense.**

## Problem Statement

With a lack of targeted advertising in Timeless Pages marketing strategy, the company is not able to reach the fullest potential of its revenue stream and overall business capture. By implementing the proposed algorithm, we will aim to solve that problem and help the company reach its fullest potential in the digital marketplace.

## Customer Summary

Timeless Pages is one of the most popular independent bookstores in the state of Florida, having been founded in 1973 and working to expand the business continually to the present day. They have achieved incredible success as a brick and mortar store by buying and selling any and all kinds of books and gaining a reputation for great selection and prices as a result. This reputation in particular has made them a pillar of the community, attracting a degree of tourism to the city in addition to their local customers. In light of their spreading reputation, Timeless Pages implemented an online storefront five years ago, allowing customers all over the world to buy and sell books from their inventory.

Our proposed application will help expand their reputation to their online presence by providing a digital replacement for the aspects of their in person business that cannot be brought directly to it – the knowledgeable staff who can recommend books to customers in person, and the impulse purchase of other books from browsing the shelves.

## Existing System Analysis

Currently, Timeless Pages uses an offline POS system in their store and an Oracle Database to maintain their inventory. With the opening of their online storefront, they have been using Shopify to power their e-commerce presence, connected to their database for inventory management and order preparation. Our solution will increase the power of the storefront to generate revenue, and in the future we hope to be able to create a fully overhauled storefront that will implement our machine learning output into targeted advertising in the storefront itself. Unt

## Data

The raw data for this project will be collected from the database of Timeless Pages itself. This proposal includes several CSV files that simulate the tables of the database, prepared in the appropriate views to service this data solution. The data will be collected automatically by the program through SQL calls and processed in the program itself. Data anomalies will be addressed manually as they are encountered, and if a common cause of multiple anomalies is found, it can be addressed then.

## Project Methodology

Code Conjurers will be using the Agile methodology to undertake this project for the reasons listed on page ## of Section A of this document. Our tenative sprint schedule for the intial development phase is shown below. Future sprints will need to be planned as the project commences, to adapt to needs that arise as we get underway.

Project Outcomes

The primary deliverable of this project will be the machine learning algorithm and its future integration into Shining Pages online storefront. Once the solution is developed, future deliverables will be adjustments to the algortihm to improve effectiveness and implementation of the algorithm into the online storefront.

## Implementation Plan

After the initial development phase, we will being a testing phase where we will use the output of the algorithm to produce targeted advertising and collect feedback on its effectiveness in a real world scenario. Based on this feedback, we will adjust the algorithm to maximize the effectiveness of its reach. Once we reach our target metrics, we will work to implement targeted advertising directly into the online storefront to customers as they purchase books.

## Evaluation Plan

The most important verification method to use for this project is market research that will be conducted after initial implementation is complete. Using the Agile methodology, we will be able to conduct testing of the software as we go through the development process, thereby evaluating the effectiveness of the process with minimal wasted time and resources.

## Resources and Costs

|  |  |  |
| --- | --- | --- |
| **Resource** | **Description** | **Cost** |
| **Personnel Costs** | | |
| Machine Learning Engineer | Full time employee, annual salary of 124,625.00 for development and maintenance of this and future machine learning projects. | 124,625.00 |
| Total | | 124,625.00 |
| **Technology Costs** | | |
| Computers and Miscellaneous Hardware | Laptop computers, storage devices, networking devices. (Usage of Code Conjurers’ in-house hardware) | 2,500.00 |
| Cloud Hosting | Microsoft Azure Cloud hosting for 12 months. (Can be integrated with the existing cost of hosting the online storefront.) | 3,756.00 |
| Development Software | Cost of development software and database management utilities, as well as any relevant licenses for use. (Usage of Code Conjurers’ in-house software) | 2,500.00 |
| Total | | 8,756.00 |
| **Subtotal** | | 133,381.00 |
| Contingency Fund (10% of budget subtotal) | | 13,338.00 |

## Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| **Sprint** | **Start** | **End** | **Tasks** |
| 1 | 10/9/2023 | 10/10/2023 | Kickoff meeting, planning machine learning algorithm |
| 2 | 10/16/2023 | 10/27/2023 | Perform primary development of algorithm. |
| 3 | 10/30/2023 | 11/3/2023 | Create data pipeline and database, integrate into AI framework. |
| 4 | 11/6/2023 | 11/10/2023 | Collect and compile data into database, begin testing data with algorithm. |
| 5 | 11/13/2023 | 11/17/2023 | Final algorithm testing, submission to sponsor for approval. |
| 11/20/2023 | | | Milestone 1 – Completion of Algorithm |
| 6 | 11/20/2023 | 1/20/2023 | Engage in market testing and adjust algorithm to meet market trends. |
| 7 | 1/21/2023 | 2/15/2023 | Work with IT to integrate algorithm into online storefront. |
| 8 | 2/17/2023 | 2/28/2023 | Creation of online advertising based on output of algorithm. |
| 3/1/2023 | | | Milestone 2 – Integration into Online Storefront |

Due to the nature of the Agile methodology, this schedule is tentative, and should be expected to change as the project is undertaken. However, the milestones set here are reasonable by our standards and we expect to meet them, pending any major setbacks.

**Part C: Application**

My application and all of its requisite files, including screenshots of the program in use and simulations of its data, are attached with this submission in JDTCapstoneProject.zip . The program can be started by running main.py from the base folder. The CSV files in the included folder simulate the database discussed in Part B.

# Part D: 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 problem that this solution was designed to address was the lack of targeted advertising in the online storefront of Timeless Pages booksellers. The solution that I have implemented will solve that problem by creating data driven targeted advertising that can be integrated into the business’s online storefront and marketing campaigns.

Ideally, the solution would be implemented into the online storefront. As a customer logs onto the site to order a book, the program will check their order history and prepare the results in the background. Then, while they browse, the recommended books can be shown to them, either as search results or inline advertisements.

## Datasets

The data that was used by the application consisted of the twenty most recently purchased items in a customer’s history and the inventory of Timeless Pages. In order to use the data, some manual manipulation was neccesary. In order to create the Tag items in the inventory database, I took the synopsis of each book in the inventory, filtered out key words that could act as similarities between multiple books, then added them to each row as a separate entry.

I chose to do this ahead of time instead of in the algorithm itself for the sake of efficiency. For each book, the tags were not going to change once chosen, so it made sense to add them to the database as strings that the algorithm could search for.

By adding the tags to the inventory database directly, I also simplified the process of compiling the user’s preference profile. I used a matrix made of nested list items to grab the tags of each book in the customer’s purchase history. Then, since the tags came from the same source, it made the process of comparing them to create a similarity score even more streamlined.

The data that I used is in the CSV files included with my submission, as outlined below.

## Data Product Code

The product that I have created follows the following procedures. First, the program asks for a customer ID number, in order to begin the program. If the user inputs a non integer or a text string, the program outputs an error message and prompts them to try again.

Once the program has a valid input, it searches the customer database for that customer. If it fails to find them, the program returns another error message and again prompts the user for a customer ID. If it finds the customer, it takes the book ID’s of the last twenty books that customer purchased and stores them in the first column of a matrix (a set of nested lists). (Each customer’s purchase history is stored in a table with their ID number.)

With the book ID’s, the program then looks at the data and collects a set of relevant information about each book in the customer’s history, such as the author, genre, and the seven plot summary tags. These values are stored in each row of the matrix with the book ID. This filled in matrix is the customer’s similarity profile.

Once the similarity profile is completed, the program then proceeds to compare the contents of each book to each item in the book database. Each book is assigned a number of points based on how many of the book’s characteristics are found in that customer’s similarity profile. For example, a book receives 8 points for each book by that author in their profile, 8 points for matching genres, and 2 points for each plot summary tag that is found in common. (These point values are arbitrary, based on my perspective of what would draw a customer’s attention. These values can be changed based on market research.) Each book’s score is stored in a list with the book ID.

