1. Abstract

This study is concerned with the socio-economic factors that affect the distribution of work from home (WFH) in England and Wales (2011-2021). The authors analysed census data to assess the influences of sex ratios, accommodation type (e.g., flats vs detached) and access to public transport and cycling infrastructure on rates of WFH across 33,000+ Lower Layer Super Output Areas (LSOAs). To reduce dimensionality, the authors used Principal Component Analysis (PCA) and t-SNE to group LSOAs into similar clusters before employing Bayesian Linear Regression to predict WFH trends up to 2030. Results are presented using interactive Tableau dashboards as an evidence base for urbanisation driven policy-making to ameliorate differences in socio-economic factors and to better assess future workforce mobility and infrastructure development as societal change remains a long-lasting feature deriving from COVID-19. The key findings show strong correlations between WFH expansion, increased population employment rates, aspects of cycling infrastructure and high proportions of flats. However, WFH is less popular and expands at slower rates in groups or neighbourhoods with high student populations or less accessibility to transport options.

2. Introduction

The COVID-19 pandemic catalyzed one of the most profound shifts in the contemporary labor market: the universal embrace of remote work, or "working from home" (WFH), as it is popularly referred to. Remote work was mostly confined to some knowledge-based industries and nations with advanced digital infrastructures before 2020. Yet, the pandemic precipitated a sudden and extensive transformation across many industries, disrupting conventional employment norms and inducing a rethink of job prospects, residential development, internet connectivity, and work-life balance.

This research explores how various socio-economic and geographical determinants influenced the ability of people to work from home in the period from 2011 to 2021. Using neighborhood-level census data for England and Wales, it explores WFH inequalities by employment sector, gender, housing type, and travel behavior. The research uses previous and current census data to contrast pre- and post-pandemic socio-economic profiles. Using quantitative modeling and data analysis, the current research illustrates underlying structural trends and temporal shifts, thereby offering a multifaceted explanation of the ways in which job types, home environments, and spatial disparities influenced the trajectory of remote work.

2.1 Socio-Economic Issue Addressed:

What are the socio-economic drivers or limiting factors that propelled or hindered the adoption of remote work (WFH) across different areas of England and Wales from 2011 until 2021, and how these trends may evolve in the near future.

I tackle this issue on four core socio-economic themes, each underscored by numerous visual and statistical components in the project:

- **Employment Status:** Examines how economic activity levels, derived from 2011 and 2021 employment/unemployment rates, influenced the adoption of remote work.

- **Gender Distribution:** By comparing gender proportions and their relationship with WFH rates over the two years,I examine possible gender-based differences in WFH availability.

- **Accommodation Type:** I evaluate if type of housing infrastructure (e.g., detached houses versus flats) is relevant to facilitating remote work, taking into account space and living conditions.

- **Transport Modes:** Commuting dependencies (i.e., car, metro, foot) are quantified to investigate whether public vs. private transport dependence determines remote work adoption.

These factors are not separate; they capture interlinked features of regional growth, economic resilience, digital infrastructure, and equity.

2.2 Target Audience:

This study is intended to inform:

- **Policy makers** prioritizing regional digital infrastructure investments
- **Urban planners** designing WFH-friendly housing and transport systems
- **Socio-economic researchers** investigating post-pandemic labor behaviors
- **Employers and HR professionals** shaping hybrid work strategies

The visualization analytics approach, utilized in Tableau, is very interactive and meant to enable such stakeholders to explore the data by themselves — area by area, socio-demographic group by group, or WFH behavior by behavior — so as to develop actionable insights.

2.3 Project Goals:

This initiative has as its main goals:

- At the LSOA (Local Super Output Area) level, compare regional WFH absorption trends in 2011 and 2021.

- Apply PCA and t-SNE projections based on employment, housing, gender, and transportation to reveal latent clusters and outliers.

- Forecasts possible WFH rates for 2030 by means of Bayesian Linear Regression, so providing a forward-looking policy viewpoint.

