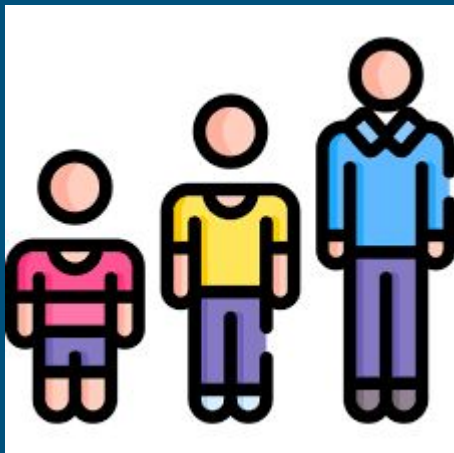# Understanding Pay Gaps Across Gender: Analysis of Income Over Two Decades

Yuchen Li, Tiancheng Yang, Yiming Wang, Qianyi Mo

Income Gap?

Problem Definition

➢ Context & Significance
  ○ Ongoing societal debate: Do gender-based pay gaps exist?
  ○ Influence of age, education, occupation, and experience on pay equity.

➢ Research Questions
  ○ Are there substantial pay differences between genders when controlling for other factors?
  ○ Do these gaps vary across different demographic groups?

➢ Data Source & Methodology
  ○ Utilizing Panel Study of Income Dynamics data (1980-2010).
  ○ Employing time-series analysis and predictive regression for in-depth insights.

➢ Objective & Approach
  ○ Aiming to understand and forecast the evolution of gender pay gaps.
  ○ Integrating diverse factors for a comprehensive analysis.

Stakeholders

➢ Women in the Workforce
   ○ Directly impacted by pay gaps.
   ○ Inequitable pay and career advancement opportunities.
➢ Employers and Corporations
   ○ Need to ensure fair pay practices.
   ○ Diversity, equity, and inclusion in the workplace.
➢ Academic and Research Community
   ○ Contribute to ongoing research and discourse on gender pay gaps.
   ○ Use findings to further understand labor market dynamics.
➢ General Public
   ○ Increased awareness of gender pay gaps issues.
   ○ Influence public opinion and societal norms on pay equity.

Data Descriptions

- The Panel Study of Income Dynamic

- 274 columns & 33,398 rows

- Hourly wage

- Gender, age, industry, occupation, etc.

Dataset Overview

Index	Feature	Description
1	intnum68	1968 Interview Number
2	pernum68	Person Number 68
3	wave	Current Wave of the PSID
4	sex	Gender of Individual (1=male, 2=female)
5	intnum	Wave-specific Interview Number
6	farminc	Farm Income
7	region	Region of Current Interview
8	famwgt	PSID's Family Weight, Used in Analyses
9	relhead	Relation to Head of Household (10=head; 20=married wife; 22=partner)
10	age	Age
11	sch	Highest Year of Schooling
12	annhrs	Annual Hours Worked
13	annlabinc	Annual Labor Income
14	degree	Agent's Degree Status (0=no degree; 1=bachelor's; 2=advanced)
15	degupd	Agent's Degree Status (Updated with 2009 Values)
16	schupd	Updated Years of Schooling
17	annwks	Annual Weeks Worked
18	union_covered	Union Coverage Indicator
19	usualhrwk	Usual Hours Worked Per Week
20	labincbus	Labor Income from Business
21	yrsexp	Experience
22	selfemp	Self Employed Status (1=Self Employed)
23	smsa	SMSA Dummy Variable
24	perconexp	Personal Consumption of Previous Year
25	hrwage	Hourly Wage
26	annhrs2	Alternate Measure of Annual Hours Worked
27	inflate	Inflation Factor for Earnings Conversion
28	realhrwage	Real Hourly Wage in 2010 Dollars
29	ft	Full-Time Work Status Indicator
30	ba	Bachelor's Degree Indicator
31	adv	Advanced Degree Indicator
32	LEHS	High School or Less Education Indicator
33	total_income	Total Income
34	Race	Race Identifier
35	Industry	Industry of Employment
36	Occupation	Occupation Type

