

TYLER, ERIC, KATIE, JON

BRIDGING THE GAP



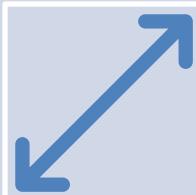
EXPLORING THE PATH TO FAIRNESS FOR ALL

NEXT

Introduction



The gender pay gap remains a significant issue in employment and social economics, affecting industries and professions worldwide. Our project seeks to analyze the current state of the gender pay gap, exploring salary differences across job roles, seniority levels, industries, and regions. We're using datasets from Kaggle and the Census to dive deep into the data and figure out the key factors that contribute to the gap, like education, industry type, or even social norms.



We'll also check out how things have changed over time and try to predict when the gap might close if current trends continue. On top of that, we're going to look at real-world examples of companies and countries that have made progress in closing the gap, like through union representation or pay equity audits. Our goal is to pull all of this together and offer some insights into how things could improve in the future.



Process

BRIDGING THE GAP

WAGE GAP OVERVIEW

IB	IC	ID	IE	IF	IG
region_1	martial status	race_1	sex_1	employee type	union type
Northeast	Never mar	Hispanic	male	Local govt employee	NIU
South	Never mar	Formula Bar	male	Wage/salary, private sector	No union coverage
Northeast	Never mar	white	male	Wage/salary, private sector	No union coverage
West	Never mar	White	male	Wage/salary, private sector	No union coverage
South	Married, spouse present	Other	male	Self-employed	#N/A
West	Married, spouse present	Hispanic	male	State govt employee	NIU
Midwest	Married, spouse present	White	male	Wage/salary, private sector	NIU
South	Married, spouse present	Hispanic	male	Wage/salary, private sector	NIU
South	Married, spouse present	Black	male	Wage/salary, private sector	NIU
Northeast	Married, spouse present	Black	male	Wage/salary, private sector	NIU
Northeast	Married, spouse present	White	male	Wage/salary, private sector	NIU
Northeast	Separated	White	male	Wage/salary, private sector	NIU
West	Never mar	Black	male	Wage/salary, private sector	No union coverage
South	Married, spouse present	White	male	Wage/salary, private sector	NIU
Northeast	Married, spouse present	White	male	Wage/salary, private sector	NIU
South	Never mar	White	male	Local govt employee	No union coverage
Northeast	Married, spouse present	White	male	Wage/salary, private sector	#N/A
Northeast	Married, spouse present	White	male	Wage/salary, private sector	NIU
Northeast	Married, spouse present	White	male	Wage/salary, private sector	No union coverage
West	Married, spouse present	White	male	Local govt employee	NIU
Northeast	Never mar	White	male	Wage/salary, private sector	No union coverage
Midwest	Married, spouse present	White	male	Wage/salary, private sector	No union coverage
West	Married, spouse present	White	male	Wage/salary, private sector	#N/A
Midwest	Divorced	White	male	Wage/salary, government	#N/A
West	Married, spouse present	Hispanic	male	Wage/salary, private sector	NIU
West	Divorced	White	male	Wage/salary, private sector	Member of labor union
West	Married, spouse present	White	male	Wage/salary, private sector	NIU
South	Married, spouse present	White	male	Wage/salary, private sector	NIU

```
from pyspark import SparkFiles
csv_file_paths = [
    "Resources/CurrentPopulationSurvey_1981.csv",
    "Resources/CurrentPopulationSurvey_1990.csv",
    "Resources/CurrentPopulationSurvey_1999.csv",
    "Resources/CurrentPopulationSurvey_2007.csv",
    "Resources/CurrentPopulationSurvey_2009.csv",
    "Resources/CurrentPopulationSurvey_2011.csv",
    "Resources/CurrentPopulationSurvey_2013.csv"
]

dfs = [spark.read.csv(path, sep=",", header=True, ignoreLeadingWhiteSpace=True) for path in csv_file_paths]

combined_df = dfs[0]
for df in dfs[1:]:
    combined_df = combined_df.union(df)

# Show DataFrame
combined_df.show(5)
✓ 5.0s
```

We began with a dataset containing over 300,000 rows and more than 200 columns. To make the data more manageable, we first performed a VLOOKUP to replace numerical codes with descriptive labels for certain columns. After identifying and removing unnecessary columns, we used PySpark to process the data efficiently. Due to the large file sizes, we split the dataset by year to ensure it could be imported into Git. The final datasets were saved and imported as CSV files.

CLEANING THE DATA PROCESS

NEXT

MILESTONES IN THE *Journey* TOWARD THE GENDER PAY GAP

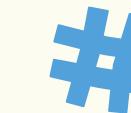


THE EQUAL PAY ACT 1963



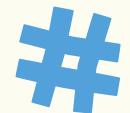
is passed by Congress, making it illegal for employers to pay women lower wages than men for work requiring the same skill, effort, and responsibility. Employees suspecting pay discrimination have legal routes for recourse, including contacting an Equal Employment Opportunity counselor or directly filing a lawsuit.

CIVIL RIGHTS ACT 1964



Title VII safeguards employees and job seekers from workplace discrimination or unfair wage gaps due to race, color, religion, sex, or national origin. It applies to all employment-related activities, from hiring and firing to promotions and working conditions, ensuring equal treatment in every aspect of employment.

RIGHT TO VOTE 1920



The 19th Amendment, ratified in 1920, marked a pivotal moment in American history by granting women the right to vote across all states and in federal elections. Given that the fight for suffrage began in the 1800s, few of the original supporters lived to see the passage of the 19th Amendment. Throughout the struggle, suffragists faced harsh opposition, including heckling, physical assaults, and imprisonment. New York State's adoption of women's suffrage in 1917 and President Wilson's eventual support were milestones.

HISTORICAL CONTEXT

NEXT

Gender Pay

BRIDGING THE GAP

Across Industries

Male-to-Female Ratio (Top 10 industries based on Pay) =

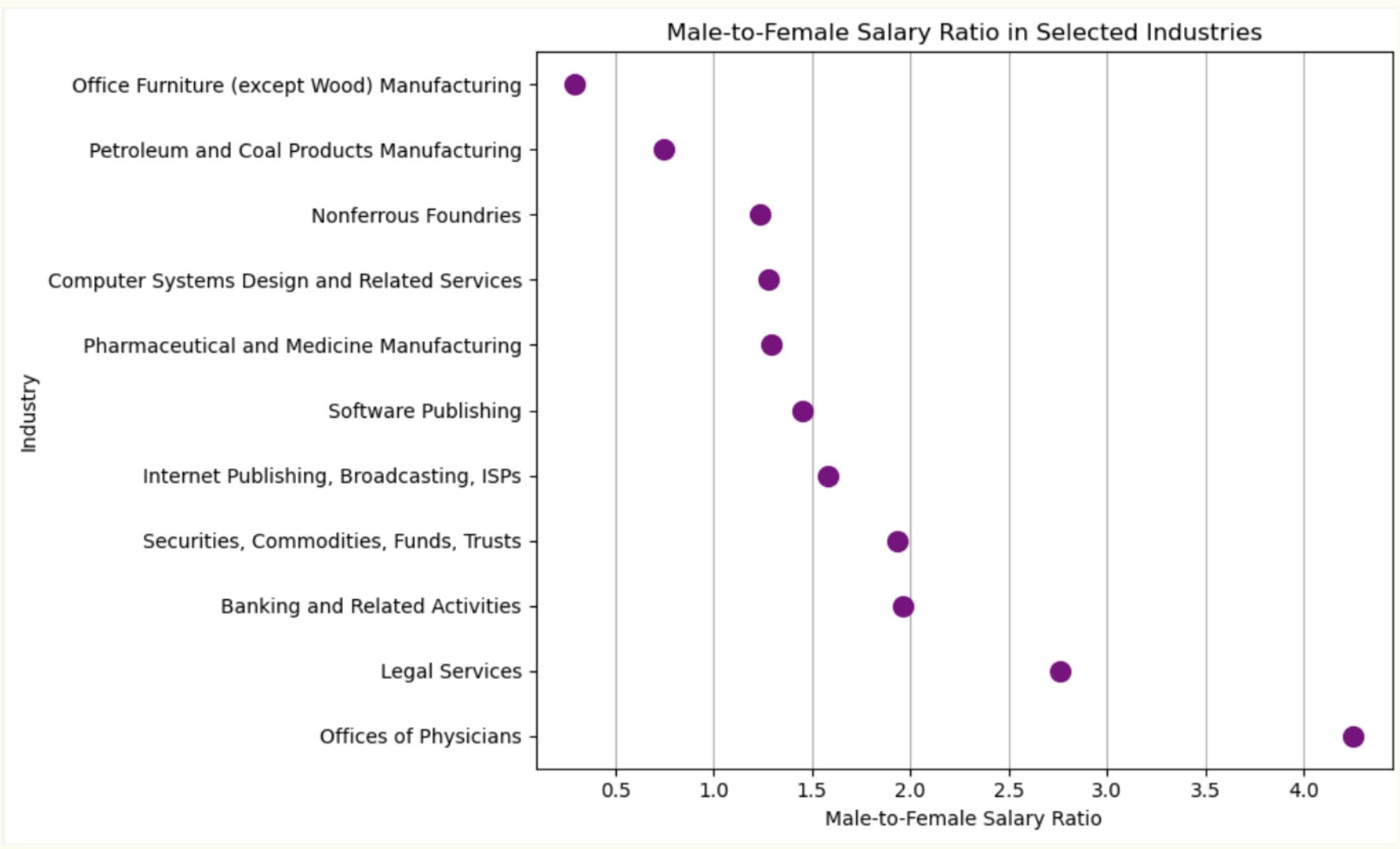
Average Salary of Females / Average Salary of Males

A ratio of 1 means that males and females earn the same average salary.

A ratio greater than 1 indicates that males earn more than females on average.

A ratio less than 1 indicates that females earn more than males on average.

Office Furniture and Gas/Coal Products Manufacturing are the only industries that females are paid more than male.



