IBM SKILLS BUILD

DATA70202 APPLYING DATA SCIENCE

Format B
Word Count: 6894

Group 13

List of abbreviations

NFER – National Foundation for Educational Research

ONS – Office for National Statistics

SOC – Standard Occupational Classification. SOC-4 indicates classification by 4-digit code, which is the most granular level. SOC-2 indicates classification by 2-digit code.

ESS – Employer Skills Survey

GMCC – Greater Manchester Chamber of Commerce

WF – Working Futures

LSIP - Local skills improvement plan

ERB - employer representative body

DfE- Department for Education

RMSE- Root Mean Squared Error

RMSPE - Root Mean Squared Percentage Error

MSE - Mean Squared Error

Introduction

Automation, AI, and technology advancements will continue to create shifts in the labour market, slowing growth in lower skilled jobs while driving growth in high skilled occupations. The digital skills gap is expected to increase and to cost the UK £63 billion per year in lost potential gross domestic product (GDP) through reductions in economic productivity, business competitiveness and employee job satisfaction. Given this sizeable cost, it is in the interest of the government and employers to improve the quality and supply of digital and non-technical skills in the labour market to meet the increasing demand.

IBM Skills Build is a platform focused on providing core technology and workplace skills to address the skills gap and improve the talent pool for high demand jobs. With the ever-evolving labour landscape, the purpose of this paper is to estimate the growth in skills demand by occupation in the immediate (2025) and long-term (2035) labour market. The analytical approach implemented involved projecting future demand for occupations with the combination of skills required for each occupation. Moreover, we were able to quantify the predicted skills gaps by regions across England at very granular occupational level. The approach was tailored to provide tangible insights that IBM Skills Build could   In line with our findings and previous literature published by various government organisations, we focused on workplace skills such as leadership, critical thinking, and digital skills as the key targets for upskilling the labour force, rather than formal qualifications or subject-matter expertise/general knowledge areas.

We achieved 6 key objectives:

- Objective 1: Developed deep understanding of the current labour market and UK trends in skill shortages, including key challenges faced by employers and establishments.
- Objective 2: Obtained a forecast for the near future of the job occupations that exhibited the most promising trend of growth and were also relevant to our industry partner.
- Objective 3: Delved into specific skills associated with the roles of the highest importance and provided detailed information about the skill composition of the current labour demand, as represented by frequency of job adverts.
- Objective 4: Identified the skills gap in the chosen relevant positions, at the highest granularity, and regional level for England, and discussed the impact of this forecast for IBM Skills Build
- Objective 5: Produced estimates of the projected labour market size by relevant occupations until 2035 to provide long-term recommendations for the growth of IBM Skills Build
- Objective 6: Developed tailored insights for IBM Skills Build based on our findings and the current structure of the platform.

Literature Review

We looked at four independent studies and published literature to gather detailed information on the current skills landscape, which were crucial for developing our methodological approach and ultimately providing IBM with tailored recommendations for their educational platform.

Employer Skills 2022

The Employer Skills Survey is a biannual nationwide survey, which examines the challenges faced by establishments in recruiting, training, and hiring employees (IFF Research, 2023). The survey also includes information on skill gaps, which are defined as employees who are not fully qualified for their positions, or vacancies that are hard to fill due to skill requirements (IFF Research, 2023). According to the 2022 findings, 36% of all establishments reported having at least one skill-shortage vacancy, compared to 10% in 2017 (IFF Research, 2023). Furthermore, 15% of establishments noted they had employees with skills gaps (IFF Research, 2023), which suggests that the skills discrepancy is more pronounced among new hires rather than existing employees. The report delivers information about skills gaps by major industry sectors and UK regions, as well as major SOC groups, which are useful as an overview of current trends, and highlight the dynamic nature of the skills market.

Crucially, the report highlights a significant challenge in the current skills landscape - lack of training (IFF Research, 2023). Only 60% of employers across the UK arranged or funded training for staff in 2022 (IFF Research, 2023). Another indicator of training performance involves quantifying businesses investment in employee training. Estimates suggest that employer investment in training is still broadly flat across England, with the spend per employee increasing by only 1.5% in 2022 compared to 2019 (IFF Research, 2023). This highlights the pressing need for strategic development and investment in upskilling the UK workforce, which is also a key concern for IBM Skills Build.

However, the public ESS dataset (GOV.UK, 2023) was too broad for our project purpose, as IBM was specifically interested in a few critical sectors. While the report investigates the specific skills that job applicants and employees are currently lacking (IFF Research, 2023), the information on specific skills is not accessible in the publicly available dataset, which combines findings from 2011 to 2022. The dataset also contains high incidence of missing values, which, combined with the very few data points (biannual), and the absence of the IT sector before 2017, meant it was unsuitable for forecasting skill shortages. Therefore, we decided to refrain from using the data in our models despite our initial intention to do so.

LSIP

Local skills improvement plan (LSIP) are policy frameworks developed by an employer representative body (ERB) for 38 narrow English geographical areas. The plans draw on feedback from local employers and statistics, to highlight the current and future local skill landscape. These findings are published on the dedicated ERB website for each LSIP, highlighting the most pressing issues in the regional labour market, such as skills gaps, recruitment, and training shortages. The LSIP reports provide rich contextual information, which is crucial for understanding local labour market trends. However, due to the amount of information available at very granular level, this literature review will focus on one example - Greater Manchester (GM) - as a case review to frame the current landscape of LSIP data.

Recent GMCC (2024, p. 22) research of the IT industry provides detailed breakdown of technical skills currently in demand for the sector (Table 1). Within GM, vacancy data also displayed demand for management skills and soft skills in the sector (GMCC, 2024).

SKILLS	DEMAND (AUGUST TIO OCTOBER 2023)	2022 – 2023 YoY % DEMAND CHANGE (AUGUST TO OCTOBER)
Software Development	3,174	31.3%
Data Analytics/ Science/Mining	290	14.2%
Cybersecurity	1,477	31.3%
Cloud Computing	3,500	53.7%
AI/ML/IOT	2,167	33.4%

Source: Adzuna

Table 1. Digital Skills Demand in GM (GMCC, 2024, p. 22)

The demand for this set of skills is not specific to the digital sector alone, but cuts across all other sectors in GM, and encompasses qualifications such as computer literacy and the ability to use software programmes like Microsoft Word and Outlook (GMCC, 2024).

However, interviews with employers are confidential and therefore not publicly available for analysis. This is the case across all LSIPs, which meant we could not incorporate this information in our own methodological approach. However, the GMCC report (2024, p. 54) provides excerpts from interviews with employers, which highlight key skill gaps prevalent across all sectors (Table 2).

SKILLS GAPS
Exhibiting the right behaviours and showing a positive attitude and willingness to learn is more important to some employers than having all the skills needed for a particular role.
Green/sustainability skills are useful in many industries – just knowing the basic principles, rather than a specific qualification.
People management and leadership skills are lacking in almost all industries – very often people are promoted through being proficient in their current role, but don't have the necessary leadership and management skills that are needed when taking a step up.
Automation, robotics and data analysts are in short supply in some industries, particularly engineering.
Fitters, tool makers, and more traditional engineering skills are still required, but lacking in this industry.
Electrical engineering and mechanical design skills missing in some engineering firms.
Lack of communication skills is a big problem for lots of sectors, with many citing this as one of the key issues when it comes to performance, management and productivity.
New recruits, in some cases, are not able to communicate effectively either in writing or in person.
Digital skills, basic data, data literacy and understanding data usage are key skills lacking in some industries.

Table 2. Key Themes from Business Interviews in GM (GMCC, 2024, p. 54)

The LSIP report reveals that there is a skills mismatch in GM - a gap between what employers need now and, in the future, and the skills the current workforce possesses (GMCC, 2024). Therefore, attention must be paid to bridging the skills gap and promoting efficient reallocation of labour, both of which require targeted training programmes that consider transferability of skills as well as sector-specific needs. Training programmes that are tailored to individual organisational needs can be effective in overcoming the skills mismatch and promoting business efficiency and performance, ultimately leading to a more productive local economy.

However, evidence suggests that the volume of training and employer investment in training has been on the decline in the UK (GMCC, 2024). The LSIP survey showed that only 4 out of 10 employers in GM provided job training in 2022 (GMCC, 2024), something which is in line UK-wide ESS 2022 findings. Employers reported challenges in accessing training - due to a lack of effective communication and awareness of available provision, as well as local availability, including in more modular, flexible courses (GMCC, 2024). This was one of the many reasons co-designing training programmes with employers was seen as essential to effective provision - not only providing the right skills, but also in the right places (GMCC, 2024).

This research supports the importance of our current study. Not only did we perform an independent research of market skill gaps, but we also provided tailored recommendations to ensure that IBM platform is a valuable tool in addressing the lack of quality training and increasing supply of talent.

The Skills Imperative 2035

Dickerson et al. (2023) investigate the future labour market demand for skills and present a forecast of the UK skills composition in the next 10 to 15 years. Their latest working paper (Dickerson et al., 2023), part of the 'The Skills Imperative 2035' programme led by the NFER, provides a comprehensive review of the overall predicted skills demand, as well as breakdown by occupational categories. 'The Skills Imperative 2035' project (Dickerson et al., 2023) utilises time series macroeconomic models to deliver projections of labour demand by occupation (4-digit SOC) and region until 2035, relying on labour force data from the ONS. These occupational projections have subsequently been mapped to skills requirements of the US Occupational Information Network (O*NET) database (Dickerson et al., 2023). The skills projections provide findings of high granularity, which are largely missing in the available literature on labour skills forecasting.

