Advanced Real Estate Analytics for Retail
Determining Express Locations

Veda Samhitha Alluri, Shubhankar D Bhajekar, Harshraj Vijaysinh Jadeja, Srinija Srimamilla, Husam Saleh, Khan, Shoaib Amjad
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
alluriv@purdue.edu; sbhajeka@purdue.edu; hjadeja@purdue.edu, ssrimami@purdue.edu; hasaleh@purdue.edu, khan180@purdue.edu

ABSTRACT

In the competitive landscape of the retail industry, the strategic selection of store locations is crucial for business success. This paper presents a comprehensive study on developing a predictive analytics model to identify the most profitable locations for new convenience store outlets within a large retail chain. Amalgamating rich external data gleaned through web scraping with in-depth proprietary data, we examine factors such as traffic volume, proximity to highways, nearby amenities, and customer sentiment metrics at a granular latitude-longitude level. The model incorporates both proprietary data from the retail chain and public domain data, including demographic insights and competitor analysis. Our research contributes to the existing body of knowledge by not only advancing the methodological approach in retail analytics but also by addressing the gap in predictive accuracy concerning the store location value. The findings are expected to underpin strategic decision-making for retail expansion, offering a scalable solution that can be adapted nationally. The implications of this study are significant, providing a data-driven foundation for optimizing store placement and enhancing customer reach.

Keywords: Retail Analytics, Predictive Modelling, Retail Site Location, Location Analytics

INTRODUCTION

In the fast-paced and competitive world of retail, the selection of store locations is a critical strategic decision that significantly influences the success of retail chains. As markets become more saturated and consumer preferences shift, the ability to choose optimal store locations based on a deep understanding of market dynamics, consumer behavior, and competitive landscapes becomes increasingly important. This paper aims to employ predictive analytics and advanced data science techniques to refine the process of retail site selection, providing retail chains with a powerful tool to enhance their market penetration and profitability.

The importance of location in retail cannot be overstated. A store's location directly affects its accessibility to customers, its exposure to potential new customers, and its ability to compete with other retailers. As such, the strategic selection of retail locations has long been considered both an art and a science. Traditional methods for site selection have relied heavily on demographic data, competitor locations, and basic traffic patterns. However, with the rise of big data and advanced analytics, there is an opportunity to revolutionize this traditional approach by integrating more sophisticated data sources and analytical techniques.

Recent trends in the retail industry underscore the need for a more data-driven approach to site selection. According to a report by Forbes, retailers are increasingly turning to location analytics to understand consumer behaviors and preferences more deeply, allowing them to make more informed decisions about where to open new stores[1]. Similarly, a Gartner study highlighted the transformative impact of integrating geospatial data with traditional market analysis, which has proven to enhance the effectiveness of site selection processes[2]. Additionally, The Wall Street Journal has reported on the growing trend of using social media and online traffic data to assess the potential success of retail locations, showcasing the expanding range of data sources that retailers can leverage[3].

The central research question this paper addresses is: How can predictive analytics refine the site selection process for retail chains aiming to expand in competitive markets? This question is particularly pertinent in the current retail landscape, where the difference between a successful store and a failing one can often be attributed to the precision of its location choice. This research seeks to develop a scoring algorithm that quantifies the potential value of various locations, integrating diverse data sets that include real-time traffic data, demographic insights, consumer behavior analytics, and competitive positioning.

This study is significant for several reasons. First, it addresses a fundamental challenge in retail management, offering a methodological improvement that can lead to more precise and informed decision-making. Second, it contributes to the academic literature on retail analytics by demonstrating how various types of data can be synthesized to forecast store performance. Lastly, for practitioners in the retail industry, this research provides a scalable solution that can be adapted to different markets and geographic areas, potentially revolutionizing their approach to expansion and strategic development.

The methodology of this research involves several key components. Initially, data collection will focus on gathering a comprehensive set of variables that influence retail success. These include traditional factors such as demographic characteristics and proximity to competitors, as well as more nuanced data such as customer online search behavior and sentiment analysis derived from social media. Following data collection, sophisticated data preprocessing techniques will be applied to ensure the quality and consistency of the data set. Exploratory data analysis will then be conducted to identify patterns and relationships within the data.