- Geospatial dashboards, KPIs, top/bottom LSOA comparisons, and demographic filters help you visualise spatial inequalities.

- Create interactive Tableau dashboards including toggles for year and metric selection to support user-driven, insight-based decision-making.

- These objectives are reached by means of careful design of interactive visuals together with rigorous data preparation and modelling. This twin approach supports clear, powerful narrative as well as exploratory study.

3. Data Preparation and Abstraction

The exploratory data analysis conducted in this project relies on a wide synthesis and abstraction of disparate census datasets from diverse sources, viz., the years 2011 and 2021. What follows is the complete data pipeline — covering steps such as data wrangling, feature engineering, normalization, merging, and dimensionality reduction — undertaken to create a high-quality dataset for use in statistical modeling as well as visual analytics.

3.1 Data Sources and Tables Used

In accordance with coursework specifications, the project draws upon three 2011 Census datasets and one 2021 Census dataset, selected on the basis of thematic relevance to remote working behavior. These datasets cover employment status, accommodation types, transportation modes, and demographic composition.

Domain	2011 Table	2021 Table
Gender	TS008: Sex	TS008 (2021) – same variable
Employment	TS066: Economic Activity Status	TS066 (2021) – Employment, Unemployment, Students
Accommodation	TS044: Accommodation Type	TS044 (2021) – Housing Type breakdown
Transport	TS061: Method of Travel to Work	TS061 (2021) – Mode of Transport incl. WFH

All datasets were provided at the **LSOA (Lower Layer Super Output Area)** level and contained a **Location_Code** column, which served as the primary join key across all data sources and time periods.

3.2 Data Cleaning Process (Python)

Each dataset was individually preprocessed using **pandas** in Python. The main cleaning steps included:

- Standardizing and renaming columns for consistency and clarity
- Filtering out suppressed or missing data entries (e.g., NaN, '-', or zero totals)
- Normalizing raw counts into rates or percentages for meaningful comparison across LSOAs:
    - Employment_Rate = Employed / Total_Pop
    - Flat_Pct = Flats / Total_Households
    - Work_From_Home_Rate = WFH / Total_Working

3.3 Merging and Alignment

After data cleaning, all of the datasets were joined on the shared Location_Code. The resulting combined dataset contained 33,647 records—one for each LSOA—maintaining geographic specificity. This combined data platform facilitated easy temporal comparison across 2011 and 2021 for all the variables.

The final and detailed dataset was exported with the filename tableau_final_combined.csv, which served as the direct input for visualization and modeling procedures.

3.4 Feature Engineering: Deriving "Change" Metrics

To capture socio-economic shifts over time, delta features were computed to quantify the change in selected metrics between 2011 and 2021. These derived variables — such as WFH_Change, Employment_Change, and Flat_Change — are central to the comparative dashboards and KPIs throughout the visual analytics workflow.

3.5 Dimensionality Reduction for Projection View

In order to reveal hidden structures and clustering tendencies existing in the dataset, two dimensionality reduction methods were applied:

- **Principal Component Analysis (PCA)**
- **t-Distributed Stochastic Neighbor Embedding (t-SNE)**

Both algorithms were run with scikit-learn. A subset of scaled numeric features was taken as input, and the resulting projections were saved to projection_output.csv. These projections were subsequently joined back to the main dataset to facilitate visual exploration of regional clusters in Tableau.

3.6 Bayesian Prediction of 2030 WFH Rates

To fulfill the Bayesian modeling component, a **Bayesian Linear Regression** model was trained to predict future remote work rates (WFH_Rate_2030). The model leveraged trends between 2011 and 2021 across key socio-economic variables. The predicted values were incorporated into the final dataset, enabling users to explore future scenarios through dedicated visualisations.

