CBRE Datathon – Image Detection

Franz Limoges, Dennis Schoenmaker and Kasia Dziewulska

Background:

Supply is one of the most poorly understood aspects of tracking the US industrial real estate market. Often data are not well marked and addresses can be a little inaccurate so CBRE is challenging you to flex your skills and come up with a way to better model the construction process.

Typically an industrial warehouse can be built in anywhere from 2 quarters (6 months) to 6 quarters (18 months) and that range can depend on size is square feet, ceiling height of the building in feet, the loan process, available construction labor, materials costs and availability and weather conditions to name just a few factors. We are going to keep this model simple and focus solely on size within the warehouse subproperty type of Industrial. The model you will apply is laid out in this table:

SubProperty Type	Size	Quarters to Complete
Warehouse	< 99,999	2
Warehouse	100,000 to 299,999	3
Warehouse	300,000 to 599,999	4
Warehouse	600,000 to 999,999	5
Warehouse	1,000,000 <	6

In this document, we have supplied some construction stages and images that we believe capture a particular stage of construction. You have been provided with data from 3 cities in the US to use as training/test data complete with latitude or longitude. Your focus will be Atlanta and our end goal is to create a file that can be used to forecast the supply (measured in Sq Ft) which is coming into the market (the construction process). You may choose to use a Convolutional Neural Network but other options available to you might be using a desktop application like QGis with Orfeo Toolbox to do supervised classification. The model you will build will then be used with an input mask to perform image detection.

Description:

To build a model for estimating time of construction for Industrial Warehouse in the Atlanta Georgia market. Requirements include:

1. Import data from the table given in a CSV format.
2. Acquire recent aerial/satellite imagery. (Google, Bing). I checked recently and Bing appears to have more up-to-date imagery than Google in Atlanta.
3. Using a convolutional neural network /QGis with plug-ins, attempt to train a model to classify the current point of construction on the points in the CSV file and the satellite imagery. The US Census block layer in GeoJson format could be used to mask and cut way down on the areas that need to be searched.

4. Classify the status of the buildings for which you are searching as one of the 5 categories listed in the stages of construction section of this document.
5. Using the table in the background section, figure out how long the buildings should take to complete. This will likely vary from the EstimatedYearQuarter field in the CSV containing construction data. In a later step you will need to aggregate all buildings with delivery dates from your work to come up with the sq footage by year and quarter in which your model expects the building to be completed.
6. Given the structure of an estimator (Atlanta Dat file), take the data aggregates you added in the last step and add them to the Data file. You may need to create rows where the year quarter is in the format yyyy.Q. Add your aggregate to the completions_sf column.
7. The stock_SF column becomes the previous YearQuarter stock_sf figure plus the current Completions and you will work stock forward across the added quarters.
8. The "stock_under_construction_sf can be left blank and so can the Under_construction_as_a_percentage_of_stock fields.
9. Make sure, if possible, to record the approximate date of the satellite imagery because adjustments might have to be made to account for the age of the image versus, the stage attained in the image and the current date.
10. Bonus consideration: are there structures not in the data file but are on a building parcel zoned as industrial that the model picked up off the imagery but are in the stages ground broken, concrete pad or framing going up?

Data Dictionary:

CSV of buildings under construction:

Column Name	Column Definition
PropertyID	The Unique Identifier for the Building
VendorID	The Vendor from which Data were Derived
Address	The Current Street Address of the Building
Size_sf	The Size in Square Feet of the Structure
Available_sf	The amount of space available for rent in the building
QuartersTrackedInDatabase	For howlong has the data appeared in our database
EstimatedCompletionYearQuarter	the Year and Quarter the vendor believes the building will come online
YearQuarterGroundBroken	The Year and quarter the building broke ground and construction began
BuildingStatusCode	The current status of the building
SubPropertyType	What type of industrial building is this? Logistics, R&D, manufacturing or other
buildingName	The name of the project while under construction
complex	The name of any building park to which this structure belongs
Latitude	The latitude at which the building resides
Longitude	The longitude at which the building resides
CeilingHeight	The distance in feet between the floor and the ceiling joists

