

# MGT 6203 Group Project Proposal Template

## TEAM INFORMATION (1 point)

**Team #:** 26

### **Team Members:**

1. Marissa Robinson, GTID: mrobin317. Currently a quantitative analyst at Bank of America, Graduated from Southern Methodist University with an undergraduate degree in Economics.
2. Emmett Drake; edrake7. Senior Data Analyst at Cardlytics; graduated from the University of Georgia with a BBA in Finance
3. William D'Onofrio; wdonofrio3  
Software Team Lead at Micron Technology with 6 years of experience and 2nd year of OMSA.
4. Jamie Hernandez Kluesner; jkluesner3  
Merchandise Planner Analyst at The Home Depot. Undergraduate degree in Neuroscience at Georgia Tech; first year of OMSA.
5. Sanjay Lindsay; slindsay32; Quantitative Analyst at Bank of America with 4 years of experience. 6<sup>th</sup> course in the OMSA program.

## OBJECTIVE/PROBLEM (5 points)

**Project Title:** A Statistical Analysis of the Gender Pay Gap

### **Background Information on chosen project topic:**

This year marks the 103-year anniversary of the ratification of the 19<sup>th</sup> Amendment. This amendment paved the way for a major change in women's rights and roles in the United States. However, the journey to full gender equality is not complete. Despite the passing of the 19<sup>th</sup> amendment, the women's rights movement of the 1960's, and that now more women are attending college than men, there is still a discrepancy between what a man and woman get paid for the same job.

### **Problem Statement (clear and concise statement explaining purpose of your analysis and investigation):**

The purpose of this analysis is to identify the major contributing factors of the gender pay gap.

### **State your Primary Research Question (RQ):**

What factors are most highly correlated with the pay disparity between men and women?

**Add some possible Supporting Research Questions (2-4 RQs that support problem statement):**

1. What departments and job titles have the highest gender pay gap disparity?
2. Does an increase in a woman's education help offset the gender pay gap?
3. For what age group(s) is the gender pay gap the largest?

**Business Justification: (Why is this problem interesting to solve from a business viewpoint? Try to quantify the financial, marketing or operational aspects and implications of this problem, as if you were running a company, non-profit organization, city or government that is encountering this problem.)**

Companies need to attract and retain top talent to stay competitive in the marketplace, period. A key aspect of attracting and retaining top talent is to offer competitive salaries and bonuses to employees. But in recent years, this alone has proven to not be enough. Now, in addition to offering the best financial incentives, companies must also ensure they employ a diverse workforce of all ages, genders, and backgrounds. By offering equal pay for equal work, companies will attract the top employees, create a more diverse work environment, and simultaneously best position the company for success.

## **DATASET/PLAN FOR DATA (4 points)**

**Data Sources (links, attachments, etc.):**

Data was found on Kaggle.com:

<https://www.kaggle.com/datasets/nilimajauhari/glassdoor-analyze-gender-pay-gap>

**Data Description (describe each of your data sources, include screenshots of a few rows of data):**

The following data was pulled from Glassdoor, a website that provides resources and information for job seekers, employers, and employees. The data was found in their research section on their site and uploaded to Kaggle.

It contains income for various job titles. It features gender, age, performance evaluation, education level, department, base pay, and bonus. Employees recorded a performance evaluation score between one to five. In addition, seniority is the number of years worked from one year as the minimum and five as the maximum. Education ranges from high school to PhD. Age varies and gender is either female or male.

Glassdoor Gender Pay Gap

JobTitle	Gender	Age	PerfEval	Education	Dept	Seniority	BasePay	Bonus
Graphic Designer	Female	18	5	College	Operations	2	42363	9938
Software Engineer	Male	21	5	College	Management	5	108476	11128
Warehouse Associate	Female	19	4	PhD	Administration	5	90208	9268
Software Engineer	Male	20	5	Masters	Sales	4	108080	10154
Graphic Designer	Male	26	5	Masters	Engineering	5	99464	9319
IT	Female	20	5	PhD	Operations	4	70890	10126
Graphic Designer	Female	20	5	College	Sales	4	67585	10541
Software Engineer	Male	18	4	PhD	Engineering	5	97523	10240

**Key Variables: (which ones will be considered independent and dependent? Are you going to create new variables? What variables do you hypothesize beforehand to be most important?)**

Dependent Variable: Income will be the dependent variable as it is influenced by several independent variables. The one in question in this project is if gender has the most significant influence on the pay gap seen between male and females making it the most important independent one.

Independent Variable:

1. Job Title
- 2. Gender**
3. Age
4. Performance Evaluation
5. Education
6. Department
7. Seniority
8. Base Pay
9. Bonus

A new variable will not be created.

## APPROACH/METHODOLOGY (8 points)

**Planned Approach (In paragraph(s), describe the approach you will take and what are the models you will try to use? Mention any data transformations that would need to happen. How do you plan to compare your models? How do you plan to train and optimize your model hyper-parameters?))**

Firstly, we will begin with data preparation: Loading and cleaning the dataset, handling missing data, and encoding categorical variables such as gender, education level, and department. Numerical features will be normalized or standardized as needed to ensure consistency in the analysis.

Next, we will conduct an exploratory data analysis (EDA) to gain insights into the dataset. Descriptive statistics and data visualization will be employed to understand the distribution of salaries across several factors such as gender, age, department, and education. Correlation analysis will help identify which factors are most highly correlated with the gender pay disparities.