The number one skill by absolute increase based on updated NFER projections is 'interacting with computers' (Dickerson and Rossi, 2024), which highlights the significance of technological and digital literacy across all occupations. Furthermore, 5 out of the top 20 projected skills by absolute increase involve analysing, processing and comprehending data (Dickerson and Rossi, 2024), which demonstrates the increasing importance of good analytical skills. Additionally, the findings of the report suggest that soft skills, such as communication, planning, and collaboration with people, will be the most utilised across all occupations and are expected to continue to be an integral part of job requirements (Dickerson and Rossi, 2024). While the data associated with forecasting specific skills is not publicly available as of now, the econometric projection of the labour market between 2025 and 2035 by region, LSIP and occupation are freely accessible (DfE, 2023). Although the data are regarded as experimental estimates, they provide a high level of granularity that is unmatched by any other UK public data source. Hence, we have incorporated these projections into our own methodological approach.

AI Labour Market

The "Understanding the UK AI Labour Market: 2020" report (Dabhi et al., 2020) delivers valuable insights for the context of skills gap analysis and projection of future skills requirements. The findings are based on research conducted by Ipsos MORI, through a survey of 118 businesses and public sector organizations, and 50 in-depth interviews, providing a comprehensive analysis that outlines the current state and future trends pertaining to the UK AI labour market (Dabhi et al., 2020). The report underlines the technical and soft skills gap on in AI and data science, which is becoming increasingly important, as businesses try to incorporate AI technologies as a source of innovation and growth (Dabhi et al., 2020). Most companies reported that there were not enough key AI skills among the currently working personnel (Dabhi et al., 2020). The report specifically indicates that 49% of the businesses had experienced disruption of activities, as applicants did not have the technical skills they required (Dabhi et al., 2020). Moreover, 35% of businesses believed they were unable to meet objectives due to a lack of technical skills among its currently employed staff (Dabhi et al., 2020). Additionally, Dabhi et al. (2020) note that soft skills were key for collaborating in interdisciplinary teams, which will be of growing importance as the UK tech landscape is increasingly diversifying.

Dabhi et al. (2020) emphasise the importance of systematic planning and investment at the strategic and operational levels to ensure that technological advancements are effectively supported and utilised through targeted training, which demonstrates the role of flexible, comprehensive educational platforms, such as IBM Skills Build.

Methodology

Our methodology consisted of 4 sequential parts, which will be the focus of this section. More details about the process and further analysis of UK regions outside of England can be found in the Appendix.

Objective 1: Obtain Occupational Demand by 2025

The first stage of our methodological approach was aimed at forecasting the labour demand by occupation (SOC-4 classification level) until 2025 and selecting the key occupations for IBM that are aligned with the industry's focus and show the strongest growth trends.

We utilised monthly job postings data collected between 01.2017-05.2023 (ONS, 2024), which contained a high level of regional and occupational granularity. Although the dataset is considered official statistics in development (and thus can be updated at a later stage), the recent publication and the relevancy of the data were key for our forecast. More outdated datasets would inevitably be less reliable in capturing dynamic labour market change, especially with the emergence of new technologies and job roles. The data notably contained imputations in the period 04.2020 to 12.2021 due to a technical issue with the web scraping algorithm (ONS, 2024).

We used a Histogram Based Gradient Boosting (XGBoost) model, which is especially suited to time series forecasting tasks (Papadopoulos and Karakatsanis, 2015), because of its improved accuracy due to the sequential fitting of residual errors of each constructed tree. The feature engineering included lag variables based on number of ads in the last 1, 2, 3 and 12 quarters, as well as GDP in the prior quarter. Furthermore, we added a COVID-19 dummy variable to account for pandemic-related disruptions. Quarterly projected GDP data for the period until 2024 was based on data from the House of Commons (2024). Beyond 03.2024, we utilised House of Commons (2024) estimates from the OBR (high threshold) and the Treasury Department (low threshold), incorporating randomness to capture potential deviations. Throughout each stage of the modelling, results were inspected to ensure the accuracy and robustness of our models. We used cross validation and RMSE to assess their performance.

As our first objective was to forecast the job demand in the most granular occupational level (SOC-4), we used XGBoost with the full dataset to obtain predictions until 2025 and subsequently performed linear regression on the results to obtain coefficients for the growth in demand of different SOC-4 categories. Out of the top 30 occupations with the most significant total forecasted demand, we isolated 4 occupations which were the most relevant to IBM.

Objective 2: Obtain Regional Breakdown of Skill Gaps in 2025

Our second objective was to get the SOC-4 occupational gap by region. To achieve this, we had to model SOC-2 data by region from the ONS (2024) web ads dataset. We utilised the same modelling approach as discussed in the previous section. Subsequently, we calculated the proportion of the relevant SOC-4 categories out of the total SOC-2 projected demand. We then multiplied this estimate by the regional predictions of the SOC-2 model, which gave us the total projected regional SOC-4 demand.

However, these steps only provided us with the relative projected demand, and we needed to seek out additional sources for market size projections. Census data from 2021 (ONS, 2023) and LSIP projections for 2025 (DfE, 2023) of the market size per SOC-4 enabled us to derive an index for the supply in the market.

We obtained the proportion of each SOC-4 occupational category in the English population from the Census, as it was the most representative dataset available on the topic. Due to this, we had to work under the assumption that the market size per occupation in the country would stay the same until 2025. We grouped LSIP data by region and used the 2025 projections of the market supply to obtain a statistic (Figure A-B-2), which we multiplied by the proportion of market supply for 2025 from the Census (Figure A-B-3). This gave us the total projected supply for 2025 for the 4 narrow categories of interest to IBM. The ultimate step of getting the projected gap was creating an index that calculates the ratio of the estimated demand and supply per region by SOC-4 (Figure A-B-4). This index allowed us to map the regions of England for the 4 relevant occupations, illustrating where demand exceeds supply, visualising the distribution of the projected skills gap.

Objective 3: Obtain Specific Skills for Relevant Occupations

To get the specific skills within those 4 occupations, we decided to use web scraping techniques of job postings, as there was no readily available data detailing this information. Automated web scraping technologies were utilised between 30.04.24 and 01.05.2024, targeting the extraction of job advertisement data from the Indeed.com website. A function was constructed to generate Indeed search URLs, enabling systematic extraction of advertisement information for related positions based on a predefined list of job keywords and a list of major UK cities. These job keywords covered a range of fields including programmers, IT business professionals, and managers. Data collected from the Indeed.com website was subsequently loaded and pre-processed. Multiple JSON data files were combined into a single data frame and deduplicated to ensure the uniqueness of each job advertisement, resulting in a sample size of 14244 ads. Textual information within the data was standardised to aid subsequent analysis.

Objective 4: Forecast Occupational Supply by 2035

Our last objective was to provide a forecast in the next 10 years. However, due to accumulated error, our monthly projections of web ads would not be reliable for modelling. Hence, our last task involved utilising historical labour market data by the from 1990 to 2022 to generate long-term predictions (GOV.UK, 2016). The data contains actual instances from 1990 – 2014 and produces projections from 2014-2022 (GOV.UK, 2016). The projections are based a macroeconomic model that contains sector, occupational and regional dimensions (GOV.UK, 2016). Although our forecast is an extension of these available projections, the time series contained 24,192 data points per region and an extensive coverage of socio-economic variables and occupations, which made it especially suited for our task at providing an estimated trajectory of the labour market until 2035.

Pre-processing steps included data cleaning and feature engineering, such as adding lag features. Various machine learning models, including Elastic Net, K-Nearest Neighbours, and SGD Regressor, were employed to predict the market size for different occupations by 2035. Model parameters were optimized through cross-validation and grid search techniques, and the models were trained and evaluated.

These 4 steps were undertaken to ensure that we capture the skill gap by 2025, in the occupational categories most relevant to IBM, while also providing details on specific skills in the current labour force. Moreover, our extended forecast by 2035 provides additional insights that can guide IBM in choosing relevant occupations to focus on long-term.

Results

Forecast of Regional Gaps by Occupation in 2025

Training Models

Our training model (SOC-4 by country) had RMSE of 0.43 on the training set and 0.76 on the validation set. Notably, the most important features were the lag of the previous and the previous two quarter values, as well as the SOC-4 code. However, the corresponding results for our training model (SOC-2 by region) demonstrated decreased performance (RMSE training-0.10, validation-0.24). Again, the lag of the previous and the previous two quarter values, were the most important predictors, but SOC-2 code was in fourth place. The lower RMSE metrics in the training set were probable cause due to the COVID-19 period, which although we accounted for with a dummy variable, was less representative than dates after the pandemic. Moreover, the imputed values mentioned in the methodology section, also fell in the training set. The lower performance of the second model might have been due to the increased complexity introduced by regional data, and the lower predictive influence of SOC2 compared to SOC-4.

Occupational Demand Index

To make the most of the limited data we had, we trained an XGBoost model on the entire dataset to forecast the period between 05.2023 and 12.2025.

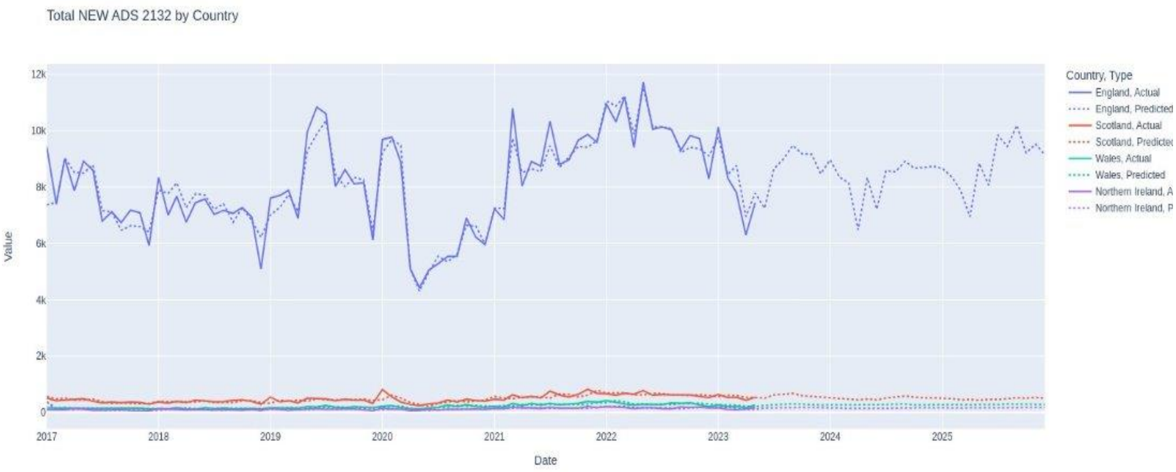


Figure 1. Forecast for IT Managers by Country

Our modelling results enabled us to generate granular forecasts by individual SOC-4 occupational categories at country level. Figure 1 displays an example of the total projected ads for IT Managers between 2023 and 2025, as well as a visualisation of the historical fluctuations between 2017 and 2023. The number of ads for this position come predominantly from England, which is also the region we have focused on due its crucial importance for the UK labour market, and the industry partner.

