Data Science Fundamentals

A description of data science techniques that can be used on both small and large data sets and an evaluation of the effectiveness of these, using examples from the UK charity sector

Executive Summary

With increasing volumes of data being produced, organisations are faced with both problems to solve and opportunities to seize. A standardised approach to collecting, manipulating and analysing data will help to unlock insights and guide management strategy. This report considers two examples from the UK charity sector, one large, one small to evaluate the data science methods used in their operations.

Analysis of the BBC Children in need dataset showed interesting differences in the grant per head by region. The Channel Islands, an area not normally identified as being 'in need' received almost twice the UK average grant. Moving from a bid-based grant allocation towards a predictive model would greatly improve transparency and help the charity proactively target those most in need.

The Trussell Trust logistic regression model for food bank usage whilst being a useful tool, is complex. The volatility in food bank driver variables and the exclusion of other possibly important variables means that the model should be viewed as a tool to support and inform management decisions in the knowledge that it will need to be assessed, monitored and adjusted over time in the light of new information.

Government socio-economic data is a vital part of developing sound predictive models in the charitable sector. A national food security measure, like the fuel poverty indicator and a national database of food bank usage would greatly help charities as well as provide the Government the means to monitor UK progress towards meeting its commitment to the UN Sustainable Goal for reducing hunger.

The ability to share data amongst charities would help them develop predictive models but this must be balanced with the need to meet the GDPR and increased public concern from recent high-profile cases of un-ethical behaviour in the charitable sector.

Contents

Introduction

By 2020, it is predicted that 1.7 Mb of data will be created every second for every human being and accumulated data will grow to approximately 44 trillion gigabytes (Marr, 2018). Increasing amounts of data along with the capacity to store and process information with faster, larger and more complex computing resources becoming more readily available, enhances the potential for organisations to unlock insights to inform their operational strategies. To take full advantage of this potential, an organization should develop a data culture, or become 'data driven'. By making clean, reliable and timely data available quickly and as widely as possible, an organization can gain significant competitive advantage. (Patil and Mason, n.d.).

Data science is the process of extracting meaningful insights from data using analytical or scientific methods. The flow of data through the different parts of this process is often referred to as a data pipeline, see Appendix 1. This report examines each element with reference to data sets from the charitable sector.

Small Data

Small data can be defined as "data in a volume and format that makes it accessible, informative and actionable." (WhatIs.com, 2018). Data is usually held locally in a structured or tabular form in databases or spreadsheets and can be used to provide answers to specific questions or problems. The rate at which new data is accumulated is usually slow and steady and hardware is vertically scalable.

Define the Problem

The first step in the data science process is to define the questions or problem you want to solve. Developing a thorough understanding of the organization and setting clear objectives, keeps the work focused and helps to identify data requirements.

The BBC Children in Need Appeal (CIN) grants allocation data is one example of a small data set. The Charity's aim is to help disadvantaged children and young people in the UK to have a safe, happy, secure childhood and reach their potential. Following an assessment process, grants are awarded to applicant organisations for projects running over 1 to 3 years. There is an annual report and an interactive map on the website showing where grants have been allocated. Appendix 2 shows examples from the report and it is evident that the information is well presented and summarized.

However, they report total grants but not grant per head. In 2017, The National Lottery was heavily criticised as the per head spend was considerably higher in Scotland than England. (Civilsociety.co.uk, 2018). The question to be investigated was whether this was also the case for the CIN appeal.

Acquire Data

The data is taken from 360 Giving, a site that publishes open grant data from 94 charities in the UK. The dataset was contained within a CSV file with 1694 lines and 15 fields. Most fields were surplus and could be removed, leaving the location, amount, project duration, start date and description.

Wrangle and Clean

Raw data usually needs converting into a useable format. This includes merging or joining information held in different formats such as CSV files, spreadsheets or databases and commonly known as 'wrangling'. Data cleaning removes duplicates, identifies errors, outliers and missing values and basic descriptive statistics can then be compiled. This process can be very time-consuming and frustrating, absorbing up to 80% of the time spent by the data scientist (V. 2018). Clean data is vital and as noted by (Patil and Mason, 2015 p.6) 'data must be organized, well documented, consistently formatted and error free.'

The dataset was read into Python, converted to a list and cleaned to remove empty columns, whitespace, unwanted data, correct any errors or numbers shown as text and get it into the format required form for analysis. Python Pandas library was used to view, group and sort the data which was read to a CSV file. The full code is in Appendix 3.

Regular expressions (Regex) were used to strip out reference numbers. Regex are useful where there is a set pattern, but the characters vary, and these were abundant in the sample dataset. Appendix 4 show examples of the expressions. Each regex was tested to check that it deleted the right data using a test set of 2000 lines from the original file.

Descriptive Analytics

The exploration of data will identify any patterns, trends or interesting features that might support the objective or direct the investigation.

The 2011 Census 2011 population figures were obtained from the Government website in order to calculate the grant per head. As the dataset listed location differently to the Census, it had to be adjusted and the figures amalgamated for each region. The results were graphed and are shown in Appendix 5.

Just over half the total grant goes to London, Scotland, North West, West Midlands and the South East but the grant per head shows a different picture. Although the grant per head figures for NI, Scotland and Wales were amongst the highest, the Channel Islands grant per head of £2.95 was the second highest and almost twice the national average. Almost £472,000 was received for 6 projects, an average of £78,627 per project compared to £55,968 for all CIN projects.

Top Five Regions by Grant Allocation for BBC Children in Need Appeal (2013-2015)

Region	Total Grant £m	Region	Grant per head (£)
London	16.0	Northern Ireland	3.38
Scotland	11.1	Channel Islands	2.95
North West	10.7	Wales	2.17
West Midlands	8.5	Scotland	2.09
South East	7.0	London	1.96
All Regions	94.7	Average all regions	1.50

Appendix 5 shows the detailed descriptive statistics and graphs.

Predictive Modelling

Predicting analytics forecasts outcomes, trends or behaviour based on historical data using statistical models and/or machine learning algorithms. A detailed description of machine learning is outside the scope of this report, but examples are shown in Appendix 6. The choice of algorithm depends on the size, quality and complexity of the data and the nature of the task. Data is split into a training set to 'train' the model and a test set to evaluate the model often on an 80:20 ratio. Overfitting problems can occur if the model matches the training data so closely that it doesn't correctly predict when presented with new data.

The BBC appeal allocates grants to applicants based on local assessment by a board so predictive analysis is not used. The most deserving organisations may not get funding because they didn't apply and those that are better at writing bids might get more which could explain the results from the Channel Islands. Predictive models can address this but are complex and require skillsets many charities do not have the resources to fund.

Evaluation

The data shows where funds are spent but little reassurance that it is targeted at those in most need. Reporting the ratio of successful applications to total applications and the source of funds raised by region would improve transparency. Further, building a predictive model, with Government open data, other child and poverty charity data would represent a proactive approach to identifying need. The following section illustrates the development of such a model.

Big Data

Big data can be defined using the three V's, of Volume, Velocity and Variety (Blogs.gartner.com, 2018). The size of the data set; the speed at which this data is generated, and its variety or complexity is also relevant. Big data is often unstructured i.e. not confined to traditional databases or spreadsheets and may be held on different servers in geographically diverse locations. Data formats can include emails, photos, videos, tweets, blog comments, phone calls, sensor data etc. and often requires distributed file systems on or off the Cloud to store and process.

Charities can generate vast quantities of data. Appendix 7 shows data sources that one large UK charity, the McMillan Cancer Trust produces gained from a search of their online presence. Unlocking insights from data can be extremely valuable for charities in terms of:

- targeting scarce resources more effectively and perhaps more controversially
- identifying potential new funding streams.

One example is Parkinson's UK, who have used data analysis, social media, and algorithms combined with Government open data to identify what people care about and the connections between them, using this to successfully identify high-value donors with whom they can build relationships. (Marrins, Stephenson and Kinloch, 2018).

