Exploring Ethnic Structure and Employment between 2011 and 2021 Analyzing Impact of COVID-19 on Employment in 2021 based on Bayesian Prediction.

Abstract

Research Objectives

This study investigates the socio-economic transformations in labour force composition and ethnic distribution across regions of England and Wales by leveraging the UK Census datasets from 2011 and 2021. In particular, we aim to assess the impact of the COVID-19 pandemic on employment by comparing observed 2021 data against counterfactual predictions derived from pre-pandemic trends (2011–2019). A secondary goal is to explore the evolving patterns of migration and ethnic diversity, and their intersection with labour market resilience and vulnerability.

Methodology

One application of Bayesian models-the Bayesian linear regression model is adopted in prediction of labor economic data in 2021. The model was trained on annual regional employment statistics from 2011 to 2019, which enabled probabilistic forecasting of 2021 labour force metrics excluding the impact of pandemic effects [1]. By comparing these predictions with actual post-pandemic census observations, I evaluated the structural shocks induced by COVID-19 and the residual and relative deviations were visualized. Dimensionality reduction was conducted using UMAP [2] and Neuroscale [3], two complementary nonlinear projection techniques. UMAP preserves local neighbourhood structure in high-dimensional spaces, while Neuroscale optimizes for global distance preservation using a Radial Basis Function network. These projections were used to support interpretable cluster analysis and interactive visualisation.

The visualization were evaluated using task-oriented inspection, focusing on the clarity, interpretability, and alignment of each view with the analytical goals. Visual encodings were examined against established design heuristics (Munzner, 2014) [4], ensuring they support comparison, trend identification, and spatial reasoning across the dataset.

Key Findings

Results suggest that due to pandemic, there are significant deviations between reality and prediction from projected employment trends, particularly in regions with historically low economic activity. The rise in economic inactivity (unemployment or non-participation) was consistently above predicted levels in these areas, indicating local labor structural vulnerabilities that were potentially aggravated by the pandemic [5]. Ethnic clustering in the projection space reveals increasing heterogeneity in regional demographic composition. Notably, some ethnic minority groups exhibited a bifurcated pattern of either heightened resilience or acute vulnerability in employment outcomes [6]. These findings highlights some complex interplay between migration, ethnicity, and labour market dynamics in before and after the pandemic COVID-19.

1 Introduction

The intersection of labour force composition and ethnicity can be a helpful perspective in interpreting socio-economic resilience and inequality in the United Kingdom. Between the 2011 and 2021 national censuses, the COVID-19 pandemic presented an unprecedented disruption to employment systems, with uneven impacts across regions and demographic groups. This study focuses on England and Wales, aiming to analyse labour market outcomes in the post-pandemic period through ethnic or regional variation. By examining employment-related census data alongside historical labour trends, the project implement visualising spatial disparities and demographic shifts, providing insight into the evolving structure of the UK workforce following a major global crisis.

2 Data Preparation and Abstraction

Census data was integrated and standardized across multiple official sources, population-level datasets from the 2011 and 2021 UK censuses were combined, covering ethnicity and labour force characteristics for individuals aged 16 to 74. To ensure structural compatibility for temporal analysis, all variables were harmonized, including unifying column names, reconnecting and reconciling geographical labels, and removing inconsistent metadata.

Bayesian Forecasting of Labour Force in 2021 without Epidemic Impact

To enable comparative analysis with 2021 ground truth data, a probabilistic forecasting model was developed based on regional labour force indicators from 2011 to 2019. A Bayesian linear regression framework was implemented, with labour indicators (e.g. in employment, unemployment) treated as responses y_t and time-indexed covariates denoted X_t:

$$y_t = X_t\beta + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, \sigma^2)$$

Priors were placed on regression weights and variance:

$$\beta \sim \mathcal{N}(0, \tau^2 I), \quad \sigma^2 \sim \text{Inverse-Gamma}(a, b)$$

Posterior distributions were sampled via PyMC to yield predictive distributions for 2021. Residuals and relative difference—defined as differences between predicted and observed employment figures—were extracted and later used in residual-based regional analysis.

