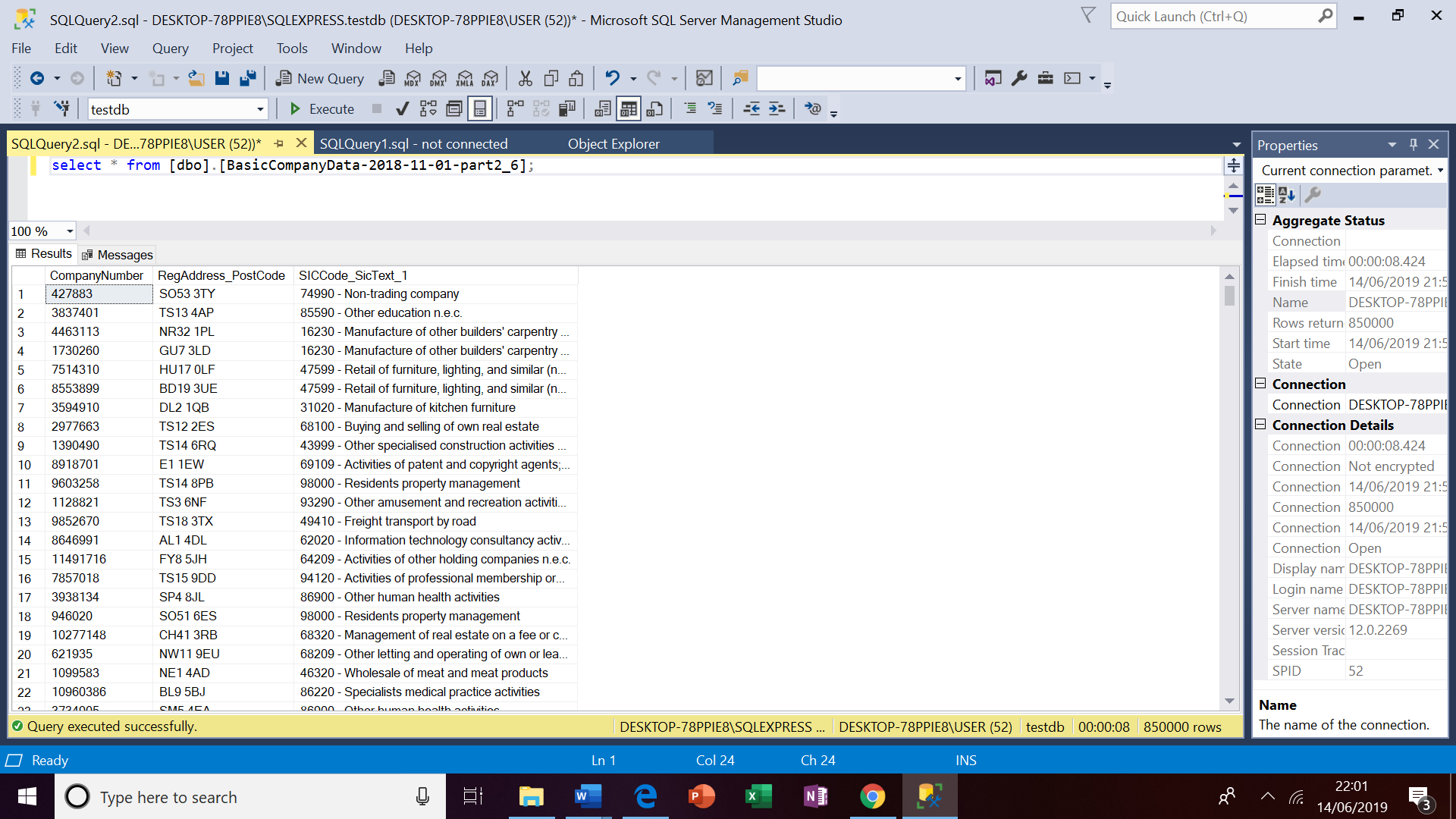
**Modelling business clusters in regions**

