Using AI-Processed News Datasets to Perform Predictive Analytics

Monday Morning Quarterbacking the Amazon HQ2 Selection

BITVORE

By Greg Bolcer
Chief Data Officer, Bitvore

Abstract

Bitvore has built a platform that continuously ingests massive amounts of unstructured data from world news and other sources and applies proprietary AI techniques to clean, normalize and organize the data. Key to Bitvore's approach is sifting through the massive amounts of noise and identifying only material business events (we call them ***business signals***) and associating them with corporate entities and municipalities. Business signals represent events that affect the bottom line of a company (e.g., legal issues, mergers and acquisitions, executive changes, labor and facilities expansion/contraction and many more). Our customers access, quantify and incorporate the resulting dataset into predictive models to support their business goals.

We are often asked for examples of how our signal-driven, precision news data can be predictive when we mention how our customers use our data. To that end, we decided to set up our own predictive experiment through backtesting. We decided to focus on one of the biggest stories of the past two years:

- September 2017 Amazon expansion announcement
- January 2018 Amazon final cities announcement
- November 2018 final selection of 2nd headquarters.

These announcements have been dubbed in the news collectively as the Amazon HQ2 Selection competition.

Our Focus

- Determine if, at each stage of the competition, we could predict what cities would advance to the next round
- Determine if we could predict the final location(s) of Amazon's HQ2 before it became public information

Predictive Experiment

There's an old saying that hindsight is 20/20. One of data scientists' favorite things to do is backtesting. Backtesting is the process of taking a predictive model and running it on historical data to help improve the results. It's a common technique used by investors, hedge funds, and financial trading firms to simulate a real system, as if it were active and online at a previous date. It's also a good way to analyze if a complex strategy, prediction, or algorithm is sound without risking any real money.

While this technique is fairly common for quantitative data traders (aka 'quants') it is seldom used for natural language or for news. Most historical data used for backtesting is comprised of spot market prices for various exchanges and commodities. Real world events outside of the numeric data are either discounted as a one-off situation or ignored as unpredictable. Another issue is that the data scientist will 'overfit' their model to the historical data, which simply means an exaggerated adherence to a small number of data points. We wanted to backtest our data against the Amazon HQ2 decisions to see if we could make an educated guess on what they were planning. Thus, the predictive experiment using all streams of Bitvore's data was born.

To support this experiment, we gathered all the Bitvore municipal, economic, corporate news, and geographic data into a common model. The HQ2 story was fascinating from the Bitvore perspective, as it required historical data across our products. Amazon laid out the requirements for what they were looking for in a second headquarters, and these requirements meshed well for things Bitvore already tracks. We had years of data available to cross-check against their announced criteria. HQ2-watching involved a large public company, emerging growth companies, and high-tech infrastructure.

It is also a competition between various city, county and state governments, all trying to prove that their education, mass transit/transportation, population centers, health centers, utilities, commerce, tax policies, and economic incentives would make them an attractive site. Besides our years of municipal, economic and corporate news, we also had public-private partnership data (P3), information on industrial development authorities (IDAs), and insights into local government fiscal health information, which could indicate how eager a municipality would be to work with Amazon. Finally, as part of our Bitvore for Munis offering, we have analyzed news data for commercial lending and commercial-backed mortgages that are good sources of early market insights for commercial real estate development in specific cities, counties, or regions.

The Plan

- Calibrate the experiment by evaluating U.S. and Canadian states involved in the request for proposals (Amazon HQ2 RFP) process and validate the prediction against states encompassing the top 20 finalists.
- Predict the Top 20 cities based on our ***geographic scoring*** and compare that to Amazon's shortlist of 20 finalist cities.
- Use the Amazon Top 20 finalists to build scoring from our ***municipal, corporate, and economic health data*** as related to the RFP's criteria.
- Score the top locations using all this data and compare them against Amazon's final selection(s).

Modeling

While most predictive models require multiple iterations based on the outcome, we had the problem that there was really only one long iteration of decision making by Amazon to figure out what location or locations they were going to locate a new headquarters. For our predictive modeling, we used the data to figure out if we would have been able to predict the locations beforehand and with incomplete information. Predictive models in general use data mining and probability to forecast a result. The inputs to the model are based on variables that are likely to influence the results. Luckily within Bitvore, we had a critical mass of data sets that mapped to Amazon's stated requirements.

Given the limitations of repeatability and iterations, we had to work with what data we had. Predictive models can be as simple as a linear equation or as complex as a multi-layered neural network architecture. There are basically two types of predictive models, classification models and regression models. Classification models predict whether some item or record is part of an indicative class. Think of a spam classification for email. It's a binary decision whether or not an email sent to you should be marked as spam or not, but the inputs to the predictive model can be complex and numerous. The other type of predictive model is a regression model. Regression models take all sorts of inputs, and instead of predicting a 'yes or no' classification, they predict a numeric variable (e.g., How much money will I make next year? How many days will it rain? How many angels can dance on the head of a pin?).

