

Predicting Intake of Applications for First Registration in the Property Registration Authority



Orlaith Mernagh
C99391341
September 2016

Dublin Institute of Technology
MSc in Data Analytics

Motivation & Research Question

Research Question:

“What are the factors that influence intake of applications for first registration to the Property Registration Authority, and can these features be used to build models to predict future intake?”

- **Real Business Need:** Reliable system for predicting intake required by the PRA
- **Aim** of dissertation is two-fold:
 1. To identify the factors that influence intake of applications for first registration
 2. To ascertain if these features could be used to build models to predict future intake.

Literature Review & Background

Prediction & the Property Market

The Irish Housing Market

- * Cyclical in nature, peaks & troughs
- * Linked to wider economy
- * Urban/Rural Divide

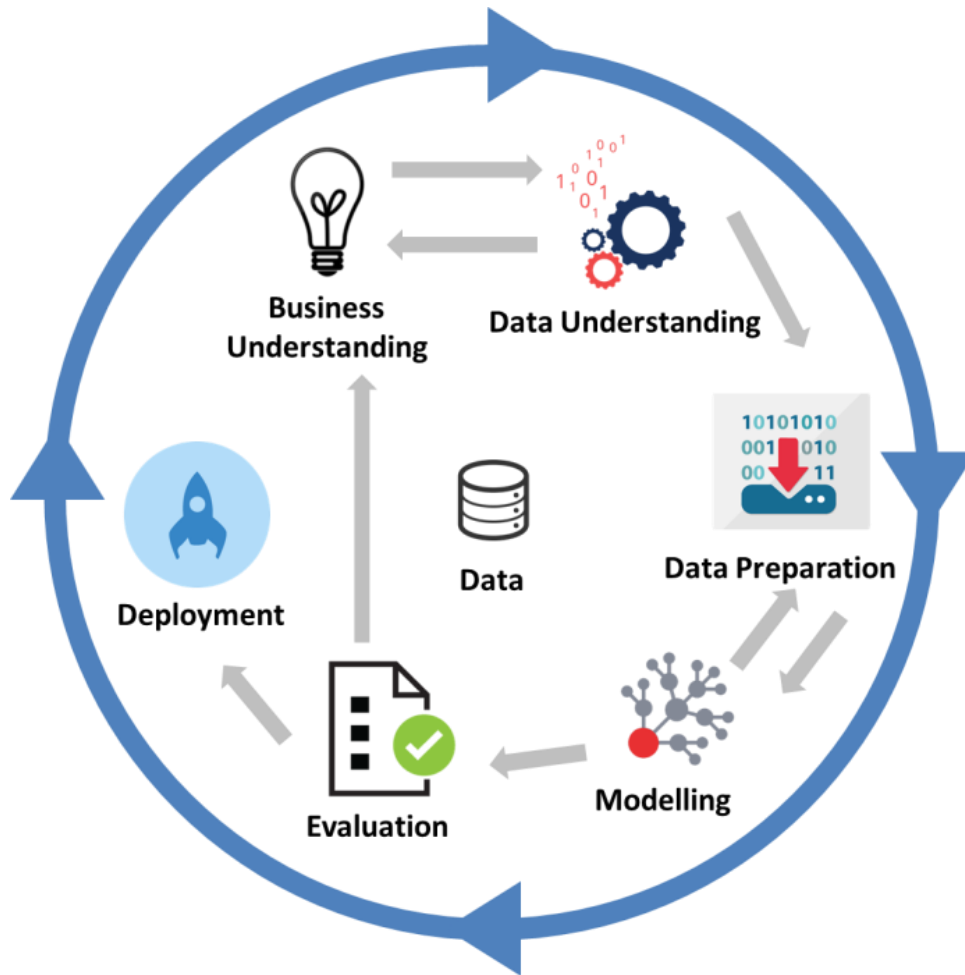
Factors that Influence Property Markets

- * GDP, Inflation & Interest Rates
- * Population Demographics
- * Unemployment Rates
- * Location
- * Property Attributes
- * Seasonality

