Columbia University - Data Analytics & Visualization Bootcamp

Group Project - Real Estate Seller Lead Generation

Outline

- Overview
- Understanding the problems
- Project objective
- Target audience
- Sales Process
- Proposed Method
  - Examples of Hypotheses
  - Statistical Modelling
  - Integration and Productionization
  - Lead Testing and Feedback Loop
- Macro Drivers to consider
- Next Steps

Overview

Residential Real Agents depend on an active sales pipeline in order to make their sales goals. A key part of pipeline management is lead generation. Leads come from all sorts of channels, for example personal networks, referrals, phone desk duty, purchased leads (e.g., Zillow), outbound calls, direct mail, and digital marketing.

Broadly, there are two types of leads – Buyer leads and Seller leads. Buyer leads are easier to come by for reasons we'll explain later, but Seller leads – who is going to sell their property and when – are more elusive. This is where data analysis and predictive modeling come in.

The goal of this project is to build a data-driven ecosystem that generates Seller leads that can then be acted upon by the agent in order to build a strong sales pipeline.

Understanding the problems

01

A consistent source of quality leads is a challenge – Some of the best lead sources for agents are not always the most consistent as far as lead volume goes. For example, word-of-mouth, repeat business and referral networks usually produce the best leads, however, often there are simply not enough of them. And, while agents can undertake lots of activities to promote themselves within their network, it is a finite set of opportunities, especially for newer agents who are building a book-of-business. What agents need are lead sources that supplement their personal sphere, especially if the goal is to grow.

Understanding the problems

02

Sellers are more elusive than Buyers– Over 80% of home buyers use the internet to research, shop, compare and narrow down their choices, including those buyers who are actively working with an agent. This gives sites such as Zillow, Redfin, Trulia and Realtor.com an opportunity to capture information on who is using their sites and sell these buyer leads to agents. On the other hand, Sellers may use the internet to compare their home to others, but they do not “look for buyers” on the internet, unless it is for-sale-by-owner, a fraction of all listings. Other ways are needed to predict who might be selling a home, why they are selling, and when.

Understanding the problems

03

Multiple sources of data need to be connected – There are many potential sources of data that could be useful in predicting who might be selling their home – public records, mortgage data, tax records, census data, demographic data, the multi-listing service, neighborhood ratings, school-district scores, crime maps, api's for web services like Zillow, purchased data such as core logic, etc. The challenge is putting it all together to develop an ecosystem that predicts just who is thinking about selling and when they will do so. And, while there are players in this space, that is, those who develop and provide seller leads for a fee, the accuracy of those leads is a challenge and no one seems to have 'cracked the code'. One of the competitive advantages is having access to actual agents and a broker to add a layer of qualitative data, which we will have in the course of this project.

Project objective

This Photo by Unknown

The goal of this project is to develop high-quality residential real estate leads that predict which Home Owners are considering selling their home, their motivations for doing so, and the timing of their actions. In addition to using a data driven approach to develop the initial set of leads, the project will include a qualitative overlay, working with actual agents and the Long & Foster Real Estate, Inc brokerage. The geographic market for developing these leads is New Castle County, Delaware (pop. 550,000).

Target audience

The Home Owner Seller Leads will be made available to Residential Real Estate Agents, who are the target audience “consumers”. The agents will test the leads by contacting them through various channels, including calling, direct mail and digital contacts. Success will be measured by lead quality as rated by the agents, and how many of the Seller Leads convert from ‘prospect’ to ‘pending or active client’.

