DSCI 5260: BUSINESS PROCESS ANALYTICS

Professor - Dr. Sameh Shamroukh

Project Title: Employment and Wage Trends Across Industries (2020-2022)

Submitted by: Group 3

Lohith Pasupuleti – 11717635

Pilli Veena Madhuri – 11722559

Veda Samskruthi Kancherla – 11750237

Shivani Nalanagula – 11448943

TABLE OF CONTENTS

ACKNOWLEDGEMENT

We would like to sincerely thank Professor Dr. Sameh Shamroukh for their help, their guidance and continuous support throughout our research project. Their knowledge and thoughtful input were essential in shaping our work and honing our approach. We would also like to thank our team members for their commitment, teamwork, and work to complete this project. Their contributions through engagement, research, analysis, and presentation were an important component of this successful study. We are grateful for each of the resources and learning opportunities we had access to that added layers and dimensions to our research.

ABSTRACT

The U.S. labor market underwent major disruptions between 2020 and 2022 because of COVID-19, rising technological innovation and elevated inflation. This paper uses Quarterly Census of Employment and Wages data to evaluate industry-specific employment changes as well as wage patterns during the major structural shift. This research investigates the impact of external events like pandemic restrictions and technological advancements which lead to remote work on the employee markets throughout various sectors. The research design used quantitative methods which included descriptive statistics along with correlation analysis and t-tests, ANOVA and regression modeling. Wage and employment statistics received analysis within both public and private sectors and specific industries and geographical locations were measured. Research results show that technology and healthcare sectors demonstrated solid growth alongside their high wage levels yet hospitality together with retail industry maintained prolonged obstacles. Wage increases occurred differently between different groups of workers and industry sectors and remote work changed the distribution of occupations. The research establishes important knowledge that helps government officials and business executives design flexible labor policies and equal pay systems and worker skill upgrading programs.

Keywords: *Employment trends, Wage growth, COVID-19 labor market impact, Economic inequality, ANOVA, Random Forest.*

INTRODUCTION

From 2020 to 2022, numerous changes happened in the U.S. labor market because of COVID-19, inflation, a technological boom, and the increase of workers working from home. Even while COVID-19 was responsible for many job losses, growth in the distinct different sectors had been uneven, with technology, healthcare, and logistics growing at a swift continuum. While many sectors in hospitality and tourism were greatly affected by COVID-19, some others faced a height of stagnation. Wages adjustments were common, and some demanded wage hikes due to rising living costs due to inflation. Other mass layoffs included customer service and manufacturing jobs through AI and automation, with some sectors witnessing major transformations due to increased participation of women and other people of color that left many to quit or take part-time positions and widening the wage gap due to the different races and gender. The ever-widening gap between the private and public sectors became even more obvious. In contrast to the state sector, the private sector was more susceptible to changes in economic conditions that arose from the COVID-19 pandemic, but salary policies offered protection for public sector employees, which also protected them from being retrenched from jobs. These changes were assisted by the **Quarterly Census of Employment and Wages (QCEW)**, which provided essential data about them and provided insight income participation in the labor by industry, wage trends, and changes in employment at the sectoral level. This will assist the policymakers, researchers, and politicians in setting out how economic, technological, and other kinds of changes affect workforce. This knowledge can be utilized for designing sound policies for workplace stability, wage equality, and worker security. It will allow policy developers to provide workforce development policies, ensuring future workers develop skills for future markets. Additionally, these data could help businesses rethink their pay and labor policies.

Data Source Link: https://catalog.data.gov/dataset/quarterly-census-of-employment-and-wages-qcew-a6fea

RESEARCH PROBLEM

The U.S. labor force has had tremendous disruption in recent times owing to adjustments of employment, wage gaps, and job security. The COVID-19 pandemic has created some layoffs and shifted demands from one labor group to almost others. In addition, due to inflation, some have questioned the sustainability of wage increases and what implications they might have for purchasing power within the working class. Also, remote work and automation have smoothed the contours of traditional employment, sending shocks among wage evolutions and due processes across various industries. The long-term impacts of such disturbances still remain in the realm of uncertainty, notwithstanding government interventions like wage reforms and stimulus packages. A better detailed study is needed to analyze detailed questions about sectoral employment trends and wage growth over time, while literature reviewed highlights short-term fluctuations in the labor force. Further investigations are required on the impact of policies put into place, on how distinct patterns of employment visibly vary between the public and private sectors, and how technological innovation(s) have molded the workforce. By building a narrative from QCEW employment and wage records, the research also hopes to identify major labor market trends and assess the results of interventions, while also evaluating factors affecting long-term wage sustainability. Understanding these processes will greatly help reduce economic inequalities, build an adaptable workforce, and inform labor policy for the future.