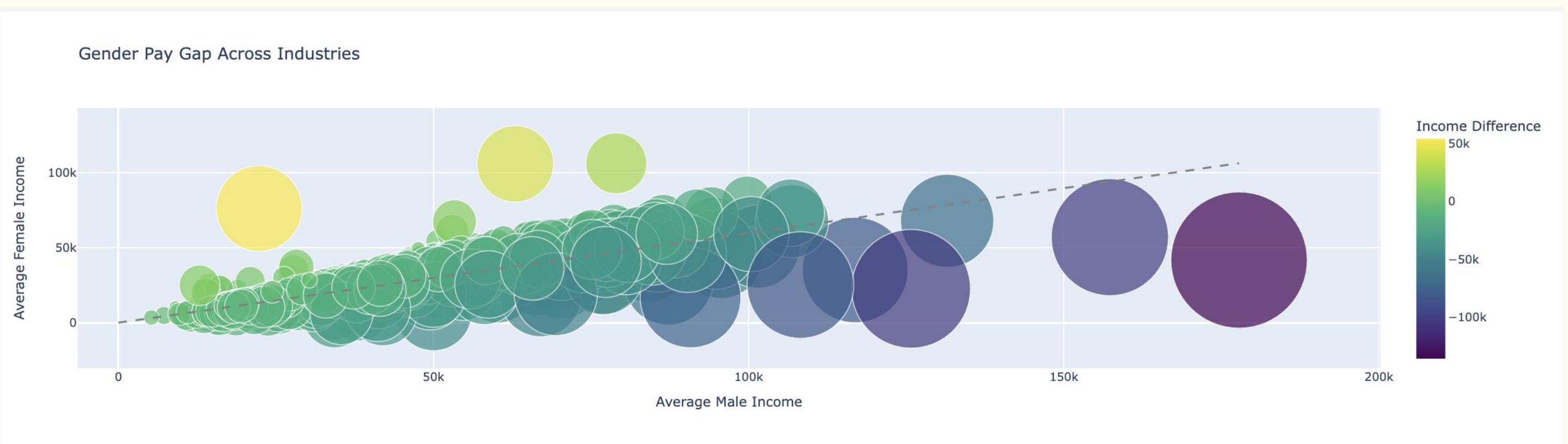
Industry Gaps

NEXT

Gender Pay Across Industries

BRIDGING THE GAP

Gender Pay Gap Across Industries

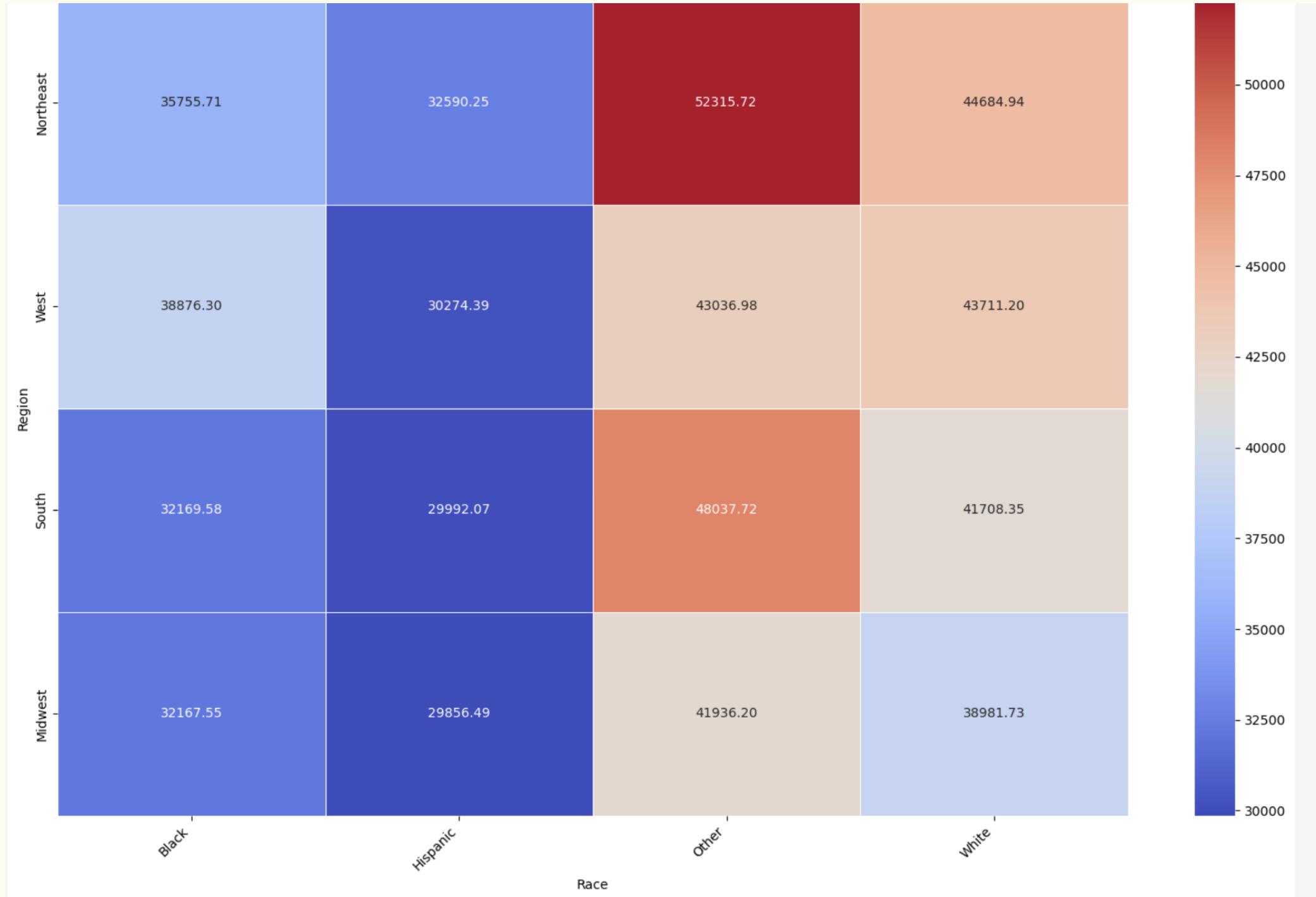


Industry Gaps

NEXT

Pay Across Region & Race

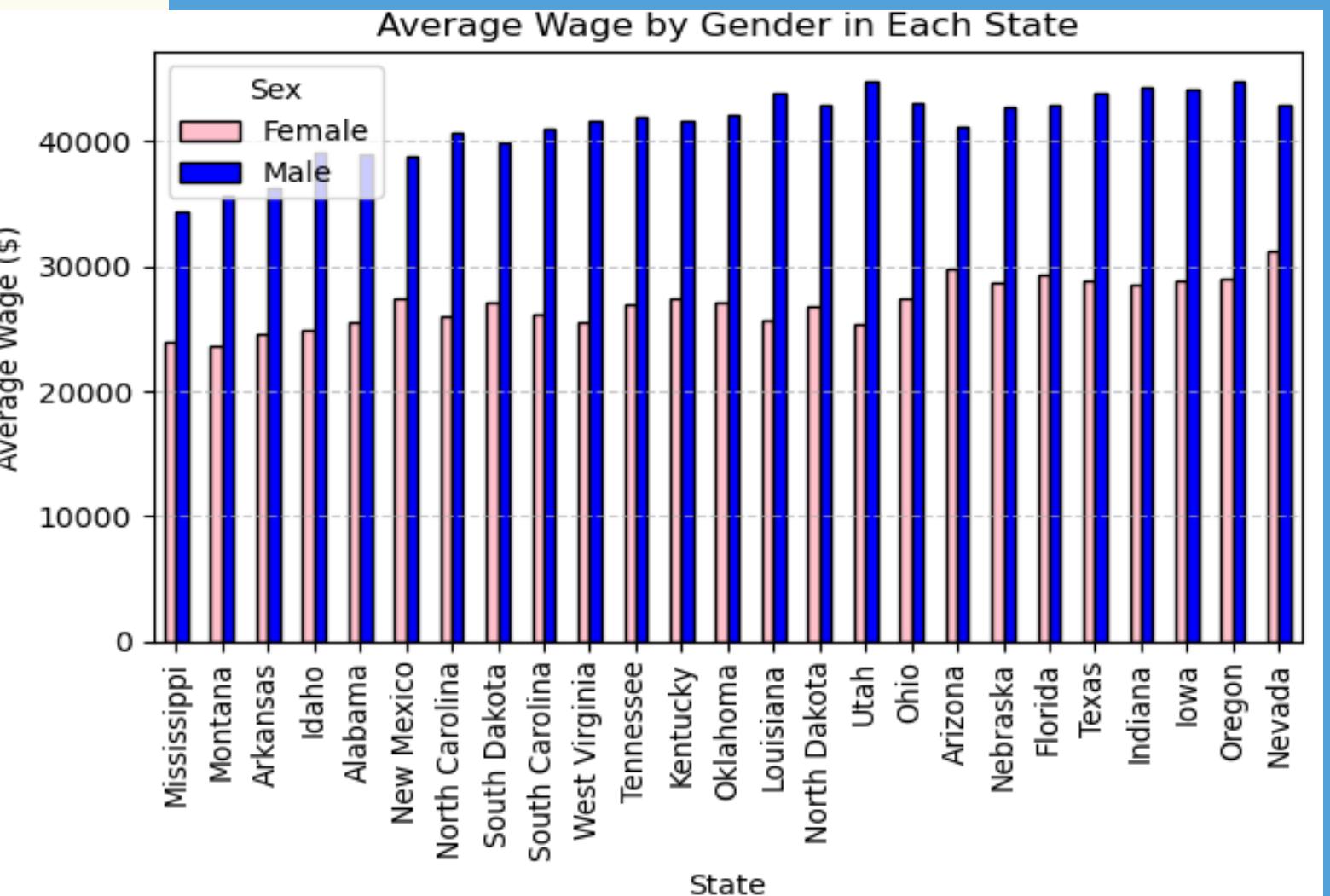
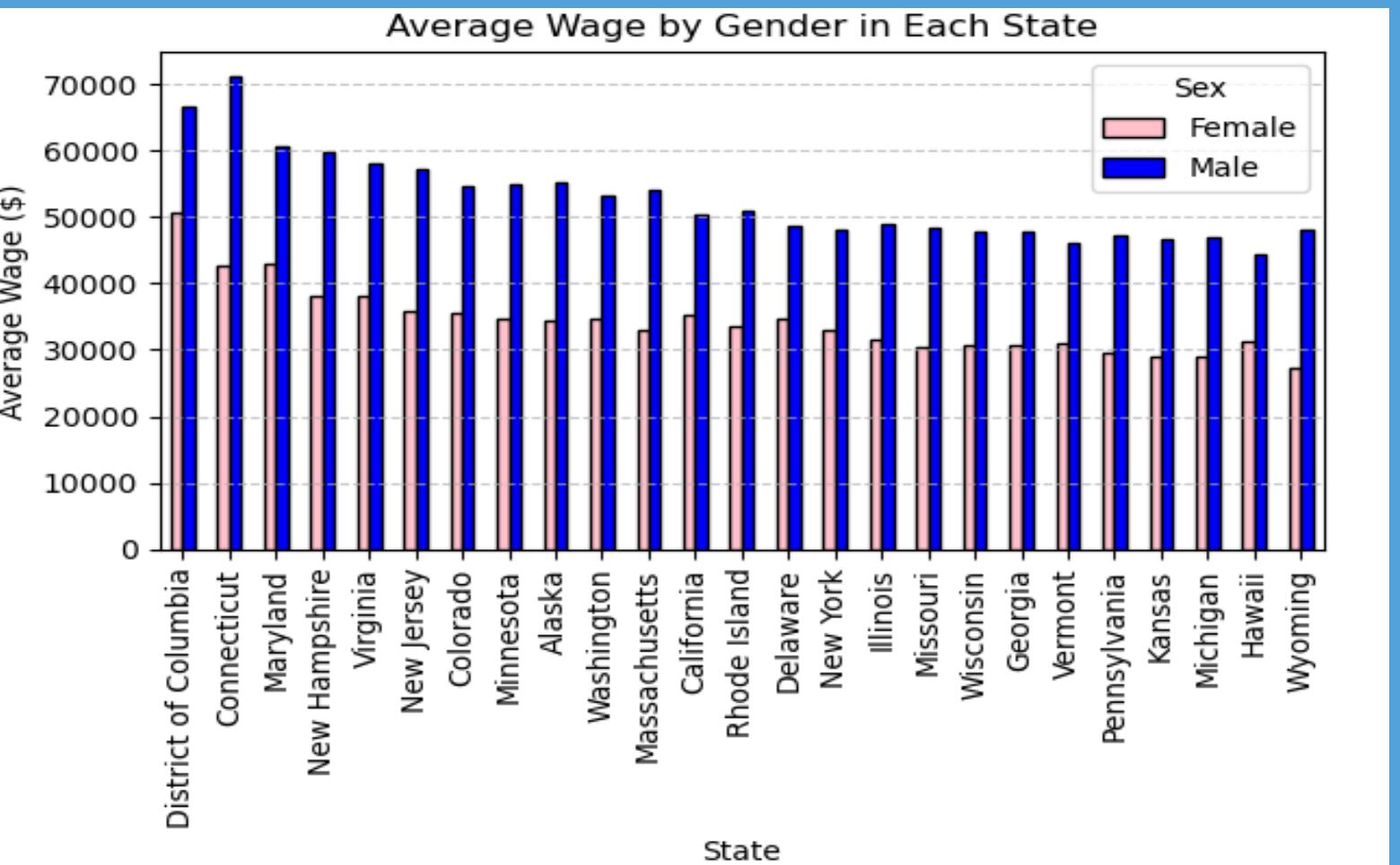
- Northeast tends to pay more in salary on average
- Other and White get paid more
- Hispanics out of all races get paid the least



