MANCHESTER 1824
The University of Manchester

ScaleUp Institute: What Determines ScaleUp Density?

Sami Bashir, BA Politics, Philosophy and Economics

Introduction

The ScaleUp Institute is a private sector, not-for-profit company focused on making the UK the best place in the world to scale up a business. The institute aims to understand what makes 36,510 businesses out of 6 million businesses expand so rapidly. This is in order to inform the public sector and the private sector on how to promote business growth in order to make Britain 'the most fertile ground for businesses'.

Objectives

To understand how socioeconomic success relates to business growth at a Local Authority District (LAD) level.

Factors that affect a nation's economy such as education, access to finance, trading openness, and infrastructure were to be defined and represented by publicly available datasets and ONS datasets. This was in order to visualise and model the factors against a variable representing high growth firms (Scaleups).

The datasets should be cleaned, merged, plotted, and modelled using R Studio. This would allow for the experiment to be reproducible as all of the saved scripts containing the code can be viewed, verified, and run again.

Methodology

Firstly, we began by importing datasets into our working environment, which was R Studio. Most of the twelve datasets had already been found at the start of the project, but we discovered a few others such as SME lending per postcode.

Secondly, we cleaned the data. Most datasets required a similar script in order to tidy it. All datasets required the removal of some columns and rows, other datasets required transformation as we attempted to form new columns that were the product of applying a function to existent columns. A minority of datasets only had data at a more specific level than LAD. An example of this was SME Lending, which only had lending figures at a postcode level. This led to the most difficult task of the cleaning process, which was aggregating the data up to the LAD level. This process was particularly difficult because it involved using 'string' functions, which were more difficult to understand and more unique to the dataset.

Thirdly, we merged the datasets. This task was straightforward; join all of the datasets using a common variable, which was the LAD code. However, it was easy to become careless at this point as the focus could have been on the relative success of the merge, instead of accounting for the missing observations. And there certainly were missing observations, which meant further cleaning. Our merged dataset included variables such as: school performance, sectoral intensity, SME lending, public transport travel times, concentration of large firms, and exports.

Fourthly, after finishing the wrangling process, we began to visualise and model. Variables were plotted and modelled both individually and together against Scaleup density. Scaleup density was our chosen dependent variable to represent high-growth firms. We found interesting relationships, some of which changed when certain variables were included in larger models. At the end of the project we collated all of the code and left notes amongst the code. This was so that the project could be repeated and developed from where it was left off.

Results and Conclusions

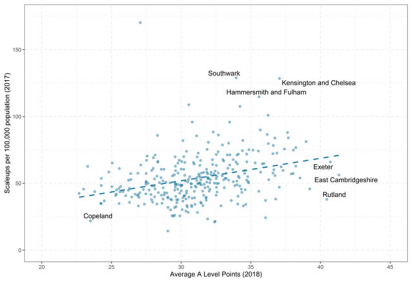
Three variables were particularly statistically significant when we evaluated what is responsible for the presence of high-growth firms in certain LADs. Mean A Level Points, Sectoral Intensity, and 'Presence of Large Firms' displayed a particularly strong correlation with Scaleup density.

Scaleup Density

This was the number of scaleups per 100,000 people. A scaleup is defined as a business that experiences an average growth rate of 20% or more over a 3-year period in either: number of employees, turnover, or both. Average Scaleup density across the 381 LADs is 79.203.

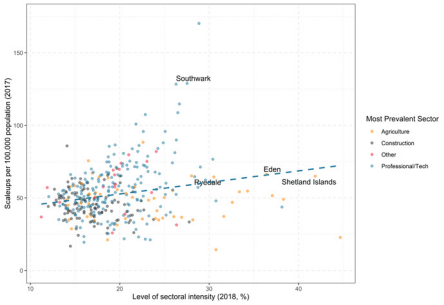
Mean A Level Points

A weighted measure of A level grades. An E equated to 16 points and an A* equated to 56 points, there was an interval of 8 points between each grade. The model found that, on average, if every student undertaking 3 A levels in a LAD were to have an increase of one grade in only one subject; this would result in an increase in Scaleup density by 0.391.



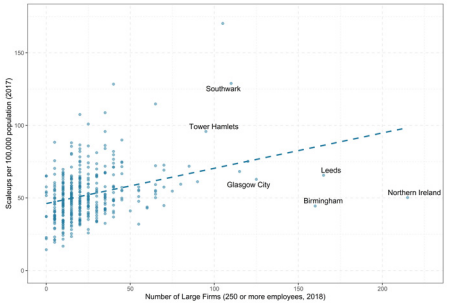
Sectoral Intensity

A measure of the proportion of firms operating within a particular sector of the economy. There were 17 sectors, some of which were defined as 'agriculture', 'infrastructure, and 'professional/tech'. An increase in the number of firms operating within a particular sector by 1% would result in an expected increase in Scaleup density by 0.812.



'Presence of Large Firms'

The number of businesses employing 250 people or more. An increase of 1 large firm within a LAD was expected to increase Scaleup density by 0.455.



Mean A Level Points, Sectoral Intensity, and the 'Presence of Large Firms' accounted for approximately 25% of the variation of Scaleup density. R^2 values typically ranged between 0.240 and 0.260 when measuring the significance of the 3 variables in different models. Furthermore, t-tests found that all 3 variables were statistically significant up to and above the 1% significance level. This implies there is more than a 99% chance that the correlation between high-growth firms, as measured by Scaleup density, and any of these 3 variables respectively is non-random. Ultimately, we have found that high-growth firms are significantly impacted by: the level of education, the concentration of firms operating within a certain sector, and the presence of large firms.

Key Skills Learnt

The main skills I learnt were: how to program in R Studio and locate useful datasets.

R Studio is a vast and powerful language that has many functions. I learnt how to: clean and transform the data, visualise the data, and model the data. Each one of these skills was implemented to a significant extent. These skills required me to gain both knowledge and experience of different packages within the programming language. This encouraged me to problem solve, think logically in the syntax of the programming language, and even think abstractly when evaluating why some models were presenting unexpected results. Another skill that was very important was communication, particularly written communication. I had to be very clear and concise as to what was happening with the code and when explaining the results.

I also located data and used databases, such as Beauhurst, in order to find relevant information. This process often required me to sift through large amounts of text, data, and web searches. Ultimately, this caused me to work much more efficiently and locate the best databases to use in future projects.

Ultimately, I benefitted greatly from the project. The main skills I acquired were: locating data, data wrangling, data visualising, and data modelling. These tasks were done in great depth such that the skills, and even the scripts for the code, will most certainly be relevant in all future data analysis projects.

Produced at the Graphics Support Workshop, Media Services, The University of Manchester

SCALEUP INSTITUTE

Contact
sami.bashir@student.manchester.ac.uk

Q-Step
A step-change in quantitative social science skills
Funded by the Nuffield Foundation, ESRC and HEFCE