PROJECT GOALS AND TOOLS

This study's central objective is to evaluate employment and wage trends in different sectors from 2020 to 2022, in view of the most important variables enforcing changes in it. This research is to give theories which will aid in corporate scrub, economic policies, and labor initiatives through the application of a data-based approach.

Determining the Main Contributing Elements to Employment and Wage Patterns:

To comprehend salary discrepancies and employment changes, perform a sector-by-sector study.

Examine how labor market patterns are impacted by external variables including inflation, technological advancements, and economic downturns.

Assessing Economic Policy Effectiveness:

Examine how government initiatives like stimulus payments, higher minimum wages, and unemployment insurance affect employment and wage growth.

Assess if these measures have had a lasting impact on incomes and job stability in various industries.

Comparing Employment Trends in the Public and Private Sectors:

Analyze wage progression and job security differences between public and private sector employment.

Determine which industries exhibit the greatest wage discrepancies and evaluate the workforce planning consequences.

Evaluating the Effects of Remote Work and Technological Developments:

Analyze the effects of automation and remote labor on pay structures and employment trends.

Investigate whether these factors have contributed to job displacement or created new labor market opportunities.

Guiding Future Workforce and Economic Planning:

Use data to make recommendations for policymakers and business leaders so that jobs can be more stable and wages more equitable.

Help in developing fair labor policies that uphold economic resilience and adaptability of the workforce.

By addressing these objectives, this research will offer actionable insights into labor market

dynamics and support evidence-based decision-making for improving employment stability and wage distribution.

TOOLS TO BE USED

Basic Excel

The data analysis task involves understanding the dataset structure and performing filtering operations and developing summary reports of employment statistics and wage distributions. The program executes basic mathematical functions along with descriptive statistical computations.

Python (Pandas & NumPy)

Data preparation includes data cleaning in addition to value handling and arrangement for analysis purposes.

Python (Matplotlib & Seaborn)

The data facilitates the production of charts and graphs which display employment patterns together with wage movements.

Python (Scikit-learn & Stats models)

Researchers used regression techniques for discovering wage-influencing factors. Methods enable grouping industrial sectors according to their career retention patterns.

R (ggplot2 & dplyr , t-tests & ANOVA)

The development of graphical displays represents employment fluctuations across different industries and geographical areas. Summarizing data for better understanding. And a wage analysis assesses differences in pay rates between institutions operating in the public sector versus those existing in the private sector. A test must verify the statistical significance between wages.

Tableau

The creation of interactive dashboard systems shows employees' wage growth patterns alongside employment situation developments. A mapping approach displays the job distribution according to state territories.

BRIEF DESCRIPTION OF THE DATA

Geographic & Time Information

The level of geographic grouping is indicated by the Area Type (such as "County," "State," or "National"). The region names, such "California" or "Alameda County," are area names.

Year: The calendar year (2020, 2021, 2022) is provided by the data.

Time Period: The quarter (e.g., "1st Qtr," "2nd Qtr," "3rd Qtr," and "4th Qtr") of the year to which the data pertains.

Ownership & Industry Classification

Ownership – Specifies the type of employer:

1. Private Industry
2. Federal Government
3. State Government
4. Local Government

The NAICS Level indicates the industry categorization level in accordance with the North American Industry Categorization System (NAICS). While Level 5 includes highly specialized sub-industries, Level 2 covers broad industry sectors.

NAICS Code: According to NAICS, each industry is given a numerical code. As an illustration, "54133" is equivalent to "Engineering Services."

Industry Name: Indicates which industry category (such as "Health Care and Social Assistance," "Engineering Services") corresponds to the NAICS Code

Employment Data

Establishments: The quantity of commercial sites (establishments) in a specific industry and geographic area.

Average Monthly Employment: The mean number of workers in the designated sector and area during the quarter.

1st Month Emp: Employment count for the first month of the quarter. **2nd**

Month Emp: Employment count for the second month of the quarter. **3rd**

Month Emp: Employment count for the third month of the quarter

Wage Data

Total Wages (All Workers): NABE Response: This is the total income paid to all employees in a sector, area, and period.

Average Weekly pay: The average weekly wages per worker can be computed by using the l pay and the number of employees.