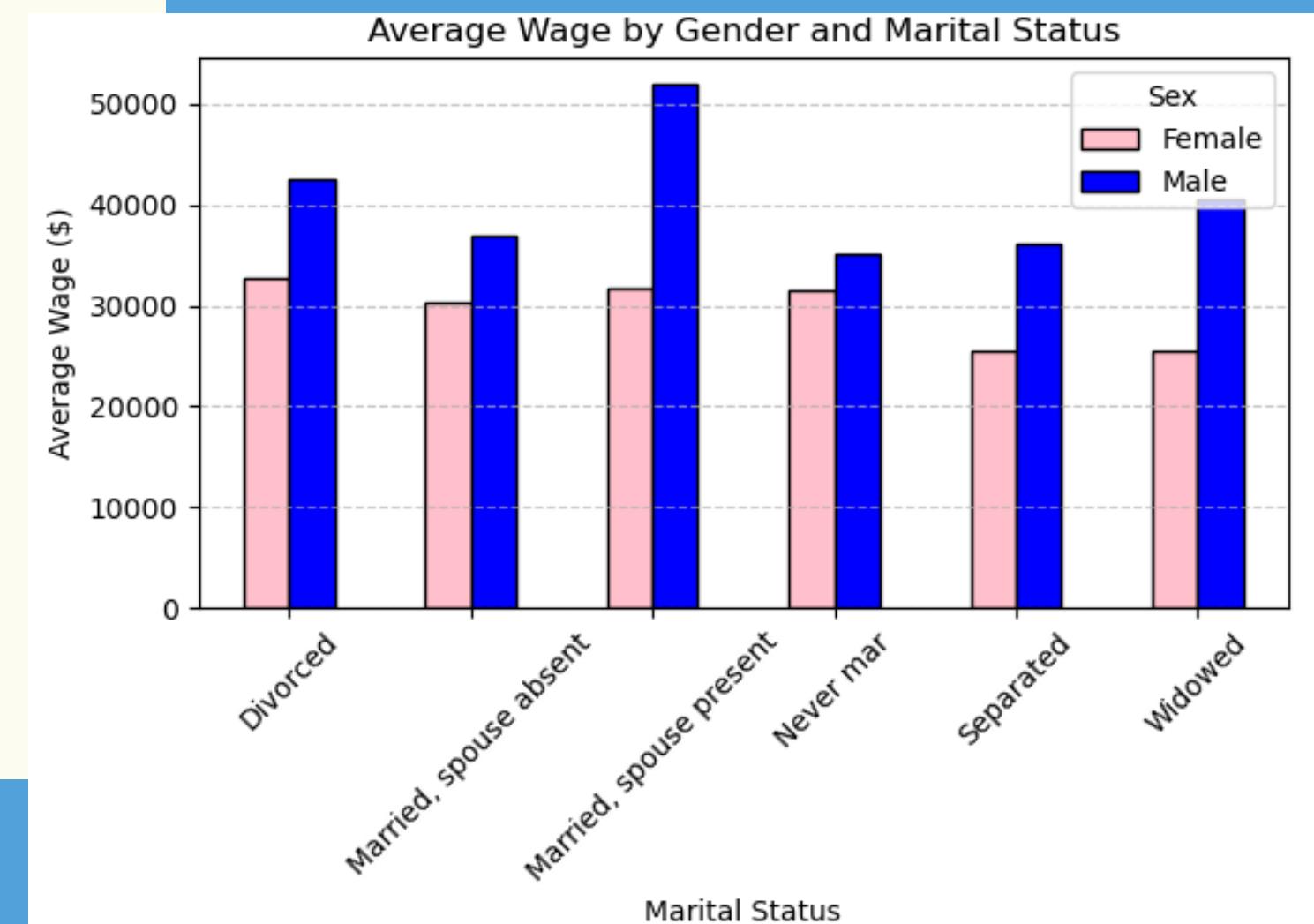
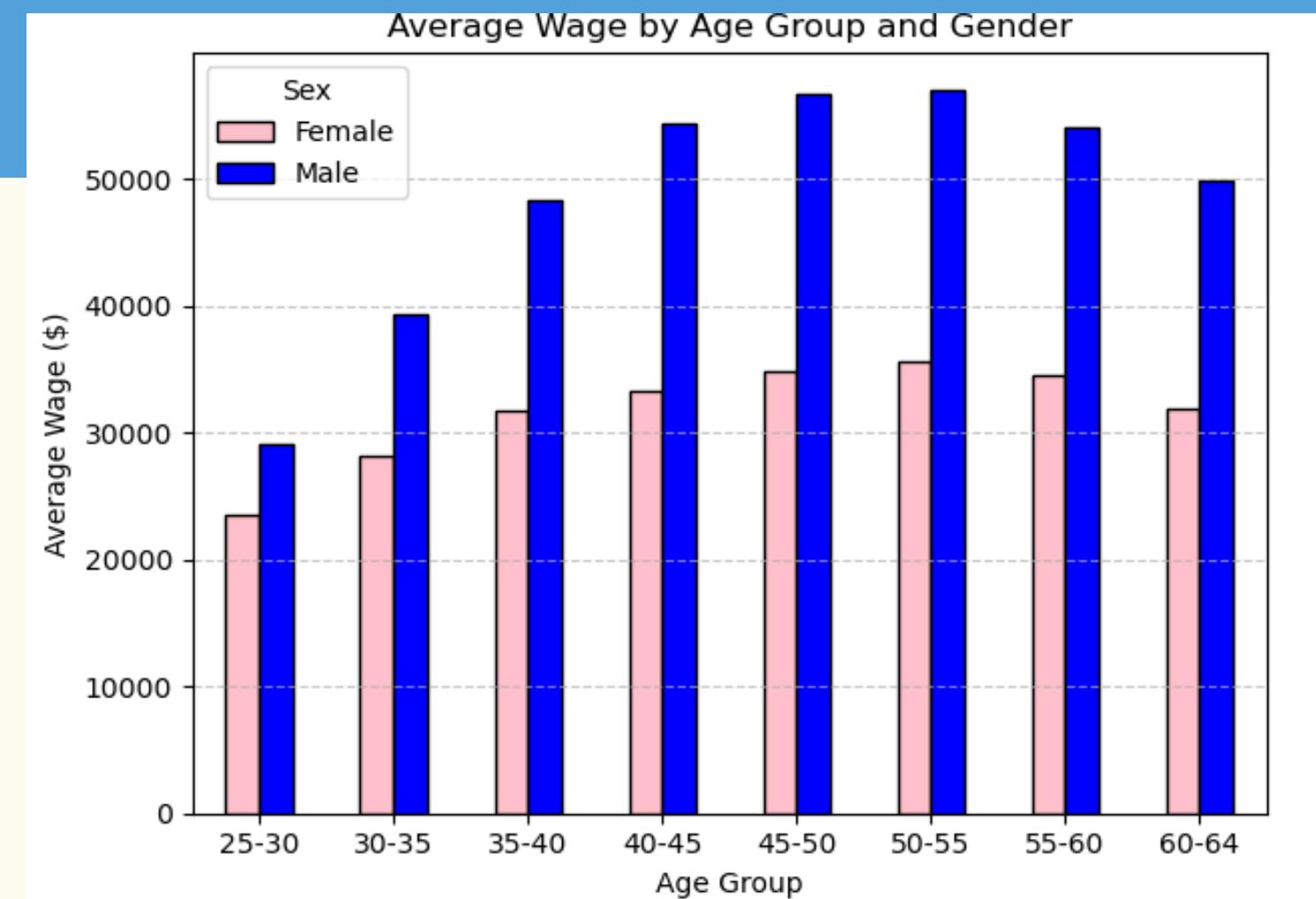
Region Gaps

