Niharika Madhadi

11659474

**ADTA 5340** **Discovery and Learning with Big Data**  
**Final Project** **Executive Summary**

In this report you will find my individual executive summary of project on **Sales Forecasting for a Retail Chain.** The main objective of the project is to predict future sales of a retail chain, we have selected liquor sales dataset for it. Hence, we are predicting the sales of liquor and exploring the features that are affecting sales in order to recommend a better inventory management and promotional planning to increase sales.

The process of this project is to collect this liquor dataset, then clean the dataset by removing null values, outliers. We have also added new features like promotional activity, holiday and temperature ranges to study the analysis of these on sales. Next, we have performed statistical analysis, various visualizations that shows the relationships between features, trends in sales, impact of promotions, weather on the sales and a correlation map. After analysing we have gone for training the data on three models like Linear Regression, XGBoost, and Random Forest Regressor. We have compared these models’ performance and decided that Random Forest is the best suited for our dataset. Using the predictions and feature importances and few other analyses from this model we have made few business insights and recommendations. With the help of the predictions of sales for next quarter using random forest model, we can plan on the inventory accordingly and reduce the risk of overstocking or understocking. For example, if approximately we have saved 10-15% on inventory losses or waste annually, that would be around $300,000 savings across all stores. Also, based on the insights if we improve the marketing of products for say with an investment of $200,000 then these promotions will in turn increase the sales and revenue by 5%, contributing an additional $1.2 million annually. From these forecasts we can check the low-demand items so we can take further action on it so that the holding cost or unsold stock will not be there that could save us up to $100,000 annually. By focusing and optimizing the stock of high-demand products could generate $2 million sales additionally.

As a next step, we could train the data on Neural networks, or we can configure the optimal hyperparameters to improve random forest performance and integrate these into the company system. Finally, our project here forecasts the sales accurately that helps in inventory management and gives insights and recommendations on promotions and other strategies that could impact the business and increase their sales. This could result in revenue growth of around $3-3.5 million annually.