Commercial Property Data Aggregation

A Major Qualifying Project proposal submitted to the Faculty of
WORCESTER POLYTECHNIC INSTITUTE
in partial fulfillment of the requirements for the
Degree of Bachelor of Science

Submitted by:

Cecily Coia
John Cotter
Samuel Ellison
Scott Lonergan
pips_mqp@wpi.edu

Date:
April 26th, 2016

Submitted to:

Project Advisor:

Professor Jon Abraham

Project Liaisons:

Brian Hayes, Quincy Mutual Fire Insurance Co
Andrew Nersessian, Quincy Mutual Fire Insurance Co

Table of Contents

Abstract

The goal of this project was to find and evaluate new data sources to be used by Quincy Mutual when evaluating and writing policies for commercial properties. First, we examined the workflow and business needs of Quincy Mutual. We then searched for public and private data sources that could be used to supplement current data collection. Furthermore, we evaluated the data for accuracy and attempted to build relationships with source providers. Following this, we developed methods for integrating the data into Quincy Mutual's business flow. Finally, we offered recommendations and plans for streamlining this process. Our final deliverables were a source list, methodology for finding sources, a list of recommendations, and contacts to data providers.

Executive Summary

Our project centered on the commercial property lines department of Quincy Mutual, a medium sized insurance company located in Quincy, MA. Living in the Information Age, our goal was to help Quincy Mutual capture more data in their workflow in order for them to better understand the risks that they are facing when writing commercial properties. Currently, Quincy Mutual relies heavily on Loss Control Surveys, which are conducted by professional inspectors who provide detailed reports on the properties they are interested in insuring. Unfortunately, these reports are expensive, time consuming, and are always gathered at the tail-end of the information collection process. Our project's focus was to find alternative data sources for Quincy Mutual to utilize and supplement the Loss Control Surveys to save time and money.

After familiarizing ourselves with the company and its workflow, we conducted a survey with the underwriters to better understand the information they find most useful. We then attempted to find this information among various third party sources. In the public space, we were successful in accessing Worcester city property records through city hall. The private sector proved more challenging. Realty sites and other data providers were difficult to contact as students, but we were able to use their information on a property by property basis through their websites.

Quincy Mutual provided us with a sample of properties in Worcester that they currently insure accompanied by Loss Control Surveys for each. Due to the reliability of the data collection for these surveys, we deemed these data points as true values. Comparing the values we collected from various sources to the Loss Control Survey allowed us to determine the accuracy of each individual source. Quantitative data such as square footage and building value

showed patterns of inaccuracy across all data sources, leading us to doubt the validity of this data.

As a final step of the process we created property profiles for these locations using exclusively third party data sources. The goal of these profiles was to mimic the format of a New Business Application or Loss Control Survey. The information gathered was then entered into Quincy Mutual's Quick Quote system, allowing us to compare theoretical premium prices to true Quincy Mutual premiums. Despite fairly inaccurate values for square footage and building value, our results were surprisingly similar to the true premiums Quincy Mutual had on their books.

We recommend Quincy Mutual use our process for generating property profiles to supplement their information. However, they should remain cautious of the various flaws present in publically available data sources when making decisions. Additionally, we recommend that Quincy Mutual continue to pursue relationships with private data companies including Renaissance Alliance, LexisNexis, and GigWalk. These companies offer higher quality data that may prove to be hugely valuable as Quincy Mutual looks to expand into new states and regions.

Chapter 1: Introduction

The world we live in is driven by data, technology, and analytics. This is especially evident in the world of business, where access to more data can provide enormous value to a company. Hopefully somewhere in this wealth of data, there is a trend or pattern that could revolutionize their business. The insurance industry is especially data driven. Insurance involves aggregating a multitude of diverse risks, collecting information on these risks, and assessing this risk based on the known characteristics. An insurance mindset says that the more relevant data you have on a specific risk, the more accurate you are in assessing its propensity to see loss. Finding ways to better access, manage, and use Information can be the difference between insolvency and prosperity in a rapidly changing industry.

Quincy Mutual Fire Insurance is looking to get the competitive advantage by incorporating more sources of data, specifically in their Commercial Lines department. This company, along with a vast majority of the insurance industry, is slowly creeping its way out of the archaic past and into a paperless, cutting-edge, and analytically focused future. With these changes comes availability to expand a company's capabilities, which is a crucial advantage in today's competitive insurance market.

Quincy Mutual currently relies heavily on the data provided to them by New Business Applications and Loss Control Surveys. The New Business Application is typically filled out by an agent about a specific property. This application is then passed on to an Underwriter to review and decide if the property fits Quincy Mutual's business appetite and adjust the price accordingly. Once the property is bound Quincy Mutual requests a Loss Control Survey which they then have 30 days the reject the property based on the Loss Control findings. As a way to improve their business flow there has been interest in looking outward to see what kind of data is

publicly or privately available. This information could prove to be more accurate, detailed, or even cost effective, but there needs to be research done to assess and compare the two streams of data.

Some solutions that have been used in the past include Underwriters using third party data sources to inquire about a property in addition to New Business Applications and Loss Control Surveys. From what they have found on these other sources such as Google Earth, Zillow, etc. It appears this could be a move in the right direction to collect data on a property. The goal of this project is to create a quicker, more reliable, way to inquire about information on a property in order to price it more efficiently and accurately. We aim to create a database of property information using solely third party resources. Through this process we also hope to create a new business flow that includes a prospective way to look for business that is desirable to expand the commercial lines within Quincy Mutual. Figure 1 contains a visualization of our proposed prospective way to incorporate third party data sources in the search for new business.
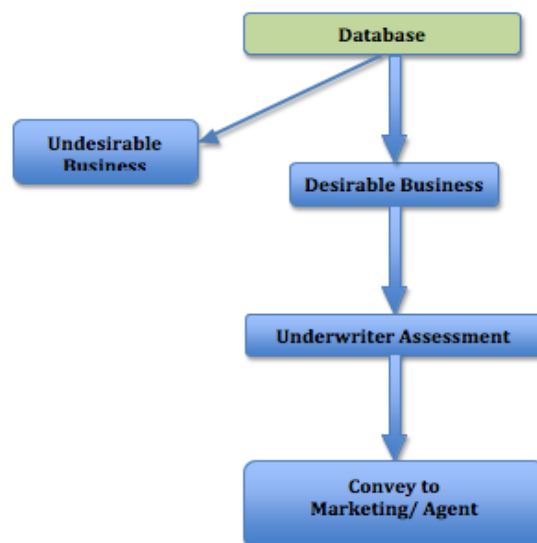
Figure 1: The proposed procedure for Quincy Mutual New Business.

The creation of the database will consequently create a new process flow involving the acquisition of new business. We want Quincy Mutual to have capability to filter through desirable and undesirable properties from the start. Once this is filtered, the desired business is passed on to an underwriter who can then make a professional assessment on these properties. Lastly the underwriter will convey the properties they are interested in to the agents to help them attain these properties.

This project is just a start to the endless opportunities third party data sources can offer to the insurance world. Our goal is to find and evaluate as many reliable sources as possible to improve the business flow within Quincy Mutual. In order to stay competitive in the insurance world this could be a huge step in the right direction.

Chapter 2: Background

2.1. Quincy Mutual

Quincy Mutual Fire Insurance is a small property and casualty insurance company and a subsidiary of the Quincy Mutual Group. Their headquarters is located in the heart of Quincy, MA. The company was founded in 1851, and today still carries their original name and have expanded their business insurance products in the private and commercial space. Quincy Mutual's personal lines operation is much larger, although their Commercial Lines business has seen consistent growth for the past several years. This project focuses on Quincy Mutual's Commercial Lines department, and all further discussion relates to this portion of their business unless otherwise specified (Welcome, 2014). Within Quincy Mutual's Commercial Lines department, insurance products are sold covering Commercial Auto, Property & Liability, and Workers' Compensation. Our project focused on the Property & Liability product specifically.

2.2. Quincy Mutual Services and Workflow

Quincy Mutual's Commercial Lines department provides coverage to New York, Massachusetts, Rhode Island, and Connecticut, with most of their business concentrated in Eastern Massachusetts. They write business through an extensive network of third-party agents. Agents act as an intermediary for the insurance company and the insured, with the reward of commission on all new business that is written. As a result, the Quincy Mutual staff gains the benefit of having business brought to them as well as never interacting with the policyholder unless they are filing a claim. To better understand the business flow of Quincy Mutual's

commercial lines, let us follow the events that unfold when Company A wants to purchase insurance.

If Company A wants to cover themselves from potential lawsuits or damages to their property, they approach an insurance agency for a policy. The insurance agent takes note of the types of coverages Company A desires and tries to find the best fit for the insured. An experienced agent may be aware that insurance companies prefer to write certain business classes or neighborhoods, and might steer the applicant towards specific companies. The agent now completes online applications for some discretionary number of insurance companies. If Quincy Mutual is one of these companies, the application enters their business flow.

Quincy Mutual's business cycle consists of two primary stages: New Business and Renewal Business. Each of these stages involves the collection or reaffirmation of distinct pieces of data. When the online application is received by Quincy Mutual staff, the potential policy is sent to the underwriter and is now referred to as New Business. The application contains an extensive questionnaire that includes desired coverage and property information. The underwriter this New Business is assigned to will review the application and determine if the property's potential risks match Quincy Mutual's appetite. If the underwriter determines the insured is a good risk, they can adjust the quote from a suggested price given by automated software and send the quote to the agent.

Upon receiving the quote, the insurance agent and Company A review their options. If they determine that Quincy has the best price and coverage for them they send a notice to the underwriter that they wish to bind the policy. After binding the policy, coverage begins immediately for Company A and Quincy Mutual orders a building inspection to be performed at all properties included in this policy. From this point Quincy Mutual has approximately 60 days,

depending on which state the property resides in, to decline the policy. Therefore, it is crucial that the building inspection is completed within this timeframe to identify risks and exposures that were not shown in the application.

After the Loss Control Survey is received it is sent to the underwriter for review. Each Loss Control Survey is stored in Quincy Mutual's databases for reference upon renewal. In the event of an unfavorable Loss Control Survey, the underwriter may decide to send a cancellation notice to the insured. This will happen primarily because of an unforeseen hazard that the MFIA inspector noted in the survey. Otherwise, the policy continues and is checked on again 150 days before the policy renewal date. If recommendations were present in the Loss Control Survey, a letter is sent to the insured notifying them that they must be corrected before the renewal date. Figure 2 below illustrates and simplifies this business flow.
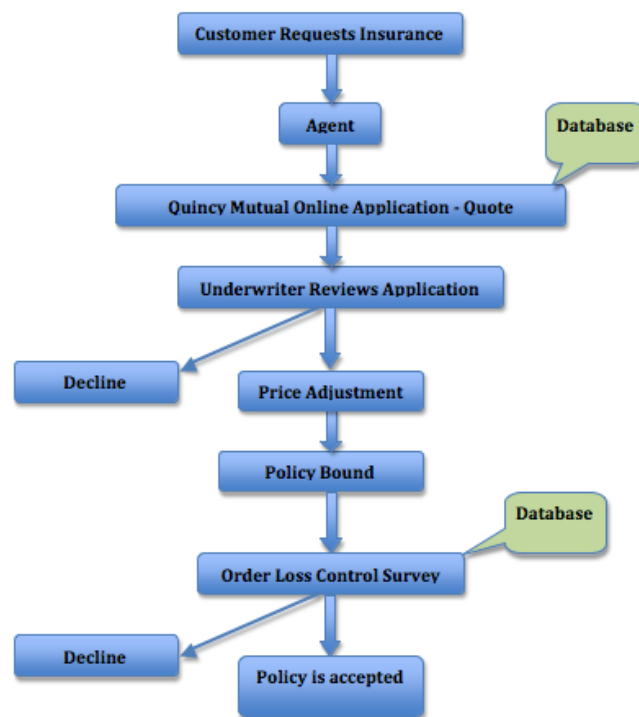

Figure 2: The procedure for Quincy Mutual New Business. Nodes with a "Database" balloon identify areas in which Quincy Mutual collects data.

