C964: Computer Science Capstone

Task 2 parts A, B, C, and D

Andrew Bright

Western Governer’s University

6/2/2023

[Part A: Project Proposal for Business Executives 3](#_Toc136427663)

[Letter of Transmittal 3](#_Toc136427664)

[Project Recommendation 4](#_Toc136427665)

[Problem Summary 4](#_Toc136427666)

[Application Benefits 4](#_Toc136427667)

[Application Description 4](#_Toc136427668)

[Data Description 4](#_Toc136427669)

[Objectives and Hypothesis 4](#_Toc136427670)

[Methodology 5](#_Toc136427671)

[Funding Requirements 5](#_Toc136427672)

[Data Precautions 5](#_Toc136427673)

[Developer’s Expertise 6](#_Toc136427674)

[Part B: Project Proposal 7](#_Toc136427675)

[Problem Statement 7](#_Toc136427676)

[Customer Summary 7](#_Toc136427677)

[Existing System Analysis 7](#_Toc136427678)

[Data 7](#_Toc136427679)

[Project Methodology 8](#_Toc136427680)

[Project Outcomes 8](#_Toc136427681)

[Implementation Plan 9](#_Toc136427682)

[Evaluation Plan 9](#_Toc136427683)

[Resources and Costs 9](#_Toc136427684)

[Timeline and Milestones 10](#_Toc136427685)

[Part C: Application 11](#_Toc136427686)

[Part D: Post-implementation Report 12](#_Toc136427687)

[A Business (or Organization) Vision 12](#_Toc136427688)

[Datasets 12](#_Toc136427689)

[Data Product Code 13](#_Toc136427690)

[Objective (or Hypothesis) Verification 15](#_Toc136427691)

[Effective Visualization and Reporting 16](#_Toc136427692)

[Accuracy Analysis 16](#_Toc136427693)

[Application Testing 17](#_Toc136427694)

[Application Files 18](#_Toc136427695)

[User Guide 18](#_Toc136427696)

[Summation of Learning Experience 18](#_Toc136427697)

Sources 19

# Part A: Project Proposal for Business Executives

## Letter of Transmittal

June 2nd, 2023

Subject: Restaurant Recommendation System Utilizing Machine Learning

Dear Executive Team at Best Travel Agency,

I am reaching out via this email with a proposition for a software solution that will benefit your organization. As we move towards an always-online and always-connected world, we will need to work towards new solutions to provide around-the-clock service. In response, I believe that what we would offer will be an excellent option.

As it currently stands, my team is proposing to design, develop, deploy, and support a restaurant recommendation system that will adhere to your current design standards within your exisiting application. The current proposal is to provide input information of either a city or a type of restaurant, and return a list of the “Top 10” restaurants with the given information. In order to acheieve this goal, we will implement machine learning to analyze a dataset provided from other components of your application. This will directly compliment the features already presented and will lend to the “one-stop-shop” aspect for all information based around planning travel.

In terms of feature cost, the current budget is projected to be $15,000, which includes $12,000 for design, development, and deployment, and $3,000 for three years of maintenance. Additionally, we will offer a lifetime maintenance add-on for an additonal $1.500.

Our team as SkillSoft is made up of a network of industry veterans and our expertise spans numerous industries. The main area of expertise is in machine learning, data science, and artificial intelligence. Previously, we have offered software solutions for companies such as Large Health Organization (LHO) and Big Financial Solutions (BFIN). Please don’t hesitate to reach out if you would like to be provided with testimonials from our former clients.

We look forward to your response and the opportunity to provide a beneficial component to your ongoing solution.