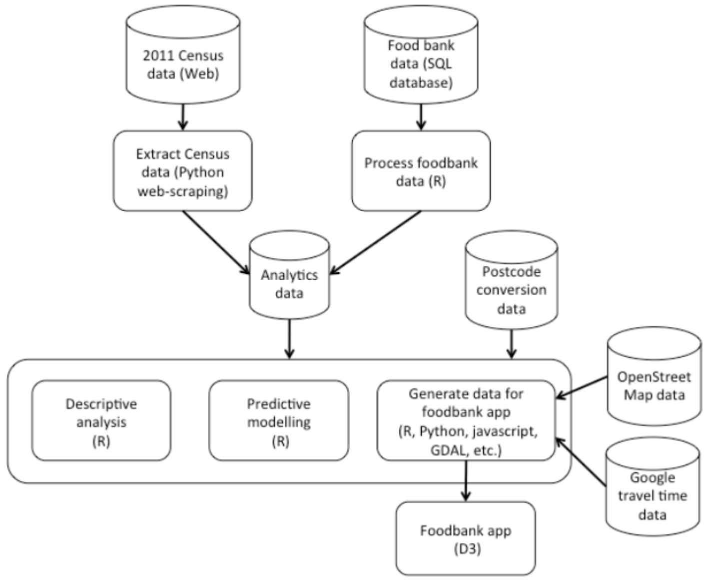
The Business Problem

The Trussell Trust is the largest organization providing food banks in the UK. The charity delivers around 1.3m emergency 3-day food packs from 420 food banks and 1200 distribution centres across the UK each year. See Appendix 8. Food bank usage has been increasing in the UK, up 13% in 2018 over the previous year. With the growth in usage, the Trust commissioned Hull University to:

- Conduct exploratory analysis of the food bank data
- Build predictive models for bank usage
- Develop a mapping app to help local and head office managers map food bank use, and identify areas of need

Data Acquisition, Preparation and Analysis

The diagram shows the process used to gather, merge, clean and analyse the data. Scraped Census data was added to the food bank usage data to prepare the model. *(Nemode.ac.uk, 2018)*



Descriptive Analytics

Descriptive statistics and graphs uncovered interesting insights for the period of the study:

- After rapid growth initially, the rate of increase in use of banks levelled off except for homeless and ethnic minorities.
- Key reasons for food bank use was unemployment in London, benefit problems were relevant in Yorkshire and in the North East, low income was key.
- Interesting outliers showed rural areas had heavy usage of food banks compared to urban.

Predictive Modelling, Deployment and Visualisation

The report identified 7 broad drivers of food bank usage:

- government policy,
- cost of living,
- availability of low skill employment,
- number and age of food banks
- referral agency provision,
- social acceptability and the
- availability of alternative emergency food

Bayesian Statistics was used to investigate food bank maturity and build a standard food bank model. This identified that a typical bank would service 50 people/week, take 6 months to reach maturity and serve double the yearly average over the Christmas/New Year period. Appendix 9 describes this model.

A logistic regression model was built to predict food bank usage geographically. Logistic regression is used to establish the relationship between a dependent binary variable (having one of two outcomes) and one or more independent variables. It assumes that the relationship between the predictor variables and the dependent variables is uniform, i.e. constant over a range of values.

Using 11 food bank variables and 30 socio-economic indicators from the 2011 Census, the results were used to develop an online interactive map which can toggle between current demand and predicted demand for an area. Appendix 10 lists the variables used in this model.

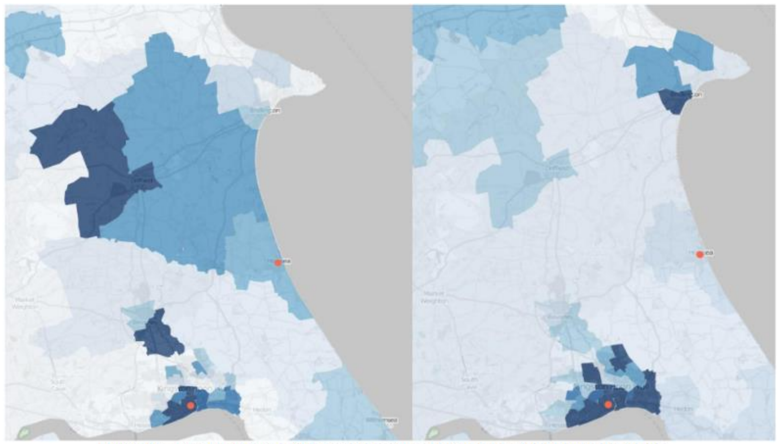


Figure 12: actual (left hand side) versus predicted (right hand side) food bank usage

Evaluation of the Model

Common problems with logistic regression models include:

- choosing the wrong independent variables so the predictions will be incorrect,
- using too many variables which can dilute associations between dependent and independent variables,
- using variables which are closely correlated with each other (multicollinearity) so the effect of each is less precise and this can dilute the model
- 'overfitting', where the training model fits the data too closely, so that when faced with new data, the model fails

(Priya Ranganathan, 2018), (EliteDataScience, 2018).

The Trussell model relies on historical data which works well in stable environments. Food bank demand is likely to be volatile since the underlying socio-economic drivers can change rapidly and this could affect the performance of the model long-term.

The model is complex with many variables but there are likely to be others not included which could affect food bank demand such as:

- weather patterns with higher fuel costs reducing money available for food (Lewis, 2018),
- seasonal patterns of migration with more homeless in summer in coastal towns (Eastbournecab.co.uk, 2018, p4),

- arrival of migrants and refugees that can't be easily predicted (Sutton, 2018),
- impact of government policies such as Universal Credit (Ft.com, 2018),
- local business expansion and contraction (openDemocracy, 2018).

It is likely that some of the Census variables that are related, will have multicollinearity. This can be addressed by removing one of the correlated variables. With overfitting, there are various statistical methods available, such as increasing the size of the training set and using cross validation (creating many subsets of training and test data to refine the model). Given the model was built by statisticians, it is assumed that if either of these issues were encountered, that they were addressed.

Building predictive models using socio-economic data can be difficult. Although a useful tool, models must be regularly reviewed for performance, adjusted for changes in the variables and monitored carefully by experts in the field.

Other Issues

There are many groups in the UK collecting poverty data which could be shared. For example, single pensioners are at higher risk of food poverty, so data collected by elderly charities could be useful (Ageuk.org.uk, 2018).

Private retail chains analyse data when locating new stores which could also help identify potential locations for food banks.

With rapidly rising use of food banks there have been calls for the Government to appoint a minister and department for food security (Church-poverty.org.uk, 2018). The Government is committed to the UN Sustainable Development Goal of removing hunger by 2030 and progress towards this could be monitored by using a national database of food bank usage and a household food insecurity measure, much like the Fuel Poverty Index. Other countries do measure food security. The US Department of Agriculture annual survey (Ers.usda.gov, 2018) generates very detailed statistics used by the 'Feeding America' foodbanks to prepare maps showing areas of food poverty which help them deliver their service.

A more controversial use for predictive analytics is to target high-value donors, a well-established business in the US. Highly sophisticated techniques are used to examine the characteristics and data of individual donors, although robust rules around data protection in Europe are likely to restrain growth of this area.

Data Protection and Ethics

This desire for data sharing by charities must be balanced by data protection. In 2017, the Information Commissioner fined 11 UK charities for breaches of the Data Protection Act including screening donors and trading donor details with other charities (Civilsociety.co.uk, 2018). Also, the General Data Protection Regulations (GDPR) came into force this year, protecting EU citizen's data and affecting how organisations store and manage that information and charities are not exempt from this. Any client or donor data that is used for analytics will need to be fully anonymised meet these regulations.

There are also ethical concerns. In 2014, the Samaritans released an App using an algorithm to monitor tweets, to indicate when someone was in distress and send emails to someone they follow. They were criticized for making people with mental health problems feel more vulnerable (BBC News, 2018).

Whilst people can regard charities more leniently than businesses, they can also feel aggrieved when charities are perceived to not behave ethically or morally. Cases such as the Oxfam scandal, the death of poppy seller Olive Cooke who had received hundreds of donation requests before her death (Dailymail.co.uk, 2018) and financial impropriety with Kids Company have resulted in calls for increased scrutiny of charity operations. This is likely to make the public warier of the use of data by charities regardless of whether it is used to target resources more effectively or to target donors.

Conclusions

This report has examined the data science process with reference to two charities in the UK. Whilst the process for big and small datasets is similar, the approach that each charity will take will vary depending on business objectives, available data and resources. Predictive models can provide important insights and using open and shared data gained legally and ethically can greatly increase the value derived from them. Models can be complex and costly to develop and should be viewed as a tool, regularly monitored and supplemented with expert and local knowledge. However, charities and government have much to gain from adopting this more proactive approach to target scarce resources to the neediest in society.

Recommendations

It is recommended that The Children in Need Appeal:

1. Increases transparency by reporting the grant allocation per head of population, as well as the ratio of successful projects compared to applications by area and funding raised by area.

2. Moves towards building a predictive analytics model, utilising government open data, national poverty indices and data held by other child centered charities, to develop a more pro-active grant allocation process which demonstrates they are targeting areas of most need.

It is recommended that the Trussell Trust:

1. Considers developing the predictive model to include other factors such as weather, migration, local economic and seasonal factors.