3.7 Final Dataset Abstraction

- Gender ratios (2011, 2021)
- Employment, student, retired, and inactive population metrics
- Detailed accommodation and transport mode distributions
- Work-from-home rates and change metrics (2011 to 2021)
- Dimensionality reduction outputs (PCA, t-SNE)
- Predicted 2030 WFH rate (Bayesian estimate)

- Each variable was clearly labeled to reflect its year of origin and standardized by type (e.g., count, percentage, rate), ensuring consistency and interpretability throughout the project.

4. Task Definition

The objective of this project is to investigate, elucidate, and document the influence of various socio-economic factors—i.e., employment, gender, location, and transportation—on the uptake of Work From Home (WFH) in England and Wales during the past decade. To tackle this multifaceted analytical challenge in an organized manner, the project employs Tamara Munzner's four-level task taxonomy (Why, What, How, Where) as the visual design foundation.

4.1 Why – The Analytical Objectives

The primary objective of this project is **exploratory data analysis and visual storytelling** to support policy formulation and urban planning. Specifically, the tasks addressed include:

- Identifying **spatial and temporal trends** in WFH adoption between 2011 and 2021
- Examining the **relationships** between employment structures, gender composition, housing types, and transport modes in shaping WFH behavior
- Highlighting **regional disparities** through comparison of LSOAs, detecting outliers, and uncovering spatial heterogeneity
- Providing **forward-looking projections** of WFH rates for 2030 using Bayesian Linear Regression
- Delivering the above insights through a suite of **interactive dashboards** targeted at policymakers, researchers, and urban planners

4.2 What – Data and User Actions

To support these goals, we defined a series of **data-centric user tasks** across the data abstraction layer. These tasks map directly to our visualisations:

Data Types Used:

- **Quantitative**: Employment Rate, Student %, WFH %, Flat %, Gender Ratios, Transport Rates
- **Categorical**: Housing Type, Transport Mode, Cluster ID
- **Geospatial**: LSOA Code (used as a unique location identifier)

Lookup	Retrieve WFH rate, Gender %, Employment rate for a specific LSOA
Filter	Use parameter controls to isolate a subgroup (e.g., Top 10 LSOAs by WFH)
Compare	Compare 2011 vs 2021 rates of WFH, employment, gender, etc.
Explore	Discover underlying clusters via PCA/t-SNE projections
Summarize	Identify general patterns across clusters, regions, and years
Predict	Use Bayesian regression to forecast 2030 WFH rates based on 2011–2021 data
Correlate	Show associations between WFH growth and factors like gender, transport

These represent realistic user goals. Keeping business rationale in mind, a policymaker would probably compare WFH rates across districts, while an urban planner would correlate types of houses with remote work trends.

4.3 How – Visual Encodings and Interactivity

We selected **visual encodings** and interaction techniques based on both **data type and analytical task**. Here's how different visual components were structured:

Visual	Encoding	Description
Choropleth Map	Color gradient	To encode WFH % change by LSOA on the map
Line Charts	Line + Year dimension	WFH trends over 2011–2021–2030
Bar Charts	Length + Color	Compare % changes in employment/gender/ accommodation
Scatterplots	Position (X,Y) + Color	Show relationships (e.g., WFH vs Gender)
PCA/t-SNE Projections	X, Y coordinates + Cluster Color	Reveal socio-economic structures
KPI Tiles	Text, color, arrows (▲▼)	Highlight changes in key indicators
Parameter Controls	User dropdowns	Enable toggle between metrics (e.g., Male vs Female)

All graphical elements feature data-revealing tooltips with raw data, indicator shift, local identifiers, and work-from-home percentage estimates in 2030. The rationale behind these design choices lies in cognitive efficiency and user-oriented information visualization principles to minimize cognitive load while maximizing clarity and confidence in the analysis.

