



# Recycling in England

ENV1 Group

Srijan Chhabda, Nelly Deprince, George Lancaster, Justin Salmon and Xiangtian Zheng

# Great Pacific Garbage Patch

- 7 million tons of weight
- Twice the size of Texas
- Up to 9 feet deep



# Aims and Objectives

---

**Understand people's recycling behaviour by:**

- Identifying causal factors that affect recycling rates
- Using those factors to predict future recycling rates

**Use predictions to motivate change**

# Research Questions

---

- Do more **densely populated** areas recycle more or less?
- Do the **collection rules** have an effect? (e.g. single bin for all recycling or separate bins for different types)
- Does the **collection frequency** (weekly/fortnightly) have an impact?
- Do councils that **spend more on recycling facilities** get better recycling rates?
- Do **wealthier** areas recycle more or less than **poorer** areas? (Index of Deprivation)
- Is there a relationship between **education** spending and recycling rate?

# Data Sets

---

Data comes from the Environment Agency and WasteDataFlow for all local authorities in England.

## From WasteDataFlow:

- Amount of waste collected
- Amount recycled (incl. type breakdowns)
- Population estimates
- Index of Deprivation

The logo for WasteDataFlow, featuring the text "WasteDataFlow" in white on an orange rectangular background. A faint circular graphic is visible behind the text.

WasteDataFlow

## From the Environment Agency:

- Local authority boundaries (shapefile)
- Local authority expenditure breakdown



# Data Wrangling

---

- Originally used summary .xls spreadsheets from EA
  - However, these were inconsistent and led to some horrendous ingestion code
- WDF has no API, manual download and parse required
  - Format is thankfully consistent year-on-year
  - Data is in an odd denormalised format, requiring some pivot table acrobatics in Pandas
- Problem: Local authorities are not static
  - Several major structural changes from 2006 - 2019
  - Careful to avoid double counting



# Data Storage

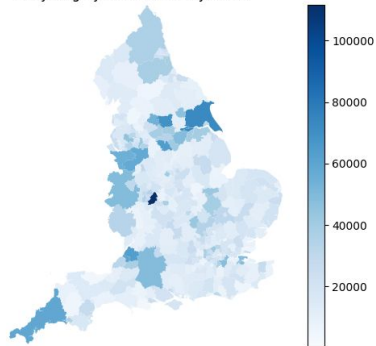
- Why - no need to pre-process every time, much faster
- How
  - Oracle cloud + MySQL Database
  - Data includes many attributes, which means tabular database can make sense;
  - Drop the irrelevant attributes;
  - Different datasets;
  - Different years data;
- What

authority	authority_type	index_of_deprivation	population	population_density	recycling	residual	year
Adur	NMD	17.0100002288818	59100	14	5304.26	17780.81	2006
Allerdale	NMD	22.9200000762939	96300	0	15543.186	44918.4	2006
Amber Valley	NMD	18.8899995422363	118600	4.12	10650.797	35458.44	2006
Arun	NMD	15.5600004196167	144500	6	18272.511	35442.55	2006
Ashfield	NMD	27.7700004577637	114000	10	12345.75	41441.79	2006
Ashford	NMD	13.4399996852875	110000	1.2225	8833.66	36698.75	2006
Aylesbury Vale	NMD	8.30000014305115	168100	1.215	12077.42	53354.08	2006
Babergh	NMD	11.289999961853	85100	1	13302.56	23018.37	2006
Barking and Dagenham	LBO	31.3199996948242	164500	45	20728.232	81433.7	2006
Barnet	LBO	16.0900001525879	329700	37	42313.161	132870.731	2006
Barnsley	MD	32.9900016784668	222100	6	33522.5	98183	2006
Barrow-in-Furness	NMD	32.9799995422363	70100	8	5789.7	25525.2	2006

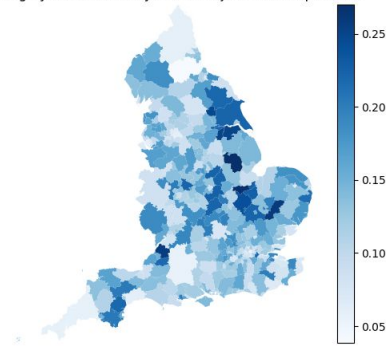
# Exploratory Data Analysis

2006

Recycling by Local Authority, 2006

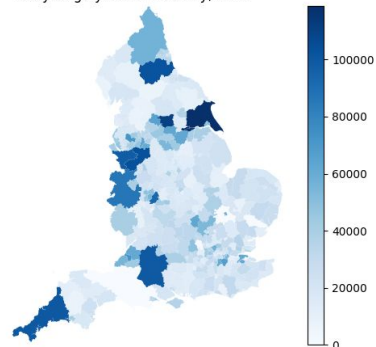


Recycling by Local Authority, 2006 (Adjusted for Population)

