EMATM0066 - Visual Analytics Coursework

Exploring Youth Unemployment in England and Wales: Insights into Gender Disparities, Industry Gaps, and Regional Variations

***Abstract*—This study explores youth unemployment in England and Wales, focusing on regional, gender and industry-level differences using data from the 2011 and 2021 censuses. It also projects unemployment counts by gender for 2030 using Bayesian Imputation and Bayesian Ridge Regression. The findings highlight persistent gender disparities in the labor force, with men more likely to be unemployed, but women more likely to be economically inactive. Regional progress has also been uneven, while some cities show improvement, districts like Rutland and the City of London remain largely unchanged. The analysis also reveals distinct patterns in youth employment rates across industries, showing which sectors consistently offer more opportunities and which lag behind.**

To support this, PCA and UMAP were used to uncover broad and local patterns in youth labor trends. All visualisations were created in Tableau, making the data accessible and easy to interact with. The dashboards were designed to be clear, interpretable, and interactive. They were intended to help end users explore patterns with ease and support more inclusive, data-driven, and future-oriented employment strategies.

Keywords: Bayesian Ridge Regression, Bayesian Imputation, Data Visualisation, Tableau, Dashboards, PCA, UMAP.

I. Introduction

As a student, I have seen first-hand how hard it can be for young people to find stable jobs. Even with the right skills, it is still tough to navigate the job market, and the uncertainty often feels overwhelming. Unfortunately, this struggle is not something that I experience alone. It is a challenge many young people face globally, and England and Wales are no exception.

As of early 2025, the UK's overall unemployment rate stands at 4.5%, while the youth unemployment rate for individuals aged 16–24 has risen to 13.3%, highlighting a persistent disparity between young people and the broader workforce. These figures reflect the scale of the problem, which is caused by a mix of gender disparities, geographic barriers, and economic slowdown.

One of the most immediate effects is the increase in student debt. Without access to stable employment, many young people struggle with loan repayments, leading to financial stress and long-term debt accumulation. The consequences of youth unemployment extend far beyond individual finances.

Socially, unemployment can lead to poor mental and physical health and increased dependence on welfare. It also contributes to a sense of disconnection and reduced self-worth among young people. These issues are particularly challenging for young women, who often face higher unemployment and less opportunities due to gendered labor roles, workplace discrimination, and male-dominated fields. Geography plays a role as well, youth unemployment tends to be higher in rural areas where job opportunities, access to education, and transport infrastructure are limited compared to urban centers.

Financially, the impact is seen both at the individual and systemic levels. Young people burdened with debt are less likely to spend, invest, or buy homes, which in turn reduces overall consumer demand and slows economic growth. Unpaid loans place additional pressure on banks through the accumulation of bad debt and tighter lending practices, which can destabilize the financial system. The strain also extends to public finances, as rising unemployment increases the demand for government support such as unemployment benefits and housing aid. This diverts resources from essential services such as education, infrastructure, and healthcare.

In the longer term, or what could be seen as a more futuristic concern, youth unemployment affects the next generation. Unemployment early in year can limit future earning potential, reduce access to opportunities, and widen existing social inequalities. The effects may also impact the next generation, restricting access to education and opportunities, therefore continuing the poverty cycle and social exclusion.

This assignment focuses on youth unemployment in England and Wales, drawing on data from the 2011 and 2021 censuses to explore how employment trends have changed over time. The past decade has seen major disruptions to the labor market, including the 2008 financial crisis, the COVID-19 pandemic, and a sharp rise in the cost of living, influenced by global events such as the war in Ukraine.

This report is aimed at policymakers, economic analysts, educators, industry leaders, and advocacy groups. It highlights the gaps young people face in the job market between regions, sectors, and gender. Understanding these gaps can help inform targeted policies, improve support through education and career services, and guide more inclusive hiring practices. The findings are intended to support action toward a more equitable and accessible labor market for future generations.

II. Problem Statement

This report looks at the trends of the 2011 and 2021 censuses, focusing on youth unemployment by region (Local Authority Districts), gender. In addition, it explores employment patterns across various industries to identify sectors where opportunities for individuals aged 16-24 years can be expanded. The aim is to identify where the gaps are and understand why they exist.