Once each book has a score, the list is sorted to pull the books with the highest scores to the top of the list. The top three books are extracted from the list, along with the similarity score, and the program uses the book ID to find the title and author. Lastly, the title, author, and similarity score of each of the top three books are printed to the CLI as output, and the program then resets, so that another customer ID may be input and the program can run again.

## Objective (or Hypothesis) Verification

The objective of the project was to create targeted advertising based on previously purchased items in a customer’s history. Based on the output of the program, I have met this objective. With more data and more tags to compare, I believe the product can be improved while using the same basic processes.

## Effective Visualization and Reporting

The descriptive method in my product contributed primarily to the efficiency of my product and its non descriptive method. I had originally considered making the program create the item tags for each book when it ran, but it became quickly apparent that that procedure would be extremely inefficient. By creating the similarity tags manually and adding them to the database as item entries for each book, I reduced the amount of time needed for the program to run, and made it much easier for the non-decriptive method to predict the similarity scores and choose the books for recommendation.

## Accuracy Analysis

The metric that I used to test the accuracy of the product was my own perspective of the elements of the books in the database, and how closely they lined up with what I, as a consumer, would want to see in a targeted advertisement. In a real world scenario, the output would need to be subjected to market testing and analysis to determine its effectiveness. However, my data product is designed in such a way that adjusting those metrics while maintaining the basic structure would not be difficult.

## Application Testing

The application was tested by running the algorithm across multiple test cases of customers with different histories of purchased books. Based on the testing, I adjusted the metrics used to create the similarity score in order to favor books that I felt were most likely to create output that the customer would be likely to respond to.

## Application Files

The following is a list of the files in my program, including the CSV files that serve as a simulated database. Indented files beneath a folder name are inside of those folders. File names begin with a > and end with a file extention, folder names do not.

JDT\_Capstone\_Project

csvData

purchaseHistories

>150.csv

>151.csv

… (files from 152-158)

>159.csv

>author.csv

>books.csv

>customer.csv

>genre.csv

Snapshots

>customerHistoryMatrixInit.png

(This is a screenshot of the customer similarity matrix, initiated without data.)

>customerHistoryMatrix.png

(This is a simulation of the customer similarity matrix for a customer, populated with data from their history, made in Excel for ease of viewing.)

>primaryIO.png

(This is a screenshot of the program in use, and its output.)

>bookProfile.py

>customerProfile.py

>main.py

>README.txt

## User Guide

1. Extract the contents of JDT\_Capstone\_Project.zip.
2. In the folder, run main.py to start the command line interface.
3. Enter a customer ID number to create a profile and get recommendations for that customer. (Only numbers 150-159 are present in the simulated database, any other integer will not find a customer.)
4. The program will output the recommendations for the customer once their ID is validated. To find another customer’s recommendations, enter their ID when prompted. The search can be repeated indefinitely with different customers.
5. To end the program, enter “Exit” instead of a customer ID. This will close the program.

These instructions can also be found in the README.txt file.

## Summation of Learning Experience

This was the first program that I had ever written in Python without detailed instruction beyond what I wanted to accomplish. I was extremely nervous as I began writing, but as I did, I found elements and lessons I had learned from all of my previous software design classes coming into my mind as I tackled each part of the design process.

Writing the documentation for this project also invoked classes I had taken who’s lessons I had not yet had a chance to put into practice. In particular, I used what I had learned in my Business of IT classes to plan the schedule of what a full scale development of this project would look like, which I found particularly encouraging given that I struggled with Business of IT-Applications.

Overall, I felt very well equipped to accomplish what I set out to accomplish with this project. In fact, I would compare the experience to when I first began my job with the TSA in 2016. After months of training and weeks of being supervised by an OJT coach, my first day on my own felt paralysing. However, all I had to do was get started, and I suddenly realized that my training and practicing had left me perfectly equipped to do the job, and my anxiety was unfounded.

Doing this capstone project, I felt the anxiety of unsupervised development fade quickly as I went. Every challenge I faced, I either knew how to complete, or I knew how to find the solution. The experience confirmed for me that I am ready to receive my Bachelors of Computer Science, and that I am prepared to enter this world and continue to improve and learn to accomplish my full potential.