Analysis Plan

Darren Carvalho (301251637)

Professor: *Bilal Hasanzadah*

Due date: July 17, 2023

Analytics Startup Plan

Synopsis: *This document provides a high-level walkthrough of the activities required to guide completion of the analysis.*

Project	*Demand Forecasting in Restaurant Industry*
Requestor	*Bilal Hasanzadah*
Date of Request	*July 4, 2023*
Target Quarter for Delivery	*Q2*
Epic Link(s)	
Business Impact	• *Better business planning and strategic decision-making based on comprehensive data analysis.* • *Enhanced operational efficiency through data-driven decision-making.* • *Optimized inventory levels and improved staff allocation based on accurate customer flux prediction.*

Table of Contents

1.0 Business Opportunity Brief

The restaurant industry relies heavily on accurate sales forecasting to optimize revenue management, improve operational efficiency, and enhance customer satisfaction. However, traditional judgmental-based methods often fall short in capturing the complex nature of restaurant sales, influenced by factors such as time, weather conditions, and economic indicators. This creates an opportunity to develop a robust sales forecasting system tailored specifically for the restaurant industry, leveraging advanced data analytics and forecasting techniques to drive better decision-making.

System Design and Development: Utilizing the insights gained from the review, design and develop a cutting-edge sales forecasting system. The system will integrate various forecasting models, including multiple regression, Poisson regression, exponential smoothing, ARIMA, neural networks, Bayesian networks, and hybrid methods. It will also incorporate association rule mining to capture valuable patterns and insights from sales data.

Tailored Implementation: Collaborate with pilot restaurants to implement and fine-tune the sales forecasting system. Gather feedback from restaurant owners, managers, and stakeholders to ensure the system meets their specific needs and provides actionable insights for optimizing operations, strategic planning, and revenue management.

Comprehensive Review: Conduct an extensive review of forecasting methods and techniques published over the last 20 years within the restaurant industry. This review will serve as a foundation to identify and classify the most effective approaches for restaurant sales forecasting.

Expected Impact:

Enhanced Operations: The implementation of an accurate sales forecasting system will enable optimized labor scheduling, inventory management, and product preparation, resulting in reduced operating costs and improved overall service quality.

Strategic Planning: Accurate sales forecasting and analysis will provide valuable insights for corporate-level decision-making. Restaurant chains will be able to assess the impact of promotional activities, measure brand recognition, identify business trends, and conduct price elasticity analysis.

Revenue Optimization: Effective revenue management based on accurate demand forecasting will enable appropriate pricing strategies, optimal resource allocation, and improved table availability, maximizing revenue potential.

Improved Customer Satisfaction: The ability to forecast demand accurately will lead to enhanced service levels, reduced waiting times, and improved overall customer experiences. This, in turn, will increase customer satisfaction and loyalty.

1.1 Supporting Insights

The demand for healthier and fresher food options is driving the growth of full-service restaurants, especially among health-conscious baby boomers and millennials. University and college students, with limited time for proper meals, prefer takeout's and quick food services, contributing to the growth of fast-food restaurants.

Global foodservice brands, such as Starbucks, McDonald's, and Domino's, are expanding rapidly in Canada. They are targeting areas near colleges and universities to cater to the student population. Yum! Brands, KFC, Pizza Hut, and other major foodservice providers are also embarking on expansion plans across targeted cities.

The Canadian food sector has the potential to undergo a revolution thanks to machine learning, which may generate profitable insights and results. Demand forecasting is one important application where machine learning algorithms may reliably estimate future client demand by analyzing past sales data, weather patterns, economic indicators, and other pertinent aspects. This makes it possible for restaurants to manage their inventory more effectively, cut down on waste, and make sure they always have the proper quantity of materials and supplies on hand, which lowers costs and boosts operational effectiveness.

1.2 Project Gains

Revenue Gains: By accurately forecasting customer demand and optimizing inventory levels, staff allocation, and sales targets, restaurants can experience revenue growth. The machine learning models will provide insights into factors such as reviews, ratings, demographics, and competitors that impact customer flux. This knowledge enables targeted marketing strategies, personalized recommendations, and optimized pricing, leading to increased customer visits and higher sales.

Quality Improvements: Leveraging machine learning algorithms allows restaurants to improve the quality of their offerings. By analyzing customer reviews and sentiment analysis, restaurants can identify areas for improvement in food quality, service, and overall customer experience. This leads to enhanced customer satisfaction, increased loyalty, and positive word-of-mouth, driving repeat business and attracting new customers.

Cost and Time Savings: The machine learning models will help optimize resource allocation, such as labor scheduling and inventory management. By accurately predicting customer demand, restaurants can minimize overstaffing during low-demand periods and avoid stockouts or wastage. This results in cost savings by reducing operating expenses and maximizing efficiency. Additionally, the automation of data analysis and insights generation through machine learning streamlines decision-making processes, saving valuable time for restaurant management.