Table 3 displays the top 30 job categories with the most significant expected growth in demand by 2025. These categories were arranged by market importance, combining coefficients, and forecasted new job numbers using weighting, which reflects the relative influence of each occupation on the future market (Figure A-1). Out of the results in Table 3, we selected the following categories: Programmers and software development professionals (SOC2134), IT Managers (SOC2132), Business and financial project management professionals (SOC2440), IT project managers (SOC2131). These professions are most closely aligned with IBM's organisational focus, as well as the targeted audience of their Skills Build platform, allowing us to deliver tailored results for the industry partner. Notably, software development professionals top the list, indicating a high market demand and significant importance for IBM's Skills Build platform.

	index	SOC4	Intercept	Coefficient	SOC 4 digit label	SOC2	Value	WeightedRoles
0	0	2134.0	5128.495161	49.436290	Programmers and software development professio...	21	303615.0	1.252369
1	3	6112.0	2502.840323	6.125907	Teaching assistants	61	128702.0	0.065784
2	1	6113.0	1938.338710	7.454133	Educational support assistants	61	103439.0	0.064335
3	2	2313.0	2101.016129	6.672379	Secondary education teaching professionals	23	110923.0	0.061754
4	4	2132.0	2303.467742	5.576613	IT managers	21	117384.0	0.054619
5	5	3574.0	2476.741935	4.133569	Other vocational and industrial trainers	35	124383.0	0.042899
6	8	2440.0	2790.166129	3.031452	Business and financial project management prof...	24	140592.0	0.035561
7	9	5434.0	2801.141935	2.914012	Chefs	54	140323.0	0.034118
8	6	9231.0	1643.527419	3.978629	Security guards and related occupations	92	84873.0	0.028175
9	14	9223.0	2966.730645	2.106552	Cleaners and domestics	92	146067.0	0.025674
10	12	2412.0	1956.946774	2.243246	Solicitors and lawyers	24	97008.0	0.018157
11	7	9265.0	1302.624194	3.255847	Bar staff	92	66789.0	0.018144
12	18	4216.0	2229.724194	1.713306	Receptionists	42	110056.0	0.015733
13	11	9264.0	1456.862903	2.526714	Waiters and waitresses	92	72677.0	0.015322
14	13	2453.0	1225.498387	2.176512	Quantity surveyors	24	61569.0	0.011181
15	36	6131.0	2915.325806	0.836492	Nursing auxiliaries and assistants	61	142942.0	0.009977
16	22	2212.0	1694.124194	1.180242	Specialist medical practitioners	22	81864.0	0.008062
17	10	3213.0	618.487097	2.531552	Medical and dental technicians	32	32918.0	0.006953
18	27	3229.0	1457.874194	1.037097	Welfare and housing associate professionals n....	32	72111.0	0.006240
19	15	2240.0	663.312903	1.950202	Veterinarians	22	34369.0	0.005593
20	16	2319.0	654.217742	1.937500	Teaching professionals n.e.c.	23	33872.0	0.005476
21	33	5241.0	1482.174194	0.894859	Electricians and electrical fitters	52	72226.0	0.005393
22	17	9267.0	685.251613	1.728226	Leisure and theme park attendants	92	35033.0	0.005052
23	32	6136.0	1328.335484	0.915020	Senior care workers	61	65514.0	0.005002
24	37	2129.0	1431.525806	0.799294	Engineering professionals n.e.c.	21	70251.0	0.004685
25	34	4142.0	1273.683871	0.894758	Office supervisors	41	62749.0	0.004685
26	59	8214.0	2644.454839	0.404032	Delivery drivers and couriers	82	127315.0	0.004292
27	19	2131.0	614.217742	1.526714	IT project managers	21	31678.0	0.004035
28	20	3414.0	628.403226	1.330645	Dancers and choreographers	34	31930.0	0.003545
29	25	3557.0	713.261290	1.081956	Events managers and organisers	35	35825.0	0.003234

Table 3. SOC-4 Weighted Importance Coefficient

Regional Demand by Occupation

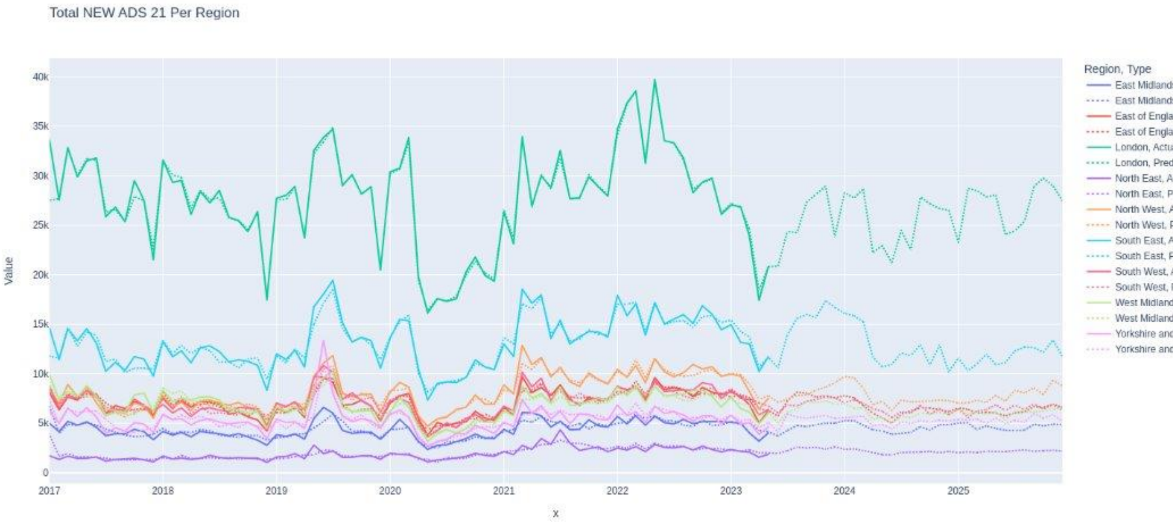


Figure 2. Forecast for Science Research and Technology Professionals by region

Our second model allowed us to predict the future demand as represented by total job ads by English regions at the level of SOC-2. Figure 2 illustrates the forecast for the broad occupational category ‘Science, research, engineering and technology professionals’. Unsurprisingly, Greater London and the South East are expected to have the biggest number of ads, which highlights their key role in the 2025 labour market for this sector.

Having combined the predictions from our two models (Figure A-B-2), we mapped the projected demand for the 4 key SOC-4 occupations by region, obtaining a detailed prediction of the English areas which are expected to see the most growth by 2025 (Figure 3). Software engineers and programmers overwhelmingly showed the biggest forecasted growth in demand among the 4 occupations across all regions. Greater London is predicted to be the region with most demand for all occupations except Business and Financial Project Managers. Interestingly, the South East and the North West had the highest forecasted demand for this position, suggesting a possible economic shift in the Business Industry at managerial level from the current hub, London, to emerging regional markets.

