Rady | UC San Diego School of Management

MEMORANDUM

To: Michael Boutillon

From: Group N

Date: November 26, 2023

Re: Croq'Pain location

EXECUTIVE SUMMARY

As part of our strategic initiative to enhance Croq'Pain's expansion process, our team developed a new statistical model for selecting the locations of new stores. This model aims to address the inconsistencies and challenges faced by our current approach, which primarily relies on expert opinion and management's subjective assessment. The goal is to improve the accuracy and reliability of our location selection process, thus enhancing the profitability and success rate of our new stores. To achieve this, we have employed linear regression with carefully selected features that align best with our business needs. This methodology ensures strategic decision-making based on quantifiable data, significantly enhancing our ability to predict and select profitable locations for new stores.

STATEMENT OF THE PROBLEM

Croq'Pain faces significant challenges with its current subjective method for selecting new store locations, leading to inconsistent performance and profitability among newly opened stores. To address this, Michel proposes an objective, data-driven model, developed using data from 60 stores. He suggests validating this model through retrospective analysis, using data from the first 50 stores to predict the performance of the last ten stores opened. This crucial step aims to test the model's predictive accuracy and its alignment with Croq'Pain's performance target of 26%, thereby assessing its practical applicability and reliability for future expansion decisions.

METHODOLOGY

First, we started by analyzing the data provided to us, using statistical methods and visualization to determine if any corrections were needed. We identified an outlier in the earnings data, where

one value was exceptionally high, likely due to a data entry error. To correct this, we divided the value by 1000, bringing it to a more plausible scale. We then checked for duplicate values. The data was split into two parts: the first part was used to develop the model, and the end rows with unknown values were set aside for prediction.

To identify the variables for the model, we examined a correlation matrix across all variables, comparing which had the highest correlations with each other. Further analysis and visualization of the more correlated variables helped us interpret the data. For instance, variables like K, SIZE, COMP, NCOMP, and NREST were skewed to the right. P15, P35, and P45 showed strong linear relationships with the total. The correlation between K and SIZE was high, which is logical considering that larger stores typically require more investment.

Based on these insights and predefined instructions, we normalized selected columns. We then built our model, initially with untransformed variables, and subsequently performed stepwise regression, also considering the VIF value to remove less significant variables. After examining the distributions of the resulting variables, we decided to log-transform those that were right-skewed. These log-transformed variables were then combined with the normalized ones to construct our second model. This model showed better fit with good R-squared values and no multicollinearity, leading us to select it for our linear regression analysis.

RECOMMENDATIONS

Using the above model described, we predicted the normalized earnings. Then we multiplied them with the total to get the actual earnings for each location. Then we got the performance ratio for all and shortlisted the ones that meet the criteria of being greater than 26%. The resulted suggestions were:

- Toulouse with a performance ratio of 41%,
- Montpellier with a performance of 27% and
- Dijon with a performance of 26%.

Based on our model and the results, we suggest the next restaurant to be opened in Toulouse.