RESEARCH QUESTIONS

1. What factors drove the changes in employment and income trends across industries between 2020 and 2022? This question assists in identifying industries that have undergone considerable increase, decline, or stability in employment and salaries, taking into account factors such as automation, remote labor, and economic upheavals.
2. What are the main disparities in employment and salary patterns between the private and public sectors (federal, state, and local) over this time period? This question seeks to examine job stability and salary growth discrepancies between the private and public sectors.

LITERATURE REVIEW

This survey of the literature uses five major sources to investigate the economic dynamics of labor markets, wage growth, and inequality. These include examinations of the effects of COVID-19 on labor markets, historical patterns in low-income workers' wages, and the connection between wage inflation and labor market tightness. Future patterns in income distribution are assessed, and the impact of economic policies on wage dynamics—particularly for marginalized groups—is examined.

A critical viewpoint on labor market dynamics, especially in the wake of the COVID-19 epidemic, may be gained by examining employment and salary patterns from 2020 to 2022. Understanding the connection between wage growth and labor market circumstances has been one of the main areas of concentration in economic study throughout this time. Labor market conditions have long been evaluated using conventional metrics like unemployment rates and vacancy-to-unemployment (V/U) ratios. Recent studies, however, indicate that these measures could not adequately reflect pay patterns in contemporary labor markets. Rather, it has been discovered that other metrics, such the quits rate and the vacancies per effective searcher (V/ES) ratio, are more accurate drivers of pay increase.

Unlike conventional measures that focus on labor market slack, these new indicators emphasize worker bargaining power and job-switching behavior, which directly influence wage adjustments. Regression analysis and econometric modeling confirm that these alternative indicators offer more precise wage forecasts than traditional unemployment-based metrics. This realization calls into question established economic theories that have traditionally connected pay increases to total employment levels. The study offers a more nuanced view of how market tightness impacts pay dynamics by utilizing these improved indicators, which is helpful information for policymakers attempting to strike a balance between wage growth and economic stability.

One of the study's principal conclusions is that, in spite of the pandemic's effects on the economy, wage growth was comparatively stable. Although short-term inflation and productivity shocks influenced salaries, they had little effect on long-term wage-setting patterns. This means employers

still largely influences the setting of wages. It serves as evidence to the more generalized economic theory, that is, the decision of companies on how to determine wage structure is mainly dependent on labor markets' circumstances instead of cyclical changes in the economy. In that sense, grasping such wage-setting mechanisms is all-deciding not only in regard to what policies should be piloted, but also in terms of addressing equity in remuneration against economic stability.

Impact of COVID-19 on Labor Markets

According to 2020—2022 Quarterly Census of Employment and Wages (QCEW) data, the COVID-19 pandemic did not have a uniform effect on employment (U.S. Bureau of Labor Statistics, 2023). The hardest hit sectors of the economy were those with the lowest pay across workers, those of color, and women. Many people in sectors where it was not possible to work remotely, such as childcare, restaurant, and hospitality among many others, lost their livelihoods when businesses closed down or reduced operations, (Piacentini et al., 2022).Because schools and childcare facilities were closed, many mothers had to leave the workforce to stay home and care for their children (Gould & deCourcy, 2023). It was also a reversal of years of progress in female labor force participation. Especially in service jobs where many black and Latino workers were concentrated as frontline employees, layoffs and exposure to the virus further amplified racial disparities in work outcomes (Piacentini et al. 2022).

Although government initiatives such as stimulus checks, prolonged unemployment insurance, and emergency rental assistance helped soften the immediate economic shock, they were unable to avert long-term job loss (Jordà & Nechio, 2023).While White-collar workers profited from won working options, their income stability was intact and in some cases their salaries increased with more competition for skilled labor (Heise et al., 2024). In addition to these factors (remote work opportunities and rising asset values) it widened the gap between the high-income earners and low wage workers, due to the pandemic. However, remote work was not universal. Lower income workers who typically lacked the digital infrastructure and employment stability to make the transition to remote work were included in administrative and supporting jobs, therefore expanding the digital divide and sustaining inequality (Piacentini et al., 2022).

