Improving Hospitality Spend Management with Machine Learning-Based Predictive Analytics

Nilabh Chaturvedy, Payal Pandya, Sarnav Chauhan, Supriya Malla, Tanveer Singh Sapra, Yang Wang

chaturvn@purdue.edu, pandyap@purdue.edu, chauha37@purdue.edu, malla2@purdue.edu, tsapra@purdue.edu, yangwang@purdue.edu

PURDUE UNIVERSITY® | Mitchell E. Daniels, Jr. School of Business

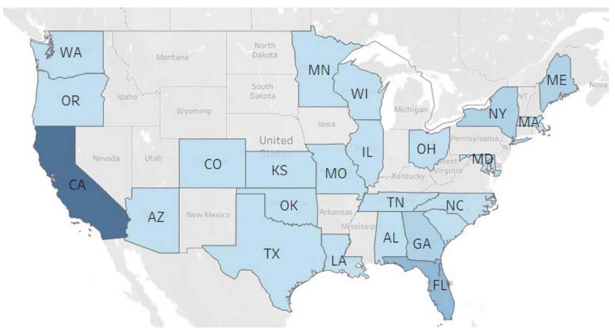
ABSTRACT

This study aims to identify and predict the annual spending capacity of various hotels across the USA. We analyzed the spending patterns of 11,000 hotels, each belonging to 12 different market segments. Analysis of the data spanning 6 years revealed that hotel spending varied across 30 different product/service categories. For each of these combinations, we implemented a customized forecasting model to accurately predict a hotel's annual expenditure. Through feature selection and variable significance analysis, we also identified key hotel characteristics that lead to greater annual spending.

INTRODUCTION

The hotel procurement industry is crucial for its operations and strategy. This project was motivated by the need for a leading hospitality procurement agency in North America to analyze partner hotels' expenditures with their supplier networks. The industry partner currently uses a 20-year-old forecasting model and aims to improve its accuracy. By analyzing spending patterns across a combination of different product/service categories and market segments, the client wished to accurately predict the potential spending.

- 6.1M Observations
- 12 Market Segment
- 11K Census Data
- 30 Product Category
- 19 Features
- 10 models each

