A black text on a white background

Description automatically generated

**Student Name: Prajwal Bharti**

**Student Id Number: A00325600**

**Course: Masters in Data Analytics**

**Title of Assignment: Data Visualization Project**

**Date: 6 April 2025**

**Declaration**

I hereby certify that the material, which is submitted in this report towards the award of MSc. in Data Analytics, is entirely my own work and has not been submitted for any academic assessment other than part fulfilment of the above named award.

Signed Prajwal Vithal Bharti Date 06/04/2025

**Aim**

To explore how professional data sourced from LinkedIn reflects disparities and trends in salary, career progression, and diversity across gender, ethnicity, and geography.

**Introduction**

In today's digital age, social media has become more than a space for connection it is a rich repository of professional and personal data. LinkedIn is the world's largest social and professional networking platform, it has offered insight into the workforce which is both real-time and has diversity overall. The platform shows professional trajectories, salary expectations, job data, and demographic details that, when correctly analysed, can show vivid picture of the changing job market.

This project aims to uncover key trends in salary distribution, job progression, and diversity metrics filtered through the lens of social media data. By leveraging a structured dataset resembling LinkedIn profiles, we examine whether social media can provide reliable signals about income inequality, progression gaps, and representation across various segments. The core focus lies in understanding how social media-derived data reflects real-world disparities in pay and progression, and what that means for businesses and policy frameworks moving forward.

With millions of users self-reporting career milestones, achievements, and transitions, LinkedIn acts as a living resume database. This opens the door to non-traditional data analysis that can help uncover hidden trends whether its which demographic groups are rising fastest in leadership roles, or how geography intersects with pay equity. The digital footprints left on platforms like LinkedIn provide more than just personal branding they are measurable indicators of systemic trends.

The analysis focuses on a unified theme: how social media reveals workplace realities. Each visualization explores an element of workforce diversity or compensation, linking it back to insights that social platforms like LinkedIn can offer. This approach is useful for strategic decision-making in human resources, diversity and data-driven organizations seeking to build equal workplaces.

**Table of Contents**

1. Title Page ..................................................................................................................................1

2. Declaration ................................................................................................................................2

3. Aim ............................................................................................................................................2

4. Introduction ...............................................................................................................................3

5. Glossary of Terms .....................................................................................................................4

6. Data Preparation and Exploratory Data Analysis ..................................................................... 6

7. Background to the Project .........................................................................................................8

8. Data Visualizations

Chart 1: Salary Distribution by Gender .....................................................................................10

Chart 2: Job Progression vs Salary ............................................................................................11

Chart 3: Salary Distribution by Ethnicity ..................................................................................12

Chart 4: Geographic Salary Distribution ...................................................................................13

Chart 5: Expected vs Actual Salary ...........................................................................................14

9. Map Dashboard Overview ........................................................................................................15

10. Overview Dashboard Summary ..............................................................................................16

11. Conclusion and Recommendations .........................................................................................17

12. Ethical Issues in DataVisualization ........................................................................................18

13. References ...............................................................................................................................19

14. Appendices

Appendix A: Raw Data Snapshot ..............................................................................................20

Appendix B: Cleaned Data Snapshot ........................................................................................ 20

Appendix C: Data Dictionary ................................................................................................... 21

Appendix D: Choosing the Data Visualizations ........................................................................22