Dimensionality Reduction and Data Projection

To make latent patterns interpretable, two complementary projection methods were applied: UMAP and Neuroscale.

Uniform Manifold Approximation and Projection (UMAP)

UMAP can effectively preserve local topological relationships by constructing graphical representation of local neighborhoods in high-dimensional space and optimizing low-dimensional embedding that maintains structures. Its objective function minimizes cross-entropy between high-dimensional and low-dimensional neighbor graphs:

$$\mathcal{L}_{\text{UMAP}} = \sum_{i \neq j} \left[p_{ij} \log \frac{p_{ij}}{q_{ij}} + (1 - p_{ij}) \log \frac{1 - p_{ij}}{1 - q_{ij}} \right]$$

UMAP was applied to the combined 2011/2021 dataset, allowing exploration of inter-year clustering and group dispersion.

Neuroscale Projection To support perceptual fidelity, Neuroscale was applied to a pairwise distance matrix derived from shortest paths over a word-association-like graph constructed from the 2011 dataset. A nonlinear mapping is defined from input space to a perceptual plane such that pairwise distances are preserved under an adaptive stress function:

$$\mathcal{L}_{\text{Neuroscale}} = \sum_{i<j} w_{ij} \left(d_{ij}^{\text{high}} - d_{ij}^{\text{low}} \right)^2$$

Unlike UMAP, Neuroscale introduces a perceptual metric through Radial Basis Function (RBF) networks, allowing flexible control over non-Euclidean dissimilarity representations, which means suited to exploring abstract structures in socio-economic datasets.

Abstraction Schema

LTLA-level records were treated as analytic units. Variables were mapped to abstract types: labour indicators as ratio-scale quantitative attributes; ethinic categorical variables retained as nominal values. Via this abstraction, mapping to various visual encodings is then feasible. (e.g., scatter plot coordinates, choropleth intensities).

3 Task Definition

Three primary tasks were designed to explore and interpret patterns in employment and ethnicity across English and Welsh local authorities in 2011 and 2021. Based on Munzner's Taxonomy the relevant details of task definition can be summarised in the table below:

Task	Intent (WHY)	Data Type (WHAT)	Encoding/Interaction (HOW)	Spatial Substrate (WHERE)
Regional employment residuals mapping	Discover	Quantitative (residuals per LTLA)	Encode (choropleth map) Filter (region selection)	Geographic space (colored channel map)
Ethnicity × employment data projection	Discover	Categorical + Quantitative (multi-dimensional)	Encode (DR-UMAP / Neuroscale) Select (highlight)+Connect (hover info)	Projected latent space (scatterplot layout)
Real vs. predicted employment comparison	Discover Present	Quantitative (prediction vs. actual values)	Encode (bivariate bar chart) Filter (by category / region)	Bars and Curve In comparison layout

Visualization Tasks according to Munzner's Taxonomy

A. Exploring regional deviations in predicted employment outcomes.

A primary objective was to identify geographic patterns in model residuals—the differences between predicted and actual employment data in 2021. According to Munzner's taxonomy, this task falls under the *Discover* intent, targeting quantitative attributes over multiple data items (local authorities). It was operationalised using geographic heatmaps, where regional residuals are encoded as colour intensity across a choropleth. This supports users in spatially navigating and filtering regions with substantial deviations, enabling further socio-economic interpretation.

B. Analysing latent socio-ethnic structures via projection space.

To explore the interplay between ethnic diversity and employment patterns, dimensionality reduction was applied using UMAP and Neuroscale, producing a 2D projection space. This task also aligns with the *Discover* intent, aiming to reveal clustering or continuity in high-dimensional data. The scatterplot encodes projected data points with employment-related attributes and ethnicity categories, supporting interactive selection and spatial exploration within the latent feature space.