Modelling Approaches

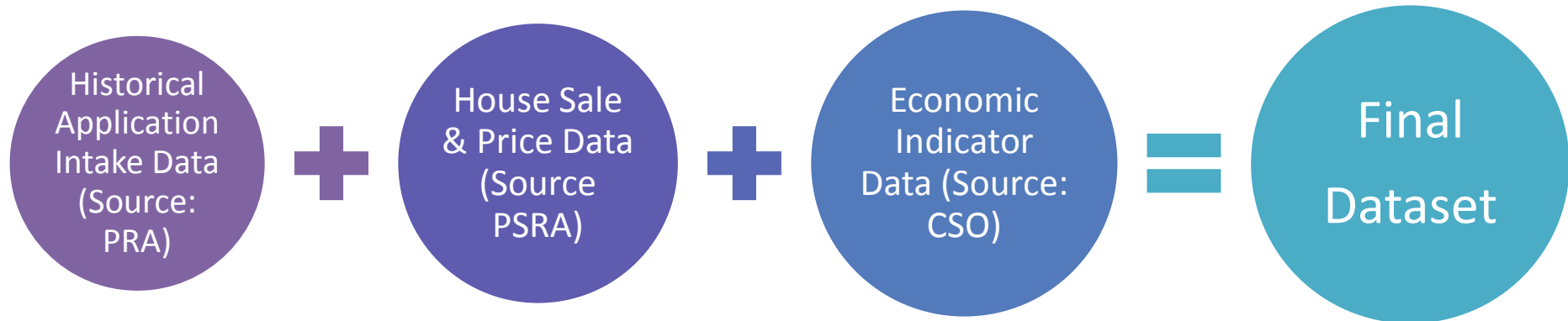
- * Decision Trees & Random Forests
- * Linear Regression
- * Artificial Neural Networks

Design & Methodology



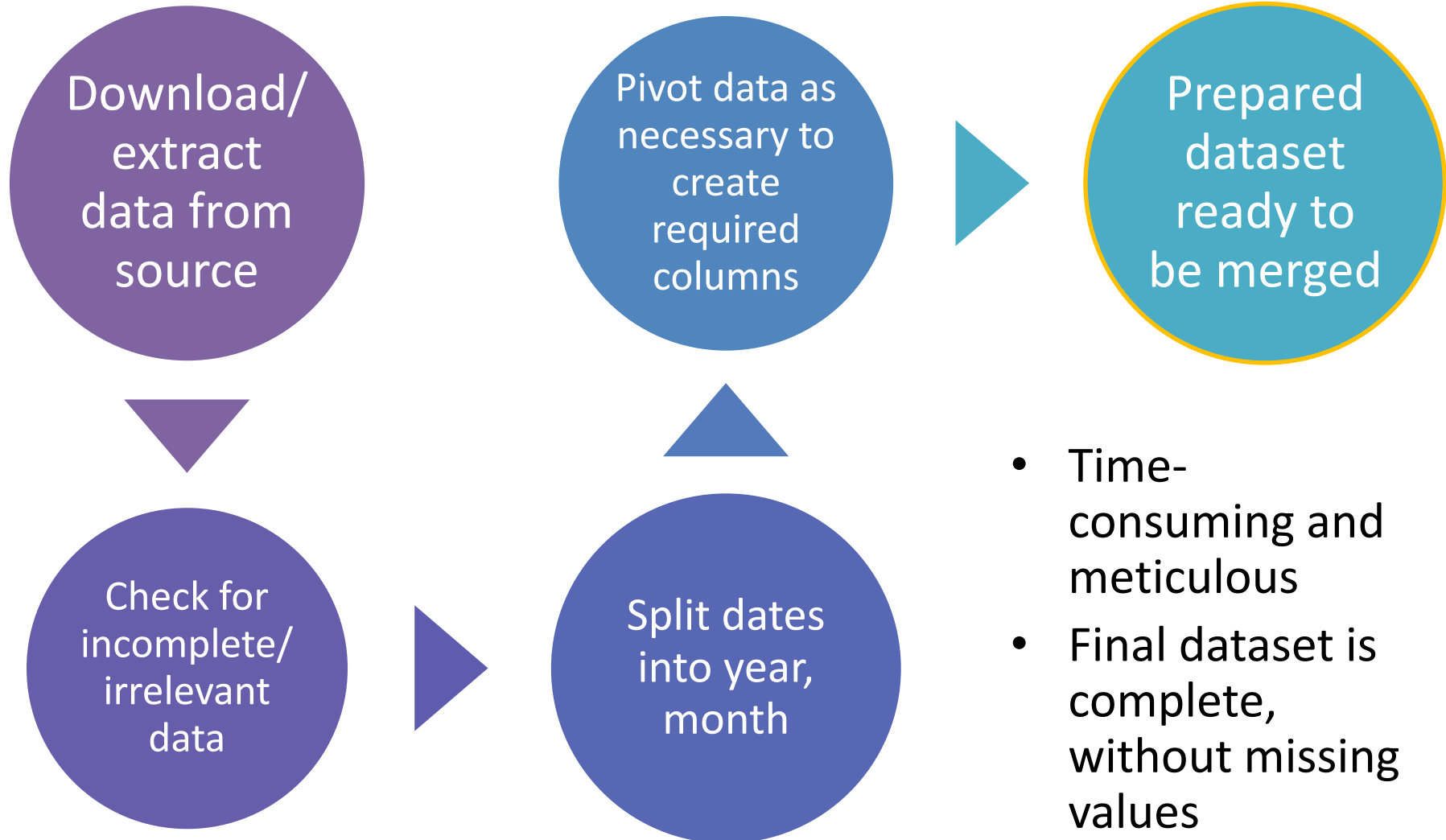
- CRISP DM Methodology
 - Industry standard
 - Creates logical structure
- Software Used:
 - Microsoft Excel
 - Tableau
 - RStudio

