Graphical Geniuses
Vaishnavi, Tanmayee, Atharva

Final Project Proposal

1. Research Questions

This project explores **gender diversity and pay equity** in the tech industry, focusing on leadership representation in major tech firms and how these trends compare to broader U.S. labor data. Using two key datasets - **Diversity in Tech Companies** (Kaggle) and **U.S. BLS Wage Data.** We aim to assess whether leadership diversity aligns with equitable pay, and whether tech companies are ahead or behind other industries in this regard. Our analysis centers on three key narratives:

Climbing the Ladder:
How does gender and ethnic diversity change across job levels in tech companies?

- Many women enter at junior levels, but representation declines in senior roles—highlighting the "leaky pipeline".

Pay Gaps in the Pay Leaders:
Is greater female leadership linked to smaller gender pay gaps?

- This narrative examines whether diverse leadership teams are linked to broader compensation equity.

Beyond Tech:
How do gender pay gaps in the tech sector compare to other major U.S. sectors?

- We place tech in a national context to evaluate its performance in gender pay equity relative to sectors like healthcare, finance, and education.

These narratives will shape a data-driven story that reveals patterns in representation, leadership access, and pay fairness across sectors.

2. Target Audience

This project is intended for stakeholders focused on workplace equity, fair compensation, and organizational change:

- DEI Officers & HR Professionals: To support audits, identify gaps, and refine diversity strategies
- Corporate Executives & Policymakers: To inform planning, reporting, and equity-related decisions
- Researchers, Educators & Advocates: For use in studies, public engagement, and policy discussions
- Students & Data Professionals: As a real-world example of applying data visualization to equity issues

Our goal is to deliver insights that are clear, actionable, and relevant across professional, academic, and advocacy spaces.

3. Dataset Description

Primary Dataset: Diversity in Tech Companies (Kaggle)

- Source: Compiled by Kaggle user *jainaru*, using public DEI data and company press releases (2014–2018).
- Format: Clean, structured CSV format.
- Structure: Includes year, company, gender (% male/female), and ethnicity (% White, Asian, Black, Latino, Multiracial, Other).
- Credibility: Based on official company reports and media coverage.
- Limitations: Ends in 2018 and lacks intersectional breakdowns (e.g., race + gender).

U.S. Bureau of Labor Statistics (BLS) Wage Data

- Provides national pay gap statistics by gender and industry, offering critical context to compare tech with other sectors.

These datasets are well-aligned with our research questions and support both **company-specific** and **cross-industry** analysis of diversity and pay equity in tech.

4. Analysis Plan

We will analyze **gender diversity** (male vs. female) and **ethnic diversity** across tech companies from 2014 to 2018, focusing on trends and company comparisons. Key variables include year, company, gender distribution, and ethnic breakdown (White, Asian, Black, Latino, Multiracial, Other). This will help us identify which companies have made progress in diversity and where disparities still exist.

To analyze these variables, we will:

- Track diversity trends over time
- Compare companies to identify leaders and laggards
- Explore correlations between leadership diversity and pay equity
- All results will be visually represented using Tableau.

5. Visualization Plan

Our visual storytelling will highlight patterns in gender and ethnic diversity, leadership representation, and pay equity across tech companies. We plan to create **6–7 visualizations** using **Tableau**, selected for its interactivity and accessibility for both technical and non-technical users. These visuals will directly support our research questions and make complex data insights easier to explore and understand.

How Visuals Will Support Research Questions:

- Bar charts: Compare gender and ethnic representation across companies
- Line graphs: Track diversity and pay trends over time
- Funnel charts: Highlight drop-offs in representation at senior levels ("leaky pipeline")
- Scatter plots: Explore links between leadership diversity and pay equity
- Combined views: Compare tech industry trends with other major U.S. sectors

Designed for a diverse audience: HR professionals, executives, researchers, and students, our visuals aim to clearly communicate key insights and allow users to interact with data filters such as company, year, and demographic group.

We anticipate a few limitations: the dataset ends in 2018, restricting analysis of more recent trends, and it lacks intersectional granularity (e.g., race and gender combined), which limits deeper equity insights. Despite this, the available data provides strong support for our research questions and enables impactful visual storytelling.

6. APA Citations

Primary Dataset:
Jainaru. (n.d.). *Diversity in Tech Companies*. Kaggle. https://www.kaggle.com/code/jainaru/eda-on-diversity-trends-in-tech-companies/

Secondary Data Source:
U.S. Bureau of Labor Statistics. (n.d.). *Wages by Occupation and Industry*. https://www.bls.gov/bls/wages.htm