**Glossary of Terms**

|  |  |
| --- | --- |
| Term | Description |
| Annual Salary | The total compensation received by an individual in one year, before deductions. |
| Expected Salary | The predicted or benchmark salary for a person based on their years of experience. |
| Job Progression Index | A custom metric that quantifies career advancement based on role transitions. |
| Salary Bracket | Grouped ranges of annual salaries (e.g., 40K–60K, 60K–80K) used for distribution analysis. |
| Experience Level | Categorization of individuals based on job seniority (e.g., Entry, Mid, Senior, Executive). |
| Ethnicity | Demographic classification referring to cultural identity, used here for diversity analysis. |
| Bullet Chart | A chart comparing actual values (e.g., salary) against expected benchmarks. |
| Box Plot | A statistical graph showing salary spread, median, and outliers by demographic group. |
| Scatter Plot | A chart used to identify trends between two continuous variables (e.g., progression vs. salary). |
| Map (Filled) | A geographical chart showing average salaries by U.S. state using colour intensity. |
| Parameter (What-If) | An interactive slider allowing users to simulate outcomes based on variable input (e.g., years of experience). |
| Filter | Interactive control that allows viewers to narrow down visualized data (e.g., by gender or ethnicity). |
| Tenure | The number of years an individual has worked in their current position. |
| LinkedIn Engagement | A metric reflecting user activity such as posts, likes, comments, or profile views. |
| Followers | The number of LinkedIn users following the profile, indicating network size. |
| Nationality | The country with which the individual identifies or holds citizenship. |
| Progression Band | A binned grouping of individuals based on their job progression index. |
| Dashboard | An interactive interface displaying visual representations of data for analysis. |
| Data Cleaning | The process of correcting or removing inaccurate, corrupted, or irrelevant data. |
| Calculated Field | A new field created using formulas applied to existing data within Tableau. |
| Outlier | A data point that significantly deviates from other observations in a dataset. |
| What-If Analysis | A technique to simulate different scenarios and outcomes by adjusting inputs. |

**Data Preparation and Exploratory Data Analysis**

Before working on the visual dashboard design in Tableau, I spent some time exploring and preparing the dataset to make sure everything was clean, consistent, and ready for analysis. The dataset was designed to simulate LinkedIn profiles and included a wide range of professional details such as gender, ethnicity, job title, company, salary, job progression, experience level, and even social media metrics like follower count.

When I loaded the dataset into Excel to check through the data it revealed some inconsistencies in the way the data was formatted. For example, some columns had missing values, certain entries were written differently (like “male” vs. “MALE”) in gender column, and a few numeric fields weren’t properly set as numbers. There were also some blank cells and placeholder values that needed to be cleaned up.

**Cleaning and Preparing the Data**

Here’s a breakdown of what I did to clean and prepare the dataset:

* **Removed incomplete records**: Any rows that were missing key information like gender, salary, experience, or ethnicity — were removed. These fields were essential to the analysis and leaving them in could have skewed the results.
* **Fixed inconsistent entries**: I standardised all text-based categories. For instance, different spellings or capitalisations for genders or ethnic groups were made consistent (so "MALE", "male", and "Male" all became just "Male").
* **Formatted numbers correctly**: I made sure fields like *Annual Salary*, *Job Tenure*, and *Job Progression Index* were set as numerical values. I also adjusted the decimal places — keeping *Job Tenure* to one decimal and *Progression Index* to two — so they would look clean in Tableau.
* **Created new calculated fields**:
  + *Job Progression Index*: A custom metric I created using the number of roles and time spent in those roles, showing how quickly someone advances in their career.
  + *Expected Salary*: A field that estimates what someone should be earning based on their experience.
  + *Salary Ratio*: This compares the actual salary to the expected one, helping highlight over- or underpaid individuals.
* **Grouped continuous values**:
  + I grouped salaries into brackets (like $20K–$40K, $40K–$60K, etc.) to make charts easier to interpret.
  + I also grouped the progression index into bands to simplify comparisons across individuals with similar career movement.
* **Removed unrealistic outliers**: A few entries showed extremely high salaries for entry-level roles (e.g., over $500,000), which didn’t seem reasonable. I removed those outliers.
* **Cleaned up locations**: For the map visualisation, I made sure all location names were aligned with standard U.S. states so Tableau could map them correctly.
* **Final formatting**: Once all the cleaning was done, I exported the dataset into a fresh Excel file with clear column headers and no hidden formatting issues — making it fully ready to use in Tableau.

**What I Noticed During Exploration**

As I cleaned and explored the data, a few interesting patterns started to emerge. I noticed differences in how salaries were distributed by gender and ethnicity, and variation in how job progression aligned with pay. These early observations helped shape the dashboard design — and are what led me to use visual tools like box plots, scatter plots, and bullet charts to bring those patterns to life.