Thanks,

Andrew Bright  
***Engineering Team Lead***  
[abright@skillsoft.com](mailto:abright@skillsoft.com) | LinkedIn

## Project Recommendation

### Problem Summary

Best Travel Agency, like every other far-reaching organization in the ever-expanding world through technology and communication, is being faced with the unique problem of being able to provide a solution around-the-clock, in any situation. This is where the team at SkillSoft is proposing an idea for a restaurant recommendation system utilizing machine learning that would integrate with the current aspects of the application already on the market from BTA. The project would look to acheieve the following:

* An app component that adheres to the current design standard to recommend restaurants where a customer is wanting to travel
* Documentation over app components, support, and a general user guide
* Three years of maintenance/support with an optional lifetime package

### Application Benefits

The proposed solution will utilize machine learning to recommend customers restaurants based on where they would like to travel, or based on the type of food. It will be designed to integrate seamlessly with the environment of the current application and will utilize data collected from the application and the external data already utilized by the organization. The project will help complete the picture of BTA being a sole source of travel planning, information, and recommendations accessbile at any time of day.

### Application Description

The application proposed within this recommendation will serve to “slot-in” as an accompanying model and feature alongside what is currently on offer. It will encorporate machine learning in order to parse through various pieces of data to result in a list of recommended restaurants that contain the location, city, type of food, reviews, number of reviews, and comments. It will also include a list of “key phrases” pulled from within the comments that highlight some of the key words from the given comment. The main types of information provided to the application are a type of food the customer is interested in and a location, however it will offer recommendations if either are missing.

### Data Description

The origin of the data could either be provided via a public database (such as Kaggle.com) of readily-available data, or a private dataset provided within the application as long as it is converted into a CSV (comma separated value) format with the various fields of information. The main data file that will be read through and cleaned to be used will be a single file, laid out as such:

* Restaurants (as a CSV)
  + Name
  + Type
  + Street Address
  + Location
  + Contact Number
  + Comments
  + Reviews
  + No of Reviews

Additionally, the appliation will utilize the “Comments” information as mentioned previously to generate a “Key Phrases” set of data. All of this information will be presented for viewing and reference within the recommendation model.

### Objectives and Hypothesis

The objectives of this project are to produce a customized recommendation system that provides relevant recommendations to a wide user base using real-world reviews in one convenient place. It will be accessible alongside the current methods that BTA implements being a mobile app, desktop and mobile websites to be accessible to anyone, regardless of device type. It is designed to be light-weight and receive data from an online source to return results quickly.

### Methodology

Due to the type of machine learning this proposed project would be implementing, we will focus on using the Waterfall methodology of developing and deploying the solution. This is the most appropriate to this project as the system utilized within the solution focuses on filtering, cleaning, and organizing the data rather than training a model for accuracy, and will not require mutliple stages of testing and deployment once the base funcitonality is achieved. Beginning with designing the project, we will focus mostly on forming the project and model around the type of data we would use, as well as the information best-suited for what we aim to do. During implementation, the base functionality will be established, and internal testing will focus on real-world scenarios as well as delivering a stable experience. After implementation, our team will provide documentation and guidance on supporting the datasets and soluton itself, as well as resolving any software defects during the length of the chosen maintenance window.

### Funding Requirements

The current financial requirements of the project will be a proposed amount of $15,000. The estimated cost of labor for designing, developing, and implementing this solution is estimated at 200 hours, billed at $55/hr for a total of $11,000. An additional $3,000 will be for the maintenance window of three years.

Breaking down the cost of this project, it will utilize open-source and free developmental tools, requring no licensing fees. Due to this being accessible by both a desktop and mobile device, the development will be primarly performed on a desktop/laptop and finalized using each type of device designed for use (being primarily a mobile device and desktop). Hosting will use BTA’s current infrastucture model, and will have an overall cost of $1,000.

There will be an option for an additional lifetime maintenance contract for $1,500.

### Stakeholder Impact

The impact would be everyone involved with BTA’s current solution of software and support. The proposed project will complement the existing application and provide a more extensive repository of information in regards to travel. The direct impacts to the stakeholders will be an increase in revenue and stock price, as BTA will have a wholly-inclusive solution to all needs of travel.

### Data Precautions

The current and possible data implications are limited. As the project can be based off of publicly-available data from sources such as Kaggle.com or the internal datasets provided by the existing application, it will either use data that is already scrubbed of personal identifiable information, or it will adhere to the existing policies in place that are used within the application. The recommendation model does not require personal information, but filtering for possibly compromising aspects such as comments would be looked into if it is deemed necessary.