4.4 Where – View Arrangement & Dashboard Design

Our views were organized into **six thematic dashboards**, each designed to support a distinct analytical goal:

Dashboard	Focus Area	Layout
1. Geographic WFH View	Shows spatial change in WFH (%) across LSOAs	Filled map + tooltips + hover
2. WFH Temporal Trends	Multi-year WFH evolution (2011–2021–2030)	Line chart with cluster filter
3. Top/Bottom LSOAs	Identify best/worst performing areas	Dual bar charts (Top 50, Bottom 50)
4. Socio-Economic PCA View	Clusters LSOAs based on high-dimensional features	PCA scatterplot with tooltip
5. Gender, Accommodation vs WFH	Compare male/female ratio vs WFH	Toggle scatterplot with parameter control
6. Transport and WFH	How commute mode affects WFH	Transport vs WFH scatter + comparison

Each dashboard was **self-contained, interactive, and linked through filters and parameters**, allowing users to drill into areas or subgroups of interest while maintaining context.

5. Visualisation Justification

This project uses a varied range of visualization methods to analyze, interpret, and present the socio-economic trends behind the uptake of remote working (WFH) from 2011 to 2021, and predict its future trends up to 2030. In each case of visual method employed, choice which was informed by the respective data types to be dealt with (quantitative, geospatial, temporal) and cognitive goals of target consumers, following established best practices of visualization design and perceptual psychology.

The visualisation choices are guided by Munzner's Nested Model, Cleveland and McGill's perceptual effectiveness ordering, Bertin's visual variables, and Shneiderman's information-seeking mantra: "Overview first, zoom and filter, then details on demand."

5.1 Choropleth Maps (Geospatial WFH Change View)

Technique Used: Filled maps with color encoding of WFH % change (2011–2021) across LSOAs.

Justification:

- Color hue and saturation are very good at encoding ordinal or ratio data on maps, especially for users to notice regional patterns.

- The map supports geospatial pattern recognition spatially embedded data is optimally interpreted intuitively in its geographic context.

- We used diverging color scales (blue to green) to naturally encode positive vs negative change in WFH, leveraging pre-attentive processing to facilitate instant comparisons.

Perceptual Design:

- Luminance contrast and saturation render differences instantly visible.
- Tooltip interactivity offers “details-on-demand”
- Designed for summarize and lookup tasks, to allow policymakers to spot regional disparities in remote work adoption.

5.2 Line Charts (Temporal WFH Trends: 2011–2021–2030)

Technique Used: Line plots of WFH % over time with cluster filter options.

Justification:

- Line plots are the gold standard for visualization of time series data because they convey value by position on a common scale, the most accurate channel articulated by Cleveland & McGill (1984).
- Our trendlines show WFH variation at three points in time, including the 2030 estimate of Bayesian Linear Regression, helping users predict future policies.

Interaction Design:

- Users can easily identify rate of change, inflection points, and growth trends using slopes.
- Colored by cluster, it supports group comparison, aiding explore and compare tasks.

5.3 Bar Charts (Top/Bottom LSOAs, Gender, Housing, Transport Comparisons)

Technique Used: Vertical bar charts encoding categorical breakdowns of % change in measures.

Justification:

- Bars take advantage of length (a high-placed channel) for effective value comparison between regions or groups.
- Bar charts are particularly good at ranking, filtering, and showing extremes (e.g., Top 50 vs Bottom 50 LSOAs), which is important for summarize and communicate tasks.

Interaction Design:

- Highlight actions and filters allow users to quickly select categories or areas of interest.
- Facilitates Shneiderman's: "Zoom and filter" and "Details on demand".

5.4 Scatterplots (Gender vs WFH, Transport vs WFH, Cluster Projections)

Technique Used: 2D scatterplots to show correlation between percentage of Male/Female and WFH, modes of transport and WFH and dimensionality-reduced socio-economic clusters (PCA and t-SNE projections)

Justification:

- Scatterplots facilitate the detection of correlation and outliers, which in turn assists in hypothesis generation.

- Encoding data through spatial location is effective cognitively for comparing two quantitative variables.

Perception Support:

- Users are able to easily see trends, clusters, and outliers visually.

- Tooltip reveals values of 2011, 2021, and 2030 WFH and socio-economic data.