**Background to the Project**

In today’s digital world, social media is not only just used for sharing personal updates or connecting and talking casually. It has evolved into as a powerful source of behavioural, demographic, and professional data. **LinkedIn** stands out as the world’s largest professional connection network, with over 900 million users globally, among all the social media platforms available. Each user profile on LinkedIn gives insights on his career through out his life. Because users often keep their profiles up to date, LinkedIn provides a present and past constantly refreshed snapshot of the global journey of a user.

This project is built on a dataset inspired by that LinkedIn ecosystem. The data has been structured to resemble real LinkedIn profiles, including fields such as gender, ethnicity, nationality, experience level, annual salary, job title, tenure, and a job progression index. It also integrates social media–specific signals like follower count and engagement activity, which add another dimension online professional presence and influence.

The main goal of the project is to explore whether professional data used from a social platform like LinkedIn can provide real-world trends in how people are paid, how they grow in their careers, and how different demographic groups are shown. In short, can we use social media to shine a light on inequality in the workplace? And if so, how can that details help businesses make more informed, fairer decisions?

In practice, most of the organisations struggle to make full use of the data that’s already available. HR departments and analysts often don’t have the tools or time to tap into social media platforms for the insights. Even though LinkedIn gives an large valuable pool of information, it’s still not more often used resource in strategic planning, especially when it comes to equity, diversity, and inclusion.

This project goal is to demonstrate how tools like **Tableau** can help unlock that potential. By converting simulated LinkedIn data into clear, interactive visuals, the dashboard enables decision-makers to spot gaps in pay, progression, and representation briefly. This information can then produce real action whether it's shaping recruitment strategies, adjusting pay policies, or revisiting promotion practices.

Further, this reflects a shift in how organisations are moving towards workforce analytics. Rather than depending solely on internal reports or survey data, businesses have started to look outward using real-time, user-generated content from social media to use it in traditional analysis. As this trend will continue, platforms like LinkedIn will play an even more key role in helping companies in building workplaces that are not just productive, but also fair.

**Data Visualizations**

**Chart 1: Breaking the Bracket – Salary Distribution by Gender**

This chart presents the distribution of male and female professionals across various salary brackets. On visualizing the salary ranges across the genders, the chart reveals a pattern is visible indicates a higher number of men in upper salary brackets compared to women.   
  
While there are relatively same number of male and female represented in the 40K–60K range, the proportion of male professionals in the 80K+ range grows significantly, which suggest a disparity in income distribution between genders. This visualization suggests the importance of addressing systemic pay gaps and ensure to give equal opportunity for compensation across the genders.

A screenshot of a graph

AI-generated content may be incorrect.

**Interpretation:**  
This bar chart reveals how salaries are distributed between male and female professionals across various income brackets. A significant concentration of male professionals appears in higher salary ranges (especially above $80,000), while female representation tends to peak in mid-range brackets (e.g., $40K–60K and $60K–80K). This visualization suggests a persistent gender pay disparity, with fewer women achieving top-tier compensation. While entry-level pay brackets appear more balanced, the skew in upper ranges highlights potential systemic barriers to equal earning opportunities for women as careers progress.

**Chart 2: Climbing the Ladder – Job Progression vs Salary**

The scatter plot maps the Job Progression Index against Annual Salary, with colors representing experience level. As expected, there is a general upward trend, showing that increased career progression often correlates with higher pay.   
  
However, a deeper look reveals that individuals with similar progression scores do not always earn comparable salaries, highlighting inconsistencies in compensation structures. The color dimension adds another layer entry-level individuals are more tightly grouped, while executives show a broader salary spread. This variability suggests differences in how industries or companies reward advancement and could indicate unequal access to high-paying roles even for those with similar progression metrics.

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretation:**  
This scatter plot presents the relationship between an individual's job progression index and their annual salary, color-coded by experience level. The upward trend suggests that higher progression generally correlates with higher salary. However, salary distribution remains highly varied, especially among individuals at the same progression level but different experience bands. This indicates that career advancement alone doesn't guarantee equitable salary increases and points toward other influencing factors, such as industry, role specialization, or demographic characteristics. It also reveals that while executives may have greater salary potential, they also experience a broader salary range, reflecting a high variance in compensation at senior levels.

