**Liquor Prediction Program: Capstone**

James Mcgill

C964: Capstone

December 2, 2023

Table of Contents

Letter of Transmittal 3

Project Proposal Plan 5

Project Summary 5

Data Summary 6

Implementation 7

Timeline 8

Evaluation Plan 10

Resources and Costs 10

Post-implementation Report 11

Solution Summary 11

Data Summary 12

Machine Learning 12

Validation 13

Visualizations 13

User Guide 14

[Reference Page](#_Toc1702353417) 17

# Part A: Letter of Transmittal

December 2, 2023

Mr. Barr Mann

Bar Management Strategies

212 Bar Lane Duluth, Georgia 30125

Dear Mr. Mann,

Bar Management Strategies, and bars in general, face a problem when it comes to deciding a price for buying and selling their liquor. Oftentimes, liquor distributors may overcharge and not provide a fair price, or potentially more damaging, a bar may undercharge and lose money on their liquor. This problem is complicated by the wide range of goals for a particular business and the liquor selections they must choose from. In creating a solution that can help predict buy prices for liquor and sell prices for each shot of that liquor, bars will benefit from sales growth and better utilization of important funds. The Liquor Prediction Program will provide several graphical depictions of data that can be useful, such as the average price per shot for each type of liquor, while also providing an excellent prediction tool to predict either buy or sell prices based on variable inputs. The standalone application will make recommendations based on an interpretation of a provided dataset and input provided by a user. The data being used is not restricted information but is based on a nationally available database that provides fair averages for price. However, in the case that a user wants to include proprietary information in the dataset, all data will already be stored in an external database. Providing price recommendations based on custom profit goals and alcohol content will have several benefits. Financial resources will be precisely allocated so overspending on buying products for distribution is not excessively allocated. Growth targets will be more easily achieved to provide reassurance to stakeholders. Customers will experience more stable and predictable pricing for drinks.

The resulting application will provide a graphical user interface that is intuitive and easy to use, predictive functionality, a demonstration database, and graphs to represent the database. The predictive functionality will accept a value between 100 and 300 percent profit and between 0 and 100 percent alcohol percentage. It will use data trends to make reasonable recommendations to achieve a target profit percentage of a target alcohol percentage liquor.

The application will require little maintenance and could be repaired or updated on a contractual basis. The initial development cost will be around $9000, which will pay a team of three developers $58 per hour for 40 hours. Labor costs will accumulate to $7000, the remaining $2000 will pay for the remaining labor required for bug fixes that must be done post-deployment, the cost of reliable equipment, and the cost of training for the application. No additional licenses or proprietary information is required, as the project does not use restricted information. If additional requirements are requested, a maintenance request should be assumed to be one-third of the original cost, which would make the price $3000 for maintenance.

To implement this project, I will use knowledge that is a culmination of courses and experiences throughout obtaining a Bachelor of Computer Science degree. This knowledge has been refined by successful independent contracting and personal projects.

Sincerely,

James McGill

Lead Engineer

# Part B: Project Proposal Plan

## Project Summary

## Bar Management Strategies, and bars in general, face a problem when it comes to deciding a price for buying and selling their liquor. Oftentimes, liquor distributors may overcharge and not provide a fair price, or potentially more damaging, a bar may undercharge and lose money on their liquor. This problem is complicated by the wide range of goals for a particular business and the liquor selections they have to choose from. The client is a bar management company whose chief executive officer is responsible for the acquisition of dozens of bars. The client needs a program that uses descriptive methods to produce visuals for the data and predictive methods that will provide reasonable recommendations to the user to help predict and manage buy and sell prices. The finished application will include a graphical user interface (GUI) that will display charts and provide intuitive interactive elements that allow the program to receive recommendations provided by an ordinary least squares algorithm(OLS). The user guide will include instructions for installation on any major operating system (Windows) as well as a helpful guide for how to interact with the program.

## There are currently many, relatively ineffective solutions to price prediction in the bar industry. Bar management has implemented daily redundant inventory counts that can be inaccurate and time-consuming. Bar management has also cooperated with other local bars to attempt to achieve fair prices, but this typically results in certain beverages being unavailable due to unwillingness to pay the asking price. The Liquor Prediction Program will provide an intuitive way to circumvent these problems by allowing for a definitive prediction based on alcoholic content and intended profit margins.