Following the conclusion of this process, each policy is considered Renewal Business. All Renewal Business is checked on a yearly basis in order to examine any changes that may have happened in the past year for rate or coverage changes. For example, if a business built an additional level on their building, they would need to adjust coverage and premium to ensure the entire property is protected. Leading up the renewal date, an abbreviated version of a Loss Control Survey is ordered to examine the property for any significant changes in the previous year. One of the primary functions of these further inspections is to check for the completion of recommendations that were outlined in the Loss Control Survey.

2.3. Quincy Data Management and Loss Control Surveys

There are several instances in which Quincy Mutual collects data points for properties that they see. Initially, information comes almost entirely from the agent and a few third party sources selected by Quincy Mutual underwriters. This information is later verified by a contracted building inspection. All of the information obtained from the New Business Application is housed in a custom-built system called Quincy Online. This program connects Quincy Mutual's employees to the agents that bring business to them, and acts as the warehouse for all information provided by the agents.

The bulk of Commercial Lines underwriting data is produced by the Mutual Fire Insurance Association of New England (MFIA N.E.). This organization has trained building inspectors distributed across New England that complete and record property inspections for the 19 companies that belong to the association (About, 2010). Quincy Mutual orders these inspections to be completed on every property that they write. Their underwriters receive the information from these inspections in Loss Control Surveys. These Loss Control Surveys contain

comprehensive information for underwriters including building valuation, identification of various hazards, and COPE (Construction, Occupancy, Protection, Exposure) fields. Quincy Mutual underwriters generally never need information beyond what MFIA Loss Control Surveys are able to provide; however, in recent years the timeliness of the inspections has wavered, and they can no longer be guaranteed for New Business written by the firm. As a result, their underwriters have had to rely more heavily on information they find on the internet or through their agents.

Several years ago, Quincy Mutual operated almost entirely on paper. Today, all of the MFIA Loss Control Surveys that Quincy Mutual receives are stored digitally through an image file program called ImageRight. Each of these inspections can be easily searched for, but the individual fields in each survey cannot be queried as a consequence of the image format. Recently, Commercial Lines has gained the ability to obtain survey information in aggregate directly from MFIA's servers. While there is not a live version of MFIA data in Quincy Mutual's systems, this new capability allows Quincy Mutual to pull inspection results for a variety of analysis purposes.

2.4. Public Property Data

States have been collecting money from its citizens through property taxes since the early 1800's (Carlson, 2004). To ensure the states are collecting the correct amount in taxes, each state has a department that assesses properties to determine their value. These records are updated periodically on an annual, five, or seven year basis depending on the state. Most states will also update the information every time the property is purchased (Assessor, 2016). There are many approaches to assessing property, but a commonly used practice is the Replacement Cost

approach where the professional examining the property uses square footage, building age, condition, and construction materials to determine the value of the property (Division, 2014). The information collected in these assessments are very important when Quincy Mutual is looking to write a policy for a property so the team will explore the possible use of public data in our project.

2.5. Private Property Data

The emergence of private property data started in the 1970's as real estate became a lucrative industry. In 1976, the first publication of Homes & Land was distributed in the US describing numerous properties with pertinent information (Homes, 2016). This could be considered one of the first times a private company compiled property data and widely distributed it in hopes of improving business. Homes & Land stayed relevant through the years and at the turn of the century was one of the major real estate players beginning to use the internet to reach more potential buyers as Homes.com. As seen in the figure below, since 2000 more and more home buyers are beginning their search on these real estate websites (Stone, 2013). This demand has resulted in the formation of big data for property information with these sites pulling from various public sources as well as having users submit information regarding their own properties. Moving forward, we will try to utilize these private firms in order to help Quincy Mutual gain a better insight into the properties they are potentially insuring before a Loss Control Survey has been ordered. We will also try to determine if these sources are accurate and reliable before recommending to Quincy Mutual to use these third party sources.
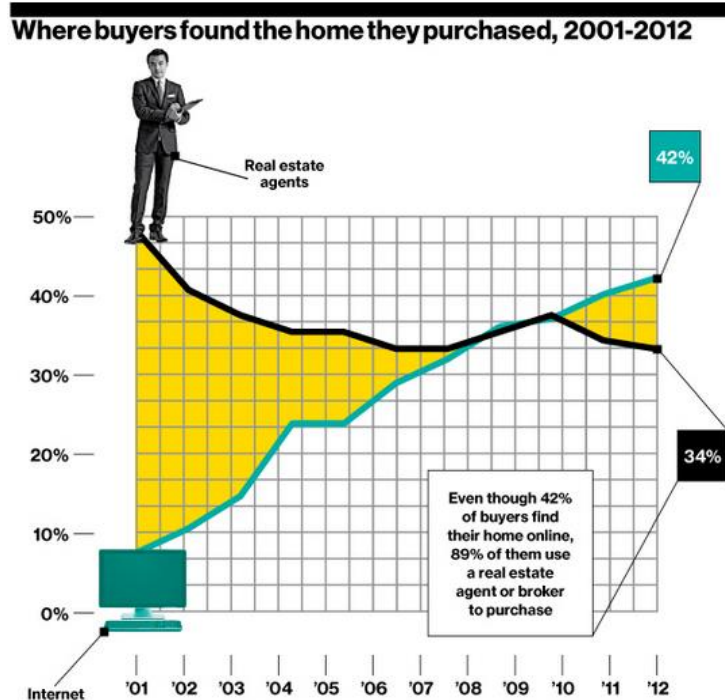
Figure 3: Rise of Technology in Real Estate (Stone, 2013)

2.6. Evaluating Data

There is a large volume of information and sources in today's technology driven world, but how do we determine what is the most reliable and accurate? Northern Michigan University developed a system of evaluating internet sources that the team referenced when analyzing new information. The guide highlights six criteria to examine for each source.

When looking at a new data source, we should determine the authority, or who is providing the information. Does this company, firm, or organization have a history of being reliable and do they have any necessary certifications to publish this information? After considering this, begin looking at the data to determine the accuracy of what they are publishing. Although we will not know the true values of certain aspects of any given property, we can cross

reference the data with other sources to see if this new site is providing similar and reasonable information (Northern, 2009).

It is also important to determine if the information has any form of bias built into it. This can be determined by learning the motives of the organization that is publishing the information as well as identifying the target audience of the data. Real estate is a constantly changing industry as properties are improved by their owners or even bulldozed to be replaced by new ones. For this reason, the sources with the most up to date information are some of the most valuable for our project. Determining the rate at which the data is revised is a very important factor we must consider when compiling sources (Northern, 2009).

The guide also directs the user to look for citations or references to where they are getting their raw data from. This aspect may be difficult for us to find on the private companies' sites because the places where they compile their information from is a key part of their competitive edge. And the last the aspect we should evaluate is the appearance of the source's website. This could include the overall layout of the site as well as the use of proper and professional wording and explanations on various pages (Northern, 2009).

2.7. Property Data in Other Industries

Property data has a variety of potential uses beyond what was envisioned during the inception of this project. The following section discusses several instances in which third party property information might find use.

The database and associated comparator tool we are creating is most applicable to the insurance industry simply because their design was oriented for it. Quincy Mutual has a plan for its use spanning multiple business operations including underwriting and marketing. However,

limiting the ability to generate this database may restrict its utility, even among only the commercial insurance industry. Insurance companies of different sizes may find ways to fit our methodology into their business operation in a separate manner from that planned by Quincy Mutual. Our methods for data validation and authentication may be the missing component to a company's customer application process. Conversely, another insurance company might seek to use the business scoring model in a similar manner to Quincy Mutual in order to boost their marketing efficiency.

In addition to other insurance companies benefiting from our data, another key industry is contracting. Items such as when a building was built or what the building is made out of are important items a contractor should be aware of when beginning a project. Knowing factors like building materials will enable the contractor to better estimate costs before even arriving on site. Details such as year built and roof materials and conditions are also fields this database will have that would benefit a contractor.

Another area that could benefit from this data is the real estate industry, a sentiment affirmed by the National Association of Realtors, who recently wrote an article about "how big data and mobile computing can be leveraged by brokers, agents and real estate portals to more effectively market homes to consumers" (DeSanctis, 2014). The most important benefit lies in agent efficiency due to the increased amount of information the real estate agents would have about a property. A more expansive and detailed source of data would allow the agents to communicate valuable and otherwise unavailable information to the consumers. Agents with this capacity could reach more customers and leave customers feeling more satisfied with the property they choose. It is getting exceedingly easier to aggregate more detailed and extensive

data on properties. Having a wealth of data in one place would provide customers with an interface to filter their property search just by the characteristics they desire.

Similar to realtors, potential or established business owners would benefit from comprehensive third party data. When looking to expand or purchase their first commercial property space, many factors play a role in their decision. Alongside location, other characteristics such as sources of potential liability or costly repairs should be considered by the buyer. The database we hope to compile will include information on the condition of sidewalks, the roof, and other key components of the structure. This will allow future buyers to have a better understanding of the property before even visiting the site.

To provide loans to commercial businesses, banks examine a variety of information including the business credit score, owner character, and capital. While it might not be possible to acquire information about a company's finances, our database will include information such as property value and credit score. This data can then be authenticated using our comparator tool alongside the bank's internally sourced information. The bank may also find data about the condition of a property useful when determining the business's quality of management. Additionally, it has been found that "geographic proximity [to a bank] improves the quality of private information" (Agarwal, 2010). As we seek to populate a database around a centralized location, this distance may become an important piece of our comparator tool.

Chapter 3: Methodology

3.1. Primary Sources Process

3.1.1. Initial Individual Research

The first search for data sources involved loosely structured individual research. All potential sources were initially divided into two distinct groups: paid sources and freely available sources. The goal during this stage was to find any data that is highly visible online, and to explore the various organizations that provide property data. Here we also investigated what format online data would be presented in. All of the sources that we found at this point were stored in an annotated bibliography.

The city of Worcester, MA was used as a point of study during the entirety of this project. This city was chosen for three primary reasons. First, because we resided in Worcester during the completion of project work, and therefore could access physical data sources and properties for verification purposes. Secondly, it is a city that Quincy Mutual has written many policies in, and therefore had recent Loss Control Surveys available for comparison purposes later on. Lastly, Quincy Mutual was looking to further expand their business in this region of Massachusetts, so the information found would be directly beneficial to their current operation. Each member of our team used a separate approach during this stage.