C. Comparing predicted and real employment figures.

A more targeted analytical task focused on examining model accuracy by comparing actual 2021 employment values against Bayesian predictions across categories and regions. This is primarily a *Present* task involving multiple quantitative attributes. Bivariate bar charts were employed to support direct comparison and identification of over- or under-estimation trends, with sorting and filtering enabling focused analysis.

4 Visualization Justification

4.1 Predicted Employment Trend Visualisation

Based on ONS labour data from 2012 to 2019, a Bayesian linear regression model was applied to generate predictions for 2021 employment distributions, therefore visualizing impact of COVID-19 pandemic on employment structures. Three primary visualizations were employed to form comparison:

Grouped bar charts, which show predicted and actual employment values for each Government Office Region (GOR). Bar charts use position encodings to support accurate comparisons of magnitudes and differences, a principle supported by empirical findings on graphical perception [7].

Intersected heatmaps, which presents the spatial distribution of residual errors (actual minus predicted). A diverging colour scale was applied to visualise direction and intensity of prediction deviations across regions, highlight anomaly detection and regional comparison through pre-attentive visual effect.

multi-line chart, as complement, was used to visualise employment trends across regions from 2012 to 2019 whereas Line charts are suited for time-series data and help users perceive trend trajectories over time [8]. Colour was used to differentiate regions, and interactive filters enabled disaggregated analysis.

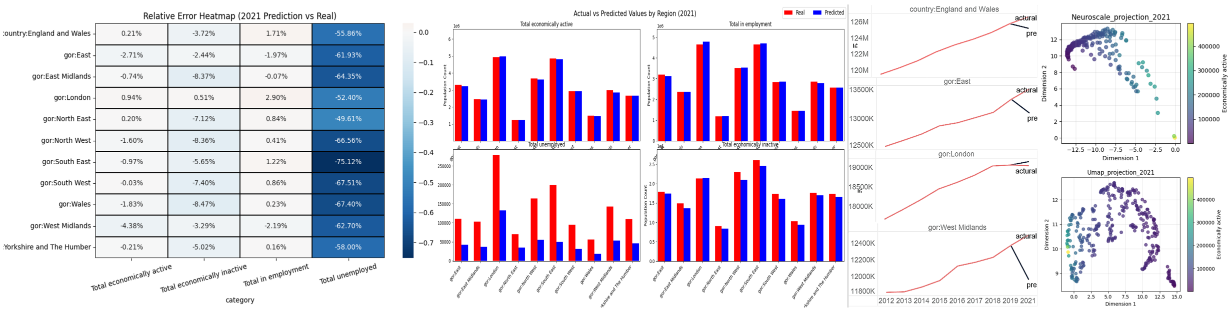


Figure 1: Sample Graph Generated by Matplotlib

4.2 Dimensionality Reduction Projections

These projections of UMAP and Neuroscale were visualised using 2D scatterplots, each point representing a GOR. As the sample graphs presents above, UMAP preserves local structure and neighbourhood relations, While Neuroscale exhibites interpretation of broader relationships and distributional patterns.

Scatterplots use spatial position to encode reduced dimensions, while colour encodes dominant ethnic or employment attributes. Tooltips provide additional attributes, implementing Shneiderman's concept [9] of "overview first, zoom and filter, then details-on-demand".

4.3 Demographic and Economic Structure Visualisation

To show population breakdowns by ethnicity and economic activity, I adopted stacked bar charts and pie charts. While Stacked bar charts are suitable for part-to-whole comparisons, enabling detailed cross-group comparisons [4], Pie charts provide an overview in summary dashboards, facilitating rapid interpretation of proportional composition for non-expert users. Their use was limited to contexts where precise comparison was not essential.

5 Evaluation

5.1 Evaluation Strategy

Drawing on Munzner's validation framework, a heuristic evaluation was performed. The evaluation focuses on three key dimensions consisted of **Encoding validity**, **Technique appropriateness** and **Task support**. In addition to this heuristic framework, further assessment is performed through comparison with external or official data sources, perspective-based usability analysis and reflections on design and implementation to identify validity and improvement potential.