NEXT

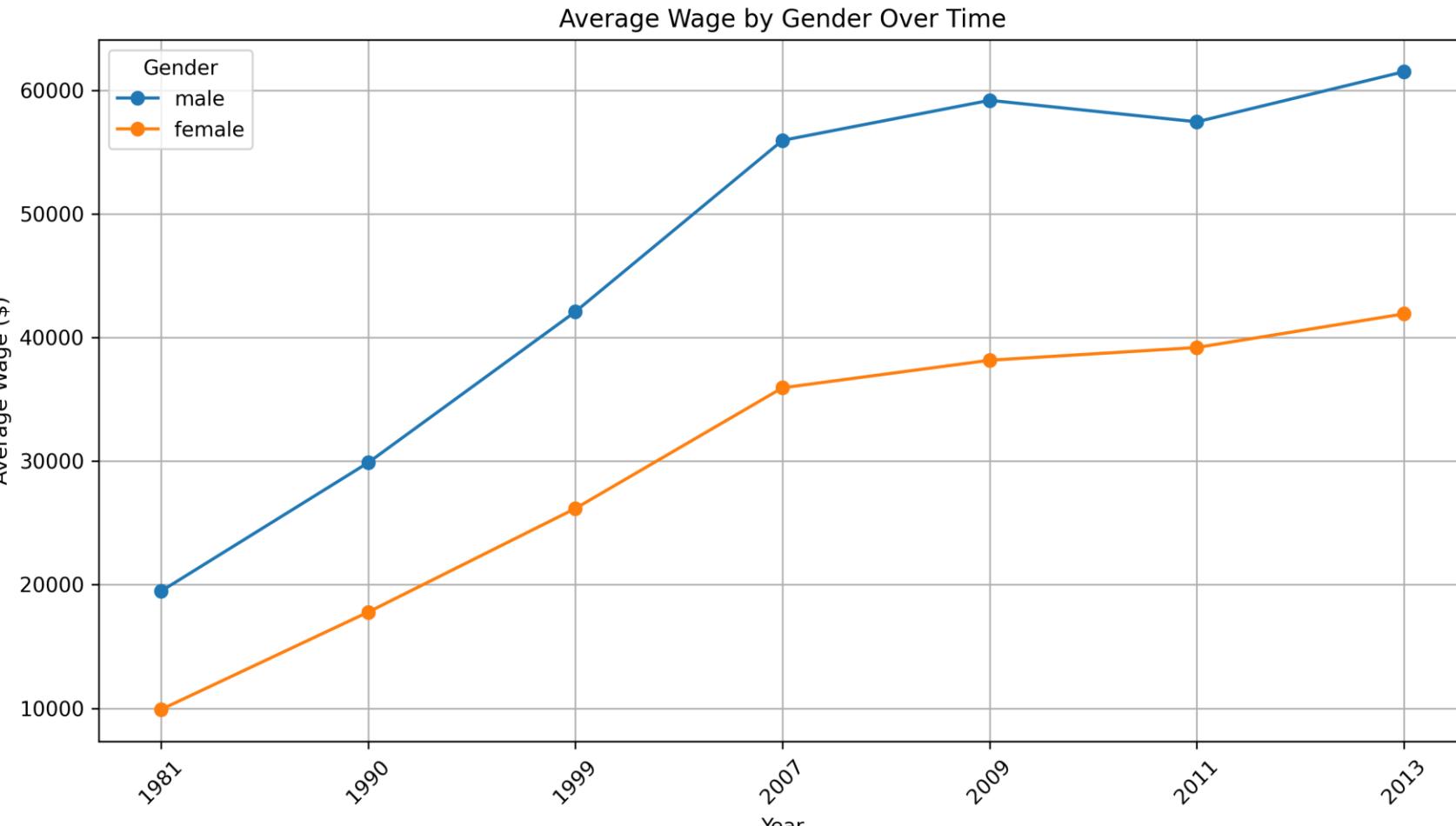
FACTORS CONTRIBUTING TO THE GENDER PAY GAP: ANALYZING BY STATE



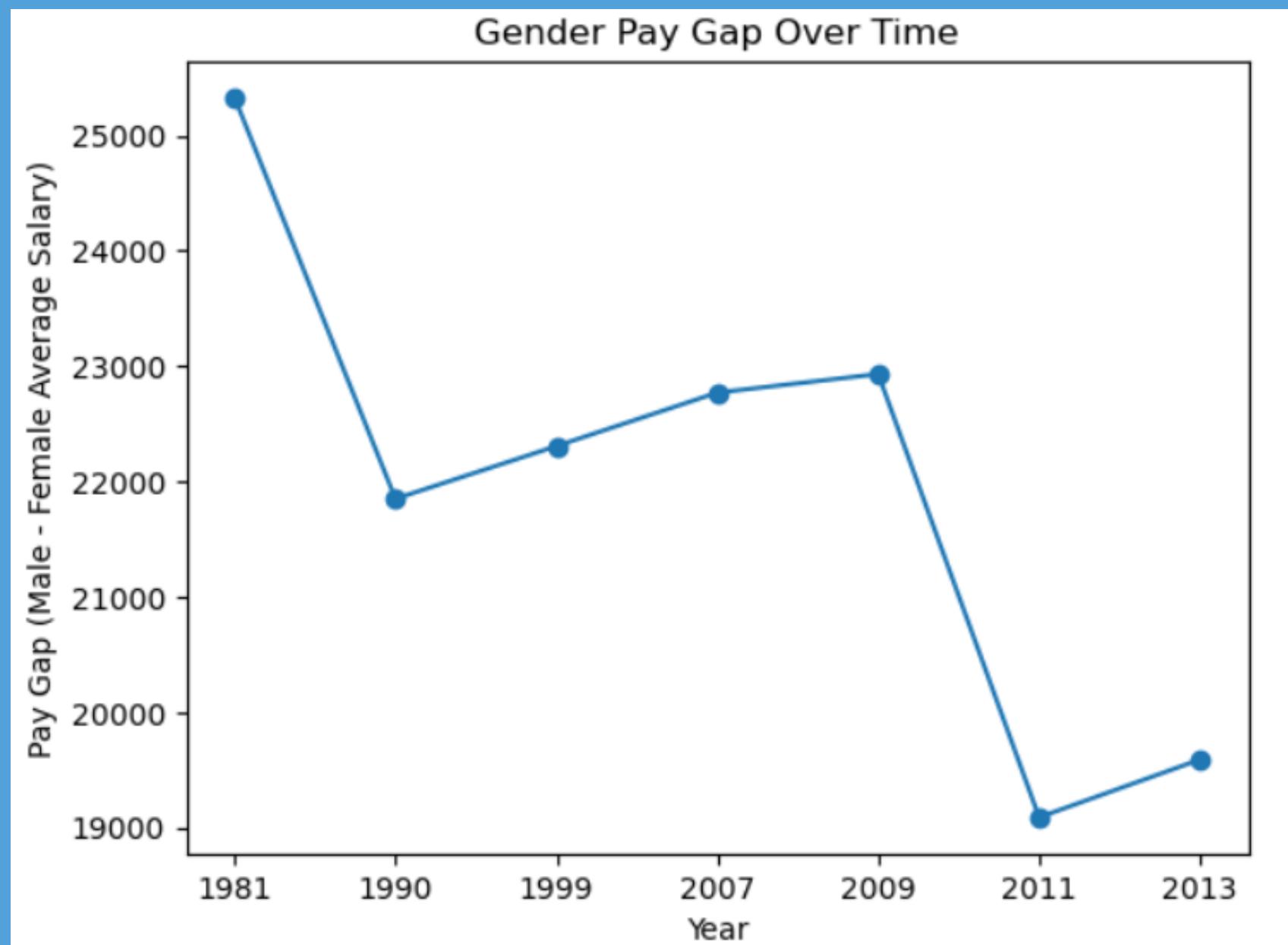
FACTORS CONTRIBUTING TO THE GENDER PAY GAP: ANALYZING BY AGE AND MARITAL STATUS



WAGE GAP OVER TIME: USD (\$) AND PERCENTAGE (%) DIFFERENCE