Subsequently, the model-building phase will involve the application of several advanced statistical and machine learning techniques. These methods are chosen for their ability to handle large, complex datasets and their proven effectiveness in predictive accuracy. The predictive model developed will then be validated using rigorous statistical methods to ensure its reliability and accuracy.

The results section of the paper will detail the outcomes of the predictive model, including its accuracy metrics and the insights it provides into the factors that most significantly impact retail site success. This will include a discussion of the model's implications for retail strategy and its potential to be integrated into the decision-making processes of retail chains.

Finally, this study will conclude with a comprehensive summary of the research findings, a discussion of the theoretical and practical implications of the study, and suggestions for future research in the area of retail analytics.

In summary, this introduction sets the stage for a detailed examination of how advanced data analytics can be generatively applied to one of the most critical aspects of retail management—site selection. By bridging the gap between traditional practices and cutting-edge data science, this research aims to provide significant contributions to both academic literature and retail industry practices, offering new insights and methodologies that can enhance the strategic decision-making process in retail location selection.

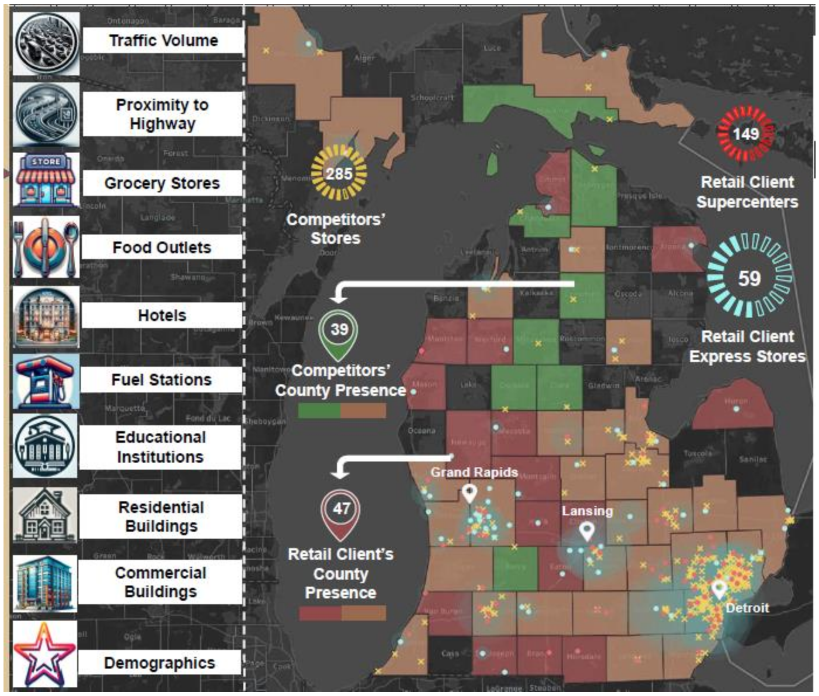

Figure 1: Retail Store Location across a state in the US
(Store Attributes generated by DALLE)

LITERATURE REVIEW

The literature study encompasses an array of scholarly works that delve into various dimensions of retail analytics, each contributing valuable insights pertinent to the development of a scoring algorithm for retail site selection. These works, collectively, provide a multi-faceted understanding of retail dynamics, from historical methodologies to modern analytical approaches. The review navigates through classic theories, contemporary data scraping techniques, and customer satisfaction metrics, all essential in shaping a robust predictive model for a retail chain's expansion strategy.

Historical Methodologies in Retail Analytics

The foundational work by Applebaum (1966) presents a pioneering exploration of retail site selection, emphasizing demographic considerations and traffic patterns. While Applebaum's approach marks the inception of structured retail location analytics, it operates within the limitations of its time, lacking the predictive capabilities offered by contemporary data science. This historical perspective, however, lays the groundwork upon which this project builds, incorporating technological advancements to enhance predictive accuracy.

Modern Analytical Techniques

In the sphere of contemporary retail analytics(Quang Thai et al., 2015), the application of web scraping and API services, as discussed in recent studies, is particularly relevant. These techniques facilitate the extraction of vast quantities of data relevant to retail site selection, such as customer reviews and competitor analysis. By integrating these real-time, granular data points, the present research goes beyond traditional static factors, adopting a more dynamic and holistic approach to predictive modeling. Le and Pishva (2015) demonstrate the use of web scraping and Google API services to optimize the distribution routes for convenience stores in Japan.