2. Regularly reviews and evaluates the model including the methodology and variables, to assess performance and adjust where necessary.

It is recommended that charities:

1. Work together to lobby Government to adopt a coordinated approach to dealing with food poverty and insecurity in the UK, including developing a national database of food bank use and a food poverty index.

Bibliography

[], H. (2018). *How to remove multiple indexes from a list at the same time?*. [online] Stack Overflow. Available at: https://stackoverflow.com/questions/11303225/how-to-remove-multiple-indexes-from-a-list-at-the-same-time [Accessed 17 Nov. 2018].

360Giving. (2018). *360Giving - 360Giving*. [online] Available at: http://www.threesixtygiving.org/ [Accessed 16 Nov. 2018].

Analytics, D. (2018). *An introduction to data science*. [online] Ibm.com. Available at: https://www.ibm.com/developerworks/library/ba-intro-data-science-1/index.html [Accessed 17 Nov. 2018].

Anon, (2018). [online] Available at: https://www.quora.com/What-are-different-models-in-machine-learning [Accessed 22 Nov. 2018].

BBC News. (2018). *Samaritans pulls 'suicide watch' app*. [online] Available at: https://www.bbc.co.uk/news/technology-29962199 [Accessed 22 Nov. 2018].

Bbc.co.uk. (2018). *BBC - Home - Children in Need*. [online] Available at: http://www.bbc.co.uk/corporate2/childreninneed [Accessed 16 Nov. 2018].

Blogs.gartner.com. (2018). [online] Available at: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf [Accessed 22 Nov. 2018].

Brownlee, J. (2018). *A Tour of Machine Learning Algorithms*. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/ [Accessed 22 Nov. 2018].

Church-poverty.org.uk. (2018). [online] Available at: http://www.church-poverty.org.uk/stepup/news/pressroom/resources/reports/stepuptotheplate [Accessed 29 Nov. 2018].

Citymetric.com. (2018). *"We mapped out food poverty across England, to see where food banks are needed most" | CityMetric*. [online] Available at: https://www.citymetric.com/horizons/we-mapped-out-food-poverty-across-england-see-where-food-banks-are-needed-most-3627 [Accessed 22 Nov. 2018].

Civilsociety.co.uk. (2018). *Big Lottery Fund criticised for spending more per head in Scotland*. [online] Available at: https://www.civilsociety.co.uk/news/big-lottery-fund-criticised-for-spending-more-per-head-in-scotland.html [Accessed 22 Nov. 2018].

Civilsociety.co.uk. (2018). *ICO fines 11 major charities for data protection breaches*. [online] Available at: https://www.civilsociety.co.uk/news/ico-fines-11-charities-for-breaches-of-data-protection.html [Accessed 22 Nov. 2018].

Dailymail.co.uk. (2018). [online] Available at: https://www.dailymail.co.uk/news/article-3083859/Shame-charities-drove-Olive-death-Organisations-exploited-pensioner-s-kind-heart-admit-sending-begging-letters.html [Accessed 25 Nov. 2018].

Datascience.berkeley.edu. (2018). *What is Data Science?*. [online] Available at: https://datascience.berkeley.edu/about/what-is-data-science/ [Accessed 16 Nov. 2018].

Docs.python.org. (2018). *re — Regular expression operations — Python 3.7.1 documentation*. [online] Available at: https://docs.python.org/3/library/re.html [Accessed 16 Nov. 2018].

Downloads.bbc.co.uk. (2018). [online] Available at: http://downloads.bbc.co.uk/tv/pudsey/Final_Appeal_Annual_Report_2016_email_version.pdf [Accessed 16 Nov. 2018].

Eastbournecab.co.uk. (2018). [online] Available at: http://www.eastbournecab.co.uk/_app_/resources/documents/www.eastbournecab.co.uk/Homelessness%20Report%20Full%20Report.pdf [Accessed 11 Dec. 2018].

EliteDataScience. (2018). *Overfitting in Machine Learning: What It Is and How to Prevent It*. [online] Available at: https://elitedatascience.com/overfitting-in-machine-learning [Accessed 11 Dec. 2018].

Ers.usda.gov. (2018). *USDA ERS - Food Security in the U.S.*. [online] Available at: https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us.aspx [Accessed 1 Dec. 2018].

Feeding America. (2018). *Map the Meal Gap*. [online] Available at: http://map.feedingamerica.org/ [Accessed 1 Dec. 2018].

Ft.com. (2018). *Food bank demand soars during universal credit rollout | Financial Times*. [online] Available at: https://www.ft.com/content/58c638f4-46c9-11e8-8ee8-cae73aab7ccb [Accessed 11 Dec. 2018].

GitHub. (2018). *'groupby' multiple columns and 'sum' multiple columns with different types · Issue #13821 · pandas-dev/pandas*. [online] Available at: https://github.com/pandas-dev/pandas/issues/13821 [Accessed 18 Nov. 2018].

int, C. (2018). *Convert all strings in a list to int*. [online] Stack Overflow. Available at: https://stackoverflow.com/questions/7368789/convert-all-strings-in-a-list-to-int [Accessed 18 Nov. 2018].

Lewis, P. (2018). *Forced to choose eating or heating, family burns furniture to keep warm*. [online] the Guardian. Available at: https://www.theguardian.com/uk/2010/jan/17/eating-heating-furniture-cold-weather [Accessed 13 Dec. 2018].

Macmillan.org.uk. (2018). [online] Available at: https://www.macmillan.org.uk/_images/macmillan-cancer-support-annual-report-2017_tcm9-326894.pdf [Accessed 21 Nov. 2018].

Marr, B. (2018). *Big Data: 20 Mind-Boggling Facts Everyone Must Read*. [online] Forbes. Available at: https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#337d902517b1 [Accessed 17 Nov. 2018].

Marrins, K., Stephenson, K. and Kinloch, N. (2018). *How analytics and big data can transform giving*. [online] JustGiving blog. Available at: https://blog.justgiving.com/how-analytics-and-big-data-can-transform-giving/ [Accessed 22 Nov. 2018].

Nemode.ac.uk. (2018). [online] Available at: http://www.nemode.ac.uk/wp-content/uploads/2016/04/Nemode-FB-project-phase-2-technical-design-V10-FINAL-2.pdf [Accessed 22 Nov. 2018].

News & Views. (2018). *Can your charity stand up to scrutiny? - News & Views*. [online] Available at: https://newsandviews.zurich.co.uk/talking-point/can-charity-stand-scrutiny/ [Accessed 22 Nov. 2018].

Ons.gov.uk. (2018). *2011 Census: Population and Household Estimates for the United Kingdom - Office for National Statistics*. [online] Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/2011censuspopulationandhouseholdestimatesfortheunitedkingdom [Accessed 20 Nov. 2018].

openDemocracy. (2018). *Why are profit-making factories closing across England?*. [online] Available at: https://www.opendemocracy.net/uk/brexitinc/sara-mahmoud/why-are-profit-making-factories-closing-across-england [Accessed 13 Dec. 2018].

Patil, D. and Mason, H. (2015.). *Data driven*.

Priya Ranganathan, R. (2018). *Common pitfalls in statistical analysis: Logistic regression*. [online] PubMed Central (PMC). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5543767/ [Accessed 6 Dec. 2018].

Regular-expressions.info. (2018). *Regular-Expressions.info - Regex Tutorial, Examples and Reference - Regexp Patterns*. [online] Available at: https://www.regular-expressions.info/ [Accessed 17 Nov. 2018].

Sas.com. (2018). *A guide to machine learning algorithms and their applications*. [online] Available at: https://www.sas.com/en_ie/insights/articles/analytics/machine-learning-algorithms.html [Accessed 22 Nov. 2018].

Sutton, C. (2018). *Food banks help migrants in Brighton and Hove*. [online] Brighton and Hove Food Partnership. Available at: https://bhfood.org.uk/food-banks-help-migrants-in-brighton-and-hove/ [Accessed 13 Dec. 2018].

the Guardian. (2018). *How data science is helping charities to fight hunger in the UK.* [online] Available at: https://www.theguardian.com/voluntary-sector-network/2016/may/09/data-science-helping-charities-fight-hunger [Accessed 22 Nov. 2018].

The Trussell Trust. (2018). *The Trussell Trust - Latest Stats.* [online] Available at: https://www.trusselltrust.org/news-and-blog/latest-stats/ [Accessed 27 Nov. 2018].