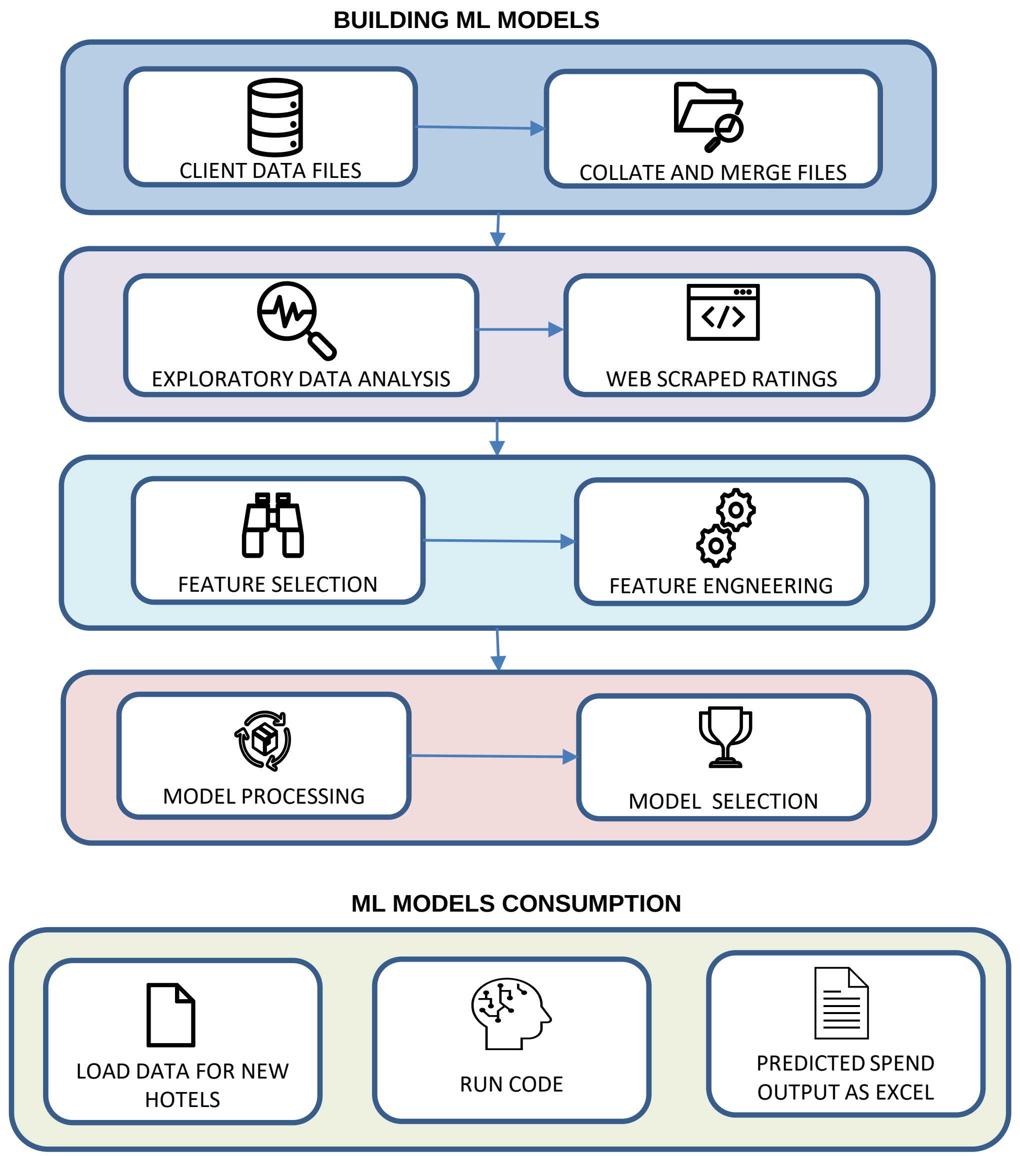


Average hospitality spend density across states in the US

RESEARCH OBJECTIVES

- Determine the features (new or existing) that impact the spending prediction.
- Build a tailored model that caters to distinct market segments individually.
- Overhaul the legacy model into an intuitive solution for streamlined user experience.