III. Data Preparation and Abstraction

For this study, data tables were extracted from the 2011 and 2021 censuses to analyze trends in youth unemployment in England and Wales. Only data for individuals aged 16–24 were used, as they are defined as youth and represent early career starters. The following is a breakdown of the specific tables extracted from both years.

A. Data Collection: 2011 Census Tables

1) DC6110EW - Economic Activity by Industry:

Use Case: This table shows youth employment across industries and regions, helping to identify which sectors employ more young people.

Reason for Inclusion: This data helps identify industries with more opportunities and highlight sectors with higher/ lower unemployment risks.

2) DC6201EW - Economic Activity by Gender:

Use Case: This table breaks down the data on employment, unemployment, and inactivity by gender for the youth population in 2011.

Reason for Inclusion: It is essential to identify regional gender disparities in youth unemployment rates.

3) DC5203EW - Highest Level of Qualification by Age:

Use Case: This table provides data on the highest level of qualification attained by individuals in the 16-24 age group, including details on qualifications such as Level 1, Level 2, Level 4, or no qualifications. It also includes individuals who chose the apprenticeship route instead of traditional qualifications.

Reason for Inclusion: By analyzing this table, we can gain insight into the educational attainment levels of young people, which is a key factor influencing their employability.

B. Data Collection: 2021 Census Tables

1) RM024 - Economic Activity by Gender:

Use Case: This table provides a gender-based breakdown of employment, unemployment, and inactivity rates for 2021. It is similar to the 2011 dataset DC6201EW.

Reason for Inclusion: It helps assess whether gender-based inequalities have improved, worsened, or remained consistent between regions over the last decade.

2) RM064 - Economic Activity by Industry:

Use Case: This table provides data on the distribution of employment in various industries for the youth population (ages 16-24) in 2021.

Reason for Inclusion: Similar to DC6110EW (2011), this table allows for a comparison that highlights how the labor market has changed due to economic changes, such as the rise of AI-driven industries and global disruptions like the COVID-19 pandemic.

C. Data Cleaning & Preprocessing

The common data cleaning steps implemented throughout the data tables involved several key actions to ensure consistency and accuracy. First, whitespace removal was applied to the 'Region' column using .str.strip(), eliminating any leading or trailing spaces that could cause inconsistencies. Capitalization standardization was also performed with .str.title(), ensuring uniform naming across region entries (e.g., "northumberland" was changed to "Northumberland"). Duplicate region entries were checked as they could have skewed the analysis, but none were found. Each column in the dataframe was carefully reviewed to ensure that the data types were correct. This was important because having the right data type ensures that calculations and analysis are performed correctly. For instance, columns like `'Employed: Total'` and `'Unemployed: Total'` were checked to confirm that they were set as `int64`, since they represent counts of individuals, it needs to be treated as whole numbers. On the other hand, columns like `'Region'`, which contains text, were appropriately classified as `object` type to handle string data.

Missing values were identified in the DC6201EW (2011), specifically in columns such as `"Employed: Total"`, `"Unemployed: Total"` and `"Inactive: Total"`. To address the missing values, Bayesian imputation was used through the IterativeImputer from the sklearn library with a Bayesian Ridge estimator. It uses a probabilistic approach to predict missing values based on patterns in the data, rather than relying on simple averages.

For example, when data for `"Employed: Total"` is missing in a region, the imputer does not only consider the missing column, but also analyzes the entire row, including related columns like `"Unemployed: Total"` and `"Inactive: Total"`. It also considers how these columns typically correlate across different rows (regions).

Once imputed, values were rounded to integers, as appropriate for employment data, and validated against overall trends.

D. Data Integration

Integrating data from the 2011 and 2021 census datasets required careful handling of regions (Local Authority Districts), as some were merged or renamed by 2021. For example, multiple districts were combined into a single region such as Dorset. To allow accurate comparisons across both years, a mapping process was used to align 2011 region names with their 2021 counterparts.

In the initial attempt, the datasets were merged using the `'Region'` column, but this left many rows empty due to unmatched region names. These missing values `(NaNs)` created problems during Bayesian imputation, which does not perform well with rows that are completely or mostly empty. As a result, many rows were dropped, leading to a significant loss of data.

