Factors Behind Gender Pay Gap

Jenny Gong

PPOL565: Final Project
04.23.2023

Outline

- Research Question and Background
- Data Sources, Target, and Important Features
- Parametric and Non-Parametric Techniques Used
- Evaluation and Interpretation
- Conclusion and Limitations

Research Question and Background

- Background
  - Median earnings for women in 2021 were 83.1 percent of the median for men U.S. Bureau Labor Statistics, 2022)
  - Pew Research Center reported that in 2022, women typically earned 82 cents for every dollar earned by men in full-time jobs (Carolina, 2023).

Research Question and Background

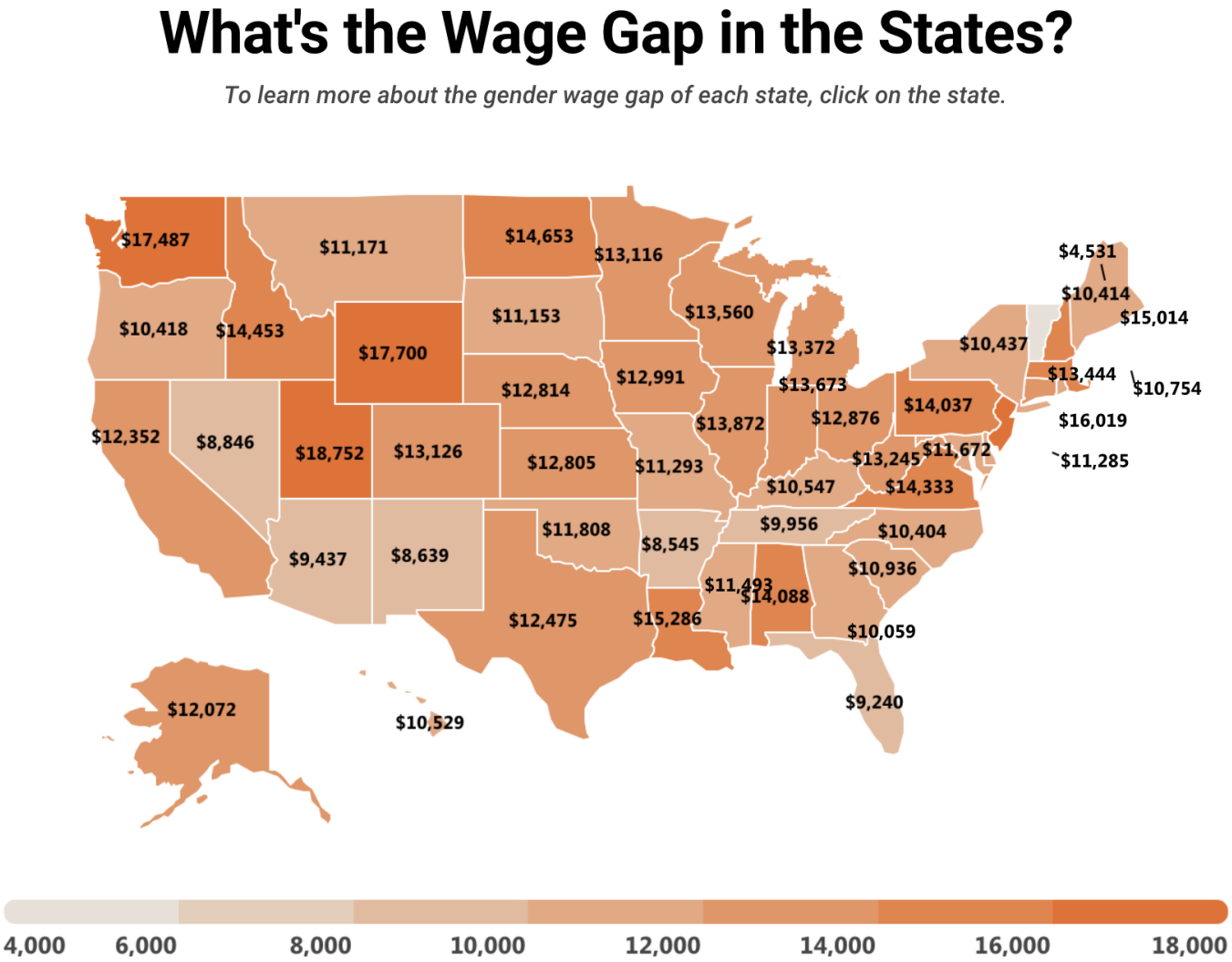

Figure 1: The wage gap in the U.S.