**Chart 3: Boxing the Bias – Salary Distribution by Ethnicity**

Using a box plot, this visualization displays salary distributions across three ethnic groups. The box height reflects the interquartile range, while lines indicate minimum, maximum, and median values.   
  
Asian professionals tend to have a slightly higher median salary, though the range of salaries varies within all groups. Notably, the presence of outliers above the upper range suggests exceptional high earners in each ethnicity.   
  
This chart is especially useful for exploring pay equity. While central tendencies are relatively balanced, the variability and spread hint at structural inconsistencies in how different demographic groups are compensated in similar roles.

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretation:**  
The box plot illustrates how salaries vary across different ethnic groups. While median salaries appear relatively close between groups, the interquartile ranges and outliers highlight notable differences. Asian professionals show a slightly higher median and a wider spread, suggesting greater earning potential but also greater inequality within the group. Meanwhile, the narrower salary spread among Black professionals may indicate a lack of upward salary mobility. This visualization helps understand that while median salaries seem to be similar, the distribution tells a more about access to high-paying opportunities and representation in top-tier roles.**Chart 4: Mapping the Pay – Geographic Distribution of Salaries**

This filled map visualizes average salaries by U.S. state. Dark shades indicate higher average salaries, whereas the lighter shades represent lower average salary.   
  
The map clearly reveals geographic disparities. For example, states such as California and New York show high salary averages, likely to indicate their strong economies and concentration of corporate jobs. Conversely, rural states show lower salary averages.

A map of the united states

AI-generated content may be incorrect.

**Interpretation:**  
This filled map highlights average annual salaries across U.S. States like California, New York, and Texas appear to be in dark green shades, indicating higher salary averages. This aligns with expectations, as these states typically have larger job markets and higher costs of living. On the other hand, lighter shades states such as Idaho and West Virginia reflect lower salary averages. The map strongly emphasizes regional economic disparities and provides important context for understanding how location influences compensation. These insights are particularly useful for remote work planning, salary benchmarking, and relocation policies.

**Chart 5: Expectation vs Reality – Are Salaries Meeting Aspirations?**

This bullet chart compares actual average salaries to expected salaries based on years of experience. Users can adjust the years of experience via a parameter slider, updating the 'expected' benchmark dynamically.   
  
The chart clearly shows that in many cases, actual salaries fall short of expectations, particularly for women. This gap illustrates a form of silent inequality, where professionals may technically progress but not be rewarded equitably.   
  
What-if analysis tools like this empower users to simulate realistic salary expectations, helping organizations develop transparent and justifiable pay structures that align more closely with experience and performance.

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretation:**  
This bullet chart compares actual salaries to expected salaries calculated based on years of experience (adjustable via a parameter). Across the board, actual salaries tend to fall below expected benchmarks especially among female professionals. This discrepancy reveals a potential mismatch between career growth and compensation, even when controlling for experience. The chart allows for dynamic "what-if" analysis by letting users simulate various experience levels, making it a powerful tool for career planning, equity audits, and strategic salary calibration.

**Map Dashboard – Exploring Salary Patterns by Location**

**Description:**  
The Map Dashboard offers a focused view at how average salaries differ across U.S. states. Using a filled map, each state is shaded based on its average annual salary darker greens represent higher earnings, while lighter shades indicate lower pay levels. This visual approach helps users instantly identify regions with higher or lower compensation trends.

The dashboard is interactive, allowing users to hover over a state to view its exact salary figure. This interactivity provides clarity and encourages exploration, making it especially useful for analysing geographic salary gaps.

A map of the united states

AI-generated content may be incorrect.

**Interpretation:**  
This dashboard gives the insight on the relationship between location and salary. States like California, New York, and Texas appears to be in darker shades, as due to the higher economy centres they reflect to be in higher wages states, also the higher costs of living. In contrast, states with lighter shades, such as Iowa or Mississippi, shows lower average salaries.

The visualization takes attention on how geography influences pay, offering valuable insights for those planning relocations, companies setting region-based compensation, or policymakers aiming to bridge regional income gaps. It demonstrates that where a person works plays a significant role in their earning potential.