To resolve this, a mapping method was applied to align the 2011 region names with their updated 2021 region names. This mapping ensured that merged or renamed areas were correctly handled, allowing for a consistent comparison between the two datasets. The data were then grouped by the updated 2021 region names, and the employment figures were summed within each group, preserving important statistics without losing data.

III. Task Definition

The primary goal is to explore and provide insights on youth unemployment in England and Wales, focusing on 2011 and 2021, with predictions for 2030 based on Bayesian imputation. Using Munzner's task taxonomy, the tasks for this analysis can be broken down into multiple key steps to address specific aspects of the socioeconomic problem and ensure the user gains meaningful insights from the visualisations.

Task 1: Understanding Regional Disparities in Youth Unemployment

Objective: This task aims to explore how youth unemployment rates vary between different regions of England and Wales. It involves grouping the data by `Region` and comparing the unemployment statistics for the 16-24 age group. By identifying the regions with the highest and lowest unemployment rates, the goal is to uncover regional disparities in youth employment.

Target Users: Policymakers, local authorities, social organizations, and youth employment agencies.

Purpose: This task aims to understand regional imbalances in youth unemployment, helping to pinpoint areas where targeted resources and infrastructure development are most needed to reduce unemployment and create more opportunities for young people.

Task 2: Analyzing Gender Disparities in Youth Unemployment

Objective: This task explores gender differences in youth unemployment by comparing unemployment rates for men and women aged 16–24. It focuses on identifying regions with significant gender gaps and understanding the factors behind these disparities.

Target Users: Gender equality advocates, policymakers, and social organizations for gender equality focused on addressing gender-based disparities in employment.

Purpose: To assess the impact of gender on youth unemployment, identifying areas where gender-targeted policies may be needed.

Task 3: Understanding Industry-wise Employment Trends

Objective: This task focuses on analyzing the employment of youth by industry, comparing employment rates in different sectors such as manufacturing, healthcare, and finance. The goal is to understand which industries provide more employment opportunities for young people and which sectors contribute more to youth unemployment.

Target Users: Industry stakeholders, workforce development agencies, educators, and career advisors.

Purpose: To highlight industry-wise employment patterns, helping stakeholders identify which industries are linked to higher youth unemployment (i.e., lower employment rate) and where there is potential for growth.

Task 4: Visualizing Temporal Changes in Youth Unemployment (2011-2021)

Objective: This task involves comparing youth unemployment rates from 2011 to 2021 to understand how they have changed in the past decade. The analysis will focus on changes in unemployment rates between regions and industries, taking into account the effects of economic slowdowns such as the COVID-19 pandemic.

Target Users: Economists and policy makers interested in tracking changes in youth unemployment over time.

Purpose: To provide a perspective on youth unemployment, assess the effectiveness of policies, and understand how external factors, such as economic recessions or global pandemics, have affected youth employment trends.

Task 5: Predicting and Visualizing Youth Unemployment Trends for 2030

Objective: This task involves comparing youth unemployment rates from 2011 to 2021 to understand how they have changed in the past decade. The analysis will focus on the shifts in unemployment rates between regions, genders, and industries taking into account the effects of economic slowdowns such as the COVID-19 pandemic.

Target Users: Economists and policy makers interested in tracking changes in youth unemployment over time.

Purpose: Provide insights into trends in youth unemployment, helping policy makers and industry leaders plan for the future.

IV. Visualisation Techniques and Justification

1) Dashboard: Educational Attainment by Youth Population in England and Wales (2011)

The dashboard is designed to analyze the relationship between educational attainment and employment rates among the youth population (ages 16-24) across regions in England and Wales.

A. Spatial Distribution Encoding Using Choropleth Mapping

Focus: Data Type & Visual Encoding (Geospatial data, Choropleth map)

The Choropleth map serves as the primary visualisation, displaying the distribution of the youth population in various regions. The regions are color-coded in shades of brown, where darker shades represent regions with larger populations. This map, as recommended by Tamara Munzner in Visualisation Analysis and Design (2014), allows users to intuitively perceive geographic patterns and disparities in the distribution. The choice of using a choropleth map aligns with human perceptual abilities, as it helps to visually identify high and low density regions quickly.

