The Impact of Educational Attainment on Skilled Labor Shortage in the UK

Abstract—**This report analyses the UK's skilled labor shortage, emphasising the correlation between education levels and employment outcomes. Using data from the 2021 census, it explores how educational attainment influences occupation levels across Westminster constituencies and age groups. Visualisations, including scatter plots, choropleth maps, and bar plots, aid in understanding correlation, regional disparities, and trends. Finally, Policy implications underscore the need for targeted educational interventions to bridge the skills gap and boost economic productivity.**

I. Introduction

A skilled workforce is vital for the UK's economic productivity and growth, as a skilled worker can perform complex tasks, innovate, and adapt to new technologies, driving economic output. For instance, the Kenan Institute emphasizes that a workforce with a wide range of skills can accomplish more tasks with greater competency, thereby boosting economic productivity [1]. Similarly, the Scottish Government's National Strategy for Economic Transformation underscores that a skilled population is key to business productivity and economic prosperity [2].

Despite the clear benefits of a skilled workforce, the UK is currently experiencing a mismatch between the skills possessed by the workforce and those demanded by employers, particularly in sectors like IT and digital technologies. The Employer Skills Survey 2022 indicates that around a quarter of all employers had vacancies, with 10% reporting skill-shortage vacancies. Additionally, 5.7% of the workforce had skills gaps, indicating a lack of proficiency in certain areas [3]. This labour and skills shortage directly impacts business operations and economic performance. According to the report by Personio, there are around 1.12 million job vacancies in the UK, and 78% of UK organizations have reported reduced output, profitability, or growth due to skills shortages [4].

This skills shortage can be linked to the education level of the individuals in the UK. According to an OECD report, individuals with tertiary education in OECD countries, including the UK, are more likely to be employed than those without such qualifications. The report states that over 80% of the population with tertiary education is employed, substantially higher than those with lower educational qualifications [5]. Additionally, the Graduate Labour Market Statistics (GLMS) for 2022 reveals that graduates and postgraduates have significantly better employment outcomes than non-graduates [6].

This report aims to strengthen the above statements using the 2021 census of the UK. It will explore the relationship between educational attainment and occupation levels, providing insights for policymakers responsible for improving the skilled labor force and boosting the UK economy.

II. Data Exploration and Abstraction

The dataset used for this report is primarily derived from the 2021 census, encompassing all usual residents of the UK. This includes individuals typically residing in England and Wales but excludes non-UK-born short-term residents and visitors. Westminster Parliamentary constituencies categorise the data with the dataset's essential features, including age groups, the highest level of qualification, and current occupation.

During data exploration, it was observed that there were no missing values. The dataset includes a total of 573 Westminster constituencies and 11 age groups (ranging from 'Aged 15 years and under' to 'Aged 65 years and over' in 5-year increments). Moreover, The dataset categorizes the highest qualification levels into seven distinct groups. The first category is "No Qualification." The second category, "Level 1 and Entry-Level Qualifications," includes 1 to 4 GCSEs (grades A* to C), any GCSEs at other grades, O levels or CSEs (any grades), 1 AS level, NVQ Level 1, Foundation GNVQ, and Basic or Essential Skills. The third category, "Level 2 Qualifications," encompasses 5 or more GCSEs (A* to C or 9 to 4), O levels (passes), CSEs (grade 1), School Certification, 1 A level, 2 to 3 AS levels, VCEs, Intermediate or Higher Diploma, Welsh Baccalaureate Intermediate Diploma, NVQ Level 2, Intermediate GNVQ, City and Guilds Craft, BTEC First or General Diploma, and RSA Diploma. The fourth category, ”Level 3 Qualifications,” includes 2 or more A levels or VCEs, 4 or more AS levels, Higher School Certificate, Progression or Advanced Diploma, Welsh Baccalaureate Advanced Diploma, NVQ Level 3, Advanced GNVQ, City and Guilds Advanced Craft, ONC, OND, BTEC National, and RSA Advanced Diploma. The fifth category, ”Level 4 Qualifications or Above,” comprises degrees (BA, BSc), higher degrees (MA, PhD, PGCE), NVQ Level 4 to 5, HNC, HND, RSA Higher Diploma, BTEC Higher Level, and professional qualifications such as teaching, nursing, and accountancy. There is also a category termed “Others” for qualifications of unknown level, indicating instances where the census governing body could not classify the inputted qualification level among the predefined categories. Additionally, the dataset includes nine occupation categories: the first category is Managers, directors, and senior officials. The second category is Professional occupations, and the third is Associate professional and technical occupations. The fourth category is Administrative and secretarial occupations, the fifth category is Skilled trades occupations, sixth category is Caring, leisure, and other service occupations, seventh category is Sales and customer service occupations, the eighth category is Process, plant, and machine operatives, the ninth category is Elementary occupations.