### Developer’s Expertise

The primary developer overseeing this project has years of experience working in software support alongside Software Engineering, Product, Sales, and Infrastructure. The developer has a BS in CS alongside an ITIL v4 Foundation and CompTIA Project+ certifications, and is aware of the aspects of the project throughout the design, development, and implementation of the proposed solution.

# Part B: Project Proposal

## Problem Statement

## In a world where instant access to any and all information is a never-ending battle, appropriate responses to customer’s needs are important. Within this proposal, we are recommending a machine learning solution to recommend restaurants based on filtered data using location and the type of restaurant. We will provide a curated list of recommendations that include contact information, location, reviews, and key phrases from comments for a concise component to the exisiting application on offer.

## Customer Summary

## Best Travel Agency is a mid-size to larger-corporation that has offered a robust travel assistance and planning application for numerous years alongside their remotely-based online support and website for various travel needs. Their client-base is in the millions, and they are a highly rated travel agency with remote staff around the country that assist with travel needs. However, having physical staff for questions can have its limitations, and providing a machine learning system to provide information will lessen the need for physical staff at all times.

## Existing System Analysis

## Currently, Best Travel Agency’s application supports hotel, car and plane travel, alongside some details in regards to the cities that their customers are staying in. Pertaining to this proposal and application, customers can also leave reviews for restaurants, however they currently are sporadic in nature and only provide a limited amount of information when researching where to visit in a new city.

## This proposal will plan to build upon this foundation and utilize the data produced within the application to filter data specific to an area (in this case a city) or a type of restaurant. The result will be a list of at most 10 filtered restaurants based off of real-world reviews to provide the best recommendation for a potential customer.

## At the time of this proposal, the current internal systems are hosted on various Linux servers that utilize AWS for data delivery with lambdas for current integrations with external systems. The proposed project will be python-based in a Google Colaboratory notebook utilizing the Pandas, NumPy, Scikit-Learn, Seaborn, IPython and Matpotlib libraries for its function. It will first be developed using a dataset from Kaggle.com, but will be designed to interpret a CSV exported by the other components of BTA’s application.

## Data

## The current dataset at the time of this proposal is provided by Kaggle.com. It contains restaurant names, locations, reviews, contact information, and comments from real-world reviews. Best Travel Agency provides this data throughout various components using various integrations with other aspects of the application.

## Data will be used while designing the application from Kaggle.com, while a custom integration will be developed to receive data from BTA in a CSV format to match what is provided within the Kaggle CSV in the main stage of development. Maintenance of this integration will be included with the agreement for the three years of mandatory maintenance, as well as thorough documentation providing the aspects of the integration and various lambas for reference.

## Data will first be interpreted and uploaded into the Python project via a CSV format. It will then be read and converted into a Dataframe with the necessary columns needed to parse through the data. During the cleaning phase of the data, reviews containing no type or number of reviews will be dropped. The number of reviews will be converted to an integer-type for final use with filtering the final dataframe that is presented. Reviews missing or not containing US states will be dropped as well, as at this time, BTA’s main application is only supported within the US. Reviews with duplicate names will have their comments combined into a list of strings, and then consequently dropped to make sure only one instance of a restaurant exists.

## In terms of the recommendation model, the restaurant type will be filtered down and separated to be comapred to the restaurant as well as the name. The restaurant with the highest amount of reviews will be set as the index for the model and be recommended first, with 9 other restaurants returned after being compared against. Comments will be filtered and weights will be assigned to key phrases, with the top three phrases returned for quick viewing.

## In the instances of missing or limited data, the interface utilizes a drop-down containing location and type data for selection to minimize the chances that malformed data is entered into the application that would cause issues.