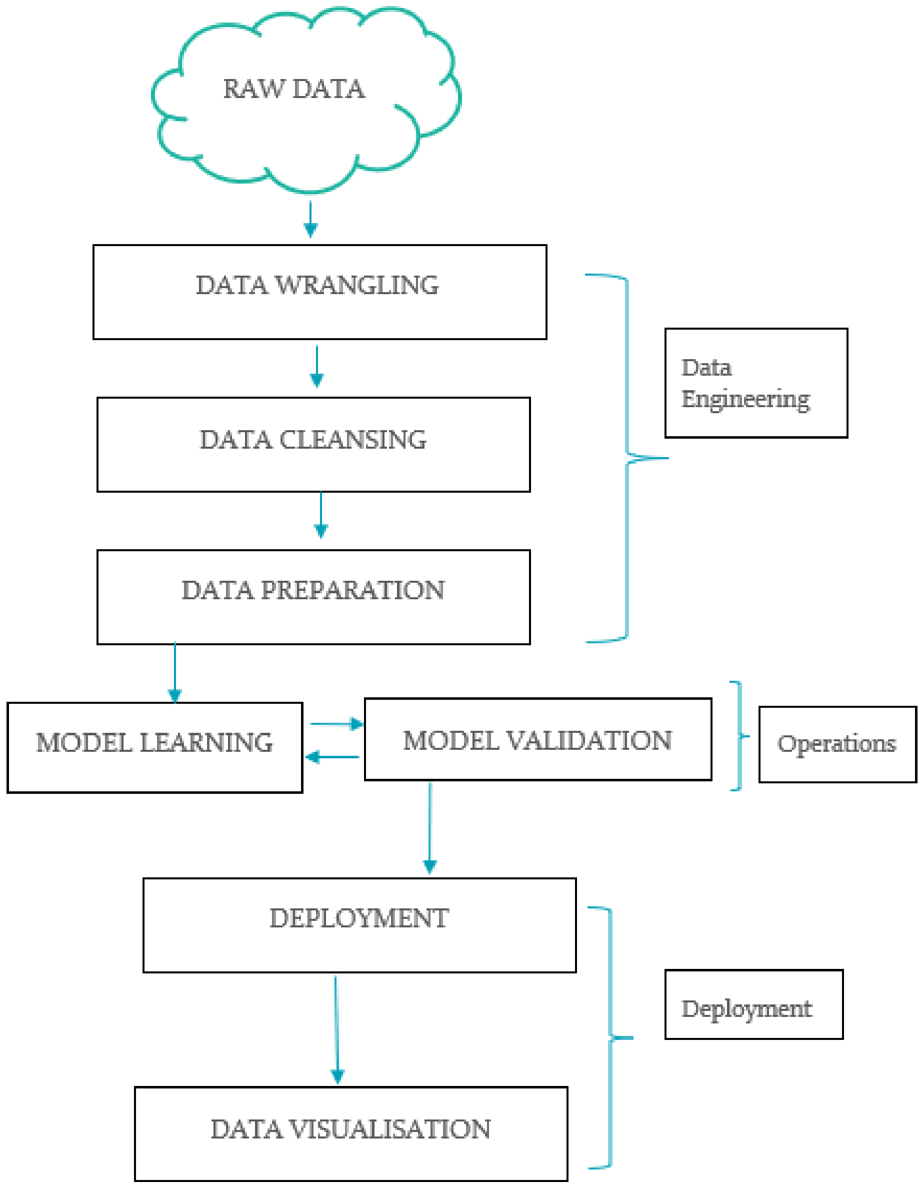
The Trussell Trust. (2018). *What We Do - The Trussell Trust.* [online] Available at: https://www.trusselltrust.org/what-we-do/ [Accessed 22 Nov. 2018].

UpGrad Blog. (2018). *A Beginner's Guide to Data Science and Its Applications | UpGrad Blog.* [online] Available at: https://www.upgrad.com/blog/a-beginners-guide-to-data-science-and-its-applications/ [Accessed 22 Nov. 2018].

V. (2018). *Data Scientists Spend Most of Their Time Cleaning Data.* [online] What's the Big Data? Available at: https://whatsthebigdata.com/2016/05/01/data-scientists-spend-most-of-their-time-cleaning-data/ [Accessed 16 Nov. 2018].

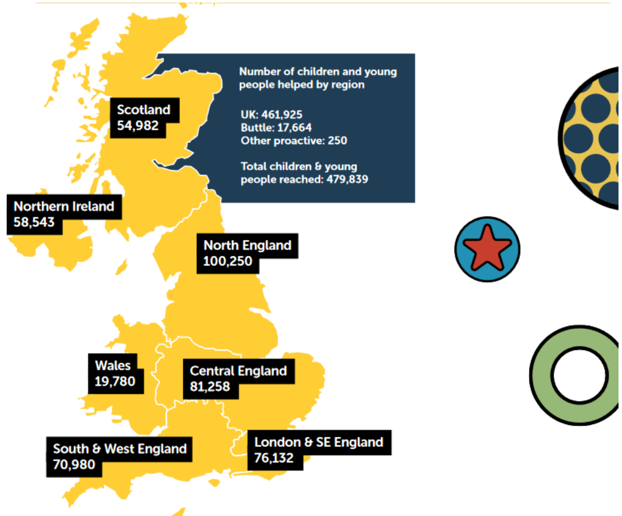
WhatIs.com. (2018). *What is small data? - Definition from WhatIs.com.* [online] Available at: https://whatis.techtarget.com/definition/small-data [Accessed 16 Nov. 2018].

Appendix 1 – The Data Science Pipeline



Source: IBM analytics, 2018

Appendix 2 – BBC Children in need report 2016/2017



Grants awarded in the year

In 2015/16 we awarded £51.4m in grants of which £48.9m related to responsive funding of 1,342 projects.

Main Grants Small Grants Total

Grants awarded in the year

In 2015/16 we awarded £51.4m in grants of which £48.9m related to responsive funding of 1,342 projects.

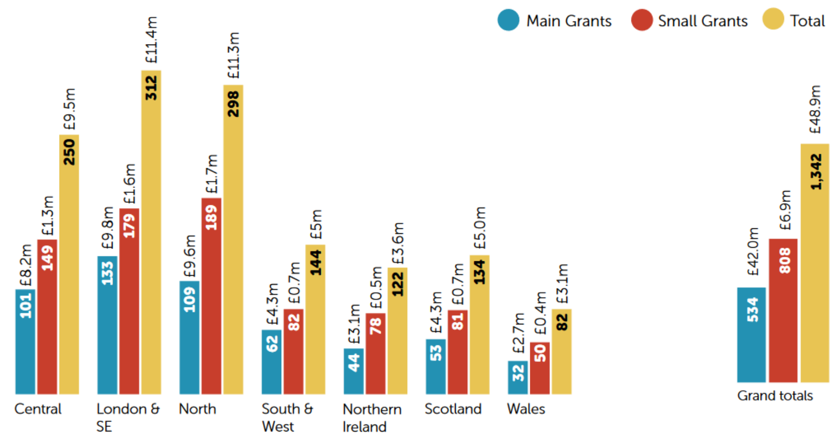


Table 2: Portfolio - Distribution of active grants by disadvantage (June 2016)

Disadvantage	Number	% of total number	Value	% of total value
Abuse/neglect	143	6%	£10,328,807	8%
Behavioural difficulties	72	3%	£4,544,930	4%
Disability	550	22%	£25,998,396	20%
Distress	254	10%	£17,168,526	13%
Illness	189	8%	£10,697,602	8%
Marginalised groups	343	14%	£18,442,044	14%
Poverty and deprivation	903	37%	£42,815,015	33%
Total	**2,454**	**100%**	**£129,995,320**	**100%**

Appendix 3 – Code Used to Clean the Small Data Set

In [2371]:
```
# open and readfile
with open('360_giving_data_02102016.csv','r') as f:
    bbc = f.readlines()
```

In [2372]:
```
#convert to string and remove newlines
mystring = ''.join(bbc)
bbc_string =mystring.replace('\n',"")
```

In [2373]:
```
#convert string to list
bbc_list = bbc_string.split(",")
```

In [2374]:
```
#Remove fields not wanted
for item in bbc_list:
    if item == 'BBC Children in Need':
        bbc_list.remove(item)
    if item == 'GBP':
        bbc_list.remove(item)
```