B. Interaction Design for Contextual Demographic Insight

Focus: Interaction Techniques (Details-on-demand)

The interactive feature of the map enhances the experience, when users hover over any given region, a tooltip appears, displaying details such as the `Total Youth Population` and the `Employment Rate` for that region. This interaction provides immediate insights into the demographic and employment data while enhancing user engagement.

C. Visual Correlation between Educational Attainment and Employment Rates

Focus: Analytical Task (Correlation)

This supports the hypothesis that higher educational attainment, particularly at Level 4 (higher education), correlates with higher employment rates. The visual design reflects this relationship by showcasing regions with more highly educated youth and higher employment rates in a more prominent, visually distinct manner.

The bar chart was chosen because it is well suited to compare categorical data across regions, making it ideal for illustrating patterns in educational attainment. Its linear structure allows users to easily compare the proportions of each qualification level side by side. To enhance this, distinct and contrasting colors are used for each category: `Apprenticeship`, `Level 1/2/3 qualifications`, `Level 4 qualifications`, and `No qualifications` making it easy to differentiate between them at a glance. This combination of chart type and color encoding supports quick visual comparisons, highlights the dominant education level in each region, and enables exploration of how qualifications are distributed geographically.

D. Integrated Dashboard Design for Insightful Analysis

Focus: Multi-View Dashboard for Comparative Regional Analysis

By combining the use of a choropleth map with a dynamic, filtering bar chart, this dashboard enhances the user's ability to explore and understand the relationship between educational attainment and employment rates. The dashboard follows strong information design principles, making the data easy to understand and visually engaging. By combining clear visuals with interactive elements, users can quickly see how different levels of youth education relate to employment rates across regions.

Overall, the goal was to create a practical, user-friendly dashboard that supports analysis and decision-making by presenting complex data in an interpretable format.

2) Dashboard: Youth Unemployment and Workforce Distribution in Major UK Cities (2011 vs 2021)

This dashboard is designed to analyze regional changes in youth unemployment rates and the evolving composition of youth employment in major cities in England and Wales between 2011 and 2021. Focusing on individuals aged 16–24, it offers a comparative view of unemployment trends over time along with industry-specific employment distributions. By integrating unemployment trends with industry-wise employment insights, the dashboard enables a comprehensive exploration of youth labor market patterns, highlighting both regional disparities and shifts in employment sectors.

The analysis focuses on major cities to improve visual clarity by reducing overplotting and making the charts easier to interpret. This scope enhances the analytical focus on urban labor markets, where youth employment challenges are more concentrated and data is more robust. By focusing on key metropolitan areas, the dashboard remains clean and ensures easy comparisons without being overwhelming.

A. Temporal Trend Comparison of Youth Unemployment Rates

Focus: Data Type & Visual Encoding (Line Chart)

The line chart includes an interactive tooltip that is activated when hovering, displaying key details such as `Region`, `Year`, and `Unemployment Rate`. This interaction enhances user engagement by providing specific information without cluttering the visual. Additionally, the use of two distinct colors, blue for 2011 and orange for 2021. Makes it easier to distinguish between the two censuses, helping users quickly compare trends across regions.

B. Distribution of Youth Employment Across Industries

Focus: Analytical Task (Comparison) & Data Encoding (Categorical Bar Chart)

Beneath the unemployment trend chart, two grouped bar charts show the employment rate by industry for 2011 and 2021. These charts highlight how youth employment is distributed across industries like `Agriculture`, `Construction`, `Distribution & Hospitality`, `Finance`, and others. The bar chart is ideal for comparing employment levels across industries as it clearly displays categorical data. It makes it easy to track changes and compare industries. The use of distinct colors for each industry helps users quickly differentiate between them. This helps users effectively analyze how youth employment has shifted across sectors and regions.

C. Integrated Dashboard Design for Insightful Comparative Analysis

Focus: Multi-View Dashboard & Comparative Analysis

This dashboard adopts a multi-view approach by integrating a line chart showing unemployment rates for the years 2011 and 2021, alongside grouped bar charts displaying industry-wise employment rates for the same years. When users hover over the line chart, a tooltip appears with key details, and the data in the grouped bar charts below are filtered accordingly. This layout allows users to explore regional unemployment data, as well as employment rates by industry, across two time points.