3.1.2. Data Scraping

One of the most significant challenges we faced was figuring out how to quickly aggregate data for many properties across multiple websites. Almost all of the websites that we found only displayed information for one property at a time. Without access to the aggregate data

behind the scenes of the website, each property would need to be searched for manually across all important websites. Building an expansive database containing thousands of properties using this technique would require exhaustive efforts and be difficult to maintain. As a response to this, we looked into a technique called "data scraping". This involves the use of a programmed script that runs on a website and quickly extracts desired data fields.

While the creation of this program was possible, it was soon discovered that nearly all of the websites we found useful specifically forbid the scraping of their websites in their terms of use. Further research showed numerous legal cases in recent years involving businesses scraping the websites of their competitors. In particular, New York had seen a sharp increase in the number of data scraping lawsuits in recent years. In order to avoid legal ramifications for Quincy Mutual in the future, it was recommended that data scraping not be used as a method of data collection.

To avoid this problem, we decided to focus on a small subset of Worcester properties. By this method, data fields were manually extracted from third-party sources for a smaller subset of the population. This allowed us to create a sample of a complete database. We also attempted to make contact with all major data publishers in order to access this data straight from the producers. If a deal could be made between the companies providing the data and Quincy Mutual, it would be easier to feed the information directly to Quincy Mutual's staff.

3.1.3. Underwriter Survey

While extracting more than 55 data fields from the New Business Application, we investigated how crucial each field is for underwriter decision making. We were interested in the following: Does Quincy Mutual rely heavily on all of these fields? Are some fields more important than others? Which fields relate specifically to pricing a property that are captured by

this application? Discovering which data points are valued most heavily by underwriters allowed us to prioritize our further research. We accomplished this by surveying the underwriters about the fields that they found important. Additionally, we wanted to interview the underwriters to discuss how they price a policy, as well as what information they feel is missing.

We wanted to prompt the interview with the Underwriters by first sending them a survey to help us better understand how they interpret the information they are given. Some of the questions we included in the surveys were:

- What information that is not currently collected through the New Business Application and Loss Control Survey would you like to know about a property?

- Are there any specific websites that you consult to find out information about a property? List them.

- Do you find it easier to find information about properties in specific states or regions than others? Provide examples.

- Could you please list the top five field selections that you look at most when underwriting a property?

We also included a table of new fields we thought might be helpful when underwriting a property that were not included on the New Business Application. We asked that they rate them on a scale of 1-10, 10 being the most useful. A full list of these fields is found in Appendix A.

Sending the surveys ahead of time, we then held a meeting at Quincy Mutual to discuss their answers further. We felt a group conversation would provide our team with a better understanding of what each individual looks at when underwriting a property. During this meeting, one of our objectives was to identify points of focus for the duration of the research. We will continue discussion of the Underwriter Survey and meeting in the Results section.

3.1.4. Cross Referencing Quincy Mutual Data to Other Sources

We then sought to examine how many fields that the underwriters felt were most important had been found through our third-party sources. The objective was to determine how much of Quincy Mutual's current data could be replicated externally. This was accomplished by cross referencing these two groups with a spreadsheet. In our data matrix, we placed each source running across the x-axis of an Excel sheet with all the data fields (year built, building materials, etc.) that Quincy Mutual currently collects or desires to collect running down the y-axis.

If the data point was found, we would mark the box corresponding to the data field and source it was found on. With the information organized in this manner, we were able to better visualize which sites offered various pieces of data. This was the first step in analyzing how useful each source was. Other objectives in the creation of this matrix were to provide us with a tool for tracking unfound data fields, and show how extensive an aggregated a database containing all sources would be.

3.1.5 Methodology for Finding Data Sources

After determining what data fields we were looking for in our search process, the focus changed to refining a methodology for finding data sources. In doing this, we wanted to determine a complete and precise process that could be replicated in the future by Quincy Mutual. The first step of this process was to create the list of data fields of interest, which is explained in detail in the aforementioned sections. The second step was to choose what search criteria we would filter on. If we tentatively found sources, we would move on to the next step of the methodology. If not, then the search terms were refined and this second step would be repeated. The third step of this process was to determine the original data sources of the search

results from step two. By finding the core dataset of each source we find, we are better able to understand the stream of information in the private and public market. This also gives us more leads on potential sources to feed into our final data stream. Once we have where the original data comes from, we then contact the source directly. We created an email template to send to companies that included our goals for the project and how their data could play a role in our project and their future flow of business. Our background research on how this data could play into other industries was instrumental in this process.

The following final two steps were the most important to our process. The first of these was to determine a number of different metrics in order to assess the data sources. Through extensive research on what makes data "good" and "bad", we compiled a list of characteristics to rate each source on. These characteristics included who the intended audience was, how often it was refreshed, completeness, who owns and is generating the database, accuracy and consistency, validity and reliability, relevance, bias, limitations, and ease of access. In summation these characteristics approach all relevant angles to assessing data for the purpose of our project. With these different features in mind, we constructed written summaries of each web source for Quincy Mutual's reference which we will discuss further in findings. The final and most important piece was integrating these data sources into one cohesive database environment, referred to as a "comparator", which will be described more thoroughly in future sections.

3.1.6 Worcester Proof of Concept and Comparisons

We then produced a proof of concept by applying the sources we found to actual properties that Quincy Mutual had written a policy on. We selected twelve Worcester properties from a dataset of Loss Control Inspections that had a thorough Loss Control inspection completed and that varied in size and business type. This allowed us to see our capability to

produce information from third parties across all property classifications seen by Quincy Mutual. The information in the Loss Control Survey was considered accurate for comparison purposes because MFIA inspectors conduct thorough examinations and measurements of the property.

Each group member took three properties and populated a document with as much information from third party sources as possible. For these fields, we collected every value that each source provided, and tracked where each value stemmed from.

Two values that were very important to Quincy Mutual were square footage and replacement value. We wanted to determine which of our sources were more accurate when reporting these values by comparing them to the true numbers which we were getting for the Loss Control Survey. To represent the difference between each sources' square footage and replacement values from Quincy Mutual's data, we used relative error. To calculate relative error, we took the difference between the values from each source and the Loss Control Survey and then divided it by the Loss Control Survey number. This allowed us to compare the errors on a defined scale. If the relative error was 0, this represented a perfectly accurate number from the source. An error equal to 1 meant the data point from the source was 100% greater than the number in the Loss Control Survey.

After calculating the relative errors, we grouped the sources together to see how consistent they performed on the twelve properties in comparison to Quincy Mutual. To visually understand these values, we graphed them with rings representing different levels of closeness to accuracy.

Lastly, we took the twelve properties and created a property database where we listed the top data fields the underwriters desired and filled out the fields we were able to find while citing where the information was originating from. Expanding from this database, we turned our focus

towards another proof of concept production which was creating mock applications in the form of property profiles.

3.1.7 Full Property Profiles

After creating our Property-Source Database and analyzing the accuracy of the information we found, we then began creating a Property Profile for each individual address. We wanted to include more than the essential items such as square footage and replacement value. In this iteration, we mirrored the design of a Loss Control Survey using solely third party data sources.

We included every piece of reasonably useful information in these property profiles. Excluding only those fields we considered frivolous such as exterior paint color. We looked at any and all websites to find any information possible for these addresses. Things such as coastal distance or the nearest fire station were all things we sought to include in these profiles. Whether they were in a flood zone or a hurricane zone were also things we included. The goal was to mimic an application in the hopes that the information would be enough to accurately underwrite a property for Quincy Mutual.

These Property Profiles only displayed a single value for each data field. Our next task was then to figure out how to determine which source, or combination of sources, provided the most accurate option. Each data field is quantified differently, and therefore required individual attention. We examined each field on their own to create a guideline for choosing which source to choose or how to combine values from multiple sources.

3.2 Auxiliary Sources Process

The above methodology outlines our techniques for approaching the comprehensive and easily accessible sources of data. Also important to the completion of this project was finding the sources that provide valuable information about specific property fields or that offered their data in a significantly different format. The following section describes the process that we underwent to accomplish this.

3.2.1 LexisNexis, MarketStance, and Renaissance

Imagining that there must exist a storehouse for the massive amount of property data across multiple states, we sought a source that focused on commercial property on a large scale. There were three companies that Quincy Mutual either had some relationship with or had been approached by for business: LexisNexis, MarketStance, and Renaissance Alliance. We explored the possibility and benefit of integrating the services provided by these companies into Quincy Mutual's operations. Using the services of any of these companies would require investment on Quincy Mutual's part, so the benefit of their addition would need to be carefully considered.

While we performed a majority of the outreach to various companies, the initial connections for these three companies was already established by Quincy Mutual. Using this advantage, we planned to speak with each company regarding the various services and resources they offer. LexisNexis is a company that is already used by Quincy Mutual's Personal Lines department to supply important information like building valuation and occupancy. Renaissance Alliance connects insurance agents and companies in the New England region to provide better services to them collectively. As Quincy Mutual's agents are a part of this network, a further relationship with the company seemed natural. Alternatively, a third party data aggregator, MarketStance, approached Quincy Mutual previously for possible partnership so the team reached out to find out more about their products.

3.2.2 Finding Current Property Pictures

One of the most important data points Quincy Mutual's underwriting team showed interest in was relatively current images of properties. They urged that these pictures would be much more useful if they were clear, incorporated many angles, and were recently taken. They further emphasized that they could ultimately make their final decision to write the property or not with this sole piece of information in most cases. This led us to pursue a variety of potential sources that could provide property images, both paid and free, that we will discuss in Findings.

3.2.3 Additional Data Sources

After identifying some of the data fields that Quincy Mutual underwriters specifically requested, we looked into additional sources that could provide the fields that had not been previously found. Some of these included Distance to Coast, Flood Zone, Loss History, and Fire Protection Class. In no particular order, these fields were important items that are not always readily available to underwriters when they initially review a piece of New Business.

As the larger data aggregator websites had already been inspected for this information, searching for these fields was done on an individual basis. We took each field and looked into other industries that might find it useful. For example, flood zone may something recorded by government agencies for public safety. In a similar manner, fire protection class might be important to local fire departments, especially in more rural areas. These and other leads directed our searches for the various data pieces.

Before the onset of this project, Quincy Mutual completed an examination of their greatest causes of loss over the past seven years. These statistics also became a point of focus as

we sought to supplement the information currently available to underwriters. We created other fields that became a wish list during the overall data search. Some of these fields include Plumbing Quality, Fire Protection Devices, and Roof Quality.

Quincy Mutual is also interested in aspects of properties that relate to the behavior of the owner as well. This information is more difficult to uncover; however, if available it may significantly alter the underwriter's opinion of a property's risk.

Chapter 4: Results

4.1 Primary Data Sources

4.1.1 Individual Initial Research Results

Our initial research provided us with many of the data sources that later became key components of our proof of concept. Most of this initial data was derived from government sources because it was the most visible. The Massachusetts state government has a variety of databases available through their website. Another important discovery during this period was that realtor websites contain a lot of building information that is largely unrealized for insurance purposes. We did not focus on evaluating any sources at this point, only noting which seemed to have information relating to the project goals.

4.1.2 Big Spreadsheet with Data Fields and Sources

All data fields in the websites we found during our initial search were entered into a spreadsheet. Organization of this form allowed us to track all currently found fields as well as estimate the usefulness of the sources. Additionally, from this we identified instances in which multiple sources provided the same piece of information. For example, eight sources contained information about the year built and the square feet for each property.