PCA and t-SNE projections in Tableau include tooltips showing the LSOA code, key socio-economic features (employment, transport, housing), and the predicted WFH % for 2030, enabling local-level interpretation.

5.5 KPI Tiles (Key Performance Indicators)

Technique Used: Numeric indicator tiles with colored arrows and delta values (e.g., ↑ +8.2% WFH)

Justification:

- **Cognitive chunking** of key metrics) helps reduce information overload.

- KPIs support **monitoring tasks**: instantly identify significant change without scanning entire charts

5.6 Dimensionality Reduction: PCA & t-SNE

These were applied in Python to map high-dimensional socio-economic data (2011 + 2021 features) into 2D scatterplot projections, which were visualized in Tableau with cluster labels.

- **PCA (Principal Component Analysis)**

Why Used: To map linear variance of multi-dimensional socio-economic features into 2D space for interpretable clustering.

Justification: PCA maintains global structure and is suitable for exploratory analysis at the beginning. It helps identify the significant socio-economic patterns responsible for variance in WFH adoption.

Scientific Basis: PCA is justified by Jolliffe (2002) and has broad application as an exploratory multivariate analysis method if variables are inter-correlated.

- **t-SNE (t-Distributed Stochastic Neighbor Embedding)**

Why Used: To uncover non-linear relationships among data and classify LSOAs based on local neighborhood similarity.

Justification: While PCA captures global patterns, t-SNE is best at discovering unobvious clusters from multi-variable proximity.

Scientific Basis: As noted by van der Maaten & Hinton (2008), t-SNE performs well with high-dimensional socio-economic embeddings, especially where non-linear separability exists (e.g., subtle variation in transport/housing profiles).

5.7 Perceptual and Cognitive Principles Applied

Principle	Application
Pre-attentive Attributes	Used color, size, and position for rapid pattern recognition in maps and scatterplots.
Cleveland & McGill Ranking	Prioritized position and length over area or angle for value encoding.
Shneiderman's Mantra	Specific ashboard provides overview, filter/zoom, and details-on-demand.
Cognitive Load Reduction	Summary indicators (KPIs) and filter actions allow intuitive exploration without overload.

6. Evaluation

I was adhering to Munzner's Nested Model for Visualization Design and Validation (2014), the project involved evaluation early during the design process, when data preparation and dashboard layout were in place but numerous refined views and interactions (e.g., PCA/t-SNE projections, prediction plots, and completed KPI tiles) were still in progress.

The evaluation consisted of four peer classmates from my discussion group who reviewed an in-progress version of theTableau dashboard.The version was built with a working layout, foundation maps, initial employment, gender visualizations and partially implemented dot plots. The more advanced dimensionality reduction results, enhanced interactivity, and tooltips were not yet included at this stage. My peers made multiple comment on trends, collaborate with filters, and provide constructive, focused feedback—especially for further detailed usability and visual appearance

Peer Feedback Summary :

Evaluation Dimension	Avg. Score (1–5)	Summary of Peer Feedback
1. Overall Impression	4.0	Peers found the theme timely and well-structured for early design. Scope was clear and relevant.
2. Usability	3.6	Basic navigation worked, but peers noted difficulty in understanding advanced views (PCA/t-SNE).
3. Design & Layout	3.5	Layout was clean; one peer suggested adopting a more appealing color palette—this led to a full redesign inspired by their work.
4. Data Representation	3.8	Peers liked the early choropleths and KPI tiles. Scatterplot views were still abstract and lacked legend/tooltips.
5. Functionality	3.7	Filters and hover interactivity were appreciated. Peers suggested adding regional comparison options.
6. Insights & Value	3.9	Even in the incomplete state, some users could spot patterns in gender/WFH and employment clusters.

A peer had advised using a bolder and more user-friendly palette so as to make the categories, such as accommodation types and transport use, more distinct. That suggestion prompted me to undertake changes in the visual theme as a whole, borrowing several cues from the peer's own visual project, to enhance visual consistency and readability.