The Residential Real Estate Sales Cycle – Seller-side

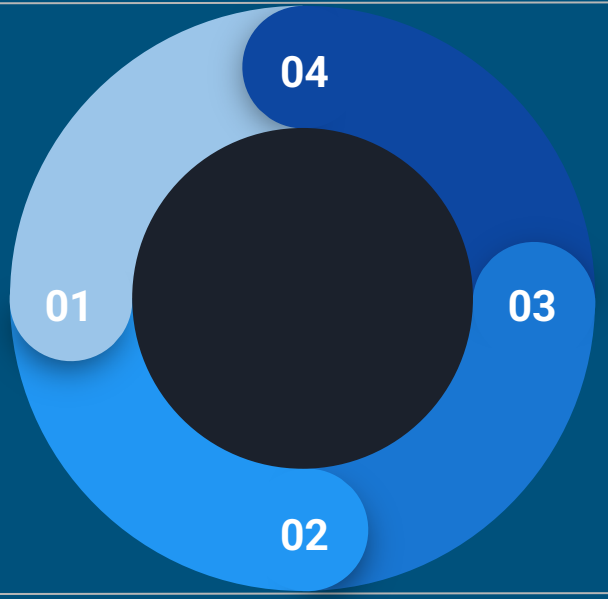
This project aims to generate quality leads in order to start the sales cycle

Lead Contact

Agent reaches out to lead and makes initial contact; His/her/their goal is to figure out which contacts are viable, and what the timing is

Getting the Listing

This is the stage where the agent makes his/her/their sales pitch; why the client should list with this agent, what are the advantages over other agents



Active Listing

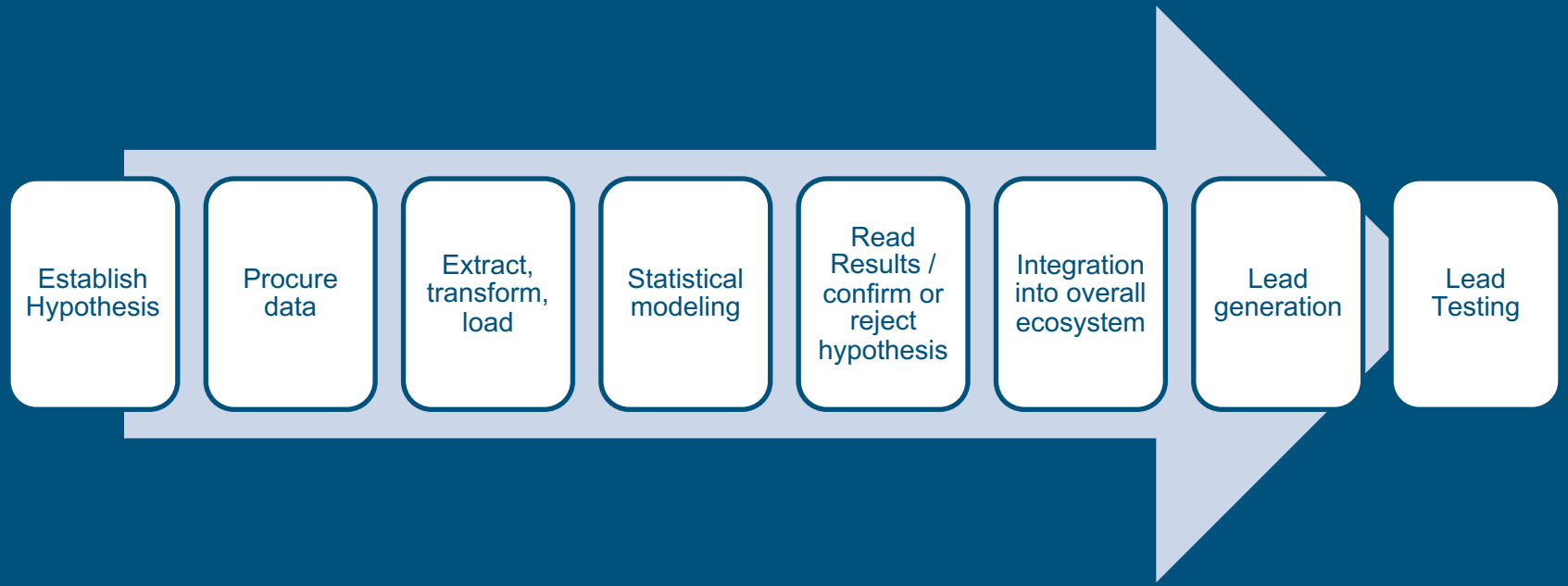
In this stage, the property is actively marketed to Buyers and their agents. Tours are conducted, and offers are received and presented.