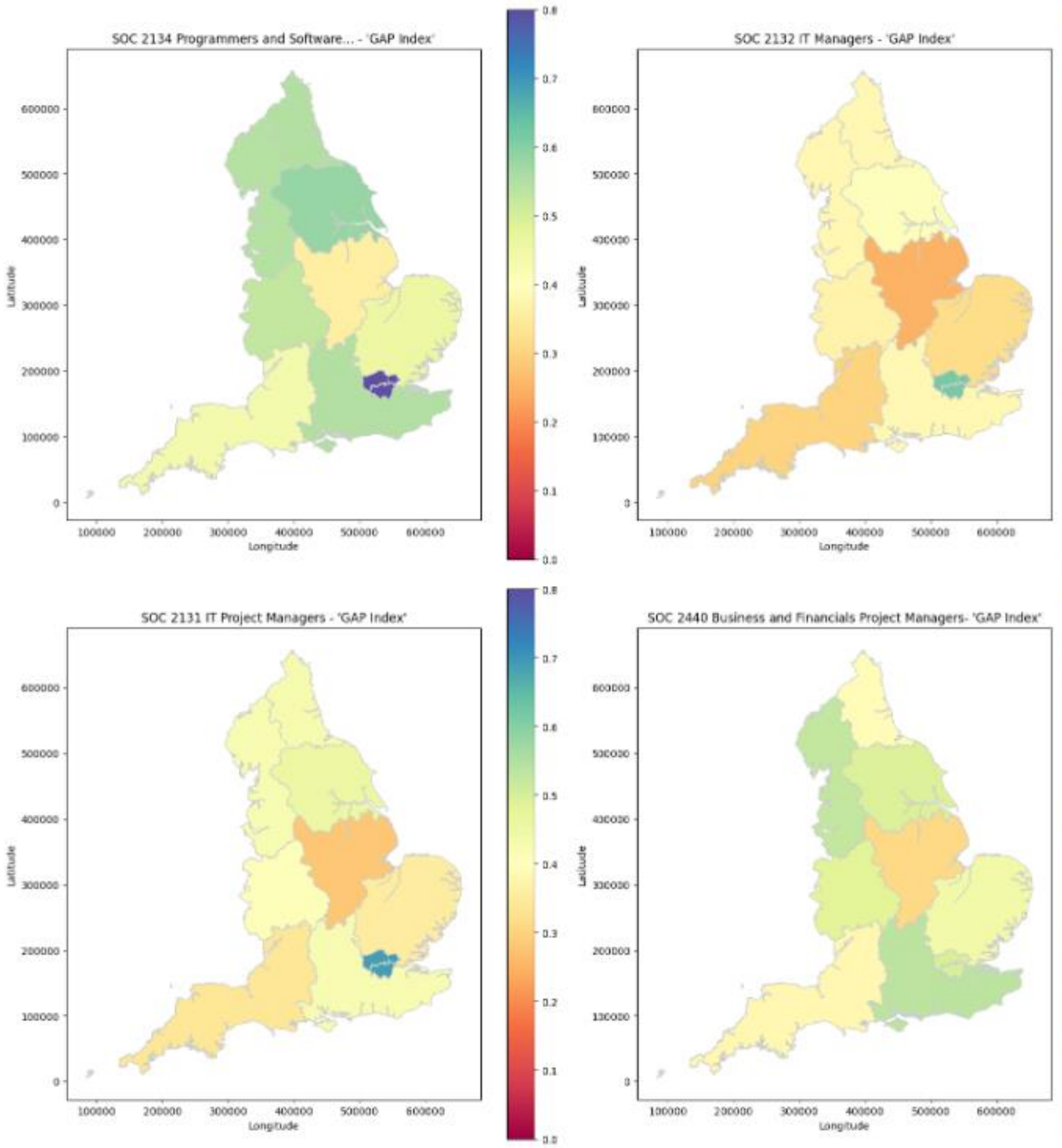


Figure 3. Index for Regional Demand per SOC-4

Skills Gap

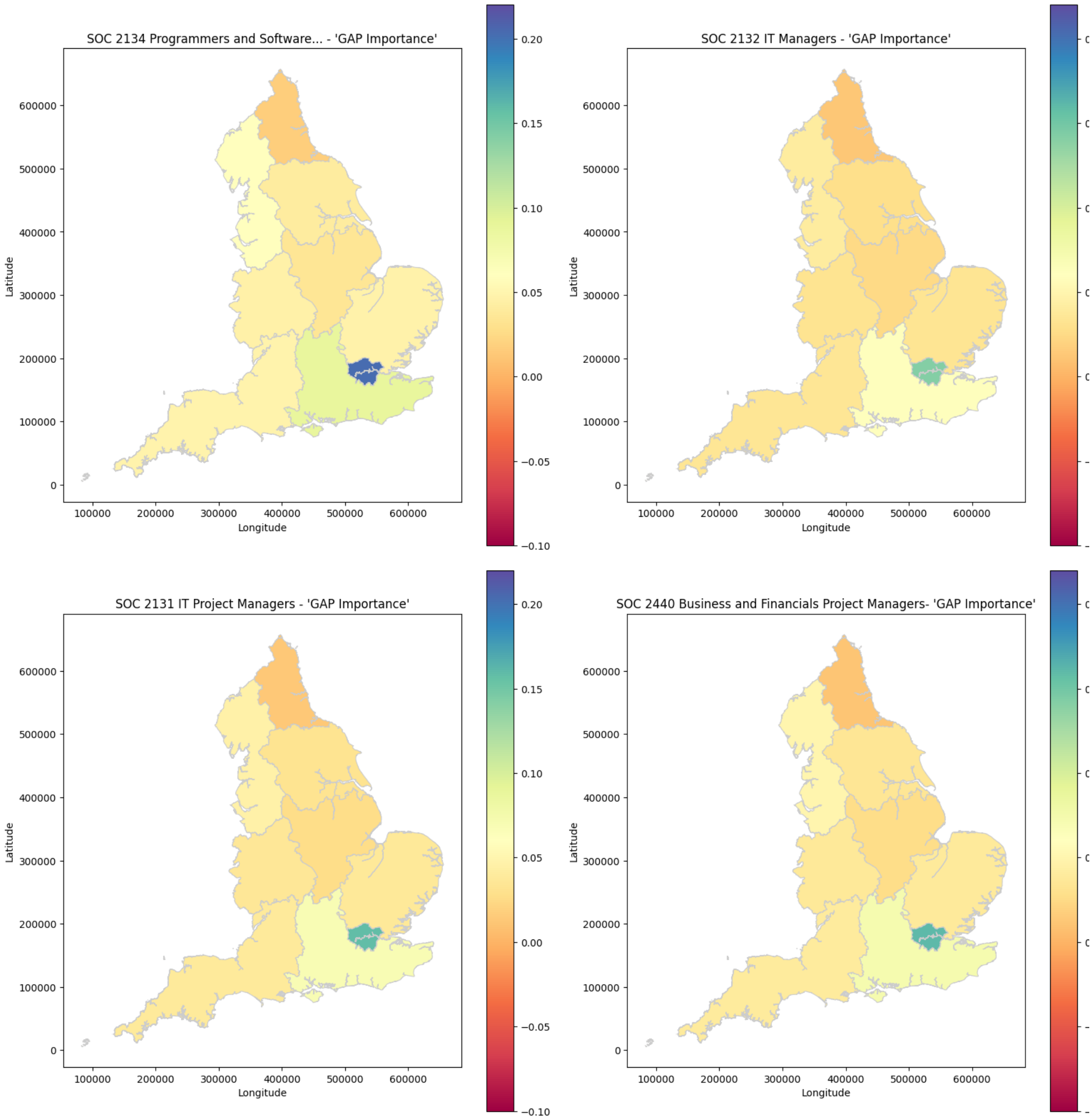


Figure 4. Regional Gap Importance by SOC-4

Figure 4 outlines the distribution of skill gaps in the chosen job categories across different UK regions, quantified by the Gap Importance Index. This index was derived from the ratio between the forecasted demand for positions and the projected size, or supply, of the workforce in these occupations, highlighting the unmet needs for specific IT and business occupations in major English regions **(**Figure A-B-3, Figure A-B-4). Greater London exhibits the most significant skill gaps across all job categories, indicating that the demand for IT and business professionals substantially exceeds supply in the region. The skills gap is most pronounced for Programmers and Software

Developers (SOC 2134), with the South East exhibiting the second largest predicted skills shortage. While most areas show a relatively balanced demand and supply for IT Managers (SOC 2132), the South East and the North West display a noticeable shortage of skills. Additionally, IT Project Managers (SOC 2131) and Business and Financial Project Managers (SOC 2440) also experience higher-than-average skill demands in certain areas, particularly in Greater London, which could pose challenges to local businesses' recruitment strategies. The skill demands in these specific roles highlights the importance of non-technical roles and skills, which are growing in parallel with pure digital skills. For example, management and business roles require skills in communication, decision-making, planning, general critical thinking and problem-solving skills. This finding aligns with the skills forecast provided by the NEFR (Dickerson et al., 2023).

These analytical results provide valuable insights into the labour market dynamics of specific IT and business professions across various UK regions, pinpointing areas where targeted educational and training measures may be necessary to bridge the gap between demand and supply. Moreover, this information is crucial for shaping regional development strategies and educational policies, especially in an economy that is rapidly evolving and heavily reliant on technology.

Current Skills

The results of the web scraping of job adverts from Indeed.com provided a detailed picture of the current skills landscape for the targeted occupations. Our analysis revealed a high demand for both technical and soft skills across these occupations. Technical skills, such as familiarity with various software and programming languages appeared frequently in IT and finance job postings, indicating a strong market need for data manipulation and software development competencies. Furthermore, soft skills, such as leadership, organisation, and customer service were universally requested across all sectors, signifying their continued impact in the current IT and business environment, for all job levels.

Figure 5 below displays the 15 most frequent skills that appear on job postings in the relevant roles. Leadership is the most common skill sought out across all job postings, followed by competency of Microsoft Excel and Office. This demonstrates that even in jobs requiring high technical expertise, foundational technical and soft skills are the highly desired. Figure 6 provides a composition of the top 15 skills, broken down by proportion of soft skills (43.6%) and technical skills (56.4%). The relative balance between the two categories suggests they are seen as equally important for employers, and thus candidates are expected to display good aptitude in both spheres. This resonates with findings from the Skills Imperative 2035 (Dickerson et al., 2023), which highlight the need for soft skills and foundational data literacy as the most important skills in the coming years. However, as we have tailored our search to highly technical professions, it is expected that we will observe slightly more demand in the technical skill categories.