7. Conclusion

7.1 Learning from the Socio-Economic Issue

From rich analysis of 2011 and 2021 census data for England and Wales, the project discovered significant patterns and differences in working from home take-up. There were several primary lessons that could be learned:

- **Employment Type & WFH:** Higher work-from-- home levels in 2021 were found in areas with higher full-time employment rates and student populations. These results imply that remote working was more adopted in university towns and areas where knowledge-intensive, white-collar businesses were highly present.

- **Gender & Work Patterns:** Visual analysis of gender composition revealed that LSOAs with higher shares of women in the working population were more likely to have lower levels of WFH take-up, especially where part-time and economically inactive statuses were also prevalent. This mirrored persistent employment flexibility and access-based gender differentials.

- **Housing & Infrastructure:** Flats (and converted flats and commercial, in particular) as forms of accommodation appeared closely associated with rises in WFH levels. They have a tendency to dominate in locations that are more economically and digitally developed in order to connect, thus catering to higher remote working adoption levels.

- **Transport Changes:** One important finding of the transport analysis is the notable decline in public transport reliance, particularly metro use, in places where work-from-home has grown. City designers must give great thought to this strong inverse correlation as they adjust to changing transport needs and possible demand for less transit infrastructure.

- **Predictive Modelling:** Through Bayesian Linear Regression, the project forecast WFH in 2030, forecasting ongoing (if slower) growth. This gives evidence-based view for upcoming policy and planning, especially regarding digital infrastructure, housing development, and employment support.

7.2 Lessons Learned Regarding Information Visualisation

Learning the methodology was very important and in addition of that, designing , deploying, and demonstrating visual analytics technology was also equally important. This course wasn't just constructing graph data, but rather taking complex, multi-dimensional data and turning it into an exploratory narrative that is visually appealing to the users—policy makers and researchers, for instance—can readily understand and act on.

Key lessons:

- Dimensionality reduction techniques (PCA and t-SNE) helped to expose hidden patterns across multiple features but required their resulting output to be interpreted carefully visually. Without interaction, tooltips, and legends, such projections are hard to understand for general users.

- Peer review at the beginning taught me how valuable visual consistency, intuitive interaction, and the power of a good color scheme are. Even when my dashboard was not yet finished, peer review gave me actual advice and creative inspiration. Incorporating iterative validation into the design process enabled me to refine not just the appearance, but the story I wanted to tell.

- I found that clarity arises through intentional curation. It is especially true in Tableau, where dashboards, filters, and calculated fields must be planned carefully for ease of cognition.

- Python integration with Tableau was also a key technical skill. Translating projection and regression results from code to Tableau required stringent data preparation and pushed me to bridge the gap between storytelling and data science.

In summary, this coursework has greatly improved my knowledge of information visualization as a critical analytical and communicative tool. I have learned that the effectiveness of visualization goes beyond simply showing trends; it is instrumental in making data-driven decisions understandable, fair, and inclusive. Through the analysis of socio-economic changes related to remote work, as well as the careful tweaking of different filters and axes in Tableau, this project has reinforced the idea that good visualization uncovers the hidden and makes the complex actionable.

References

I. Cleveland, W.S. and McGill, R., 1984. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. *Journal of the American Statistical Association*, 79(387), pp.531–554.

II. Jolliffe, I.T., 2002. *Principal Component Analysis*. 2nd ed. New York: Springer.

III. Munzner, T., 2014. *Visualization Analysis and Design*. Boca Raton: CRC Press.

IV. Shneiderman, B., 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: *IEEE Symposium on Visual Languages*. Boulder, CO: IEEE, pp.336–343.

V. van der Maaten, L. and Hinton, G., 2008. Visualizing Data using t-SNE. *Journal of Machine Learning Research*, 9(Nov), pp.2579–2605.

VI. Canva, 2024. *Color Theory for Designers*. [online] Available at: https://www.canva.com/colors/ [Accessed 7 May 2025].