CSV of the market history:

Column Name	Column Definition
MarketCode	Code delineating the market
MarketName	name of the market
YearQuarter	Year and quarter of history
Stock_sf	The total size of the market in square feet
Completions_sf	the size of all the square footage of building that were completed in this quarter
Stock_Under_Construction_sf	the total future size of buildings that have broken ground
under_construction_as_a_percentage_of_Stock	the percentage of the total future size of buildings that have broken ground / stock_sf

<u>Stages of Construction:</u>

Undeveloped land:

- Free of structures
- No sign of land having been cleared or bulldozed
- Vegetation or sand might appear to be unmoved
- Reached in Quarter 0 – 0% of the way through the process

Ground Broken:

- Free of structures
- Discernable movement of land with possible temporary roads
- May be construction vehicles present
- Vegetation cleared
- Reached 1-25% of the way through the process

Concrete Pad:

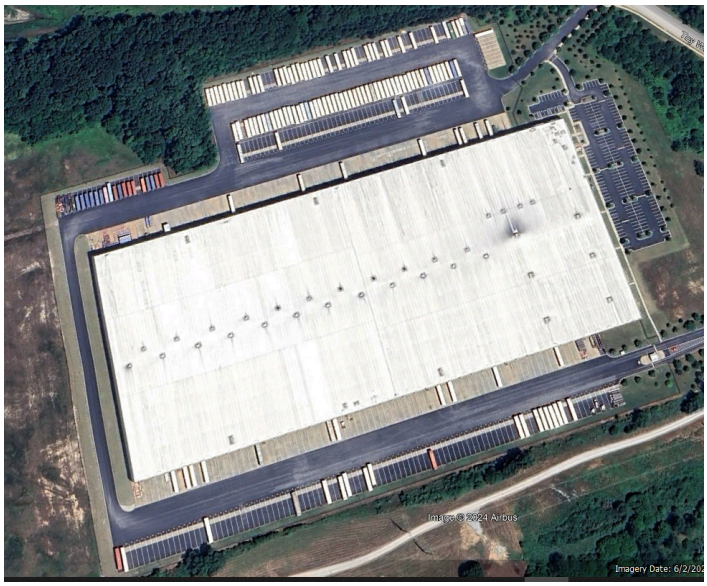
- Land contains a large flat surface of poured concrete
- The land around that pad is still unfinished surface usually dirt and free of vegetation
- May be surrounded by construction vehicles
- Reached 25-45% of the way through the process

Framing Going up:

- Contains many of the same elements of concrete pad
- Instead of a large flat surface you begin to see more variation in pixel color as the complexity of the structure begins to take shape
- Walls may be casting shadows depending on the time of day of the satellite photo.
- There may be a roof present but there is no paved parking lot present.
- Reached between 45-85% of the way through the process

Near completion or completed:

- A polished appearance.
- Structure has a roof (often white but not always
- Freshly paved parking lot. May have vehicles in the lot or trailers (containers) backed up along at least one side of the building.
- May have attractive plantings to demarcate portions of the parking lot.
- Reached 85-100% of the way through the process



Judging Criteria:

Business Efficacy:

1. How will your solution help the business by solving this need?
2. How will the work help the business get a handle on real estate supply?
3. How can it be used by brokerage?
4. How effectively can it be used to model?

Accuracy:

5. Consistency with past completion levels.
6. How accurately did the model classify buildings?
7. Bonus: can the model pick up buildings in the satellite image that appear in one of the stages: ground broken, concrete pad or framing going up but wasn't part of the set provided in the CSV file?

Predictability:

8. Using labeled data, were you able to combine that with the provided market history to create a model of construction out over the next 6 quarters (180 days)?
9. Can the model be used on another market with similar results? (second market provided)
10. How well can you explain any issues that arose due to discrepancies between performance in these markets.

Storytelling:

11. What was the story you were attempting to investigate?
12. Can you explain how your model works clearly and concisely?
13. In what ways do you believe your model is actionable?