

Price Optimization in Pharma via Machine Learning

Srividya Inampudi, Sai Krupa Jangala, Soham Joshi, Vedant Rajeev Kumar, Victoria Edwards

Data Science Institute, Columbia University and Novartis Pharmaceuticals

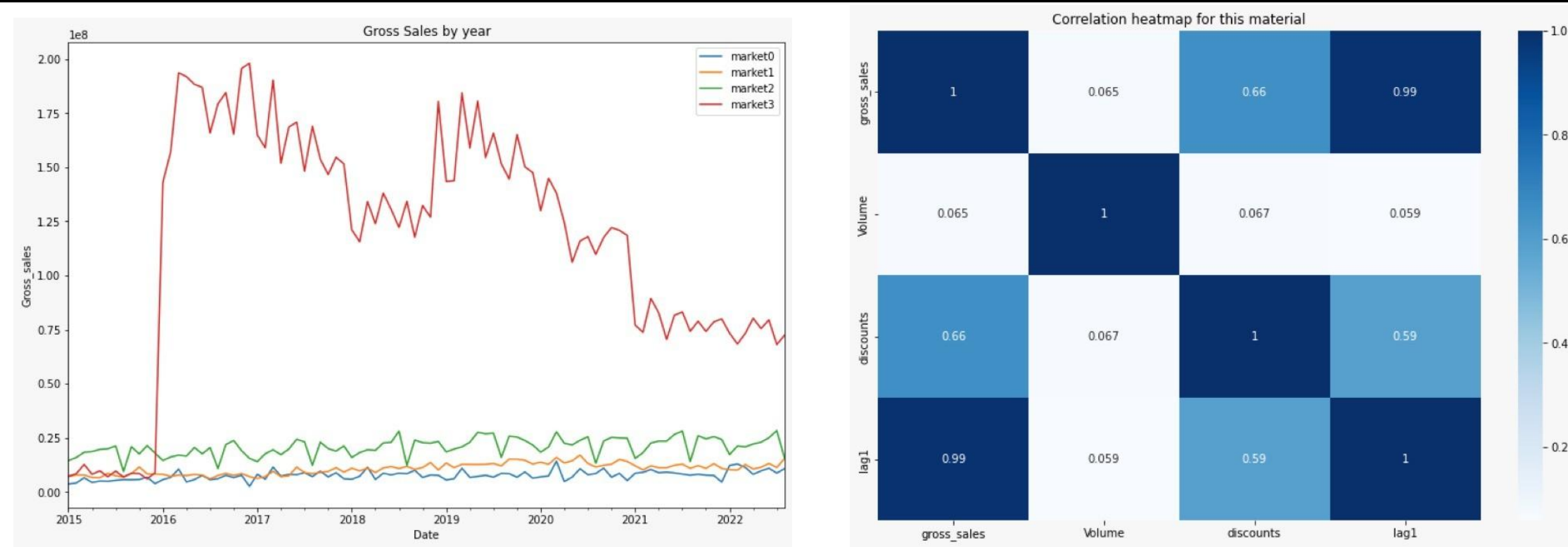
Problem Statement

- Novartis is a Swiss-American pharmaceutical corporation
- Problem: Identify main key drivers of sales and profit
- Deliverable: Evaluate scenarios to maximize sales by efficiently identifying key contributors