The Dataset

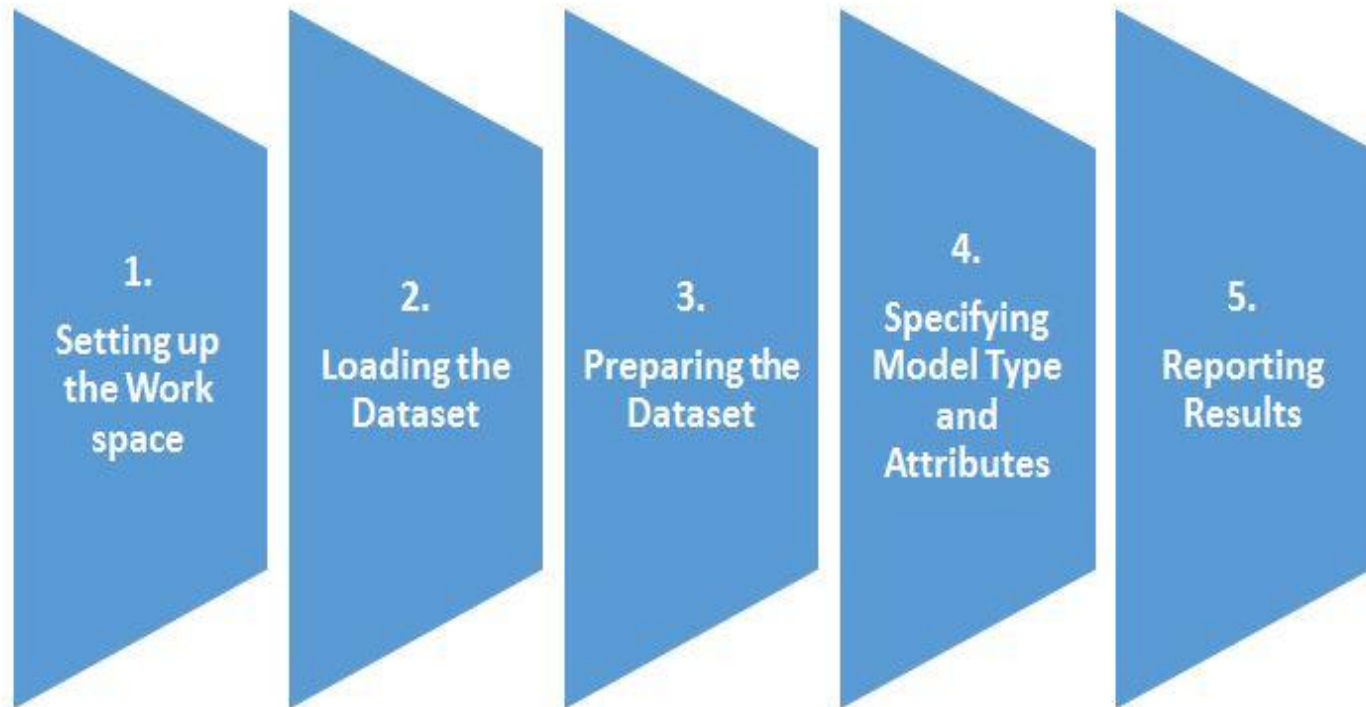


- 3 individual sources
- All datasets cover 6 years 2010-2015 inclusive
- Process of data extraction, data preparation & consolidation into a single dataset
- Final dataset of 14 variables with the Target variable 'Count' and comprises a total of 1,872 rows
- Dataset explored thoroughly and a series of exploratory visualisations created

Data Preparation Process



Implementation & Experiments



- A series of models created using RStudio
- Dataset assessed for Normality, to identify Outliers and to check for Associations

Implementation & Experiments

Model #	Description
1	Regression Tree, no MinSplit, automatic pruning
2	Regression Tree, MinSplit = 25, manual pruning
3	Random Forest
4	Multivariate Linear Regression using stepwise regression
5	Neural Network – one hidden layer, only continuous inputs
6	Neural Network – three hidden layers, only continuous inputs
7	Neural Network – one hidden layer, categorical and continuous inputs
8	Neural Network – three hidden layers, categorical and continuous inputs

- 8 models built in total
- Each model run against the training set and results extracted for evaluation