Of course, any model is only as good as the question it is trying to answer. Our model was tailored specifically to answering the Amazon HQ2 Selection question. We used decision trees to capture and eliminate various candidate geographies, linear and logistic regression to score economic and geographic frequency, clustering and outlier algorithms to help align stated company criteria when compared with actual status of prospective sites, and time series algorithms to monitor momentum, trends, and key patterns in the data. Different techniques were used in different parts of the process to either produce scoring, momentum, sentiment, or a final prediction.

Because we didn't have enough repeat occurrences to train a model, we depended on some intuitive, visual/graphical techniques also to help properly frame the data and garner insights. Histograms are useful ways to make sense of large data sets by spotting how similar things happen with similar frequency. We were able to use histograms to understand how sentiment, economic health, and frequency of Bitvore business signals contributed to individual scores. We also used calculated trend lines. While we didn't have enough data points or data points over time to draw perfectly separated data, we were able to find clusters of points and outliers. Finally, for a lot of the scoring, we used a standard deviation method to help derive more accurate scores according to Amazon's criteria.

Amazon's HQ2 Timeline of Events

September 2017

On September 7th, 2017 Amazon announces they are planning on investing up to $5 billion, employing 50,000 workers, and filling up to 8 million square feet in a new location that would be an equal to their current headquarters. Amazon releases a request for proposal (RFP) outlining their criteria.

October 2017

The city of Seattle, facing budget issues and trying to find a way to address the homeless/affordable housing issues in the city, decides to propose a large-company "head tax." Council members propose a tax of anywhere from $100 to $275 per employee/year for high grossing companies within the city. While the first revision of the tax fails to garner enough support, it continues as a mainstay on the city council's agenda going forward, sharpening Amazon's resolve. Amazon receives over 238 proposals from various cities. The RFP deadline ends on October 19th.

January 2018

On January 18th, Amazon announces a shortlist of 20 finalist cities. This sets off a firestorm of speculation that would last the better part of 2018. Everyone from real estate and housing experts, economists, bankers, journalists, and of course, high tech armchair quarterbacks, have their say in the public square.

February-April 2018

Amazon signs NDAs with city governments. They also visit each prospective location, find specific sites, negotiate agreements on land, formalize procurement processes, and finalize incentive packages.

May 2018

Amazon announces they are halting construction of a new high-rise building near its headquarters while awaiting the outcome of Seattle's city council decision on the head tax. A day later, Amazon announces they are expanding their Boston office by 2,000 jobs and Vancouver by 3,000.

September 2018

Jeff Bezos, CEO, Amazon, publicly announces Amazon would make known their decision before the end of the year.

November 2018

Early in November, Amazon states they are close to announcing their final decision. They follow up a few days later stating that they will be adding two new headquarters instead of just one.

A November 13th, 2018 announcement follows, stating the Long Island City neighborhood in Queens, New York, and the "National Landing" area in Northern Virginia that includes Arlington, Crystal City, and Pentagon City, are the selected locations for their new headquarters.

New York agrees to give Amazon tax breaks of at least $1.525 billion, cash grants of $325 million, and other incentives. Virginia will provide $573 million in tax breaks, $23 million in cash, and other incentives. (We now know that Amazon cancelled the selection of the Long Island location in February, 2019 after political opposition. More about that and its possible predictability later.)

As a consolation prize, Nashville is chosen as an expansion site for an Amazon operations center for customer fulfillment, customer service, transportation, and supply chain support.

Experimental Design

In order to design the experiment, we split the data and scoring into several phases based on the timelines. We wanted to be sure that we only used data that would have been available at the time before each Amazon announcement.

- **Phase 1 data:** Amazon's September 7th, 2017 announcement until the October 19th, 2017 submission deadline.
- **Phase 2 data:** October 20th, 2017 until January 18th, 2018 announcement of the 20 shortlist finalists.
- **Phase 3 data:** January 19th, 2018 until the November 13th, 2018 headquarters announcement

Geographic Scoring

We generated a Geographic score based on frequency of precision news articles mentioning HQ2. In Bitvore, our Geographic labeling (aka 'tagging') is done through a rigorous process of scoring, cross-checking, relevancy, and disambiguation. To get a valid location on a news record requires multiple points of validation, such that we get valid geographic tagging, while avoiding issues like casual mentions or non-targeted news items.

First, we looked at state-level data, collating the list of states ranked based on the scoring that will be compared to the states represented in the 238 submitted proposals for Phase 1. It's interesting to not only look at the relative strength of each state but figure out which states were not included.

Second, we looked at the scored city and county data over the course of Phase 1 and Phase 2, to figure out who the top 20 shortlist finalists will be as compared to those picked.

Third, we used the Phase 3 data to score city and county data to predict the top 3 locations.

Municipal and Economic Health Scoring

- Sector-Sentiment weighted ranking
- Material Event-Sentiment ranking

Bitvore has over 223 material events that it uses to evaluate various municipalities separate from corporate or general economic news. Likewise, Bitvore has eight different sectors for categorizing news about municipal information, as well as several subsectors.