Research Question and Background

Figure 2: Payment Group Distribution

Research Question and Background

- Research question
  - 1. What is the extent of the wage gap between women and men in different age groups and pay-level groups? Furthermore, are women earning equal pay as men in the same department?
  - 2. What are the factors that influence the gender pay gap, and how significant are they?

Data Sources

- Data sources
- GlassDoor Economic Analysis Lab
  - The study aimed to assist employers in analyzing the gender pay gap within their organizations.
  - The data is stratified from a large dataset.
  - The dataset has 1000 rows and 9 features.

Target, and Important Features

- Target:‘BasePay‘
- Important Features: 'Age', 'JobTitle', 'Gender', 'Education', 'Dept', 'Seniority', 'PerfEval'.

Variable Label	Description
JobTitle	The specific job title held by each employee in the company.
Dept	The department within the company where each employee works, such as operations, management, or administration.
Seniority	An employee's level of experience and position within the company hierarchy, often designated by job titles or "tiers."
PerfEval	The performance evaluation score of each employee.

Parametric and Non-Parametric Techniques Used

- Parametric model: Multiple Linear Regression
- Non-Parametric model: Decision Tree

Evaluation and Interpretation

- Multiple Linear Regression

	('Model Men', 'Coefficient')	('Model Men', 'P-value')	('Model Women', 'Coefficient')	('Model Women', 'P-value')
Intercept	83827.92204720061	0.0000***	94872.18678737635	0.0000***
C(Education, Treatment(reference="High School"))[T.College]	4837.256858204929	0.0922*	1607.3309022981139	0.5653
C(Education, Treatment(reference="High School"))[T.Masters]	9653.571474938235	0.0003***	6646.42654918909	0.0206**
C(Education, Treatment(reference="High School"))[T.PhD]	9925.579954251394	0.0004***	10397.497514750039	0.0003***
JobTitle[T.Driver]	2888.2762385752426	0.5229	-8549.233298401188	0.0537*
JobTitle[T.Financial Analyst]	4747.954473750195	0.2627	1121.5632789179515	0.7971
JobTitle[T.Graphic Designer]	-552.6440018130179	0.9001	-2753.2634261453313	0.5313
JobTitle[T.IT]	1452.7013037879915	0.7464	-4442.461726598883	0.3049
JobTitle[T.Manager]	35067.4003949616	0.0000***	31003.704194297745	0.0000***
JobTitle[T.Marketing Associate]	-7906.124797589391	0.2845	-18766.736397281962	0.0000***
JobTitle[T.Sales Associate]	6040.518973270986	0.1655	-3426.4129960912123	0.4482
JobTitle[T.Software Engineer]	17224.517907738777	0.0000***	-1437.2376677230409	0.8629
JobTitle[T.Warehouse Associate]	-2007.00565742271	0.6581	-1850.7647290975474	0.6764
PerfEval	-227.45723106347418	0.7438	-1405.6897206937645	0.0506*

Figure 3: Multiple Linear Regerssion (1)

Evaluation and Interpretation

- Multiple Linear Regression

	('Model Men', 'Coefficient')	('Model Men', 'P-value')	('Model Women', 'Coefficient')	('Model Women', 'P-value')
Intercept	25283.226131046555	0.0000***	15527.216836215968	0.0000***
C(Dept, Treatment(reference="Operations"))[T.Administration]	363.04821861678283	0.8688	-1451.0784405537952	0.4893
C(Dept, Treatment(reference="Operations"))[T.Engineering]	6096.9236892874615	0.0054***	1789.566154171181	0.4017
C(Dept, Treatment(reference="Operations"))[T.Management]	5094.536411331481	0.0165**	4087.380449649927	0.0573*
C(Dept, Treatment(reference="Operations"))[T.Sales]	8638.306550715875	0.0001***	2108.3615710636486	0.3088
Age	1040.2735535793595	0.0000***	1006.5182095881821	0.0000***
Seniority	9025.223368601914	0.0000***	10307.046451125589	0.0000***

Figure 4: Multiple Linear Regerssion (2)