Data Preprocessing

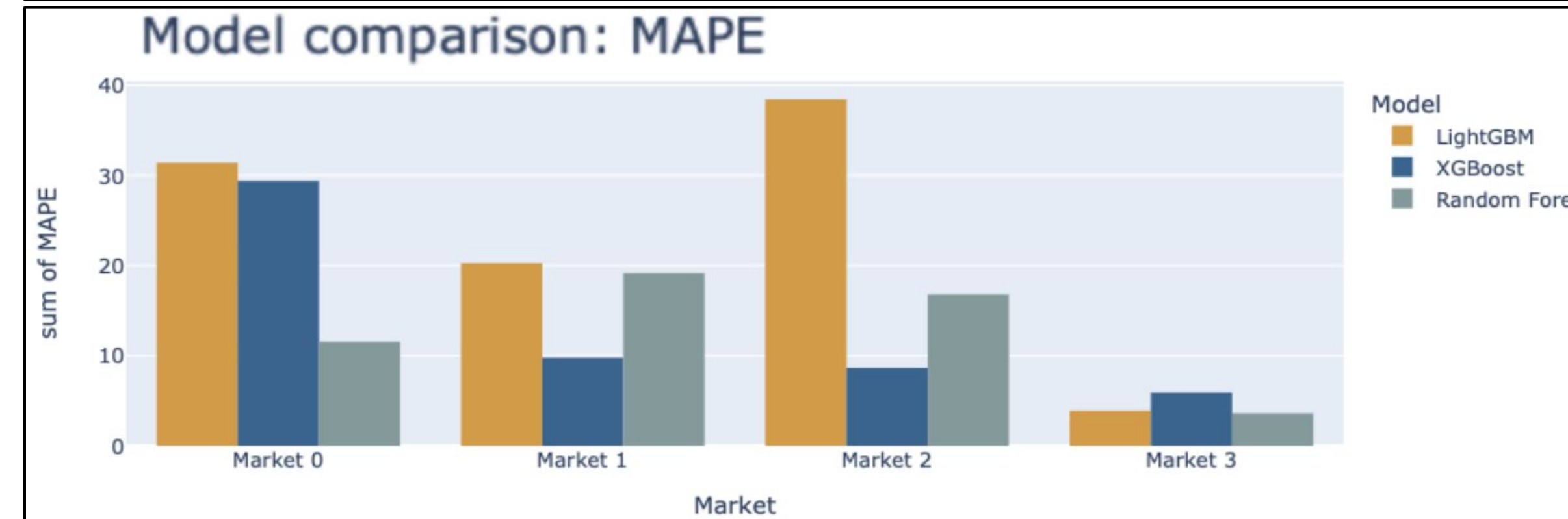
- Target Encoding - Material and market ciphers
- Dropped rows where sales or discounts are zero, and maintained thresholds to consider deviations between gross and net sales
- Took absolute values of discounts relative to Volume
- Feature Engineering - Lag, Percentage change in sales for past months - for each material and market combination

Exploratory Data Analysis



- Gross Sales for Market 3 are much higher and might be a possible outlier comprising of high volume and cost sales
- Correlation matrix shows that gross sales are correlated with the discounts, volume and the lag parameters

