



# QUANTIFYING THE ADVANTAGES OF ARTIFICIAL SWARM INTELLIGENCE WHEN BUILDING PREDICTIVE MODELS

Through a divide-and-conquer strategy, artificial swarm intelligence produces predictive models that surpass those created with traditional techniques.



## Summary

The US property market is highly diverse, with different market dynamics in different areas. A single model that predicts home prices for the entire country is not as accurate as an ensemble or swarming model that is tailored to specific areas. Swarmalytics has created an artificial swarm intelligence engine that breaks down the dataset into homogenous subsets and builds sub-models for each subset. To quantify the benefits of the swarming methodology, the swarming model was compared to a traditional single model, and it was found to be significantly more accurate, making the investment into splitting and recombining the dataset worthwhile.

## Introduction

With over 140 million housing units spanning 3 million square miles, the US property market is incredibly diverse. Each property also has a wide variety of characteristics that can be compared to other homes. Real estate market dynamics also vary greatly from one part of the country to the next. As a result of these differences, using a single model to predict home prices in both Malibu, California and Jackson, Mississippi reduces the accuracy of both predictions. Since one model only has a finite number of predictors, it is not possible to have enough variations to provide quality predictions for all markets.

Swarmalytics has created an artificial swarm intelligence engine to overcome this problem. By breaking the datasets into homogenous subsets, specialized models can be built for each subset in a way that provides greater accuracy. This allows for the sub-models to have better customized predictions for the subsets of data, leading to more accurate predictions for the data set as a whole.

However, this methodology comes with a cost. Using the swarming modeling technique increases the computation cost and effort by splitting the dataset, creating a model for each segment and then recombining the output model. In this whitepaper, we are going to compare the performance of a swarming model versus a traditional single model in order to provide context as to whether the increased accuracy that comes from a swarming model is worth the additional effort.

## Model Development

To compare the effectiveness of the swarming intelligence process, two different models were built. Both models were trained to rank residential properties in the United States according to their likelihood of being sold, renovated, and sold again for a profit in a short amount of time (this process is often called 'flipping'). The datasets used to train and validate the two models were the same, with the only difference being that one model employed the swarming approach while the other used a single model for all predictions.

Starting with a database containing every US residential property, their characteristics, and their transaction history, Swarmalytics then appended a wide variety of public and proprietary data sources to develop predictive models. Public data sources included listing data from multiple listing services (MLS), financial and economic data from the Internal Revenue Service and the Securities and Exchange Commission, and geographic proximity data from various mapping sources. Swarmalytics' proprietary data included behavioral data that is highly predictive of an incoming property sale. Together, more than 3,300 candidate variables entered the modeling process.

After all additional features were added to the dataset, each of the two models was built using a different approach. One approach used data from all properties in the US to build a singular model. The