By combining comparative, categorical, and regional data, the dashboard offers a comprehensive view of how youth labor dynamics have evolved between 2011 and 2021. The integration of different data types with appropriate visual encoding, along with interactivity, enhances the interpretability of the dashboard without compromising clarity.

The goal was to create a user-friendly dashboard that offers clear, insightful analysis of youth employment patterns, presenting complex data in a way that supports both detailed exploration and broader comparisons. Beyond visual analysis, the dashboard helps identify areas with high youth unemployment and sectors lacking entry-level opportunities, guiding targeted re-skilling efforts. For policymakers and educators, the data can inform regional employment strategies, while career advisors can use the insights to align young people with growing sectors and future-proof career paths.

3) Dashboard: Youth Labor Participation and Unemployment in the UK (2011)

This dashboard visualizes youth labor participation and unemployment trends in the UK for 2011, also focusing on gender disparities. It integrates two primary elements: a choropleth map and a bar chart. The map displays regional unemployment rates, as well as including the number of economically active youth and the total youth population in each region. The bar chart complements this by providing a gender-specific breakdown of unemployment rates across regions.

Together, these visualisations offer a comprehensive view of youth unemployment patterns across the UK, highlighting both geographic and gender-based differences while offering insight into the size and engagement of the youth labor force in each area.

A. Spatial Distribution Encoding Using Choropleth Mapping

Focus: Data Type & Visual Encoding (Geospatial data, Choropleth map)

It helps visualize the distribution of youth unemployment across regions. The regions are color-coded based on the population of economically active youth, where darker shades represent a higher concentration and vice versa. This method, as discussed by Munzner in Visualisation Analysis and Design (2014), allows for easy identification of regional patterns and disparities. It aligns with human perceptual abilities, making it easy for users to spot areas with high or low concentration of active youth at a glance.

B. Interaction Design for Contextual Demographic Insight

Focus: Interaction Techniques (Details-on-demand)

An interactive feature is implemented within the map, when users hover over a specific region, all other regions dim down, while the region being hovered over is highlighted. This dynamic interaction increases clarity, allowing users to focus on the specific region of interest. In addition, a tooltip appears on hover, displaying key information such as `Active Youth in the Workforce`, the `Unemployment Rate`, and the `Total Youth Population` for that region. These details provide context and offer valuable insights that are crucial for future analysis. This interaction follows the details-on-demand principle, enhancing user engagement by offering focused, region-specific data while maintaining an uncluttered visual layout.

C. Visual Correlation Between Regional Youth Unemployment Rate

Focus: Analytical Task (Comparison & Correlation)

The bar chart shows the gender breakdown of the youth unemployment rates for each region. As mentioned previously in this report, this chart is the well-suited format for comparing categorical data, particularly when dealing with distinct groups such as male and female unemployment rates. Its structure and clear visual separation between categories make it easy to interpret and ideal for side-by-side comparison. A color scheme is used to represent female (pink) and male (blue) data respectively, helping users visually distinguish between the two. This format supports quick identification of gender-based disparities within each region and enhances overall clarity.

D. Integrated Dashboard Design for Insightful Analysis

Focus: Multi-View Dashboard for Comparative Regional Analysis

By combining the choropleth map with the bar chart, the dashboard allows users to explore regional youth unemployment alongside gender-based insights. This coordinated layout supports direct comparisons across regions and between genders, helping users understand key patterns quickly and clearly.

E. Enhancing Readability and Interpretability

Focus: Interactive & Dynamic Design

When users hover over a region, the other regions dim and only the selected region is highlighted, improving the clarity of the data. This filtering mechanism allows users to focus on specific areas and see the corresponding unemployment data on the bar chart. The combination of these interactive elements helps users to more effectively analyze the relationship between regional youth unemployment trends and gender disparities, without overwhelming them with excessive data.

Overall, this dashboard supports exploratory analysis with a user-friendly interface, dynamic filtering, and comparative insights into youth unemployment across UK regions. The interactive highlighting of the hovered region, along with the bar chart filtering, enhances clarity and focus.