**Data collection and analysis:**

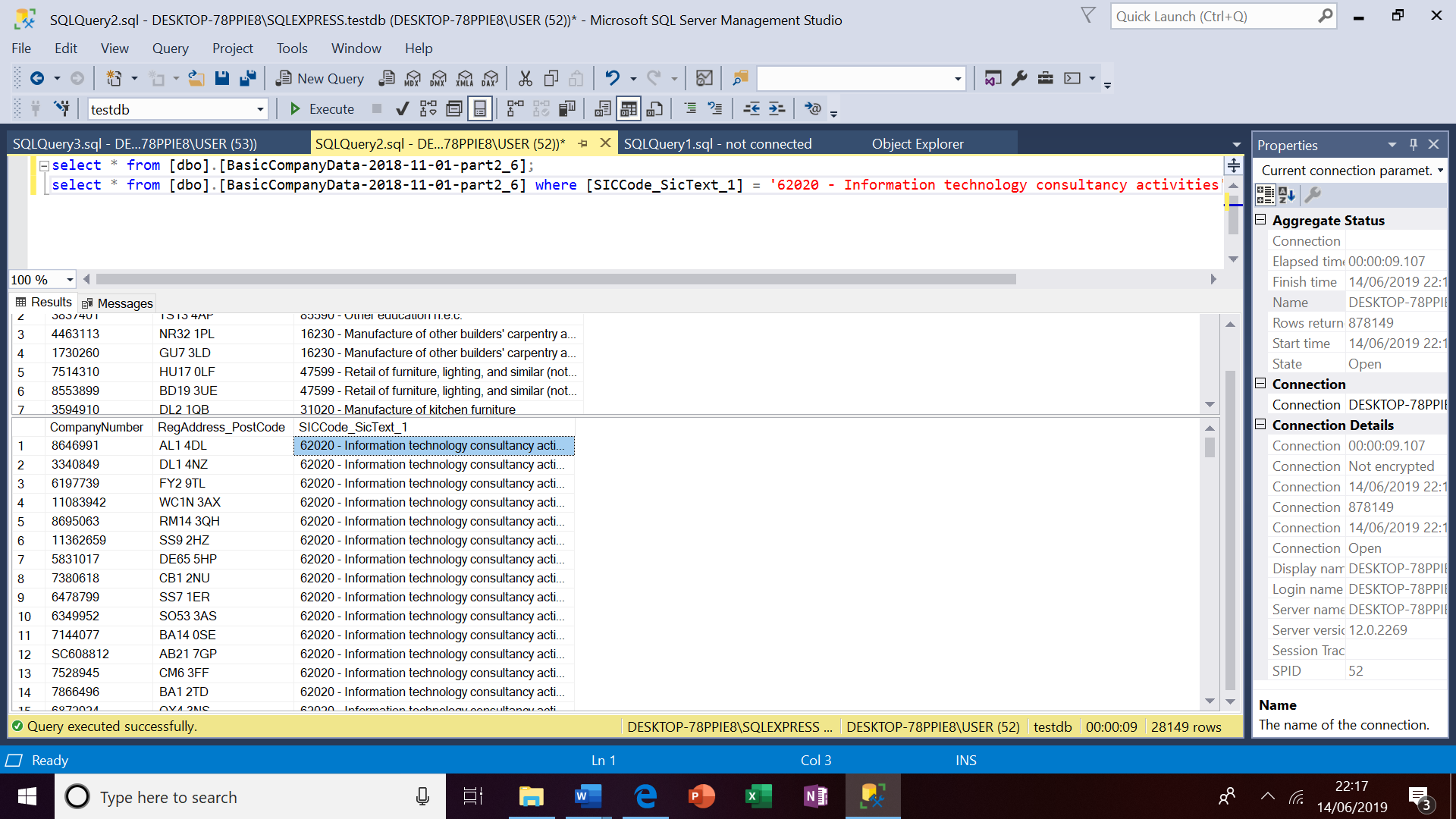
Data is available from the open Gov.UK company House, UK Office of National Statistics data of companies and data from UK Science Park Association (UKSPA), the International Association of Science Parks (IASP) and Catapult technology centres (Innovate UK). Data will be extracted cleaned then both by serial iteration through (outliers separated for individual analysis).

At a micro-level, Geographical Information Systems will be used with real data obtained from the UK Office of National Statistics and data.gov.uk to create GIS mashups of the actual (post code) locations of ‘prey’ firms onto maps and then superimpose real "predators", using data available pertaining to mature clusters (obtained from UKSPA, IASP etc.) as well as younger nascent initiatives (from innovate UK).