2.0 Analytics Objective

The main purpose of this study is to use machine learning techniques to identify the proportion of customers that will visit a restaurant to optimize inventory levels, staff allocation and sales target. The major factors that affect this are history data on sales, demographics, competitors in the vicinity, and social media insights about the restaurant.

Models will be supervised for the problem statement with the target being the total number of customers visiting the restaurant in a continuous or categorical format by using a threshold to determine if the customers are significant enough or not.

Model outputs can assist the management in making informed decisions to optimize the business targets as well as get insights on the performance on the restaurant and areas of improvement.

Key Questions:

1. Does the number of reviews, context of reviews, ratings received impact the customer flux.
2. Does the demographics of the restaurant like location, operating hours, cuisine type impact the performance of the restaurant.
3. What are the factors that make successful restaurants perform better than others.

2.1 Other related questions and Assumptions:

1. History data of sales will be the among the most important factors but the availability of such data to be open source is difficult. As a result, a combination of the other factors will be considered in performing the analysis.
2. The yelp check-in count will be considered as the number of customers entering the restaurant. Attempts will be made to understand the proportion of this count with respect to other parameters like online delivery, eat in, take out to extrapolate the actual number of customers visited.
3. Reviews are unbiased.

2.2 Success measures/metrics

Regression Metric: RMSE is a commonly used metric in regression models that measures the average deviation between the predicted and actual values, considering the square root of the mean squared differences. By using RMSE as a metric, the business can evaluate the accuracy of the customer visit predictions. A lower RMSE indicates a smaller average deviation between predicted and actual values, reflecting a more precise forecasting of customer demand. This metric helps the business optimize targets by providing insights into the accuracy of the predictions.

Classification Metric: Precision is a classification metric that measures the proportion of true positive predictions out of all positive predictions made. In this project, precision can be used to assess the accuracy of identifying whether the number of customers visiting a restaurant is significant or not. A higher precision score indicates a better ability to correctly identify significant customer visits. This metric is valuable for the business as it ensures efficient allocation of resources during high-demand periods, optimizing inventory and staffing levels and enhancing customer satisfaction.

Classification Metric: Another crucial classification parameter is recall, which counts the percentage of correctly predicted positive cases among all real positive cases. Recall can be utilized in this project to assess how well significant customer visits can be recorded and identified. A higher recall score indicates a better ability to capture all significant customer visits, minimizing the risk of missing potential business opportunities. This metric enables the business to effectively allocate resources and maximize customer visits, increasing sales and overall revenue.

2.3 Methodology and Approach

Type of Analysis: Supervised machine learning model (Classification or Regression)

We would test a regression and tree-based model on the dataset for a regression model first and then based on the performance and model parameters, decide whether to create categories of the target variable that is customer flux to improve the performance. The reason for doing both approaches is since the check-in value for yelp is significantly lower than the actual number of customers visiting the restaurant. As a result, a categorized target variable with the bins set to respective threshold based on insights and domain understanding will be adopted. The targets for classification can be binary or multiclass in nature.

Methodology

Data Exploration:

1. Conduct exploratory data analysis on the Yelp reviews dataset to gain insights and determine the characteristics.
2. Analyze the number of ratings received and average ratings for each restaurant.
3. Decide whether to reduce the number of restaurants based on these criteria, considering the dataset's significant size.

Feature Engineering:

1. Perform text processing on the reviews to extract relevant information.
2. Explore user statistics, such as the number of reviews given by each user and the duration of their activity on the platform.

3. Create a variable indicating the confidence level in each review based on user activity and review frequency.
4. Utilize techniques like word cloud analysis to visualize common terms in the reviews.
5. Apply sentiment analysis techniques to preserve the context and determine the sentiment of each review.
6. Incorporate additional numerical variables, such as price and value for money, that can contribute to the model's predictive power.

Data Enrichment:

1. Utilize the OpenStreetMap API to gather location-based insights for each restaurant.
2. Retrieve information about nearby amenities, such as libraries, other restaurants, and popular visiting spots.

Incorporate External Data:

1. Utilize the dine safe dataset that evaluates restaurants, to obtain additional insights.
2. Explore the dataset to extract relevant information about each restaurant's safety rating or compliance with regulations.
3. Incorporate external factors into the model to improve its ability to predict the target variable.

Feature Selection:

1. Utilize statistical techniques, such as correlation analysis and tree-based models, to identify the variables that have a significant impact on the target variable.
2. Select the most relevant features that contribute to the predictive power of the model.

Data Preprocessing:

1. Perform summary statistics of the variables, such as mean, standard deviation, and quartiles.
2. Handle outliers using appropriate techniques, such as normalization or outlier removal.
3. Normalize the variables using scaling methods like standardization or min-max scaling.
4. Create dummy variables for categorical variables, if applicable, to represent them numerically in the model.