In [2375]:
```
#Regex expressions to remove code fields not wanted
import re

re1 = re.compile(r'360G\-CIN\-[a-z]+')

for item in bbc_list:
    match = re1.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2376]:
```
#Regex expressions to remove code fields not wanted
re2 = re.compile(r'360G\-CIN\-[0-9]+')

for item in bbc_list:
    match = re2.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2377]:
```
#Regex expressions to remove code fields not wanted
re3 = re.compile(r'[A-Z]{1}\d{7}')
for item in bbc_list:
    match = re3.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2378]:
```
#Regex expressions to remove code fields not wanted
re4 = re.compile(r'[GB\-[CHC|COH|EDU|NIC|]\-[\w]+')
for item in bbc_list:
    match = re4.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2379]:
```
#Regex expressions to remove code fields not wanted
re5 = re.compile(r'[GB\-[SC]\-[\w]+')
for item in bbc_list:
    match = re5.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2380]:
```python
#Regex expressions to remove code fields not wanted
re6 = re.compile(r'Grant to [A-Za-z0-9\w]+')
for item in bbc_list:
    match = re6.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2381]:
```python
#Regex expressions to remove code fields not wanted
re7 = re.compile(r'#N\\A|#N\/A|LGD|LONB|MD|NMD|UA')
for item in bbc_list:
    match = re7.findall(item)
    if match:
        bbc_list.remove(item)
```

In [2383]:
```python
#Removing the Grant Type items

for item in bbc_list:
    if item == 'Small Grants':
        bbc_list.remove(item)
    if item == 'Main Grants':
        bbc_list.remove(item)
    if item == 'zPositive Destinations':
        bbc_list.remove(item)
    if item == 'zFun and Friendship':
        bbc_list.remove(item)
```

In [2384]:
```python
# remove headers for columns no longer needed (indexes 0,1,3,6,9,13,14) -
#in reverse order so that it doesn't need to be reindexed each time. (ref
#https://stackoverflow.com/questions/11303225/how-to-remove-multiple-indexes-from-a-list-at-the-same-time )

indexes = [0,1,3,4,6,9,10,13,14]
for index in sorted(indexes, reverse = True):
    del bbc_list[index]
```

In [2385]:
```python
#Rename the headers

bbc_list[0]='Name'
bbc_list[1]='Location'
bbc_list[2]='Description'
bbc_list[3]='Amount'
bbc_list[4]='Duration'
bbc_list[5]='Date'
```

In [2386]:
```python
#Use indexes to pick out each field. Eg, name, start at 0 to the end and pick each 6th element

name = bbc_list[0::6]
location = bbc_list[1::6]
description = bbc_list[2::6]
amount = bbc_list[3::6]
duration = bbc_list[4::6]
date = bbc_list[5::6]
```

In [2387]:
```python
print(amount)
```

```
, '9995', '9967', '9900', '10000', '9500', '3450', '9860', '3350', '7961', '4915', '9861', '9989', '9000', '9660',
'10000', '7080', '5000', '8236', '10000', '9996', '9752', '9500', '8749', '9700', '5000', '7898', '6690', '9985', '
7951', '9912', '9904', '9728', '10000', '10000', '10000', '7733', '3000', '9995', '3838', '10000', '9600', '6000',
'9820', '9865', '8605', '9454', '9900', '10000', '9000', '9975', '10000', '6000', '10000', '10000', '8000', '9450',
```

In [2389]:
```
#Remove items which are blank from location list
indexes = [-1,-2,-3]
for index in sorted(indexes):
    del location[index]
```

In [2390]:
```
#Remove items which are blank from amount list
indexes = [-1,-2]
for index in sorted(indexes):
    del amount[index]
```

In [2391]:
```
#Remove headers
del amount[0]
del location[0]
```

In [2392]:
```
#convert number strings to integers https://stackoverflow.com/questions/23375606/converting-list-items-from-
#string-to-intpython#/23375633

for i in range(len(amount)):
    amount[i] = int(amount[i])
```

In [2393]:
```
#check length is the same
print(len(location))
print(len(amount))
```

```
1693
1693
```

In [2394]:
```
#create dataframe and view top 5 lines
import pandas as pd
df = pd.DataFrame()
df['Location'] = location
df['Amount'] = amount
print(df.head())
```

```
              Location  Amount
0    Rhondda Cynon Taf   26468
1          Dundee City   80340
2           Portsmouth   29745
3             Knowsley   53694
4  Newcastle upon Tyne   75000
```

In [2395]:
```
#check total agrees to spreadsheet
df.Amount.sum()
```

Out[2395]:
```
94753818
```

In [2399]:
```
#sort by Location
df.sort_values('Location')
```

```
              Location  Amount
0    Rhondda Cynon Taf   26468
1          Dundee City   80340
2           Portsmouth   29745
3             Knowsley   53694
4  Newcastle upon Tyne   75000
```

```
In [2400]: # adding amounts by location  ref: https://github.com/pandas-dev/pandas/issues/13821
df.groupby(['Location'], as_index=False).sum()
```

Out[2400]:

	Location	Amount
0	Aberdeen City	285371
1	Allerdale	254644
2	Antrim and Newtownabbey	123460
3	Argyll and Bute	357643
4	Armagh Banbridge and Craigavon	549068
5	Ashfield	59587
6	Aylesbury Vale	226874
7	Barking and Dagenham	331072
8	Barnet	506148
9	Barnsley	102612
10	Barrow-in-Furness District	214967
11	Basildon District	99952
12	Basingstoke and Deane District	41050
13	Bassetlaw	94528
14	Bath and North East Somerset	268954
15	Bedford	219730
16	Belfast	2986363
17	Bexley	93030

```
In [2401]: #Export to csv
df.to_csv('bbcappeal.csv')
```

Appendix 4 – Regular Expressions

['Identifier', 'Recipient Org:Identifier', 'Recipient Org:Name', 'Title', 'Recipient Org:Location:Geographic Code Typ
e', 'Recipient Org:Location:Name', 'Recipient Org:Location:Geographic Code', 'Description', 'Amount Awarded', 'Curren
cy', 'Grant Programme:Title', 'Planned Dates:Duration (months)', 'Award Date', 'Funding Org:Name', 'Funding Org:Ident
ifier', 'Last modified 360G-CIN-64988', '360G-CIN-nipitinthebud', 'Nip It In The Bud', 'Grant to Nip It In The Bud',
'UA', 'Rhondda Cynon Taf', 'W06000016', 'This project will provide a programme of activities for children and young p
eople in an area of high deprivation. This will develop their personal and social skills and provide positive opportu
nities and role models.', '26468', 'GBP', 'Small Grants', '36', '01/11/2011', 'BBC Children in Need', 'GB-CHC-802052'
, '2016-10-02T18:00:00Z 360G-CIN-68854', 'GB-SC-SC038627', 'Boomerang Community Centre', 'Grant to Boomerang Communit
y Centre', 'UA', 'Dundee City', 'S12000042', 'This project will run youth clubs for children who live in an area of l
ittle opportunity and could otherwise be involved in anti-social behaviour or substance abuse. The project will incre
ase their confidence and integrate them into the community.', '80340', 'GBP', 'Main Grants', '36', '01/02/2013', 'BBC
Children in Need', 'GB-CHC-802052', '2016-10-02T18:00:00Z 360G-CIN-68857', '360G-CIN-connorstoylibraries', 'Connors T
oy Libraries', 'Grant to Connors Toy Libraries', 'UA', 'Portsmouth', 'E06000044', 'This project provides two communit
y based toy library sessions each week for pre-school children in a deprived area. Through play the children are happ
ier and develop better social skills.', '29745', 'GBP', 'Small Grants', '36', '01/02/2013', 'BBC Children in Need', '
GB-CHC-802052', '2016-10-02T18:00:00Z 360G-CIN-68881', '360G-CIN-homestartknowsley', 'Home-Start Knowsley', 'Grant to
Home-Start Knowsley', 'MD', 'Knowsley', 'E08000011', 'The project will train volunteers to undertake home visits to s
upport families with young children.', '53694', 'GBP', 'Main Grants', '36', '01/02/2013', 'BBC Children in Need', 'GB
-CHC-802052', '2016-10-02T18:00:00Z 360G-CIN-68914', '360G-CIN-newcastleunitedfoundation', 'Newcastle United Foundati
on', 'Grant to Newcastle United Foundation', 'MD', 'Newcastle upon Tyne', 'E08000021', 'The project will provide a 36
week programme of football fun clubs in 10 different venues including fitness and mobility work throwing passing bal
l control and shooting skills for young disabled people', '75000', 'GBP', 'Main Grants', '36', '01/02/2013', 'BBC Chi
ldren in Need', 'GB-CHC-802052', '2016-10-02T18:00:00Z ']

Target 1: GB-CHC-802052 **Target 2:** Any of MD, LGD, LOND, NMD, UA, #N/A or #N\A **Target 3**: E06000044

Regex: [r'GB\-[CHC|COH|EDU|NIC|] \-[\w] +')

Note: Match 'GB- 'with a backslash before the hyphen to escape it.