For instance, **Education** is a sector that is identified through analysis of the content we collect. It has several subsectors that correspond to **Higher Education** (universities and colleges) and **Primary/Secondary Education** (high schools, charter schools, trade schools, and elementary schools/ school districts). One of Amazon's stated requirements included "a highly-educated labor pool" and a "strong university system." Bitvore tracks the health of higher education as tied to specific geographies. Union actions, unfunded pensions, liabilities, legal risks/lawsuits, construction projects, ability to pay, admissions, closings, deficits, emergency actions, and dozens more material situations contribute to our assessments.

Amazon likewise is looking for a "business-friendly environment and tax structure that can provide tax breaks and exemptions, fee reductions, workforce grants, utility incentives, and other incentives". These requirements map nicely to our Local Government sector and material events.

"A high quality of life for the 50,000 employees to be located there" meshes well with tracking special taxes, bonds, financing of parks, recreation, museums, sports venues, and art under our municipal data and economic information, combined with crime, traffic, pollution, and other collective measures.

Amazon's requirement for "the ability to travel to/from an airport within 45 minutes" and "mass transit options located on-site" also map well to our Transportation sector and material events.

"Fiber connectivity and acceptable cell phone service" is discoverable in our corporate data through products/markets launches and the high-tech sectors.

For each geography, we put together a score based on the combination of the city/count, the municipal data related to Amazon's criteria, and the economic data from across our product lines.

In the full list of Amazon requirements, Amazon discusses what the drivers are for their decision-making process. With this criterion in mind, we started analyzing the data.

Calibrating States

Taking all Amazon HQ2 related stories from 9/1/2017 through 1/1/2018, which was two weeks before the finalist announcements, we scored them according to geographies. Because we use precision intelligence and not phrase matching, a calculated geography in this Bitvore experiment is a metric which represent a high relevance to the topic (in this case 'HQ2'). Bitvore showed news items from 68 states or territories from both the U.S. and Canada during that time period, compared to the 54 listed on Wikipedia. In Bitvore, we treat Washington D.C. as a state, not a city, equivalent to other U.S. territories like Guam, Puerto Rico, and U.S. Virgin Islands.

Another humorous note is Washington State scored very highly, but we eliminated it from the rankings, as Amazon stated they would not be expanding there.

We then took the top 20 states from both Bitvore's prediction and Amazon's top 20 finalist list and compared them to calibrate the experiment. While Bitvore ranked their predictions, Amazon did not provide any ranking for their values. In Figure 1 below, the center blue box contains all the state names that appeared on the Bitvore top 20 as well as Amazon's list. The number in parentheses is the ranking in the Bitvore prediction.

Bitvore Only	Bitvore + Amazon		Amazon Only
Michigan (6th)	**Texas** (1st)	**New Jersey** (10th)	**Tennessee** (22nd)
Missouri (15th)	**California** (2nd)	**Maryland** (11th)	**Ontario** (25th)
Arizona (16th)	**New York** (3rd)	**Ohio** (12th)	
Alabama (18th)	**Pennsylvania** (4th)	**Florida** (13th)	
	Illinois (5th)	**North Carolina** (14th)	
	Massachusetts (7th)	**Washington D.C.** (17th)	
	Georgia (8th)	**Virginia** (19th)	
	Colorado (9th)	**Indiana** (20th)	

Figure 1 – Top 20 States (Bitvore vs. Amazon)

Given Bitvore found 16 out of 18 (adjusted for Texas and Pennsylvania), that puts the prediction at 88.89% precision. The dark blue box on the right represents states only Amazon picked. Both Tennessee and Ontario, Canada were on the Bitvore state lists, but below the top 20 cutoff. Finally, the left represents four states that made Bitvore's top 20, but Amazon passed over. 6th ranked Michigan was the one surprise that had high marks but didn't make the cut. We were curious why that was, and discovered that the state of Michigan and the city of Detroit did offer Amazon a $4 billion incentive package through a Renaissance Tax Zone, but ended up not making the cut for municipal and economic health reasons, which we'll dive into later.

Predicting the Top 20 Cities

Similar to the state analysis, we took the geographic scores from all the Amazon HQ2 topics up until about a week before the finalists were announced. Amazon did not rank their cities, but Bitvore did rank our predictions. As seen in Figure 2 below, Bitvore correctly predicted 10 of the city finalists out of 20 for a score of 50% precision. Cities that Bitvore only predicted did have very strong proposals in the bidding round but fell out for various reasons.