A careful look at the data fields we were unable to find at this stage fed into the development of the survey sent to underwriters. We also divided all of the data fields into six distinct categories:

1. Expected Third Party - data that we hoped to obtain from third party sources

2. Obtained by Loss Control Survey - data fields that are present in loss control surveys that we did not expect to find in third party data

3. Unknown Source for Data Field - data fields that are determined internally by Quincy Mutual's programs or by the respective underwriter

4. Manual Input Expected - data fields that are chosen by the insured and their agent

5. New Fields for Quincy Mutual - data fields that Quincy Mutual does not collect but would find useful to their operations

6. New Fields found in Research - data fields that Quincy Mutual had not expressed known interest in but may find of use

This analysis showed us that the fields in the "Obtained by Loss Control Survey" and "Manual Input Expected" categories were unreasonable to collect or unobtainable from third party sources. This led us to remove almost all of the fields in these categories from our future research.

Data Field	Description	How Many Sources Contain This Field	worcesterma.gov	PropertyShark.com	homefacts.com	publicrecords.onlinesearches.com	Realtor.com	Netronline	Zillow.com
Miles from Fire Department		1			X				
Feet from Hydrant		0							
Construction Codes		0							
Construction Type	(Frame, Joisted Masonry, Non Combustible, Masonry Non Combustible, Modified Fire Resistive, Fire Resistive)	4	X			X	X		X
Sprinkler/Protective	Spinkler or alarm s	0							
Alarm System		1					X		
Year Built		8	X	X	X	X	X	X	X
Number of Stories		6	X	X	X	X	X		
Square Feet		8	X	X	X	X	X	X	X
Free Standing	Adjacent buildings	2							X
Number of apartments		6	X			X	X	X	X
Apartment/Restaurant class specific endorsement		1	X						
Apartment or restaurant		3	X				X	X	
Rating State		3	X				X	X	
Employee Count		0							
Number of Locations		1						X	
Insured Type		0							
FEIN	Federal Employer	0							
Billing Plan		0							
SIC Code	Business code	0							
Owner Occupied	Does owner live ir	0							
Price per Square Foot		6			X	X	X	X	X
ACV	Actual Cash Value	4					X	X	X

Figure 4: Data matrix to organize sources and determine what fields each provides

4.1.3 Outcome of Underwriter Survey

As a next step in our research, we wanted to determine what was most important to an underwriter when pricing a property. Through surveys and discussions, we narrowed our list of 55 data fields to make our continued research more efficient. The questions that we asked the underwriters were first sent in individual surveys, then discussed further in a focus group.

Our first question to the underwriters during our interview referred to a "wish list" of prospectively found data points they believed to be most crucial when underwriting a property. The main response to this question was that more pictures received before the Loss Control Survey is completed would be the most helpful. Second on their list was discovering loss history or prior claims, which can highlight obvious tenant or structural risk. Other information they wanted included property purchase dates and COPE fields.

Our second question was to identify the public third party data sources they have used in the past or currently rely on. We felt it would be a good idea to evaluate these sources against our own if we had not already found them. Due to the emphasis on pictures, Google Earth and Google Maps were among the top sources identified by the Underwriters. Other sources used include the Tax Assessor's database as well as other realtor websites such as Zillow.com that we had already taken into consideration.

In our third question, we asked if they noticed any among data sources for property information in different states. From our survey and discussion we found there were few significant differences between states. Many underwriters relied on the Tax Assessor's database, except in New York, which does not have an accessible Assessor's database.

Our final two questions asked underwriters to assign a numerical rating to the importance of the fields they already used, as well as the new fields we found through our research. Among

the existing fields, items such as construction, occupancy, building value, year built, and square footage were all important in underwriting a property. The new fields that were highly rated included roofing quality/condition, heating system type, and maintenance. Overall, our survey and discussion with the Underwriters ensured a positive projection for the remainder of our project.

4.1.4 Contacting Various Data Companies

We contacted all the sources we had found using the contact information listed on their website. We hoped to uncover from each their process for gathering information. Beyond this, we also wanted to begin a partnership between them and Quincy for the use of their information. All of the realtor oriented sites failed to respond after inquiring multiple times. To view the template for the email sent to each site, please see Appendix B.

We did have success in gaining access to public information through Worcester City Hall. After a short correspondence with the city's assessor department, we were provided a data dump of property data on over 35,000 addresses. This file will serve as an additional deliverable that a future intern will explore further.

We also established contact with Vision Government Solutions, the firm that specializes in collecting and maintaining the software for property assessments for the purposes of taxing for city government. They serve all of New England, New York, Pennsylvania, and Virginia. After discussing their roles in the data collection, we learned they were unable to provide us any information because the city and towns individually own the data. The best option for us to collect all the property data for these states is to contact each town and city's assessor as we did for Worcester.

4.1.5 Property Data Matrix

After the team had gathered ample amounts of data and inquired about what is important to an underwriter when writing a property, we felt the next step of the project was to bring this data to actual properties. Each member of the team took three properties in the Worcester area and collected data on each property in the form of a data matrix as discussed in our methodology. We wanted to take the information the underwriters identified in our surveys as the most important from only third party data sources. The properties we chose already had Loss Control data that was supplied to us from Quincy Mutual. Figure 5 contains a sample of one of the properties we collected data on:
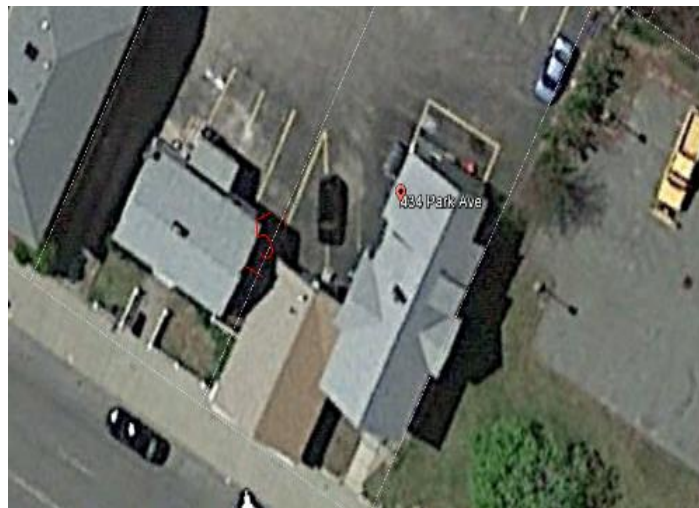
Source	Square Footage	Year Built	Construction	Occupancy	Building Value	Loss History	Roofing Material	Nearest Fire Department/Hydra	Heat Source
434 PARK AVE.									
Worcesterma.gov	4698	1930	Wood Frame	2 - Mixed Use/Reside	192200		Asph/Fbgl/Comp		
Zillow.com	3385	1930			234338		Asphalt		Heat Pump
Realtor.com	3385	1930	Wood Frame		195998		Asphalt		Heat Pump
Propertyshark.com	3385	1930							
Homefacts.com	3385	1930			194000	No Fire Incidents		2 Within 1 Mile	
Homes.com	3935	1930			568100				
Summary	4698	1930	Wood Frame	2 - Mixed Use/Resi	276927.2	No Fire Incide	Asphalt	2 Within 1 Mile	Heat Pump
Compare QM	2100	1910	Frame/Joisted Mas	Apartments	453450		Architectural Shingle	.1 Miles	Unknown/Gas

Figure 5: Example of one property in our Property Data Matrix. Blue shows chosen data fields, red shows a summary of sources, and green shows Quincy Mutual Loss Control Survey data for comparison purposes.

We chose a list of sources that appeared to have the data fields we identified as important in earlier stages of our research. Next we chose data fields (Blue) that were of importance after speaking with underwriters. We then searched each site and pulled all the information we were looking for and summarized it in the red. Lastly, as a comparator tool, we also included Quincy Mutual's Loss Control data highlighted in green. We felt confident in comparing our information with Loss Control data because that information was collected by a surveyor who was physically at the property. From what we found most of the data from third party sources was fairly consistent with the information from the Loss Control Survey.

4.1.6 Individual Property Profiles

After we pulled information about individual properties into our data matrix, we wanted to format our aggregate property data in a manner similar to a New Business Application or Loss Control Survey. The goal of this step was to determine if our third party data provided enough information to generate a quick quote or make an underwriter feel comfortable to bind a policy on the property. We wanted to include any and all information obtained using free, public data sources. We tracked of where our information came from by including a subscript number which referenced the website used for that field. For an example of a property profile see Appendix C.

In order to test these property profiles we ran them through the quick quote pricing system at Quincy Mutual. We took two properties and used solely the information gathered in the profile to see what quote we would get in comparison with what they correctly price the property at now. 17 Rockdale Street quoted $987 using our property profile and is currently quoted by Quincy mutual at $1,088, less than a 10% error. Lincoln Street quoted $3,401 from the profile and Quincy Mutual Quoted $2,799, a larger error around 20%. We found that for Lincoln Street our building value varied much more from source to source potentially throwing off our quoted number. The next step we felt was to narrow down how to get a more accurate building value as well as square footage but most other fields all seemed to align correctly.

4.1.7 Accuracy Graphs

Using the relative errors as discussed in the Methodology, we created accuracy bullseye graphs to better visualize the data. Using actual Quincy Mutual insured properties, each point on these graphs represent the errors in both square footage and replacement value. As the point

travels farther from the origin, the more our source's data varies from the Loss Control survey from Quincy Mutual.

As the point moves right on the x-axis, this represents the square footage value reported by the third party site is larger than the Loss Control number. Similarly, as the data point moves up vertically, the third party replacement value is higher than the value Quincy Mutual has on file for that property. Together the two errors give us a coordinates that are plotted.

In order to improve judgements of each source, we added accuracy rings in order to give perspective of the distances from the origin for each of the points. In green, we have a circle with radius 1 to represent a close estimate for both values. The yellow circle has a wider radius where values are possessing large errors that are not ideal for our use in the project. Lastly the red is any value that we cannot consider to be an adequate estimate.
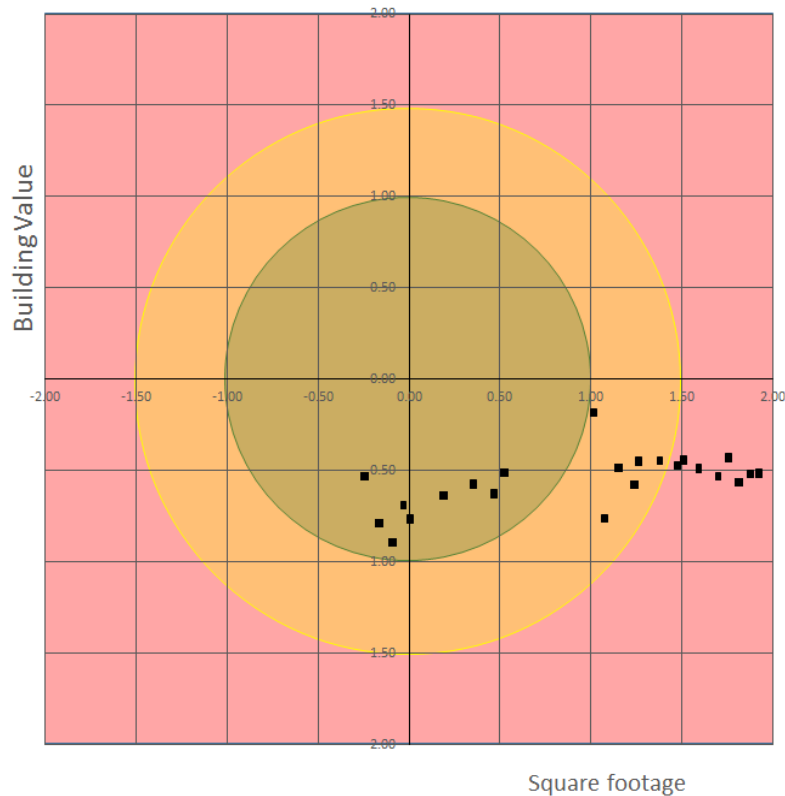

Figure 6: Relative Error Accuracy Graph

Figure 6 contains accuracy graphs for Worcester's public assessments. The graphs for all the other sources can be found in the Appendix D. The graphs all shared a similar shape with the building's replacement value underestimated by the source and the square footage varying randomly. We found the replacement value estimations are usually about half the valuation in the Loss Control because most sites only report the price to buy the property as opposed to the cost to rebuild if a fire were to completely destroy it which is what Quincy Mutual wants to know. To better estimate square footage will prove to be more difficult because many sources report living area or total area without listing which they are referring to. Another obstacle facing these two values is Quincy Mutual may only cover one office within a larger building so the sources we pull from deliver estimates for the entire building which represents more area and value than the policy is intended to cover.