Customer Satisfaction and Retail Success

Udokwu et al. (2020) assess various techniques for capturing customer satisfaction data, with a focus on using social media and review websites as primary data sources. The importance of customer satisfaction data in retail supply chains, as evaluated in certain studies, underscores the critical role of consumer feedback in determining store performance. This project acknowledges the need to incorporate customer sentiment, extending these techniques into the realm of retail site analytics to predict store success more accurately. By doing so, it addresses a gap in the literature, where the direct application of customer satisfaction to site selection is not thoroughly explored.

Spatial Analysis in Retail Site Selection

Spatial location analysis provides a comprehensive review of the application of GIS and location intelligence in retail site selection. The research highlights the integration of spatial data with business analytics to inform strategic decision-making. Aboulola (2017) explores the strategic importance of spatial location analysis in retail site selection, emphasizing the integration of Geographic Information Systems (GIS) and multi-criteria decision-making models. The review identifies that location plays a pivotal role in the sustainability and profitability of retail outlets, highlighting cases where poor location decisions have led to the closure of stores.

Deep Learning in Retail Sales Forecasting

Eglite and Birzniece (2022) comprehensively examines applications of deep learning (DL) for retail sales forecasting across 19 recent studies. Their analysis reveals that specialized DL architectures, particularly recurrent neural networks like LSTM, demonstrate significant predictive accuracy improvements over conventional machine learning and statistical approaches. More than half the papers develop custom frameworks tailored to the business forecasting context. While confirming the power of DL models, the authors also highlight lingering challenges like complexity and interpretability. The review covers retail sectors like grocery, e-commerce, and apparel, citing evidence of DL's versatility. However, opportunities remain in applying DL for geographic sales projections and enhancing model transparency. As such, this paper endorses the potential of specialized DL models in retail forecasting while recognizing gaps our location-aware, interpretable approach can fulfill.

Optimal Site Selection

Ahmad and Chua (2018) explores the use of location analytics to determine optimal sites for new retail establishments. This research employs modern data collection techniques like Google Maps to assess location characteristics relevant to different types of businesses, such as visibility from main roads and parking availability. It leverages data mining methods, including association rule mining, to identify key location attributes that correlate with successful business types. The findings suggest that certain physical characteristics significantly influence customer inflow, which can be crucial for businesses in selecting new sites. This work contributes to the broader understanding of strategic retail placement, providing a more data-driven approach compared to traditional methods, which rely heavily on manual surveys and demographic analysis.

Geographic Information Systems (GIS) in retail location decision-making

Namangale (2022) highlights the application of GIS technologies in retail property investments. This approach utilizes spatial data to visualize and analyze the geographic patterns and relationships that influence retail site selection. The study underscores the utility of GIS in identifying optimal locations by evaluating factors such as customer demographics, traffic patterns, and proximity to other businesses. Such insights are pivotal for your project, demonstrating how spatial analysis tools can be leveraged to improve the accuracy and strategic value of retail location decisions.

Retail Store Location Recommendation using Location-based Services

Chen (2020) explores the role of location-based services (LBS) and mobility analytics in the retail sector. The paper presents a methodological framework for utilizing real-time mobility data to recommend retail store locations, emphasizing the importance of dynamic consumer data in retail site analytics. This research complements traditional location decision strategies, suggesting that the integration of LBS can provide a more responsive and informed approach to selecting retail sites, aligning closely with your project's use of real-time traffic and customer movement data to enhance location decision models.

New Boundaries of Retail Location Decision Making

In a broader context, the transformative impact of big data in retail location decision-making is discussed extensively in the literature. Studies have shown how big data analytics can uncover hidden patterns and correlations that significantly affect retail strategies (Aversa et al., 2018). This body of work illustrates the critical role of big data in improving retail location decisions, providing a comprehensive view of customer preferences and market dynamics that can inform more effective retail strategies. For your project, this underscores the importance of integrating a variety of data types, including demographic insights, traffic volumes, and customer behavior patterns, to optimize retail location selection.