Descriptive Analysis

Distribution of Salaries by Sex

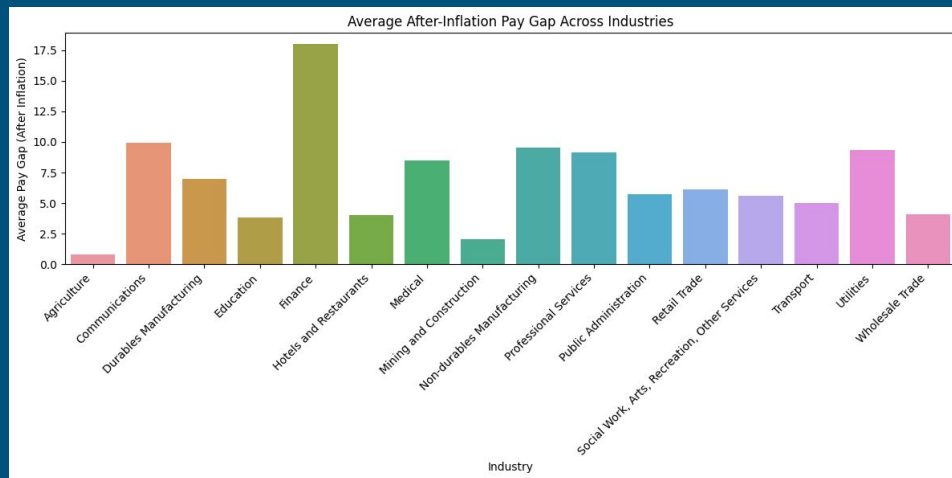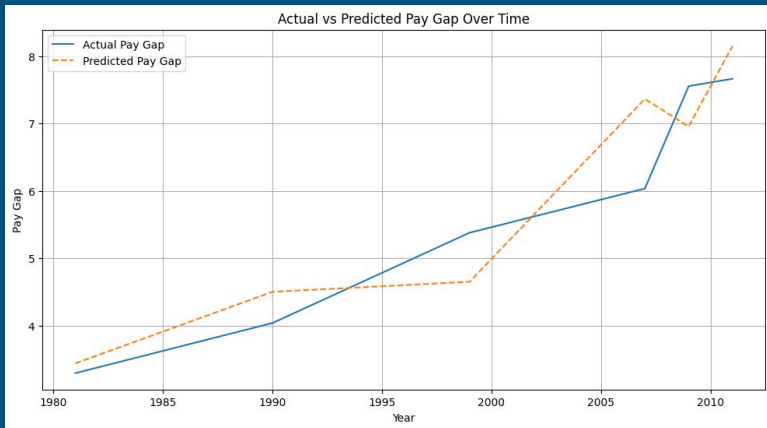
Time vs. Wage Gap
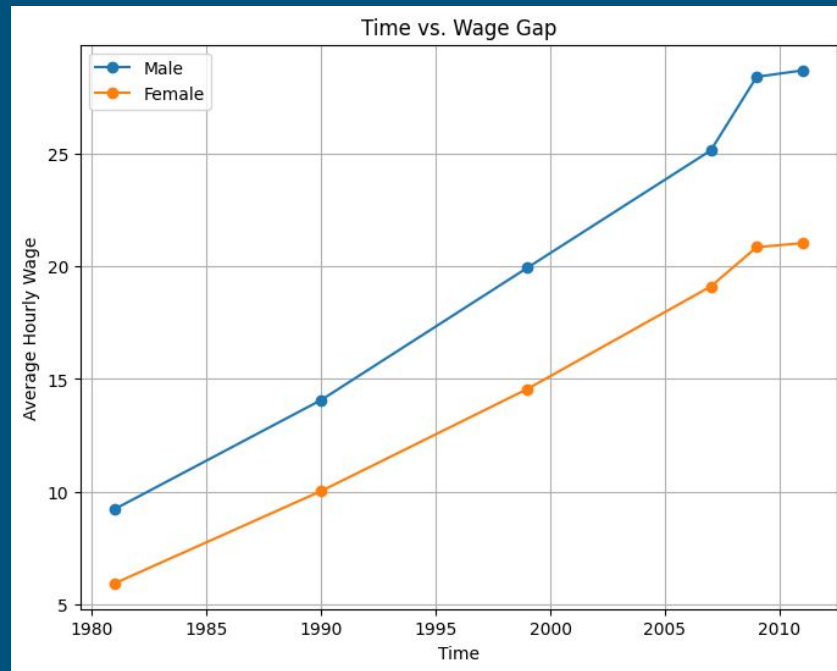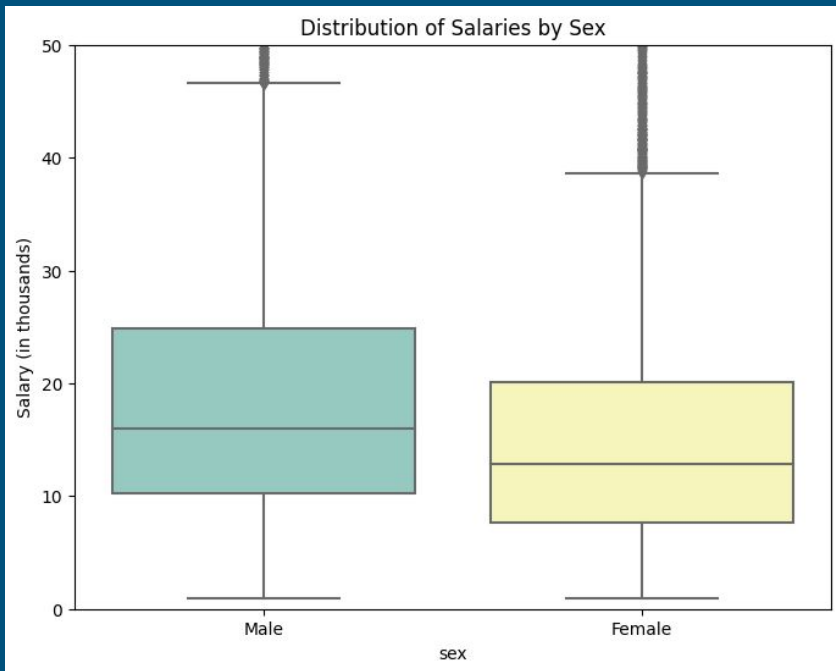
Descriptive Analysis
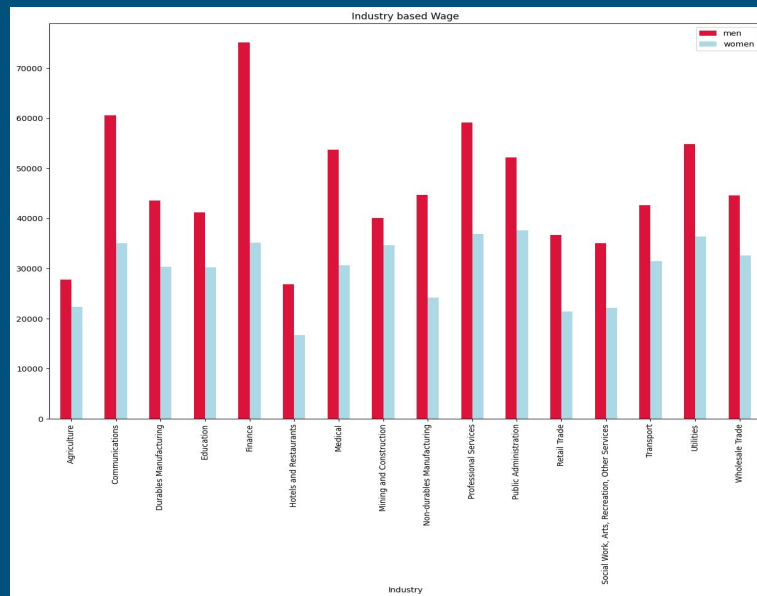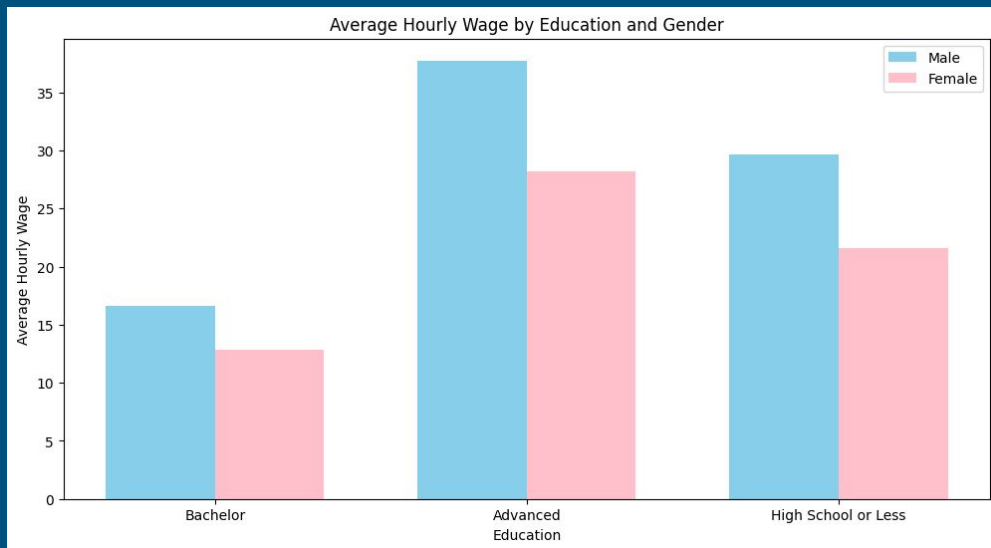
Model Results

- Linear Regression
- Random Forest Regressor
- Support Vector Regressor
- Gradient Boosting Machine
- Neural Network
- KNN
- Lasso/Ridge

Challenges

- Huge & outdated dataset

- Too many categorical features

- Income varies by industry, occupation, individual capabilities, etc.

Conclusions

Link to Collab File

https://colab.research.google.com/drive/1aun0OjBp90Ra2SizwvO6dLTFG65N5LSe?usp=sharing