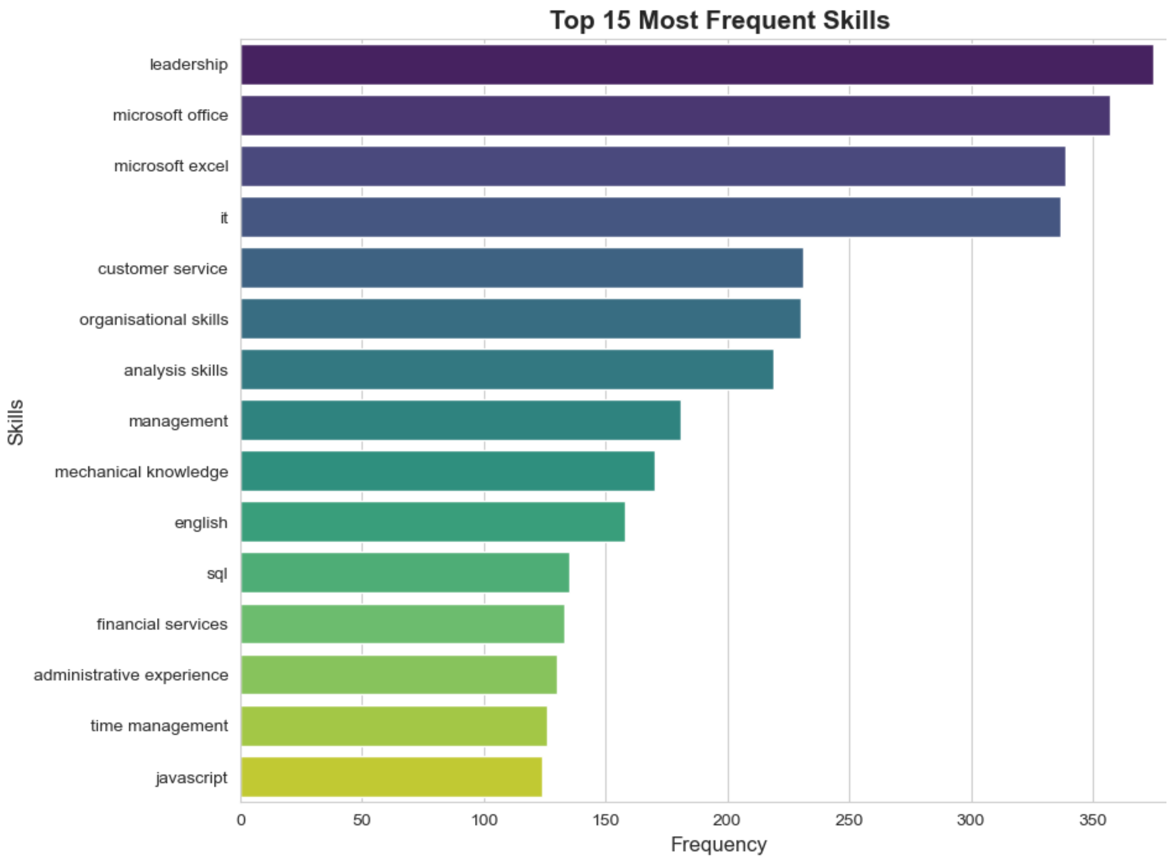


Figure 5. Top 15 Most Frequent Skills

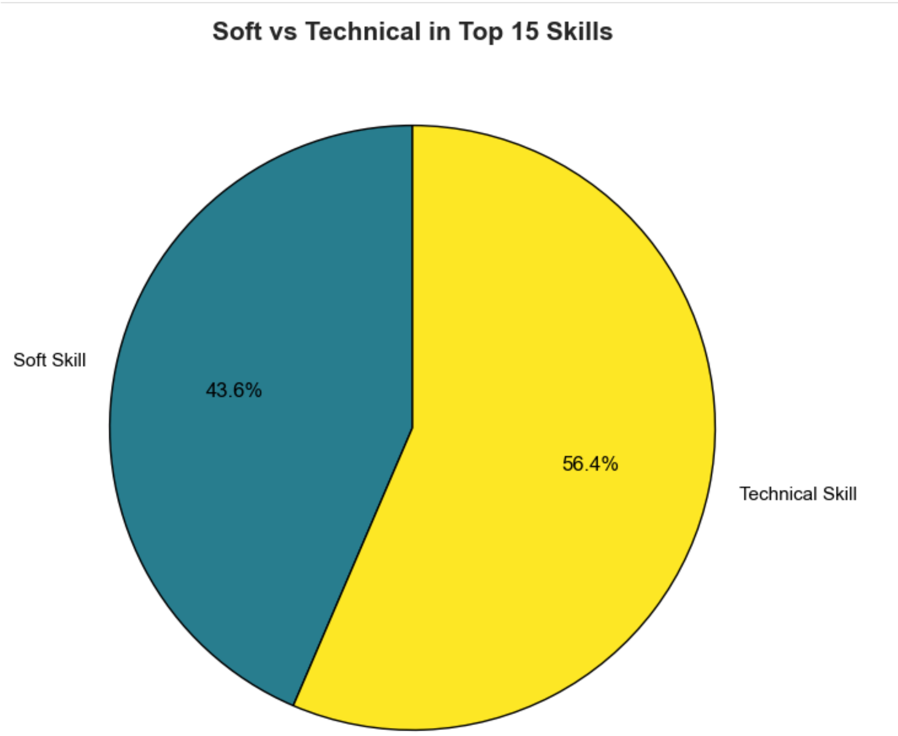


Figure 6. Soft vs Technical in Top 15 Skills Overall

Figure 7 shows the top 10 most frequent job positions. Software engineers and developers are the positions with most overwhelming demand currently.

However, it should be noted that due to the different wordings of adverts, it was a challenging task to normalise all job postings. For example, more niche professions which are emerging in the last few years, such as AI and ML engineers might appear under various names despite sharing similar characteristics. Moreover, due to the limited scope of our project and computational resources, we were not able to capture an extended period of job postings with our model.

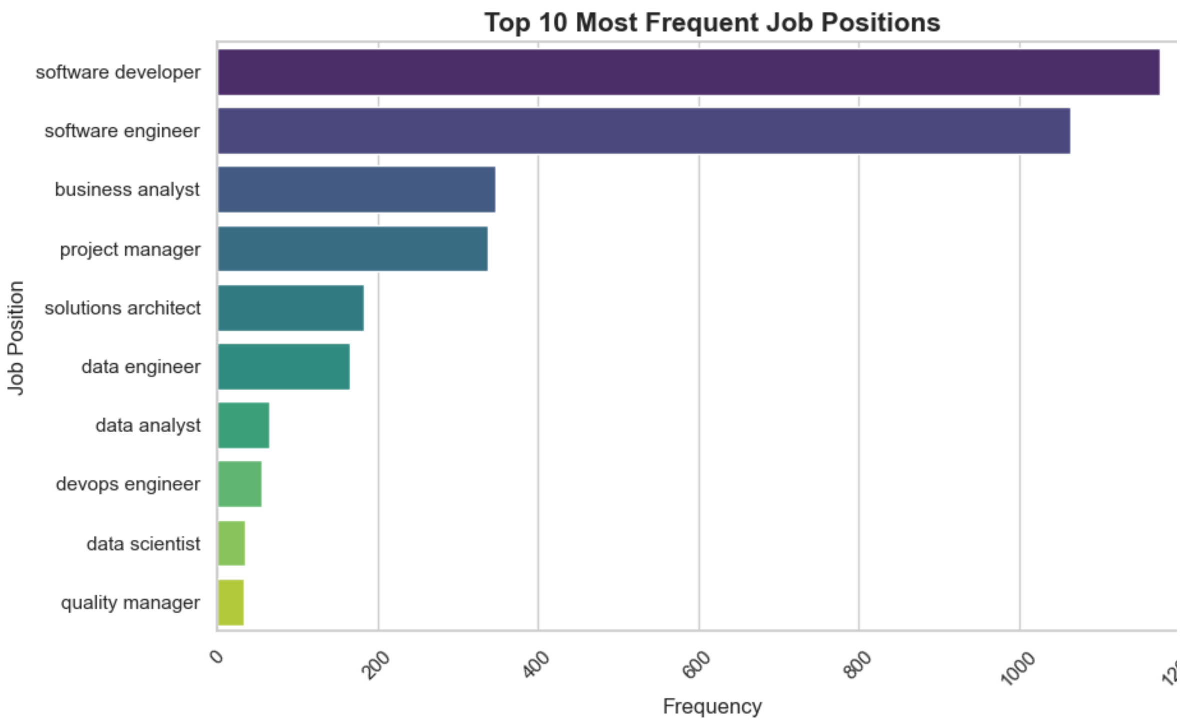


Figure 7. Top 10 Most Frequent Job Positions Advertised

Figure 8 provides a breakdown of the top 3 skills in the most frequently advertised position. Notably, the most desired skill for software developers was JavaScript (44.5%), while C++ was at the top for software engineers. However, business analyst and project manager positions required more soft skills, such as data analysis and management. Although soft skills are crucial among all occupational levels, they are especially prominent at senior positions due to their key role for successfully managing teams and projects. Cloud computing and architecture were particularly desired for solution architects, and SQL and Python were most prominent among data analysts.

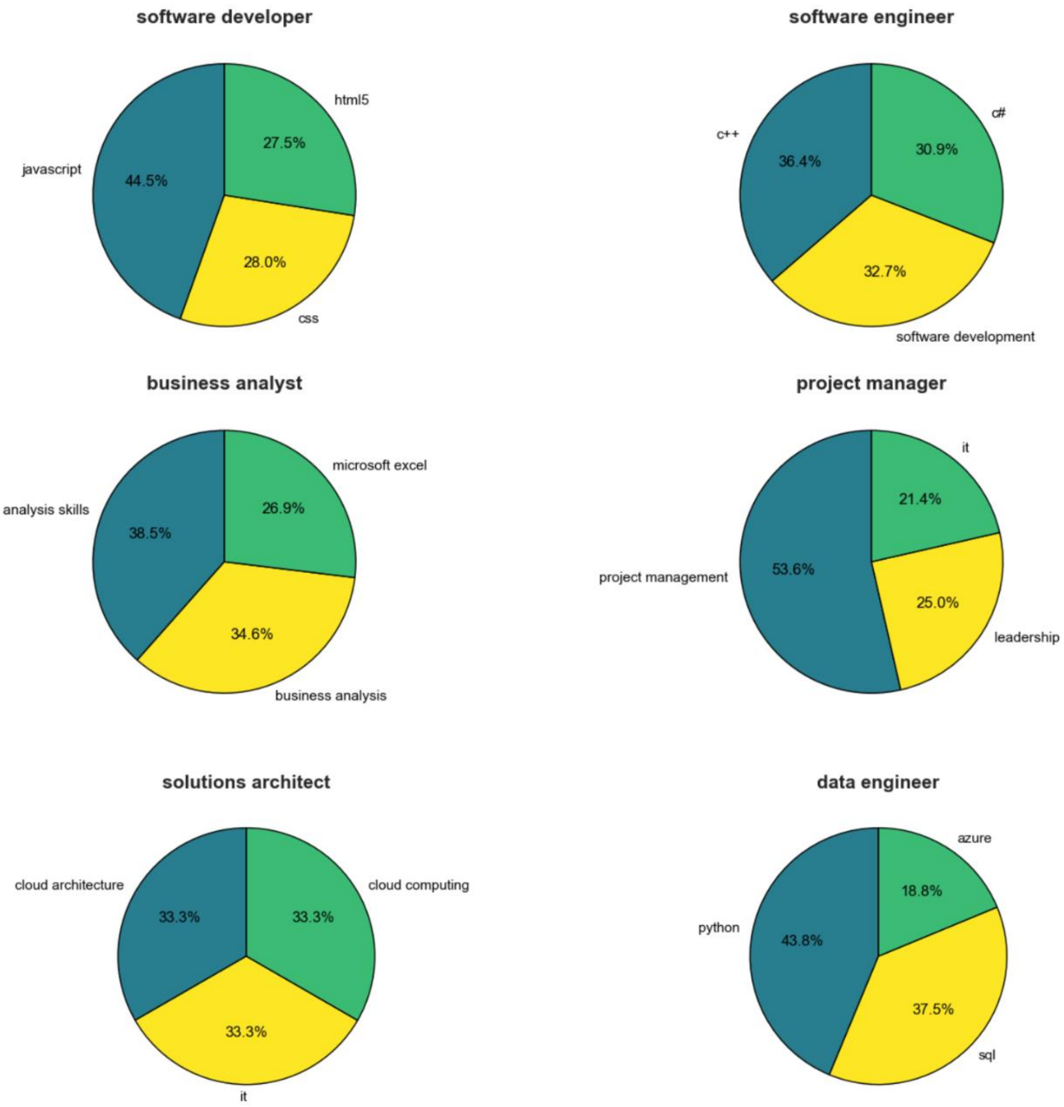


Figure 8. Top 3 skills in the top 6 positions

2035 Labour Market Forecast

We selected Elastic Net, KNN, and SGD Regressor to assess performance on labour predictions. As shown in Figure 9, KNN and SGD performed better with a lower RMSPE and MSE.