Bitvore Only	Bitvore + Amazon	Amazon Only
Houston, Texas (4th, 59.33)	**Washington D.C.** (1st, 128.95)	**Indianapolis, Indiana** (26th, 17.77)
San Francisco, California (7th, 41.10)	**Chicago, Illinois** (2nd, 87.05)	**Nashville, Tennessee** (22nd, 19.23)
San Diego, California (9th, 29.31)	**New York City, New York** (3rd, 72.09)	**Miami, Florida** (23rd,19.10)
Fargo, North Dakota (13th, 25.77)	**Los Angeles, California** (5th, 42.90)	**Newark, New Jersey** (28th,15.79)
Detroit, Michigan (14th, 24.91)	**Boston Massachusetts** (6th, 42.50)	**Columbus, Ohio** (30th,15.70)
Hartford, Connecticut (15th, 23.88)	**Atlanta, Georgia** (8th, 30.53)	**Montgomery County** (N/A,N/A)
Columbia, South Carolina (16th, 23.38)	**Denver, Colorado** (10th, 27.47)	**Northern Virginia** (N/A,N/A)
Kansas City, Missouri (17th, 21.98)	**Dallas, Texas** (11th, 26.01)	**Pittsburgh, Pennsylvania** (39th,13.77)
Portland, Oregon (18th, 21.80)	**Philadelphia, Pennsylvania** (12th, 25.79)	**Raleigh, North Carolina** (53rd,N/A)
St. Louis, Missouri (19th, 21.60)	**Austin, Texas** (20th, 19.48)	**Toronto, Ontario** (N/A,N/A)

Figure 2 – Top 20 Cities (Bitvore vs. Amazon)

Some of the Amazon picks didn't map well to our cities. We treat Washington D.C. as a state, so the data had to be meshed. A lot of the news items talked about the "D.C Area" which covered parts of Washington D.C., Northern Virginia, and Maryland.

Because "Northern Virginia" and "Montgomery County, Maryland" (a county) aren't cities, we had to re-run the data picking out county information from both Virginia and Maryland. We were somewhat relieved that Loudoun County and Fairfax County were prominent in the results for Virginia, as well as Montgomery County, Maryland. These are the locations that Amazon included on their list. For the next stage of the experiment, we used those as geographies.

Finally, some of the Amazon city picks were ranked highly on the Bitvore geography scores, such as Indianapolis, Nashville, and Miami, but they didn't make our top 20 cut off. Taking the top 30 as well as including County information mappings for Virginia and Maryland would have covered 85% of Amazon's city picks. Armed with our validated set of geographies, we set about pulling together the municipal and economic health of each geography.

Virginia Counties + Amazon HQ2 Topic
Loudoun County (51.2%)
Fairfax County (19.5%)
Arlington County (17.1%)

Maryland Counties + Amazon HQ2 Topic
Montgomery County (31.6%)
Prince George's County (23.7%)

Figure 3 – Mapping Amazon Picks to Bitvore Cities

Geographic Scoring of Finalists

A full ranking of counties is below. For geographic mentions, New York counties didn't even register. Some of the other counties had already been eliminated well before the announcement too.

- Cook County, Illinois (1st, 1013.39)
- **Fairfax County, Virginia** (2nd, 201.71)
- Montgomery County, Maryland (3rd, 141.15)
- Fulton County, Arkansas (4th, 134.40)
- **Arlington County, Virginia** (5th, 88.84)
- **Loudoun County, Virginia** (6th, 80.20)
- San Diego County, California (7th, 76.06)
- Santa Clara County, California (8th, 64.48)
- Denver County, Colorado (9th, 52.16)
- Miami-Dade County, Florida (10th, 50.21)
- Los Angeles County, California (11th, 44.68)
- Napa County, California (12th, 41.96)

The state geographic ranking likewise was influenced by lobbying. California was over-represented, as two municipalities who had missed the deadlines continued public lobbying. This includes Napa Valley and Santa Clara. San Francisco, Oakland, Fremont, Richmond and Concord had put in a joint bid to offer Amazon locations in each of those cities, while San Jose had put in a separate bid of its own. None of these cities made the shortlist of finalists.

- California (1st, 721.34)
- Illinois (2nd, 473.19)
- **Washington D.C.** (3rd, 223.70)
- **New York** (4th, 203.41)
- Florida (5th, 183.71)
- New Jersey (6th, 174.89)
- Texas (7th, 295.200)
- Massachusetts (8th, 121.72)
- Colorado (9th, 120.03)
- Georgia (10th, 105.69)
- Indiana (11th, 5104.78)
- Minnesota (12th, 103.57)
- Virginia (13th, 98.13)
- Colorado (14th, 95.51)
- Pennsylvania (15th, 90.09)
- Ohio (16th, 70.20)
- Maryland (17th, 67.01)
- Michigan (18th, 63.10)
- Connecticut (19th, 50.56)
- **Tennessee** (20th, 46.45)

Municipal Scoring

To calculate the economic health, we used our tax, revenue, and economic activity events from Bitvore data. Using Amazon's RFP criteria, we came up with a scoring system based on what we interpreted as their intent, determined by public statements of the company, executives, and the RFP.

To calculate the municipal score, we took all the sector information in our system dealing with local governments and weighted them according to importance to Amazon. Using those weightings, we split them into primary, secondary, and tertiary goals. The theory behind the model is positive news in certain municipal sectors for the government is positive for Amazon to various extent, and negative news is negative. For each of the municipalities, we were able to come up with a score for each requirement based on importance.

