Technology Strategy Board
Driving Innovation

Look before you leap

MODELLING PROPENSITY TO BUY FOR UK BUSINESSES

Level Business Limited / Nothumbria University

The need

Dealing with 'big data' in the business world only through traditional database techniques is almost impossible but coming to terms with it, and then making use of it, is vitally important.

Level Business provides UK company information so that businesses can make informed decisions about who they trade with and supply their services to. On its own raw data are not that useful; the focus of the internship was to find ways to process, visualise and analyse these data in order to provide value and utility both to the company and its users. In order to do this Level Business wishes to derive new information by combining and aggregating existing data sources, using this to motivate the search for new sources of data.

The outcomes

The project has given Level Business insight into their data sources. These include the patterns of how information is processed at Companies House, the first steps in visualising relationships between companies and departments is Local Authority spend data, implementing an algorithm to give a company director a 'rating' and combining data from Companies House and Local Authorities to provide new classifications of the type of business a company conducts.

The project has exposed the student to a number of new techniques for processing and manipulating data such as the querying language Hive and the use of Python for natural language processing and machine learning. It has also built on the student's existing knowledge of dimension reduction and graph/ network analysis and visualisation.

It has also allowed the student the opportunity to develop a greater understanding of the business information world, experience life in a start-up business as well as improving the student's ability to communicate ideas clearly to non-specialists in the subject area.

> *"[Helen] has brought with her experience in visualising big data and applied it to a number of real problems we have been looking to tackle. [The project] required clear thinking and accurate execution. Helen has been outstanding at both these."*
>
> **James Dobree**
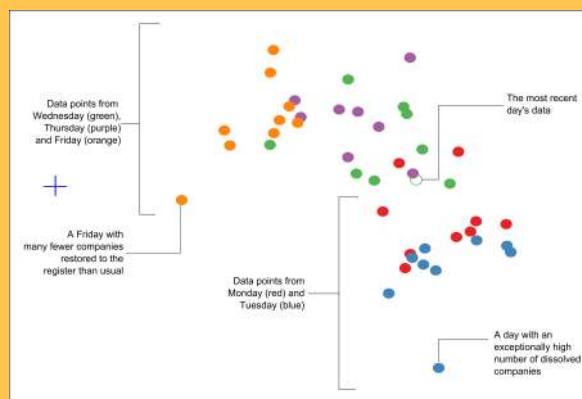> **Level Business Limited**

Technical summary

In the UK alone there are now over three million registered businesses with Companies House. Each day Level Business receives status updates on these companies; the first task involved visualising these status updates using a dimension reduction technique and finding patterns such as a distinct 'fingerprint' of update combinations for each day of the week.

Companies House is not the only source of UK company data - through Local Authority transparency spend data we can understand how companies are conducting their business. For the City of Westminster, creating dynamic networks of relationships between companies and departments, we were able to track how departments were trading with companies in terms of frequency and spend.

Further discussion, explanation and images can be found at this links http://bit.ly/fe6emK and http://t.co/SXPCZ9l.

Level Business also holds data on company directors. Combining director data, such as number and type and duration of positions held, and using the querying language Hive an initial model of how we may rate directors in comparison to

each other was produced.

Data points from Wednesday (green), Thursday (purple) and Friday (orange)

The most recent day's data

A Friday with many fewer companies restored to the register than usual

Data points from Monday (red) and Tuesday (blue)

A day with an exceptionally high number of dissolved companies

Standard Industry Classification (SIC) codes are used to describe the type of business a company conducts. Through using natural language processing on the text that describe each of these codes and using machine learning techniques to learn which words are associated with which codes this was applied back to the Local Authority data and the classifications they give to businesses to generate new code for some companies.

EPSRC
Engineering and Physical Sciences
Research Council

Project Details

Partners
Level Business Limited
Northumbria University

Project investment
£15,000

For further details
on the technology:
James Dobree
Level Business Limited
james.dobree@levelbusiness.com

For further information
on internships and
other collaborations:
Lorcán Mac Manus
Industrial Mathematics KTN
lbmm@industrialmaths.net
+44 (0) 1483 579108