There is also a button in the dashboard which allows user to navigate from the map dashboard to overview dashboard. Also, If the user clicks on any location the overview dashboard shows only particular data to that location.

**Overview Dashboard – A Holistic View of Workforce Trends from Social Media**

**Description:**  
The Overview Dashboard brings together multiple key charts to provide a broad perspective on salary distribution and workforce dynamics. It includes visualizations that cover differences in pay by gender and ethnicity, the link between job progression and income, geographic salary differences, and the comparison between actual and expected salaries.

This dashboard is fully interactive, with filters for gender, ethnicity, and experience level that update all charts simultaneously. A “Back to Map” button offers seamless navigation between the broader overview and the more location-specific analysis.

A screenshot of a computer

AI-generated content may be incorrect.

**Interpretation:**  
This Overview Dashboard serves as the central narrative of the project, bringing together multiple key insights from the dataset in a single, unified space. Each chart reflects a different lens through which workplace trends—such as salary inequality, progression disparities, and representation gaps—can be examined. By curating these views side by side, the dashboard encourages both a detailed and comparative exploration of data inspired by LinkedIn’s professional ecosystem.

Top Left – Gender Divide Across Salary Brackets:

This bar chart clearly exposes a gender imbalance in higher salary ranges. Males are disproportionately represented in brackets above 80K, while females dominate the lower brackets. This visualization not only shows the distribution but calls attention to how structural disparities still limit women’s access to top-paying roles.

Top Right – Salary Expectations vs. Reality:

The bullet chart compares actual average salaries to expected benchmarks by gender. Across the board, both genders fall slightly short of expected pay—but females appear to be further behind. The interactive 'Years of Experience' slider offers a dynamic way to simulate expected pay, making this an engaging tool to explore how experience gaps contribute to earnings disparity.

Bottom Left – Pay Gaps by Ethnicity:

The box plot explores salary ranges across different ethnic groups. While median salaries appear close, the spread and outliers show deeper variation. For example, the White group shows the widest distribution, suggesting greater earning potential but also increased inequality. The visualization highlights the need to look beyond averages and understand the full distribution of opportunity.

Bottom Right – Job Progression and Pay:

This scatter plot shows the relationship between the Job Progression Index and Annual Salary, segmented by experience level. The concentration of points around lower progression values reflects that many professionals remain at stagnant levels despite years of service. The outliers are few—showing that rapid or upward progression is rare, especially among entry- and mid-level roles. This insight is particularly valuable for organizations looking to diagnose bottlenecks in internal mobility.

**Conclusion**

This project has highlighted the potential of social media platforms particularly LinkedIn as powerful tools for uncovering workforce patterns and inequalities. Through a carefully constructed Tableau dashboard, we explore how salary, career progression, and demographic factors intersect across the different professional scenarios. The analysis revealed persistent gender and ethnic pay gaps, geographical disparities and inconsistencies between expected and actual salaries. These patterns emphasize the need for organizations to adopt more inclusive and data-driven approaches to human resource management.

To address the findings, several actionable recommendations can be made. First, companies should implement transparent pay structures and conduct regular audits to ensure salaries are equitable across demographics. Additionally, targeted mentorship programs and clear progression pathways should be introduced to support underrepresented groups in advancing their careers. HR departments can also benefit from using data-driven recruitment strategies informed by regional salary and diversity insights. It is equally important for organizations to align expected salaries with actual compensation benchmarks to manage both internal expectations and fairness. Finally, companies should begin integrating social media data into their HR analytics strategies to gain real-time insights into diversity, progression, and compensation trends.

For future research and development, several avenues can be explored. A time-series analysis could track how inequalities evolve over time and whether diversity initiatives are making a measurable impact. By segmenting based on industry level can uncover whether some fields are making more progress than others. Additionally, incorporating social media influence metrics such as follower count or endorsement levels could reveal correlations between online presence and professional outcomes. Further, comparing salary trends at the company level could enable benchmarking against industry averages. Lastly, adding sentiment analysis from public posts or endorsements could offer qualitative depth to the numerical trends.