## Data Summary

## The data will be collected from <https://youthalcoholbrands.org/price-database/> and copied to a Python data cleaning script as raw text. The data will then be processed by the Python script that will remove unnecessary entries for each row of data and convert it into a MySQL usable values script. The data from then on will be stored in the relational database. During the design of the Software Development Life Cycle(SDLC), the data will not be stored locally and will only be accessed once the program is ready to begin using it. Then, during the development stage of the SDLC, the data will be stored on a local text file, processed, and then transferred to the MySQL database using pyconnect. The data will remain in the MySQL database during the rest of the application's life. Should the application need to be updated in a way that requires the data to change, for example during maintenance, the python processing script would be able to handle the new data in that form and is easily tile able to account for more columns or rows. The data meets the needs of the project because it provides a central and publicly available resource that can be used to determine pricing. The data provides both the price per ounce of each liquor as well as the alcohol content, both of which are essential to the application functionality. One issue with data anomalies pertains to some of the liquor names containing apostrophes. Apostrophes in a name, such as in a possessive context, if not preceded by an escape character will cause the MySQL data to be formatted incorrectly and throw unexpected errors. Therefore, excluding at the beginning and the end of any data entry, if an apostrophe is encountered, the program will precede it with a backslash, the escape character for MySQL. There will fortunately be no ethical or legal concerns. The data is not proprietary, and it does not contain information about any groups that could be considered protected or otherwise requiring sensitivity.

## Implementation

## To develop this project, I will use an incremental software development approach. The software will be developed in steps, each one adding to the functionality and being tested before starting the next. This will allow us to reduce time spent debugging after development.

## The high-level steps I will take to develop the project are as follows:

## Create a technical design (table rows, graphical layout, RL algorithm specifications) plan from which I can create a lower-level implementation design.

## Create a low-level design for each area of the project:

## Database (MySQL)

## Backend (Python and NumPy)

## Frontend (pyQT5) Each of these respective plans should include class designs, fields, and methods to obtain the required results.

## iii. Create the database including all required tables and columns, no rows will be present yet.

## iv. Develop a simple program that uses regular expressions to process 100 copied and pasted data entries and translate them into a predefined MySQL values entry. These entries then will be used to create an initial setup file that can be used to reset or set up the database with the appropriate data consistently.

## v. Develop the backend, or model, code by following this process:

## Create methods to load data from persistence. Perform several tests.

## Create methods for each graphical visual display to describe the data (descriptive bar chart using mean and scatter plot). Display graphs and ensure the data displays correctly for multiple data sizes.

## Implement the RL algorithm to make specific predictions using an ordinary least squares machine learning algorithm.

## vi. Develop the frontend, or view, user interface.

## vii. Test the final application to ensure requirements are met successfully.

## Timeline

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestone or deliverable** | **Duration**  **(hours or days)** | **Projected start date** | **Anticipated end date** |
| Design   * Low-level design | 6 hours | 12/4/2023 | 12/6/2023 |
| Development   * Data processed and stored in a custom MySQL database | 4 hours | 12/7/2023 | 12/8/2023 |
| Development   * Backend completed including machine learning OLS algorithm implementation | 65 hours | 12/9/2023 | 12/23/2023 |
| Development   * Front end GUI completed connected to backend code, application now has all required functionality | 35 hours | 1/1/2024 | 1/6/2024 |
| Testing   * Quality Assurance tests performed | 10 hours | 1/7/2024 | 1/9/2024 |
| Delivery   * Final application deployed | 2 hours | 1/10/2024 | 1/10/2024 |

## Evaluation Plan

## During the Design phase, the success of the plan will be defined as whether the plan accurately addresses each of the client's requirements.

## During the Development database and backend phases, at each increment completion, an assertion test, known as assert in Python, will be performed to determine if each predictive function is providing accurate results. The tests will pass if the value provided, and the expected value are equal for 100 percent of the tests.

## The GUI will be tested for accessibility upon completion. The GUI will be considered successful if it allows the user to read input from each required predictive requirement.

## Upon project completion an accessibility and accuracy test will be performed. The project will be considered successful if the user can inspect each visual, interact with each input method, and acquire relevant predictions based on input fields. The project must also return the expected value for each relevant prediction to be considered successful.

## Resources and Costs

## Hardware:

## One Dell desktop setup (including all necessary I/O devices) installed with a Microsoft operation system: $1300.

## One portal laptop for increased program accessibility: $700.

## Software:

## The software and all of its dependencies will be free.

## Labor:

## Labor cost: $7000 at $58 per hour.

## Labor time: 120 total hours.

## Environment:

## Deployment will be free, as it is a modular standalone application and does not need to be hosted.

## Maintenance is unnecessary unless additional requirements are requested, if they are requested, the assumed cost should be one-third of the initial development cost: $3000.