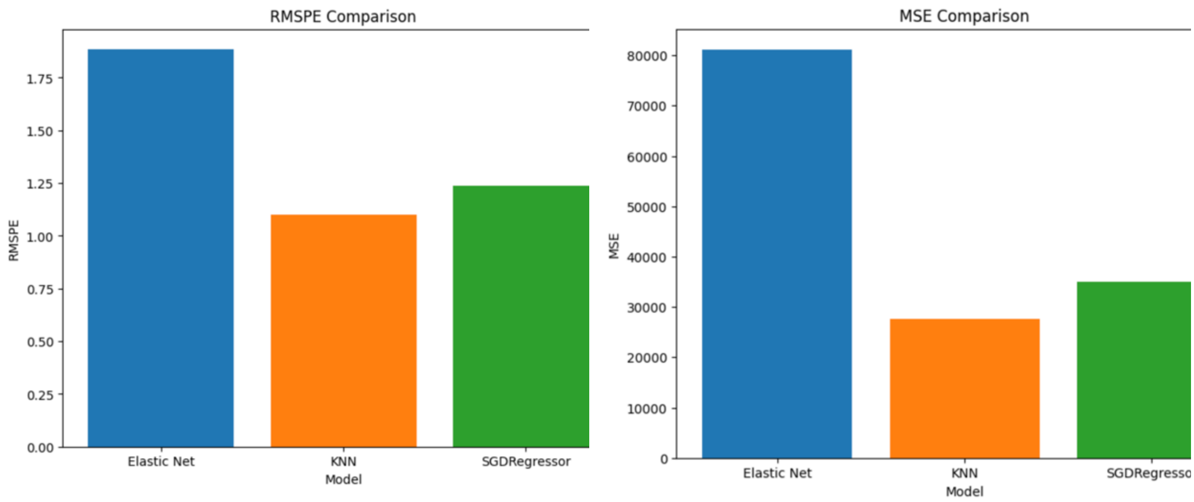


Figure 9. Performance Metrics Comparison

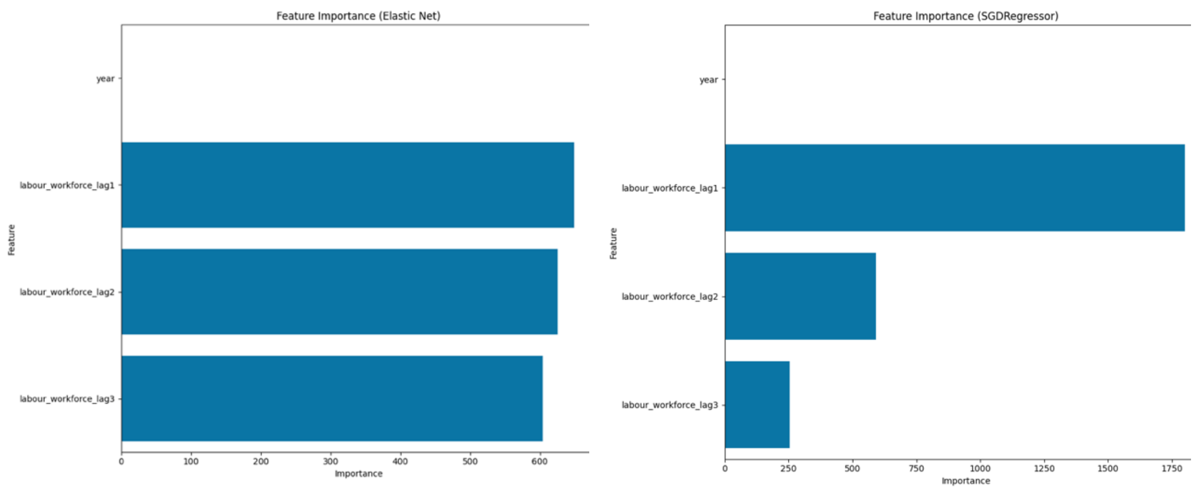


Figure 10. Feature Importance Comparison

Figure 10 shows the feature importance for Elastic Net and SGD Regressor. The lag features for the previous 3 years are all significant predictors for the model, with the most recent (lag1) being the most important. The balanced importance of features is due to the regularization, which can reduce overfitting without significantly limiting model complexity. Alternatively, SGD Regressor places the most significant emphasis on lag1. Predictions are more heavily impacted by recent observations, implying that the model is less suited to capture longer-term trends. However, due to the inherent unpredictability of forecasting market size until 2035, any model would represent an uncertain estimate of the future. As our models forecasts data based on past fluctuations, they should be treated as an expected trajectory of the market, rather than an exact forecast, which is why we utilised 3 different models to provide a comparison of the different estimates obtained. Nevertheless, this was the most feasible way to achieve our 4th objective and provide guidance for the IBM Skills Build Platform in the next 10-15 years.

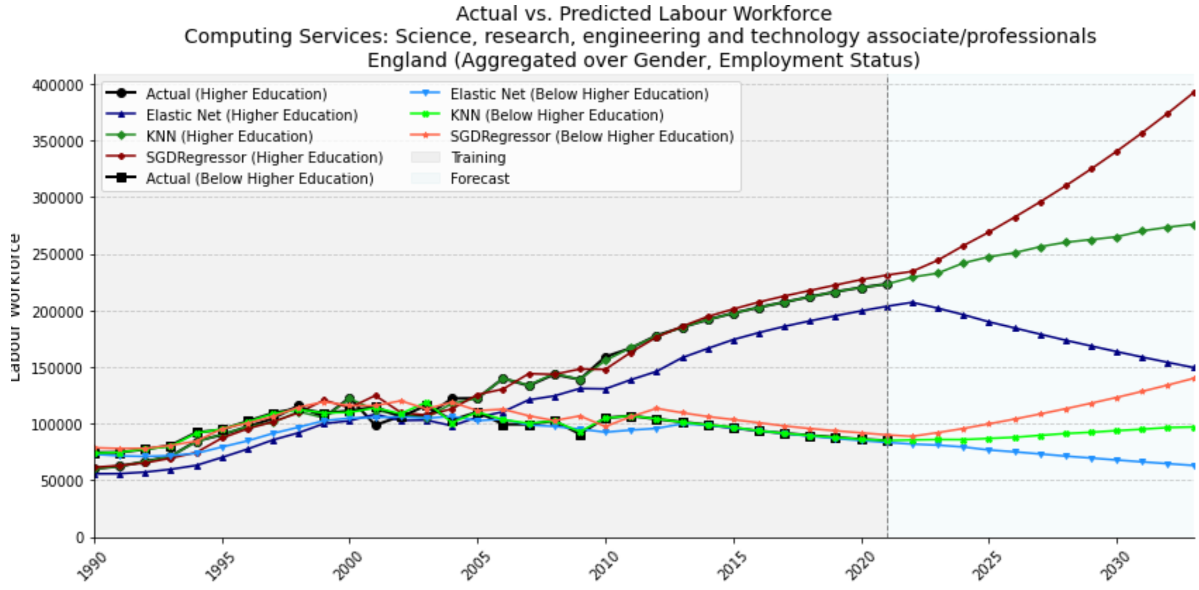


Figure 11. 2035 forecast for STEM associates and professionals

Figure 11 demonstrates the predictions for STEM associate and higher professionals in Computing Services up to 2035. Similar graphs were generated for the same occupational category, along with corporate managers and business professionals for the Computer Services, IT, Scientific Research, and Financial Services sectors (provided in the Appendix). The SGD model provides the best fit for the market growth among people with higher education and above, while KNN performs best among those with lower education. This trend is repeated across all the findings, likely due to KNN's ability to capture repeated patterns, which are possibly more pronounced in the lower education category. Additionally, SGD captures trends in the higher education and above category, which potentially integrates more diverse career paths, due to the model's efficiency at handling high dimensional data. Elastic Net fails to account for patterns in this category due to regularisation techniques preventing it from learning complex trends.

Overall, our findings suggest that business and science professionals/managers with higher education will experience a steady growth until 2035 across the 4 relevant sectors. This resonates with findings from the NFER (Dickerson et al., 2023) which project that highly skilled occupations will dominate the 2035 market. Thus, training tailored to Computer Services, IT, Scientific Research, and Financial Services sectors, encompassing soft skills and long-term digital upskilling, should be the focus of IBM Skills Build in the long term, evidenced by the projected market growth of these industries by 2035.