## Project Methodology

## Due to the nature of this application, we will utilize the Waterfall methodolgy. Since the machine learning model proposed does not require a formal training and re-training, a standard model of developing and deploying components in large stages is what we recommend. Because the interpreted data will be designed to integrate and refresh on updated recommendations automatically at pre-determined times, maintenance will be sparse unless major infrastructural or data processing changes are made within the application.

## The initial design and basic data analysis will be performed on the public datasets provided by Kaggle, with main development focusing on the integration and data collections for use within the recommendation model. Maintenance will consist of monitoring the integration for any issues due to data changes, etc. as well as providing minor changes to the internal structure of the code as needed if bugs are found.

## Project Outcomes

## During the designing and development of the software solution, information in regards to what data is needed from BTA’s current application and where the data would best be served from will be provided. Various forms of visualizations from the collected data will also be provided to allow for an overview of various aspects of interpreting the data after it has been cleaned. A project schedule will also be provided to allow for a concise timeline of project deadlines and a visualization of components of the process.

## Upon project completion and deployment, access to the final application will be accessible via Google Colaboratory. User and official documentation will be provided that oversees how to use the application and providing general support for reference on the aspects of the project.

## Implementation Plan

## During design of the project and utilizing the publicly-available data from Kaggle, tests will be performed to ensure proper recommendations are returned from the initial application testing. This will be enhanced upon as the integration is developed to work with BTA’s current model to ensure the best possible data and reviews are fed to the recommendation model. As changes are made and new reviews are introduced, the datafeed from BTA to the model will be updated to include newer reviews and provide more relevant recommendations. Dependencies will include Python and various free libraries such as: Pandas, NumPy, Scikit-Learn, Seaborn, and Matpolib. The main environment will be Google Colaboratory, with the necessary appliaction-specific model being adhered to and developed alongside this model. After completion, access to the Colaboratory widget will be accessible for use.

## Evaluation Plan

## During design and development, the number of reviews as well as the content are the main informaitonal system used within the recommendation model propsed within this document. During the design phase, the public data will be referenced and tested until an acceptible amount of results are returned. Finally, the information provided by the integration will build upon this model, and will utilize the number of reviews as the basis of providing a sound recommendation. It will take into account all types of reviews to provide the best recommendations as possible. This recommendation system will benefit from a larger amount of data provided to the model, and will only provide more curated results with time.

## Resources and Costs

## Specific hardware needed for this project will be a desktop, laptop, and mobile device for deployment and testing. Designing, testing, and development will take place within a Google Colaboratory notebook uzing Python and various libraries such as: Pandas, NumPy, Scikit-Learn, Seaborn, and Matpolib. These are all free software and do not require licensing.

## Hosting of this application will be using BTA’s current model of Linux servers and AWS for hosting the main application and feeding the data to both the recommendation system and customers. The cost of this will total around $1,000. Labor costs for designing, testing, and deploying will total $11,000 billed at $55/hr for 200 hours, and $3,000 will be used for the required maintenance window of three years.

## The total cost of this project will be $15,000. There is an additional option for a lifetime maintenance agreement for an additonal $1,500.

## Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Time (Hours) | Resources | Estimated Dates |
| Proposal Delivery and Approval of Requirements | 12 | Project Mananager & Stakeholders | 6/5/2023 – 6/6/2023 |
| Analysis of Existing System | 20 | Software and Database Engineer | 6/7/2023 – 6/12/2023 |
| Data Collection and Initial Design | 20 | Software Engineer & Data Scientist | 6/13/2023 – 6/15/2023 |
| Web Model Implementation and Integration Development | 40 | Software and Datbase Engineer | 6/16/2023 – 6/26/2023 |
| Application Component Development | 40 | Software Engineer | 6/27/2023 – 7/5/2023 |
| Final Component Testing | 20 | Software Engineer, QA Tester | 7/6/2023 – 7/10/2023 |
| User Acceptance Testing | 20 | Software Engineer, QA Tester, Stakeholders | 7/11/2023 – 7/13/2023 |
| Application Component and Web Model Deployment | 20 | Software Engineer | 7/14/2023 – 7/18/2023 |
| Project Documentation and Final Overview of Application Components | 8 | Project Manager & Stakeholders | 7/20/2023 – 7/21/2023 |

# Part C: Application

The submission for this project includes a link for accessing the Google Colaboratory Notebook, as well as a compressed ZIP file containing the notebook file (ipynb) as well as the required CSV file for data processing.