4.1.8 Mock Premiums

Finally, we took all the information collected in the property profiles and inputted them into Quincy Mutual's QuickQuote system. This system is used by agents to get a rough idea of the premium price Quincy Mutual would quote a certain property. We entered the information we found for free, doubling the building values from the assessor data (due to the results of the accuracy graphs), and keeping the types of coverage, policy limits, and deductibles constant. We compared the quote we received to the actual premium amount Quincy Mutual was insuring the property for. Seen in Figure 7, we graphed our quote value vs the actual premium and a line representing where the two values are equal.
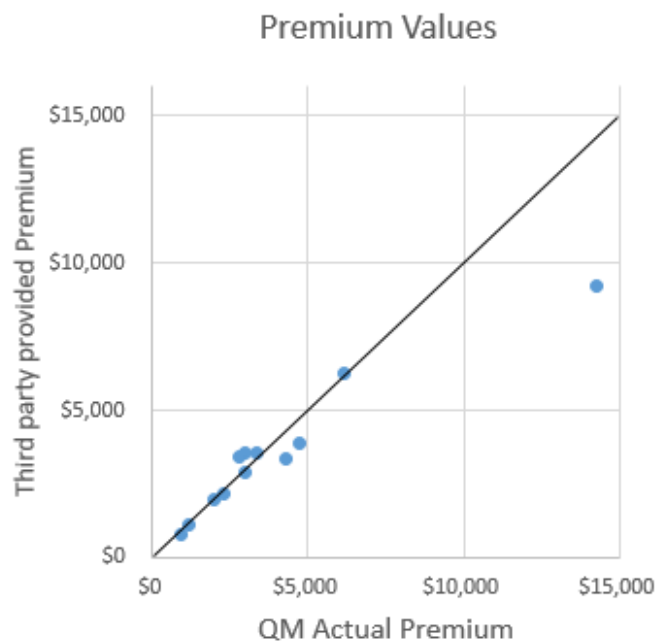
Figure 7: Graphed premiums from Quincy Mutual against the quotes we produced from third party

information

As you can see, our quotes come generally close to Quincy Mutual's values for premiums that are less than $4,000. When writing these smaller properties, Quincy Mutual can save time and money by using the sources we provided instead of waiting on the results of the Loss Control Survey. For larger properties, we would not recommend relying on the free sources only and instead still get an inspector to the property to ensure the values are accurate.

4.2 Auxiliary Data Sources

4.2.1 MarketStance

MarketStance's sales team had approached Quincy Mutual in the past on several occasions, and their business model seemed to align with the goals of this team's project. The company provides insurance companies with property information with special attention to the projected growth and decline of industries and locally segmented by units as small as zip code.

We arranged an interview and a presentation of their product tools and discussed with them the ideas we had for our project.

Unfortunately, MarketStance aggregates data on a higher level than we had hoped, most typically by census tract or zip code. This did not fit Quincy Mutual's vision of having information for each property by address level, 123 Main Street for example. So although MarketStance has a lot of great information, they did not fit the project's needs. In the future, Quincy Mutual may want to revisit MarketStance because their products offer great advantages for companies looking to expand into new regions and decide what industries to target in those specific areas. Their data wealth of loss history and business growth projections are great assets when looking to grow in areas Quincy Mutual has not written in before. To use MarketStance, Quincy Mutual will have to have less of a focus on specific addresses and instead prospect on a larger scale.

4.2.2 Renaissance Alliance and LexisNexis

Renaissance is a firm who is compiling data collected by insurance agents when they enter information into quotes. Their main goal is to have the foundation to begin having an entire database of properties in the Northeast with most of the essential information insurance companies desire already listed. Quincy Mutual had a series of meetings with Renaissance to discuss the development of this idea. Unfortunately, Renaissance is still in the early stages of formation and we were unable to meet with them.

4.2.3 Property Pictures

Finding current and clear pictures of properties was overwhelmingly a priority from the underwriting team. The first source we pursued was Google Earth because Quincy Mutual

already heavily used this service. It provides clear images from multiple angles as requested. However, the images are not as frequently updated as the underwriters would prefer. Properties may have seen a change in occupancy, significant additions, or a change in maintenance in the years since the most recent picture was taken. Despite this problem, this tool does have many advantages for an underwriter's use. The pro version allows the user to measure distances between buildings and determine the square footage of a building. It also shows images of rooftops, which are generally difficult to see. Through this research we also found out that Google Earth Pro, which was previously $399/year, is now free to download.

We then pursued sources that could provide pictures on a more frequent basis. The two main results that came from this was GigWalk and TaskRabbit. We reached out to both of these organizations to better understand their business and explain to them the purpose of our project. In our discussion with GigWalk, their representative explained that under their old business model, our project would have fit well into their mission. This mission was to allow the public to post their "gigs" as needed and in no restrictive manner. This business model was changed at the end of 2015, however. The current model, the representative explained, now focuses on larger customers with much higher demand for the uses of these "gigs". He explained that they are looking for work that involves 4-6 data points on a couple hundred locations, whereas our work would only be one data point on a small, varying number of locations. At this time there did not seem to be alignment between the visions of our two projects, but the contact remains there in the future if anything were to change.

When we learned we could not get this information from Gigwalk, we researched other companies offering similar services that may be able to assist. This led us to TaskRabbit. This

company was not looking for any potential new business prospects at this moment and were not interested in a discussion of integrating our work with theirs.

4.2.4 Behavioral Data

The main focus of this project involved information about the properties themselves. One additional analysis we wanted to do to supplement this was finding information about the people that live in or around these properties as well. This behavioral data would add a creative layer to our project that would allow Quincy Mutual to consider other factors in their rating. One of these main behavioral factors was smoking, which can be a serious hazard to both people and properties.

The first behavioral data source we looked into was the Americans For Non-Smokers' Rights database. This data set is creating using state and county-wide public records that are aggregated for a more complete picture of smoking laws. The purpose of this information is to show a number of different views and aspects of smoking data from county to county. This information includes which states and counties ban smoking in their restaurants and businesses. This group is a lobbying organization intended to inform and protect nonsmokers from environments that allow smoking, inherently exposing them to secondhand smoke. This data was last refreshed February 18nd, 2016 and seems to be refreshed monthly since 1986. It also seems that the data is comprehensive of the entire United States. Since they are pulling their data based on state law, there is very little error for incompleteness. The company owns and is generating the database, but they depend on representatives that do research in each state. Again, the accuracy and consistency as well as the validity and reliability are all excellent. This site is simply reporting state and local laws so there is not much error in this process. As far as relevance, this could directly affect the underwriting guidelines for businesses in these states.

This would be very effective information to have as a rater. Since this is a lobbying organization for non-smokers, they may be biased and misrepresent what is truly happening with smoking laws. However, if there is suspicion of bias, researching individually into state and local laws can solve this problem. There seem to be very little limitations for what this database wants to achieve, and it is very easy to work with. There are a number of helpful views, both in graphic and list form, of various smoker data. Finally, and most importantly, this data could be integrated very effectively into the workflow of Quincy Mutual. Since this data can be broken down to county level, and Quincy Mutual has these corresponding counties, it could be very easily aligned.

Similar to this source, we looked into the Behavioral Risk Factor Surveillance System (BFRSS) run by the Center for Disease Control. This is the "the nation's premier system of health-related telephone surveys that collect state data about U.S. residents regarding their health-related risk behaviors, chronic health conditions, and use of preventive services. Although this operation was wide in scope, we focused our efforts on the smoking aspect of their data. We found that most of their data is at county level. However, there is information showing the percentage of individuals that smoke in each town relative to the rest of the state. If we can ultimately provide Quincy Mutual with a source that outlines which town have higher rates of smoking, they may be able to adjust their rates to reflect the added risk of the policyholders in these areas.

The final behavioral data source we looked was Yelp. Our focus shifted on this source because we wanted to better understand the environment of the businesses that Quincy Mutual writes or could potentially write. We wanted to find a way to tap into Yelp's database, not just on a page by page basis where information would have to be manually found. The idea was for us to

filter on search terms in the reviews such as "non-smoker", "clean", "safe" or other qualities Quincy Mutual would deem a property to become a safer risk. Our main lead was Yelp's dataset challenge. This allows outside parties to download Yelp's data and enter a competition to complete something innovative in return. We ran into two issues with this. The data provided was not easily compatible with basic software, so we did not have the technology necessary to analyze this data. Furthermore, we did not meet the qualifications to enter this competition nor did the data match as well as we would have liked for our project. Although we would not get full access to Yelp's data the way we had hoped, there is still value in the information on this site and other commercial business rating sites that Quincy Mutual should continue to consult on an individual basis.

4.2.5 Other Data fields

Of the data fields that Quincy Mutual underwriters suggested, Loss History and Fire Protection Class were two that we placed priority in finding. Loss History was a piece of information that was only obtainable through the agent of the insured. The only source that we were able to find indications of loss on a property was Homefacts.com. While the completeness of the information is most likely up to the discretion of the realtor, it does provide some basic details surrounding damaging incidents.

Additional data fields Quincy Mutual was interested in was flood zone information including distance to the coast. We found that the government provides a wealth of information on this topic. As part of the National Flood Insurance Program, FEMA has developed and maintained a database of public flood data. As explained on their website, their goal is: "Through FEMA's flood hazard mapping program, Risk Mapping, Assessment and Planning (MAP), FEMA identifies flood hazards, assesses flood risks and partners with states and communities to

provide accurate flood hazard and risk data to guide them to mitigation actions." These maps are at a state, county, and town level and are available on a user-friendly interface where you can filter based on your specific search criteria.

The most interactive source and the source that will provide the most supplemental data to Quincy Mutual, however, is OLIVER, Massachusetts' GIS online mapping tool. This tool is a visual aggregation of a number of data files publicly available through governmental sources. Some of the information that is mapped includes: census data, coastal and marine features, conservation information, cultural resources, environmental monitoring, infrastructure, physical resources, and regulated areas. The information most useful to our project is mainly FEMA's flood data, infrastructure information about local police stations, fire stations, schools, prisons, data on coastal zones, as well as population densities. When the team was looking for flood zone data we found this tool, and immediately recognized its value. Although valuable, this interface is not yet available in other states Quincy Mutual writes in. Nonetheless, this information is all from public and government sources, it is just not aggregate in as nice of a way as OLIVER has done.

