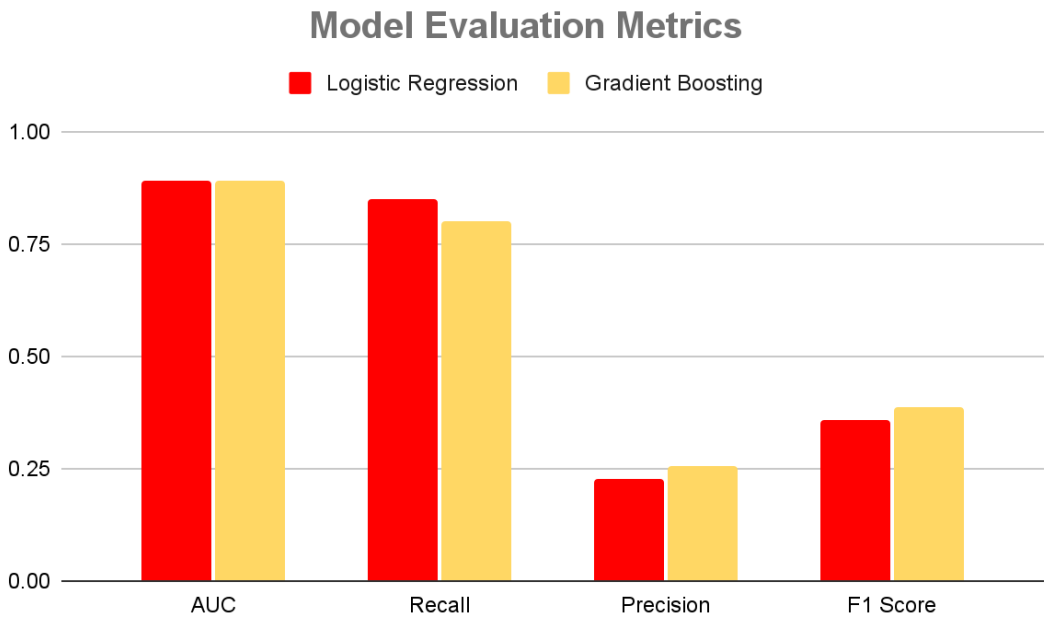
Executive Summary

This project focused on modeling and understanding customer behavior in the Moreno Valley, CA real estate market by simulating user engagement data and predicting the likelihood of users contacting an agent. We engineered a realistic dataset that mimics raw user browsing sessions across multiple devices, tracking actions such as saving, sharing, and scheduling viewings for listings.

To address the class imbalance in contact behavior, we applied SMOTE oversampling and evaluated several classification models, including Logistic Regression, Random Forest, and Gradient Boosting.

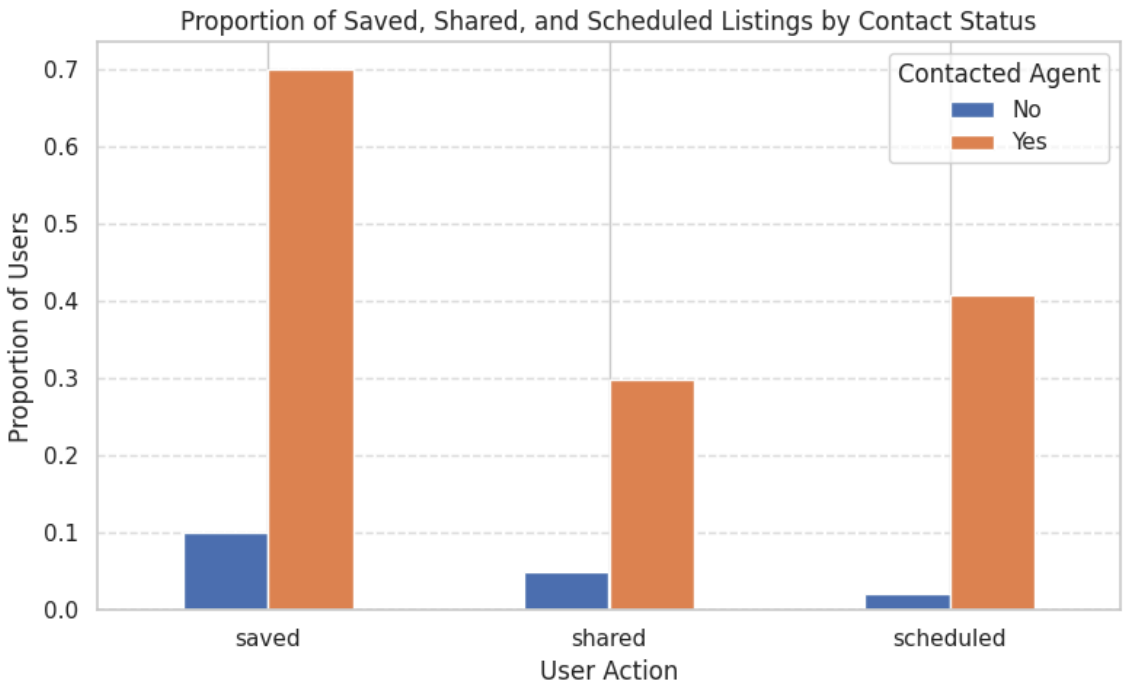
Key Results:

- **Logistic Regression** achieved the highest **recall (0.85)**, making it the best model for capturing most potential leads, despite a **low precision (0.23)**.
- **Gradient Boosting** delivered a slightly lower recall (0.80) and similar precision (0.26), but added complexity without significant performance gain.
- High **recall was prioritized** to ensure interested users are not missed, aligning with business goals of maximizing lead capture.



Behavioral Insights:

- Users who **saved listings** were **3x more likely** to contact an agent.
- **Sharing listings** and **scheduling showings** also had a strong positive relationship with contact rates.
- **Longer session lengths** slightly increased contact likelihood, but viewing **more listings** showed a marginal **negative effect**, suggesting diminishing returns.



Recommendations:

- Deploy the **Logistic Regression model** for lead prioritization, given its interpretability and performance.
- **Enhance site features** that encourage users to save and share listings to boost conversions.
- Investigate and refine the user experience for those who browse many listings without converting.
- Leverage engagement metrics to build **user segments for targeted follow-ups**.

- Monitor **city-level patterns** (e.g., Moreno Valley and Perris) to optimize regional marketing and agent allocation.

This approach enables us to translate user engagement data into actionable strategies, improving lead conversion and guiding smarter resource allocation.