Link to project:

<https://colab.research.google.com/drive/1q5YYLmvySYPKb7UCjIxE2pVgRMDx1xuT?usp=sharing>

The ZIP file containing the project will be laid out as such:

C964 – Andrew Bright.zip

\C964 – Andrew Bright.ipynb

\Restaurants.csv

# Part D: Post-implementation Report

## A Business (or Organization) Vision

## Best Travel Agency has offered a travel planning companion application and website for quite some with numerous components useful for someone who is planning a vacation, business travel, or a casual trip. However, with the issues around providing support and information at all times, providing a solution that utilizes machine learning for a recommendation system can alleviate the need for additional personnel that would otherwise assist during business travel. The system has been designed to integrate with the existing data structure to provide a robust system of restaurants that contain comments, types of food, and information about the establishment for reference. It complements the application to work seamlessly alongside and provide a world-class experience.

## In the sense of providing an example, if someone were to plan a short trip to Dallas, TX, and was interested in Seafood, they could choose from either of those options to be provided with the top 10 seafood restaurants in Dallas. However, if they were curious as to what the results were without a type of food, they could search with only the city, and have a list of restaurants returned via the recommendation system. Alongside the recommendation, they would have instant-access to key comments pointing out aspects of the restaurant, the address, and contact information.

Datasets

The dataset used within this application was from Kaggle.com and consists of one large CSV containing all of the information present.

A screenshot of a computer

Description automatically generated with medium confidenceHere is a sample of the raw data:

A screenshot of a computer

Description automatically generated with low confidence

Through processing, only the necessary data was initially selected and placed into a primary dataframe. Reviews containing no information for the type or number of reviews were dropped, as those are used amongst the primary recommendation model. Empty comments were filled with an empty string to remove missing values, and the number of reviews was stripped down to only the number characters and transformed into an integer value.

Further along with the data cleaning, the location data primarily containing cities, states, and ZIP codes were used, but the ZIP code was removed for a cleaner presentation. Additional reviews that were not US-based were dropped to put the focus on restaurants within US states.

A picture containing text, screenshot

Description automatically generatedHere is an example of the processed data prior to comment filtering:

Finally, comments are combined into one list of strings, and duplicate names of restaurants are dropped from the list. This is performed after the visualization models are presented, as reviews containing no comments can still be useful for showing the number of reviews. Comments later on are processed to produce the key phrases, which will be returned within the final dataframe.

For the purpose of an example, here is the final set of data that will be filtered and used by the recommendation model:

Here is a direct link to the dataset used from Kaggle.com:

<https://www.kaggle.com/datasets/siddharthmandgi/tripadvisor-restaurant-recommendation-data-usa>

The data was renamed as “Restaurants.csv” for a cleaner presentation.

## Data Product Code

## As stated previously, the data used within this application was extracted from the primary CSV, placed into a dataframe to provide concise information. Data objects within the dataframe were kept as they are to minimize processing time unless needed, and was referenced as the data type while cleaning. This was true all for one column within the dataframe, being the number of reviews, as the integer value was directly accessed and used as a filtration method for the restaurants.

## For the descriptive methods used within this project, visualizations were generated using the provided informaiton given in the CSV and the main dataframe.