Structural Shifts in the Labor Market: Automation and Digitalization

The most lasting impact of the pandemic was the fast wave of automation and digitalization of all industries (Piacentini et al., 2022). Technology became part of the solution for businesses to keep operations going with fewer human workers in manufacturing, logistics, retail and food services. Self-service checkouts and automated inventory management systems were extended by retailers to the extent that this reduces the need for cashiers and stock clerks (Piacentini, et al., 2022). Like fast food chains and restaurants, digital ordering and food processing technologies were adopted by the fast food chains and restaurants to reduce dependency on labor. E-commerce rose dramatically, reducing the number of employment opportunities in the traditional retail stores, moving demand towards warehouse and logistics jobs that usually require new technical skills. Without extensive retraining of displaced low skill workers, they had enormous barriers to reentry into the labor force. However, if education and training initiatives targeting the specific group are not pursued, the ongoing trend toward automation could only expand further the labor market inequality gap.

The Widening Income Gap and Policy Implications

The pandemic has only reinforced longstanding economic inequities as there has been divergence between the employment outcomes for low wage and high-income earners (Gould & deCourcy, 2023). While high income earners weren't only kept in their jobs, they also saw wage growth and financial market gains, low wage workers lost jobs, had income instability and lost career advancement opportunities. Post pandemic QCEW data for 2020 – 2022 shows that income inequality could persist in the post pandemic labor market with the trend and automation of remote work (U.S. Bureau of Labor Statistics, 2023). As a result, many low wage jobs lost in the pandemic are unlikely to return, increased reliance on gig, part time and temporary work arrangements, which tend to provide lower wages and fewer benefits (Piacentini et al., 2022).To facilitate displaced workers returning into the ever-changing economy, reentering workforce development strategies might focus on providing opportunities for reskilling, vocational training, and digital literacy (Gould & deCourcy, 2024). From a policy perspective, in addition to minimum wages being increased, labor protections are expanded, and funds are allocated to build infrastructure and renewable energy initiatives, these actions would provide stable employment opportunities as well as reducing economic inequality over the long run (Gould & deCourcy, 2024).

Government Policy Response and Labor Market Recovery

Jordà and Nechio (2023) point out that government fiscal interventions, such as expanded unemployment benefits, stimulus checks, and small business support programs, played a major role in mitigating the immediate economic effects of the pandemic. These measures temporarily stabilized the financial situation for some workers to search for a better job opportunity.But once the benefit expired, many of these workers were unable to establish financial stability again. Long term job insecurity was the lot of workers that didn't possess enough skills to easily adapt to a digitally driven economy, whereas higher skilled workers were able to easily move to remote or resilient lines of work (Piacentini et al., 2022). Government stimulus was able to prevent a deeper economic crisis, but not structural inequalities made worse by the pandemic. Therefore, present policy efforts will need to strengthen social safety nets, broaden reskilling programs, and reinforce labor standards in order to foster a more inclusive recovery (Gould & DeCourcy).

Wage Growth and Low-Wage Workers

As recent research has shown, low wage workers saw large improvements in real wages from between 2019 and 2022, with the 10th percentile hourly earnings increasing by about 9 percent, or historically as the rise in the first chart (Gould & deCourcy, 2023). This trend was also motivated by policy interventions in combination with tight labor markets.But while minimum wage gains artificially increased wages of the lowest earners, most low-income workers remained in a financially insecure position. In many areas, wage growth lagged behind rising living costs, consequently increasing the difficulty of workers to achieve lasting economic stability (Gould & deCourcy, 2024). The study also found that geographic variation in wage growth also highlighted the influential role that local economic conditions played alongside state-level policy interventions, including an extremely aggressive minimum wage system. These helped in the short term to reduce income inequality, but challenges remained. Purchasing power was eroded by inflation and the gap in wages based on race and generation was still wide (Gould & deCourcy, 2024). Supporting labor protections like paid sick leave, good scheduling, and workforce development can help to drive more sustainable gains for low-wage workers.

Labor Market Tightness and Wage Inflation

Traditional metrics like the unemployment change rate cannot explain the wage growth during the post pandemic recovery (Heise et a., 2024). However, a more accurate picture of labor market tightness is given by alternative indicators such as the quits rate and the vacancies per effective searcher (V/ES) ratio. High quits rates and persistent vacancies indicated that workers had more bargaining power, in particular in lower wage industries (Heise et al., 2024). But there is still doubt as to whether inflationary pressures will strengthen or erode wage growth. Similarly, some economists warn that aggressive interest rate hikes aimed at curbing inflation could hurt low-income earners disproportionally by decreasing job availability (Jordà & Nechio, 2023). Some argue that moderate inflation is desired for sound economy. Further, the attainment of long-term income equity will necessitate a continued policy effort that maintains an acceptable inflation control and allows for the corresponding rate of wage growth.