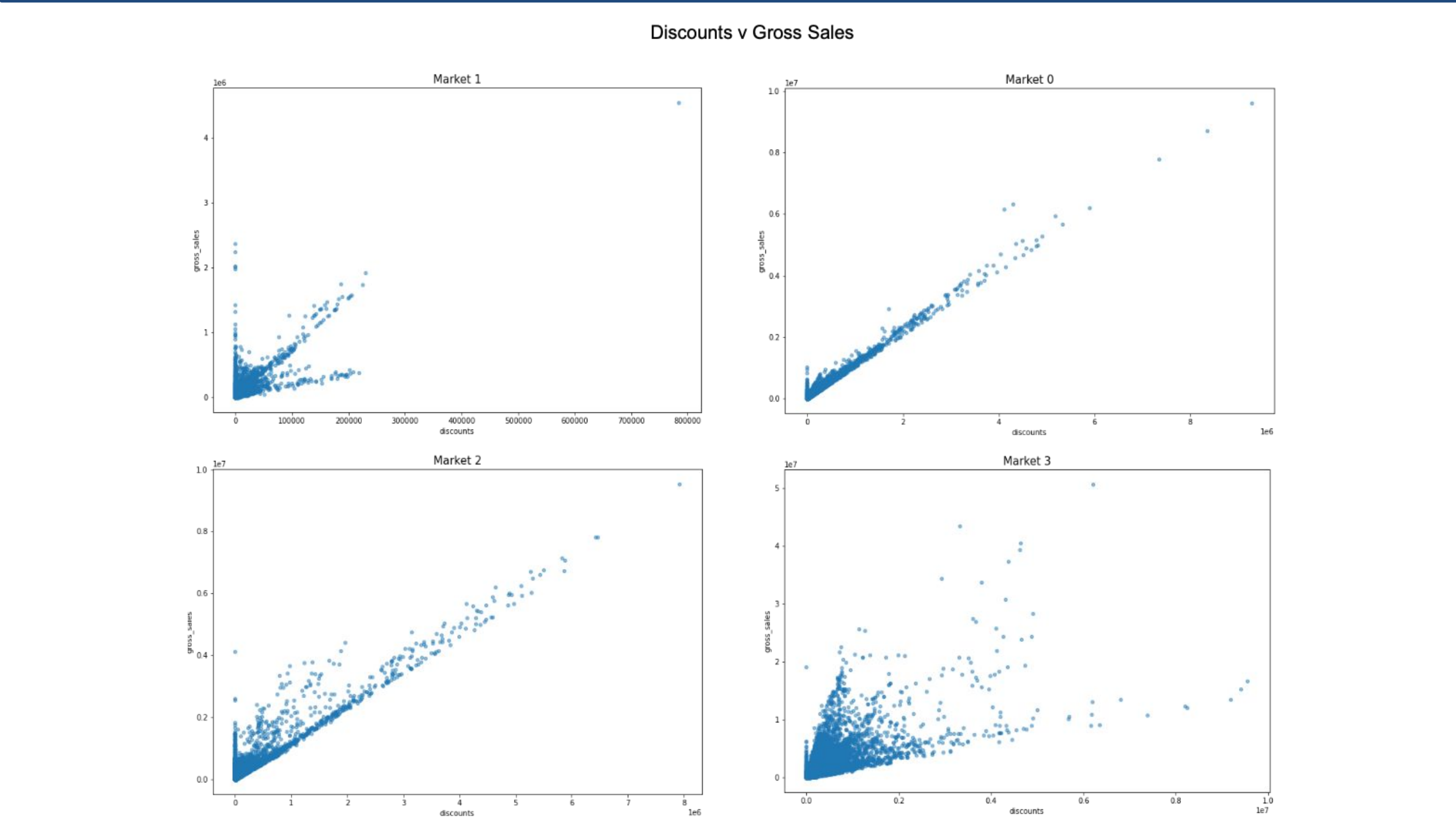
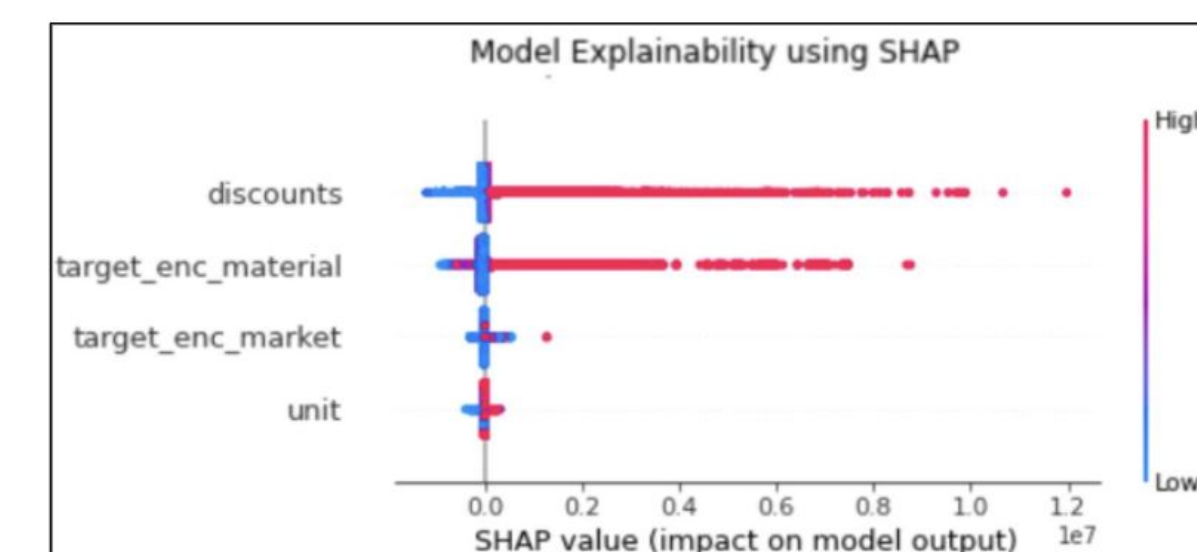