Limitations

Defining Skill Gaps

One significant challenge we encountered was devising an appropriate methodological framework to address skills gaps given the various perspectives on definitions and measures in the literature. There is no uniform definition of a 'skill gap' - for example, Employer Skills Survey report (IFF, 2023) addresses skill gaps as "skill-shortage vacancies" - job positions that are hard to fill due to applicants lacking necessary skills, qualifications, or experience. Dickerson et al. (2023) define it as occupational demand exceeding the supply of appropriately skilled workers, which we adopted in our methodology. However, it was challenging to compare findings of different reports and assess their reliability, as well as their relevance to the scope of our project.

Data Availability

Publicly available data rarely contains information on specific skills. Reports focus on predicting skills shortages in industries and occupations, but the research falls short in the discussion of detailed breakdown by skills, e.g. specific programming languages and certifications. Furthermore, businesses actively utilise skills gap information for profit and are reluctant to share this data publicly. However, the Skills Imperative 2035 report (Dickerson et al., 2023) provides detailed predictions of UK skills gap in the next 10 years. As the UK does not have an official guideline on mapping occupational codes to skills, the authors had to utilise the American O*NET schema to incorporate skills in their projections (Dickerson et al., 2023). Numerous contributors and official bodies collaborated on this research over the course of years, and despite the appeal of their methodological insights, we were not able to replicate the framework due to the extremely limited scope of this project. Additionally, we tried to obtain data directly from the authors but unfortunately did not receive a response.

Data Quality

The ESS dataset specified in the project assignment posed a significant issue for our analysis due to many missing value instances, insufficient number of data points for predictions, and inconsistencies in key variables, such as job sectors, across the time series data.

This led us to seek out different datasets, which also came with their unique data quality discrepancies. The web advertisement dataset we utilised for the prediction of skills gaps lacked the exact occupational and regional granularity we needed to efficiently complete our project objectives. This resulted in building separate models and indexes that accumulated different errors and brought a further degree of uncertainty. Additionally, the web ads data also contained imputed values which likely affected our results.

Moreover, neither of the available datasets contained detailed breakdown of specific skills, which was a key requirement of the partner. Hence, we resorted to identifying skills shortages and then obtaining information for skills by using web scraping technologies. However, without extensive computational resources and a limited schedule for the project, the results of the web scraping only provide a small snapshot of the current skills demand.

Uncertainty in Future

Due to the inherent uncertainty in predicting the future, producing a realistic assessment of what the labour market might look like in 2035, based on what we know now, is a challenging task. Unforeseen events, such as global shifts and emerging modern technologies could alter the future skills landscape drastically. One example for such an occurrence was the COVID-19 pandemic, which had a significant impact on the labour market and job demands. Hence, any forecasts based on current data will likely paint a biased picture of actual skills gaps in the next 10-15 years. Therefore, we are aware of the inherent limitations of our modelling results and their ability to reflect the actual future fluctuations of the skill gap landscape. Nevertheless, we have applied flexible ways to analyse a multitude of diverse sources and obtain tangible results that can serve as a guideline for the IBM Skills Build platform, despite the uncertainty associated with predicting skills gap trends in the future.

Recommendations

We identified specific challenges (Table 4) through our literature review and independent study. We have formulated the following recommendations to IBM Skills Build based on our understanding of these challenges and our assessment of the platform today.

Table 4. Recommendations by Theme

Challenge	Recommendation
Significant skills gap in soft skills (i.e., leadership, management, communication)	IBM Skills Build has a strong repository of non-technical skills (i.e., Design thinking, project management, etc.). The nature of soft skills is inherently social. We recommend introducing a **social aspect** to soft-skills focused courses through: • Group-based projects. • Simulation based on real business problems. • Access to seminars and conferences • Mentorship programs
Lack of employer investment in training	Reduced investments in on-the-job training for employees implies that employees' skills may become obsolete or in-adequate with changing technology and business processes. We recommend focusing on: • **Generalization:** Courses that focus on specific technologies should also incorporate core principles (i.e., course on C+ includes software engineering principles transferable across languages) • **Application:** There are extensive educational resources on theory and technical knowledge areas. Where possible, learned should be encouraged to apply concepts through short and long form exercises (i.e., review a research paper and apply approach to a business problem)

Current platforms do not fully address the needs of employers and workforce	**Employer needs:** Skills training needs to closely align with the requirements of employers. We recommend ensuring IBM skills tracks (i.e., data analyst, cloud computing, etc.) meets those need by assessing the training against WorldSkills Occupational Standards (WSOS). WSOS provides “rubrics” by occupational role with the knowledge, skills, and capabilities that underpin the role. In addition, WordSkills also tests the quality of external training programmes and can support IBM in a review. **Workforce needs:** Learners are keenly interested in platforms that engage and support their learning. In order to increase engagement, we recommend exploring different formats for IBM courses (i.e., bootcamp styles, virtual classrooms, etc.) and incorporating user-friendly learning tools (i.e., memorization widgets, smart agents, etc.).
One size fits all approach in training platforms may not be the right fit for all learners	IBM Skills Build has a variety of IT and AI tracks, such as Cloud Computing and Data Science. However, these career tracks currently offer a very broad overview of the topics. As such, we recommend IBM tailor the content to different levels of expertise. For example, content could be separated into beginner, intermediate, and advanced levels or specific **knowledge tracks** could be created based on **learner archetypes** (experienced CS grad, career switcher, etc.). Advanced and specialist digital skills, over and beyond low-level AI, are of huge importance. For specialist and sector-specific skills, IBM could consider **partnering with specific employers** to co-create more advanced courses (i.e., advanced NLP methods).

Project Reflections and Management

In this project, we collaborated with an industry partner from IBM to achieve the ambitious goal of predicting future skill gaps and identifying areas of improvement for the IBM Skills Build platform by applying preprocessing, cross validation, modelling and critical analysis of various data sources.

Project Workflow

The first stage of the project involved an introductory meeting with our industry partner, helping us get acquainted with the objectives and the key priorities for IBM. We implemented the MoSCoW method to prioritise tasks and plan the project efficiently.

We chose an initial project manager and divided into two teams – one dedicated to exploring the ESS dataset, and one focusing on literature review and supporting datasets. As we progressed with the project, we implemented a rotational project manager schedule, where each week a different person would set priorities and guide the team. This helped us distribute the workload fairly and stimulated team members to take responsibility and improve their leadership skills. It also helped us manage individual deadlines, and personal arrangements, accommodating every team member's current needs.

We organised our workflow with a detailed Gantt chart, allowing us to keep track of project milestones and deadlines, as well as the teams responsible for each task. Updates and feedback on key deliverables were facilitated through Teams, Google Collab, and weekly working sessions and progress meetings in person.

Key Challenges and Adaptations

One early challenge involved locating viable datasets to complete the analysis. The original dataset was impractical four our objectives, leading us to seek alternative sources. However, this was a particularly daunting task which was very time consuming, delaying our milestone timeline. We consulted both our industry partner and our academic mentor to communicate this issue and get guidance on how to proceed with our approach.

After thorough analysis, we opted for datasets that better met our criteria, though this increased our workload significantly in terms of data collection and processing. However, these additional datasets posed other limitations that we needed to address rigorously. Additionally, the dynamic nature of the labour market and unforeseeable factors presented challenges to predictive modelling, which unfortunately limited the certainty of our forecast. However, this bottleneck forced us to engage in creative and flexible techniques to accomplish our tasks. For example, we incorporated new data analysis techniques, such as web scraping and allocated the people with most suitable skill sets to these tasks. However, the lack of clear-cut-path also forced the team members to leave their comfort zone and acquire new technical skills and in-depth knowledge on the assigned topic.

Accomplishments

We successfully utilised advanced data analysis and modelling techniques to forecast labour demand and identify skills gaps, while acknowledging the realistic limitations of our approach in a critical manner. By integrating multiple data sources, we provided a comprehensive overview of the future landscape of the skills gap and delivered tangible insights to our industry partner. This project highlighted the importance of methodological flexibility in adapting to available data and ever-changing market conditions. It also emphasized the critical role of ongoing skills development and training in addressing mismatches in the labour market, which will help IBM enhance the quality of its educational products and better serve future labour market needs.

Bibliography

Dickerson, A., Rossi, G., Bocock, L., Hilary, J. and Simcock, D. (2023). *An analysis of the demand for skills in the labour market in 2035. Working Paper 3*. Slough: NFER. Available at: https://www.nfer.ac.uk/publications/the-skills-imperative-2035-an-analysis-of-the-demand-for-skills-in-the-labour-market-in-2035/

Dickerson, A. and Rossi, G. (2024). *An analysis of the demand for skills in the labour market in 2035: Revised tables and figures*. Slough: NFER. Available at: https://www.nfer.ac.uk/media/i25c3hxz/an-analysis-of-the-demand-for-skills_revised-projections_revised-tables-and-figures.pdf

Dabhi, K. et al. (2020). *Understanding the UK AI Labour Market: 2020*. Ipsos MORI. Available at: https://assets.publishing.service.gov.uk/media/60992b39d3bf7f2888d18f84/DCMS_and_Ipsos_MORI_Understanding_the_AI_Labour_Market_2020_Full_Report.pdf

DfE (2023). *Labour market and skills projections: 2020 to 2035.* Available at: https://www.gov.uk/government/publications/labour-market-and-skills-projections-2020-to-2035

IFF Research (2023). *Employer Skills Survey 2022: Research Report.* Available at: *https://assets.publishing.service.gov.uk/media/65855506fc07f3000d8d46bd/Employer_skills_survey_2022_research_report.pdf*

GMCC (2024). *Greater Manchester Local Skills Improvement Plan: Update Report.* Available at: https://www.gmlsip.co.uk/reports

GOV.UK (2023). *Calendar year 2022: Employer Skills Survey.* Available at: https://explore-education-statistics.service.gov.uk/find-statistics/employer-skills-survey/2022

GOV.UK (2016). U*K labour market projections: 2014 to 2024.* Available at: https://www.gov.uk/government/publications/uk-labour-market-projections-2014-to-2024

Papadopoulos, S. and Karakatsanis, I. (2015). *Short-term electricity load forecasting using time series and ensemble learning methods*. 2015 IEEE Power and Energy Conference at Illinois (PECI). doi:https://doi.org/10.1109/peci.2015.7064913.