METHODOLOGY

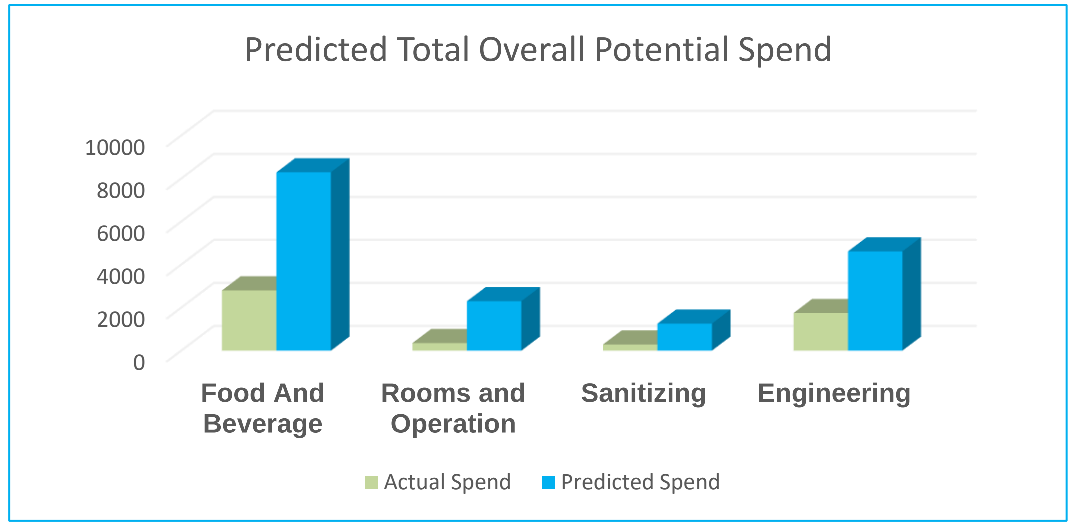


STATISTICAL RESULTS

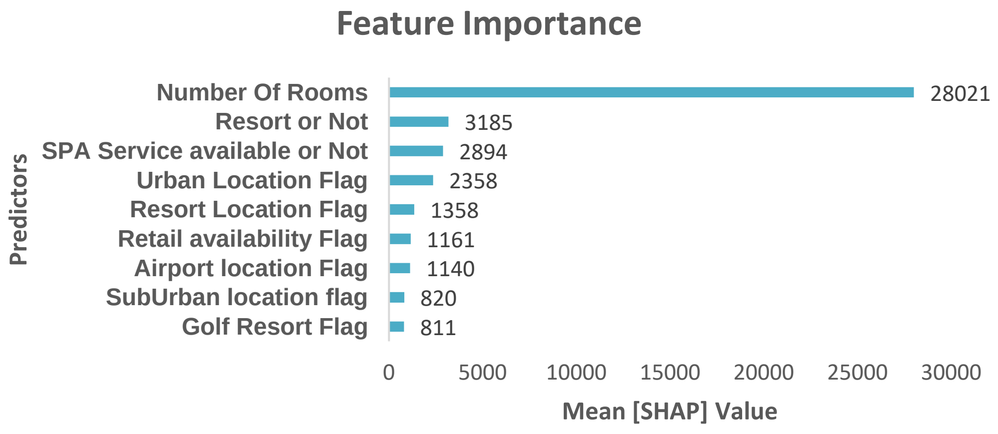
We used Root means Squared Error as the evaluation metric for selecting the best model. For each category and market segment combination, we compared the metric against baseline multivariate Linear Regression model(legacy model being used by client) to benchmark the performance on the validation set.

Product Category	Best Model Name	Linear Regression RMSE	Best Model RMSE
Food And Beverage	Ensemble	8290	2802
Operation Management	XG Boost	2299	354
Cleaning Equipments	XG Boost	1257	287
Maintenance Services	CAT Boost	4614	1751

For the above categories, the below table depicts the average actual vs predicted spend.

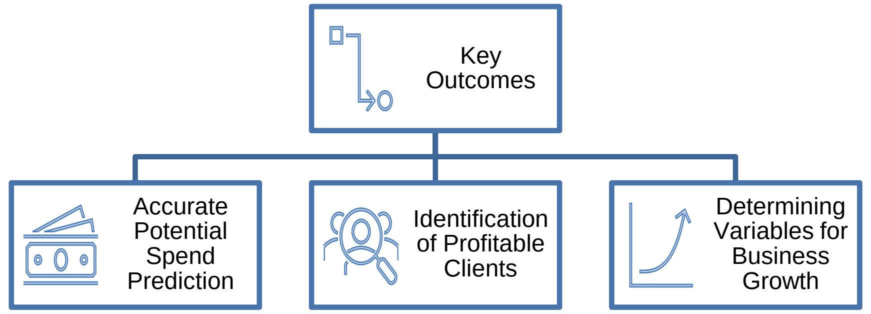


We extracted the feature importance of the best model using mean SHAP value, that gives us the feature that are important for predicting the spend for that category. This can help clients to identify the attributes of property which can be focused on, for increasing the overall spend of the property



EXPECTED IMPACT

- Enabled our industry partner to obtain predicted spend stratified by geography, facilities like swimming pool, resort, etc. and number of rooms.
- We built a solution **66% more accurate** than the legacy model, identifying and leveraging the critical features for each product category and market segment.
- An accurately predicted potential spend can help our client to identify the most profitable businesses and minimize lost opportunity cost.



CONCLUSION & RECOMMENDATIONS

Our solution is closely aligned with our industry partner's strategy to procure more business opportunities and be better prepared for annual planning.

This model is feasible, scalable and can be applied in the future as well.

Our suggestion to the partner was to include more variables that impact potential spend so that the model accuracy can be further improved. These are summarized below:

- Add real-time Occupancy data, and Average Daily Rate for each property.
- Include property attributes that do not change frequently for the property. Examples include the age of the property, acreage, number of lawns and gardens.

ACKNOWLEDGEMENTS

We would like to thank our industry partner for their guidance and support on this project as well as the Purdue MS BAIM program for partially funding this work.

informs BUSINESS ANALYTICS CONFERENCE