## Part D: Post-implementation Report

## Solution Summary

## Bar Management Strategies, and bars in general, face a problem when it comes to deciding a price for buying and selling their liquor. Oftentimes, liquor distributors may overcharge and not provide a fair price, or potentially more damaging, a bar may undercharge and lose money on their liquor. This problem is complicated by the wide range of goals for a particular business and the liquor selections they have to choose from. The client is a bar management company whose chief executive officer is responsible for the acquisition of dozens of bars. The client needed a program that uses descriptive methods to produce visuals for the data and predictive methods that provide reasonable recommendations to the user to help predict and manage buy and sell prices. The finished application includes a graphical user interface (GUI) that will display charts and provide intuitive interactive elements that allow the program to receive recommendations provided by an ordinary least squares algorithm (OLS). The user guide includes instructions for installation on any major operating system (Windows) as well as a helpful guide for how to interact with the program.

## Data Summary

## The data was collected from <https://youthalcoholbrands.org/price-database/> and copied to a Python data cleaning script as raw text. The data was then processed by the Python script that removed unnecessary entries for each row of data and converted it into a MySQL usable values script. The data from then on was stored in the relational database. During the design of the Software Development Life Cycle(SDLC), the data was not stored locally and was accessed once the project entered development. Then, during the development stage of the SDLC, the data was stored on a local text file, processed, and then transferred to the MySQL database using pyconnect. The data remained in the MySQL database during the rest of the application's life. Should the application need to be updated in a way that requires the data to change, for example during maintenance, the Python processing script can handle the new data in that form and is easily tile able to account for more columns or rows.

## Machine Learning

## The Machine Learning method used is an ordinary least squares regression plot. This plot takes a pair of data sets, plots them each on either the x or y-axis of a graph, and draws a regression line through the data. The line is then used to make predictions of y based on x. The method was developed using Python’s NumPy and stats models libraries. The NumPy library was used to assess the range of values for the regression line and the statsmodels library was used to get the ordinary least squares prediction formula and to plot the line visually. This algorithm was chosen because the requirements expect the predictive function to provide the same y for a given x value every time it is called. Therefore, the prediction needs to be consistent and does not require adaptation or training. The algorithm was developed using Python libraries to utilize abstraction and make future maintenance easier, deployment simpler, and response time faster.

## Validation

## The above method was tested using accuracy testing. The test would run the prediction method and assert the result to an expected value—all the tests needed to pass for the algorithm to be considered successful. The algorithm was successful.

## Visualizations

## As soon as the user runs the application, two of the three visuals are visible:

## A screenshot of a computer Description automatically generated

## By scrolling down, the user can view the remainder of the scatter plot as well as the OLS plot.A screenshot of a computer Description automatically generated

## User Guide

**Use the following instructions to install appropriate libraries and run the application:**

1. Unzip the file containing the project code.
2. Place the unzipped file on your desktop, ensure it is named “Capstone.”
3. Open the unzipped file, navigate to application and then MySQL.
4. Run the .sql file in MySQL workbench.
   1. Open MySQL workbench
   2. Log into your root connection
   3. Select “file” and then “open SQL script”
   4. Navigate to the directory: Capstone/Capstone/Application/MySQL
   5. Select the .sql script
   6. Run the sql script in MySQL workbench
5. Navigate to the directory Capstone/Capstone/Application/LiquorPredictionProgram/model
6. Open the load\_data.py in an IDE of your choice and modify the password to your root connection workbench password.
7. Open the command line and run the following commands to ensure the necessary libraries are installed:
   1. pip install matplotlib
   2. pip install NumPy
   3. pip install stats models
   4. pip install pyqt5
   5. pip install MySQL-connector-python
8. In the command line, navigate to the main folder of the project using this command:
   1. cd Desktop/Capstone/Application/LiquorPredictionProgram/main
9. In the command line, run the following command to run the application:
   1. py run.py

**An example of how the application should be used:**

A bar has just received a new shipment of liquors containing one liquor, named Lily’s Liquor Whiskey. This Liquor has an alcohol percentage of 70%. The bar management needs to determine how much to sell a single show of this liquor for to achieve their goal of 150% profit. The application would make this determination easy; the user would follow these steps in the already running application:

1. Scroll to the bottom of the screen.
2. Set the first scroller to 70%, the label below should read, “Alcohol Percentage: 70%.”
3. Set the second roller to 150%, the label below should read, “Target Sales Growth Percentage: 150%.”
4. Press the button labeled, “Get price per shot based on alcohol content and target sales growth.”
5. The label above the first button should now read, " Recommended Price per Shot: $9.85, at 70% Alcohol Percentage to achieve 150% Sales Growth Percentage!”
6. This indicates to the user that they should sell the new alcohol for $9.85 per shot.

# Reference Page

*Price database*. Youth Alcohol Brands. (n.d.). https://youthalcoholbrands.org/price-database/