EARL conference

insight Hub

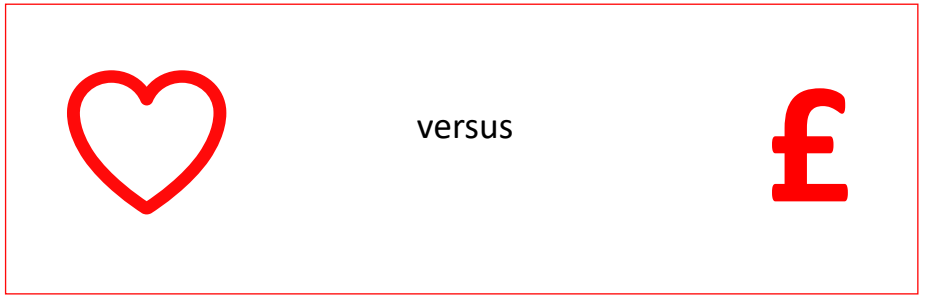
For Love or Money?

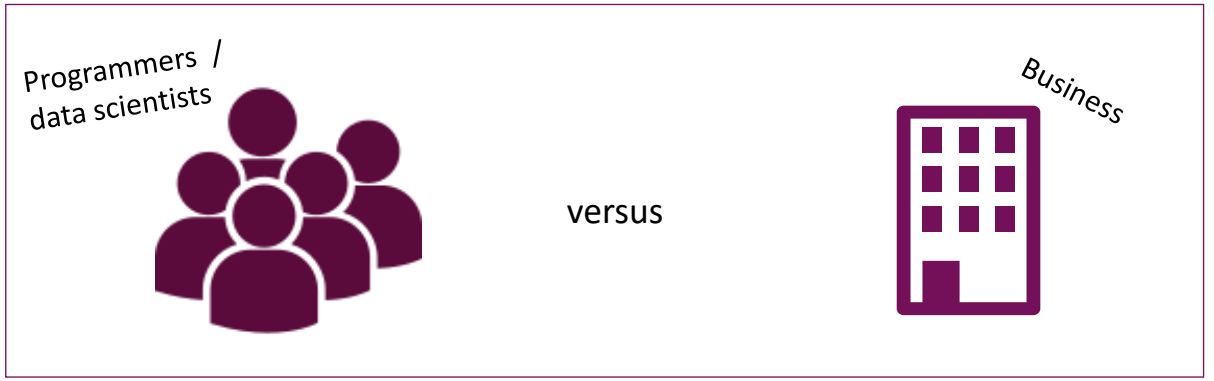
How to generate high revenue from a team of R developers

Dr Lisa Clark

Principal Data Science Manager

The tension between fun and value?





A misnomer!

I will show this does not have to be the case using:

- Experience
- Examples

Use it to influence upwards, sideways and/or downwards!

It’s just me, me, me…

MPhys Physics with Nuclear Astrophysics

PhD Theoretical Physics (Cosmology)

Postdoc in Cosmology

Founded and ran 3 companies

Research Fellow

- Solar Panels
- Renewable Energy

Economic Policy Analyst

Data Scientist

Principal Data Science Manager at VM

Data
Programming
Analysis
Business Reqs

Working at the Sheffield Solar Farm, The University of Sheffield

Virgin Media at a glance…

LIBERTY GLOBAL

One of the world's leading converged video, broadband and communication companies

14.9m

homes passed

3.1m

mobile customers

45k*

businesses served

5.9m

cable customers

*This doesn't include our SoHo customers in the UK and Ireland

My experience…

…of leading analytical teams

Business 

- Demands ROI…but how should VALUE be determined?
- Often wants quick solutions
- Views skilled analysts and data scientists as an expensive resource
- Doesn't always understand "AI"

Managers 

- Can struggle to define the "right" projects and priorities
- Are stuck between delivering high value and maintaining an engaged team
- Can find it difficult to engage a team of (mainly) introverts

Data Scientists 

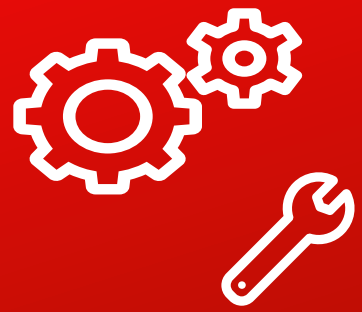
- Want to develop exciting code
- Want to learn new tricks
- Hate too much data engineering!
- Data science projects aren't quick!

