“Quick” Sales of Condominiums in Arlington, VA

Chris Frohlich
Darik Gossa
James Haas
Roger Moncarz
Jeff Robinson

ROBERT H. SMITH SCHOOL OF BUSINESS
Leaders for the Digital Economy

LONG & FOSTER REAL ESTATE, INC.

The Problem

- Determine the factors that lead to a ‘quick’ sale for condos in Arlington, VA
  - Useful in a slowing real estate market
- Answer such questions as:
  - Does the community and surrounding area affect the time a property is on the market?
  - How do the attributes of condos affect quick sales?

The Data

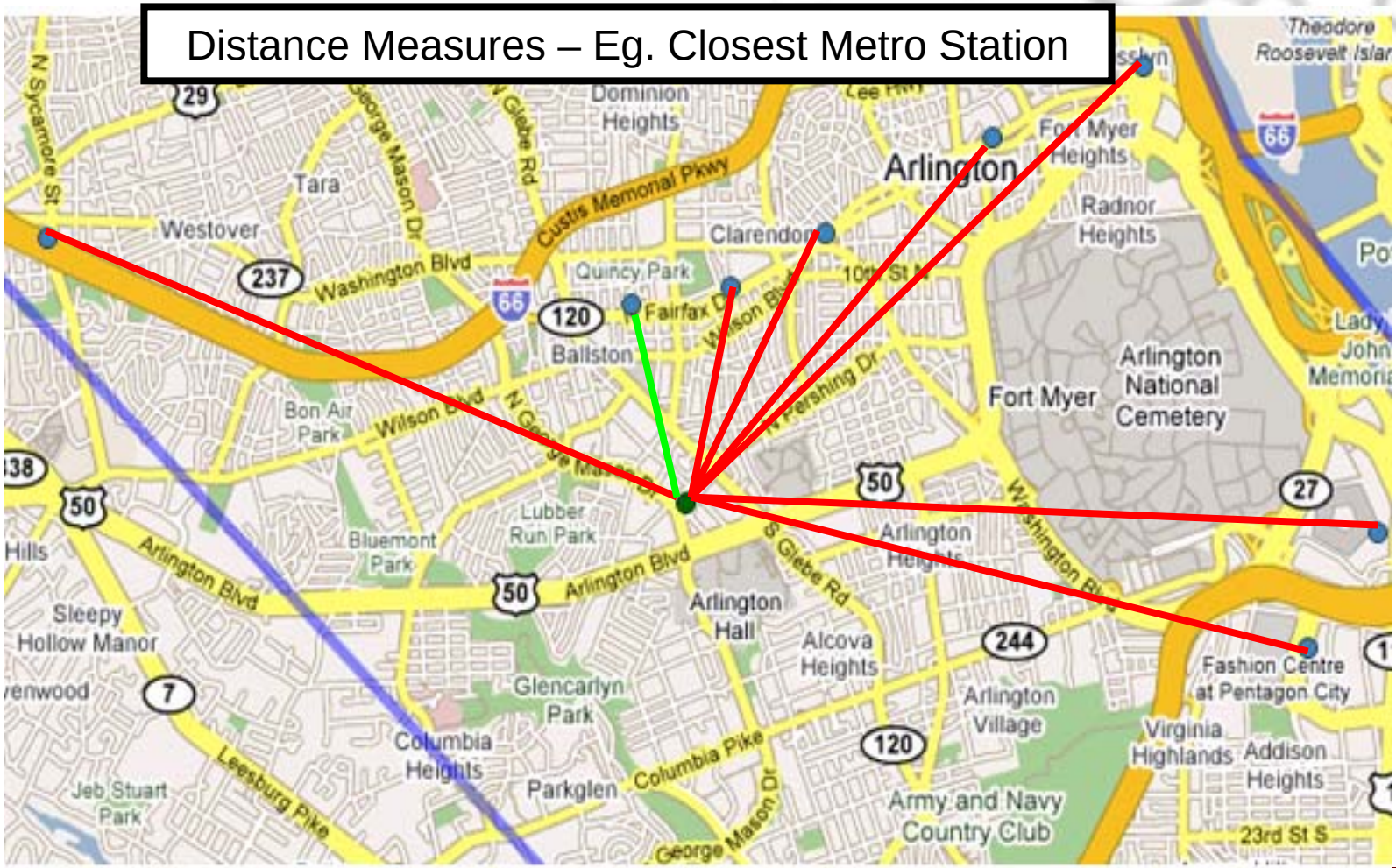
- Condominium Data: MRIS Database
  - ~1,500 records of all condos sold in Arlington from 10/1/06 to 11/1/07
  - Worked with a local Long & Foster real estate agent to identify fields from MRIS most likely to affect quick sale

- Community Features
  - Identified and located addresses for features likely to contribute to quick sales
  - Crimes, Bars and Restaurants, Convenience and Grocery Stores, Parks, Metro Stations

Data Details & Preprocessing

- Created binary “Quick Sale” Outcome Variable
  - Based on Days on Market: < 15:=1 | ≥15: = 0
- GeoCoded addresses of all condos and all community features
  - Obtained lat. & long. from www.batchgeocode.com
  - Used “Great Circles” formula for “Crow Flies” distances
  - Generated Density/Count Variables
    - Eg Number of Crimes w/in .5 miles
  - Generated Distance Variables (distance in miles to closest feature)
    - Eg. Distance to closest metro station (miles)

Geocoding Process



ROBERT H. SMITH SCHOOL OF BUSINESS
Leaders for the Digital Economy

LONG & FOSTER REAL ESTATE, INC.