Property Preparation, Marketing and Listing

In this stage, the listing agreement is signed, the property is prepared for market, a listing price is determined, and the marketing plan is defined.

Proposed Method

The process starts with making a sound hypothesis - what might motivate a home owner to sell a property – then uses a data and statistical approach to develop leads, which are subsequently tested in the real world by agents working the leads.



Examples of Hypotheses

Hypothesis	Data Sources / Considerations
Owners with children entering school will consider moving to get to a better school district; they must have equity in their current home in order to do so.	Demographics, birth records, mortgage information, school district ratings.
Owners at retirement age will consider selling to move to a retirement community or destination.	Demographics, regional migration patterns, current home size, availability and price of homes at destination, Zillow APIs
Changes in interest rates will motivate or inhibit owners to sell – both from a Buyer demand and purchase opportunity perspective.	Interest rate history and forecasts, correlated to past home sales; attempt to isolate effect
Single owners getting married will sell, respectively, then purchase a new home in joint tenancy	Demographics, marriage records, mortgage information
Changes in crime patterns will motivate owners to sell	Crime maps, historical trends, neighbor insights

Statistical Modelling

The goal of the statistical modelling is to see if there is a correlation between the hypothesis and home sales. The proposed method is to look at past sample data in order to determine whether a correlation exists for the population as a whole (i.e., determine if there is statistical significance – use of Pearson correlation, p-values and null hypothesis.)

The outcome will be enhanced by use of a judgemental overlay. Integration of multiple statistically significant factors will need to be tested for redundancy in order to get a streamlined ecosystem model.

Integration and Productionization

Once the indicators that are statistically significant have been identified, they will need to be integrated into the ecosystem and productionized. This may involve machine learning. Certainly the data extraction, transformation and loading processes should be productionized. This should also be the goal of the statistical modelling component, as well as the production and ‘packaging’ of finished sales leads. If possible, an automated lead delivery system using a web portal would be ideal.

Lead Testing and Feedback Loop

Once leads have been developed, they need to be appended with contact information, and agents will need to contact the home owner in order to test them. This is a critical step. Key performance indicators (usually expressed as conversion rates) will need to be established and measured.

Agents will need to report back on the relative effectiveness of the leads. Differences in agent contact and sales approach will need to be factored into the overall results. Some contact methods will be better than others (e.g., outbound call versus direct mail), although there could be challenges with certain contact channels (e.g., availability of good phone numers, ‘do-not-call’ lists, and other constraints.)

Macro Drivers of Home Sales

There are a lot of factors that go into the market for home sales at any point in time. Real estate leads can not be measured in a vacuum. They need to be evaluated against the back drop of these factors, which include:

- Market inventory (number of homes for sale)
- Interest rates (currently at around 7%, a multi-decade high)
- Job market
- Recessionary factors

The project will need to consider these and other factors in measuring relative success.

What needs to happen next

This presentation is designed to see if there are any students interested in helping with the project. Here are some of the suggested next steps:

- Team formation
- Build out hypothesis
- Articulate approach
- Align the project plan with project goals and requirements of the class
- Begin the work

What needs to happen next

If you're interested in joining the project, shoot me a slack, email to text:

Vince

vjtrombetta@gmail.com

347-715-0256

I'm a licensed real estate agent in Delaware and affiliated with my broker Long & Foster Real Estate, Inc. I have access to public record, multi-listing service and other data sources. In addition, I am willing to consider purchasing some lead sources, especially api's from real estate sites like Zillow, as well as to use government data, free-commercial use data, and proprietary data from my broker.

Thank you!