Primary Goals

Sector/Subsector	Importance
General	25
Education/HigherEducation	25
Housing	25
LocalGov	25
TaxRevenue/EconomicDev.	25
Transportation	25
Transportation/MassRapidTransit	25
Education	20
TaxRevenue	20
LocalGov/GenPurpose/ PubImprovement	18
Transportation/Airports	15
Development	12
Health	10
Housing/SingleFamilyHousing	10
LocalGov/ParksZoosBeaches	10
Utility	10

Secondary Goals

Sector/Subsector	Importance
Development/LandPreservation	8
Development/OfficeBldg	8
Education/OtherEducation	8
Development/IndustrialDev.	5
Development/OtherDevelopment	5
Development/Redevelopment/ LdClearance	5
Education/PrimarySecondary	5
Health/Hospitals	5
Housing/SingleMultiFamilyHousing	5
Transportation/OtherTransportation	5
Utility/PollutionControl	5
LocalGov/OtherPublicService	4
Transportation/Airlines	3
Housing/MultiFamilyHousing	2
LocalGov/LibraryOrMuseums	2
TaxRevenue/OtherRecreation	2
TaxRevenue/StadiumsSportsComplex	2
TaxRevenue/Theaters	2

Tertiary Goals

Sector/Subsector	Importance
Dev/CivicConventionCenters	1
Education/StudentLoans	1
Health/OtherHealthcare	1
LocalGov/Courts	1
LocalGov/FireStationEquipment	1
LocalGov/GovtPublicBuildings	1
LocalGov/PensionFundingRetire.	1
Transportation/Bridges	1
Transportation/TollRoad&Hwy	1
Utility/FloodCtlStormDrain	1
Utility/Gas	1
Utility/PublicPower	1
Utility/OtherUtilities	0.7
Dev./CorrectionalFacilitiesJails	0.5
Health/NurseHomes	0.5
Health/Veterans	0.5
Housing/RetirementHome	0.5
LocalGov/PoliceStationEquip	0.5
Transport/SeaportsMarineTerm	0.5
Transportation/Tunnels	0.5
Utility/WaterAndSewer	0.5
Utility/Sanitation	0.2
Utility/SolidWaste	0.2

Figure 4 – Municipal Scoring Goals

Pulling It All Together

Image Courtesy of [WRAL TechWire]

To calculate a final score, we used the Bitvore data between the announcement of the finalists and Bezos' announcement in early November that they would be selecting two sites, not just one. These scores are a combination of:

- City, state, and county geographic rankings
- Economic health scores of the geographies
- Municipal scores weighted by Amazon's criteria

Amazon HQ2 Finalists

Rank	Site	County	State	Final Score
1	Fairfax County	Fairfax County	Virginia	106.5762292
2	Toronto	York County	Ontario	98.97696744
3	Dallas	Dallas County	Texas	85.59466958
4	Denver	Denver County	Colorado	84.99287573
5	Indianapolis	Marion County	Indiana	83.77261629
6	Pittsburgh	Allegheny County	Pennsylvania	81.89380888
7	Atlanta	Fulton County	Georgia	79.88506026
8	Boston	Suffolk County	Massachusetts	78.97458307
9	Newark	Essex County	New Jersey	78.42523709
10	Raleigh	Wake County	North Carolina	77.95892231
11	Loudoun County	Loudoun County	Virginia	74.79144404
12	Montgomery County	Montgomery County	Maryland	72.3138568
13	Nashville	Davidson County	Tennessee	71.21252844
14	Columbus	Bartholomew County	Ohio	70.66383665
15	Austin	Travis County	Texas	67.68931698
16	Los Angeles	Los Angeles County	California	66.25609360
17	Arlington County	Arlington County	Virginia	65.60353329
18	Miami	Miami-Dade County	Florida	63.10437949
19	Chicago	Cook County	Illinois	62.65039693
20	New York City	New York County, Kings County, Bronx County, Richmond County, Queens County	New York	57.74520573
21	Washington D.C.	District of Columbia	Washington D.C	57.0736693
22	Philadelphia	Philadelphia County	Pennsylvania	56.23267411

Figure 5 – Final Results

Fairfax County: Fairfax County ranked first on our list and was chosen as one of the sites for Amazon's HQ2 as part of the November 13th, 2018 announcement.

Toronto, Ontario: Toronto claims that participating in the contest raised the profile of the region and the ability to gain industry growth without extreme tax breaks. Toronto scored very high in our model due to excellent education and transportation scores.

Dallas, Texas: One Amazon site selector is rumored to have said that top-5 contender Dallas didn't have a chance due to the volume of jobs and real estate.

Denver, Colorado: Denver had issues with transportation and difficulty in finding tech workers although Amazon still plans one-time investments in the area.

Indianapolis, Indiana: The Indianapolis proposal remains under NDA, although it's being reported that unlike other locations, they won't even get a consolation prize.

Pittsburgh, Pennsylvania: Pittsburgh's offer of $6.7 billion failed to entice Amazon. City and business leaders said it demonstrated Pittsburgh is an exciting place to do business, but the proposal process also put their weaknesses on display. These included a lack of tech-ready employees, poor transportation, unfavorable state taxes, and mediocre public schools.