## A picture containing screenshot, rectangle, text, square Description automatically generatedThe first visualization is a breakdown of how many reviews there were per the state they were based in, using a method of slicing the location data to only contain the states, and performing a count on each object. This was presented as a histogram:

## The second visualization used was a count of the amount of reviews versus the top 50 types of food (to eliminate very low values). This was performed by expanding and exploding the list of types contained within the primary dataframe. This was organized in descending order within a countplot:

## (the full graph can be expanded within the Colaboratory notebook to more easily see the values)

## The final visualization was a pie chart to show the relationships between the average rating of a restaurant. This was performed by counting the number of reviews against the average rating:

## A picture containing screenshot, text, diagram, circle Description automatically generated

## The recommendation and predictive method used a grouping style of prediciton utilizing the cosine-similarity algorithim provided in Scikit-Learn. Within the final components of the recommendation model, a secondary dataframe is created pulling information for use from the parameters provided (locaiton or city, and type of food).

## That dataframe then has the key phrase extractor ran to assign weights or ranks with the Rake-Nltk algorithm to phrases, and return only the top three. Afterwards, it is first sorted by the highest-rated overall restaurant (given by the number of reviews provided), and the types of restaurants are converted into a matrix and used with the cosine-similarity algorithm. Finally, the names of restaurants are set as an index for the list that will be compared against when locating the first 10 closest matches from the top restaurant. This list is returned as a view of the final dataframe, containing the recommendation data to the user.

## Links to all sources will be provided in a Sources list at the end of this document.

## Objective (or Hypothesis) Verification

## The primary objective with the project was to provide a concise list of recommended restaurants and information around how others felt about the restaurant through comments, ratings, and key phrases. The acceptance of the system was mostly reliant on the following: Integrate with existing systems or allow for an ease-of-use method of interpreting data, and provide the customer a curated list of results based on either a city, type of food, or a combination of either to further drive engagement with the existing customer base.

## After development and deployment, the system was set in-place to work with exisiting infrastructure and data, allowing for a cost-effective method of transforming data that was available elsewhere. The recommendation model was designed to accept updated data as it is generated, to make sure it is always using relevant information.

## Since implementation, Best Travel Agency has reported an increase in customer satisfaction and has become an effective solution to provding any and all necessary information for planning travel.

## Effective Visualization and Reporting

## The primary objective of exploring data in the methods provided were to provide a wider picture fo the useful data, such as where most restaurants that received rewiews were located, the most popular type of food in given reviews, and the relationship between the numbers of average ratings.

## In the provided Kaggle dataset, most reviews were based in CA, NJ, NY, TX, and WA. The most popular types of food by far were American, Vegetarian Friendly, Italian, and Bar. Most reviews listed were either 4.5 or 4 out of 5, providing a good picture of how the restaurants rated. This information was helpful in determining to use the restaurant type and location as recommendations rather than the number of reviews or review rating. The visualizations mentioned can be located on pages 14-15 of this document.

## Accuracy Analysis

## Due to the nature of this application, there is no formal mesaurement of accuracy given for the machine learning model. Since it is already based on reviews and sorting by ratings, a limited amount of filtration would be needed for data collection to be used to ensure that spam or unrelated content appears within the dataset. As this application was designed to pull from various other application components automatically, ideally there will be no maintenance as the filtration methods were put in place during the development of the integration.

## In regards to the accuracy of the application, unfortunately recommendations are only as such: a recommendation. The application has been designed to group similar reviews together using machine learning and provide a list that we feel is “best-suited” for the end user. This type of project relies on reliable data for the model, and that is where the integration was designed to provide.

## All in all, the recommendation model serves to produce a list of restaurants given the choices made based on where you are or will be, and the type of food you would like to have compared. The final dataframe of results using an example city and food type can be seen below, presented in the cell of the interactive widget within the Colaboratory notebook:

## A picture containing screenshot, text Description automatically generated

## The returned data returns a list of American-like restaurants that the end user could then experience for the first time, or list similar restaurants to a possible favorite included within the list.

## Application Testing

## During devlopment, unit testing was performed on each component to ensure funcitonality before the application was assembled. Various types of data were tested that either listed everything within the initial CSV, or different lists from the final one contained within the dataframe. As development went on, methods were devised to clean the data in different ways for easier processing and presentation later.