In summary, this project demonstrates how structured, user-generated social media data can be transformed into powerful workforce intelligence. With growing demand for inclusivity and transparency, platforms like LinkedIn when analysed thoughtfully offer a unique window into the realities of modern work and can drive meaningful change in organizational practices.

**Ethical Issues**

While data visualisation is a powerful tool for communication and decision-making, it also has some ethical responsibilities attached especially when working with data that relates to people, such as gender, ethnicity, salary, and career progression. In this project, which uses simulated data inspired by LinkedIn profiles, several key ethical considerations were identified during the visualisation process.

**1. Misrepresentation Through Design Choices**  
The most important ethical responsibilities in data visualisation is to ensure that the visuals accurately reflect the used data. In this project, it was essential to choose chart types that communicated both trends and limitations clearly. For example, showing only averages could have concealed real disparities within demographic groups. Instead, using box plots and scatter plots allowed for a more nuanced view. Misleading visual design like manipulated axes or omitted data points can distort the truth and lead to biased conclusions (Kirk, 2019).

**2. Oversimplification of Demographic Data**  
Working with categories like gender and ethnicity always carries a risk of oversimplification. In this project, users were grouped into broad demographic labels to aid analysis, but such grouping can unintentionally erase the diversity within each category. For example, the term "Asian" can include vastly different cultures, experiences, and economic conditions. While necessary for analysis, such simplifications must be acknowledged so viewers understand the limitations of the visualised data (D’Ignazio & Klein, 2020).

**3. Privacy and Anonymity Concerns**  
Although this dataset is simulated, real-world applications using LinkedIn or similar platforms would involve personal data that users may not expect to be analysed at scale. Even anonymised data can sometimes be reverse-engineered to identify individuals, especially when combined with geographic or company-level details. Ethical visualisation must prioritize data protection and avoid presenting information that could indirectly expose individuals (GDPR Portal, 2023).

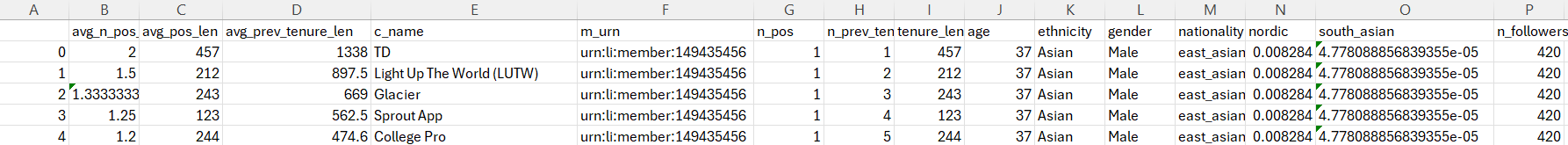
**4. Reinforcing Bias Through Data Gaps**  
The visualisations created in this project rely on available fields such as salary, job title, and experience level. However, not all groups are equally represented on platforms like LinkedIn—certain populations, such as blue-collar workers, older individuals, or those from rural areas, may be underrepresented. Building visuals based on incomplete or skewed data can unintentionally reinforce existing social and economic inequalities (O’Neil, 2016). Being transparent about these limitations is essential to ethical storytelling.

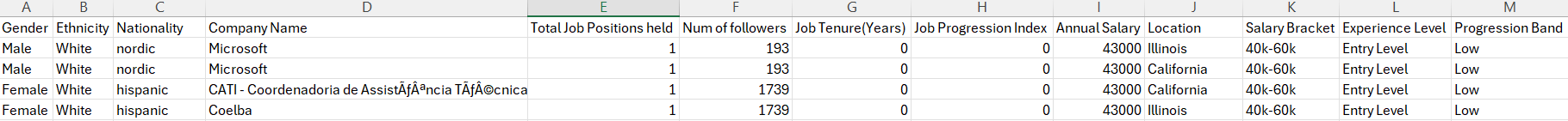
**5. Emotional Impact and Interpretation**  
Data related to income, identity, and inequality is deeply personal. The way visuals are designed can influence how viewers interpret these issues, sometimes evoking emotional reactions. For example, seeing one’s own group consistently underpaid or underrepresented can be frustrating or disheartening. Ethical data visualisation requires sensitivity—not just in how charts are built, but in the narrative that accompanies them. Designers have a responsibility to foster understanding without alienating or blaming specific groups.