Atlanta, Georgia: After the NY fallout, Atlanta was rumored to be one of the three cities under re-consideration, but nothing ever came out of their $5 billion in incentive offers. Some claim that the inability to negotiate silently with city leaders and real estate developers was one of the key issues driving Amazon's decision.

Boston, Massachusetts: Boston has world class talent, but it missed out as the top choice. Despite announcements that Amazon was expanding jobs in their Boston offices, it was passed over for reconsideration due to its smaller population and dense real-estate.

Newark, New Jersey: Newark offered Amazon $7 billion reasons to pick them, but unfortunately it didn't work out. Despite its proximity to New York City, Amazon employees weren't enthralled with living in the Garden State and coordination between city leaders and real estate developers proved difficult.

Raleigh, North Carolina: Raleigh offered a competitive incentive package, but city leaders and tech triangle executives say Amazon cooled on the area early on. While leaders were never told they were formally out of the running, the lack of communication made Raleigh unlikely to be selected.

Loudoun County, Virginia: Along with Arlington and Fairfax Counties, National Landing is the chosen site for HQ2 with higher education, housing, and transportation scoring very high.

Montgomery County, Maryland: Amazon passed over Montgomery County for the nearby Virginia counties. However, they will still be close enough to reap some economic benefit, without having to provide any tax incentives. Loudoun County expects up to 8% of the Amazon HQ2 workforce to live in their boundaries.

Nashville, Tennessee: Despite losing out in the initial picks, Amazon decided to greatly expand their employment and investment in Nashville after the NY withdrawal. Even though Amazon is not building an actual HQ2 in the city, the expanded office space and investment provides all the benefits, as if they were a consolation prize.

Columbus, Ohio: Columbus offered an incentive package of around $2.6 billion, but it wasn't enough to overcome their lack of infrastructure, economic base, transportation plans, and quality of life issues that Amazon employees were looking for.

Austin, Texas: According to news reports, Austin breathed a sigh of relief after not being chosen. The region is already experiencing rapid technology and economic growth, and some city leaders worried that Amazon would push their transportation, schools, and municipal services well past their ability to absorb new growth.

Los Angeles California: Los Angeles felt it was already a long-shot due to Amazon's current headquarters being located on the West Coast and in the same time zone. Lack of affordable housing, quality of life issues including traffic, competing for talent with other Silicon Valley tech companies expanding into Southern California, and other issues pulled them out of the final decision.

Arlington County, Virginia: Northern Virginia was chosen as the split-site for Amazon's HQ2 as part of the National Landing project. After Amazon reconsidered its NY deal, Arlington has had some pushback on livable wages, hotel occupancy taxes, and labor agreements surrounding occupied real estate space. As of now, Amazon is still moving forward here.

Miami, Florida: Miami, like a few other cities, did not get chosen nor receive any consolation prizes of increased employees or investment. They pitched their proposal as a gateway to Latin America, but Amazon, despite Miami's diverse and global appeal, didn't score them as well on world-class workforce recruiting and education.

Chicago, Illinois: Chicago scored well on a lot of Amazon's requirements. It was eliminated due to quality of life, economic growth, and tax issues that put them at a disadvantage to other sites.

New York (Cities and Counties): This was the big surprise. Although it scored low in Bitvore's rankings, Amazon chose New York as one of the sites for their HQ2. Despite having the ability to bring $13.5 billion in tax revenue and 40,000 new jobs to the site, city, county, and state leaders pushed back against the $3 billion in economic incentives. The backlash caused several real estate developers to block permits for development sites that were already agreed upon. Residents complained the deal would drive up housing prices, a criticism supported by the fact that some Amazon employees purchased housing before the final announcement. There were some efforts made to rescue the deal, but Amazon categorically stated that success comes from mutually-beneficial partnerships, and many local politicians adamantly opposed their presence.

Washington, D.C: D.C.'s proposal was an amalgam of properties cobbled together to form the Capitol Riverfront project. Amazon chose a D.C. suburb with better access, higher quality of life and better transportation, located across the river from D.C. proper.

Philadelphia, Pennsylvania: Philly offered $1.1 billion in incentives to Amazon only to be rejected. Unlike other sites, Philly also offered up to $4.6 billion in long term financial incentives for a total of $5.7 billion. Despite one of the largest incentive packages, Amazon didn't feel Philadelphia had a large enough tech economy, tech recruiting environment, and would require far too much investment in transportation, schools, and other services.