Evaluation and Interpretation

- Feature Importance
  - Education
    - College: Men
    - Master: Men, Women
    - PhD: Women, Men
  - Job Title
    - Driver: Women
    - Manager: Men, Women
    - Marketing Associate: Women
    - Software Engineer: Men
  - Department
    - Engineering: Men
    - Management: Men, Women
    - Sales: Men

Evaluation and Interpretation

- Decision Tree

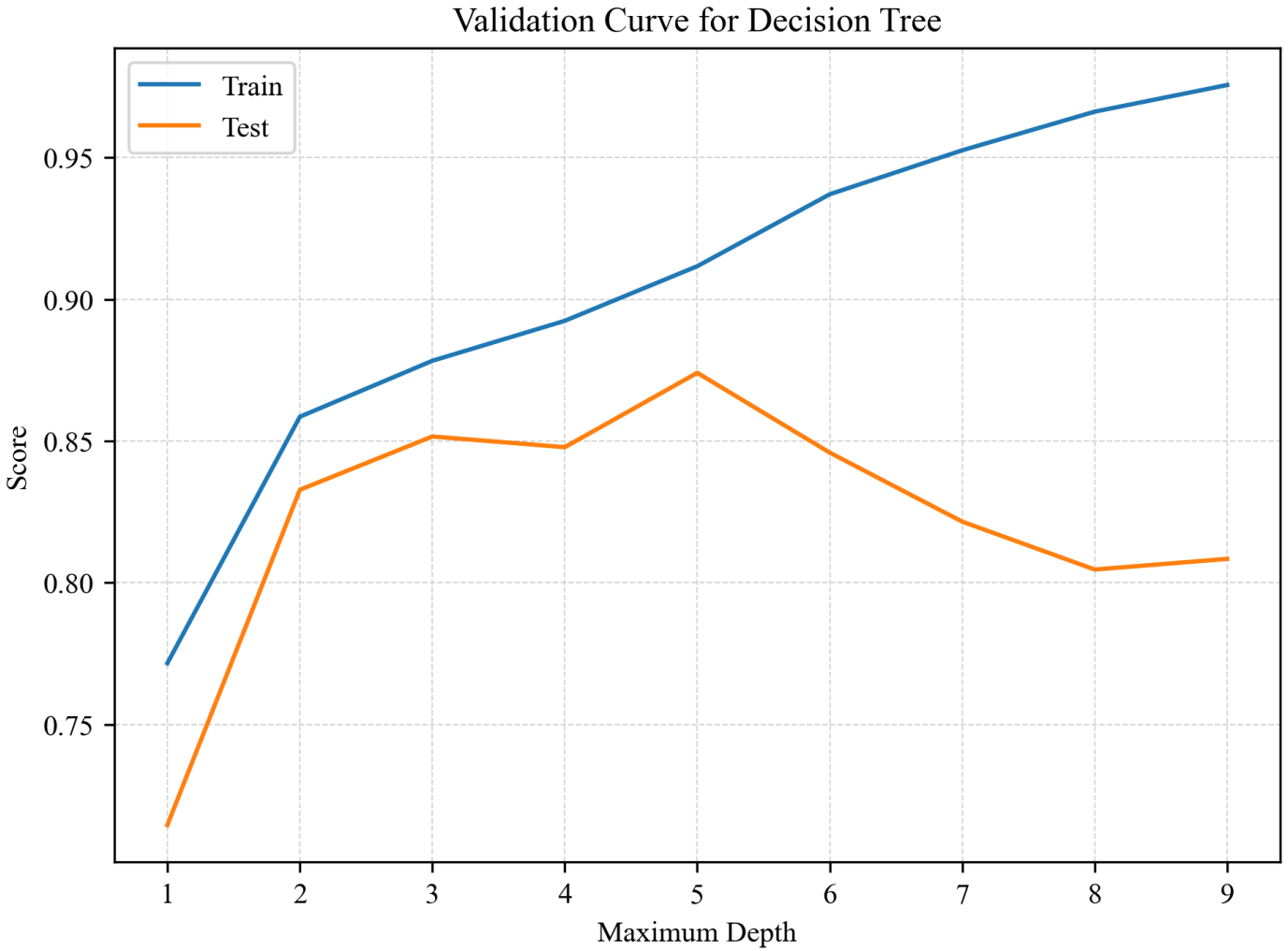

Figure 5: Validation cure - Decision Tree for Men's Data