Chapter 5: Recommendations and Conclusions

5.1 Public Data

Reviewing the sources of the private companies, they mainly pull from public data so we recommend using the Assessor data from the cities and towns. This data is very extensive but can vary property to property. A majority of the information provided by the Loss Control Surveys can be found on the public assessor's report. Unfortunately, we have found that for the Worcester assessors' data can provide significantly different than the values provided to us by Quincy Mutual.

We found square footage varies randomly due in part to the fact Quincy may only write one floor or part of the building where the assessors are reporting the total living area. The other significant value that the underwriters were interested in was replacement value. Unfortunately we found that the assessed replacement value was around half of the cost Quincy Mutual had documented. With this in mind, we recommend that Quincy Mutual doubles this replacement value and uses it as a rough estimate.

5.2 Private Data

Pursuing a relationship with property websites that specialize in real estate would give Quincy Mutual access to square footage, property values, and other desirable information. Unfortunately, after several attempts, the team was unable to get in contact with any of these companies and help them realize the potential of their data. Another downfall of this data is the inaccuracy as we found in the accuracy graphs discussed in Chapter 4.1.7.

Despite these shortcomings, HomeFacts is a great source providing data such as square footage, building value, year built as well as a history of loss. Although the amount of data varies

property to property, HomeFacts should be checked by underwriters because it provides the same types of information as other realty sites but also includes information about the neighborhood, distance to fire stations, and the ever important loss history.

5.3 Behavioral Data

We recommend that the behavioral data found by our team throughout this project be utilized to supplement typical rating factors. Specifically, we believe this will help Quincy Mutual to understand smoking risk of the areas and business they write. With websites like the Americans for Non-Smokers' Rights database, they can incorporate smoker data at zip code or county level. This will provide valuable underwriting knowledge that is easily compatible to their workflow and rating process. Additionally, Behavioral Risk Factor Surveillance System (BFRSS) should continue to be reference in partnership with the first smoker database. Lastly, we recommend Quincy Mutual continue to pursue information from Yelp when insuring businesses and restaurants. This provides valuable information that fits with the "gutalytics" mindset of commercial underwriting. Ultimately, finding a way to access their entire database is the end goal.

5.4 Data Scraping and Data Use

Due to the legal ramifications of scraping most publically available websites, the technique should not be used for the collection of data. Instead, the company providing the data should be approached for access to the complete dataset. As a final resort, the data should be gathered on an individual property basis.

For all publically available information, the Terms of Use of this data should be examined to ensure that its use by Quincy Mutual will not violate this. The details in a Terms of Use can vary immensely between cities for public data and companies for privately held data.

5.5 Renaissance and LexisNexis

The potential partnerships with Renaissance and LexisNexis seem very promising in accomplishing the task of reducing the need of Loss Control Surveys. Unfortunately, the groundwork for these projects and firms were still being developed when we were finishing up our project. We recommend staying in touch with Renaissance and help bring this database of properties from the agents into the workflow as soon as possible to help prospectively write more properties.

5.6 Aggregate Data with MarketStance

If Quincy Mutual desires prospective, model-built property data in the future, they should consider MarketStance. While this company does not accurately capture the details of a specific property, they are able to forecast various growth and valuation metrics for properties in an area. Quincy Mutual could examine projections of growth among specific business classes as they look to grow their book of business among these classes. Quincy Mutual might find this information appealing as they expand to new towns and cities in New England.

5.7 Gigwalk

We recommend that Quincy Mutual follow up with the correspondents at Gigwalk. Our contact throughout this process was Cameron Fillmore (cameron@gigwalk.com). Their business services, although not currently in alignment with Quincy Mutual's, could serve an important

role in the future of getting up to date property pictures for underwriting. Following up with them to understand what steps Quincy Mutual needs to take to bridge their business plans is a vital role. GigWalk is looking for high-volume deals so Quincy Mutual must find a way to accommodate this need. We further recommend that a member of the Quincy Mutual Commercial Lines executive team reach out to Gigwalk.

5.7 Provide Underwriters with Source List

In order to ensure all underwriters are aware of the resources available to them, they should be provided with our document containing a list of all useful sources. This contains the sources that we found useful alongside the purpose that each source serves. Rather than a structured approach, this document will merely act as an easy-to-use reference guide. A copy of this source list can be found in Appendix E.

5.8 Our Structure for Property Profiles

We found that our Property Profile's served as a great tool in underwriting a property. These profiles mimicked that of a New Business Application or a Loss Control survey in an abbreviated form. As a test of these profiles we used Quincy Mutual's quick quote system to see if we had enough information as well as if the information was accurate to what Quincy Mutual had previously quoted them at. Although there are some questions on accuracy we were able to fill in all the necessary fields with the ones we provided on the Property Profile from free third party data sources. A brief discussion of each data source we found can be found in Appendix F.

In creating these profiles we also tracked which sources we used the most. For pictures, we relied heavily on Google Earth Pro which provided us with aerial pictures, frontal views, as well as tools to find the distance between two buildings. For the majority of property information

such as square footage, building value, replacement cost we resorted to the city assessor's database. Lastly another source we relied heavily on was homefacts.com. This website provided us with loss history, nearest fire/police stations, crime rate in that area and flood/hurricane risks. Of the many sources we researched these three seemed to stand out the most in creating our Property Profiles. Some next steps to improve this process we foresee would be narrowing down the accuracy of square footage as well as replacement cost on a property.

5.9 Conclusion

As a team, we hoped to find public and private data to help Quincy Mutual reduce the reliance on Loss Control Survey in the underwriting process of Commercial Lines Property. The surveys are performed thoroughly by an inspector, but that comes at a cost of time and money. We sought to find the information found on these surveys that underwriters or operations would be able to find on their own, rather than wait for inspection results. Unfortunately, we found that data varies wildly from both private and public sources so eliminating the detail and accuracy provided by the inspector would result in writing on bad information.

If Quincy Mutual feels it is time to move away from Loss Control Surveys, we recommend surrendering some of the fine detail and partnering with a company like MarketStance who can provide accurate and very useful data, just not at address level. If Quincy Mutual can use their tools and information to identify profitable industries in specific areas, they will be able to write more in less time.

If Quincy Mutual wishes to continue to writing properties they feel are strongly good risks, we recommend they continue to issue inspections as well as continue the new process of holding onto those surveys for future use. Preliminary checks can be performed by using sources

such as the assessor's data, HomeFacts, and Google Pro, but to ensure this data is accurate and the quote is appropriate we recommend continuing the inspections because of how often the values of our sources vary from the surveys.

Bibliography

About MFIA N.E. (n.d.). Retrieved September/October, 2015, from
    http://www.mfiane.com/about.htm

Agarwal, S., & Hauswald, R. (2010, June 16). Distance and Private Information in Lending.
    Retrieved October/November, 2015, from http://ushakrisna.com/2757.pdf

Assessor and Property Tax Records Resources. (2016). Retrieved March 02, 2016, from
    http://publicrecords.onlinesearches.com/Assessor-and-Property-Tax-Records.htm

Carlson, R. H. (2004, September 1). A Brief History of Property Tax. Retrieved
    September/October, 2015, from
    http://www.iaao.org/uploads/a_brief_history_of_property_tax.pdf

DeSanctis, A. (2014, November 11). Big Data Will Play an Important Role in the Future of Real
    Estate and for Realtors®. Retrieved September/October, 2015, from
    http://www.realtor.org/news-releases/2014/11/big-data-will-play-an-important-role-in-
    the-future-of-real-estate-and-for-realtors

Division of Taxation Property Administration. (2014, September). How Property Is Valued For
    Property Tax Purposes. Retrieved September/October, 2015, from
    http://www.state.nj.us/treasury/taxation/pdf/lpt/ptassessment.pdf

History of Homes.com. (n.d.). Retrieved September/October, 2015, from
    http://connect.homes.com/about/

Northern Michigan State University. (2009). Evaluating Internet Sources. Retrieved
    September/October, 2015, from http://library.nmu.edu/guides/userguides/webeval.htm

Stone, B. (2013, March 7). Why Redfin, Zillow, and Trulia Haven't Killed Off Real Estate
    Brokers. Retrieved September/October, 2015, from
    http://www.bloomberg.com/bw/articles/2013-03-07/why-redfin-zillow-and-trulia-havent-
    killed-off-real-estate-brokers

Welcome to the Quincy Mutual Group. (2014). Retrieved September/October, 2015, from
    https://www.quincymutual.com/about-quincy.htm

Appendix A: Data Fields

Data fields inquired about in the survey to Quincy Mutual underwriters:

- Plumbing information - quality and year
- Roof - geometry, use (if flat)
- Roof - Material
- Roof - Quality/condition
- Siding - Type
- Siding - Condition
- Windows - Type
- Windows - Condition
- Heating system - Type
- Heating System - Maintenance
- Space heaters - existence of
- Fire suppression systems - maintenance, type
- Staircases - condition, quality
- Handrails - condition
- Snow/ice removal - frequency, conditions required, impact
- Floor Type
- Area crime history and reports
- Theft and vandalism probability
- Tax year
- Tax Amount
- Land Value
- Building Value
- Color
- Sidewalk ownership/type
- Building Style
- Existence of Fireplace
- Fire Department Response Time
- Garage
- Business Name
- Photos - Exterior
- Photos - Interior
- Hurricane Risk

Appendix B: Email Template

Hello <u>Source</u>,

My name is _____ and I am a student at Worcester Polytechnic Institute. Myself and three other students are working on a project for a local insurance company. The goal of our project is to compile property data on buildings in the Worcester area from public and third party sources to help them better understand properties before writing an insurance policy. Upon on early research, we have found that <u>Source</u> has a wealth of information and we were hoping to schedule an interview to discuss the techniques you've used to compile this data and for a possible sample of the database to better understand all the fields <u>Source</u> has to offer. It is possible after our analysis that this insurance company may try to arrange a deal to gain permanent access to the information.

Anything will help!