What makes a programmer happy?

A computer
(of their choice)

High Speed Internet

A problem to solve

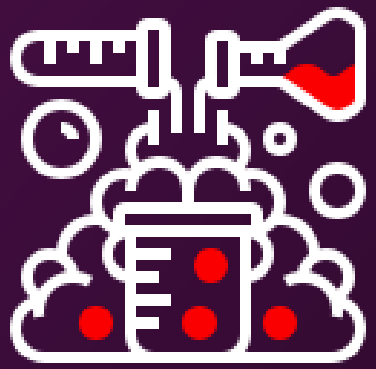


An empty room!

What tricks have I learned to get my team engaged?

- Managers need to know all the individual personalities and their irritation points
- For me personally: Listen more!

- “Capability Building” = Provide Learning & Development opportunities



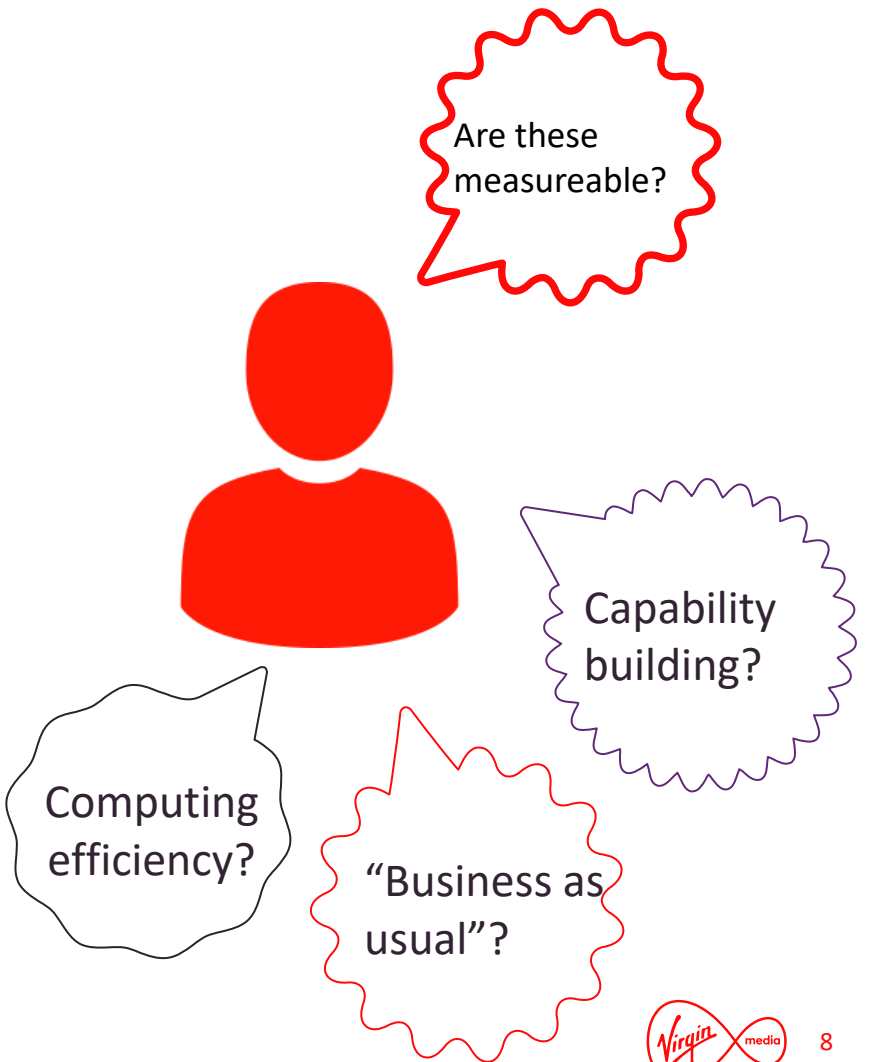
Give a data science team space and time with the right tools

- Use the correct tools & remove IT barriers
- Fewer meetings!
- Design the environment to feel more academic
- Initiate “Lunch & Learn” sessions

- During Covid19: games sessions during work hours!

Definition of value?