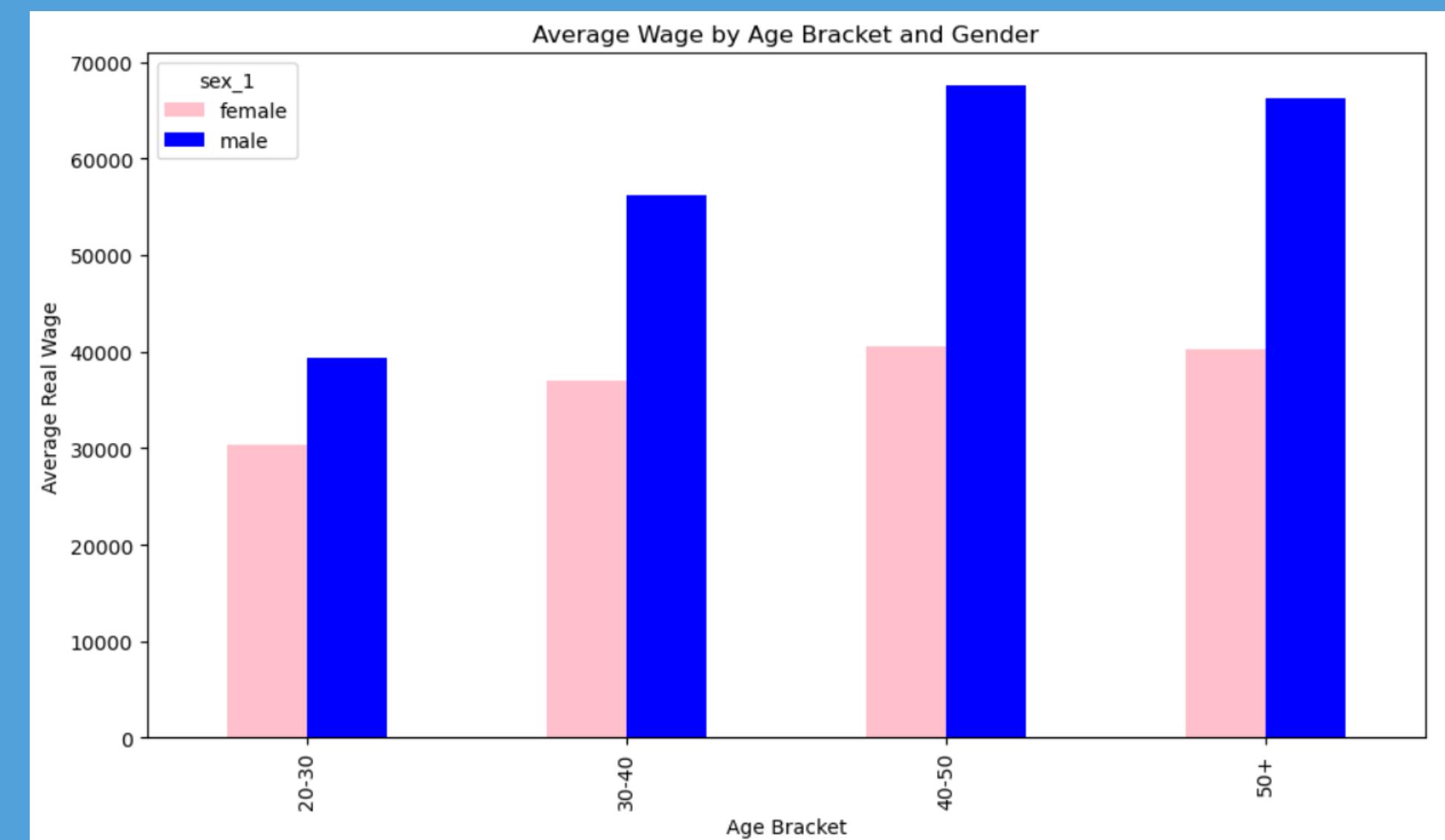


HISTORICAL TRENDS AND CAREER STAGES



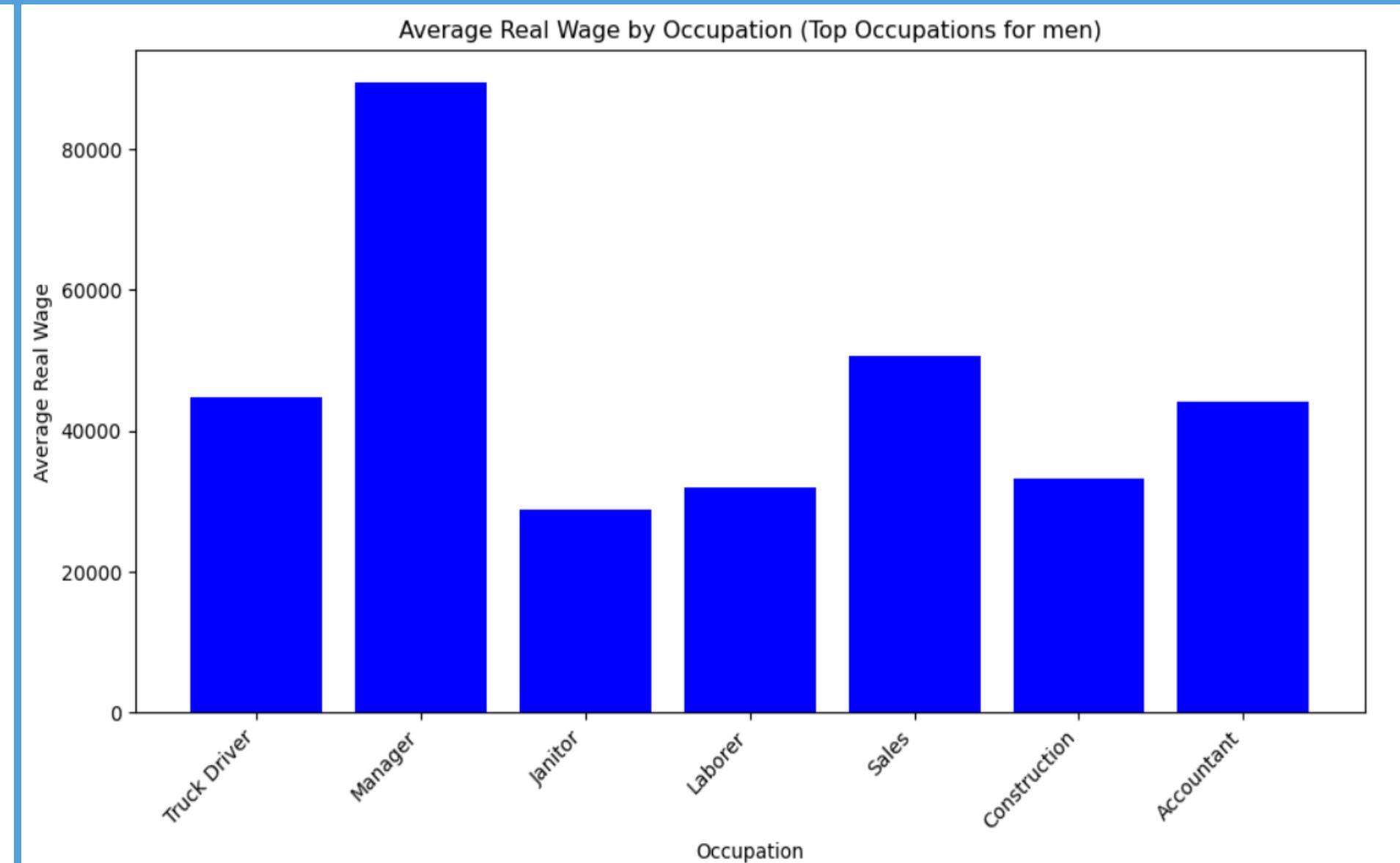
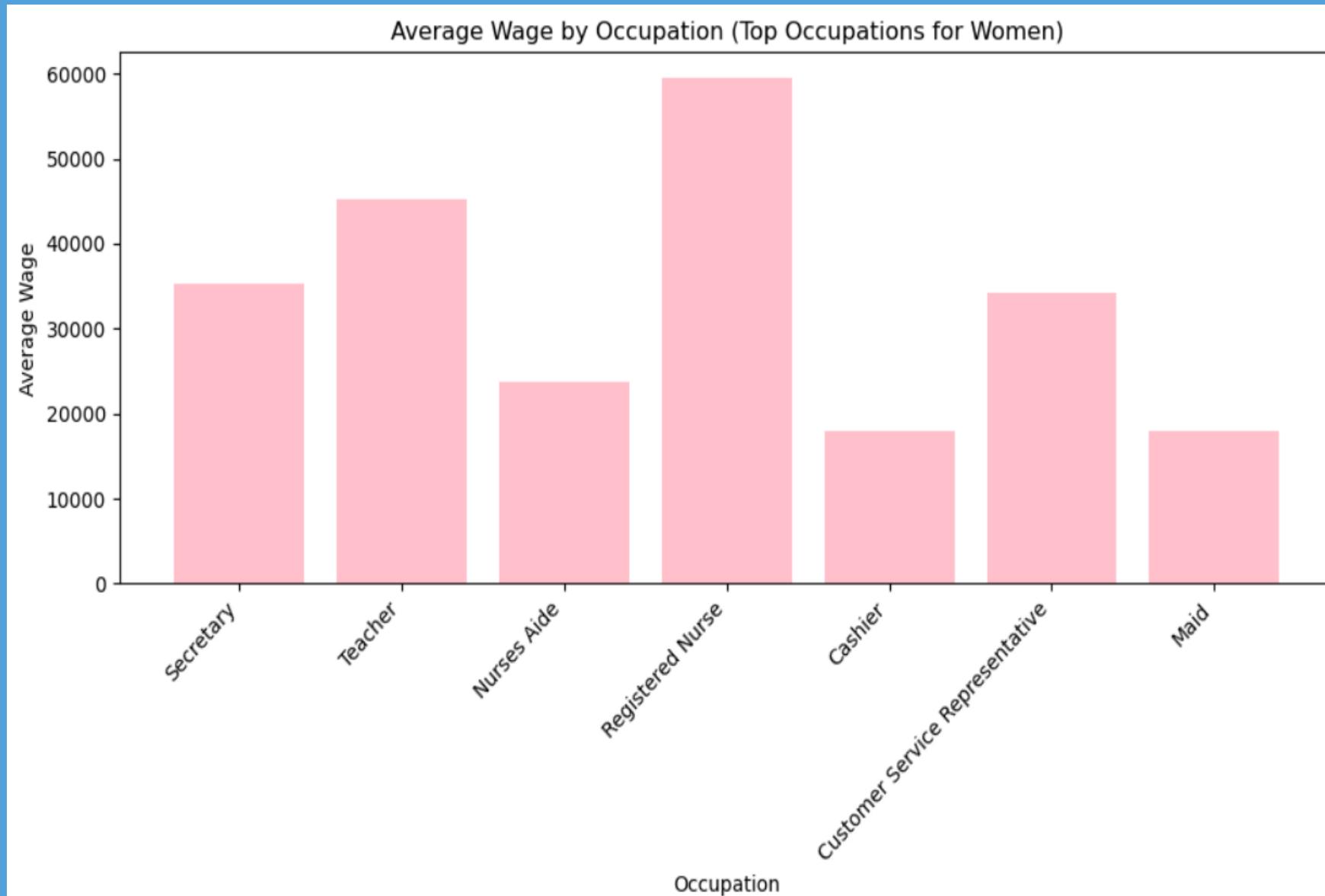
Despite the general downward trend, the pay gap is still substantial by 2013, indicating ongoing gender disparity in wages.