Thank you and we look forward to hearing from you,

Appendix C: Property Profile Example

Address: <u>434 Park Avenue, Worcester, MA. 01610</u>
Pictures:
Aerial View[1]

Front View[1]

Owner: <u>Winchedon Park LLC</u>[2]
Last Sold: <u>August 2013</u>[3]

Distance to nearest building: <u>5 Ft.</u>[1]____ Building Value: <u>$276,927</u>[7]   Replacement Cost: <u>$272,569</u>[2]_____ Year Built: <u>1930</u>[2]
Total Square Footage: <u>4698 sq. ft.</u>[2]   Livable Area: <u>3385 sq. ft.</u>[2]
Occupancy: <u>Mixed Use/Residential</u>[5]
Number of units: <u>1</u>[2]_____ Stories: <u>2</u>[2]
Loss History: <u>3 gas leak and 2 carbon monoxide investigations totaling $0 loss</u>[5]
Construction Type: <u>Wood Frame</u>[6]
Exterior Wall 1: <u>Aluminum/Vinyl</u>[2]____ Exterior Wall 2: <u>Brick/Stone</u>[2]
Roofing Material: <u>Asphalt___</u>[2]_____ Roofing Structure: <u>Gable/Hip</u>[2]

Heating Source: <u>Heat Pump</u>[2]
Air Conditioning: <u>Cooled</u>[2]
Nearby Fire Station: <u>2 Within 1 Mile</u>[5]

Nearby Police Station: <u>0 Within 1 Mile</u>[5]
Nearest Coast/Body of Water: <u>38.30 Miles</u>[4]

Other Information:
Interior Floor Quality: <u>Average</u>[2]
Crime in Area: <u>Moderate</u>[5]
Exterior Quality: <u>Average</u>[2]
Interior Quality: <u>Average</u>[2]
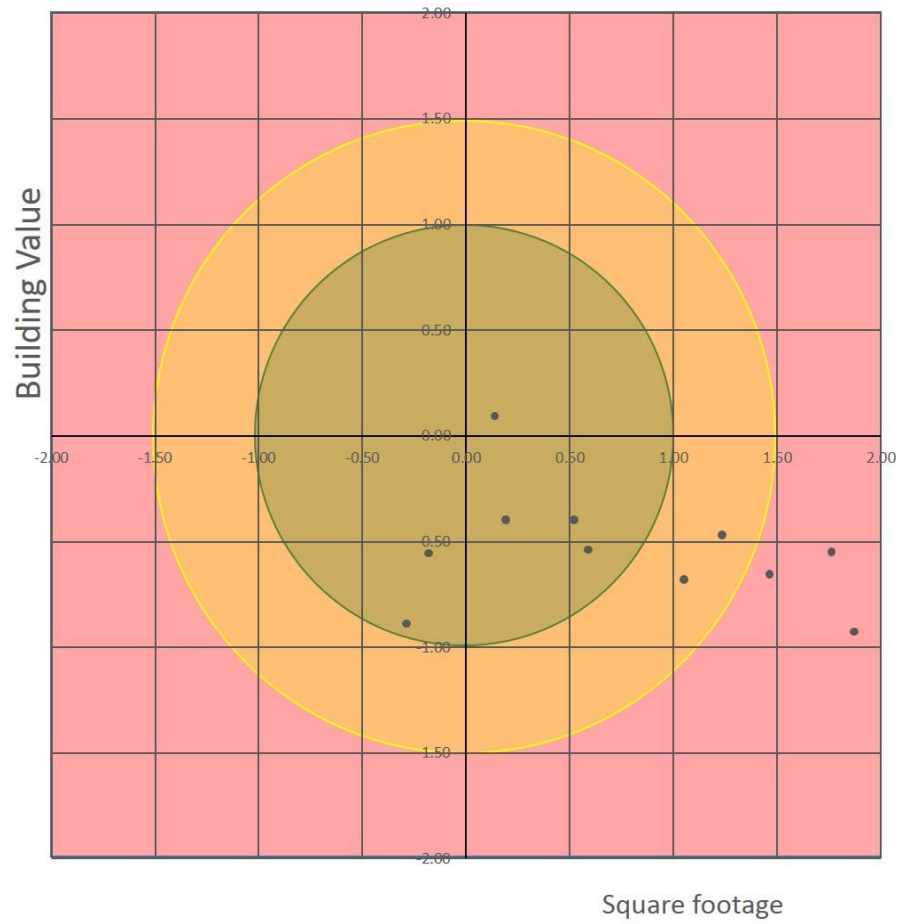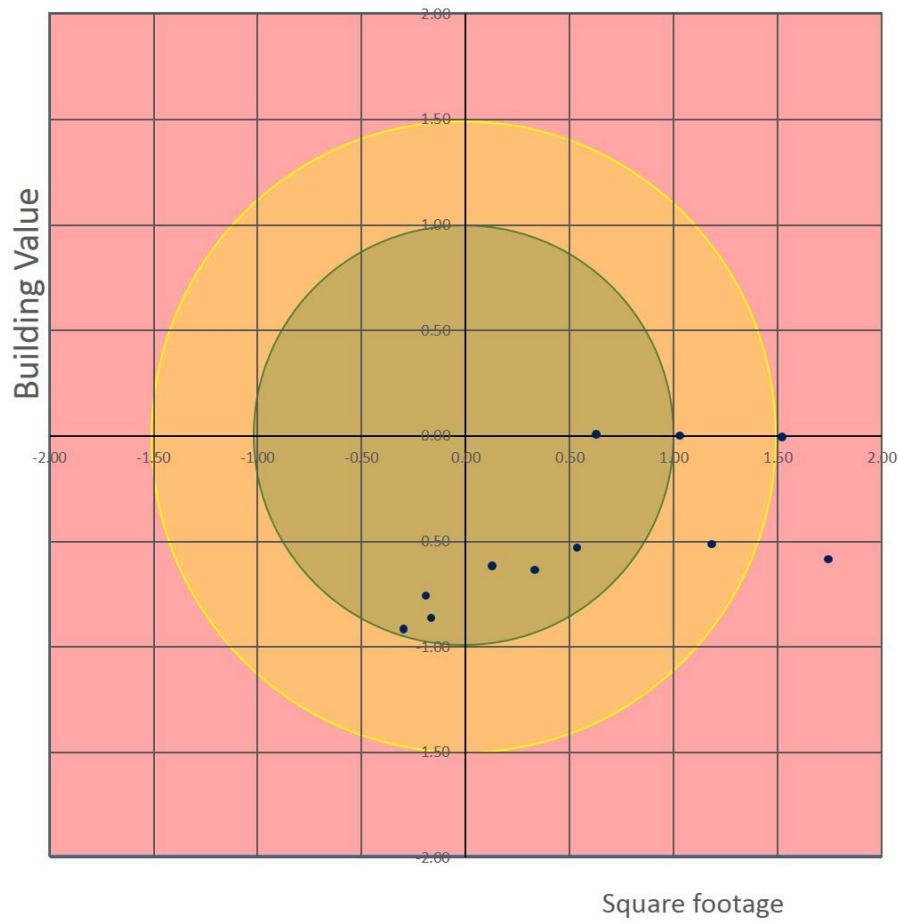Bath/Plumbing Quality: <u>Average</u>[2]
Flood Risk: <u>Low</u>[5]
Fireplace: <u>None</u>[6]

1 - Google Earth; 2 - WorcesterMA.gov; 3 - Homes.com; 4 - MapRisk.com; 5 - HomeFacts.com;
6 - Realtor.com; 7 - Average of WorcesterMA.gov, Homes.com, Zillow.com, Realtor.com,
PropertyShark.com, HomeFacts.com.
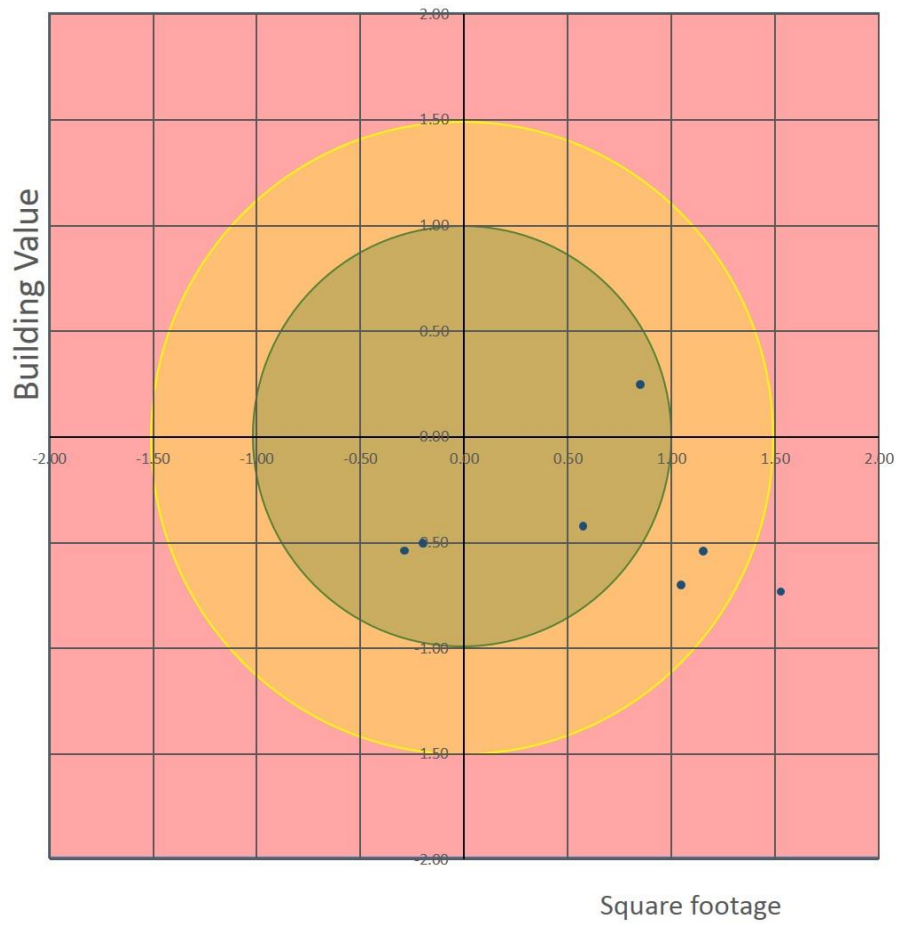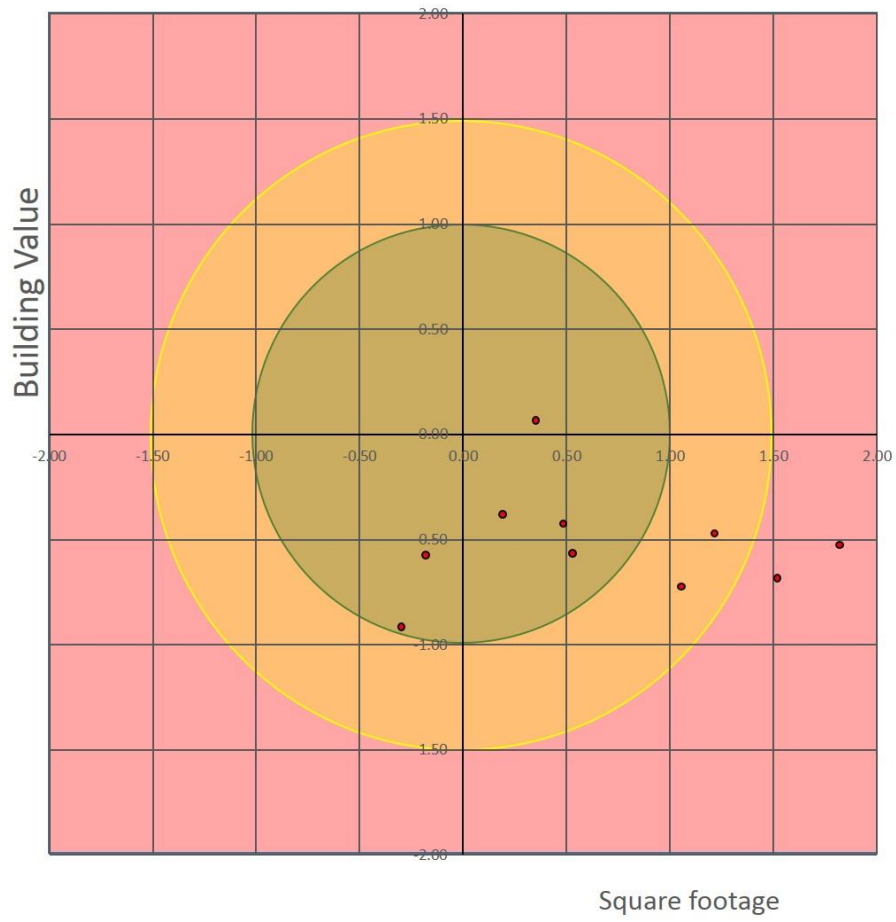
Appendix D: Accuracy Graphs