Value is in the eye of the beholder



Tension between code and ROI

When is code “good enough”?

```
for (i in 1:100){
  if(i%%3 == 0 & i%%5 == 0) {
    print('FizzBuzz')
  }
  else if(i%%3 == 0) {
    print('Fizz')
  }
  else if (i%%5 == 0){
    print('Buzz')
  }
  else {
    print(i)
  }
}
```

- Quick solution
- Possibly a little “dirty”
- Does the job!

```
# define the function
fizz_buzz <- function(x) {
  x[as.numeric(x)%%3==0 & as.numeric(x)%%5==0] <- "FizzBuzz"
  x[as.numeric(x)%%3==0] <- "Fizz"
  x[as.numeric(x)%%5==0] <- "Buzz"
  x
}

# apply it to the numbers 1 to 100
fizz_buzz(1:100)
```

- Use of a function might be better for production code
- Easier for others to understand (possibly)
- Scalable
- Does the job!

So how to obtain value and make it fun?

Integrate all aspects above into day to day project work!

The “right” projects

Environmental set-up

Capability build

Learning & Development

The Knapsack Algorithm

Case Study

Capital Expendi… The knapsack problem …ch

Requirement:

1. Maximise the revenue from projects whilst keeping within Capital Expenditure (CAPEX) budget
2. Each project has a given installation cost and associated revenue

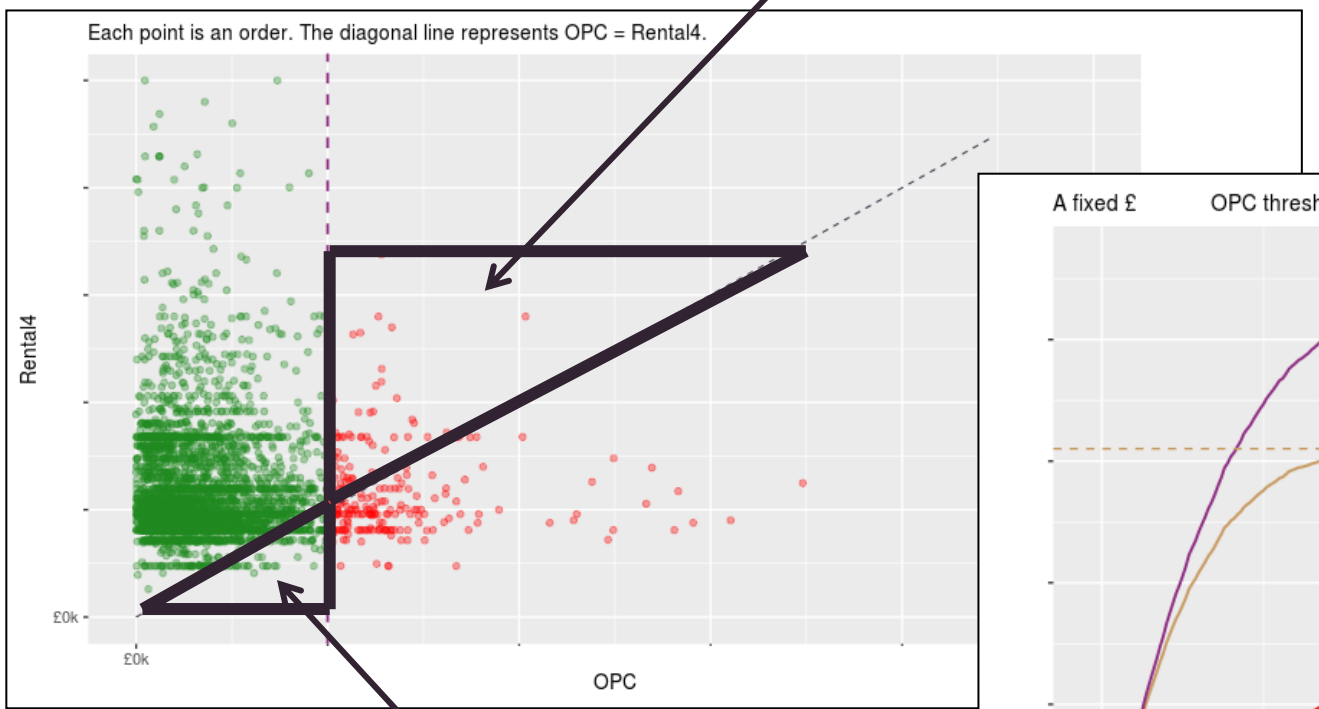
Business question:

- Apply a simple threshold for each project
- Above this threshold, projects might be cancelled
- What would be the best threshold?