The gender pay gap seems to increase with age, indicating that women earn comparatively less than men as they progress in their careers.



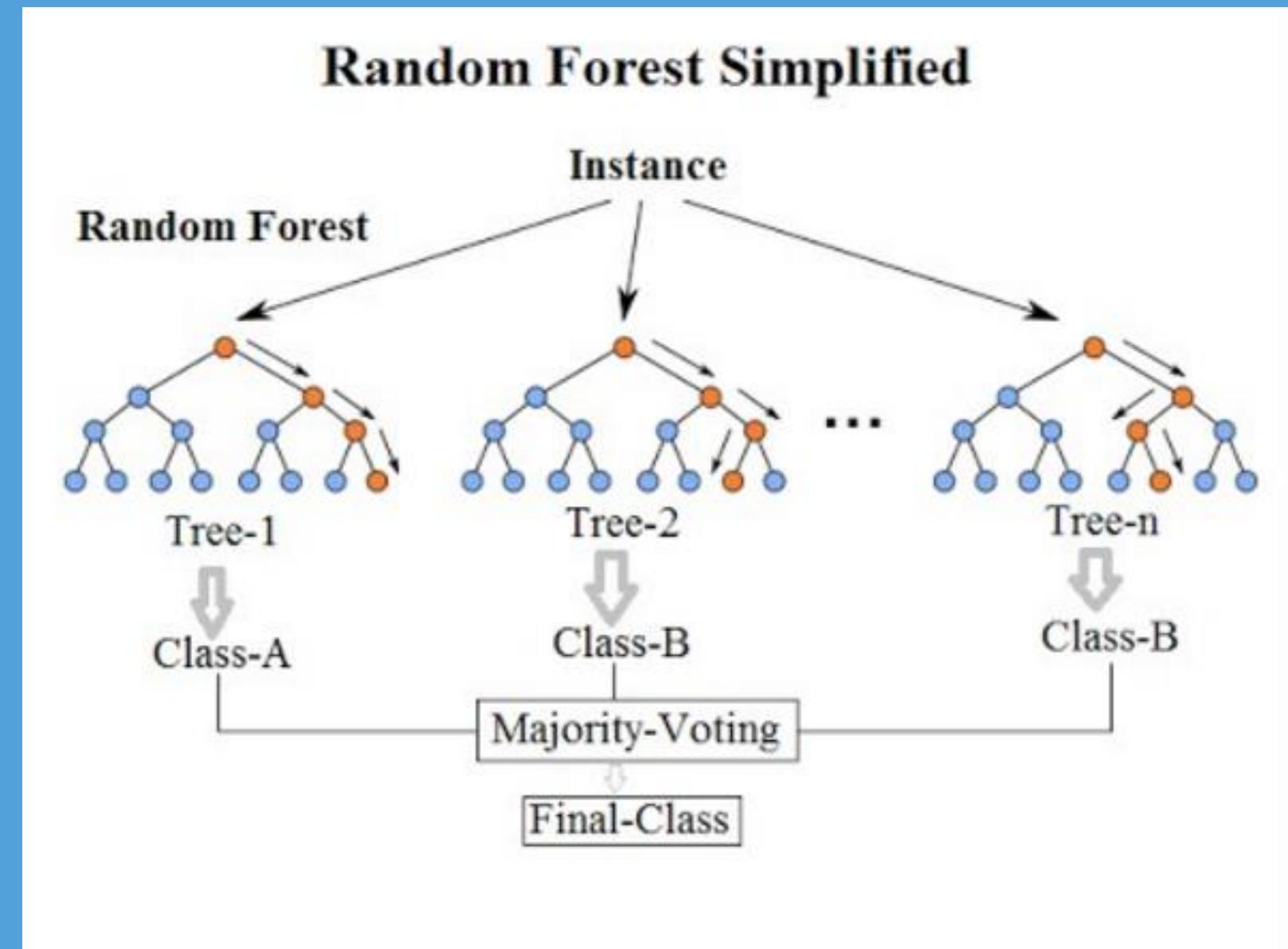
OCCUPATIONAL SEGREGATION

Women are overrepresented in lower-paid roles such as service, caregiving, and support positions, whereas men dominate higher-paid technical and managerial roles.



Machine Learning Application

- Model goal: analyze and predict the gender gap %
- Model type: regression
- Model chosen: random forest regressor
- Benefits to this model:
 - Ensemble learning
 - Bootstrap sampling (sampling with replacement)
 - Random feature selection
 - Averaging to final prediction
- Downsides of other model options:
 - Linear regressions sensitivity to outliers
 - Overfitting risks of decision tree modeling

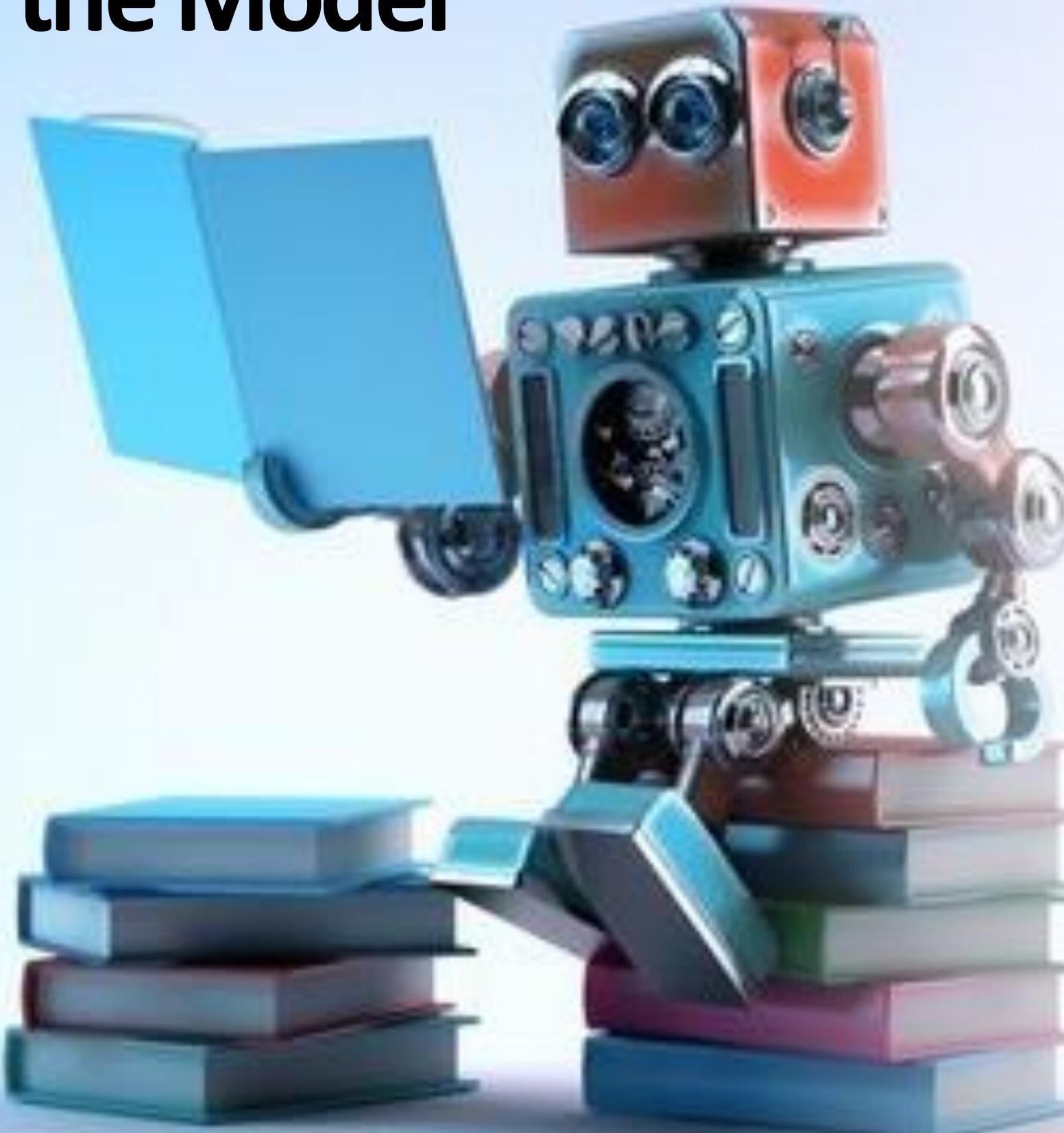


Building the Model



- Data transformation:
 - Identification of features:
 - Input features: year, state, race, age, union type, marital status, and sex
 - Calculation of average income as grouped by input features,
 - Pivot data for separated male and female columns
- Pay gap percentage calculation:
 - Difference between male and female averages for each feature grouping
- Cleaning and model preparation:
 - Checking for NaN values, dtype conversion
 - One-hot-encoding categorial features
 - State, race, union type, marital status