Data Exploration & Selection of Variables

Step 1

- Retail
- Parks
- Attributes
- Metro
- Pricing
- Crime
- Nightlife

Step 2

Reduction Process

- Domain Knowledge Test
- Visual Exploratory Analysis
- Compare AVE and STDEV
- Correlation Test
- Accuracy Test

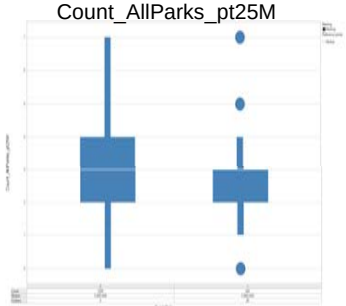
Result

Remaining Variables for Profiling Model

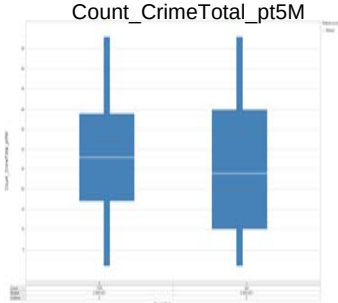
Variable	DataType	FieldType	Description
Count_AllParks_pt25M	Integer	Community Features	Number of parks within .25 miles of the condo
Dist_ClosestMetro_M	Decimal	Metro	Dist (miles) to closest metro station (any line)
Count_RestBarHotel_pt25M	Integer	Nightlife	Number of restaurants/bars/hotels within .25 miles of condo
Dist_ClosestConv_M	Decimal	Retail	Dist (miles) to closest convenience store
Count_CrimeTotal_pt5M	Integer	Crime	Number of crimes (all types) within .5 miles
PriceMovement	Money	Pricing	Equal to ClosePrice - Subsidy - ListPrice. If positive, condo sold above list, if negative, condo sold below list.
TotSqFt	Integer	Condo Attributes	Square footage size of the condo
CondoAge	Integer	Condo Attributes	Age of the condo (years). 2007-YrBuilt
PricePerSqFt	Money	Condo Attributes	The Real Close Price (ClosePrice-Subsidy) amount divided by Total Sq Ft

Visual Exploratory Analysis

Count_AllParks_pt25M



Count_CrimeTotal_pt5M



ROBERT H. SMITH SCHOOL OF BUSINESS
Leaders for the Digital Economy

LONG & FOSTER REAL ESTATE, INC.

Modeling Process

- **Investigated different types of models**
  - Discriminant Analysis, Classification Trees, Logistic Regression

- **Decided on a logistic regression model**
  - Clearest in terms of explaining results
  - Considered many combinations of 9 input variables
  - Ultimately selected a 6-predictor model.
  - All predictors statistically significant using a 95% confidence interval

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-3.70876455	0.47157034	0	*
Count_AllParks_pt25M	-0.14204456	0.05453635	0.00919857	0.86758262
Dist_ClosestMetro_M	0.26891452	0.08197037	0.00103575	1.30854332
Dist_ClosestConv_M	0.98084885	0.45526034	0.03120263	2.66671896
CondoAge	0.01129689	0.00324709	0.00050314	1.01136088
PriceMovement	-0.00003258	0.00000497	0	0.9999674
PricePerSqFt	0.00639065	0.00087736	0	1.00641108

Evaluating Model Performance

- Model demonstrated a poor fit to the data!
  - Low Multiple R^2 - 7.30%
    - Very little improvement over naïve model
  - High Overall Error Rate
    - 28.31% across all records
  - Low Sensitivity in Identifying Quick Sales
    - 15.26% Sensitivity, 84.77% Error Rate

Residual df	1459
Residual Dev.	1660.669312
% Success in training data	30.01364256
# Iterations used	9
Multiple R-squared	0.07297314

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	1	0
1	67	373
0	42	984

Error Report			
Class	# Cases	# Errors	% Error
1	440	373	84.77
0	1026	42	4.09
Overall	1466	415	28.31

ROBERT H. SMITH SCHOOL OF BUSINESS

Leaders for the Digital Economy

LONG & FOSTER REAL ESTATE, INC.

Conclusions

Overall our model did a poor job in explaining quick sales of Condo's in Arlington…

- More & different variables are needed to improve model fit
  - Perhaps economic indicators: Foreclosures, Employment, Etc.
  - If possible link these to the period Condos were on market
- May need to reconsider the threshold for defining a "quick sale"

However, with statistical significance, our model did find that condos that were more likely to be "quick sales":

- Were farther from convenience shopping
- Were farther from Metro Stations
- Were older rather than newer properties
- Had fewer parks w/in .25 Miles
- Displayed less of a drop between List Price and the total amount paid by the buyer
- Commanded higher close prices per square foot