Not all projects can be implemented within CAPEX budget so need to prioritise

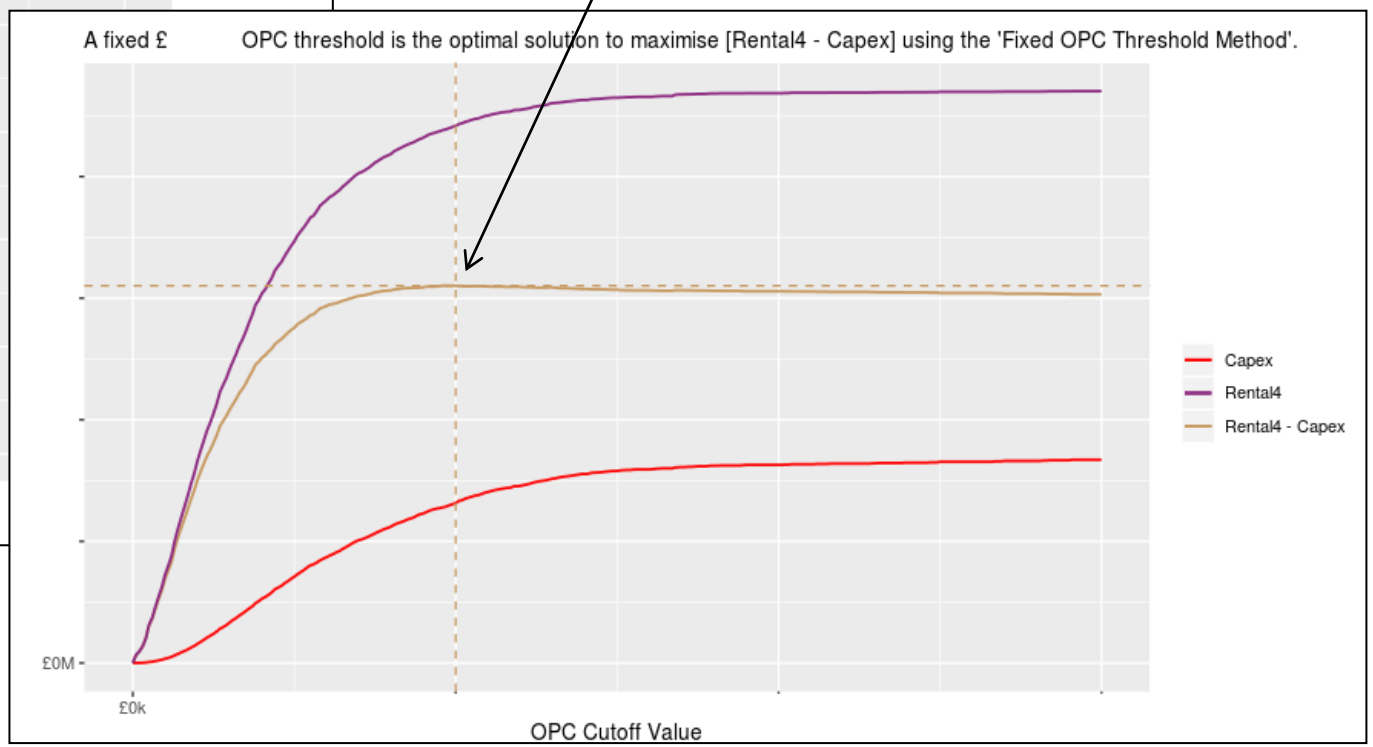
Simple CAPEX threshold

Some “profitable” projects are excluded!



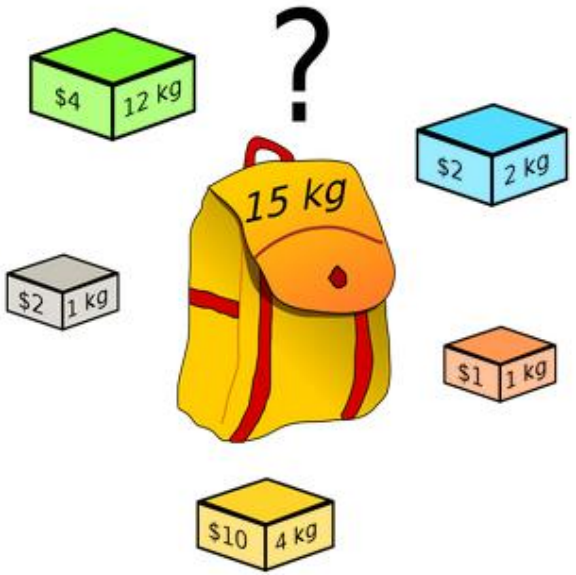
But loss making projects are still allowed!