Model Validation

Amazon HQ2: Bitvore Predictive Model

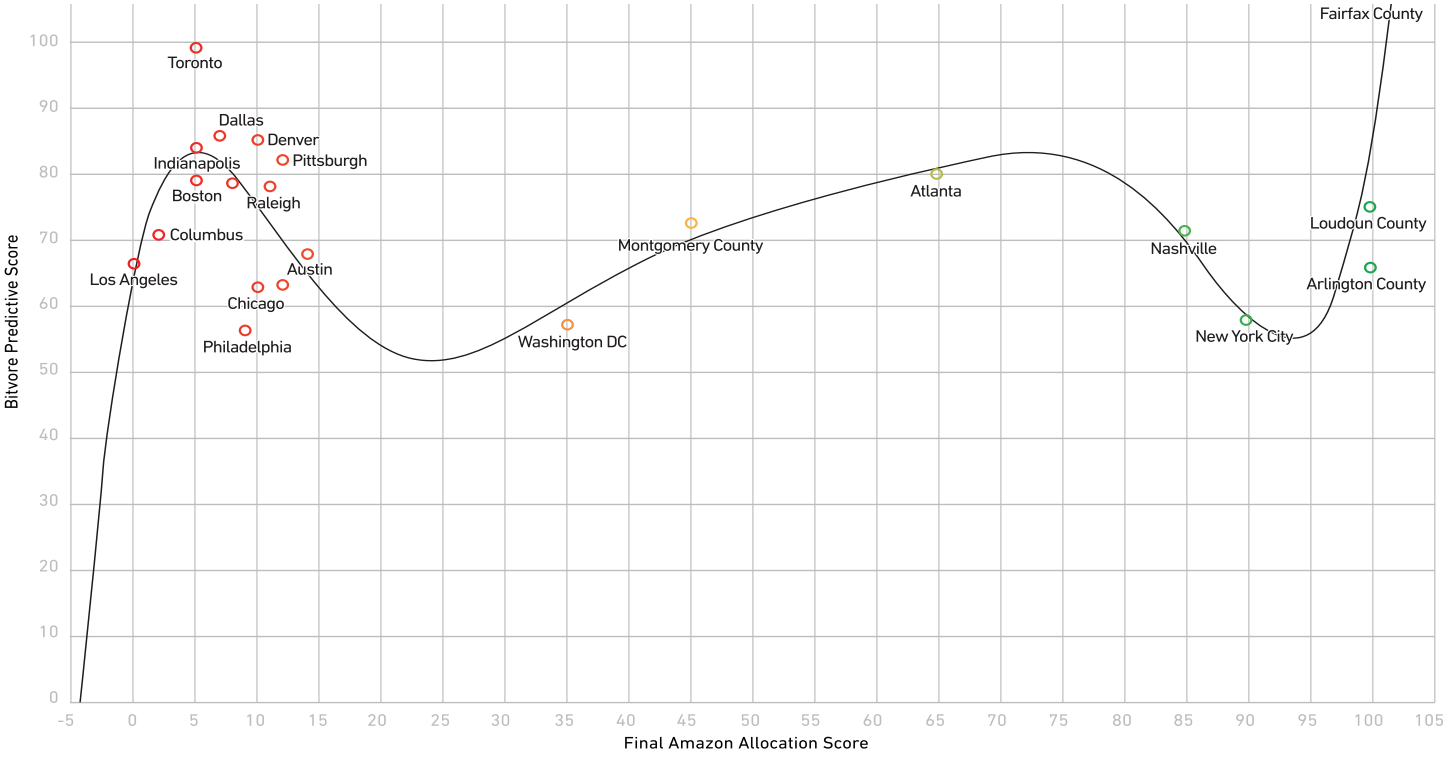


Figure 6 – Model Validation

The final thing we did was to validate the accuracy of our predictive model using a least-squares algorithm. We assigned a score to each site based on announced benefits to the area. For example, neither Nashville nor Boston's Seaport District were chosen as sites for HQ2, but Amazon announced generous investment and increased employment for those areas.

Likewise, Culver City and parts of Texas are being expanded for various line of businesses. We used a polynomial trend line, which is simply a curved line that tries to minimize the distances between the actual scores and the predicted scores based on some equation. It is very useful for categorizing data and minimizing the number of "logical backflips" that have to be done to explain the data or excuse the model for making wrong predictions.

