Executive Summary: Maven Roasters Project

Slide Deck: <https://docs.google.com/presentation/d/1q3n9MDYk1tvd7PisICls28vKyR-ajpgZ6cf6F6oF9yM/edit?usp=sharing> .

The coffee industry is dynamic and competitive, with an increasing focus on sustainability, quality, and customer experience. Our project aims to help position Maven Roasters as an industry leader, meeting the evolving needs and preferences of NYC coffee, tea, and baked good enthusiasts.

Project Overview and Goals: The Maven Roasters project was a data-driven initiative aimed at identifying opportunities to enhance operational efficiency and performance of Maven Roasters’ coffee shops. The primary objectives were to identify best-selling products, understand variations in sales across different locations, analyze seasonal trends in sales, determine peak business hours, and identify the best performing location(s).

Data Source: The choice of Kaggle as a source for our project data was based on several factors, including that it is a popular platform for data science and machine learning, and it offers a diverse range of datasets contributed by the community or provided by organizations. The platform maintains high standards for the quality and consistency of datasets which ensures the data utilized is reliable, well-documented, and adheres to best practices in data management.

Data Collection and Processing: Our approach began with collecting the transactions from the first 6 months (Jan 1 – June 30, 2023) of operations across all 3 Maven Roasters locations. We then cleaned and prepared this data to ensure accurate, relevant analysis. We then performed exploratory data analysis (EDA), employing various analytical and statistical techniques to scrutinize popular products, sales, and growth patterns.

Approach to Achieving Goals: The team applied statistical techniques and data visualization tools for an in-depth analysis. We segmented data based on location, time, and product categories to understand sales dynamics. Additionally, comparative analyses were conducted to help optimize pricing strategies and identify peak sales periods.

Unanticipated insights and resolutions: Some initial datasets sourced from Kaggle contained inconsistencies, missing values, or inaccuracies. To address this, we chose a relatively clean dataset and implemented a data cleaning process. Additionally, our dataset contains data from the initial 6 months of three locations opening and growing in performance. Thus, time series analysis results in a pattern of unchecked growth. We combatted this by looking at alternative methods of performance such as growth parameters. Further, given the relatively small sample size and non-normal distribution of data, non-parametric statistical analysis had to be employed.

Results and Conclusions: Our findings revealed key products driving sales and customer footfall across locations and seasons, along with the most profitable hours for each store. The analysis also suggested tailored pricing and staffing strategies to maximize revenue while maintaining competitiveness and operational capabilities. Further, our analysis demonstrated that all three stores are performing very similarly both in total revenue and growth metrics. We would like to reevaluate these metrics once the locations’ growth becomes relatively stable to further evaluate optimal allocation of resources, and identify if location-specific attributes are driving performance, as well as if those attributes can be applied to the other locations.

Future Research and Development Plans: During our analysis, several further areas of interest emerged, including customer loyalty trends, new markets to identify opportunities for growth, and opportunities to employ advanced technologies to enhance customer experience and streamline operations. Given more time and data, our team is interested in exploring these areas. We plan to integrate customer feedback mechanisms and market comparison studies in our future research to enhance our dataset and refine our strategies.

The analysis of Maven Coffee Shop reveals several key insights: transaction patterns show uniformity across days but vary significantly during morning hours, with Lower Manhattan peaking early from 6-8 AM, Hell's Kitchen leading from 8-10 AM, and Astoria gaining momentum in the afternoon until closing. The six-month sales data, while indicating a promising linear upward trend, is insufficient for accurate long-term forecasting due to its limited scope and failure to account for seasonal variations and broader market dynamics. These findings suggest the need for location-specific staffing strategies, with emphasis on morning hours for Lower Manhattan and Hell's Kitchen, and afternoons for Astoria. Additionally, cautious future planning is advised, supplementing limited sales data with extensive market research to develop adaptable and well-informed business strategies.

This summary highlights key aspects of the Maven Roasters project including our approach, findings, and future plans, while also reflecting on its effectiveness. For further details, please see our github repository <https://github.com/cedetraz96/Coffee_Crew> containing our in-depth analysis of the Maven Roasters project.