Threshold defined by plateau



Knapsack Model

Using the Adagio package



https://en.wikipedia.org/wiki/Knapsack_problem

The **knapsack problem** or rucksack problem is a problem in combinatorial optimization.

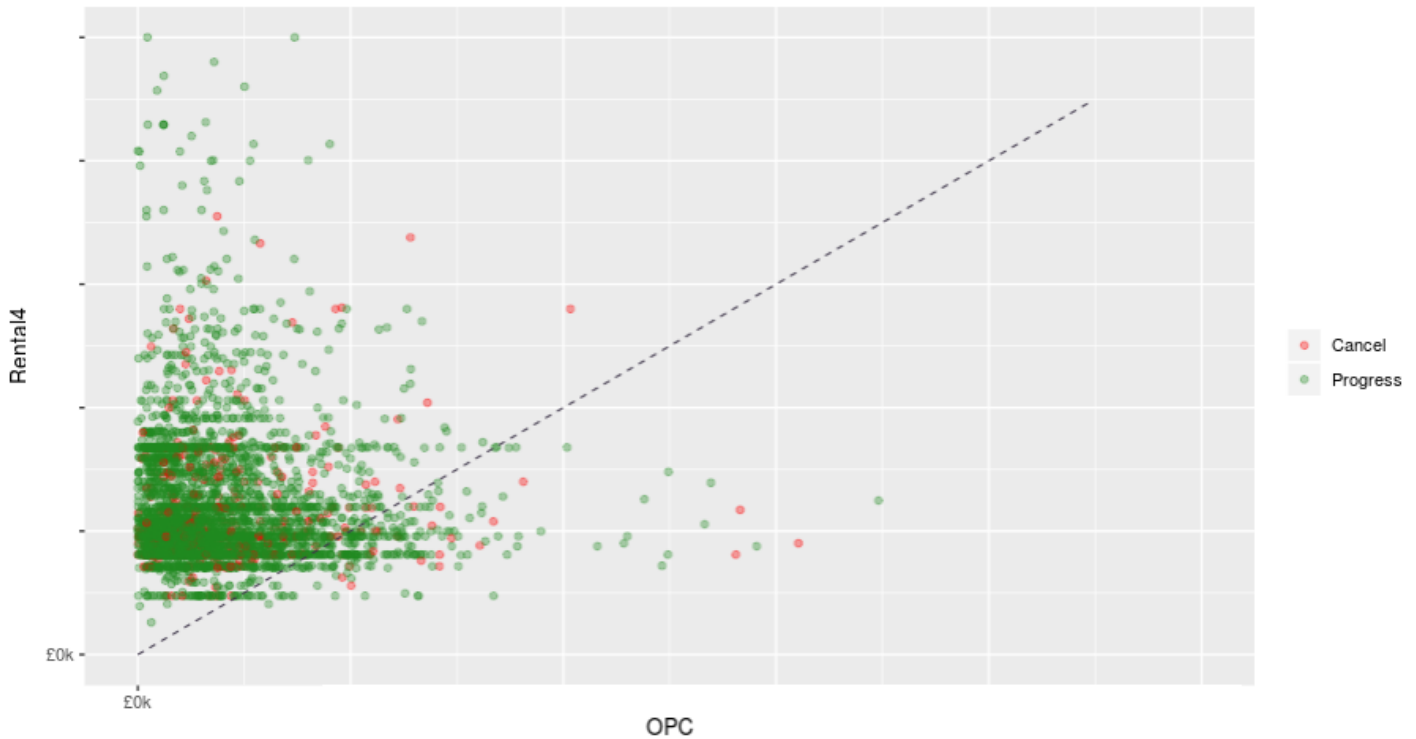
Given a set of items with a given size/weight and value, determine which items should be included in a collection in order to maximise the value by utilising the greatest total size/weight.