Results

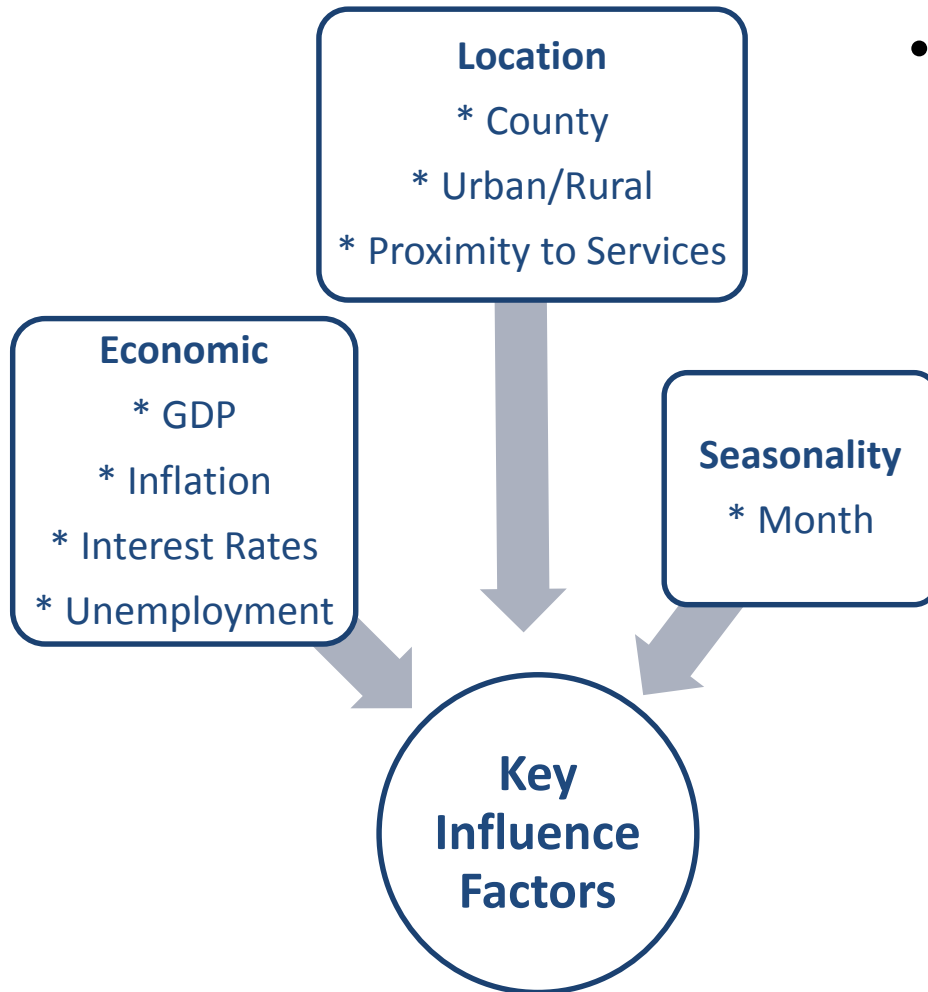
- 2 x **Evaluation Metrics** used:
 - Percentage Mean Absolute Error (%MAE) as a measure of accuracy
 - Percentage Root Mean Squared Error (%RMSE) as a measure of the average magnitude of error
- **Variable Importance:**
 - Houses Sold is key variable
 - Population & Year also significant
- **Most successful:** Random Forest (Model 3), in terms of predicting Count and with the smallest values of %RMSE and %MAE
- **Least Successful:** one of the Neural Network models (trained with three hidden layers and including the County and Month variables) & Regression Tree models

Challenges

However, several challenges encountered:

- Difficult to access **consistent data** for certain potentially relevant variables – e.g. interest rates
- To put theory into practice as a useful forecasting tool for the PRA, **further development** is required:
 - Models are accurate in prediction of current intake based on current house sale data; further development required to predict future intake based on previous data
 - There is a delay in extracting up to date data from some sources; eg rates of GDP
- Recent mass retirements amongst **Management Team** have placed extra demands on remaining staff; not possible to discuss findings - will have an opportunity at the end of this month at upcoming Divisional Managers Meeting.

Contribution & Impact



- Significant insights gained:
 - In depth analysis of the housing market in Ireland, its cyclical nature and the urban rural divide
 - Identification of key factors that influence property markets & application intake
 - Recognition of the challenges facing the PRA and the need for an accurate prediction instrument
 - Study of typical models employed in house price prediction, and exploration of which are suitable for intake prediction

Future Work

- International comparison with similar jurisdictions
- Embellishment of the Dataset:
 - Addition of Interest Rate data
 - Removal of Consumer Price Index data
 - Arrange quicker access to regularly updated data (monthly/quarterly)
- Further analysis of outlier impact and experimentation with removal
- Alternative Modelling techniques
 - Time Series Forecasting
- Alternative Software
 - SAS

Questions