REM Part 2: Building a Dataset To Capture Real Estate Value Creation



May 24, 2025

Charles Calomiris, Chief Economist

This is the second in our series of blog posts reviewing some of the technical aspects of our Real Estate Modeling (REM) framework. Our goal in sharing information about our REM framework with prospective

clients and investors in these blog posts is to allow them to see a bit inside th
modeling and share our enthusiasm and confidence in the REM framework w

In my first blog post, I summarized some of our basic discoveries about the effects of green certification on office building lease revenue, highlighting four discoveries from our Real Estate Modeling (REM) framework. In the next blog post, I will discuss how we use artificial intelligence (AI) and natural language processing (NPL) to enhance the reach of those data to provide "soft" data concepts that augment more typical "hard" data measures, and why and how we make use of cutting-edge statistical validation techniques to guide our variable selection process in building the models.

In this blog post, I will explain how we constructed our proprietary database, why and how we allow thousands of building attributes and locational attributes to interact with one another when identifying the value consequences of certification or amenity choices.

When I began doing statistical analysis as a student and young professor, the amount of information one could process was severely limited by computational and storage capabilities. In the current era of Big Data, it is now commonplace to describe one's data resources in terabytes. At Mina Analytics, our goal was to be able to provide a rich analysis of the value contributions of each attribute of each building in each location we are studying. Only by doing so, can we isolate the value contribution of the characteristic in question.

For example, suppose a client wants to know how much the rent per square foot of an office building is affected by having a beautiful rooftop shared space for tenants in the building to use. Or the client might want to know how much revenue is affected by having a parking garage, or by giving tenants access to an exercise facility. Knowing those answers can help a developer decide whether to change some feature of the building, and can help a prospective buyer determine the value of the property.

Note that each of these amenities decisions may have different value-creating
of the other aspects of the building under consideration. Furthermore, the va
amenity decision may differ across locations. The potentially relevant building
variables and their interactions that one should consider may number in the tens of thousands.

And it is important not to constrain a valuation model by excluding variables from consideration that might be important. When trying to determine the value creation from adding an amenity it is just as important to collect accurate data about control variable characteristics, as it is about the particular amenity characteristic that the client is considering. For example, if a client wants to know whether a green certification would add substantially to the value of a building, one must be sure to control for all the variables that might be correlated with a green certification; if one fails to do so, one may over-estimate the effect of green certification.

To see why, note that recently constructed buildings are often built to be green-certified by the outset. In any dataset green certification and building age are highly negatively correlated. If one does not control properly for age of the building, one may therefore exaggerate the influence of green certification. In fact, previous careless statements by people in the real estate industry have sometimes exaggerated the green-certification premium because they made precisely that mistake. When constructing a model to isolate the effect of certification or amenity characteristics, one should collect as many characteristics of buildings and their locations as possible.

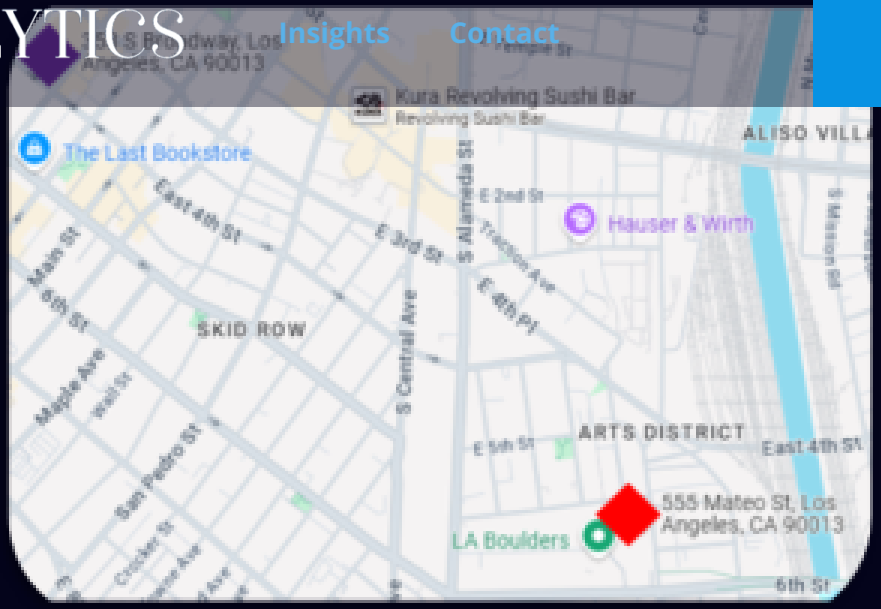
But, you might wonder, how can any model incorporate thousands of explanatory variables? The answer to that question is, by adopting a forward-selection methodology. Forward selection allows one to begin with thousands, or even tens of thousands, of possible explanatory variables, but only choose a small number of variables from that long list based on which variables work well to capture all the influences on a buildings value.