Table 1: Literature Study Summary

Title	Author(s)	Summary
Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales	William Applebaum (1966)	Pioneering retail site selection, demographic considerations
Application of Web Scraping and Google API Service to Optimize Convenience Stores' Distribution	Quang Thai Le and Pishva (2015)	Web scraping, API services, real-time data
Evaluating Technique for Capturing Customer Satisfaction Data in Retail Supply Chain	Udokwu et al. (2020)	Social media, review websites, customer sentiment
A Literature Review of Spatial Location Analysis for Retail Site Selection	Aboulola (2017)	GIS, location intelligence, multi-criteria decision-making
Retail Sales Forecasting Using Deep Learning: Systematic Literature Review	Eglite and Birzniece (2022)	Deep learning, recurrent neural networks, LSTM
Location Analytics for Optimal Business Retail Site Selection	Ahmad and Chua (2018)	Location analytics, data mining, association rules
Location Analytics for Retail Property Investment- A GIS Based Approach	Namangale (2022)	GIS, spatial data, retail property investments
Retail Store Location Recommendation using Location-based Services	Chen (2020)	Location-based services, mobility analytics, real-time data
Big Data Analytics: The New Boundaries of Retail Location Decision Making	Aversa et al. (2018)	Big data analytics, customer preferences, market dynamics

DATA

This research focused on leveraging publicly available data, meticulously gathered through web scraping techniques and API integrations, to construct a preliminary dataset. This dataset encompasses variables critical to retail site selection, such as traffic volume, distance to highways, and proximity to key amenities including groceries, restaurants, and other local services. This externally sourced dataset is a foundational layer for our analysis, enabling initial insights into location attractiveness and potential customer reach. This approach underscores the value of accessible digital data in informing retail strategic planning. The data was structured into three distinct data frames to facilitate analysis: one for the Retail chain, one for competitors, and a master data frame that integrates all collected data.

The data was divided into two main categories:

Geospatial Data: This includes traffic volume, distance, and duration to highways, and nearby places. Traffic volume data was sourced from state Departments of Transportation, using tools like the Department of Transportation website where specific latitudes and longitudes could be input to retrieve traffic counts. This data provides insights into the volume of potential customers passing by each location daily. Additionally, using OpenStreetMap and the Overpass API in Python, the nearest highway exit ramps were identified, and distances as well as durations from these points to the store locations were calculated.

Customer Data: This includes county demographics such as population, number of households, income levels, household value, age, and gender. This data helps in understanding the potential customer base in the vicinity of each store.

Proximity Analysis

The proximity to various amenities was categorized based on radii around each store location—1 mile, 3 miles, and 5 miles. Each radius was analyzed for the presence of different amenities:

- Within one mile: grocery stores and fuel stations, which are essential for daily needs and can drive regular foot traffic.
- Within three miles: apparel stores and hospitals, catering to broader shopping and healthcare needs.
- Within five miles: larger attractions like tourist destinations and airports, which can influence sporadic but significant traffic increases.

Each store's data, whether Retail chain or competitor, was methodically integrated with the geospatial and customer data to build a comprehensive view of the store environment. This structured approach not only simplifies the analysis but also ensures that data-driven decisions are based on a holistic view of each location's potential to attract customers.

Table 2: Geospatial and Customer-centric data

Variable Name	**Variable Type**	**Variable Description**
Name	String	Name of the location/store
Address	String	Physical address of the location
Latitude	Float	Geographic latitude coordinate
Longitude	Float	Geographic longitude coordinate
County	String	County in which the location is situated
Distance to Nearest Retail chain (Straight-line)	Float	Linear distance to the closest Retail chain store
Near Supercenter	Boolean	Indicates proximity to a supercenter
Nearest Motorway Entrance Lat	Float	Latitude of the closest motorway entrance
Nearest Motorway Entrance Lon	Float	Longitude of the closest motorway entrance
Driving Distance to Nearest Motorway	Float	Road distance to the nearest motorway
Duration to Nearest Motorway	Integer	Time in minutes to the nearest motorway
Traffic (AADT)	Integer	Average Annual Daily Traffic near the location
Grocery Stores	Integer	Count of nearby grocery stores
Food Outlets	Integer	Count of nearby food outlets
Tourist Destinations	Integer	Count of nearby tourist destinations
Hotels	Integer	Count of nearby hotels
Educational Institutions	Integer	Count of nearby educational facilities
Hospitals	Integer	Count of nearby hospitals
Airports	Integer	Count of nearby airports
Residential Housing Density	Float	Density of residential housing in the area
Commercial Housing Density	Float	Density of commercial housing in the area
Fuel Stations & Car Wash	Integer	Count of nearby fuel stations and car washes
Apparels	Integer	Count of nearby apparel stores
Height Restrictor	Boolean	Presence of height restrictors
Border Zone	Boolean	Indicates if the location is in a border zone
Public Transportation	Integer	Count of nearby public transportation options
radius_miles	Float	Radius in miles used for data collection

METHODOLOGY

The methodology employs a rigorous analytical framework to evaluate potential locations for retail expansion. It is structured into distinct phases: data collection, data preprocessing, exploratory data analysis (EDA), model building, and deployment, with a continuous feedback loop for improvement.