4) Dashboard: Youth Labor Participation and Unemployment in the UK (2021)

This dashboard presents a 2021 equivalent to the previous dashboard (Youth Labor Participation and Unemployment in the UK (2011), showcasing youth participation, including economically active youth, unemployment rates, and a gender split of unemployment rates. It can be compared with the 2011 data to analyze regional changes over the past decade. The comparison highlights changes in youth unemployment trends, gender disparities, and regional variations, providing insights on the evolving youth labor market.

This can help policymakers, educators, and industry leaders assess changes and identify patterns to inform future youth employment strategies.

5) Predicted Youth Unemployment Rates by Region and Gender (2030)

This dashboard is similar to the previous two (Youth Labor Participation and Unemployment in the UK (2011 & 2021)), but instead of displaying historical data, it shows predicted youth unemployment rates for 2030. It includes a gender split, providing insight into gender disparities in youth unemployment. It can be compared with the 2011 and 2021 data to track trends and predict future regional disparities in youth unemployment. These projections help inform policy makers and guide proactive measures to address potential unemployment issues.

The predictions are generated using Bayesian imputation, which draws on trends from the 2011 and 2021 census data to forecast future youth unemployment patterns across regions and genders.

6) Dashboard: Youth Economic Activity Projection - PCA & UMAP Clustering

This dashboard compares two-dimensionality reduction techniques: PCA and UMAP to analyze gender-based youth economic activity across UK regions. These methods simplify complex data while preserving key patterns in employment, unemployment, and inactivity, enabling clearer comparisons of youth labor engagement by region and gender in 2011 and 2021.

A. Dimensionality Reduction

Focus: Dimensionality Reduction Methods (Linear vs. Non-linear Techniques)

PCA is a linear technique that projects the data into a lower dimensional space by finding the directions (i.e., principal components) that maximize the variance. PCA is ideal when we want to understand the main directions of variability in the data, which is particularly useful for detecting broad patterns across regions. In this case, PCA is used to explore general labor engagement trends, capturing patterns in employment intensity and levels of participation in the labor market.

Whereas, UMAP is a non-linear dimensionality reduction technique that is more flexible in capturing complex, non-linear relationships between data points. It is particularly useful for preserving local structures, such as regional similarities and differences in youth unemployment patterns, which might not be fully captured by PCA. UMAP provides a clearer, more intuitive view of how regions relate to each other based on economic activity and gender.

In this study, both PCA and UMAP have been implemented to observe how each method represents the data differently.

B. Tableau Visualisation:

Focus: Visual Encoding & Task Support (2D Scatterplot for Cluster Detection)

This dashboard uses scatterplots to visualize the results of the two techniques: PCA for 2011 and UMAP for 2021.

In both cases, a scatterplot is the most appropriate visualisation technique. As noted by Munzner (2014), scatterplots are ideal for encoding quantitative multivariate data using 2D position. It is the most effective visual plot to support tasks such as pattern recognition, cluster detection, and similarity comparison. Each point represents a region, and the spatial arrangement reveals clusters of regions with similar economic characteristics.

Observation: In 2011, PCA classified Bristol as a district/region with strong overall labor participation (PC1 = 1.11) but lower employment intensity (PC2 = -2.57), indicating high participation alongside significant unemployment. By 2021, UMAP placed Bristol within a moderately distinct cluster (UMAP1 = 5.84, UMAP2 = 0.62), reflecting greater local variation and potential gender disparities in youth economic activity. This comparison shows how PCA captures broad trends, while UMAP reveals more nuanced, localized shifts in youth employment patterns.

