

CSC 478 Final Project Proposal - Housing Rental Business Analysis

Group Members

Yan Yu, Wenyi Yan, Fangjing Zhan

Project Type

Application Development - Business Analytics

Subject Area

This is a business scenario about sales cycle of senior housing communities. The sales cycle of senior housing communities often starts with a referral partner recommending a community to a potential future resident (prospect). And then the community sales team of a housing company makes an effort to turn the prospect into a permanent resident of the community by different sales activities. When a prospect is first connected with the community, the sales team will identify the likelihood a prospect will become a permanent resident and within what time frame. The sales team uses generic terminology, Hot (move in within 30 days), Warm (move in within 90 days), and Cold (just browsing – no set move in date) to classify prospects and determine which sales activities are necessary. We would like to create a process using data provided by referral partners to automatically determine a prospects classification prior to the sales team's first meeting. This automation will help sales teams more efficiently staff and optimize sales activities.

Source of Data

The data is real-world data from a data analysis competition held by a housing management company, Senior Lifestyle. Senior Lifestyle operations over 180 senior housing rental communities across 28 states.

Datasets link:

https://drive.google.com/open?id=1JMm6_dN2Vvm0W-NwLJTccMeQqtRZ1fUD

Datasets Description

Dataset 1: Prospect Demographics. The first dataset encompasses basic demographic information and additional disclosed preferences a prospect submits to referral agencies when starting their search for a senior housing community and associated significant dates within the sales cycle process.

Dataset 2: Sales Activities: The second dataset is a full list of activities performed by the sales team members when trying to convert the prospect into a resident.

Dataset 3: Resident Dates: The third dataset displays the significant dates for prospects that have converted into residents.

Dataset 4: Revenue Data: The fourth dataset is a full list of services provided to residents during their entire tenure at the community. Services are complete with dates and amounts.

DataSet1: 178,478 Observations, 24 Variables		
Field	Type	Definition
CommunityCode	String	Three digit code to represent unique community location
CommunityZipCode	Numeric	US Zip code location associated with unique community location - can be referenced against prospect zip code to identify distance variable
ProspectID	Primary ID	Unique ID associated with prospect record
AgeatInquiry	Numeric	Age of prospect at time of inquiry
CareLevel	Categorical	Identifies care needs requested by prospect at time of inquiry
Gender	Categorical	Identifies prospect gender
Marital	Categorical	Identifies prospect marital status at time of inquiry
Veteran	Boolean	Reference identifying if the prospect is a veteran
CurrentResidence	Categorical	Reference identifying the living situation for the prospect at time of inquiry
ZipCode	Zip Code	US Zip code location associated with prospect at time of inquiry - can be referenced against community zip code to identify distance variable
UnitPreference	Categorical	Identifies housing preference of prospect at time of inquiry
ProvidedEmail	Boolean	Reference identifying if the prospect provided an email address for contact
ProvidedPhone	Boolean	Reference identifying if the prospect provided a phone number for contact
Budget	Categorical	Reference identifying financial range prospect willing to pay per month
RequestedVisit	Boolean	Reference identifying if the prospect requested a community visit at time of inquiry
NoCall	Boolean	Reference identifying if the prospect opted out of receiving phone calls
NoEmail	Boolean	Reference identifying if the prospect opted out of receiving emails
InquiryDate	Datetime	Date of first inquiry
ReferralCompany	Categorical	List of common referral companies associated with connecting prospects to communities
ProjectedMoveIn	Datetime	Date prospect was projected to move in at time of inquiry
DeactDate	Datetime	Date of prospect deactivation (no longer a working lead)
DeactReason	Categorical	References that identifies why a prospect record was deactivated
NotQual	Categorical	List of reasons a prospect is deactivated in the system when they don't meet the qualification standards of the community
CurrentStatus	Categorical	Current status of prospect in the database
DataSet2: 711, 844 Observations, 3 Variables		
Field	Type	Definition
ProspectID	Primary ID	Unique ID associated with prospect record
Activity	Categorical	Reference that identifies what activities the sales team initiates during the sales process
dtActivity	Datetime	Date of activity
DataSet3: 5, 738 Observations, 6 Variables		

Field	Type	Definition
ProspectID	Foreign ID	Unique ID associated with prospect record
ResidentID	Primary ID	Unique ID associated with resident record
dtmovein	Date	Date prospect moved into community (becoming resident)
dtmoveout	Date	Date prospect moved out of community
istatus	Categorical	Current status of resident
ireason	Categorical	Reference that identifies why a resident moved out of the community
DataSet4: 23, 123 Observations, 6 Variables		
Field	Type	Definition
ResidentID	Primary ID	Unique ID associated with resident record
ServiceType	Categorical	Reference that identifies the types of services a resident uses. More than one service can be associated with a resident.
RateType	Categorical	Reference that identifies if the service amount is billed at a monthly or daily rate
FromDate	Date	Date service started
ToDate	Date	Date service ended, if null - service is still active
Amount	Numerical	Amount of service

Business Goal

1. Provide and analyze the following business information:

- Churn Rate - The proportion of contractual customers or subscribers who leave a supplier during a given time period.
- Determine best combination of actions to maximize conversions and reduce labor activities
- Resident Lifetime Value - A prediction of the net profit attributed to the entire future relationship with a customer

2. Using the prospect data, create different models and choose the best one to predict the following:

- Likelihood a prospect will convert and estimated time frame to convert
- Group prospects into three buckets (hot/warm/cold) based on typical sales cycles by care level.
 - Hot for the quickest conversions, Cold for least likely conversions

Method

For technical skills, we will apply the following (not limited):

- K nearest neighbors
- Logistic regression
- Decision tree
- Cross Validation