Strategic Thinking - Capstone Project

US Candy Distributor: Sales Prediction and Demand Forecast

CCT College Dublin May 2025

ABSTRACT

This project was carried out to help a US-based sweet distributor improve sales, stock management, and profits using data analysis and machine learning. By studying sales trends, and forecasting demand, it provided clear insights for better decisions.

The analysis showed strong seasonal sales peaks during holidays like Halloween and Christmas. The Gradient Boosting model was best for predicting sales, while the SARIMA model effectively forecasted future demand by spotting trends and seasonal changes. It also found that cheaper products had slightly better profit margins, and shipping methods had little impact on profits. Regions like California and Washington showed strong sales, offering chances for targeted marketing and stock planning.

These insights enable the distributor to plan stock and promotions better, reduce waste, and increase profits. Future work could focus on real-time demand tracking and customer segmentation for smarter decisions. Using data-driven methods helps the business stay efficient and competitive

RESEARCH QUESTIONS

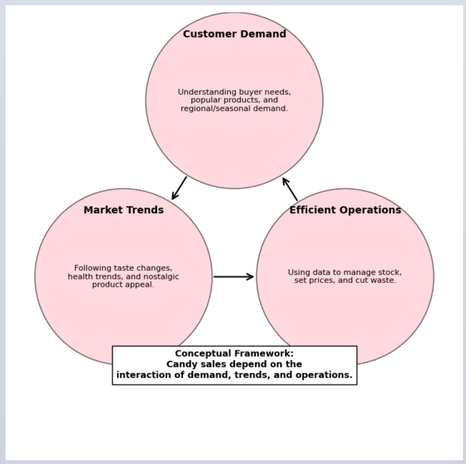
This study focuses on key questions to help a US-based candy distributor improve sales, stock management, and profits:

- Q.1. How do sales and profits vary across different regions, products, and customer groups?
- Q.2. Which factors, such as shipping methods, product pricing, or seasonal trends, most impact profitability?
- Q.3. Can machine learning and time-series forecasting accurately predict future demand to optimise stock levels?
- Q.4. What strategies can the business adopt to increase sales of low-performing products and maximise profits in high-performing regions?

CONCEPTUAL FRAMEWORK

This framework is based on the idea that successful candy sales depend on three key factors: customer demand, market trends, and efficient operations. These elements interact to shape sales performance, stock management, and profitability.

- Customer Demand – Understanding what buyers want, including popular products, seasonal preferences, and regional differences.
- Market Trends – Tracking changes in tastes, such as demand for healthier options or nostalgic treats, to stay competitive.
- Efficient Operations – Using data-driven decisions to optimise stock levels, pricing, and shipping, reducing waste and costs.



RESEARCH PARADIGM

This study follows a pragmatic research paradigm, combining both inductive and deductive approaches to analyse sales data and improve decision-making. The pragmatic approach allows for practical, real-world solutions by using data-driven insights rather than relying solely on theory or subjective interpretations.

Key aspects of the paradigm include:

- **Data-driven analysis**: Using historical sales data to identify trends, test hypotheses, and build predictive models.
- **Mixed-method evaluation**: Combining statistical analysis, machine learning, and time-series forecasting to derive actionable insights
.
- **Practical application**: Focusing on solving real business challenges, such as stock management, pricing strategies, and demand forecasting.

Research Methods

This study used **a data-driven approach** to analyse sales patterns and improve decision-making for a US candy distributor. The methods **combined quantitative analysis** with **machine learning techniques** to uncover trends and predict future demand.

Category	Method/Tool	Details
Data Analysis	Python (Pandas, NumPy)	Examined sales, profits, and regional performance
Machine Learning	Linear Regression, Decision Trees, Gradient Boosting	Predicted sales using optimised models
Time-Series Forecasting	SARIMA (Seasonal ARIMA)	Forecasted seasonal demand (e.g., Halloween, Christmas)
Hypothesis Testing	ANOVA, t-tests	Tested impact of shipping methods & product costs on profits
Process	Data Collection	Real-world sales data (2020-2024) from Maven Analytics
	Exploratory Analysis	Identified trends (best-sellers, peak seasons)
	Model Development	Optimised sales/stock predictions
	Validation	Accuracy testing & refinement

Data Collection

This study used real-world sales data from a US candy distributor to analyse trends and improve business decisions. The data was collected over four years (2020–2024) and included:

Category	Details	Analysis Type
Sales Transactions	Order dates, quantities sold, revenue, costs, and profits	*Quantitative*
Product Information	Types of sweets (e.g., chocolate, gummies), production costs	*Quantitative*
Customer Locations	Countries, regions, and cities where orders were placed	*Geographic*
Shipping Methods	Delivery types (e.g., Standard, Express), costs, and times	*Operational*
Factory Performance	Production sites, regional sales, and efficiency metrics	*Comparative*
	Primary Dataset: Candy_Sales.csv (from Maven Analytics)	*Data Sources*
	Tools: Python (Pandas, NumPy) for cleaning and analysis	*Methodology*
	Why This Data? Helps track best-selling products and profit margins	*Business Value*

