Exploring Socio-Economic Trends in England and Wales

tq24748

8 May 2025

Abstract

This project explores the socio-economic patterns of migrant populations in England and Wales using visual analytics techniques, drawing on 2011 and 2021 census datasets. Our analysis focuses on the employment status, occupational roles, qualifications, gender, age, and regional distribution of migrant groups. We used Tableau to construct interactive dashboards and implemented PCA and t-SNE in Python for dimensionality reduction. These techniques revealed latent structures in employment profiles linked to countries of origin. Bayesian linear regression was applied to model and forecast migrant economic activity trends through 2031. Key findings include the dominance of migrant employment in urban hubs, significant age and gender-based disparities, and strong associations between region of origin, qualifications, and occupational distribution. The dashboard explores key questions such as: Where do migrants work? How do occupation patterns differ by origin and arrival year? Are certain groups more economically inactive?

1 Introduction

Migration continues to shape the demographic and socio-economic landscape of the UK. In England and Wales, migrants make up a substantial portion of the labour force, contributing to both skilled and unskilled sectors. However, their employment outcomes vary based on factors such as age, gender, education, region of birth, and time of arrival. Understanding these dynamics is crucial for policy-makers, local authorities, and researchers who aim to design equitable labour policies, support integration, and plan services effectively.

This visual analytics project focuses on understanding patterns in migrant employment using census data from 2011 and 2021. Through Tableau dashboards and data projections, we aim to provide an exploratory, interactive analysis of how migrants participate in the UK labour market and how their demographic characteristics influence their employment pathways. The visualisation is primarily intended for use by policymakers, regional planners, integration programme coordinators, and researchers in migration and labour economics. It is also accessible to general users interested in exploring migrant demographics through a guided, interactive narrative.

The choice of this topic is motivated by my own experience as an immigrant and the need to understand migrant integration not just at a snapshot level, but across time and in relation to broader labour force characteristics. While previous studies have highlighted disparities in employment between UK-born and foreign-born populations, our visual analytics approach enables a more granular, multidimensional understanding of these disparities by linking demographic, temporal, and geographic attributes. The results can aid in identifying vulnerable migrant subgroups, recognising occupational segregation, and forecasting future employment trends.

We also incorporate dimensionality reduction techniques such as Principal Component Analysis (PCA) and t-distributed Stochastic Neighbor Embedding (t-SNE) to reduce high-dimensional socio-economic data into interpretable visual projections. Additionally, Bayesian modelling is used to produce credible trend forecasts.

2 Data Preparation and Abstraction

2.1 Raw Data Overview

This project integrates multiple datasets from the UK 2011 and 2021 Censuses, each providing a unique dimension of migrant-related employment and demographic data. The following datasets were used, each linked to specific visualisations in the Tableau dashboard:

- `Economic activity by country of birth by age` – used for the choropleth map and bar charts analysing employment status by age group and region of birth. Also forms the basis of the 2031 economic activity forecast.
- `Occupation by year of arrival in the UK by country of birth` – used for the treemap and stacked bar charts comparing occupation types by region of birth and arrival cohort.
- `Occupation by highest level of qualification by age` – used to analyse occupation distributions by age group and qualification level.
- `Occupation by hours worked by sex` – used in visualisations comparing working hours by gender and occupation types by sex.
- `Highest level of qualification by year of arrival by age and origin` – used to explore qualification levels by migrant subgroup and temporal cohort.
- `2031 prediction` – derived using Bayesian regression on economic activity groupings to model and visualise employment trends from 2011 and 2021 into 2031.
- `Migrant activity PCA/tSNE` – used to visualise latent clusters in migrant economic profiles based on country of birth using dimensionality reduction.

These datasets were sourced from Nomis and ONS bulk census exports. Preprocessing included unpivoting columns, splitting concatenated metadata, harmonising field names, deriving new attributes (e.g., economic activity grouping), and filtering relevant rows for migrants only.

2.2 Data Abstraction

Following Munzner's model:

- **Categorical:** Economic Activity, Country of Birth
- **Ordinal:** Age Group
- **Quantitative:** Count/Value
- **Derived Task Mappings:** Comparison across groups, filtering by demographic features, and prediction over time

These abstractions map to Munzner's "data type" categories, and informed the design of appropriate visual encodings and interactions across the dashboard.

2.3 Cleaning and Missing Data Handling

No missing values were found in the datasets used, as confirmed by exploratory diagnostics during preprocessing. Cleaning steps focused on improving attribute clarity by removing redundant text prefixes (e.g., "Age:", "Country of Birth:") and ensuring consistent formatting across fields. All datasets were transformed into tidy, long-format structures compatible with Tableau visualisations.

2.4 Outcome

The processed dataset provided a clean, structured view of migrant economic activity by age and region of birth. It enabled comparative analysis and interactive filtering within Tableau dashboards.

3 Task Definition

To ensure that the visualisations support diverse analytical needs, we defined user tasks using Munzner's taxonomy (why, what, how). The intended audience includes policymakers, local authority planners, and migration researchers who may want to explore, compare, and forecast migrant employment trends.