Model Testing and Evaluation:

1. Employ Regression / classification model with K-fold cross-validation to assess the performance of the model.
2. Compare the performance of the model using scaled data versus non-scaled data.

3. Determine if scaling the data improves the model's predictive accuracy and efficiency.

3.0 Population, Variable Selection, considerations

Audience/population selection: Restaurant reviews data, Demographics data

Observation window: Until 2019 (Yelp API will be used to attempt to get real-time data)

Inclusions: N/A

Exclusions: N/A

Data Sources: Kaggle, Open Street Map API

Audience Level: N/A (any customer)

Variable Selection: Text analysis to create variables, demographics parameters, variable importance through models.

Derived Variables: Point of interest factor that is a result of demographics and user ratings.

Assumptions and data limitations: Reviews are assumed to be unbiased. Having generalized data for all restaurants will be difficult.

4.0 Dependencies and Risks

Risk	Likelihood (based on historical data)	Delay (based on historical data)	Impact
Data Availability and quality	*Medium*	*High*	*The availability of historical data, demographics, competitor information may differ for each restaurant. This could lead to biased or unreliable models.*

Model performance and Generalization	*High*	*Medium*	*The models used for the following project may not capture the complex relationships between the predictors and target variables and there may be other external factors which impact the performance which could result in a non-generalized model.*
Assumptions is that the reviews are unbiased, however, there are some extreme cases of positive or negative feedbacks which may skew the results. Group check-in are also potential risks as it would be technically under a single user but the check-in count may vary.	*Medium*	*Medium*	*If these assumptions prove to be invalid, they may introduce bias and impact the accuracy and reliability of the model that could lead to in accurate decision making.*
Return on Investment and Implementation challenges.	*Low*	*Medium*	*There may be challenges in implementing the recommended strategies due to operational cost, resistance to change or limitation on resources.*

Risk Mitigation Strategies

1. Thorough data validation and cleansing processes to address data availability and quality issues.

2. Exploring alternative models and techniques to improve model performance and generalization.
3. Regularly reassessing assumptions and conducting sensitivity analyses to identify and address biases.
4. Employing model explain ability techniques and clear communication of results to enhance interpretability and stakeholder buy-in.

5.0   Deliverable Timelines

Item	Major Events / Milestones	Description	Scope	Days	Date
1.	Kick-off / Formal Request	Initiate the project and establish project goals and objectives.	Project Initiation	3	July 4, 2023
2.	Assessment / Triage	Evaluate the feasibility and potential impact of the project. Determine the resources and timeline required.	Project Assessment	7	July 12, 2023
3.	Prioritization	Prioritize tasks and activities based on their importance and impact on project outcomes.	Project Planning	7	July 19, 2023
4.	Data Exploration & Analysis • Issues with duplicates • Issues with Spend data	Conduct exploratory data analysis on the Yelp reviews dataset and perform feature engineering to extract relevant information.	Data Preprocessing and Analysis	6	July 26, 2023
5.	Story Board 1	Present initial findings and insights from the data exploration phase.	Data Analysis and Visualization	3	July 30, 2023

6.	QA Output	Review and quality assurance of the data analysis and visualization outputs.	Quality Assurance	2	August 1, 2023
7.	Internal team Presentation	Share the findings and progress with the internal team for feedback and input.	Internal Communication and Collaboration	4	August 5, 2023
8.	Go/No Go	Assess the project's readiness to proceed to the next phase based on the initial findings and feedback.	Project Evaluation and Decision Making	3	August 8, 2023
9.	Story Board 2	Refine and enhance the data analysis and visualization based on feedback and additional insights.	Data analysis and visualization refinement	2	August 10, 2023
10.	Pilot	Implement the developed machine learning models and evaluate their performance.	Model implementation and testing	2	August 12, 2023
11.	Delivery & sign-off	Finalize the project deliverables, obtain client sign-off, and hand over the project outputs.	Project completion and closure	3	August 15, 2023

6.0 References

(2023). Retrieved from https://www.mordorintelligence.com/industry-reports/canada-foodservice-market

Khachatryan, S. (2023). Retrieved from https://orders.co/blog/how-to-use-data-and-analytics-to-improve-your-restaurants-marketing-performance/#:~:text=Analyzing%20data%20such%20as%20sales,insights%20gained%20from%20data%20analysis.

Lasek, A., Cercone, N., & Saunders, J. (2016). Restaurant Sales and Customer Demand Forecasting: Literature Survey and Categorization of Methods.

Liu, L., Bhattacharayya, S., Sclove, S., Chen, R., & Lattyak, W. (2001). Data mining on time series: an illustration using fast-food restaurant franchise data. *Elsevier*, *37*, 455–476. doi:22 Jan , 2001

Mok, T. (2020). Retrieved from https://www.blogto.com/eat_drink/2020/09/new-study-shows-torontos-restaurant-reservations-have-been-climbing-but-still-drastically-lower-than-last-years/