Data Collection

The first phase involves an extensive data collection process where information is systematically gathered via web scraping and APIs. This encompasses a detailed compilation of Retail chain and competitor store data across various states. Key attributes such as traffic volumes, proximity to highways, and local amenities are aggregated alongside geospatial data to create a robust dataset. This dataset includes separate data frames for the Retail chain stores, and competitors, and a master frame that integrates all collected data, providing a comprehensive view for subsequent analysis.

Data Preprocessing

Following collection, data preprocessing is undertaken to ensure data quality and usability. This stage addresses the cleaning of data, treatment of missing values, and removal of outliers. Highly correlated features are identified and eliminated to prevent multicollinearity, enhancing the predictive accuracy of the model. This process also involves the integration of data from various sources into a master data frame, ensuring that the data is consistent and reliable.

Exploratory Data Analysis (EDA)

EDA is conducted to uncover underlying patterns, trends, and correlations within the data. This analysis is critical in understanding the factors that impact the success of retail locations, aiding in the effective feature selection for the predictive model. The insights gained from EDA guide the strategic direction of the model development, focusing on variables that significantly affect customer visits.

Model Building

The model building phase is centered around the development of a predictive scoring model. This model employs advanced statistical and machine learning techniques to forecast the potential success of new retail locations. Techniques such as decision trees, random forests, and gradient boosting are utilized. The process involves feature engineering to enhance model inputs, selection of the most predictive features based on their collinearity, and hyperparameter tuning to optimize the model's performance.

Performance Metrics

Several statistical and business performance measures were considered to evaluate the model. These included RMSE (Root Mean Square Error), MAPE (Mean Absolute Percentage Error), and R^2 score. RMSE and MAPE were chosen because they provide clear metrics to quantify the error magnitude between the predicted and actual values, which is crucial for assessing the accuracy of predictions in a retail context. The R^2 score was used to measure the proportion of variance in the dependent variable that is predictable from the independent variables, offering insights into the effectiveness of the model in explaining the variations in store visits.

Continuous Improvement

Throughout the project, a feedback mechanism is incorporated to capture learnings and improve the methodology. This iterative process ensures that the model remains accurate and relevant, adjusting to new data and evolving market dynamics. We have implemented a multi-stage process tailored to the retail chain's strategic expansion needs. The initial phase involves comprehensive data collection via web scraping, utilizing APIs to extract pertinent location-based metrics such as traffic volumes, proximity to highways, and neighboring amenities. This is followed by rigorous data preprocessing to ensure data quality and structure.

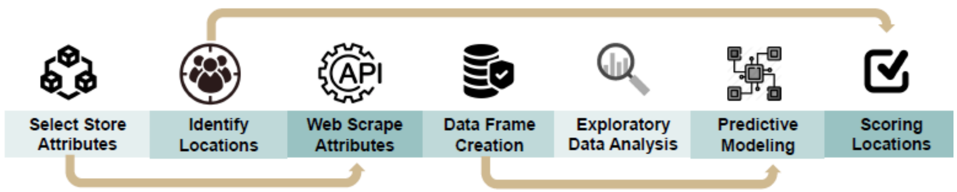

Figure 2: Analytical Process Framework

MODEL

The model-building process was methodically structured to ensure robustness and accuracy in predicting the success potential of Retail chain store locations. The process was divided into two principal approaches: developing a predictive model and creating a heuristic scoring model.