Given a set of projects with known CAPEX costs, maximise the Rental for a given CAPEX budget:

```
knapsack_orders <- adagio::knapsack(w = cost_df$capex,
                                    p = cost_df$rental,
                                    cap = capex_budget)
```

Knapsack Model Result

Using the Adagio package



Provides an upper limit to the potential "profit" given a set CAPEX limit

9.5% of projects cancelled

only 92 fewer projects vs. simple threshold

£1.5M "profit" gain keeping within CAPEX

£0.2M gain over simple threshold model

Clustering with CLARA

Case Study

~~Customer se~~ Clustering problem

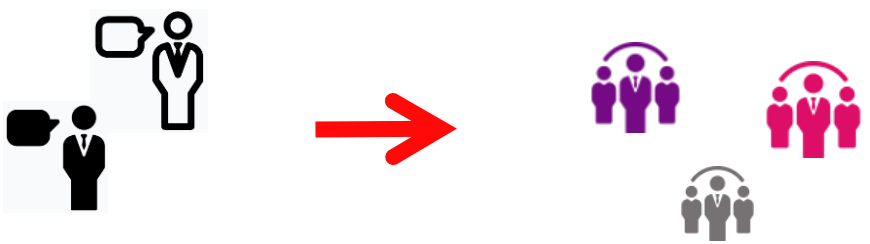
Requirement:

1. Identify groups of business customers who are alike in their characteristics
2. Use a holistic (non-human) approach to avoid human-led bias

Business question:

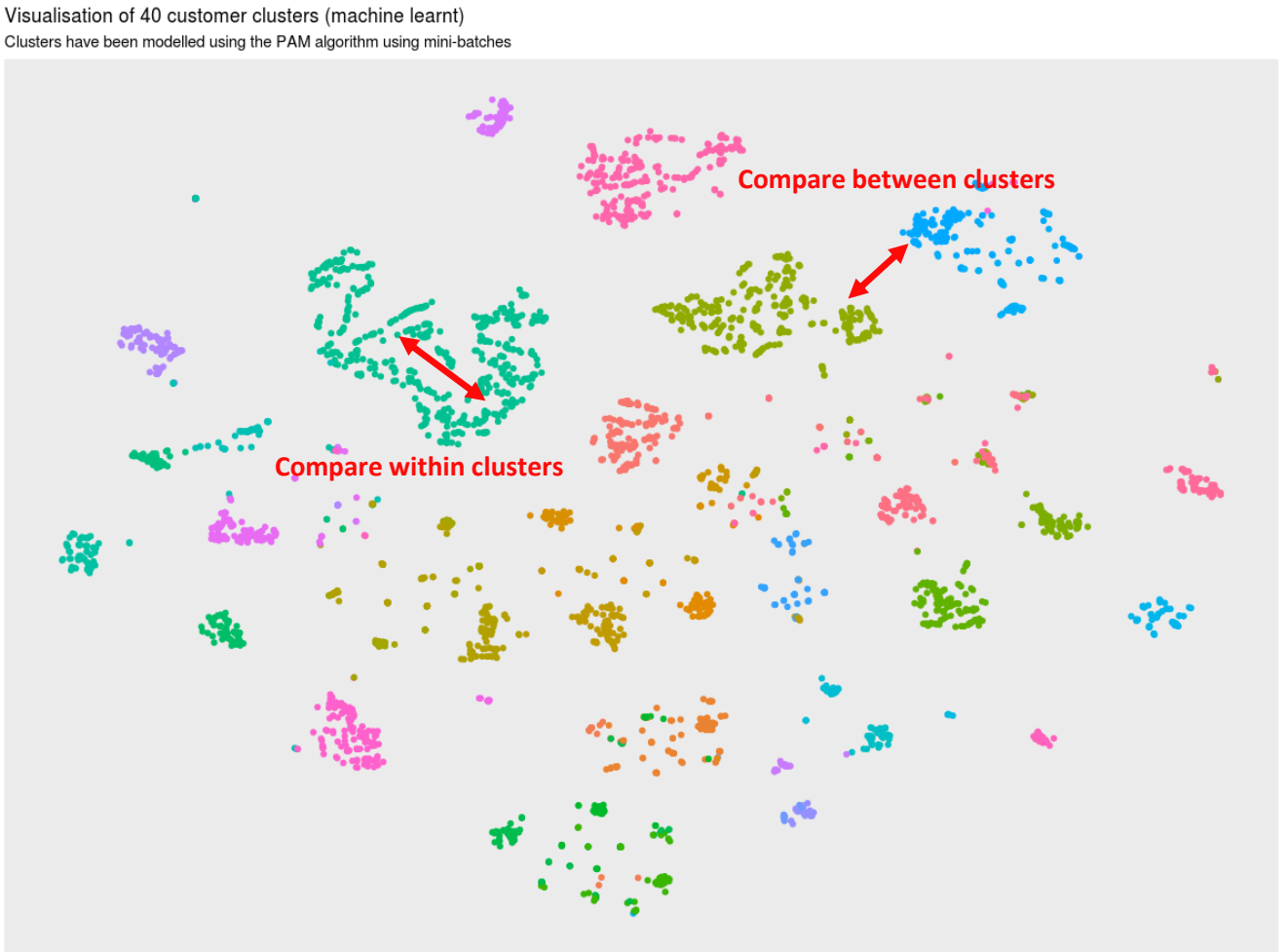
- Can we enhance our understanding of customers and their wants/needs
- Enable us to focus marketing campaigns towards specific customer groups