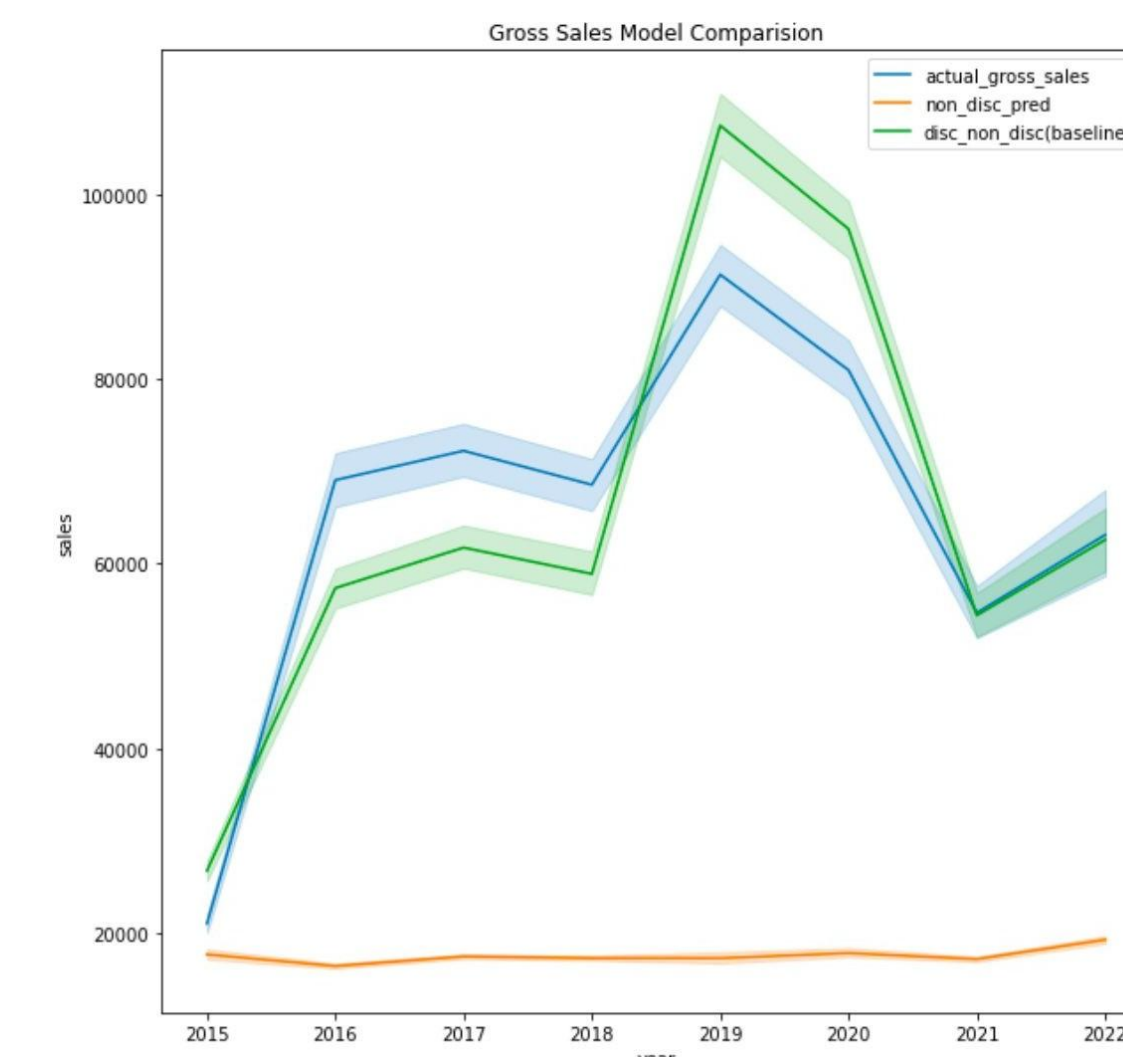
Modelling