Predictive Model Development

The initial step involved one-hot encoding of categorical columns to transform qualitative data into a format suitable for modeling. Following this, the data was split into training and test sets using a 70-30 ratio, a common practice that balances the need for adequate training data while retaining enough data to validate the model's performance accurately. After splitting, the data underwent scaling to standardize the range of independent variables, which helps in enhancing the performance of many machine learning algorithms. To refine the model further, recursive feature elimination was employed to iteratively remove the least important features based on their performance impact, optimizing the feature set for better accuracy. Hyperparameter tuning was executed using GridSearchCV, which iteratively explored various combinations of hyperparameters to find the configuration that yielded the best results. This rigorous approach ensured that the model was not only tailored to the data at hand but also tuned to perform optimally across varying scenarios. The effectiveness of the predictive model was assessed through extensive cross-validation, specifically a six-fold cross-validation, to test the model's stability and reliability. This validation process confirmed the model's robustness, with a Mean Absolute Percentage Error (MAPE) of approximately 16.49%, closely aligning with the pre-validation estimate.

Heuristic Scoring Model

The second approach involved the development of a heuristic scoring model. This model was designed to categorize and weigh various store features such as traffic, proximity to essential amenities, and demographic factors. Each category was assigned a weight based on its perceived impact on store success, derived from correlation analysis and feature importance assessments from the predictive model. The weighting system was structured to emphasize traffic and proximity highly, reflecting their critical role in attracting customer visits. The final model produced a score on a scale from zero to five, where a higher score indicated a more favorable location for new store development. This score is dynamically represented on the dashboard, allowing stakeholders to adjust input parameters and immediately see the implications for potential store performance.

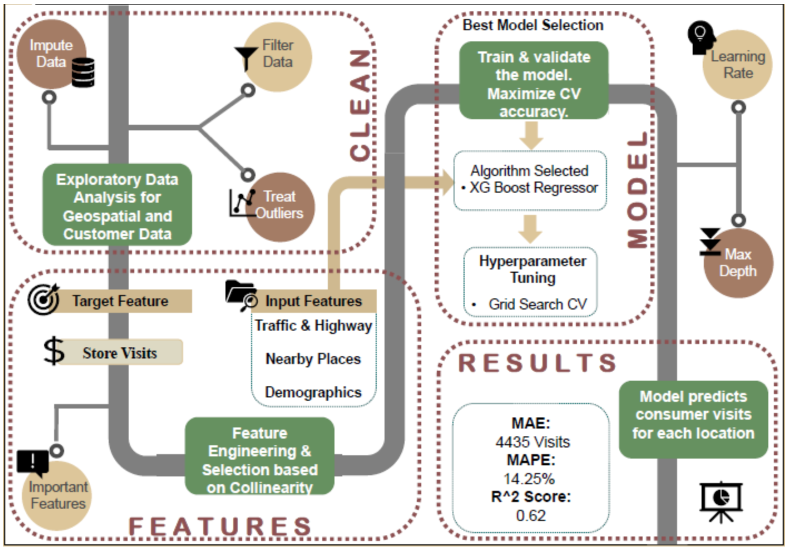

Figure 3: Model Building Flowchart

RESULTS

The evaluation of the predictive model developed in this study was comprehensive, encompassing both a traditional 70-30 train-test approach and extensive 6-fold cross-validation to ensure robustness and reliability.

- **Train-Test Evaluation**: The model achieved a Root Mean Square Error (RMSE) of 6335.45 and a Mean Absolute Error (MAE) of 4779.50, with a Mean Absolute Percentage Error (MAPE) of 15.80%. These metrics indicate the model's general accuracy in predicting store visits based on various location-based factors.

- **6-Fold Cross-Validation**: In this more robust testing scenario, the model showed a slight increase in error, with an RMSE of 6503.98 and MAE of 5062.84, and a MAPE of 16.49%. This slight increase in error metrics during cross-validation suggests some variability in the model's performance across different subsets of the data, highlighting areas for further optimization.

Table 3: Results for Retail Chain

	Method	**RMSE**	**MAE**	**MAPE**
XG Boost Regressor (With the best parameters)	**Train-Test (70-30)**	6335.45	4779.50	15.80%
	6-Fold Cross Validation	6503.98	5062.84	16.49%

Analysis Across Different Training Sizes

The model's performance was also evaluated across various training sizes, revealing that larger training sizes generally result in lower RMSE and MAE, indicating improved accuracy with more extensive training data. This trend highlights the benefits of leveraging large datasets in training predictive models for retail analytics.