This report focuses on individuals aged 25 to 64. This age range is chosen because those below 25 are primarily students and do not typically qualify as skilled employees, while those above 64 are generally at or beyond the state pension age and thus are unlikely to be working in high-skilled jobs. After filtering the dataset to include only the 25 to 64 age group, no records remained in the highest qualification category with unknown levels.

Additionally, the qualification codes and their descriptions were refined for clarity, making it easier for users to understand the order of levels (0-4). The naming of occupations was also improved, changing the category ”Doesn't Apply” to ”Unemployed or Student Living Away (During Term)”.

Data from the 2011 census was also extracted to provide a comparative analysis, allowing for an examination of changes in qualification levels over the past decade. This historical comparison helps contextualize the current state of educational attainment and its impact on the skilled labor force in the UK.

III. Task Definition

This report is designed to support multiple analytical tasks related to understanding education attainment levels and occupation levels and their interrelationships across different geographic constituencies and age groups. According to Munzner's task taxonomy [7], these tasks can be categorized as follows: Firstly, understanding the correlation between education attainment and occupation level falls under the ”Discover” category in Munzner's taxonomy, as the goal is to find new knowledge and insights about the relationship between two key attributes - education attainment levels and occupation level. The target is understanding the correlation or dependency between these attributes across the entire dataset. Furthermore, The query scope is to summarize all instances to discern patterns and clusters that reveal how changes in one attribute (e.g., higher education level) may influence the other (e.g., occupation level).

Secondly, understanding the comparison of qualification levels in the last 10 years is a discovery task aimed at identifying patterns in the qualification level attribute over a specific time period (the last 10 years). The target is to characterize the high-level behavior and trajectory of this attribute, such as whether qualification levels have increased, decreased, plateaued, or exhibited more complex dynamics over the decade.

Finally, the task is to find constituencies and age groups with low education levels. This task involves an exploratory search without a known target location. The targets are outliers or anomalies in multiple attributes simultaneously - in this case, low education attainment levels across the geographic constituency and age group dimensions. The query scope is to identify the specific instances (constituencies and age groups) that exhibit these outlier characteristics, likely to prioritize them for targeted interventions or policy actions.

IV. Visualisation Justification

To fulfill the aforementioned tasks, three dashboards were created in Tableau to assist policymakers.

The first dashboard aims to demonstrate the relationship between educational attainment and skilled labour employment. To begin with, It showcases a statement in a text box from a government website, elaborating that graduates and postgraduates have significantly better employment outcomes than non-graduates. Research supports the inclusion of text boxes in visualisations to facilitate storytelling and provide essential context and metadata, ultimately improving the effectiveness of visual data communication [8].

Furthermore, The dashboard provides a quick overview of key data points, helping users understand the broader data context. By highlighting the population surveyed, the percentage with graduate-level qualifications, and the percentage of skilled workers, these metrics show that a very low percentage of the population has high education, thus resulting in a low percentage of skilled workers. Summary metrics efficiently communicate key information, allowing viewers to grasp essential data points quickly.

Additionally, scatterplots are used to reveal the strong correlation between education and occupation levels through different data projection techniques (PCA & UMAP). The paper "Clustering Scatter Plots Using Data Depth Measures" [9] emphasizes that scatterplots are useful for identifying clusters or groups of data points, indicating meaningful subpopulations or trends within the data. By analyzing the dispersion and clustering of the dots, valuable insights about potential trends can be deciphered. Additionally, scatterplots can visually highlight outliers, which is crucial as they can significantly impact analysis and decision-making [9].

The clustering of different occupations based on an individual's highest qualification level was created using the Principal Component Analysis (PCA) dimensionality reduction technique. PCA was chosen above UMAP for this subset of the data as the occupation and qualification levels showcased simpler, linear relationships. Moreover, Research has established that PCA is a linear dimensionality reduction technique that projects data onto new orthogonal axes and captures maximum variance in the data. Thus, using linear transformation is advantageous for simple linear data.

Additionally, various age and qualification combinations were clustered according to different occupation levels using the Uniform Manifold Approximation and Projection (UMAP) dimensionality reduction technique. UMAP was used for this dataset as it has more features and is more complex in nature to project using PCA.