To address the research questions, we will implement several models. First, a linear regression will be used to examine the relationship between pay and factors like education, age, and performance evaluation. We will experiment with log and other variable transformations to help improve model performance as well. Based on our learnings from building the linear regression model, we will then build a logistic regression model to predict gender-based pay disparities as a probability. Lastly, Random Forest or Decision Trees will be used for capturing complex interactions and feature importance.

To ensure the robustness of our models, we will split the dataset into training and testing sets for evaluation. We will use appropriate evaluation metrics to compare and measure model effectiveness, such as: R-Squared, Adjusted R-Squared, Accuracy, and ROC AUC.

For model optimization, we will experiment with hyperparameter tuning techniques. Grid search or random search will be used to systematically explore hyperparameter combinations, optimizing model performance. Cross-validation will be used to help prevent overfitting and ensure generalizability.

**Anticipated Conclusions/Hypothesis (what results do you expect, how will your approach lead you to determining the final conclusion of your analysis) Note: At the end of the project, you do not have to be correct or have acceptable accuracy, the purpose is to walk us through an analysis that gives the reader insight into the conclusion regarding your objective/problem statement**

1. What factors are most highly correlated with the pay disparity between men and women?

Hypothesis: We hypothesize that factors such as job performance evaluation scores, education level, and seniority (years of experience) will be highly correlated with the pay disparity between men and women. Specifically, we expect that women with lower performance scores, lower education levels, and fewer years of experience will experience a greater pay disparity compared to men.

2. What departments and job titles have the highest gender pay gap disparity?

Hypothesis: We anticipate that certain departments and job titles may exhibit more significant gender pay gap disparities than others. Our hypothesis is that male-dominated departments and job titles, such as engineering or management, will tend to have higher gender pay gaps compared to more gender-balanced or female-dominated departments and roles.

3. Does an increase in a woman's education help offset the gender pay gap?

Hypothesis: We hypothesize that an increase in a woman's education level will have a positive effect on reducing the gender pay gap. Specifically, women with higher levels of education, such as a master's or PhD, will experience a smaller pay gap compared to women with lower educational qualifications, such as a high school diploma or bachelor's degree.

4. For what age group(s) is the gender pay gap the largest?

Hypothesis: We anticipate that the gender pay gap may vary across different age groups. Our hypothesis is that the gender pay gap will be most significant among older age groups, possibly among employees aged 40 and above, as this group may have experienced historical gender pay disparities that have carried forward into their careers. Younger age groups may exhibit a smaller gender pay gap due to changing societal norms.

**What business decisions will be impacted by the results of your analysis? What could be some benefits?**

**Recruitment/Talent Acquisition:**

Business Decision: The company may need to review and revise its pay equity policies and practices to ensure equal pay for equal work, regardless of gender.

Benefits: Ensuring pay equity can improve the company's reputation and attractiveness to top talent.

**Employee Retention and Engagement:**

Business Decision: Implementing strategies to reduce the gender pay gap can improve employee satisfaction, morale, and retention rates.

Benefits: Retaining top employees leads to cost savings associated with recruitment, training, and increased productivity across the business.

**Performance Evaluation and Promotion Practices:**

Business Decision: Reviewing and revising performance evaluation and promotion practices to ensure they are fair and unbiased.

Benefits: Fair performance evaluation and promotion practices motivate employees, increase employee retention, and improve overall company performance.

**Compliance with Legal Requirements:**

Business Decision: Ensuring compliance with gender pay equity laws and regulations.

Benefits: Avoiding legal penalties and reputational damage associated with non-compliance.

## PROJECT TIMELINE/PLANNING (2 points)

Project Timeline/Mention key dates you hope to achieve certain milestones by:

### October 11<sup>th</sup>

Data has been cleaned and prepared for the models that will be tested.

- Conduct exploratory data analysis (EDA) to gain insights into the dataset.
- Handle missing data, outliers, and any other data preprocessing steps as needed.
- Document data preprocessing techniques for transparency in the final report.

### October 15<sup>th</sup>

Models' creation kicks off. Any unexpected problems, challenges, or findings will be recorded for the progress report.

- Linear Regression Models including evaluation
- Logistic Regression Models including evaluation
- Assess if additional models could be helpful
- Compare models and select best option

### Nov 4<sup>th</sup>

#### Progress Report Due

#### Hypothesis Testing:

- Conduct hypothesis tests to determine the significance of factors identified in the analysis.
- Identify variables that have a statistically significant impact on the gender pay gap.

### Nov 11<sup>th</sup>

- If more models were deemed needed, we would have them completed by this point.

#### Data Visualization:

Create visualizations that effectively communicate your findings, such as scatter plots, histograms, and heatmaps.

Use visualization to illustrate the gender pay gap disparities across different factors.

Begin writing final report. Necessary background information is completed. An overview of the problem and the general approach is recorded along with initial hypotheses.

**Nov 18th**

Final Report:

Overview of Data is completed for final report. The cleaning process, key variables, and dataset are described.

**Nov 25th**

Final Report:

Overview of modeling is completed for final report. The models used and the selection process are described. An in-depth discussion of how the models perform is recorded.

**Dec 3<sup>rd</sup>**

Final Report Due

**Dec 6th**

Teammate Evaluation Due