Allow the computer to learn similarities between customers not otherwise identified



Clusters: most relevant characteristics

Using packages: cluster & Rtsne



££ gained in high revenue marketing campaigns

Model run with 40 clusters

19 features

Used Gower distance to account for non-numeric features

Due to the large number of customers, we also developed a bespoke CLARA algorithm for clustering

```
gower_dist <- daisy(df, metric = "gower")
gower_mat <- as.matrix(gower_dist)
pam_fit <- pam(gower_dist,
               diss = TRUE,
               k = 40)
```

Visualisation

Case Study

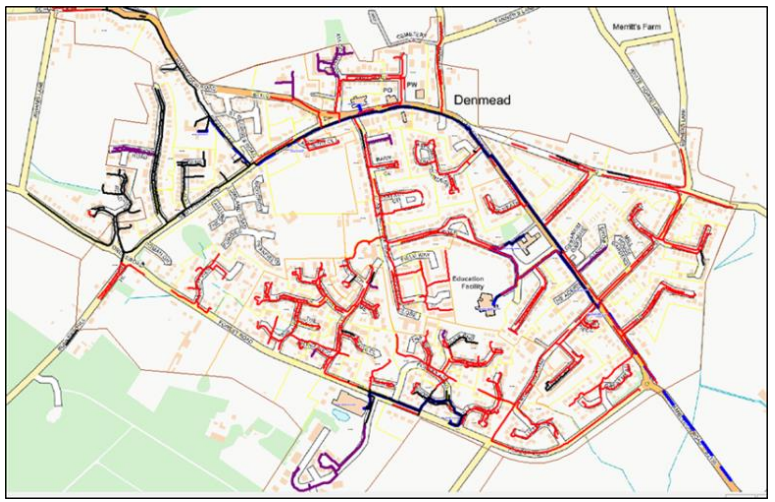
Presenting a message quickly

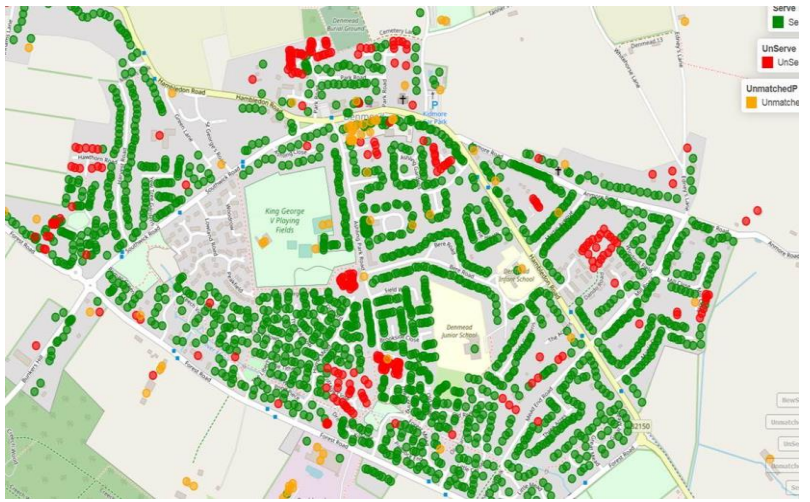
Using Leaflet (and Shiny) for premise serviceability