Evaluation and Interpretation

- Decision Tree

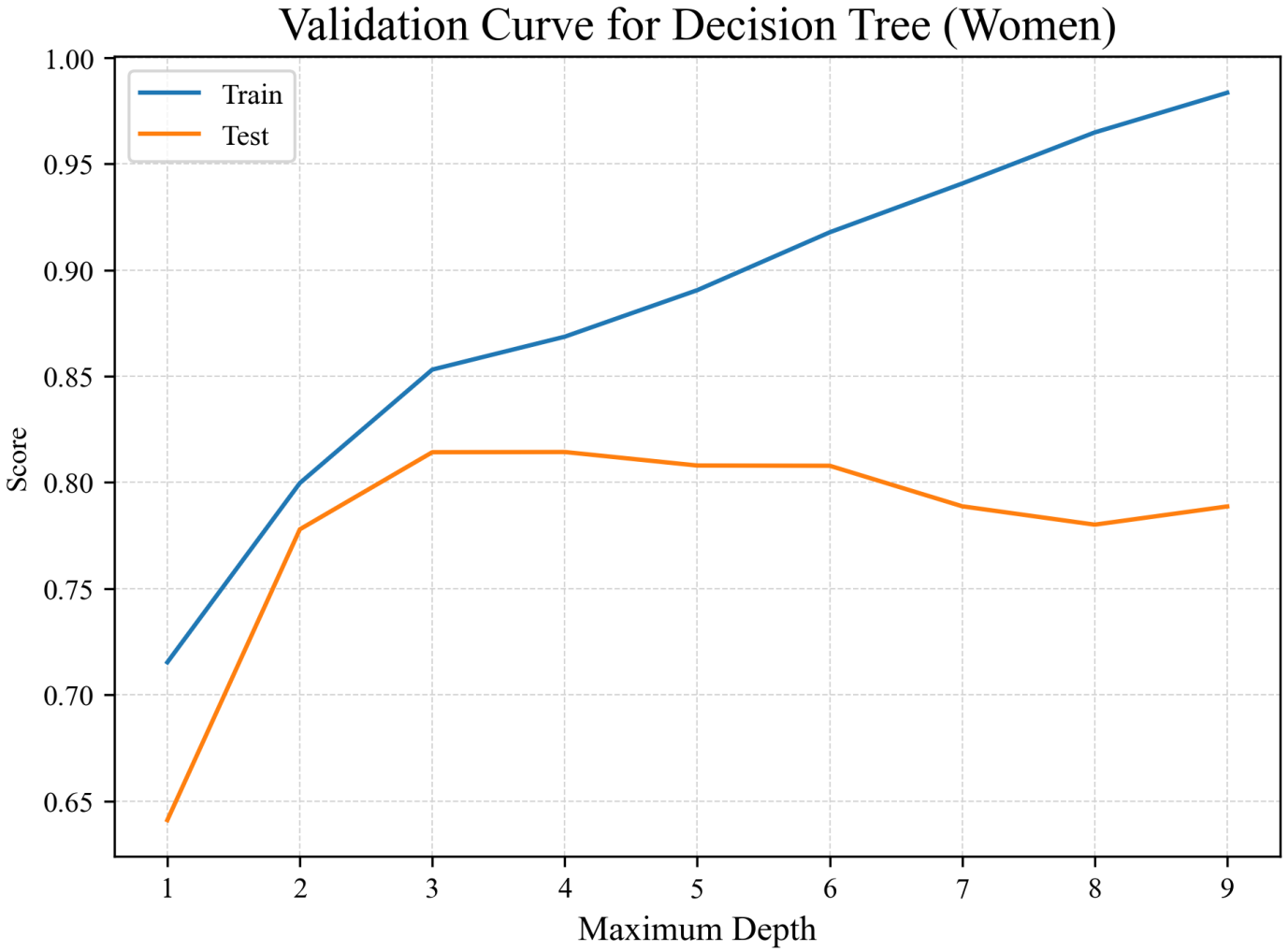

Figure 6: Validation cure - Decision Tree for Women's Data

Evaluation and Interpretation

- Decision Tree - Men

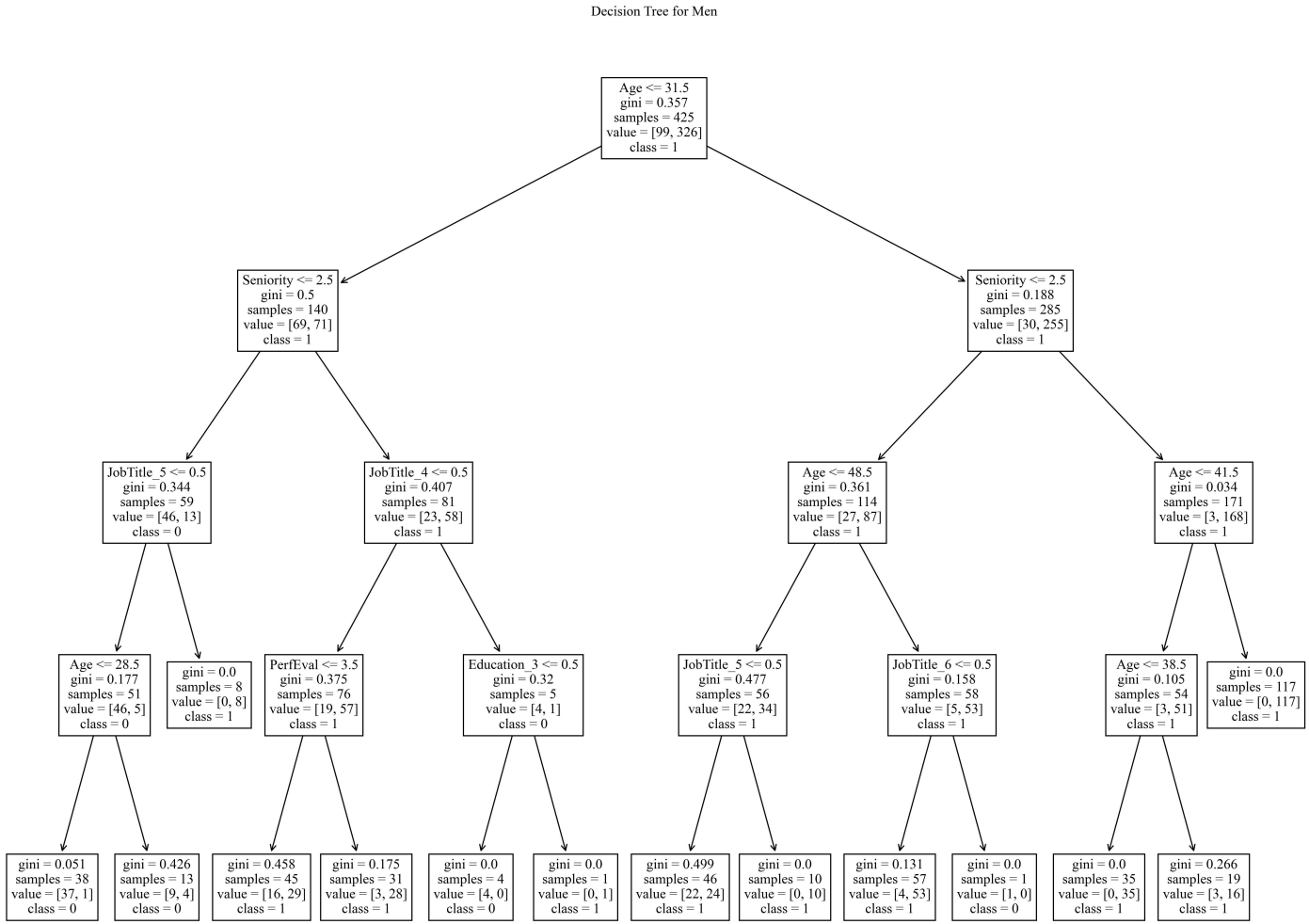

Figure 7: Decision Tree for Men's Data

Evaluation and Interpretation

- Decision Tree - Women

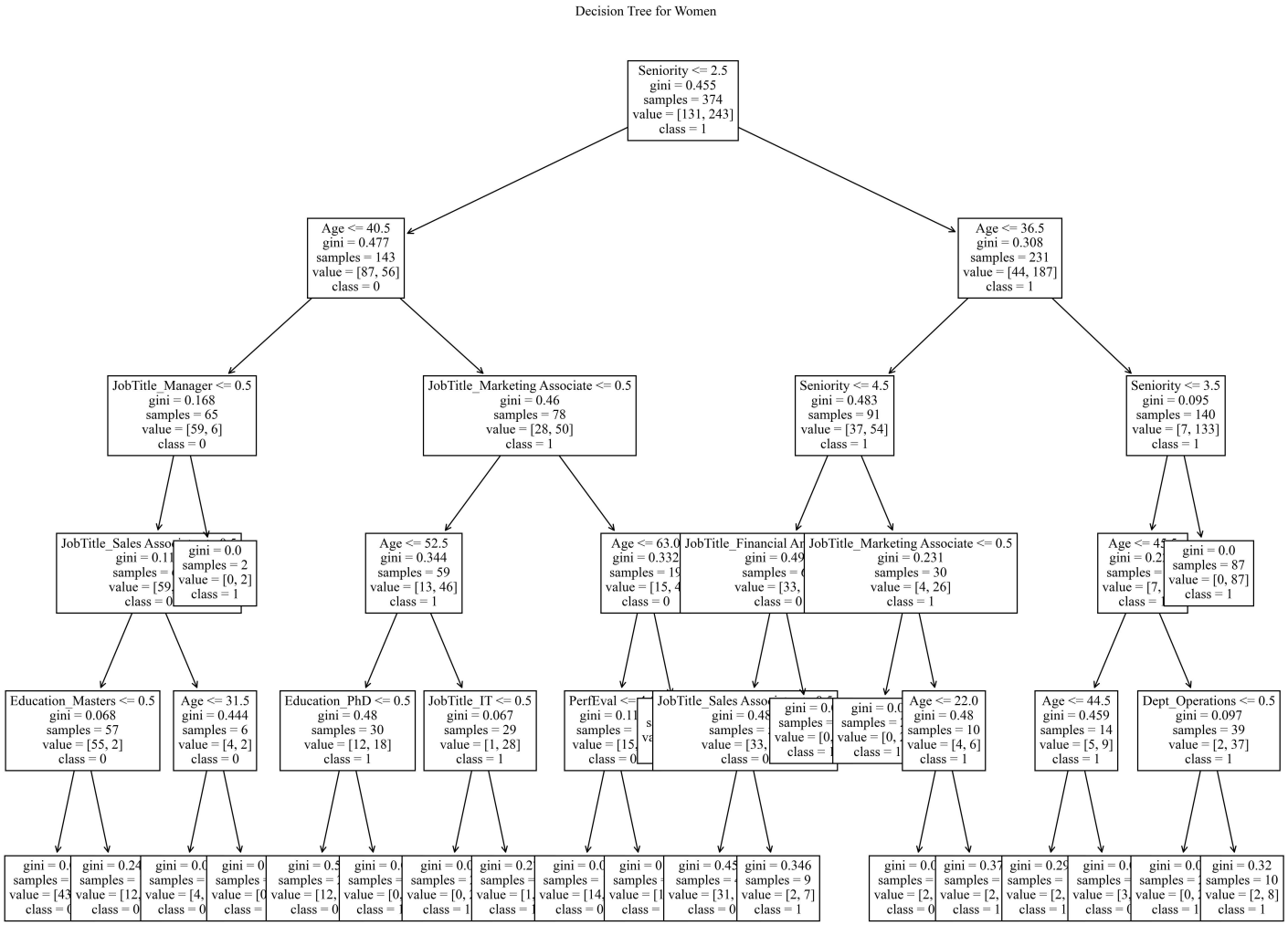

Figure 8: Decision Tree for Women's Data

Evaluation and Interpretation

- Applying the Decision Tree to Switch Gender Data

Model	**Lower Pay**	**Higher Pay**
Tree_women (data_women)	207	325
Tree_men (data_women)	75	457
Tree_women (data_men)	191	277
Tree_men (data_men)	69	399

Table 1: Decision Tree Results

Limitations

- The dataset has limited features.
- Important features such as motherhood and marital status are not included in this dataset.
- Further exploration using non-parametric techniques may enhance the validity of the results.

Conclusion & Future Analysis

- The gender pay gap exists in specific industries and departments.
- Switching the gender variable from men to women in the dataset results in a notable shift of female employees towards higher pay groups and male employees towards lower pay groups.
- Including additional variables such as motherhood, marital status, and pregnancy leave, as highlighted in the limitations, can potentially enhance our comprehension of the wage gap and aid in its reduction.