Market	Inferences
Market 0	R2 is consistent for all the models, however Random Forest has the lowest MAPE value
Market 1	XGBoost is better in terms of R2 and MAPE as compared to other models
Market 2	Although Random Forest has descent MAPE, it has the highest R2
Market 3	MAPE is consistent across models, XGBoost and Random Forest have similar R2 values

Model Explainability

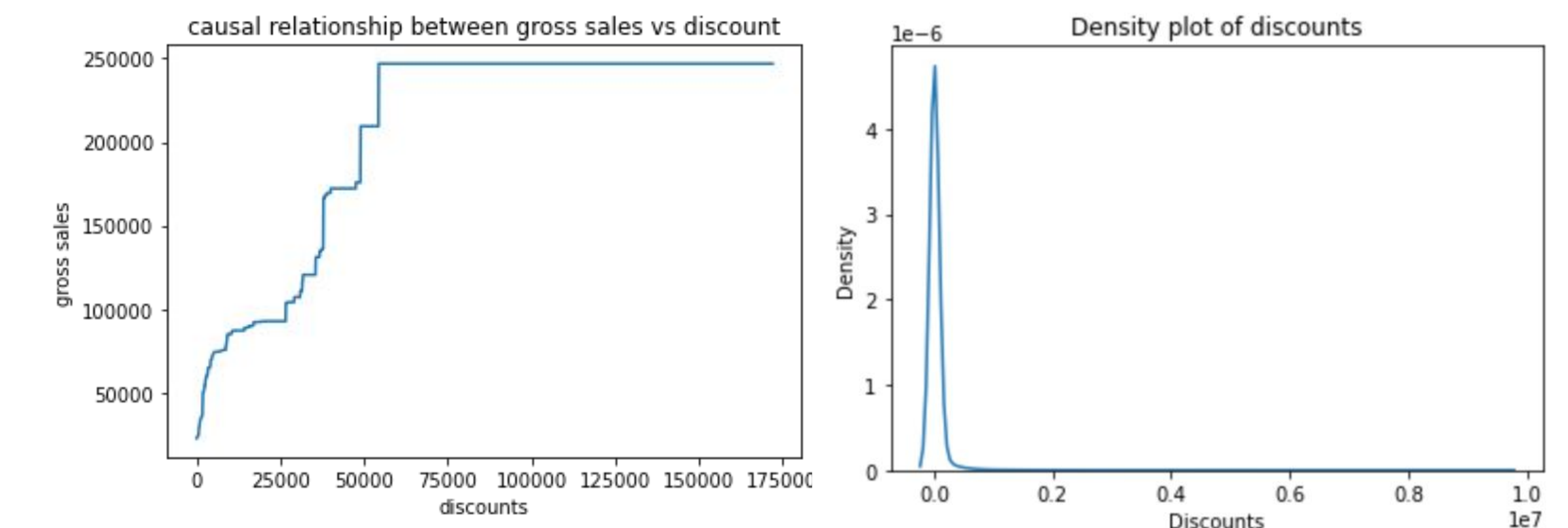
- Two models were trained - One using the entire dataset and other only on sales where the discounts were not applied, gap in performance was noted on the test set
- Applied SHAP values to understand feature importance. Observed that Discounts and Encoded Material columns have a positive effect in predicting gross sales



- Market0: Linear trend, with many lower values of discounts and sales
- Market1: 2 trends, high variance in low values
- Market2: Linear trend in outliers and high discount values. Higher variance in lower discounts
- Market3: Highest discounts and most sales