[CHC|COH|EDU|NIC] – For each of the characters which can be found after the hyphen, with a pipe character to indicate 'or', so any of these combinations will satisfy the expression. Another hypen with backslash, then '\w to indicate a match with any alpha-numeric character. The '+' sign is a 'greedy match', so one or more matches with any character to the left of it will be included. Note: the '\w' is used here instead of '\d' for digits or [0-9] in case letters have been used for numbers eg. For zero and O.

Regex: (r'#N\\A|#N\/A|LGD|LONB|MD|NMD|UA')

Note: To match all the codes used, the pipe character is used as before. Since one of the codes is #N/A and had been entered as #N\A in some cases, both were included, and each slash or backslash had to be escaped

Regex: (r'[A-Z] {1} \d {7}')

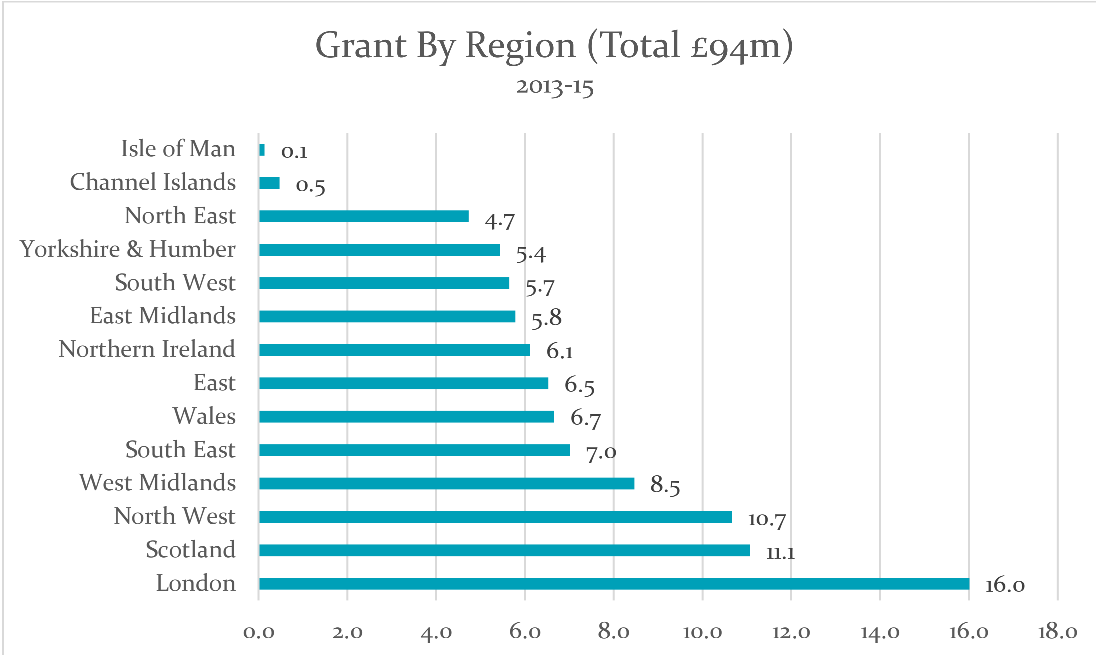
Note: Match any capital letter once, indicated by the number in curly brackets, then any seven-digit number.

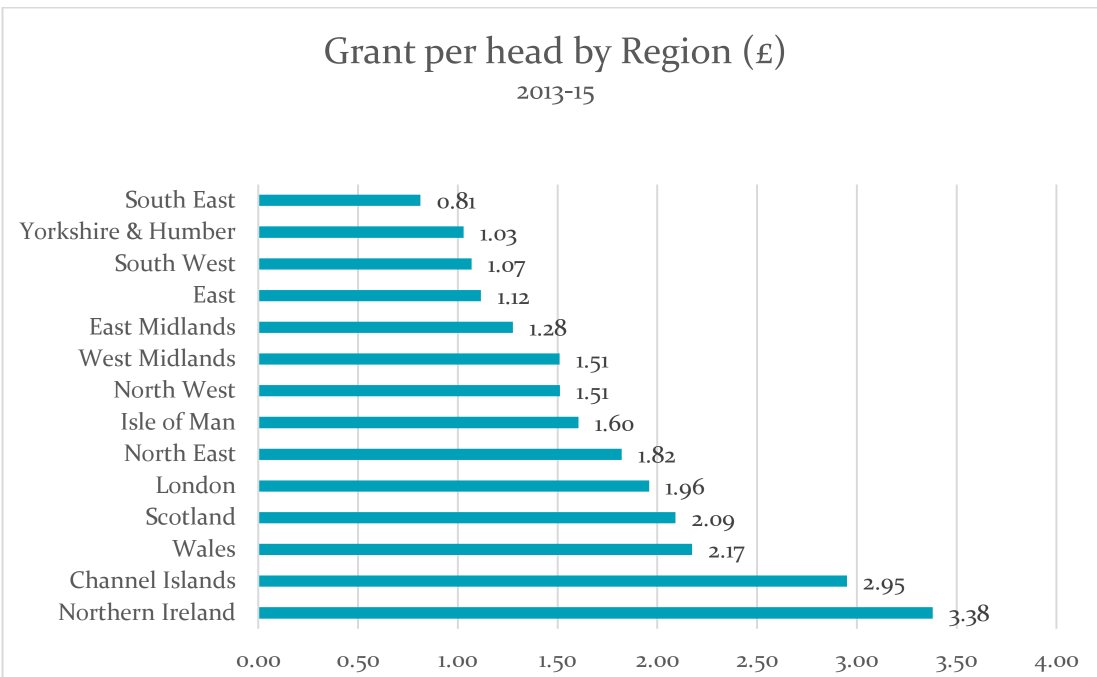
After the Regular Expressions were run, the cleaned text is shown below.

['Identifier', 'Recipient Org:Identifier', 'Recipient Org:Name', 'Title', 'Recipient Org:Location:Geographic Code Typ
e', 'Recipient Org:Location:Name', 'Recipient Org:Location:Geographic Code', 'Description', 'Amount Awarded', 'Curren
cy', 'Grant Programme:Title', 'Planned Dates:Duration (months)', 'Award Date', 'Funding Org:Name', 'Funding Org:Ident
ifier', 'Nip It In The Bud', 'Rhondda Cynon Taf', 'This project will provide a programme of activities for children a
nd young people in an area of high deprivation. This will develop their personal and social skills and provide positi
ve opportunities and role models.', '26468', '36', '01/11/2011', 'Boomerang Community Centre', 'Dundee City', 'This p
roject will run youth clubs for children who live in an area of little opportunity and could otherwise be involved in
anti-social behaviour or substance abuse. The project will increase their confidence and integrate them into the comm
unity.', '80340', '36', '01/02/2013', 'Connors Toy Libraries', 'Portsmouth', 'This project provides two community bas
ed toy library sessions each week for pre-school children in a deprived area. Through play the children are happier a
nd develop better social skills.', '29745', '36', '01/02/2013', 'Home-Start Knowsley', 'Knowsley', 'The project will
train volunteers to undertake home visits to support families with young children.', '53694', '36', '01/02/2013', 'Ne
wcastle United Foundation', 'Newcastle upon Tyne', 'The project will provide a 36 week programme of football fun club
s in 10 different venues including fitness and mobility work throwing passing ball control and shooting skills for y
oung disabled people', '75000', '36', '01/02/2013', '2016-10-02T18:00:00Z ']

Appendix 5 – BBC Children in Need Descriptive Analysis

(based on cleaned data and using Census 2011 population by region)