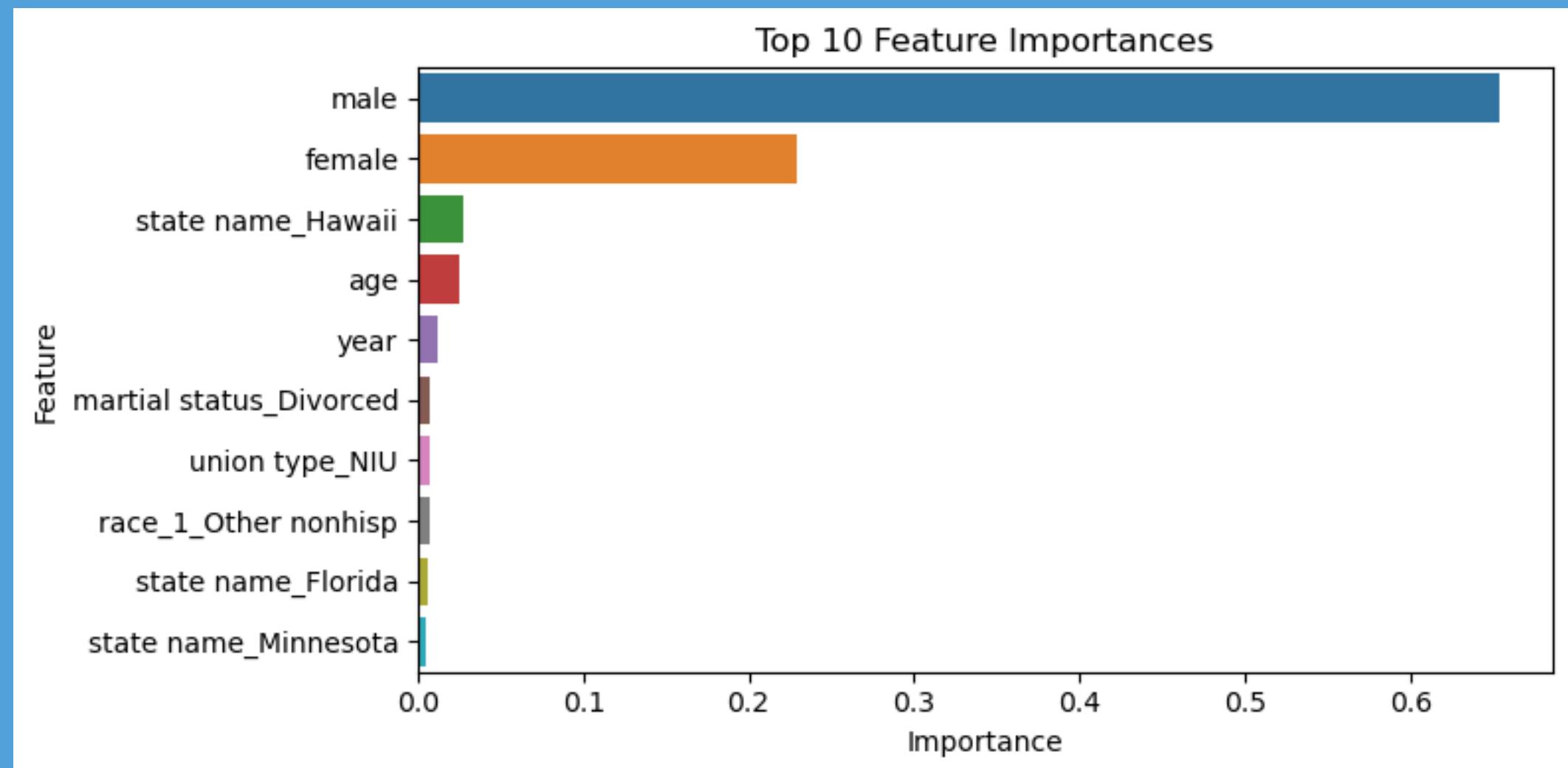
Building the Model



- Split and scale the data:
 - Separation of target vector (pay gap) and feature matrix
 - Train, test, split for separated training and testing sets
 - Use of standard scaler to standardize values for comparison
- Initialize model and fit to data:
 - Initialize random forest regressor
 - Fit model to training data set
- Make predictions:
 - Use model to predict the pay gap on the test set

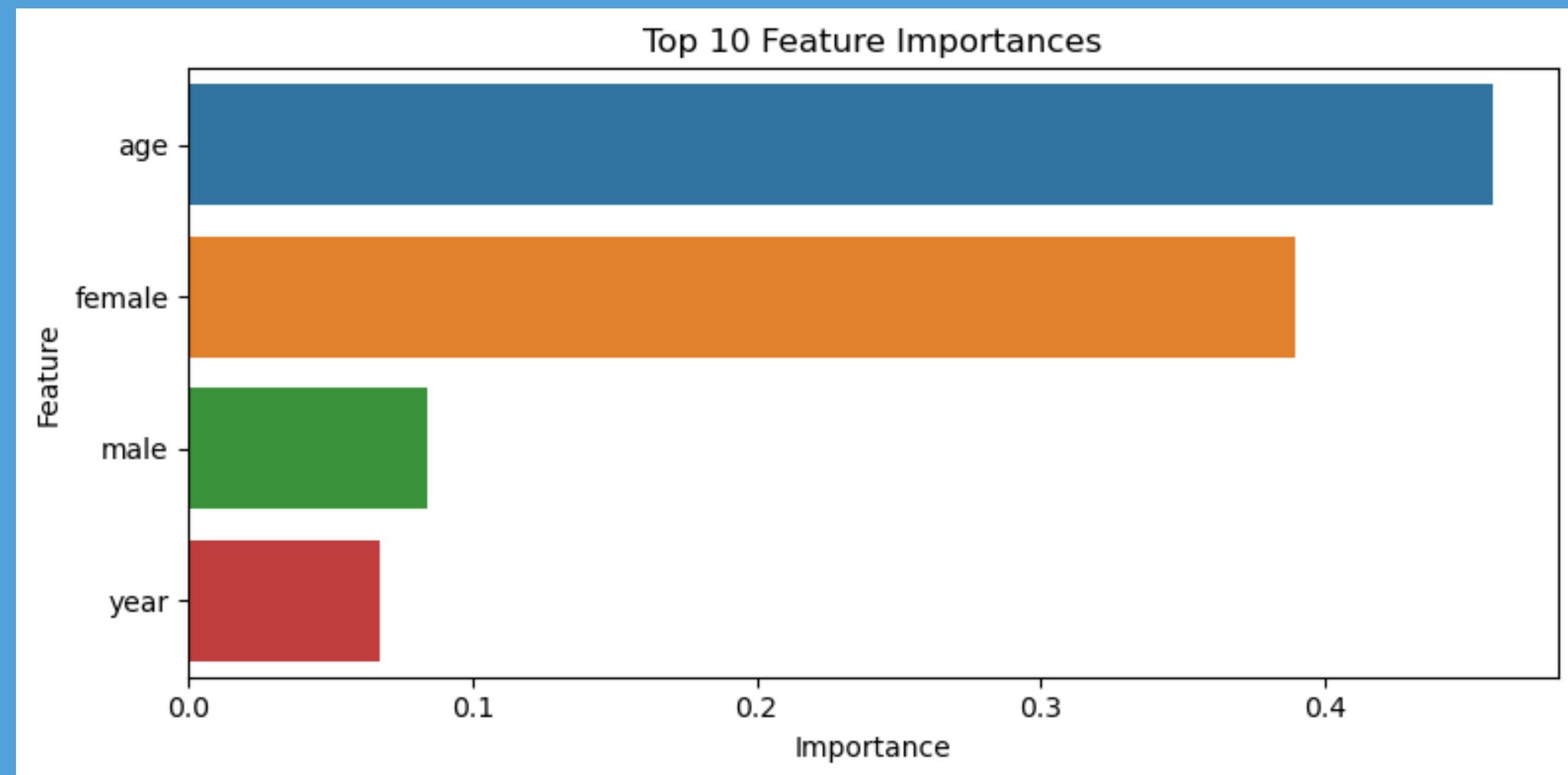
Evaluating the Model

- Model evaluation included review of the output R^2 and MSE (Mean Squared Error) values
- First model iteration:
 - $R^2 = 0.8439$
 - $MSE = 10150.0645$
- Room for improvement!
- Feature importance Analysis:
 - Method of extracting and sorting features by impact on the model
 - Helped identify features possibly hindering the model's performance