5.2 Heuristic Evaluation Based on Munzner's Validation Types

Encoding validity:
Position was used for quantitative comparison in bar and line charts, colour hue for categorical variables, and diverging colour scales in heatmaps for residual error intensity. These mappings align with perceptual recommendations in the literature [7, 10].
Technique appropriateness:
Comparing visualisation types to their analytic roles, Bar charts enabled side-by-side comparison of predicted versus actual values. Choropleth maps were used to highlight regional anomalies. Dimensionality reduction scatterplots enabled the exploration of high-dimensional patterns across census data.
Task support:

- Identify regions with largest prediction error in employment.
- Detect clusters and outliers in UMAP and Neuroscale projections.
- Compare economic activity distributions by ethnicity in stacked charts.

Above Visualization tasks could be completed using the dashboard, though projection interpretation required prior knowledge of dimensionality reduction techniques.

In addition, to summarize heuristic evaluation across key components of the dashboard, the following table presents a comparative scoring of each visualization type towards core usability dimensions derived from Munzner's framework.

Graph Type	Encoding Accuracy	Task Fitness	Interpretability	Cognitive Load	Interactivity	General Rate (Up to 3)
Grouped Bar Chart	+++	++	+++	++	+++	2.7
Choropleth Heatmap	++	++	++	++	+++	2.5
UMAP Projection	++	+	-	--	++	1.5
Neuroscale Projection	+	+	+	--	++	1.3
Stacked Bar Chart	+++	++	++	+	++	2.4
Overview Pie Chart	++	++	++	-	+	1.8

Table: Heuristic Evaluation of Key Visualisation Components

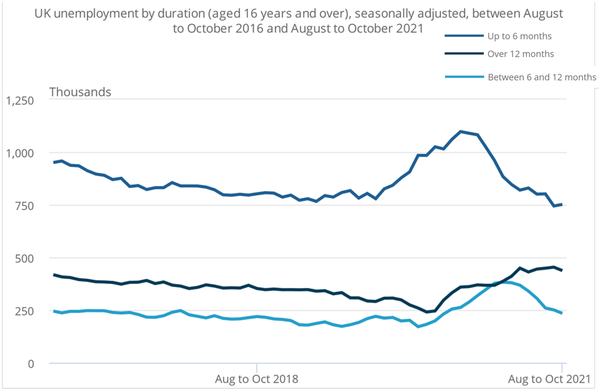
5.3 Comparison with External Visualization Research

To evaluate the reliability and realism of the dashboard outputs, selected findings were compared with external sources, including official statistics from the UK Office for National Statistics (ONS) and research reports from the House of Commons Library.

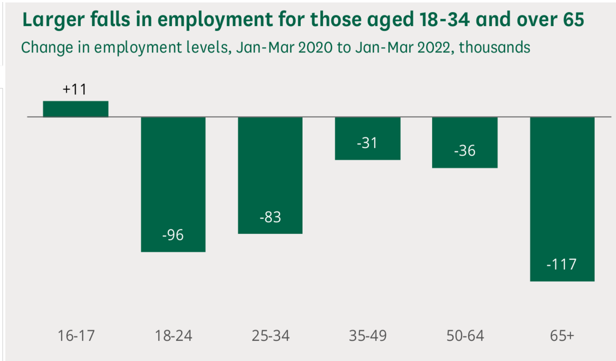
In particular, the report *Employment in the UK: November 2021* [11] provides post-pandemic labour market statistics that serve as a reference baseline. According to the

ONS, between early 2020 and mid-2021, employment rates declined while economic inactivity increased, especially in regions such as London, the North West, and the West Midlands. These patterns are consistent with the prediction difference heatmap presented in the dashboard, where the same regions exhibited the highest deviations between predicted and actual 2021 employment figures.