* Serviceable = if a premise can be serviced with VM products

Visually displaying serviceability on a map

- Quick to identify live network
- Visible premise serviceability
- Identification of “unusual” listings
- Identification of business clusters



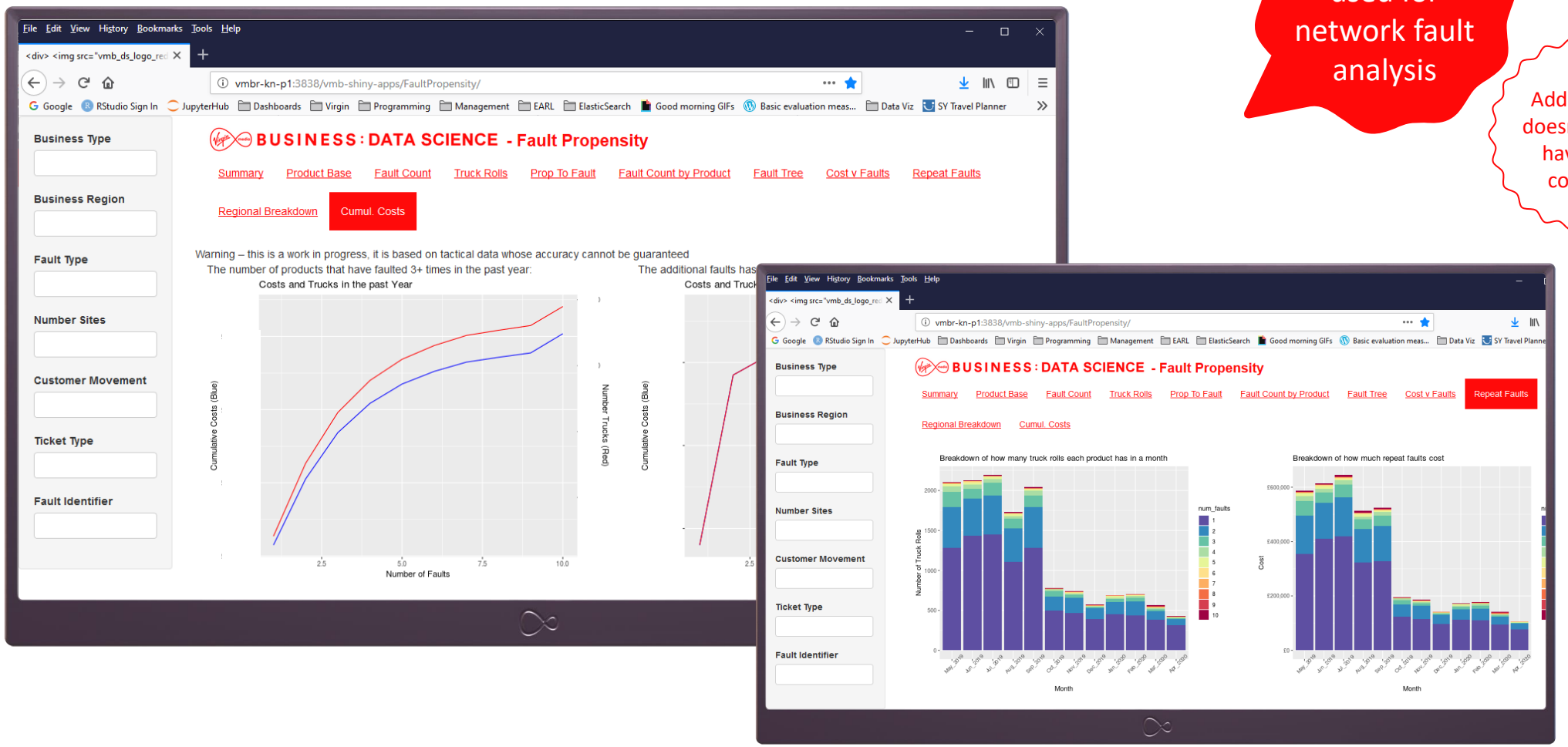


Using Shiny…

A good clear way of presenting dynamic data to non-analysts quickly

Successfully used for network fault analysis

Adding value doesn't always have to be complex!



Achieve high revenue gains by motivating the real asset:

The programmers!

Give a data science team <u>space</u> and <u>time</u> with the <u>right tools</u>

DAY TO DAY

Thank you!

Dr Lisa Clark

Principal Data Science Manager

Lisa.clark3@virginmedia.co.uk