Door-to-Door Charity Collection

Draft

What is the problem

- The charity marketplace is extremely competitive in AU
- Accentuated by diminishing government and corporate funding
- Large-scale fundraising campaigns essential to maintain viability
- Although largely executed by volunteers, there are significant costs involved in the administration and promotion of door-to-door fundraising
- To efficiently target resources (including marketing) and maximise ROI, need to:
    - Understand profile of collectors
        - Are there any defining characteristics or distinct groups?
    - Identify
    - Understand key factors that predict whether a collector is ‘profitable’

What data do I have?

- 2015 collection data from *anonymous* national charity
  - Postcode, age, number of streets covered, new or existing volunteer, total collected, total donated (by the individual collector), total received (collected + donated), profitability (binary whether or not return exceeds average cost)
  - Private
- 2013/14 ATO Postcode Data
  - No. of individuals, salary, income, tax deductible donations
  - Public
- 2011 ABS Census Community Profile - Postcode level
  - Comprehensive demographic information to build out postcode level data
  - Public

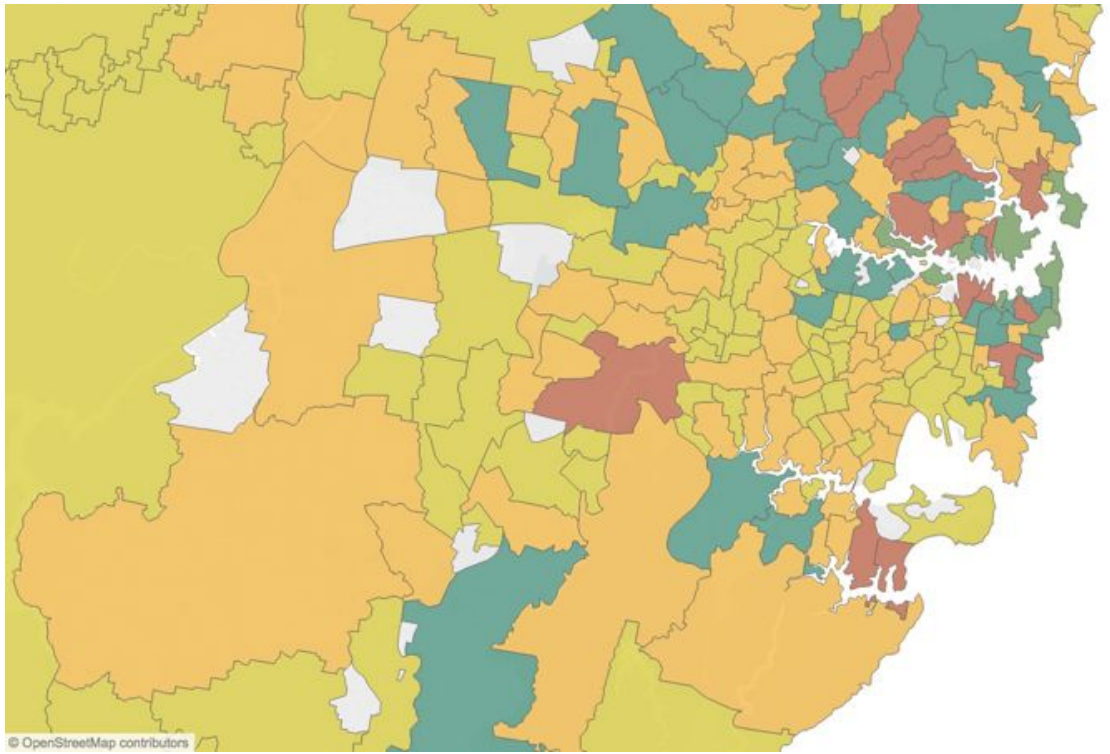
What did I do?

- Collated and prepared the data for testing/analysis
  - Needed to create a business relevant metric not present in the data - ‘profitability’
- Tried the following models:
  - **K-means clustering**
  - PCA
  - Linear Regression
  - Logistic Regression
  - Decision Trees
  - **Random Forest**

What were the results?

- Clustered collectors into 5 distinct groups
  - Gave them descriptive labels
  - Mapped them
  - Ready to send to outbound call centre/media agency
- Model that predicts profitability based on 5 features:
  - Age
  - Number of streets
  - Postcode
  - New or returning collector
  - .97 accuracy on training data (tick), generalises at .57 (cross)
    - Classif overfit?

Collector Clusters



Profitability Prediction

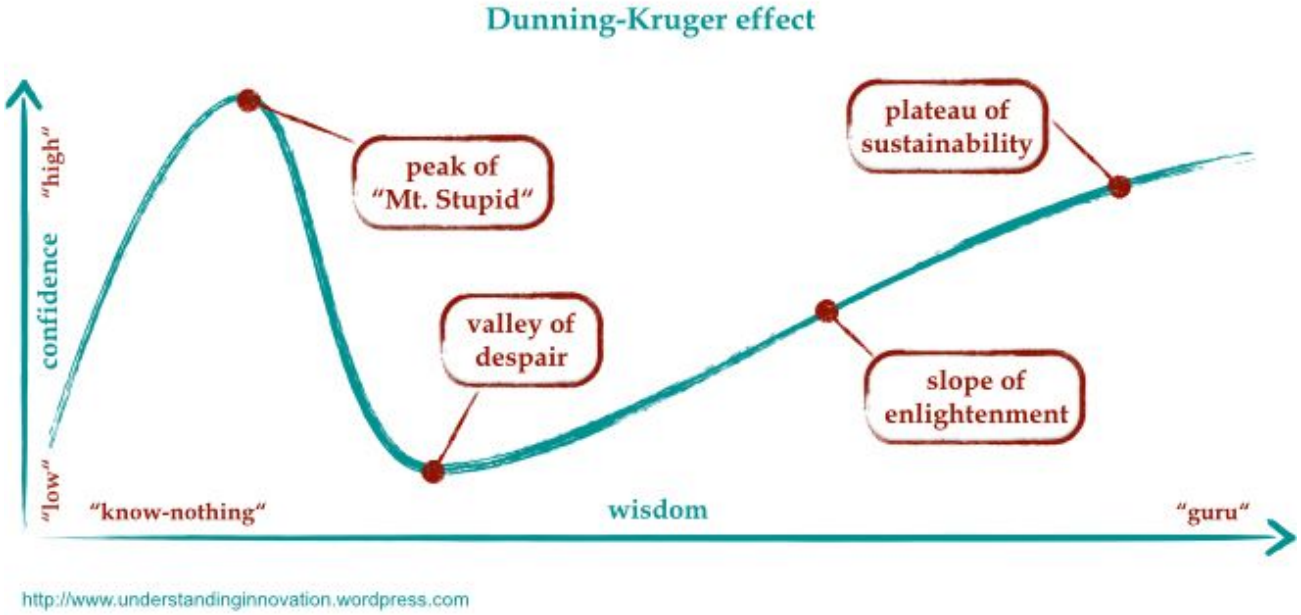
Did I Achieve What I Set Out To Do?

- In a way...
  - Near MVP product
  - Not usable (at this level)
  - Successfully implemented the techniques across the data
  - Established a potential use case

What Did I Learn?

- Not to lose the wood for the (decision) trees
  - Desire to build a model that predicts nicely moved me away from actionable business insights
- Better understanding of characteristics of data and at what level DS techniques can be employed

I Took an All Too Familiar Path



What I Will Do Next

- Go back to the data
- Work hard to find the most meaningful features and develop them
- Test it more against other ‘knowns’ in the dataset
- Create more interpretable clusters
- Look to improve prediction model or reframe the problem
-

Appendix