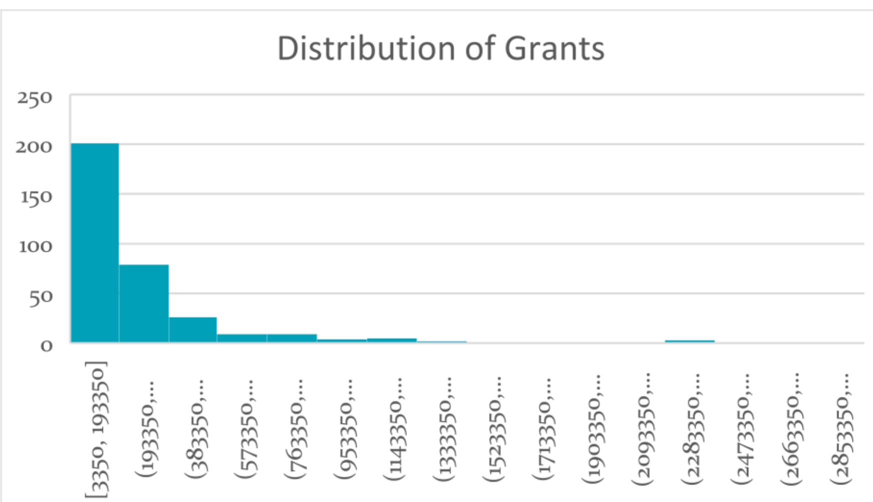


Average £1.50/head

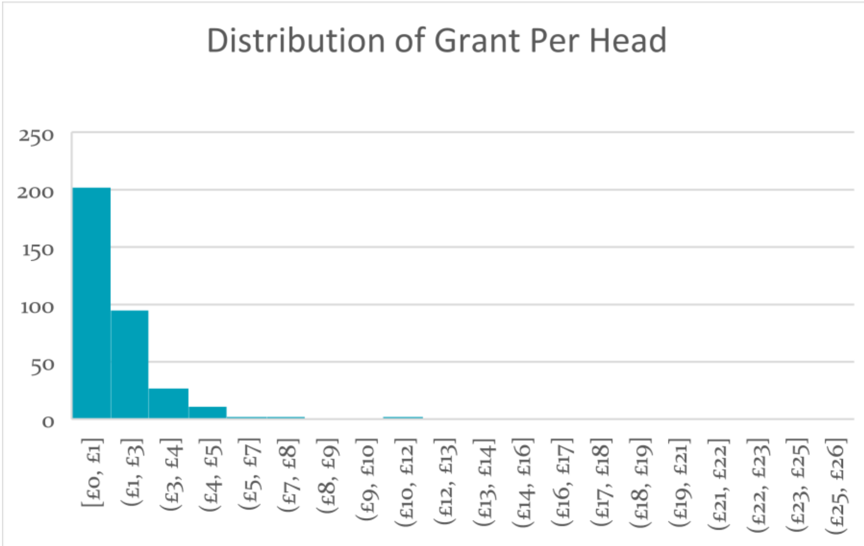
Appendix 5 continued

Top 10		**Bottom 10**	
Camden	£10.86	East Renfrewshire	£0.04
Belfast	£10.63	Calderdale	£0.05
Islington	£7.69	Chiltern	£0.05
Derry and Strabane	£6.73	Blaby	£0.05
Norwich	£6.66	New Forest	£0.06
Burnley District	£5.33	Telford and Wrekin	£0.06
Dundee City	£5.26	West Berkshire	£0.07
Hackney	£5.01	Havant District	£0.08
Carlisle	£4.92	Teignbridge	£0.08
City of Edinburgh	£4.89	Inverclyde	£0.08
Total Towns/Cities			342
Mean Grant			£277,034
Median Grant			£162,459
Mode Grant			£10,000
Max Grant Given			£2,986,363
Min Grant Given			£3,350
Std Deviation			£389,608
Semi interquartile range			£100,847
Average (£/Head)			£1.50
Std Deviation (£/Head)			£1.37
Semi interquartile range			£0.75

Distribution is positively skewed so the semi-interquartile range is a better measure of dispersion.



Histogram 1: Most grants given are up to £200,000



Most grants given are up to £1 per head

Analysis of Project Duration by Region.

			months		
Region	Count	12	24	36	Total
Channel Islands	6	0%	17%	83%	100%
East	104	19%	4%	77%	100%
East Midlands	99	30%	6%	64%	100%
Isle of Man	2	0%	0%	100%	100%
London	289	31%	6%	63%	100%
North East	84	29%	4%	68%	100%
North West	187	28%	6%	66%	100%
Northern Ireland	121	33%	7%	60%	100%
Scotland	185	28%	5%	67%	100%
South East	134	31%	2%	67%	100%
South West	114	42%	3%	55%	100%
Wales	112	29%	1%	70%	100%
West Midlands	159	32%	3%	65%	100%
Yorkshire & Humber	97	31%	3%	66%	100%
Total/Average	1693	30%	4%	65%	100%

Most projects are 12 and 36 months duration in total and across the regions

Appendix 6 – Review of McMillan Cancer Support Data Sources

Sourced from a Google Search

Sample of Variety of Data Sources for McMillan Cancer Support

Data Source	Figures/Details	Data Type	Collection Method
Twitter	669,000 followers	Tweets, photos, short videos, direct messages	Analytics, API, manual data collection
Facebook (1)	724,000 likes	Comments, photos, video, direct messages	Analytics, API, manual collection.
Youtube	9,238 subscribers	Comments, direct messages	Youtube analytics, API, manual data collection
Instagram	56,200 followers	Comments, direct messages	Instagram insights, API, manual data collection
Snapchat	N/A	Snaps, videos, photos	No API, manual data collection
Pinterest	3,394 followers	Photos and comments	Analytics, API, manual collection
Flickr	276,700 views	Testimonials	Flikr Stats Page, API, manual collection
Linked In	24,139 followers	Likes, comments and connections	Analytics, API and manual collection
Website including forum	6 million hits of which 1.5 million forum interactions(2017)	Visitors, online survey, online shop, blog post comments, online forums,	Collection from site analytics, manual collection of forum, blog comments and surveys
Phone calls, emails and questions (via website)	152,000	Voice conversations, emails, contact forms	Call logs, email records, contact form submissions
Mobile information centres	181,000 people(2017)	Face to face feedback, comments and questions	Locally kept records, brochure and leaflet takeup
High Street Shops & partnerships	No available data	Retail data, business organisation data, feedback from customers	Databases, shop records, bank records, in store surveys, partly automated and partly manually collected
Events eg. World Coffee Morning, marathons etc	No available data	Feedback, after event surveys, photos, videos, news items	Manual collection
Local Information and Support Centres	138 centres helping 304,000 people (2017)	Client health data, client and family feedback, support staff feedback	Collection from statutory or other methods, survey responses
Local Groups	No available data	Local group records, support volunteer data and feedback from clients and staff	As above

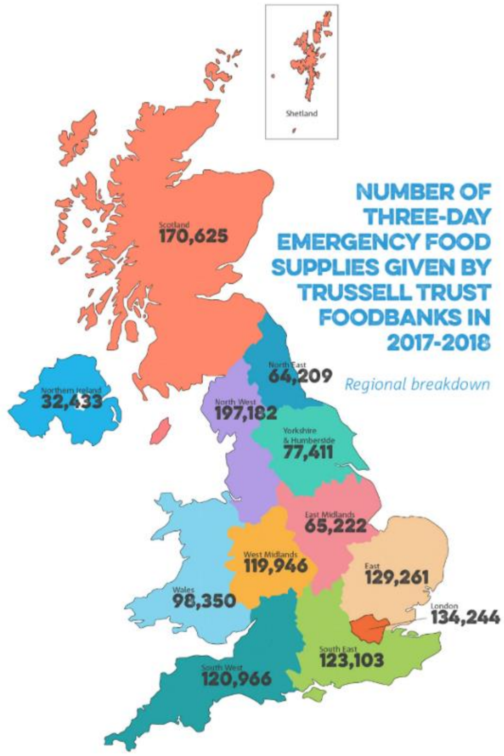
1.Also local groups on Facebook not included here