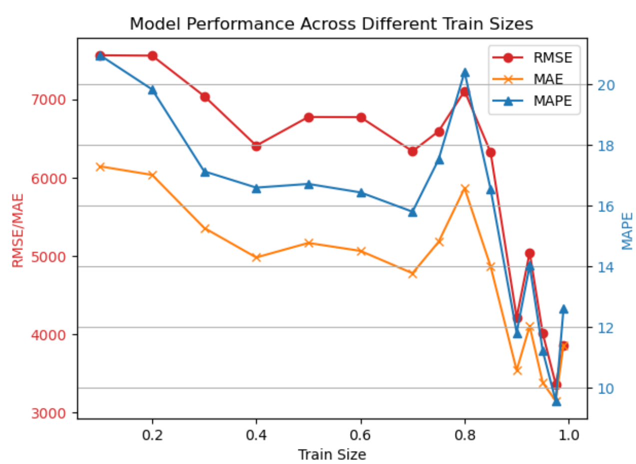

Figure 3: Model Performance Across Different Train Sizes

Metrics for Different Train-Test Splits:

	Train Size	RMSE	MAE	MAPE
0	0.100	7559.969853	6142.863310	20.969785
1	0.200	7555.250234	6034.142352	19.842451
2	0.300	7039.299653	5357.849130	17.138101
3	0.400	6407.489177	4980.781035	16.600626
4	0.500	6772.685188	5166.082340	16.722130
5	0.600	6770.718400	5064.775583	16.443013
6	0.700	6335.455550	4779.503227	15.805915
7	0.750	6589.533636	5178.919768	17.524008
8	0.800	7104.732405	5864.234375	20.417659
9	0.850	6321.450644	4867.311311	16.551389
10	0.900	4214.203519	3540.750244	11.810539
11	0.925	5044.753785	4105.700846	14.029566
12	0.950	4013.796400	3384.315430	11.215833
13	0.975	3363.471436	3150.906738	9.559551
14	0.990	3864.897397	3864.485352	12.606104

Figure 4: Metrics for Different Train-Test splits

Cross-Validation Performance Metrics

Further cross-validation with varying numbers of folds showed a general decrease in RMSE and MAE as the number of folds increased, affirming the model’s improved stability and performance with extensive validation. This suggests that the model's predictive accuracy is reliable even when subjected to different subsets of data.

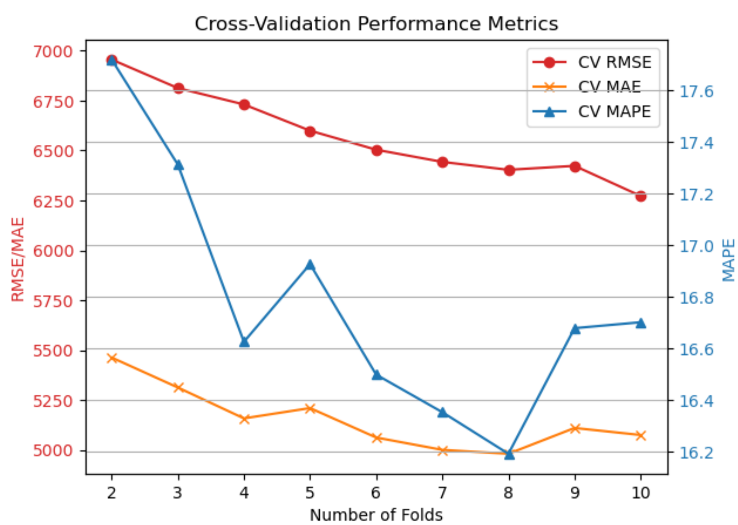

Figure 5: Cross-Validation Performance Metrics

From a business perspective, the XGBoost Regressor model equipped with the best parameters provides a robust tool for predicting store visits, crucial for strategic decision-making in retail location planning. The insights gained from the traffic and proximity analysis, in particular, suggest that locations with higher traffic volumes and optimal accessibility are likely to experience higher customer visits, thus validating the model's practical application in a real-world setting.

The model allows for a nuanced evaluation of potential retail locations, providing a composite score that integrates both quantitative traffic data and qualitative assessments of local amenities. This model can serve as a reliable decision-support tool, enabling the Retail Chain to allocate resources efficiently and target locations with the highest potential for success.

