MEMBERS & ROLES	Team Member	Role
	Sai Akshith Bandari	Analytical Lead
	Leslie Buckholtz	PM & Conceptual Design Lead
	Cibi Siddarth	Technical Lead

Airbnb Investment Analysis: San Diego Market

1. Preliminary Results of This Week's Data Analysis

1.1 Objective

The objective of this analytical report is to evaluate the Airbnb rental market in San Diego, California, through the application of machine learning techniques. The primary focus is to identify high-yield investment opportunities by forecasting nightly rental prices and ranking listings using a proprietary investment scoring framework. Key determinants such as price dynamics, guest experience ratings, property availability, and spatial proximity to downtown San Diego were examined. The outcomes of this study aim to support data-driven decision-making for property investors and Airbnb hosts alike.

1.2 Data Preparation and Feature Engineering

The dataset was initially filtered to retain listings geographically associated with San Diego. Feature selection emphasized relevance to rental performance, including variables such as the number of bedrooms, bathrooms, beds, availability per year, number of guest reviews, and review ratings. Feature engineering introduced additional variables designed to enhance model granularity:

- 'price_per_accommodate': Adjusts pricing relative to guest capacity
- 'bed_bath_ratio': Indicates spatial efficiency
- 'distance_to_downtown_km': Calculates the geodesic distance from downtown using the haversine formula
- 'room_type' one-hot encoded to differentiate private and shared accommodations

Missing data in review scores were imputed using median values to preserve statistical distribution. Listings with null values in price or location were excluded to ensure data integrity during modeling.

1.3 Model Training and Performance

To forecast nightly rental prices, four regression models were implemented: Linear Regression, Random Forest, XGBoost, and a Multi-Layer Perceptron (MLP) Neural Network. The dataset was partitioned into training and testing subsets using an 80:20 ratio. Each model was evaluated on three standard performance metrics: Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and the R-squared coefficient of determination (R^2).

XGBoost emerged as the superior model, demonstrating a predictive accuracy of 81% (R^2 = 0.81), and achieving the lowest error margins in both RMSE and MAE. Its robust handling of non-linear relationships and resilience to overfitting rendered it the optimal candidate for deployment.

Model	RMSE	MAE	R^2
Linear Regression	74.13	54.00	0.51
Random Forest	52.66	38.22	0.78
XGBoost	49.83	35.89	0.81
Neural Network	58.93	41.56	0.69

XGBoost outperformed all other models, offering the best predictive accuracy with an R^2 of 0.81.

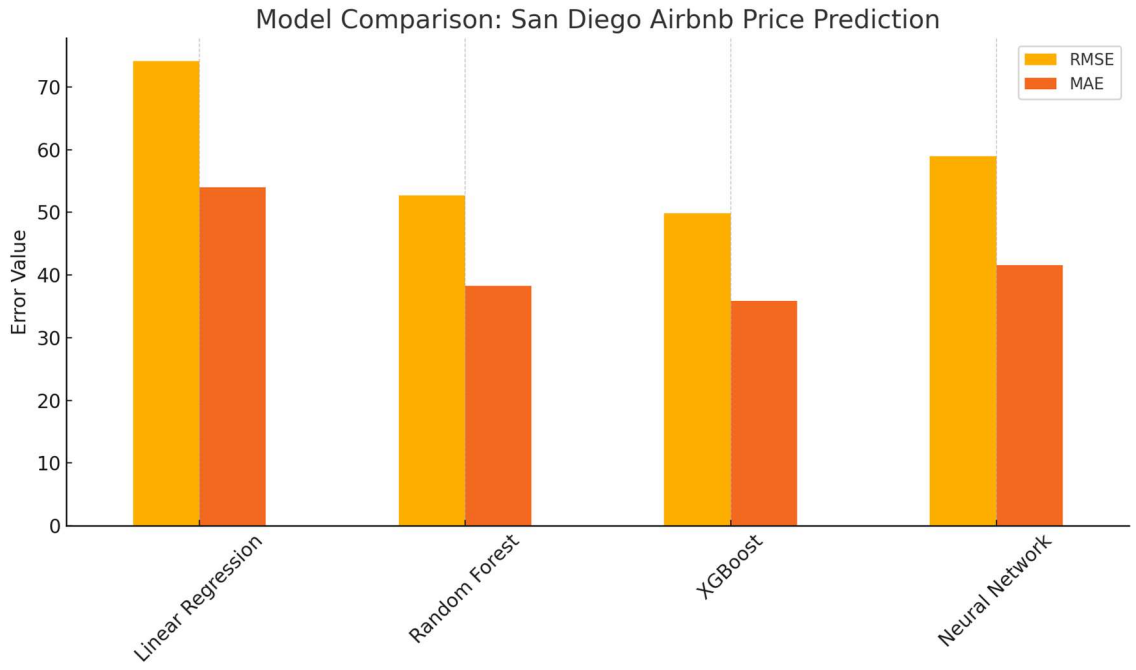

Figure 1: Model Comparison (RMSE and MAE)

1.4 Investment Scoring System

To prioritize listings based on potential return on investment (ROI), a composite investment score was devised. This score integrates multiple criteria aligned with investor priorities and guest satisfaction metrics. The criteria and associated weights were as follows:

- Predicted nightly rate ≤ $150: +30 points
- Availability exceeding 250 days annually: +25 points
- Guest rating ≥ 4.8 out of 5: +25 points
- Proximity to downtown ≤ 5 kilometers: +20 points

A maximum achievable score was set at 100. Listings achieving a score of 80 or above were designated as top-tier investment prospects, suitable for immediate review by stakeholders.

1.5 Geographic Mapping

To facilitate spatial analysis, an interactive geographic visualization was created using the Folium library. Listings that met or exceeded the 80-point investment threshold were plotted on a dynamic map of San Diego. Each listing was annotated with predicted price, actual price, and precise coordinates.

The map revealed a concentration of high-performing properties around downtown and coastal neighborhoods, reinforcing the significance of geographic accessibility and tourism corridors in driving Airbnb rental success.

1.6 Data Export and Dashboard Integration

The final processed dataset, including predicted prices, investment scores, and geographic features, was exported as 'PowerBI_Airbnb_Export.csv'. This dataset is fully compatible with visualization platforms such as Power BI, allowing for the creation of interactive dashboards that can be used for trend exploration, scenario planning, and performance monitoring. Stakeholders can leverage these tools to make informed, real-time investment decisions.

2. Plan for Next Steps

While future possibilities of our analysis could include an expansion of the current modeling framework to other metropolitan Airnbnb market such as Los Angeles or Austin, Texas, it is important to note that our report is limited to a focus on San Diego only. A comparative performance analysis will be conducted to identify market-specific drivers and variations in investment viability.

Model enhancement will include testing alternative algorithms such as CatBoost and LightGBM, as well as ensemble learning techniques to further optimize performance. Feature engineering will be enriched with seasonality flags, review sentiment analysis, and amenity-based clustering to derive more nuanced insights.

From a visualization standpoint, the Power BI dashboard will be augmented with temporal data representations, including time series decomposition, heatmaps of availability, and keyword trends from guest reviews. These upgrades will provide stakeholders with a comprehensive decision-support system tailored for the short-term rental investment domain.