Appendix 7 – Machine Learning Models

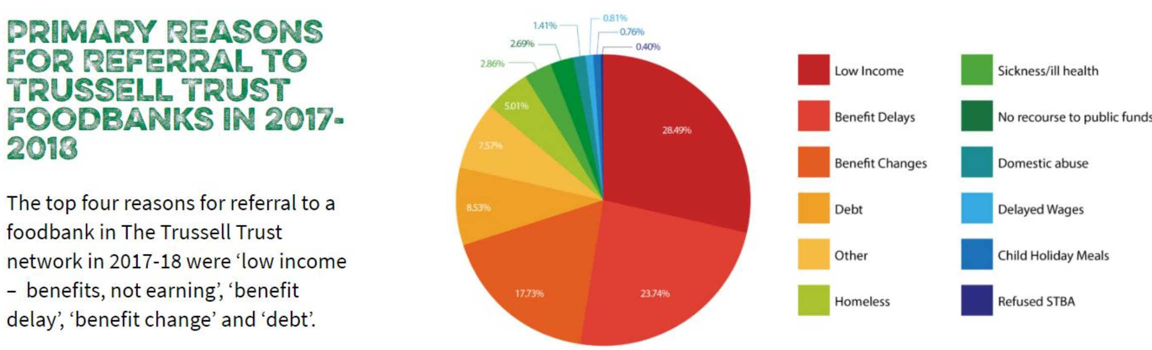
Types of Machine Learning Algorithms		
Regression Algorithms Ordinary Least Squares Regression (OLSR) Linear Regression Logistic Regression Stepwise Regression Multivariate Adaptive Regression Splines (MARS) Locally Estimated Scatterplot Smoothing (LOESS) 2. Instance-based Algorithms k-Nearest Neighbour (kNN) Learning Vector Quantization (LVQ) Self-Organizing Map (SOM) Locally Weighted Learning (LWL) 3. Regularization Algorithms Ridge Regression Least Absolute Shrinkage and Selection Operator (LASSO) Elastic Net Least-Angle Regression (LARS) 4. Decision Tree Algorithms Classification and Regression Tree (CART) Iterative Dichotomiser 3 (ID3) C4.5 and C5.0 (different versions of a powerful approach) Chi-squared Automatic Interaction Detection (CHAID) Decision Stump M5 Conditional Decision Trees	5. Bayesian Algorithms Naive Bayes Gaussian Naive Bayes Multinomial Naive Bayes Averaged One-Dependence Estimators (AODE) Bayesian Belief Network (BBN) Bayesian Network (BN) 6. Clustering Algorithms k-Means k-Medians Expectation Maximisation (EM) Hierarchical Clustering 7. Association Rule Learning Algorithms Apriori algorithm Eclat algorithm 8. Artificial Neural Network Algorithms Perceptron Back-Propagation Hopfield Network Radial Basis Function Network (RBFN) 9. Deep Learning Algorithms Deep Boltzmann Machine (DBM) Deep Belief Networks (DBN) Convolutional Neural Network (CNN) Stacked Auto-Encoders	10. Dimensionality Reduction Algorithms Principal Component Analysis (PCA) Principal Component Regression (PCR) Partial Least Squares Regression (PLSR) Sammon Mapping Multidimensional Scaling (MDS) Projection Pursuit Linear Discriminant Analysis (LDA) Mixture Discriminant Analysis (MDA) Quadratic Discriminant Analysis (QDA) Flexible Discriminant Analysis (FDA) 11. Ensemble Algorithms Boosting Bootstrapped Aggregation (Bagging) AdaBoost Stacked Generalization (blending) Gradient Boosting Machines (GBM) Gradient Boosted Regression Trees (GBRT) Random Forest 12. Other Algorithms Computational intelligence (evolutionary algorithms) Computer Vision (CV) Natural Language Processing (NLP) Recommender Systems Reinforcement Learning Graphical Models

Brownlee, J. (2018). *A Tour of Machine Learning Algorithms* . [online] Machine Learning Mastery.
Available at: https://machinelearningmastery.com/a-tour-of-machine-learning-algorithms/

Appendix 8 – Trussell Food Bank Statistics



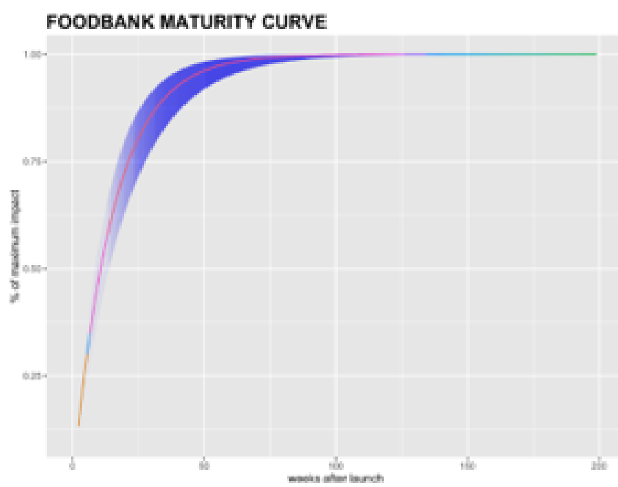
1,332,952 three-day emergency food supplies given to people in crisis.



(The Trussell Trust, 2018)

Appendix 9 – Predictive Forecast Models Used by Trussell Food Bank

1. A Hierarchical Bayesian Negative Binomial State Space Model to produce the food bank standard model to show the impact of bank maturity on food bank usage.
    - Bayesian approach – used as the data was non-linear with a lot of detail.
    - Hierarchical - Data from one food bank allowed inferences to be made about others. Seasonality factors and time taken to reach maturity is assumed to be have the same impact countrywide.
    - Negative Binomial - The number of people served was not modelled as a Poisson distribution because the mean was not the same as the variance, hence a negative binomial distribution was used to allow for the greater measure of dispersion.
    - State Space Model – is a mathematical model used for time series which considers all possible configurations. This was used as the model was observing food bank maturity over time.



2. A logistic regression model for predicting food bank demand for use in the mapping tool, which included:
    - Whether there is demand or not is based on 30 socio-economic variables and 11 food bank variables.
    - The dependent variable is predicted based on the independent variables.
    - Logistic regression is used instead of multiple linear as the dependent variable is either a 'yes' or 'no'.
    - Socio-economic variables included such factors as economic activity, people in the household, method and distance to work, social grade, household composition, ethnic group, religion, occupancy per bedroom etc.
    - Food bank data included postcode, number of children and adults, ethnicity, crisis type, age group, whether in employment.

Appendix 10 – Variables Used in Logistic Regression

Census Data 2011

QS103EW	Age by single year
QS104EW	Sex
QS105EW	Schoolchildren and full-time students at their non term-time address
QS106EW	Second address
QS108EW	Living arrangements
QS110EW	Adult lifestage (alternative adult definition)
QS111EW	Household lifestage
QS112EW	Household composition - People
QS113EW	Household composition - Households
QS114EW	Household composition (alternative child and adult definition) - People
QS115EW	Household composition (alternative child and adult definition) - Households
QS116EW	Household type
QS117EW	People aged 18 to 64 living in a one adult household
QS118EW	Families with dependent children
QS119EW	Households by deprivation dimensions
QS121EW	Armed Forces
QS201EW	Ethnic group
QS202EW	Multiple ethnic groups
QS203EW	Country of birth (detailed)
QS204EW	Main language (detailed)
QS205EW	Proficiency in English
QS206WA	Welsh language skills
QS207WA	Welsh language skills (detailed)
QS208EW	Religion
QS210EW	Religion (detailed)
QS211EW	Ethnic group (detailed)
QS212EW	Passports held
QS213EW	Country of birth (expanded)
QS301EW	Provision of unpaid care
QS302EW	General health
QS303EW	Long-term health problem or disability
QS401EW	Accommodation type - People
QS402EW	Accommodation type - Households
QS403EW	Tenure - People
QS404EW	Tenure - Household Reference Person aged 65 and over
QS405EW	Tenure - Households
QS406EW	Household size
QS407EW	Number of rooms
QS408EW	Occupancy rating (rooms)

QS409EW	Persons per room - Households
QS410EW	Persons per room - People
QS411EW	Number of bedrooms
QS412EW	Occupancy rating (bedrooms)
QS413EW	Persons per bedroom - Households
QS414EW	Persons per bedroom - People
QS415EW	Central heating
QS416EW	Car or van availability
QS417EW	Household spaces
QS418EW	Dwellings
QS419EW	Position in communal establishment
QS420EW	Communal establishment management and type - Communal establishments
QS421EW	Communal establishment management and type - People
QS501EW	Highest level of qualification
QS502EW	Qualifications gained
QS601EW	Economic activity
QS602EW	Economic activity of Household Reference Person
QS603EW	Economic activity - Full-time students
QS604EW	Hours worked
QS605EW	Industry
QS606EW	Occupation (Minor Groups)
QS607EW	NS-SeC
QS608EW	NS-SeC of Household Reference Person - People aged under 65
QS609EW	NS-SeC of Household Reference Person - People
QS610EW	NS-SeC of Household Reference Person (HRP) - HRP Aged under 65
QS611EW	Approximated Social Grade
QS612EW	Year last worked
QS613EW	Approximated social grade

QS701EW	Method of travel to work
QS702EW	Distance travelled to work
QS703EW	Method of Travel to Work (2001 specification)
QS801EW	Year of arrival in the UK
QS802EW	Age of arrival in the UK
QS803EW	length of residence in the UK

Food Bank Data

Field	**Definition**
voucherNo	Voucher unique identifier
referringAgency	Agency issuing the voucher
foodbankID	Food bank unique identifier
date	Date the food bank voucher is used by the client
clientPostCode	Six digit post code of client (or no fixed address)
noAdults	Number of adults in the household to be fed by the food package
noChildren	Number of children in the household to be fed by the food package
crisisType	One of: Benefit changes, Benefit delays, Child holiday meals, Debt, Delayed wages, Domestic violence, Homeless, Low income, Sickness, Unemployed, Other
ethnicity	One of: White, Mixed, Asian, Black, Chinese, Other
ageGroup	16-24, 25-64, over 65
paidEmployment	Y/N