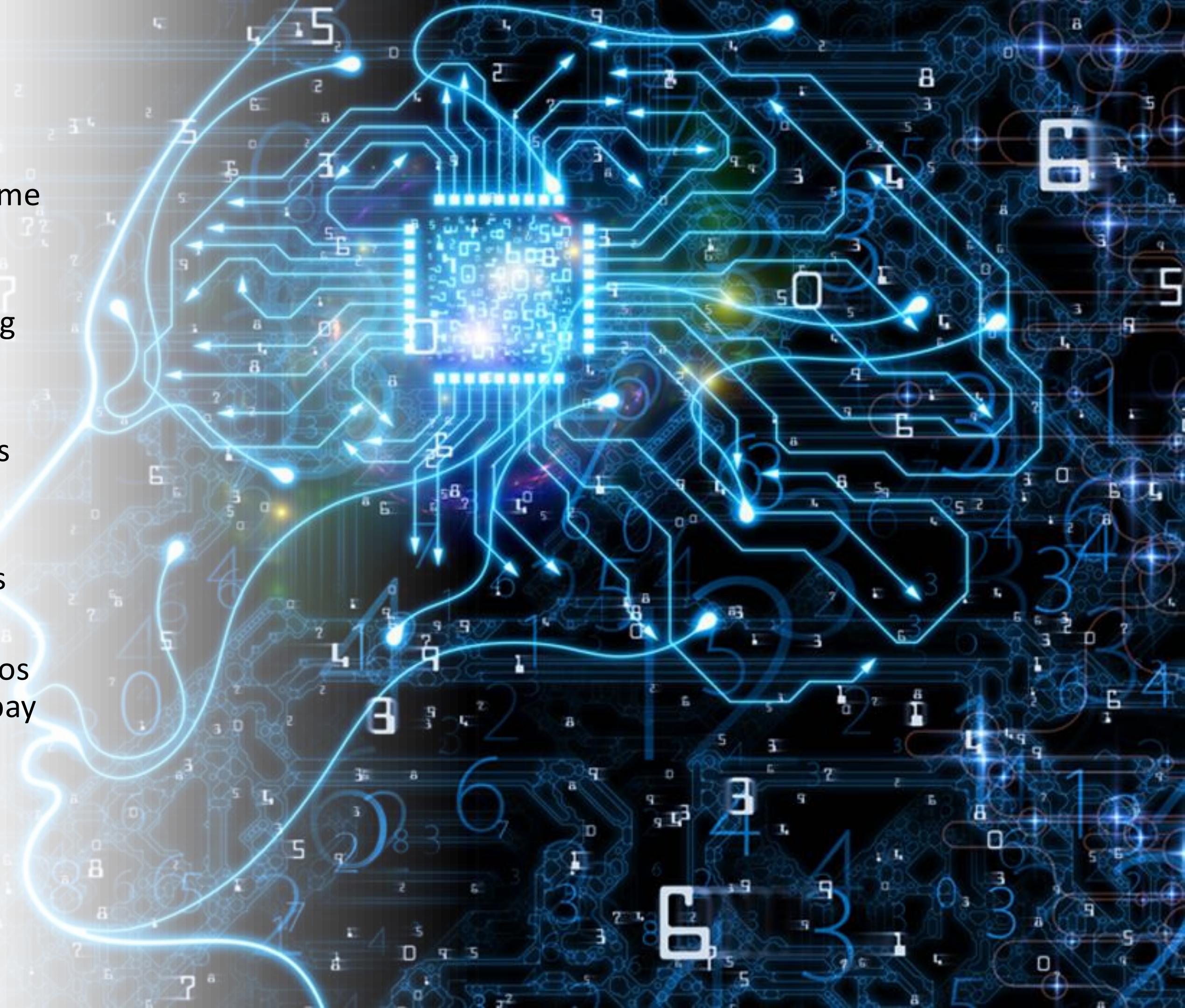
Evaluating the Model

- Model steps repeated with reduced features
 - Sex, age, and year
- Improvement was made in both scores:
 - $R^2 = 0.92405$
 - $MSE = 8.0346$
- Focusing on these inputted features improved model accuracy and reduced prediction error
- This achieves our goal of creating a model that can accurately predict pay gap



Using the Model

- Observation of pay gap trends over time
- Evaluation of policy changes through measurement of the pay gap following policy implementation
- Draw insight on demographic features that may affect gender pay gap
- Identify the pay gap of specific groups
- Experiment with hypothetical scenarios that would allow us to see potential pay gap effects



TRANSPARENCY AND PAY EQUITY AUDITS



Research shows that gender-based pay differences often cannot be fully explained by objective factors like occupational segregation, work experience, or education. Unexplained components of the wage gap are linked to biases, stereotypes, and lack of transparency in pay-setting. Studies indicate that mandated disclosure of company-level pay data is an effective policy for reducing this gap. Key reasons why transparency matters include:

- 1. Awareness and Accountability:** Open pay practices increase awareness of disparities, compelling organizations to justify unexplained differences.
- 2. Identification of Inequities:** Mandated reports prompt employers to analyze pay structures and collect standardized data, uncovering previously ignored inequities.
- 3. Competitive Pressure:** Public disclosure creates incentives for companies to improve their standings relative to industry peers to avoid criticism over unexplained pay gaps.
- 4. Bargaining Power:** Transparency enhances employees' ability to negotiate fair wages, especially for women who may lack access to pay information.

SUCCESS Stories

Several countries have introduced legislation requiring private sector employers to submit annual reports on compensation, with notable examples:

- **Iceland:** Companies with 25+ employees must obtain government certification of equal pay every three years, facing fines if they fail.
- **United Kingdom:** The Equality Act 2010 mandates private firms to report their overall gender pay gap annually.
- **Australia:** The Workplace Gender Equality Act requires non-public sector agencies with 100+ employees to disclose gaps in salary by gender.



GENDER EQUALITY

Notable companies have embraced pay transparency through audits and open reporting:

- **Salesforce:** Conducts annual pay audits and has spent over \$16 million to correct disparities.
- **Buffer:** Publicly shares all employee salaries, fostering trust and equity.
- **Adobe:** Regular audits led to \$1 million in salary adjustments to achieve pay equity.
- **LinkedIn:** Conducted audits resulting in the closure of the global gender pay gap.

US Pay Reporting Requirements



In the U.S., some federal and state-level pay reporting requirements exist:

- **EEOC Reporting:** Employers with 100+ employees submit EEO-1 reports on employee demographics, indirectly contributing to pay equity analysis.
- **State Regulations:** States like California and Colorado have stringent laws requiring pay data reporting, including median and mean wages by gender and ethnicity.
- **Local Ordinances:** Cities like New York have laws mandating salary range disclosures in job postings, promoting transparency.
- **Proposed Legislation:** The Paycheck Fairness Act aims to strengthen reporting requirements but has not yet passed.

RECAP OF KEY POINTS



Although progress has been made in narrowing gender pay differences, significant gaps persist due to various factors. Mandatory public reporting of pay data has proven effective in addressing inequities by encouraging accountability and competition among firms. Organizations can adopt practices similar to successful reporting requirements—such as conducting pay audits, enhancing transparency, training leaders, and committing to ongoing evaluations—to promote pay equity.

Thank you!

CONCLUSION

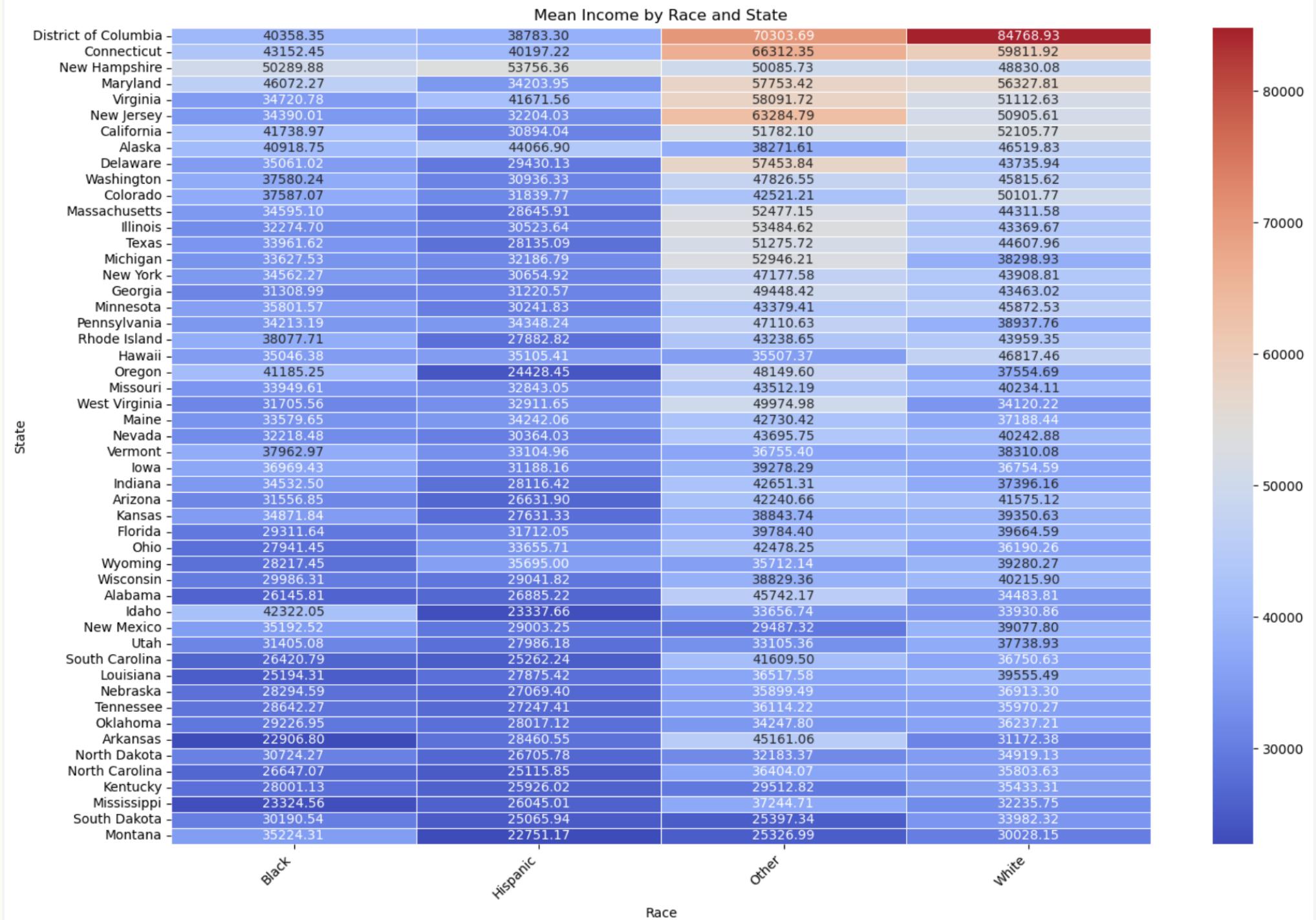
Questions?

Links References

- Blau, F. D., & Kahn, L. M. (2017). *The gender wage gap: Extent, trends, and explanations*. Journal of Economic Literature, 55(3), 789–865.
- Goldin, C. (2014). *A grand gender convergence: Its last chapter*. American Economic Review, 104(4), 1091–1119.
- Gupta, N. D., Poulsen, A., & Villeval, M. C. (2018). *Gender matching and competitiveness: Experimental evidence*. Economic Inquiry, 58(1), 162–178.
- Levanon, A., England, P., & Allison, P. (2009). *Occupational feminization and pay: Assessing causal dynamics using 1950–2000 U.S. Census data*. Social Forces, 88(2), 865–891.
- Mandel, H., & Semyonov, M. (2014). *Gender pay gap and employment sector: Sources of earnings disparities in the United States, 1970-2010*. Demography, 51(5), 1597–1618.
- O'Neill, J., & O'Neill, D. M. (2012). *What do wage differentials tell us about labor market discrimination?* In Handbook of labor economics (Vol. 3, pp. 793–862).
- Weichselbaumer, D., & Winter-Ebmer, R. (2005). *A meta-analysis of the international gender wage gap*. Journal of Economic Surveys, 19(3), 479–511.
- World Economic Forum. (2024). *Gender Gap Accelerators: Benefits Over 850,000 Women*.
- Psico-Smart. (2024). *Case Studies: Successful Implementation of Salary Equity Strategies in Various Industries*.

BRIDGING THE GAP

Pay Across Race & State



State Gaps

NEXT