**References**

* Tableau Software (2024) *Tableau Desktop User Guide*. Available at: https://www.tableau.com/learn (Accessed: 6 April 2025).
* Harvard Business Review (2020) *The Real Value of LinkedIn Data for Employers*. Available at: <https://hbr.org/2020/06/the-real-value-of-linkedin-data> (Accessed: 6 April 2025).
* Microsoft (2024) *Excel Support*. Available at: <https://support.microsoft.com/en-us/excel> (Accessed: 5 April 2025).
* LinkedIn Learning Tableau Course(2024) at: <https://www.linkedin.com/learning/tableau-2024-1-essential-training>
* IBM (2023) *What is Exploratory Data Analysis (EDA)?* Available at: https://www.ibm.com/topics/exploratory-data-analysis (Accessed: 6 April 2025).
* Kirk, A. (2019) *Data Visualisation: A Handbook for Data Driven Design*. 2nd edn. London: SAGE Publications.
* D’Ignazio, C. and Klein, L.F. (2020) *Data Feminism*. Cambridge, MA: MIT Press.
* GDPR Portal (2023) *General Data Protection Regulation Summary*. Available at: <https://www.eugdpr.org/> (Accessed: 7 April 2025).
* O’Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Crown Publishing.
* Statista (2024) *Number of LinkedIn Members Worldwide*. Available at: https://www.statista.com/statistics/274050/quarterly-numbers-of-linkedin-members/ (Accessed: 6 April 2025).

**Appendices**

**Appendix A: Raw Data Snapshot**

**Appendix B: Raw Data Snapshot**

**Appendix C: Data Dictionary**

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Data Type** |
| Gender | Gender of the individual | Nominal |
| Ethnicity | Ethnic background category | Nominal |
| Nationality | More specific nationality or region | Nominal |
| Company Name | Current employer of the individual | Nominal |
| Total Job Positions held | Total job roles held across career | Discrete |
| Num of followers | LinkedIn follower count | Discrete |
| Job Tenure (Years) | Years of experience in current job | Continuous |
| Job Progression Index | Custom index of job progression | Continuous |
| Annual Salary | Annual income in currency (e.g., USD) | Continuous |
| Location | Assigned geographic location for mapping | Nominal |
| Salary Bracket | Categorized salary range | Ordinal |
| Experience Level | Grouped career level by tenure | Ordinal |
| Progression Band | Category of job progression score | Ordinal |

**Appendix D: Choosing the Data Visualizations**  
When I was designing the data visualizations for this project, my primary goal was to clearly translate complex, multidimensional data into visuals that anyone, from an HR executive to a analyst, could interpret the data easily. Since the dataset reflected simulated LinkedIn data, the chosen visualizations should be able to communicate differences in salary, job progression, and diversity without overwhelming the viewer.

The process began by mapping out the key questions I wanted to explore:

* How does salary differ across gender and ethnicity?
* Is there a gap between expected and actual earnings?
* Does job progression influence income?
* Are there geographical patterns in compensation?

With those goals in mind, I explored various chart types for each question. Here's how the selection unfolded:

**1. Bar Chart – Salary Distribution by Gender**

Initially, I considered using a pie chart to visualize different salary bracket proportions by gender, but it quickly became evident that pie charts would make it hard to interpret information and it was showing a lot of eye travel. Instead, a bar chart was selected for its simplicity and effectiveness in showing comparisons across gender and salary ranges side by side. This visual quickly reveals disparities in how males and females are distributed across different income brackets.

**2. Scatter Plot – Job Progression vs Salary**

To examine the relationship between job progression and salary, a scatter plot was the natural choice because the variables are continuous variables. Even though I initially considered using a line chart, but since the data points weren’t sequential or cumulative, that would have been misleading. The scatter plot allows everyone's data to be plotted based on their progression index and salary, revealing clusters and outliers that would otherwise be hidden.