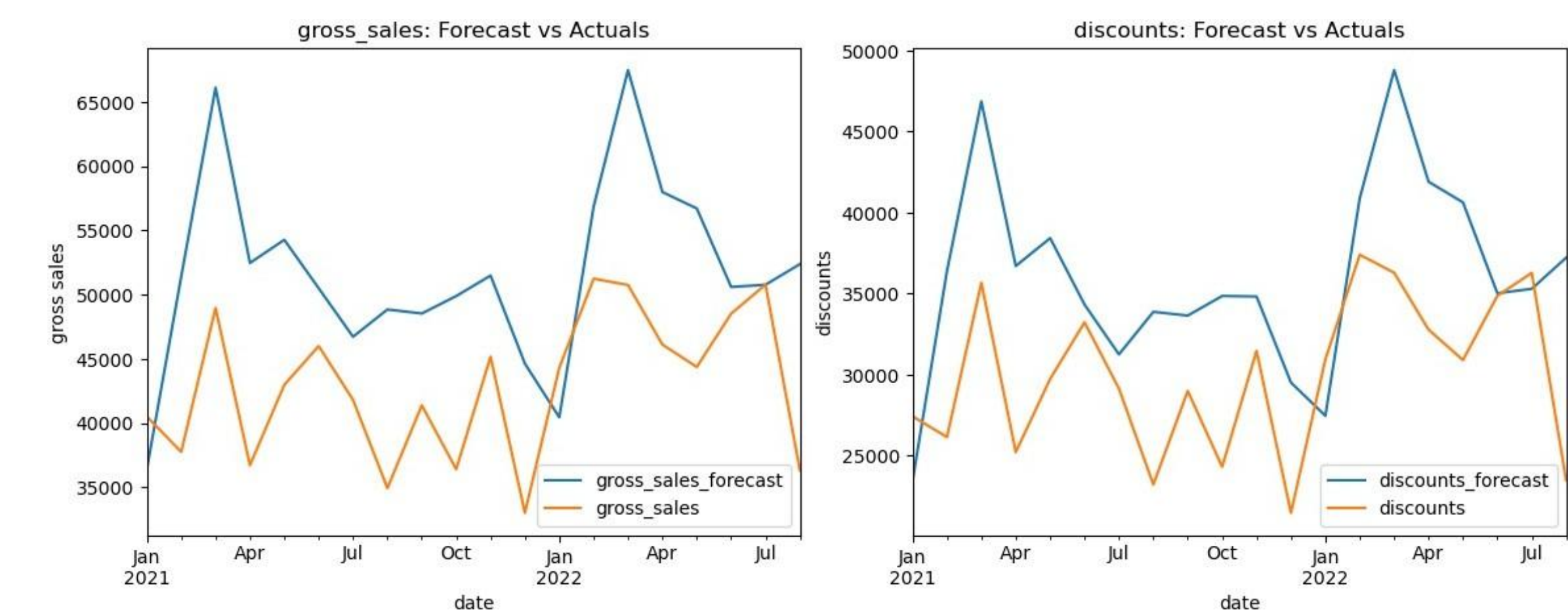
Causal Inference Approach

- Causal exploration was aimed to validate if discounts and gross sales are causally related
- Observed that discounts allocated after certain threshold do not contribute to any change in the gross sales



Time Series Approach

- The VAR model forecasts on the basis of previous values and does not allow additional features
- Incorporating additional features using the Facebook prophet model didn't improve the performance



Conclusion

- Causal analysis confirmed a causal relationship between gross sales and discounts
- After feature engineering, boosting algorithms obtained best results on a market level by reducing overall bias
- Time series modeling results didn't perform well as expected due to unavailability of data on a daily basis

Acknowledgements

Sincere thanks to Dr. Laura Sanchez Garcia, Julio Antonio, Dr. Adam Kelleher and Dr. Eleni Drinea for their constant support, expert guidance with the project and resources

References

- Sontag, David & Shalit, Uri & Johansson, Fredrik. (2017). Estimating individual treatment effect: Generalization bounds and algorithms
- Lundberg, Scott & Lee, Su-In. (2017). A Unified Approach to Interpreting Model Predictions