Square footage

Building Value
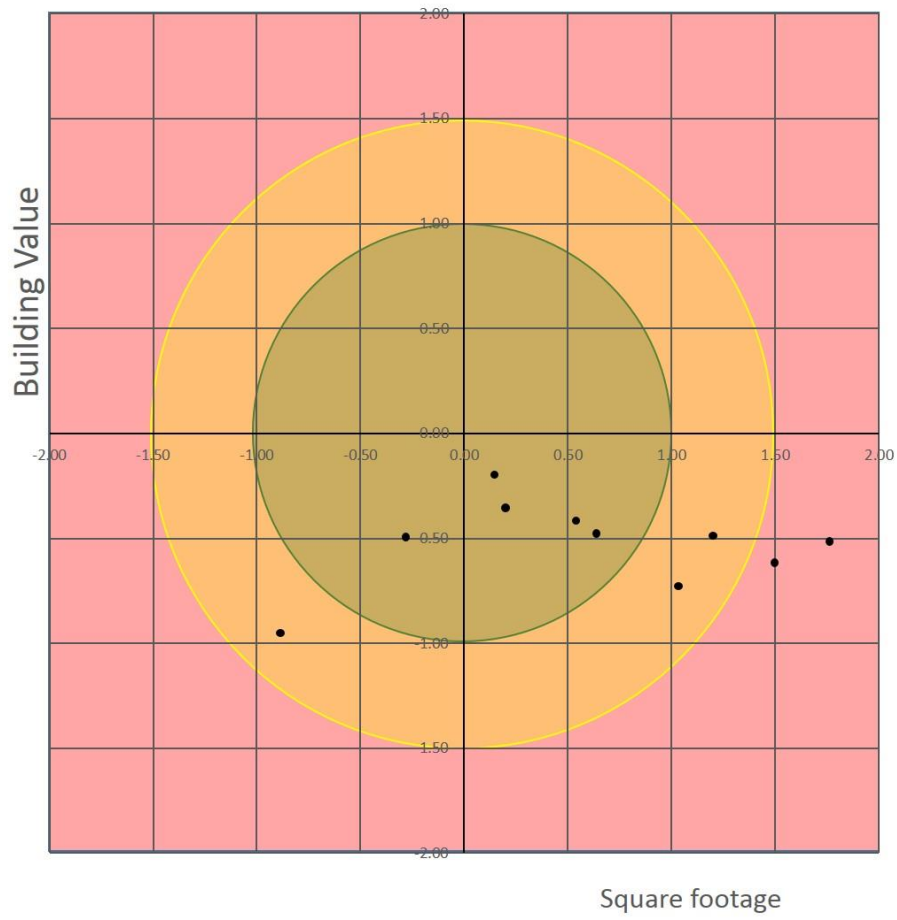
Square footage

Homes.com

Building Value

Square footage

Realtor.com

Building Value

Square footage

Zillow.com

Square footage

Appendix E: Important Source List

Source	Fields

Google Earth Pro	• Pictures
	    ◦ Aerial
	    ◦ Frontal
	• Distance between buildings

City Assessor Data	• COPE • Building valuation • Square footage • Ownership records
Maprisk.com	• Coastal Distance
Homefacts.com	• Nearest Fire/Police Station • Crime Rate • Flood Risk • Hurricane Risk
Oliver GIS	• Flood Zones
Americans for Non-Smokers' Rights	• Smoking laws by county
Behavioral Risk Factor Surveillance System (BRFSS)	• Smoking behavior of individuals in a town relative to the state

Appendix F: Data Source Write-ups

Americans for Non-Smokers' Rights

This data set is creating using state and county-wide public records that are aggregated for a more complete picture of smoking laws. The purpose of this information is to show a number of different views and aspects of smoking data from county to county. This information includes which states and counties ban smoking in their restaurants and businesses. This group is a lobbying organization intended to inform and protect nonsmokers from environments that allow smoking, inherently exposing them to secondhand smoke. This data was last refreshed October 2nd, 2015 and seems to be refreshed monthly since 1986. It also seems that the data is comprehensive of the entire United States. Since they are pulling their data based on state law, there is very little error for incompleteness. The company owns and is generating the database, but they depend on representatives that do research in each state. Again, the accuracy and consistency as well as the validity and reliability are all excellent. This site is simply reporting state and local laws so there is not much error in this process. As far as relevance, this could directly affect the underwriting guidelines for businesses in these states. This would be very effective information to have as a rater. Since this is a lobbying organization for non-smokers, they may be biased and misrepresent what is truly happening with smoking laws. However, if there is suspicion of bias, researching individually into state and local laws can solve this problem. There seem to be very little limitations for what this database wants to achieve, and it is very easy to work with. There are a number of helpful views, both in graphic and list form, of various smoker data. Finally, and most importantly, this data could be integrated very effectively into the workflow of Quincy Mutual. Since this data can be broken down to county level, and Quincy Mutual has these corresponding counties, it could be very easily aligned.

Homefacts.com

Home Facts is a data aggregator that targets those who are home buying or selling, as well as the real estate professional. The site is owned and maintained by Hidden Rocks LLC who assume no responsibility for updating or validating the accuracy of the information on the site. They rely on their users and the public to find errors and report them to be fixed by the site's administrators.

The information Home Facts is very extensive in perspective of their target customers but lacks a lot of the key data our group is interested in. A lot of basic property data is listed such as size of the property, estimated value, year built as well as number of stories, bedrooms, and baths. One of the best pieces of information Home Facts provides is history of incidents (fire or theft) which can give an underwriter an idea of the potential risks may be present when writing a policy for this property. The site also lists the history of earthquake and severe weather such as tornados in the area of the property that further provides insight in risks of loss for that policy.

This source of data has some negatives with respect to our project. As mentioned above, the target of the information is for home buyers and sellers meaning it lacks a lot of the information we are looking for. There is a strong emphasis on the benefits and drawbacks of the area including local schools, unemployment, and political views of the area.

Homes.com

Homes.com contains information for those seeking a new home and for real estate agents. The website has only residential properties and only those that are currently being listed. Some of the fields include basic information such as size, number of rooms, and value of the property.

A large portion of the website is focused on calculators for monthly and initial payments. It is noted that incomplete information can be manually entered by the user, which is not validated by Homes.com. The data is stated to be generally reliable but not always accurate. All data is owned by Dominion Enterprises and is gathered and compiled by two services called SmartZip Analytics and Fast Forward, Inc. The data management clauses in their Terms of Use specify that any programmed data collection for commercial use is strictly prohibited. Additionally, much of their data appears to be gathered from other realty websites. While the information available on this website is free, it does not appear to be highly valuable, as the same data has been found through other sources more reliably.

Netronline.com

Netronline.com provides direction to a variety of government websites through links at no charge. Additionally, they offer seven different products for real estate use that include various levels of information about a property. The most useful of these products is a "Property Detail Report" that includes the size, value, and occupancy information for the property. While this information is available for only $3.50, it is fairly easily obtainable by individual means. Also, it does not appear that the data in these reports is always complete, and may be out of date in many cases. The website used to access these reports is less user-friendly than most other websites viewed thus far, and looks out of date. All information that Netronline offers through their various products has been gathered through a variety of government sources. This may be a useful tool as a comparator for information taken from other websites, but should be considered after other alternatives. Netronline's primary mission and product seems to be copies of legal

documents and ownership histories, which lie outside what is important for Quincy Mutual's business needs.

PropertyShark.com

PropertyShark.com is a website that caters primarily to the real estate industry, including both real estate companies and potential home buyers. Basic housing characteristics are listed including area, title history, tax information, and occupancy information. Their primary tool provides a valuation for each property within a confidence interval. The website offers an easy-to-use search field that can direct the user to either an individual property or street. Accessing the full range of information that the site has requires a subscription plan. Multiple payment levels are available depending on the desired level of access. The commercial package allows the user to pull data from the website with specific characteristics into Excel. All data gathered for the entirety of the website is taken from government sources, and is therefore only as accurate or as current as these sources. PropertyShark does no work to ensure the validity of any of the information on their website. Additionally, not all properties within a city are listed. It does have both habitational and non-habitational businesses, and includes even those properties that are not currently for sale.

Realtor.com

Realtor.com appears to be another top competitor in our search for property information. Of the 57 fields in the New Business Application realtors.com contains 11 of those fields which is the most out of all other sights. In terms of our additional fields we think could be useful in underwriting a property this website contains 7 of the 32 fields. This source contains important

fields of information such as year build, alarm system, number of floors, and many more. Another useful tool it contains is that it breaks down the rooms in the building and tells what they are made of and where they are located in terms of floors. It tells about the heating and cooling systems as well as utilities and appliances. Realtors.com also incorporates a comparator of public records at the bottom which could be interesting to see how they collect their data for this as well as their other fields.

The data appears to be accurate as well as reliable since this is a very well-known sight. There is also a time stamp on each page stating when the property was listed as well as the last time it was refreshed with for most properties appears within the last few days and weeks. A downside is that only properties that are for sale appear on the website.

Worcester.gov

After contacting the assessment office in the Worcester city hall we were provided with a very large excel database of all the properties in the Worcester area. In this file there were 28 columns which could also include more useful data fields after further research. This was just a sample of what they had so there could be potential for even more data fields. Of the data fields we have extracted from the New Business Application we were able to find 9 of the 57 fields from these Worcester public records. These fields include construction type, year built, number of stories, square footage and a few others, all of which appear to be key pieces of information when underwriting a property.

We also proposed a list of new data fields we could see being useful that were not on the New Business application or Loss Control survey. Of this proposed list of 32 fields the worcesterma.gov website had 11 of them. The database appears very complete and from a

reliable source that holds no biased. The purpose of this data is mainly for public record of Worcester so it is not hindered to any type of audience. After briefly comparing items found in this database to information given to us from Quincy Mutual's Loss Control Survey with a few discrepancies.

Zillow.com

Zillow is a data aggregator that focuses on real estate, especially those that are on the market currently. Zillow is owned by Zillow Group Mortgage Inc. and rely on users to input accurate data of the properties they are selling and claim no responsibility to false information. Although the site does not allow data scraping it is not needed for Zillow provides aggregate data of the site's information at varying levels from state to neighborhood.

For buildings on the market, Zillow has a wealth of information that will benefit our project like the materials used to make the building, the property's value, square footage, and type of heating. As mentioned above, Zillow provides much of this data at an aggregate level for free meaning it will be a great resource for our project. Once we are able to narrow down the fields the underwriters find most useful, we can begin to really dig into this data even more.

Some barriers with Zillow's data is there are quite a bit of gaps, most likely from properties that have not been put on the market in a long time. It is also more focused on the home portion of real estate which means it may prove difficult to find data on commercial properties in their data.