**3. Box Plot – Salary by Ethnicity**

The objective in this chart was to show not just averages, but the **spread and concentration** of salaries among different ethnic groups. I initially tested a bar chart showing mean salary per ethnicity but found it insufficient it didn’t show outliers or variability and then I remember that I have used the bar chart before so charts will get repetitive. The box plot, with its ability to highlight medians, interquartile ranges, and outliers, gave a more complete picture of salary equity across groups.

**4. Bullet Chart – Expected vs Actual Salary**

This chart was meant to compare how real salaries measured up against expectations. A dual-axis bar chart was considered, but it cluttered the view. The bullet chart, on the other hand, provided a clean and direct comparison, with the benchmark (expected salary) which is a calculated field I created with a parameter for handling different years of experience in the background and actual values in the foreground. This made it easier to spot gaps.

**5. Map (Filled) – Geographic Salary Differences**

To explore geographic patterns, I needed a chart that could link data to specific locations. A table or bar chart could show salary by state, but it lacked geographic context. A filled map was ideal it displayed average salaries by state using colour gradients, making regional disparities immediately visible.

**Rejected Chart Types and Why They Didn’t Work**

Several chart types were trailed but ultimately not used. For example:

* I rejected **Pie charts** for any comparison-based visuals because of its limitations in showing the small differences.
* **Stacked bar charts** were considered for ethnicity and gender comparisons but led to confusion when multiple categories were presented.
* I also considered using a **heat map** to display job progression levels, but it didn’t add much value in this case. Heat maps work best when you have a high number of categories on both axes to compare, and our dataset didn’t have that kind of structure. The variables were too limited for the chart to reveal any meaningful patterns, so I decided to go with more focused visuals instead.
* Initially, the project was designed to display all visualizations within a single dashboard. The idea was to create a one-stop view that would combine all insights in a cohesive manner. However, during the development phase, it became apparent that fitting multiple complex charts each with its own filters, legends, and contextual meaning into a single canvas created visual clutter. To resolve this, I changed the layout and split into **two different dashboards**: an **Overview Dashboard** focusing on diversity, salary distribution, and progression; and a **Map Dashboard** dedicated solely to geographic salary differences. This change significantly improved user experience by decluttering the visuals, it enhanced the user readability, and allowed viewers to focus on one theme at a time. The result is a cleaner, more intuitive flow that supports analytical storytelling and enables easier navigation between charts.

**Rejected Dashboard Layout:**

A screenshot of a computer

AI-generated content may be incorrect.

**Parameter & Filter Descriptions**

To enhance interactivity and make the dashboard more meaningful for users, both parameters and filters were incorporated. These features allow viewers to interact with the data, explore different scenarios, and uncover personalised insights, rather than just passively viewing the static charts.

**Parameter: Years of Experience Slider**  
A parameter was created to allow users to simulate different values for **years of experience**. This slider plays a key role in the bullet chart, where expected salary is calculated based on the selected experience level. By adjusting the parameter, users can see how the expected salary shifts in comparison to actual salaries. This “what-if” analysis feature makes it easier to answer questions like: *What should someone with 10 years of experience be earning?* or *how does actual pay align with industry expectations for mid-career professionals?* It adds a predictive element to the dashboard and makes the visualisation more interactive and scenario driven.

**Filters: Gender, Ethnicity, and Experience Level**  
Interactive filters were added for **gender**, **ethnicity**, and **experience level** to give users the ability to dive deeper into specific segments of the dataset. These filters are especially important when analysing pay gaps or representation, as they allow viewers to isolate one group at a time. For example, a user might want to look only at salaries for senior-level women or compare job progression among different ethnicities. These filters are designed to be intuitive and can be applied with just a click, making the dashboard flexible and user-friendly for different audiences—whether it’s an HR analyst looking into diversity trends or a recruiter exploring compensation by region.

**Global Application of Filters**  
All filters were applied **globally across the dashboard**, meaning that once a user selects a filter, it updates every chart in the view. This was done intentionally to keep the analysis consistent and prevent confusion. It ensures that the story being told by one chart is reinforced by the others, creating a cohesive and aligned exploration experience.