**IS 6813 CAPSTONE COMPLETION**

**MAVERIK STORE DAILY SALES FORECAST USING TIME SERIES & QUALITATIVE DATA**

Maverik is a large convenience store chain with over 400 locations in the western United States. It recently acquired Kum & Go, doubling its store count. Maverik is multiplying and plans to open 30 new stores each year. This makes planning new stores a crucial part of the business. Maverik wants to use various statistical and predictive analysis methods on active stores' historical sales data from 2021 till 2023 to forecast the daily sales of 4 category items, namely 'daily\_yoy\_ndt.total\_inside\_sales', 'daily\_yoy\_ndt.total\_food\_service', 'diesel,' 'unleaded', for the upcoming year, for any new store they open.

The organization analyzes its time-series data to analyze and predict the daily sales of new Maverik stores. It holds the sales data from other stores from 2021 until August 2023 of the four categorical items crucial to assessing a store's sales performance. Maverik is also parallelly analyzing and correlating the sales data with qualitative data representing various items that store stock and sale. Does having a specific item in the store increase overall sales? Are two or more items in the store highly correlating and can be a factor for the causality of purchase? Does the seasonality affect the stores' overall sale or sale of a specific item? How does gas price fluctuation impact on a store's overall sales? These are the few questions that Maverik may want to know before opening a new store during a specific time of the year. To statistically analyze the available data yield answers and forecast daily sales records, Maverik can use statistical tools like R, perform extensive exploratory data analysis data preprocessing, and develop Time-series models to predict the outcome.

The solution, forecasting sales for the next financial year for the new store, is crucial for Maverik. Using statistical analysis, Maverik can deliver a running total of daily sales numbers in dollars; using the daily sales and total overall forecasted sales amount, Maverik can perform revenue forecasting and budgeting for the new stores. They can develop an expense projection plan, work on cash-flow and gross margin analyses, and obtain operating profit margins. These metrics will help them determine the Return on Investments for the newly opened stores.

The Business data scientist team members will be working on the data provided by Maverik data management teams to perform the analysis and provide the numbers as the solution, which are the daily sale value of the four products (target variables) 'daily\_yoy\_ndt.total\_inside\_sales', 'daily\_yoy\_ndt.total\_food\_service', 'diesel,' 'unleaded.' Business data analysts or data scientists can perform exploratory data analysis, to begin with and divide the data first into train and test sets. They can develop multiple predictive models and work with the train and test set to evaluate which model has higher RMSE, MAE, MAPE; generally, they can consider the Receiver operating characteristic to determine the appropriate model to move forward with.

This project will run for 14 weeks, starting in September and ending in early December. The actionable business insights will be conveyed to the management for model buy-out through a pitch presentation.