3.1 Why: High-Level Analytical Goals

Based on the project's objectives and user needs, we identified the following goals:

- **Discover:** Identify patterns in economic activity across age, gender, and region of birth.
- **Compare:** Examine differences between migrant subgroups—e.g., early vs. recent arrivals, males vs. females.
- **Summarise:** Aggregate and contextualise socio-economic metrics, such as employment and qualification rates.
- **Predict:** Model future trends in migrant labour force participation.

3.2 What: Data and Features

Each visualisation supports different task types through the following data abstractions:

- **Geospatial:** Region-based data for mapping employment concentrations.
- **Categorical:** Occupation, qualification, economic activity, gender.
- **Ordinal:** Age groups, qualification levels.
- **Quantitative:** Counts and percentages for comparison and forecasting.

3.3 How: Interaction and Encoding Techniques

Visual encoding and interaction design support core exploratory tasks:

- **Overview:** Dashboards provide high-level summaries of migrant employment patterns.
- **Filter:** Users can drill down by demographic categories using dropdown filters.
- **Details-on-Demand:** Tooltips reveal exact figures and categories on hover.

- **Compare:** Stacked bar charts, treemaps, and dot plots allow side-by-side subgroup analysis.
- **Pattern Recognition:** PCA and t-SNE plots reveal clusters in economic activity by origin.
- **Predict:** A Bayesian time series chart projects employment trends up to 2031.

These tasks directly informed the choice of chart types and the layout of each dashboard story point, ensuring alignment between user questions and visual outputs.

3.4 Insights Gained Through Task Analysis

Task analysis helped identify user priorities and informed dashboard design. From this, we concluded that:

- **Comparative tasks** were the most critical—users would want to contrast employment status, occupation, and qualifications between migrant subgroups.
- **Pattern discovery** through projection methods would be useful for understanding origin-based similarities.
- **Filtering** by key variables (e.g., region of birth, age group) enables meaningful subgroup analyses.
- **Predictive tasks** requires temporal alignment and clearly annotated projections.

These insights ensured that each story point addressed a distinct question aligned with user needs. Filters, labels, tooltips, and text boxes were incorporated to support Munzner's principles of multilevel exploration and user-guided interpretation.

4 Visualisation Justification

This section provides a justification of the visual encoding and interaction techniques used, grounded in Munzner's nested model and principles of information visualisation. The design decisions were informed by data types, user tasks, cognitive efficiency, and expressiveness of the chosen visual encodings.

4.1 Bar Charts, Stacked Bars, and Treemaps

Bar charts were selected for comparing values across categorical variables, such as occupation, qualification level, and economic activity. According to [1], bar charts support high accuracy for quantitative comparisons across ordered items, particularly for tasks such as "read exact value" and "compare".

Stacked bar charts were used when representing subgroup comparisons across multiple dimensions (e.g., gender within occupations, or arrival year within region of birth). While stacked bars introduce some limitations for cross-segment comparison, they are effective for showing part-to-whole relationships within categories when sorted and annotated carefully.

Treemaps were used to visualise hierarchical categorical breakdowns such as occupation types by region of origin. Their space-efficient design supports immediate overview of group dominance and can be read using area perception [2]. Labels and colour encoding ensured accessibility and interpretability.

4.2 Choropleth Map

The choropleth map was used to show spatial distribution of migrant employment by region. Geographic data is most intuitively understood using spatial encoding, which aligns with principles outlined by [3]. To support fair comparisons across regions with varying population sizes, we visualised employment percentages rather than raw counts. This aligns with best practice for regional choropleth maps and addresses guidance in the coursework brief.

4.3 Line Graph with Forecasting

For temporal comparison and prediction, a line chart was used to display economic activity in 2011 and 2021, along with a forecast for 2031. The forecast was derived using Bayesian linear regression, and 95% credible intervals were added to communicate prediction uncertainty. This approach aligns with best practices for visualising uncertainty in forecasts, as described by [4].

4.4 Dimensionality Reduction: PCA and t-SNE

To explore hidden structure in migrant economic profiles by country of birth, two projection techniques were used:

- **Principal Component Analysis (PCA)**: A linear projection method that reveals global structure and direction of variance in high-dimensional data.
- **t-distributed Stochastic Neighbor Embedding (t-SNE)**: A non-linear technique that preserves local neighbourhoods and reveals clusters more clearly.

Both methods were applied to vectors summarising economic activity proportions (e.g., employed, unemployed, inactive) across age groups for each country of birth. This enabled a task of "pattern discovery" and "cluster detection". PCA revealed overlapping global structure, while t-SNE separated region-based groupings more clearly. Tooltips in Tableau identify the country represented by each point, supporting interpretation.

4.5 Interaction and Story Design

Following Munzner's "how" level, interactivity was designed to support exploration and comparison. Filters for age, gender, origin, qualification, and year of arrival allowed users to zoom and filter based on their analytical needs. Text boxes at each story point answered a specific question, aligning with "overview first, zoom and filter, details on demand"—a classic mantra in visual analytics [5].