First of all the raw company data for preys was downloaded from open source – gov.uk/ company house in the excel format and then it was loaded on the MSSQL server where the simple sql query was executed to fetch total number of 850,000 rows with 3 fields named as Company Number, Post Code and SIC Code as shown below:

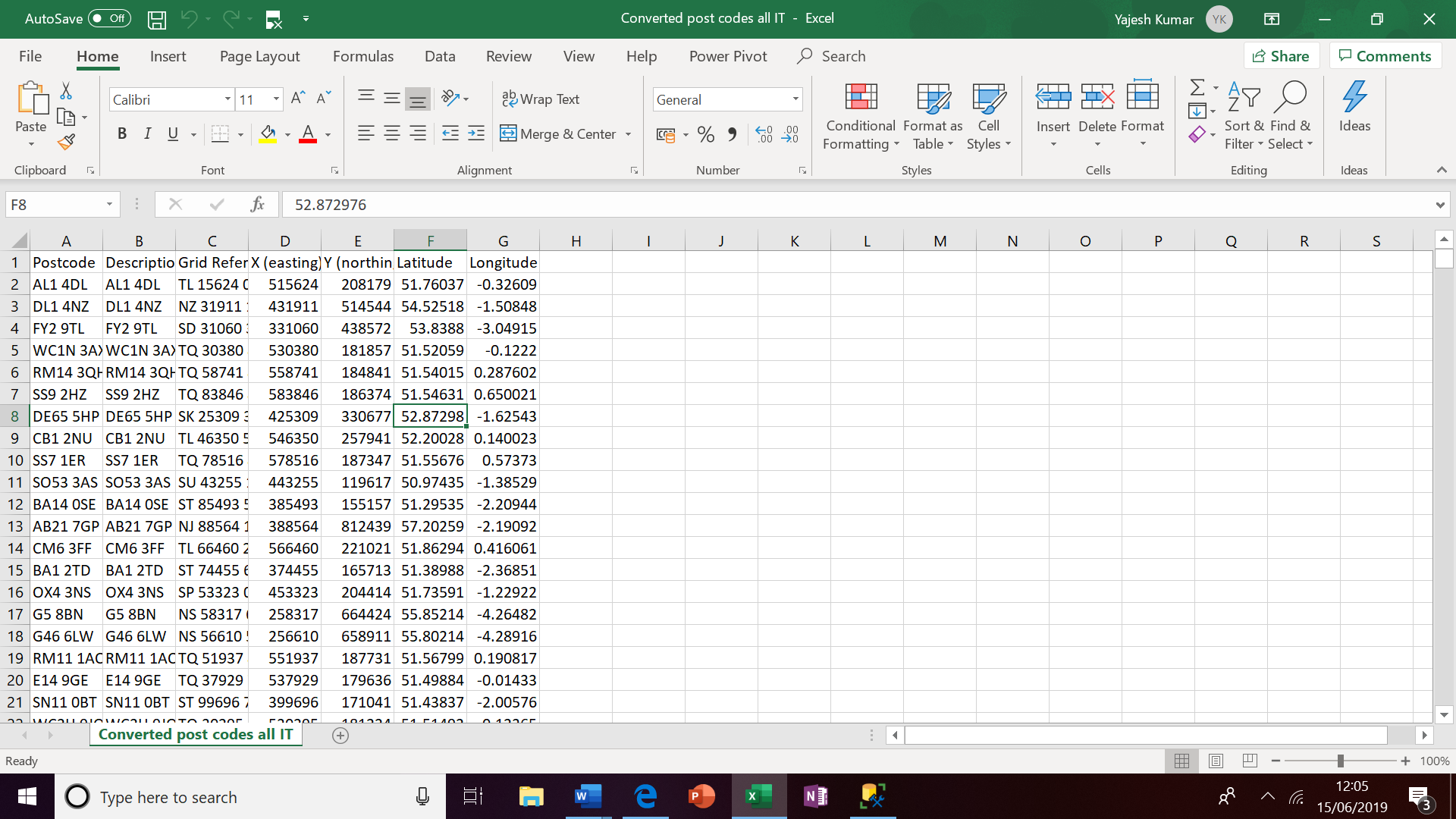


In the next step, companies with SIC code '62020 - Information technology consultancy activities' were identified making another sql query which fetched 28149 records as shown below:

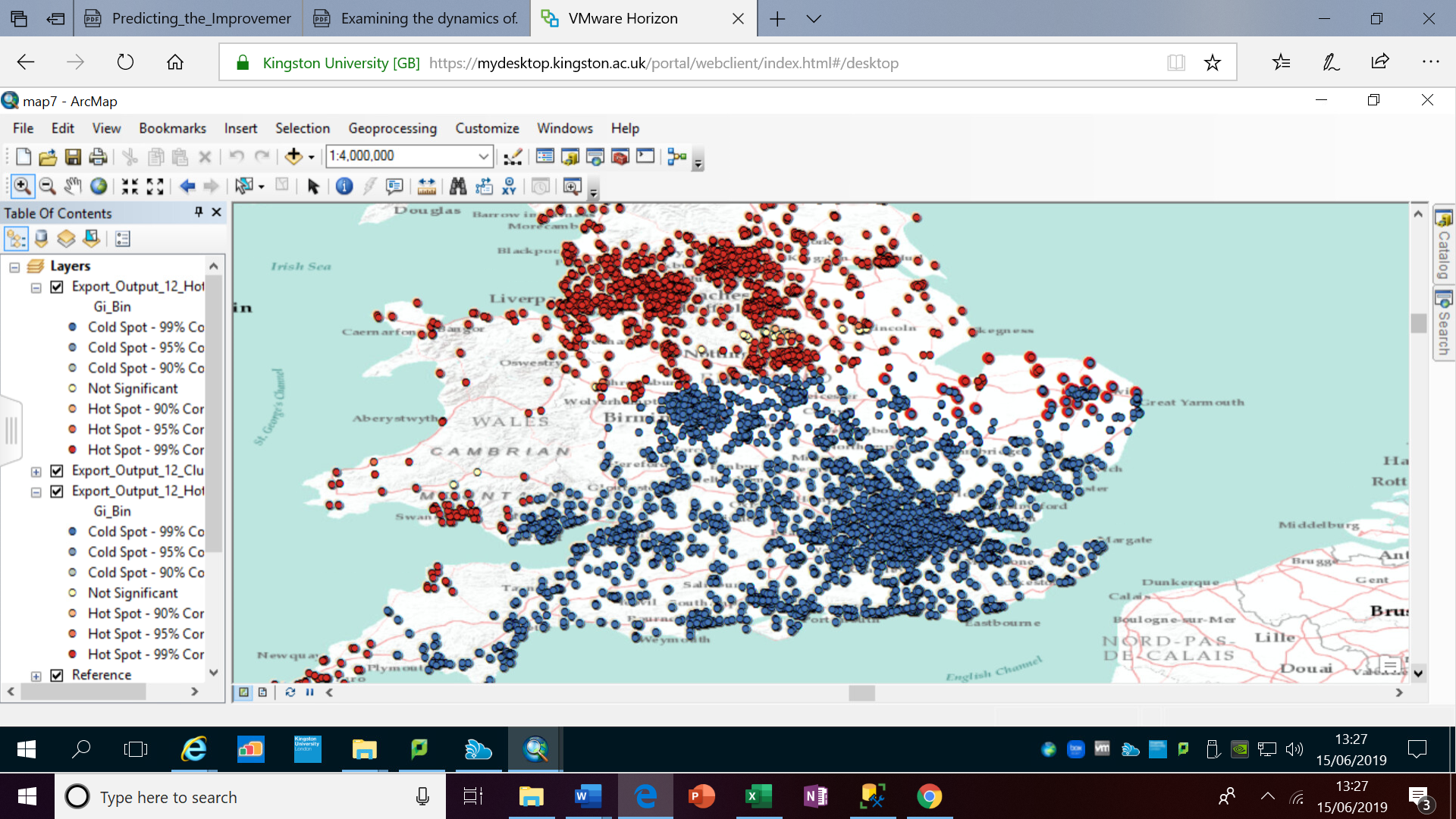


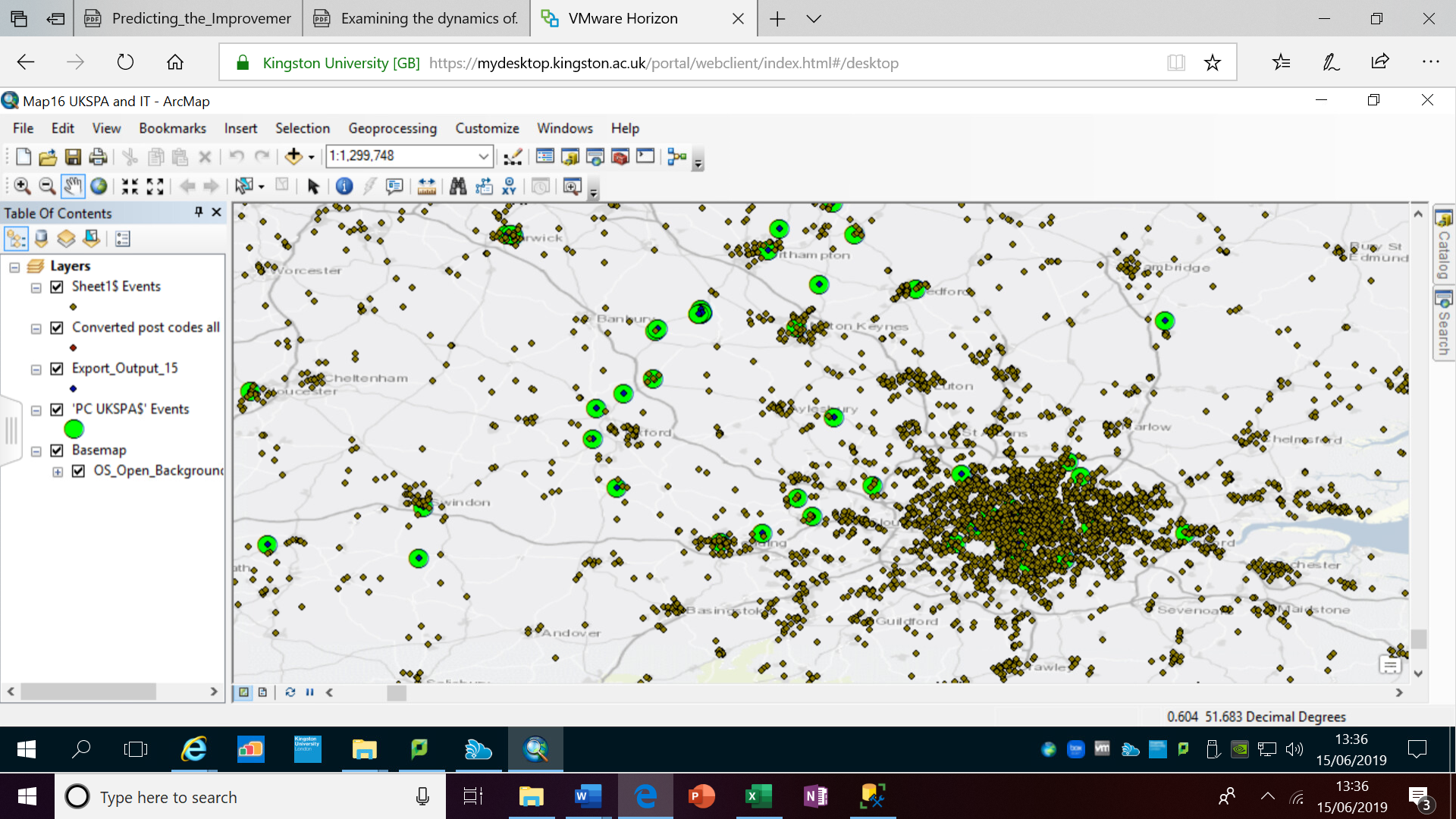
In the next step, these companies data with SIC code '62020 - Information technology consultancy activities' will be cleaned, any NaN values rows deleted and made ready to be mashed up using Geographical Information System – GIS’s application *ESRI’s Arc GIS.*

Before mashing up the data on map using *ESRI’s Arc GIS*, data of companies with SIC code '62020 - Information technology consultancy activities' will be converted in the longitude and latitude format as shown below:



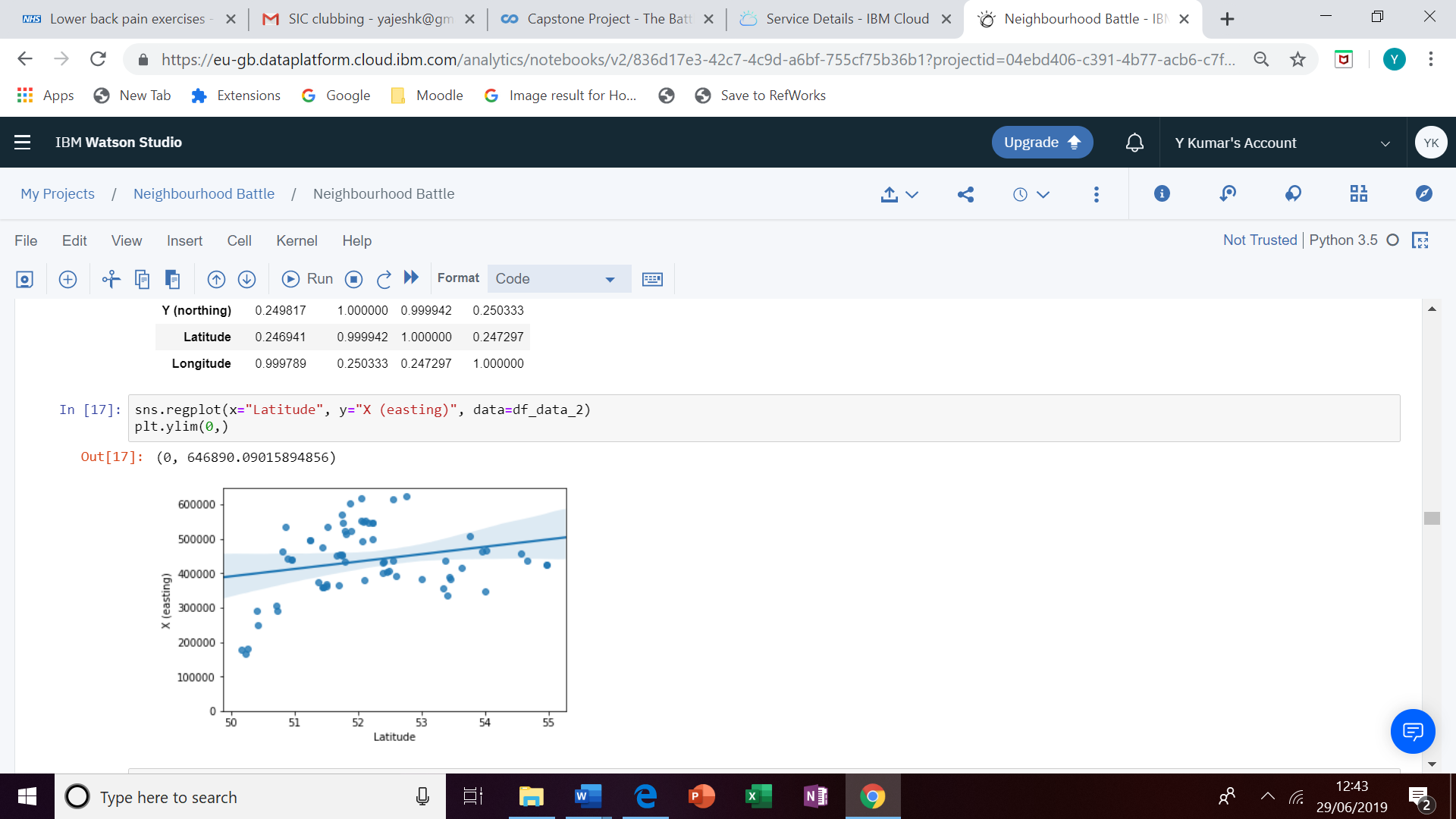
Now this data is ready to be mashed up using using *ESRI’s Arc GIS* application. On mashing up the data the map with related statistical information is presented below:





Converted post codes all IT and UKSPA mashed up together.

Also regression analysis will be performed using Jupiter notebook python to find out the best suitable location for the next UKSPA and it would be further supported by using Foursquare application.



**References and applications:**

Gov.uk

UKSPA

Company house

ESRI ARC GIS

Watson studio

MSSQL studio