In addition, the House of Commons research briefing on UK unemployment [12] provides detailed regional employment statistics and time-series trends, which were used to validate the temporal analysis component of our visualisation. The general trends visualised in the dashboard (e.g., rising employment inactivity from 2019–2021) align closely with the data change illustrated in these official documents.



Source: Office for National Statistics – Labour Force Survey



Source: ONS, A05 SA: Employment, unemployment and economic inactivity by age group, 17 May 2022

5.4 Reflections and Limitations

This visual analytics project provided substantial insights into both the socio-economic topic and the design of interactive visual systems. Careful selection of visual encoding is critical to ensuring clarity and supporting intended analytical tasks. And a compatible statistical modeling integrated with visualization is helpful to enrich exploratory analysis and fosters hypothesis generation. At the same time, dimensionality reduction methods are powerful for revealing latent structures, while interpreting projections requires additional metadata or overlays to aid comprehension.

The visualization work also exists several limitations: The absence of formal user testing limits the generalisability of the evaluation findings. Some inconsistencies in geographic definitions between the 2011 and 2021 census datasets may affect comparability despite preprocessing efforts since during the 10 years some local authorities experienced changes. In addition, prediction models do not fully account for confounding policy factors or economic shocks beyond pandemic context, which means the mothod provides limited view to analysis the content.

6 Conclusion

This project has provided opportunity to examine the evolving socio-economic features of England and Wales through perspective of ethnicity and employment, with a particular focus on topic of COVID-19 pandemic impact. By combining census data from 2011 and 2021 with longitudinal labour market statistics, it was possible to analyse structural changes in economic activity across different demographic groups and geographic regions.

The visualisations revealed several salient patterns. Bayesian prediction residuals highlighted regional disparities in employment structure and development, notably in urban and economically diverse areas such as London and the West Midlands. Dimensionality reduction techniques exposed shifts in the socio-economic clustering of regions over time, and visual correlations between ethnic composition and economic inactivity were observable. These findings underscore value of longitudinal and comparative analysis in identifying latent structural inequalities and provide potential directions for policy intervention.

In terms of methodological reflection, the coursework offered substantial learning on the design and evaluation of visual analytics systems. The process of selecting appropriate visual encodings for multiple data types, implementing interactivity for exploratory tasks, and balancing interpretability with complexity highlights the importance of user-centered design.

Overall, deepening both domain-specific and methodological understanding, this project demonstrates that visual analytics is not only a means of data presentation, but also a tool for hypothesis generation, insight discovery, and informed decision-making in the context of complex social systems.

References

[1] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. CRC Press, 2013.

[2] Leland McInnes, John Healy, and James Melville. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint arXiv:1802.03426*, 2018.

[3] Christopher M. Bishop, Gareth D. James, and Nasser M. Nasrabadi. Neuroscale: Novel topographic projection technique for multivariate data. *Neural Networks*, 11(2):277–290, 1998.

[4] Tamara Munzner. *Visualization Analysis and Design*. CRC Press, Boca Raton, FL, 2014.

[5] OECD. Employment outlook 2021: Navigating the covid-19 crisis and recovery, 2021. Available at https://www.oecd.org/employment-outlook/.

[6] Office for National Statistics. Ethnic group, national identity and religion – census 2021, 2022. Available at https://www.ons.gov.uk/releases/ ethnicgroupnationalidentityandreligioncensus2021inenglandandwales.

[7] William S Cleveland and Robert McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association*, 79(387):531–554, 1984.

[8] Stephen Few. *Now You See It: Simple Visualization Techniques for Quantitative Analysis*. Analytics Press, 2009.

[9] Ben Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*, pages 336–343. IEEE, 1996.

[10] Colin Ware. *Visual Thinking for Design*. Morgan Kaufmann, 2010.

[11] Office for National Statistics. Employment in the uk: November 2021, 2021.

[12] House of Commons Library. Unemployment by constituency, 2021.