## From an early point in development, it was determined that the dataframe used in the final stages of the recommendation would pull from a filtered list or slice from the larger dataframe to reduce processing time, cost, and complexity. Fortunately this method is able to yield a good list of results if it is provided a dataset with enough data. However, there was an additional step added onto the recommendation system that still results in a list if there is not enough to recommend a restaurant for a given city, which then uses the overall highest-rated restaurant for that city to recommend others.

## Throughout development, the filtering of comments provided to be interesting using the dataset provided by Kaggle, as it only contains a partial comment entry from an online review. Because of this, the comment filtering will assign weights to phrases that aren’t entirely “clean” or provide a clear picture of the comment(s) it was filtered from.The comment filtering system was changed from intially selecting each comment to combining comments for better results. However, using the comments provided by BTA, the Rake algorithm was able to assign weights to phrases that better encompassed a restaurant.

## Finally, testing was performed on both a laptop/desktop and a mobile device, both in the application and browser. The solution cohered to the design aspects of the mobile app, and provides a seamless experience for restaurant recommendations for a user both on the website and the mobile app.

## 

## Application Files

## This section is similar to the overall Part C of this document, but here is the link to the Google Colaboratory application:

## <https://colab.research.google.com/drive/1q5YYLmvySYPKb7UCjIxE2pVgRMDx1xuT?usp=sharing>

## The ZIP included with this submission will be titled and contains the contents:

## C964 – Andrew Bright.zip

## \C964 – Andrew Bright.ipynp \Restaurants.csv

## User Guide

## Extract the “C964 – Andrew Bright.zip” compressed folder to view its contents

## Open the link to the [Google Colaboratory notebook](https://colab.research.google.com/drive/1q5YYLmvySYPKb7UCjIxE2pVgRMDx1xuT?usp=sharing)

## If necessary, create or login to a Google account to access the notebook

## Within the notebook, either all cells can be dropped down and ran individually, ran as a collapsed set, or all at once using “Runtime” > “Run All” (or CTRL + F9)

## If it is the first time executing the application via the notebook, the “Restaurants.csv” will need to be provided within the Import Data Files cell. There is an option to Browse, where the file can be located from File Explorer.

## After all cells have executed, the Recommendation Model Widget can be used to interact with the project from a contained cell of code.

## Summation of Learning Experience

The experiences I have gained throughout this degree program as well as my professional life have prepared me for this project. With my previous work experience spanning a few organizations of different sizes, it has allowed me to understand and produce a valid business case for a machine learning model like the one in this document. Working alongside Engineering, Product, Sales and QA have lent to an understanding of various components within a software solution.

In preparation for this project, the C951 Intro to Artificial Intelligence provided most of the experience needed for this course, including the models and proposal paper. I also utilized courses on Udemy and LinkedIn over data science and various machine learning aspects in regards to Python.

Throughout the development process of this application, numerous posts on Stack Overflow, GeeksForGeeks, as well as the various documentation for Pandas, NumPy, Scikit-Learn, Seaborn and IPython.

This project and overall degree program have taught me how to perform effective independent research with various interesting syntax issues or a certain way of wanting the program to interpret the data in the way I envisioned. Alongside providing experience within the ever-expanding and relevant fields of machine learning, data science, and artificial intelligence, this program and project has prepared me for what lies ahead.

Sources

*The absolute basics for beginners*. NumPy. (n.d.). <https://numpy.org/doc/stable/user/absolute_beginners.html>

*API reference*. scikit. (n.d.). <https://scikit-learn.org/stable/modules/classes.html>

*Pandas documentation*. pandas documentation - pandas 2.0.2 documentation. (n.d.). <https://pandas.pydata.org/docs/>

*API reference*. API reference - seaborn 0.12.2 documentation. (n.d.). <https://seaborn.pydata.org/api.html>

*Rake-NLTK*. PyPI. (n.d.). <https://pypi.org/project/rake-nltk/>

*A computer science portal for geeks*. GeeksforGeeks. (n.d.). <https://www.geeksforgeeks.org/>

*Where developers learn, share, & build careers*. Stack Overflow. (n.d.). <https://stackoverflow.com/>