## Are Top Universities More Attractive To Tech Startups?

## Introduction

Do we see more tech startups around the top Universities in the UK? There is evidence that tech startups cluster together and become incubators for innovation but does that also correlate with where certain Universities are located?

The purpose of this project is to find out if certain Universities have a higher proportion of tech startups in close proximity and the similarities between those Universities.

## Data

To look at this problem we are going to need two data sets:

- A list of UK universities with ranking data
- A count of tech startups in close proximity to each University

For the university ranking data we will use The Times Higher Education World Rankings data

To get tech startup proximity data we will use the FourSquare API

UK university ranking data

The data is not easy to scrape from the THE website, luckily this data has already been collected by a contributor on kaggle

The data can be found in <a href="https://www.kaggle.com/joeshamen/world-university-rankings-2020/download">https://www.kaggle.com/joeshamen/world-university-rankings-2020/download</a>

Credit must be given to THE for producing this data and Joe Shamen for making it available in this form.

To make the data useful for our purposes we drop columns we are not interested in. Remove the rank field because we will use the index instead.

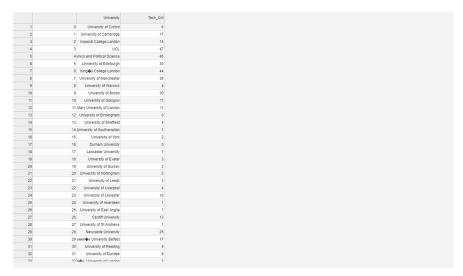
The data is formatted and then scaled by dividing each number by 100 so we can compare them later on. This leave us with a dataset that looks like this:

	University	International_Students	Percentage_Female	Percentage_Male	Teaching	Research	Citations	Industry_Income	International_Outlook
0	University of Oxford	0.41	0.46	0.54	0.905	0.996	0.984	0.655	0.964
1	University of Cambridge	0.37	0.47	0.53	0.914	0.987	0.958	0.593	0.950
2	Imperial College London	0.56	0.38	0.62	0.845	0.876	0.970	0.699	0.971
3	UCL	0.52	0.57	0.43	0.778	0.887	0.961	0.427	0.962
4 Lo	ondon School of Economics and Political Science	0.71	0.53	0.47	0.690	0.830	0.928	0.351	0.932

## Tech Startup Data

Before we can get the count of tech startups near a University we need to find the location data for each University. For that we use the geopy package and the Nominatim geocoder.

Using the location data we use the FourSquare search API to find tech startups within a kilometre of each University. This count is then stored in a new dataframe.



That completes the data gather and wrangling stage so now we can move onto the analysis.