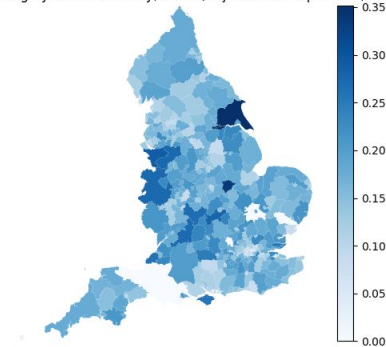


2017

Recycling by Local Authority, 2017



Recycling by Local Authority, 2017 (Adjusted for Population)

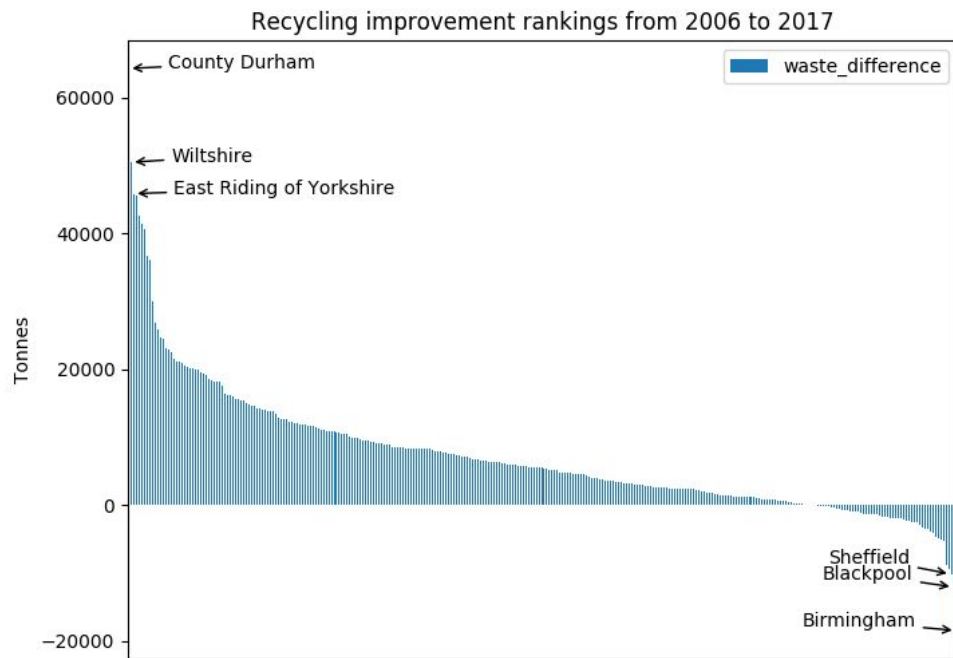


[Interactive Version](#)

Backed by Flask based REST API hosted on Oracle Cloud



# Improvement Rankings



# How can we use the data?

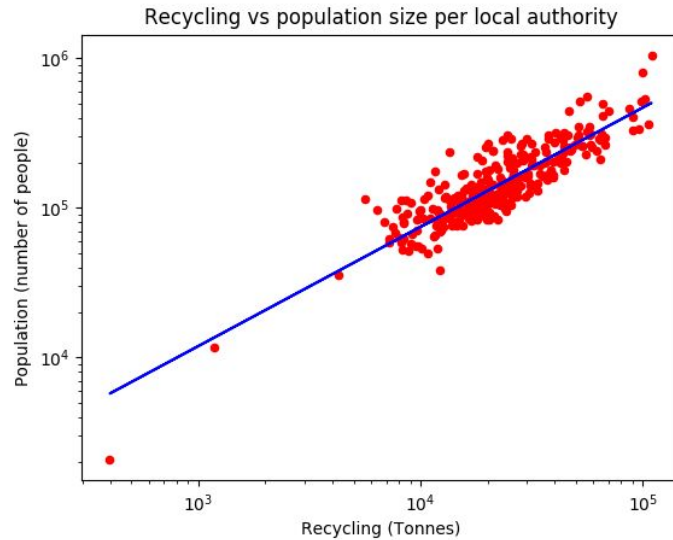
---

- Can we use data to make some predictions about future recycling rates?
  - Pearson correlation coefficient
    - Population
    - Environmental spending
- } Recycling rates