Together, these visualisation choices ensure cognitive clarity, high expressiveness for the task, and alignment with visual perception theory and the goals of exploratory analysis.

5 Evaluation

Following Chapter 4 of Munzner's framework, we conducted a qualitative evaluation of the visualisation system through observation and task-based assessment with three students from the assigned discussion group. Each participant was asked to complete three analytic tasks using the Tableau story:

1. Identify which region has the highest proportion of employed migrants.
2. Compare occupation types between early and recent arrivals from the Middle East and Asia.

3. Interpret the 2031 employment forecast for migrants.

5.1 Evaluation Methodology

The evaluation was conducted using a lightweight task-based protocol. Participants were given a short description of the dashboard and a set of structured questions. Their task completion time, observed navigation behaviour, and verbal feedback were recorded.

In line with Munzner's nested model, the evaluation spanned three levels:

- *Domain-level:* Users were asked to interpret socio-economic patterns related to migrant integration.
- *Operation-level:* Tasks involved filtering by region, interpreting bar heights, reading tooltips, and switching views. This tested the usability and cognitive fit of the visual encodings.
- *Encoding-level:* Participants interacted with visual features such as bar length, area size (treemaps), spatial position (choropleth), and projection clusters (t-SNE).

5.2 Key Findings

- Participants successfully completed all tasks within 3–5 minutes and found the map and treemap views intuitive.
- One user misinterpreted stacked bars without hovering for tooltip values, reinforcing the importance of explicit labelling—this issue falls within Munzner's encoding-level and interaction usability.
- All users noted that the t-SNE projection was harder to interpret without prior knowledge, pointing to challenges at the operation level.
- The use of text boxes under each story point was positively received and helped clarify interpretation at the domain level.
- One user suggested adding tooltips to show country details when hovering over regions. This was implemented to improve encoding-level clarity for comparisons.
- Reviewers suggested replacing the stacked bar chart, but it was retained due to its strength in part-to-whole comparison. This decision aligns with encoding-level expressiveness principles in Munzner's framework.
- A suggestion to use pie charts instead of bar charts was made, but not adopted, as bar charts offer higher accuracy for comparing magnitudes [2].
- Another improvement implemented was adding economic activity proportions as a tooltip-based pie chart in the choropleth map, supporting "details on demand" [5].

This evaluation confirmed that the dashboard effectively supported analytical tasks across all levels of Munzner's model and enabled users to derive meaningful insights from complex, multi-dimensional data. Minor refinements based on feedback enhanced clarity, encoding accuracy, and overall usability.

6 Conclusion

6.1 Socio-Economic Insights

The visual analytics process revealed several key insights about migrant employment and demographic trends in England and Wales. First, migrant populations are concentrated in urban areas such as London and Birmingham, with employment levels and inactivity varying significantly by region of birth. Younger migrants tend to be more economically active, while those over 50 are more likely to be inactive. Gender disparities are clear, with women more likely to work part-time and in care or administrative roles, and men dominating full-time and skilled trade occupations.

Additionally, the time of arrival significantly influences occupational integration. Earlier arrivals are better represented in professional and managerial roles, while recent migrants are more concentrated in elementary jobs. This reinforces the role of social integration time and credential recognition in shaping migrant labour trajectories. The PCA and t-SNE projections further illustrated that economic profiles often cluster by country of origin, suggesting cultural, educational, or systemic factors affecting employment pathways. Bayesian modelling showed an expected increase in both employment and unemployment by 2031, indicating a growing labour force but also increased job-seeking pressures.

6.2 Lessons in Information Visualisation

From a visualisation perspective, the coursework reinforced the importance of grounding design choices in established theory. Applying Munzner's nested model ensured that every visual encoding was supported by appropriate data abstraction and user task goals. The use of Shneiderman's mantra (overview first, zoom and filter, details on demand) provided a structure for exploration, while encoding principles from Ware and Cleveland guided chart selection for accuracy and usability.

Working with high-dimensional data for PCA and t-SNE illustrated the importance of clearly communicating what variables were used in projection and annotating outputs effectively. Forecasting with Bayesian methods challenged us to present uncertainty meaningfully using confidence bands and clear annotation.

Most importantly, this project demonstrated that thoughtful integration of design, data, and domain context leads to dashboards that are not only visually effective but also impactful in delivering socio-economic understanding to a non-technical audience.

References

[1] T. Munzner, *Visualization Analysis and Design.* Boca Raton, FL: CRC Press, 2014.

[2] W. S. Cleveland and R. McGill, "Graphical perception: Theory, experimentation, and application to the development of graphical methods," *Journal of the American Statistical Association*, vol. 79, no. 387, pp. 531–554, 1984.

[3] C. Ware, *Information Visualization: Perception for Design.* Burlington, MA: Morgan Kaufmann, 2012.

[4] J. Hullman, P. Resnick, and E. Adar, "Why authors don't visualize uncertainty," *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, no. 6, pp. 2397–2406, 2015.

[5] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," *Proceedings 1996 IEEE Symposium on Visual Languages*, pp. 336–343, 1996.