The second dashboard compares observations for the highest qualification levels between 2011 and 2021. The visualisations used include a grouped bar chart and a diverging bar chart. The main part of the visualisation employs a grouped bar chart, which effectively compares values across different qualification levels—for the two time periods, 2011 and 2021. This type of chart makes it easy to discern the relative proportions of each qualification level and detect shifts over time. The side-by-side arrangement of bars for different years makes it easy to identify trends and patterns in the qualification level distributions over time.

The diverging bar chart on the right showcases the percentage changes, with bars extending to the left for decreases and to the right for increases. This diverging layout effectively highlights the direction and magnitude of change for each qualification level, complementing the grouped bar chart. The explicit display of percentage changes in the diverging bar chart mitigates the potential for overlooking shifts, as viewers may not readily detect changes from the raw values alone [10].

Combining the grouped bar chart and diverging bar chart provides different perspectives on the same data, allowing viewers to explore the raw values and the relative changes effectively. This comprehensive visualization approach ensures that the trends and patterns in qualification levels are clearly communicated to the viewers [11].

Finally, the third dashboard highlights Westminster constituencies and their highest qualification levels, displaying the population percentage in each constituency that falls into each qualification category.

The visualisations used include a choropleth map, a table, and a bar chart. The choropleth map displays the geographic distribution of qualification levels across different constituencies in the UK, effectively visualising spatial data and identifying regional patterns or variations. This technique leverages the human ability to perceive and process spatial information, facilitating an understanding of geographic patterns and distributions [12].

Additionally, A table is used to display the distribution of the highest qualification levels across constituencies, showing the percentage of individuals in each qualification level relative to the total population of each constituency. Tables effectively present precise numerical data in a structured manner, leveraging the human ability to scan and compare numerical values in a grid format [13]. Furthermore, A bar chart illustrates the distribution of the population according to age groups, providing insights into the demographic makeup. This visualization helps identify the particular age groups in each constituency that may need more attention for educational attainment.

Finally, Color coding is utilised in the map, table, and bar chart to distinguish between values. A higher observation value leads to a darker shade; consequently, the lower the observations, the lighter the shade. This technique takes advantage of the human ability to perceive and differentiate colors, facilitating quick visual recognition and comparison. Using color, length (in bar charts), and spatial positioning (in the map) serves as effective visual encodings, capitalising on the human perceptual system's ability to process and interpret these visual cues [14]. It also leverages preattentive processing, enabling viewers to quickly perceive and distinguish different values or categories without conscious effort.

V. Evaluation

The visualisations were evaluated by a peer discussion group using Munzner's four levels of validation. The positive feedback highlighted several strengths. Firstly, the visualizations clearly targeted the domain of education and employment statistics, providing relevant information for policymakers and stakeholders in the UK. The data abstractions, such as categorical qualification levels, age groups, occupation types, and Westminster constituencies, were deemed appropriate for comparing educational attainment, exploring employment patterns, and identifying target groups for policy interventions. Additionally, using familiar chart types, such as bar charts and scatter plots, facilitated understanding and interpretation of the data. The map visualization in the third dashboard was particularly praised for providing geographical context and aiding in understanding regional variations. The color coding across the visualisations was also well-received for its effectiveness in distinguishing different categories and values.

However, the group provided several suggestions for improvement. Firstly, the dashboards were noted to be very text-heavy, requiring users to read a lot rather than visualize the information. Secondly, the occupation versus qualification scatterplot had labels within the plot, making it difficult to understand the clusters. Finally, the text for qualification levels was not trimmed, resulting in a cluttered appearance that made the visualizations feel claustrophobic. Based on this feedback, the dashboards were updated to reduce text density, improve label placement in the scatterplot, and streamline the qualification-level text for better clarity and visual appeal.

VI. Conclusion

This report provided a comprehensive analysis of the UK's socio-economic problem of skilled labor shortages, particularly emphasising the relationship between education levels and employment outcomes. The findings underscore a clear correlation between higher educational attainment and skilled occupation roles, reinforcing the importance of education in enhancing workforce competency and economic productivity. Furthermore, The data revealed that areas like Birmingham and its adjacent constituencies, such as Warley and West Bromwich, have the highest percentage of the population with no qualifications, highlighting regions that may require targeted educational interventions. Additionally, the regions in the west of the UK are less educated than the east. Moreover, Over the past decade, there has been a positive trend with more individuals attaining higher qualification levels, indicating gradual progress in educational attainment and, consequently, skilled labour availability.