CONCLUSIONS

The retail industry faces the crucial challenge of optimizing site selection to maximize profitability and customer reach. Strategically, locating new stores is vital to ensure accessibility and attract a significant customer base, which directly influences the chain's success and expansion capabilities. The importance of this project lies in developing a data-driven approach to identify the most promising locations for new stores, thus supporting informed decision-making and enhancing the Retail chain's market competitiveness. This project sought to answer how data analytics could be leveraged to optimize retail store placement. Through the integration of Geographic Information, web scraping for real-time traffic data, and customer sentiment analysis, the study provided a comprehensive model that predicts optimal store locations. The results highlighted the potential for significantly improving store performance by choosing locations with high accessibility, favorable customer demographics, and positive sentiment.

ASSUMPTIONS AND LIMITATIONS

The study assumes that the availability and accuracy of data from web scraping and free APIs are reliable, which may not always be the case. Limitations include the exhaustion of free API trial searches, and preventing a comprehensive sentiment analysis. Additionally, unavailability of direct traffic data downloads for most states, which could affect the model's accuracy and applicability in some regions.

To enhance the robustness and effectiveness of the location selection model, further research is necessary in integrating comprehensive customer feedback, expanding the data det, and continuous model evaluation.

AI RESEARCH TOOL REFLECTION

ChatGPT: Provided quick summaries of research papers, related work and highlighted the major points. Helped in generating code according to the requirement and DALLE was effective in generating images for data attributes while sometimes it generated complicated and unrealistic diagrams.

ResearchRabbit and SciSpace: Helped in providing related research papers and scholars. Both the platforms provided personalized recommendations based on user interactions, making discovery tailored and efficient.

REFERENCES

[1] Forbes. (2023). "How Big Data is Shaping the Future of Retail Site Selection."

[2] Gartner. (2022). "The Role of Advanced Analytics in Retail Location Decisions."

[3] The Wall Street Journal. (2023). "Data-Driven Retail: The New Frontier in Store Location Strategy."

[4] William Applebaum(1966) Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales Methods for Determining Store Trade Areas, Market Penetration, and Potential Sales on JSTOR

[5] Quang Thai LE, Davar PISHVA(2015) Application of Web Scraping and Google API Service to Optimize Convenience Stores' Distribution Application of Web Scraping and Google API service to optimize convenience stores' distribution | IEEE Conference Publication | IEEE Xplore

[6] Chibuzor Udokwu(2020) Evaluating Technique for Capturing Customer Satisfaction Data in Retail Supply Chain (PDF) Evaluating Technique for Capturing Customer Satisfaction Data in Retail Supply Chain (researchgate.net)

[7] Omar I. Aboulola(2017) A Literature Review of Spatial Location Analysis for Retail Site Selection (PDF) A Literature Review of Spatial Location Analysis for Retail Site Selection http://aisel.aisnet.org/amcis2017/DataScience/Presentations/9 (researchgate.net)

[8] Linda Eglite and Ilze Birzniece(2022) Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (PDF) Retail Sales Forecasting Using Deep Learning: Systematic Literature Review (researchgate.net)

[9] Ahmad Murad Bin Mohamed Rohani and Fang-Fang Chua(2018) Location Analytics for Optimal Business Retail Site Selection Location Analytics for Optimal Business Retail Site Selection | SpringerLink

[10] Desmond Namangale,Grace Dunga(2022) Location Analytics for Retail Property Investment- A GIS Based Approach (Open Access) Location Analysis for Retail Property Investment -A GIS Based Approach: Case Study of Blantyre City (2022) | Desmond Namangale (typeset.io)

[11] Yuh-Min Chen,Tsung Yi Chen,Lyu Cian Chen(2020) On a method for location and mobility analytics using location-based services On a method for location and mobility analytics using location-based services: a case study of retail store recommendation (2020) | Yuh-Min Chen | 4 Citations (typeset.io)

[12] Joseph Aversa, Sean Doherty, Tony Hernandez(2018) Big Data Analytics: The New Boundaries of Retail Location Decision Making Big Data Analytics: The New Boundaries of Retail Location Decision Making: Papers in Applied Geography: Vol 4 , No 4 - Get Access (tandfonline.com)

[13] Hernandez, T., Bennison, D.: The art and science of retail location decisions The art and science of retail location decisions | Emerald Insight

[14] Gartner. (2022). "How Advanced Analytics Is Transforming Retail Site Selection."

[15] Forbes. (2023). "The Role of Big Data in Retail Expansion and Market Penetration."

[16] The Wall Street Journal. (2023). "Data-Driven Decisions: The Future of Retail Location Strategy."