It can be very important to make sure to include all the relevant building char
controls (which may reduce the effect of the characteristic being analyzed) ar
treatment variable (that is, factors that make the treatment effect smaller or l

consequential the inclusion of building characteristics can be, consider Figure 1 below. The two buildings shown in this figure are located in the same zip code. All the location-specific characteristics that influence the estimated magnitude of the green-certification premium, therefore, are the same for these two buildings. And yet, the estimated value consequence of a LEED certification for the building on the left is to raise the revenue per square foot by a huge 13.1%, while for the building on the left, the LEED value uplift is smaller at 7%. The difference is attributable to the different building characteristics that the model finds are important interactions with the treatment effect of a LEED certification.

Of course, there is such a thing as being too focused on easily measured “hard” variables at the expense of harder-to-measure “soft” information. And there is also a risk adding too many explanatory variables to a model (the problem of so-called “over-fitting”). Next time, in our third and final blog post, we complete our discussion of the REM framework by explaining our use of AI and NPL to expand the range of variables the REM framework considers (i.e., specifically, to allow it to capture “soft” information). We will also discuss the importance of model validation, which ensures the robustness of model forecasts, and minimizes the risk of over-fitting.

Figure 1: LEED Premium for Two Buildings in Downtown LA

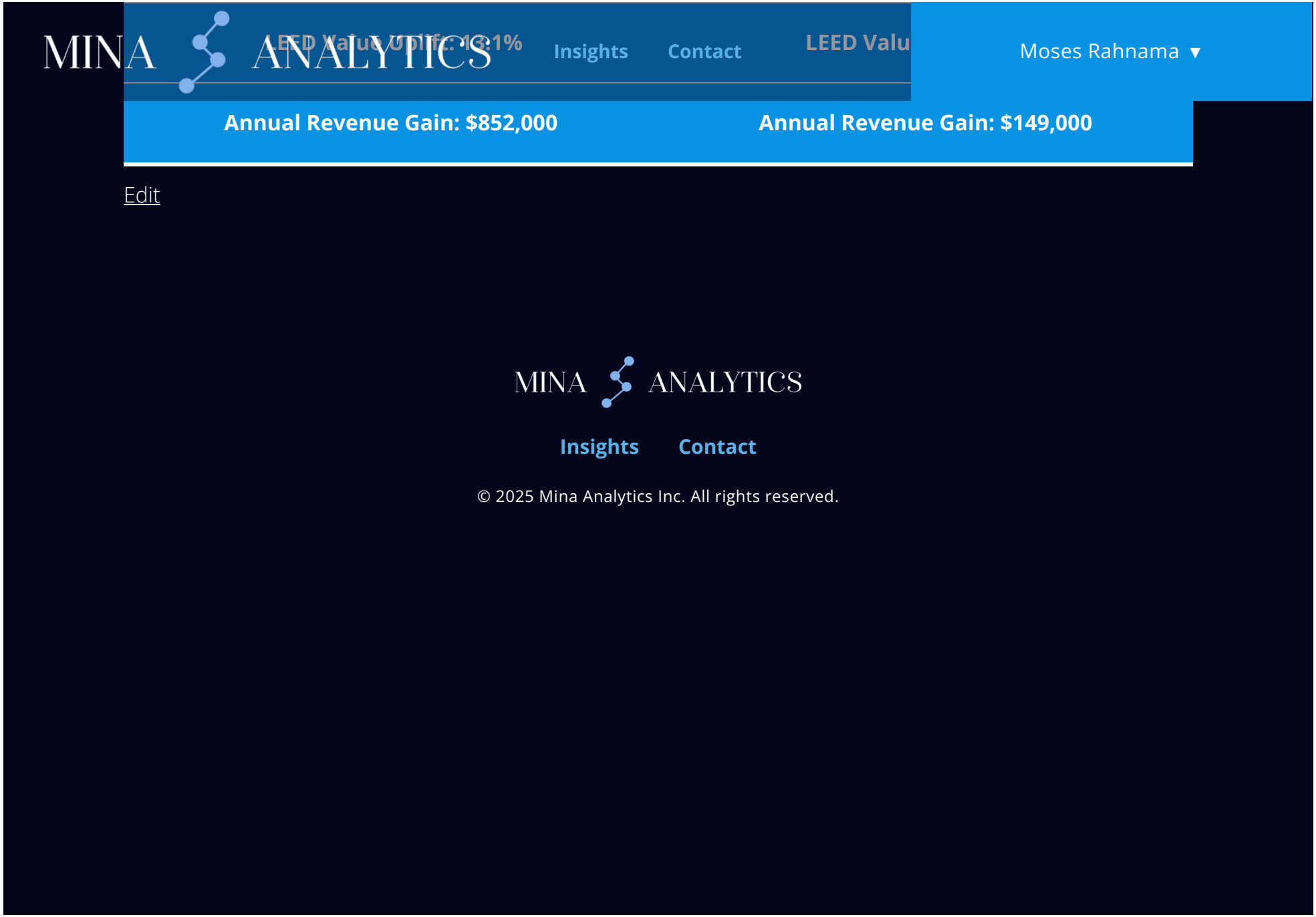


353 S Broadway, LA

555-581 Mateo St, LA






MINA ANALYTICS
LEED Value Uplift: 13.1%
Insights
Contact
LEED Valu
Moses Rahnama ▾
Annual Revenue Gain: $852,000
Annual Revenue Gain: $149,000
Edit
MINA ANALYTICS
Insights
Contact
© 2025 Mina Analytics Inc. All rights reserved.