TABLE I
PCA CLUSTERS BY EMPLOYMENT CHARACTERISTICS

Cluster	Type	Cluster Description
0	Low Activity	Low employment and unemployment; smaller economically active population
1	Mid-Range	Moderate employment and inactivity
2	Outlier	Exceptionally high employment, unemployment, and inactivity (urban centers)

TABLE II
UMAP CLUSTERS BY EMPLOYMENT CHARACTERISTICS

Cluster	Type	Cluster Description
0	Outlier	Highest employment, unemployment, and inactivity (economically dense)
1	Low Activity	Lowest employment and population metrics
2	Mid-Range	Moderate employment and activity

C. Enhancing Readability and Interpretability

Focus: Visual Cues & Interactive Features (Color, Size, Tooltip)

The scatterplot uses color to differentiate clusters of regions and visually represent gender-specific economic activities, helping users identify patterns across the data at a glance. Different marker sizes and shapes provide additional context by reflecting each regions unemployment rate, making it easier to compare. To support exploration, the dashboard includes an interactive tooltip that appears when hovering over a region. This displays key information such as the `Region`, `Cluster`, and Unemployment Rate without overcrowding the chart. Together, these visual and interactive elements make the data more accessible, allowing users to explore specific regions while maintaining a clear overview of broader trends.

Overall, this dashboard enables users to explore and compare youth economic activity across UK regions using PCA and UMAP. The interactive scatterplots help users identify both broad trends and gender-based disparities, making it easier to understand regional differences in employment patterns.

V. PREDICTED YOUTH UNEMPLOYMENT COUNTS FOR 2030 (BY GENDER): BAYESIAN IMPUTATION AND BAYESIAN RIDGE REGRESSION

This approach draws on historical data from the 2011 to 2021 censuses to predict future youth unemployment patterns, focusing on both male and female youth unemployment for each region. Bayesian imputation was implemented as it allows us to incorporate prior data, the observed unemployment data from previous years, and update it as new information becomes available. This method enables the prediction of future unemployment rates while quantifying uncertainty, which is essential when forecasting future trends.

Mathematical Formula:

$$\beta = \arg\min_{\beta} \left(\|Y - X\beta\|^2 + \lambda \sum_i \beta_i^2 \right)$$

Where:

- Y represents the target variable.
- X is the matrix of input features
- β represents the regression coefficients, which are the parameters being estimated.
- λ is the regularization term to avoid overfitting.

In the context of this model, Y represents `Unemployed_2030` (i.e., total predicted unemployed youth), `Unemployed: Male_2030` (i.e., predicted unemployed males), and `Unemployed: Female_2030` (i.e., predicted unemployed females).

X consists of log-transformed unemployment data from 2011 and 2021, specifically the unemployment counts for males, females, and the total youth population. The log transformation is applied to normalize skewed distributions, ensuring the data is more linear and suitable for regression modeling. This improves the models ability to detect relationships in the data, which are critical for accurate future predictions. After prediction, the log-transformed results are inverse-transformed to return them to their original scale, making them interpretable as actual unemployment counts.

VI. Evaluation of Visual Design and User Interpretation

As part of the evaluation process, I asked a group of peers in class to review my dashboards and provide feedback based on five targeted questions focusing on clarity, understanding, and interactivity. The majority of responses indicated that the dashboards were not cluttered and that the layout made it easy to navigate and follow the visual flow. They agreed that the charts used throughout the dashboards clearly communicated the key patterns and insights and that they were able to understand and connect the different visual elements to carry out analytical analysis.

In particular, they could easily derive trends in youth unemployment across regions, genders, and industries. What stood out most to them was the line chart showing unemployment over time, especially when it was linked to industry-wise employment rates helping them identify which sectors offered the most opportunity for young people and where improvements could be made. The interactive tools like tooltips, dynamic filtering, and highlighting were also well-received for being intuitive and informative, allowing users to focus on specific areas of interest without being overwhelmed.

Overall, the feedback confirmed that the dashboards successfully supported exploratory analysis and made complex patterns in the data visually engaging and interpretable.

VII. Conclusion

A. Data-Driven Reflections on Regional, Gender, and Industry Gaps

This study provided a detailed exploration of youth unemployment in England and Wales using census data from 2011 and 2021, uncovering persistent disparities by region, industry, and gender.

What gender-based patterns in youth labor force participation and unemployment became evident through the visualisation? One of the most critical findings is the distinction between male and female labor market engagement. While men show higher unemployment numbers, this is largely due to a greater proportion of them being economically active. In contrast, young women exhibit higher inactivity rates, suggesting that many do not participate in the labor force at all. This may be due multiple factors, including ongoing illness or disability, caregiving responsibilities, cultural expectations, limited access to information or opportunities, or simply a lack of exposure to career pathways early on. This dual challenge calls for more inclusive employment strategies: creating better pathways for re-engaging inactive women, while also strengthening support for men already in the workforce but struggling to secure stable employment.