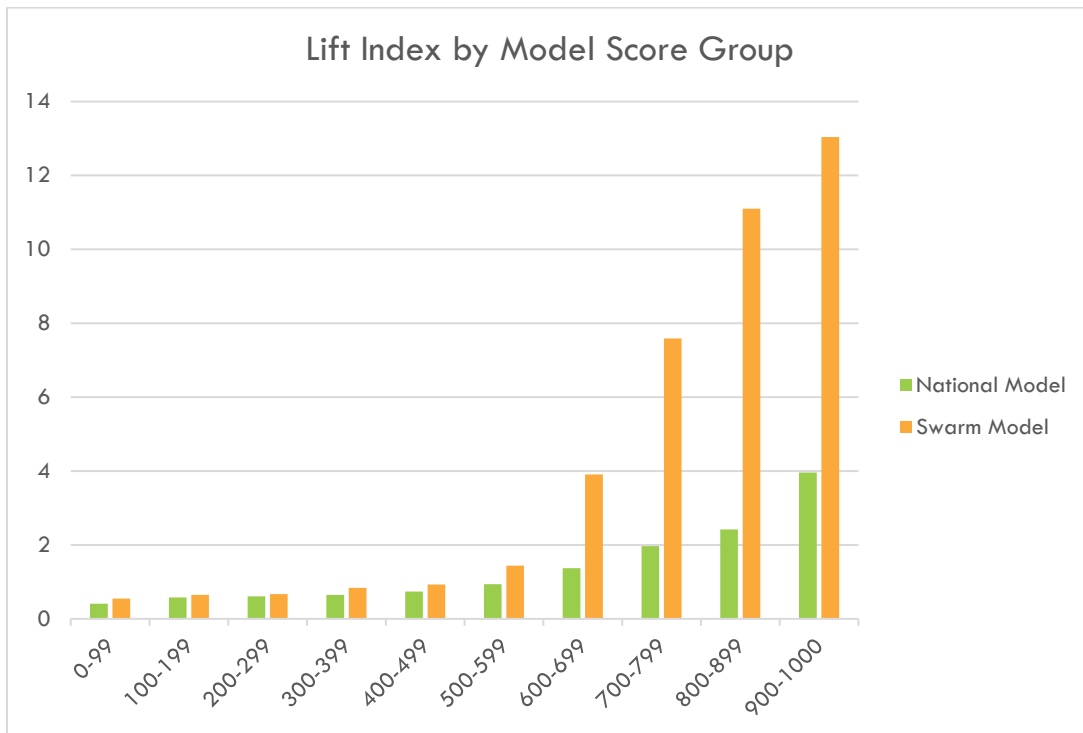


swarming approach separated the dataset into several homogeneous subsets and built a model for each subset. After these sub-models were built, they were combined to create a comprehensive model.

Apart from the difference in swarming versus traditional model training methodologies, the models were constructed in identical fashion in order to provide a reliable comparison. Using machine learning, each model created 1.2 million candidate equations which competed against each other based on predictive power. Once a final equation was selected by each of the two models, a validation dataset was scored by both equations.

## Model Performance Comparison

Properties in the validation dataset were assigned values from zero to 1,000 by each model indicating how likely they were to be flipped. By splitting the validation set into ten groups, calculating the rate of flips in that group, and dividing that rate by the rate of flips in the validation set as a whole, Swarmalytics was able to calculate a lift index for each group. These lift indices allow us to see how much more accurately each model can predict whether a property will be flipped compared to a process of random selection.



The swarming model significantly outperformed the single national model. Properties in the top score decile as determined by the swarming model were 13 times as likely as average to flip compared to only four times as likely according to the national model. The performance of groups assigned score values between 600 and 899 by the swarming model also greatly outperformed those same groups when assigned by the national model. The swarming model can better identify which property is likely to be flipped and assign a higher score accordingly.



## Conclusion

The US property market's complexity and diversity pose challenges for accurately predicting home prices using a single model. In this white paper, we introduced Swarmalytics' artificial swarm intelligence engine, which addresses these challenges by employing a swarming methodology. By breaking down the dataset into homogenous subsets and building specialized models for each subset, the swarming model outperforms the traditional single national model in predicting the likelihood of property flipping.

Our analysis compared the performance of the swarming model with the single model approach, using a dataset encompassing millions of residential properties across the United States. The swarming model demonstrated remarkable accuracy in identifying properties with a higher probability of being flipped for profit. Properties ranked in the top decile by the swarming model were found to be 13 times more likely to be flipped compared to the average rate, surpassing the performance of the national model by a significant margin.

While the swarming methodology entails additional computational costs and effort due to the splitting and recombination of datasets, our findings indicate that the increased accuracy justifies these investments. By tailoring models to specific subsets of data, the swarming model leverages the diverse market dynamics across regions and property characteristics, resulting in more precise predictions that can prove to be invaluable to actors at all levels of the real estate market.

As the property market continues to evolve, embracing innovative methodologies like swarm intelligence can unlock new possibilities for understanding and predicting market trends. Further research and development in this field hold the potential to enhance decision-making processes, optimize investment strategies, and maximize returns in the ever-changing landscape of the US property market.

## About Swarmalytics

Swarmalytics was founded in 2021 by CEO Doug Newell and CTO Dan Koehler with the goal of using artificial swarm intelligence to power insights and create value. To learn more about Swarmalytics, please visit [www.swarmalytics.com](http://www.swarmalytics.com).