For policymakers, investing in education, particularly in areas with lower qualification levels, is crucial for addressing the skilled labor shortage. Policies focusing on enhancing educational infrastructure, providing vocational training, and creating pathways for higher education can help bridge the skills gap, thereby boosting economic productivity and growth.

Finally, the coursework illuminated the importance of using effective visual tools to convey complex data insights. Familiar visualisation techniques such as bar charts, scatter plots, and choropleth maps were instrumental in making the data comprehensible and actionable for policymakers. Additionally, the importance of an iterative design process was learned. Creating initial visualisations, receiving feedback, and making necessary refinements are crucial to ensure that the final product effectively communicates the intended message. Peer reviews and discussions reveal blind spots and provide new perspectives that enhance visualisation. Finally, Providing contextual information helps users interpret the visualisations accurately.

To conclude, this coursework has highlighted the importance of effective information visualisation in understanding and addressing complex socio-economic issues. By leveraging clear, contextual, and interactive visualisations, policymakers can make more informed decisions to enhance educational outcomes and address skilled labor shortages, ultimately driving economic growth and productivity.

References

[1] G. Cohen, "Skills in the Workforce: Why They Matter for Economies," kenaninstitute.unc.edu, 2023. https://kenaninstitute.unc.edu/kenan-insight/skills-in-the-workforce-why-they-matter-for-economies/ (accessed May 19, 2024).

[2] Scottish Government, "Skilled workforce," www.gov.scot, Feb. 2022. https://www.gov.scot/policies/economic-growth/skilled-workforce/ (accessed May 19, 2024).

[3] GOV.uk,“Employer Skills Survey , Calendar year 2022,” explore-education-statistics.service.gov.uk, Sep. 28, 2023. https://explore-education-statistics.service.gov.uk/find-statistics/employer-skills-survey/2022 (accessed May 19, 2024).

[4] Personio, “Understanding the Labour Shortage in the UK,” www.personio.com, 2023. https://www.personio.com/hr-lexicon/labour-shortage-uk/ (accessed May 19, 2024).

[5] OECD, “How does education affect employment rates?,” OECD iLibrary, Sep. 09, 2014. https://www.oecd-ilibrary.org/education/education-at-a-glance-2014/how-does-education-affect-employment-rates_eag_highlights-2014-13-en (accessed May 19, 2024).

[6] GOV.UK, “Graduate labour market statistics, Reporting Year 2021,” explore-education-statistics.service.gov.uk, Jun. 29, 2023. https://explore-education-statistics.service.gov.uk/find-statistics/graduate-labour-markets (accessed May 19, 2024).

[7] T. Munzner, Visualization analysis and design. CRC press, 2014.

[8] S. L. Franconeri, L. M. Padilla, P. Shah, J. M. Zacks, and J. Hullman, “The Science of Visual Data Communication: What Works,” Psychological Science in the Public Interest, vol. 22, no. 3, pp. 110–161, Dec. 2021, doi: https://doi.org/10.1177/15291006211051956.

[9] Z. Zhang, “Clustering Scatter Plots Using Data Depth Measures,” Journal of Biometrics & Biostatistics, vol. 04, no. 03, 2013, doi: https://doi.org/10.4172/2155-6180.s5-001.

[10] R. A. Rensink, “Change Detection,” Annual Review of Psychology, vol. 53, no. 1, pp. 245–277, Feb. 2002, doi: https://doi.org/10.1146/annurev.psych.53.100901.135125.

[11] C. Ware, “Chapter ten - interacting with visualizations,” in Information Visualization (Third Edition), Third Edition.Morgan Kaufmann, 2013, pp. 345–374. doi: https://doi.org/10.1016/B978-0-12-381464-7.00010-7.

[12] D. R. Montello, “Cognitive Map-Design Research in the Twentieth Century: Theoretical and Empirical Approaches,” Cartography and Geographic Information Science, vol. 29, no. 3, pp. 283–304, Jan. 2002, doi: https://doi.org/10.1559/152304002782008503.

[13] R. Spence, Information Visualization, 3rd ed. Cham: Springer International Publishing, 2014. doi: https://doi.org/10.1007/978-3-319-07341-5.

[14] C. Ware, “Chapter Four - Color,” in Information Visualization (Fourth Edition), Third Edition.Morgan Kaufmann, 2021, pp. 95–141. doi: https://doi.org/10.1016/B978-0-12-381464-7.00004-1.