Which regions showed the greatest improvement in youth unemployment, and how should this progress be interpreted in relation to their population size? According to the data, major cities such as Birmingham, Leeds, Manchester, Sheffield, and Liverpool have shown the largest reductions in youth unemployment numbers between 2011 and 2021. For example, Birmingham saw a decrease of over 13,000 unemployed youth during this period. However, this improvement should be questioned in the context of Birmingham's overall youth population, which remains among the largest in England and Wales. Despite the reduction, the city's employment challenges remain persistent, and relative to its size, the progress may not be sufficient to indicate meaningful structural change.

What did the data reveal about districts with little or no improvement, and what underlying challenges might explain these trends? In contrast, districts like City of London and Rutland have seen either minimal improvement or no change in youth unemployment levels. These stagnant figures reflect ongoing challenges that may differ in nature but lead to similar outcomes such as overcrowding, high competition, and cost of living pressures in places like City of London, and limited local opportunities or industry diversity in smaller areas like Rutland. Despite their contrasting sizes and profiles, both districts demonstrate how economic growth alone does not always translate into equitable access to employment for young people.

How does higher educational attainment among youth relate to employment outcomes, and what factors may disrupt this relationship? Education continues to play a crucial role in shaping labor outcomes. The 2011 data shows a moderate negative correlation (r = -0.42) between Level

4 educational attainment and youth employment rates. This suggests that even in regions with a higher number of highly educated young people, employment rates were not necessarily higher. Factors such as skills mismatch, limited local job opportunities, or lack of access to appropriate career pathways may contribute to this disconnect.

Which industries continue to offer the most opportunities for young people, and which sectors need to improve to provide accessible employment? From an industry perspective, Distribution & Hospitality and Public Administration, Education & Health remain the most significant youth employers. However, Construction and Manufacturing continue to lag in terms of youth engagement, signaling the need to create more accessible opportunities in these fields, particularly in regions where these sectors are key to local economies.

B. Insights from Applying Information Visualisation Techniques

Through this coursework, I gained practical insight into the process of building dashboards, visual sheets, and interactive charts not just from a technical perspective, but from the point of view of the end user. It really got me thinking more like an analyst. I designed each part of the dashboard with a specific audience in mind, always considering what would make it more visually engaging, easier to navigate, and more useful in drawing insights. I now understand how to choose the right chart types depending on the kind of data I am working with. For example, categorical data is best visualized using bar charts, trend data is suited to line charts, and high-dimensional data is best represented with scatterplots. I also learned how to structure multiple sheets into a clean, layered dashboard that tells a coherent story.

I also learned how to improve readability by filtering out noise, simplifying complex datasets, and using features such as filters and highlights to guide the user's focus. Making the dashboard interactive, exploring tooltips, dynamic filtering, and hover effects was especially interesting and, honestly, a lot of fun to experiment with.

REFERENCES

[1] ONS. "Labour Market Statistics - Youth Unemployment Timeseries," Office for National Statistics. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/timeseries/mgsx/lms

[2] UK Government. "Drivers of Social Mobility: Work Opportunities for Young People," Social Mobility Commission. Available at: https://social-mobility.data.gov.uk/drivers_of_social_mobility/work_opportunities_for_young_people/youth_unemployment/latest

[3] UK Census Data. "Census Data for 2011 and 2021," Office for National Statistics. Available at: https://www.census.gov.uk (All data used for analysis in this report is sourced from this database.)

[4] Munzner, T. (2014). *Visualisation Analysis and Design*. CRC Press.

[5] Ware, C. (2010). *Visual Thinking for Design*. Elsevier.

[6] Tufte, E. R. (2001). *The Visual Display of Quantitative Information* (Vol. 2). Cheshire, CT: Graphics Press.

[7] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer. Available at: https://www.microsoft.com/en-us/research/people/cmbishop/prml-book/

[8] Nabney, I. T. (2004). *Netlab: Algorithms for Pattern Recognition*. Springer.