

Executive Summary

PROJECT OBJECTIVES

The Data Science and Business Analytics graduate students partnered with Hikma Pharmaceuticals and developed a data science solution to address revenue leakage. At the onset of the project, expired returns accounted for approximately 2 percent of gross sales; equating to millions of dollars in annual loss. Hikma used the project to leverage predictive analytics capabilities to diminish potential profit loss due to expired returns.

PROJECT DATA

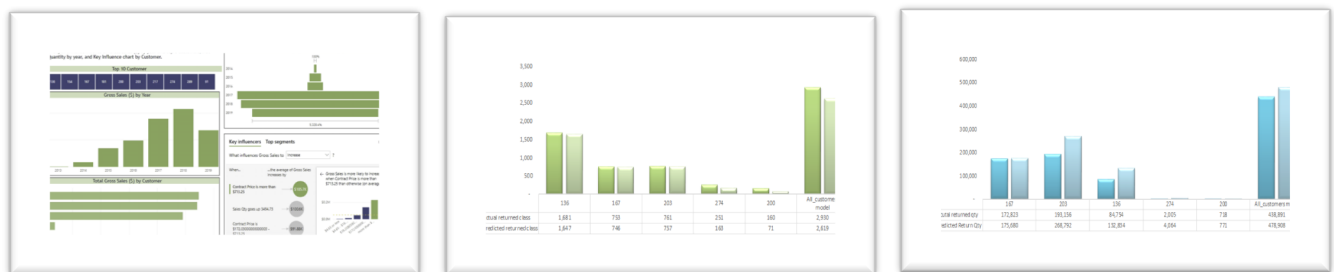
The team was provided with Sales and Returns Data from 2015 to 2020. The data required cleaning and pre-processing techniques to prepare for model deployment. Ultimately, we aggregated only the key features to develop a Sales and Return dataset.

TECHNICAL DESIGN AND DELIVERABLES

The solutions provided were [1] returns prediction model, [2] Rate of Returns estimation model, and [3] dynamic dashboards in Power BI. For these three solutions, the team focused on providing key insight and understanding for the top five customers. This was executed by creating one model for each customer, resulting in five models for regression and five models for classification. Additionally, two main models were developed for all customers, one for classification and the other for regression. The scope of the returns data totals to approximately 25 million in dollar value from the top 5 customers of Hikma Pharmaceuticals.

MODEL PERFORMANCE AND RESULTS

To choose the optimal performing model, accuracy and sensitivity were used to compare performance of machine learning classification and regression. The Decision Tree Classifier was compared to the XGBoost Classifier and it was found that a simpler decision tree model yielded better results. The Decision Tree Regressor was compared to the Support Vector Regression model and model performance was compared using the Root Mean Squared Error (RMSE) value. Observing those values, it was found that the Decision Tree Regressor had a lower RMSE. Decision Tree Classification and Regressor model outperformed the rest with average testing accuracy of 85%, sensitivity of 81% and the lowest compared Root Squared Mean Error. The results for the descriptive analytics dashboards and the two main overall models are shown below.



FUTURE OPPORTUNITY

As future considerations, the graduate students recommended Hikma to further evolve the models and dashboards based on business needs and data availability. During the course of the project, the team encountered missing or inconsistent data which led to recommendations on refining data entry practices at Hikma which will go a long way in to ensure that there is consistency across the various tables that record sales and returns.