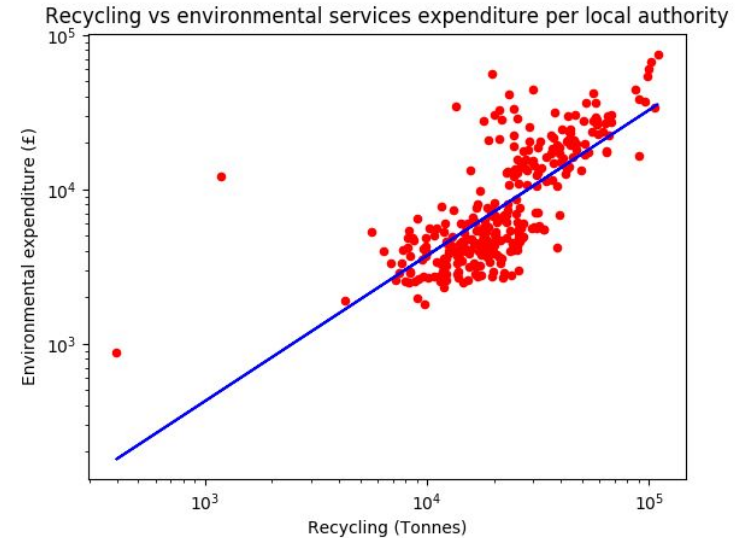
# Variable Correlation

- Effect of population and public spending

Correlation: 0.85



Correlation: 0.78



# Prediction

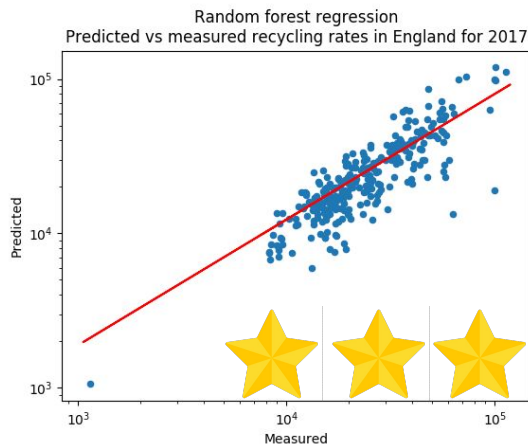
---

- Three regression models:
  - Multiple linear regression
  - Random forest
  - Neural network
- Population and environmental spending as inputs
- Used data from 2006-2016 to train
- Tested on 2017 data

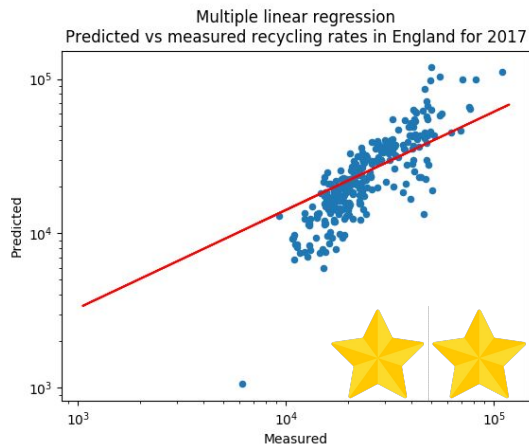
# Results

- Model evaluation
  - $R^2$  score
  - Root mean squared error

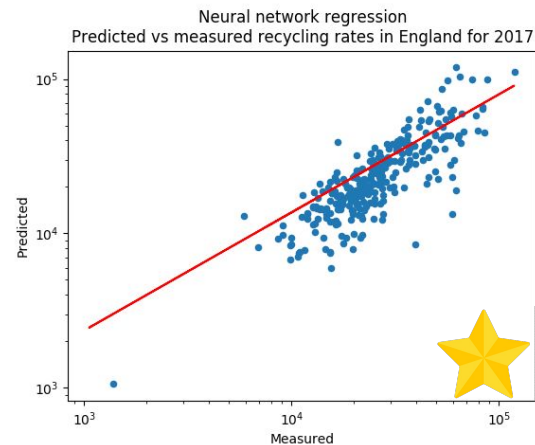
## Random Forest



## Multiple Linear Regression



## Neural Network



# Conclusions

- Authorities that spend more on recycling have higher recycling rates.
- We can exploit this to make predictions on future recycling rates.





# Future Work

---

- Can we identify more predicting factors?
- How do our predictions line up with official government predictions?
- Can we extend our work to the rest of the U.K?
- What about the rest of Europe, or even the world?

# **Any Questions?**

Thank you!