FINDINGS

Reflective Learning

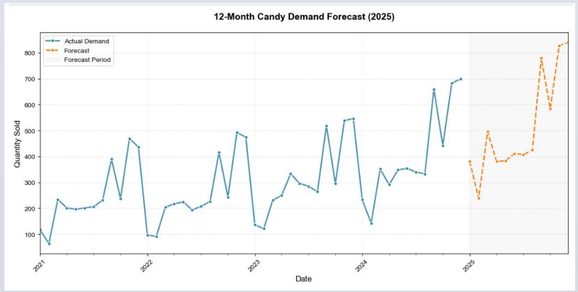
This project used a data-driven approach to analyse candy sales and improve business decisions. By examining sales patterns, testing hypotheses, and forecasting demand, we gained valuable insights into what drives profitability and customer behaviour.

Section	Insight	Details
Key Learnings	*Data is Powerful*	Historical sales data revealed clear trends (seasonal spikes like Halloween/Christmas, regional differences e.g., US vs Canada)
	Machine Learning Works Best	Gradient Boosting outperformed other models, showing how advanced tools uncover hidden patterns
	Small Changes Matter	Cheaper products had slightly higher profit margins (86% vs 84%), suggesting pricing strategies need fine-tuning
	Shipping Methods Don't Affect Profits	Surprisingly, faster shipping didn't increase profits—customers were willing to pay extra for speed
Challenges & Improvements	*Data Quality*	Cleaning and preparing data took significant time but was crucial for accuracy
	Model Tuning	Hyperparameter adjustments helped, but feature engineering (e.g., identifying bulk orders) was equally vital
	Future Steps	Real-time tracking and customer segmentation could refine forecasts further
Reflection	*Project Outcomes*	Analysis provided actionable insights but required balance to avoid overcomplicating methods
	Practical Takeaways	Focus on high-margin states (e.g., Washington) and boosting low-sellers (e.g., Fun Dip) delivered most value
	Overall Lesson	Combining data science with business questions reduces waste, optimises stock, and increases profits

Machine Learning Model, Best performance>)

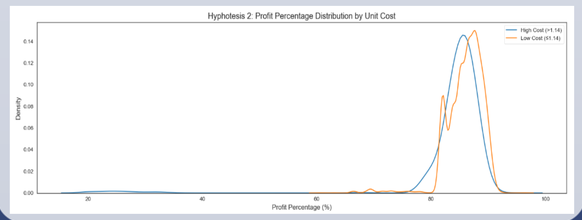
Model	Train RMSE	Test RMSE	CV RMSE Mean	CV RMSE Std	Train R2	Test R2
Gradient Boosting	0.303	0.314	0.321	0.009	0.693	0.677
Random Forest	0.263	0.325	0.333	0.008	0.769	0.652
Decision Tree	0.323	0.323	0.335	0.009	0.651	0.657

Time Series Demand Forecast



Social Learning (PBL and PAL activities)

that cheaper products currently perform better in terms of profit percentage. This insight could be important for Candy Distributor when deciding whether to adjust pricing or operations depending on the cost tier of their products.



Recommendations:

- **Stock Management:** Prepare for holiday surges and reduce orders in quieter months.
- **Pricing Strategy:** Adjust prices on low-performing products and focus on high-margin items.
- **Regional Focus:** Invest more in top-performing states like California and Washington.
- **Future Improvements:** Explore real-time demand tracking and customer segmentation for finer insights.

CONCLUSIONS

This study used data analysis and machine learning to help a US candy distributor improve sales, manage stock, and increase profits. The key findings and recommendations are summarised below:

Key Findings:

Sales Patterns:

The US market outperformed others, with California and Washington being top states for sales and profit margins.

Seasonal peaks occurred in September, November, and December (holiday seasons), while demand dropped in January and February.

Product Performance:

Cheaper products had slightly higher profit margins (86%) than expensive ones (84%).

Fun Dip and Nerds were the lowest-selling products and may need promotions or price adjustments.

Shipping & Costs:

Shipping methods did not significantly affect profitability—customers paid extra for faster delivery, balancing costs.

Wicked-Chacys factory generated the highest sales, suggesting other factories could learn from its strategies.

Predictive Insights:

The Gradient Boosting model was most accurate for sales predictions, focusing on customer spending habits.

SARIMA forecasting reliably predicted seasonal demand, helping the company plan stock and reduce waste.

REFERENCES

Frank, A. and Scott (2015). *Capstone Title Marketing Research and Analytics Database*. [online] Available at: https://coeng.uaa.alaska.edu/innovation/wp-content/uploads/2020/10/2015-Marketing-Research-and-Analytics-Database-Final_Report_Complete_V2.pdf

Mordorintelligence.com. (2019). *Mordor intelligence*. [online] Available at: https://www.mordorintelligence.com/industry-reports/candy-market.

Haveignition.com. (2024). *Go-to-Market Strategy for candy*. [online] Available at: https://www.haveignition.com/industry-guides/go-to-market-strategy-for-candy.