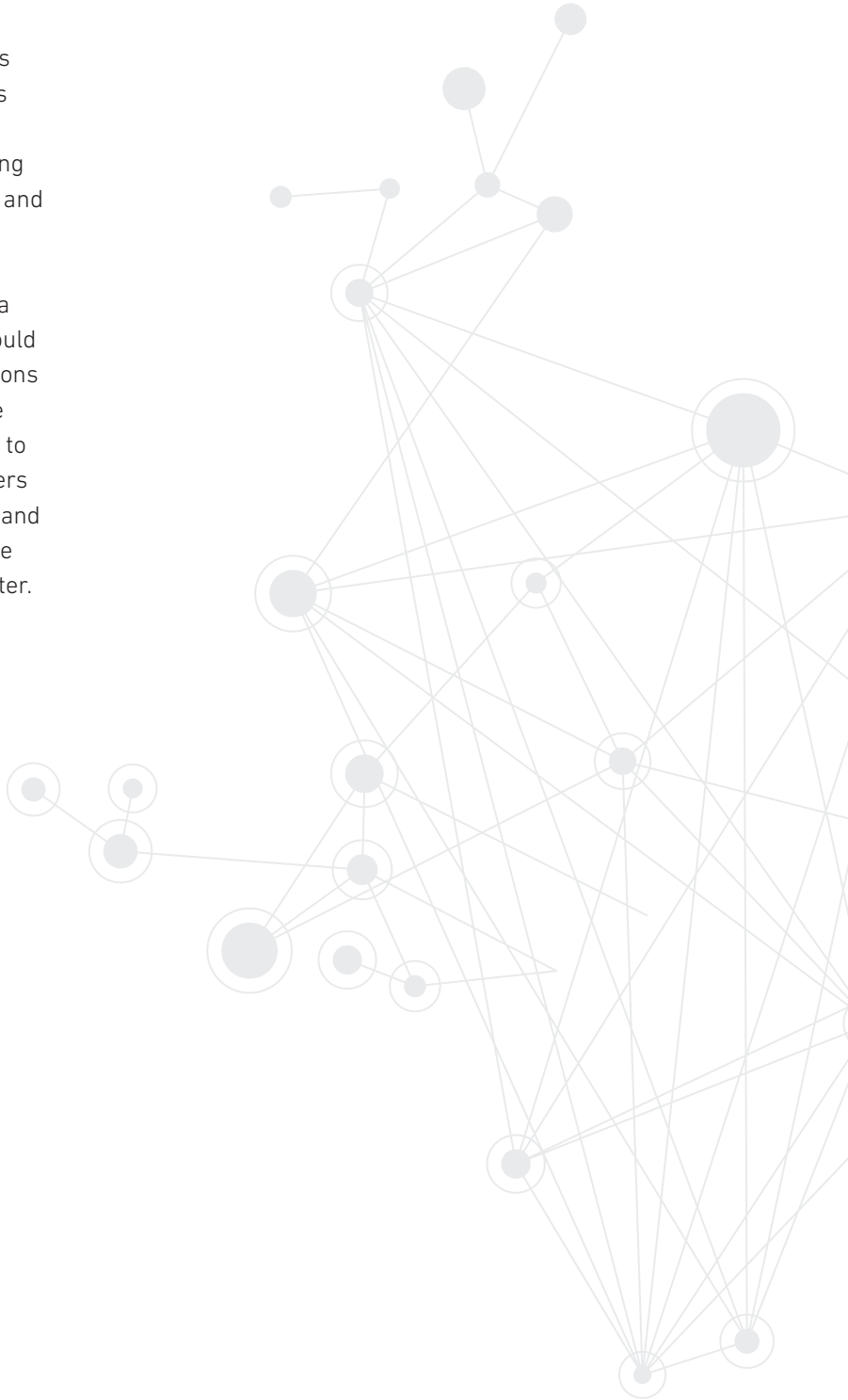
We took the final scores and calculated the degree of inflections using a line-fitting algorithm. In Figure 6 above, you can see the results of the polynomial trend line with a degree of 7 (based on 50 candidate cities). The bottom line is, the sites that were selected both had a Bitvore prediction score above 50 and an Amazon post-selection allocation score above 50. Montgomery County will enjoy some of the economic benefits due to potential workers living and commuting to the neighboring counties. While New York City scored very high initially with Amazon, Bitvore ranked it fairly low, mostly due to municipal and governmental tax issues.

It would be misleading to say that Bitvore would have seen the New York HQ2 falling through based on our insights into the track record of municipal tax, development, and property issues, but the warning signs where there.

Conclusion

Bitvore obviously had a rich aggregated, cleansed, normalized and tagged news data set relevant to the Amazon HQ2 competition that we could quantify for this experiment. It included tens of thousands of data points on 54 states in the U.S. and Canada and hundreds of cities. While our predictions were not perfect, our scoring models, which utilized Bitvore's geographic, municipal, and economic data, gave very good results.

All the data we used was from news, however, other data sources could have been added to the analysis, which could have made our predictions better. This is one of the reasons that Bitvore continues to aggregate, clean and normalize data from other, non-news sources of unstructured data to develop alternative investment data. In fact, our customers data science teams often use our data with other public and proprietary data sources to build very accurate predictive models, so their analysts can make better decisions, faster.



About Greg Bolcer, CDO Bitvore

Greg is a serial entrepreneur who has founded three angel and VC-funded companies. He's been involved at an early stage or as an advisor to at least half a dozen more. Greg has a PhD and BS in Information and Computer Sciences from UC Irvine and a MS from USC. He started his career at Irvine as a researcher in Web protocols, standards, and applications under a series of DARPA-funded grants. He formerly was the Intel Architecture chair for the Peer to Peer working group and was awarded the Distinguished Alumni of the Year in 2004 from UCI.

Read more from Greg on the Bitvore Blog.

About Bitvore

Bitvore provides precision intelligence derived from world business news and information. Our products are deployed in over sixty of the world's largest financial institutions, allowing them to rapidly create augmented intelligence solutions to address their unique business requirements. Augmented intelligence solutions assist employees in making faster and more effective decisions, so they outperform the competition.

To learn more, visit bitvore.com or contact us at Bitvore.com/contact-us/.

Precision Intelligence.
Proactive.
Personalized.

866-869-5151
info@bitvore.com
bitvore.com

BITVORE