ONS (2023). *Occupations of those in employment, by age and sex, England and Wales, Census 2021.* Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/occupationsofthoseinemploymentbyageandsexenglandandwalescensus2021

ONS (2024). *Labour demand volumes by Standard Occupation Classification (SOC 2020), UK.* Available at: https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/employmentandemployeetypes/datasets/labourdemandvolumesbystandardoccupationclassificationsoc2020uk

House of Commons (2024). *Economic Indicators (02783), April 17, 2024*. London: House of Commons Library. Available at: https://researchbriefings.files.parliament.uk/documents/SN02783/SN02783.pdf

Appendix A

GitHub Repository

https://github.com/rk-izak/DATA70202-IBM-2024/tree/main

Contains all the utilised datasets and Jupyter notebooks, as well as user guides and technical requirements for the execution of the code.

Appendix B

Importance Index:

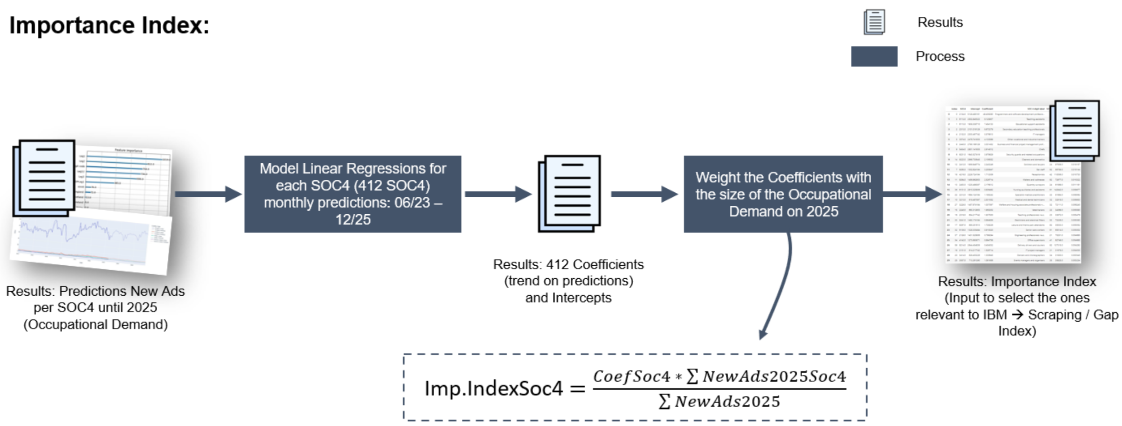


Figure A-B-1. Importance Index SOC-4

Regional Gap *Index (Part A) – Getting the Regional SOC4 demand*

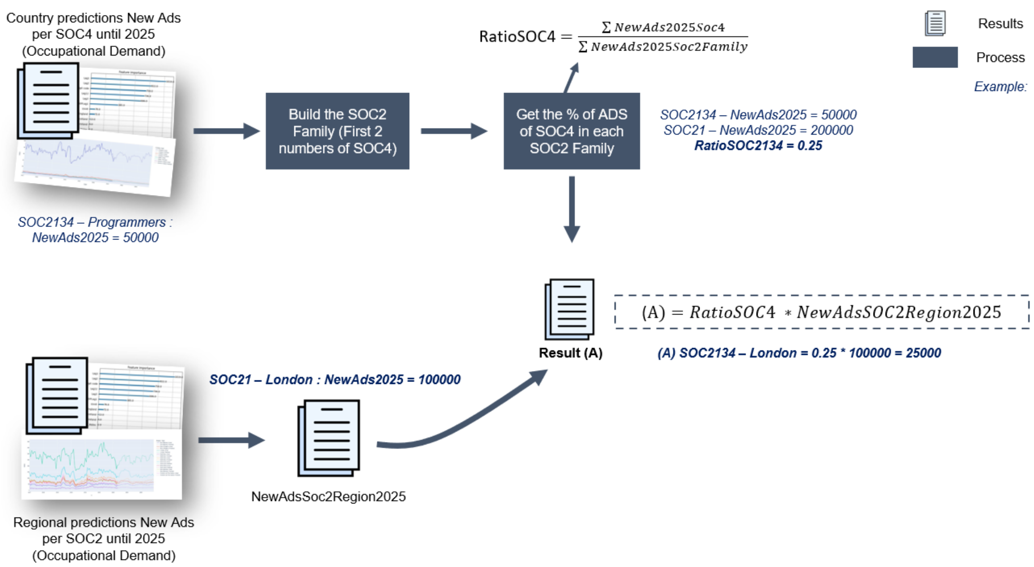


Figure A-B-2. Regional Gap Index A

Regional Gap *Index (Part B) – Getting the projected size of the labour market* by SOC4

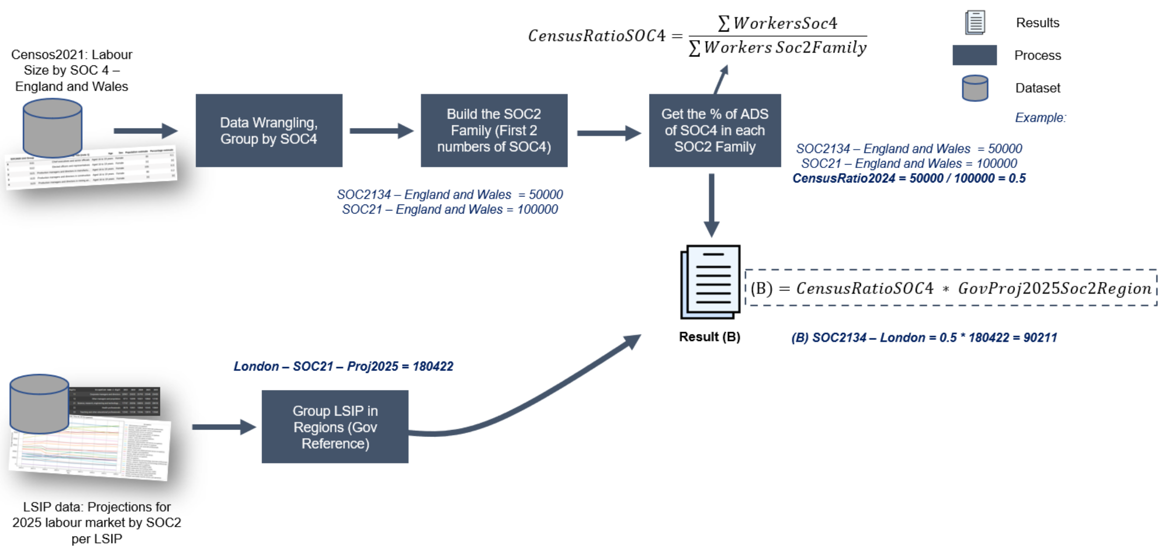


Figure A-B-3. Regional Gap Index B

Regional Gap *Index Equation* –

$$\mathrm{RegionalGAP}\mathit{IndexSOC4} = \frac{RatioSOC4 * NewAdsSOC2Region2025}{CensusRatioSOC4 * GovProj2025Soc2Region} = \frac{(A)}{(B)}$$

Regional Gap *Importance Equation* –

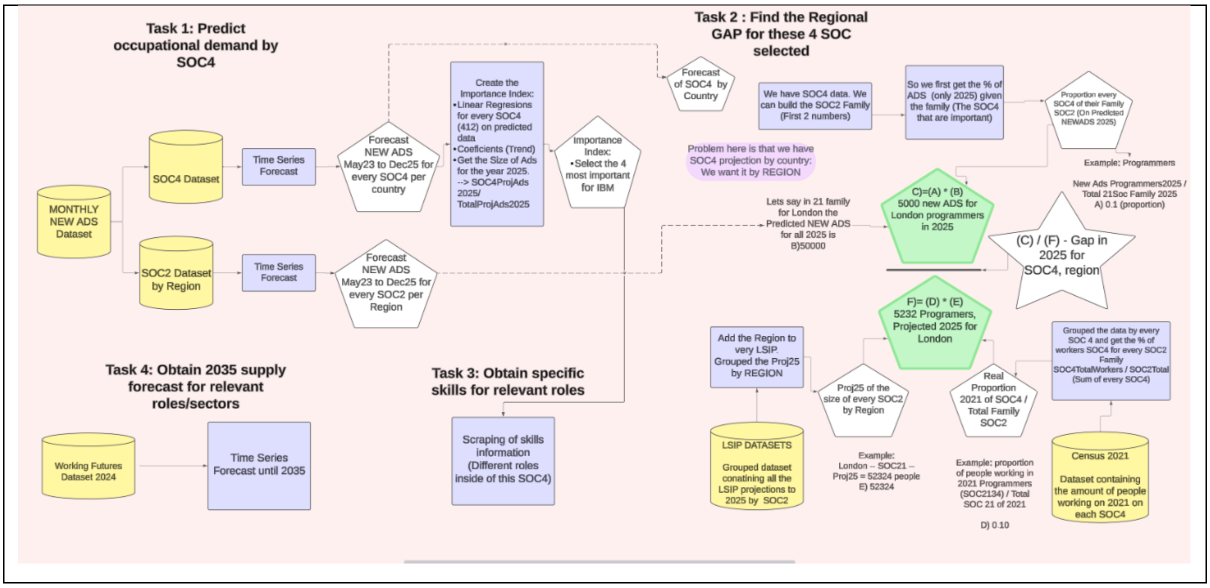
$$\mathrm{RegionalGAP}\mathit{ImportanceSOC24} = \frac{RegionalGAPIndexSOC4 * GovProj2025Soc4Region}{GovProj2025Soc4AllRegions}$$

Figure A-B-4. Regional Gap Equations

Appendix C

List of Tables and Figures

Methodology Flow Chart



- This flow chart visualises the steps undertaken in our methodological approach. We found that the complexity and detail we included were easier to follow with a graphic aid. Hopefully this will provide further clarification for the user base.
- Due to the image size, please zoom in to read the contents better