The Relationship Between Inflation and Wage Growth

Expansive fiscal policies and shifts in the labor market through the pandemic created a relationship between inflation and wage growth during 2020-22 (Jordà and Nechio, 2023). Government relief efforts boosted demand from consumers and paid for wage increases, but it also encouraged inflation in the recovery. In fact, data reveal that for every five percentage points increase in fiscal transfers, wage growth increased by almost three percentage points (Jordà & Nechio, 2023). Living standards were to some extent improved temporarily by higher wages, but so were vitality, since higher wages enhanced consumer demand and had inflationary implications. In the short run, gains in incomes through rising wages safeguarded workers economic positions at tolerable levels as they could afford essential goods and services (Gould & deCourcy, 2023). But inflation had hollowed out making it difficult for low wage workers to buy necessities because they had spent a bigger share of their income on necessities. Wage growth is unrealistic if there are no guarantees of sustainability. Thin profit margins may discourage future wage increases, if inflation is persistently high (Jordà & Nechio, 2023). Thus, along with other measures aimed at stabilizing prices, protecting vulnerable workers, supporting productivity, effective wage policies must be pursued.

Short-Term Benefits of Wage Growth During Inflation

In the short term, higher wages during the post-pandemic period of inflation brought near-term relief for many workers. During a constrained labor market climate following the COVID-19 pandemic, employers were frequently required to raise wages in order to attract and retain workers (Heise et al., 2024). The wage gains particularly benefited the lowest-paid workers who had experienced substantial earnings losses during the pandemic recession. Low-wage workers particularly enjoyed unprecedented gains—for instance, workers at the 10th percentile of the wage distribution saw approximately a 9% gain in real wages between 2019 and 2022, the most rapid of average rates of expansion on record (Gould & deCourcy, 2023). The average wages of the lowest-paid workers had grown more than 13% altogether by 2023 since 2019, the most dramatic spurt on record (Gould & deCourcy, 2024). The wage gains allowed numerous families to more easily meet the cost of basic needs such as rent, food, health care, and transportation and level their finances. In addition, since the earnings of workers holding lower earnings grew proportionally more rapidly than did the earnings of higher earners, recent wage gains modestly reduced certain income inequalities (Gould & deCourcy, 2023). Greater compensation also supported consumer confidence, and families increased discretionary spending. The consumer binge fueled the demand for goods and services across the economy and helped propel the businesses from the pandemic recession and created new jobs within a virtuous circle. Furthermore, the mix of generous government assistance (such as stimulus payments) and rising wages brought the economy out of recession more rapidly, enabling employment within numerous industries to return more rapidly than expected (Jordà & Nechio, 2023; Piacentini et al., 2022).

Long-Term Concerns: The Sustainability of Wage Growth Amid Rising Inflation

Despite these short-term benefits, growing inflation has created concerns about the sustainability of wage gains. As consumer prices rose, the purchasing power of workers' wage gains began to be eroded. Many families could not afford the rising costs of housing, food, and utilities, particularly if these prices outgained wage benefits (Gould & deCourcy, 2024). The heavy toll of this was on lower-paid workers, who spend a greater proportion of earnings on basic needs and have fewer buffers against price rises. In fact, the evidence indicates that post-pandemic wage and price trends

have benefited more higher-paid workers, such that low earners have enjoyed relatively smaller gains in real incomes once the impact of inflation has been considered (Jordà & Nechio, 2023). Even where these significant nominal wage gains have existed, they have not always kept pace with the rising costs of living among lower-paid workers (Gould & deCourcy, 2024). Sustained inflation also placed considerable strain on businesses by hiking the cost of goods sold. With rising labor costs and other costs of inputs (e.g., materials and supply chain disruptions), many businesses have had to make tough decisions. Some employers have curtailed hiring, cut back on employee hours, or sped up investments in automation as a means of keeping a lid on labor costs. Others have charged higher prices as a way of transferring the increased cost on to the consumerism that accelerates the cycle of price pressures and diminishes the purchasing power of families. There are increasingly serious concerns that eventually wage gains will slow. Unless inflation drops back to modest levels, employers (most particularly those of tight profit margins) may be deterred from continuing to keep up wage offers. Furthermore, the Federal Reserve's measures to slow inflation by hiking interest rates have started cooling the labor market, thus moderating wage growth. Although a necessity for curbing prices, increased interest rates dampen the demand for labor and have the potential to undermine some of the recent wage gains. Should the wage growth trail far behind the price growth for a considerable time, wages will deflate, having the potential of negating many of the post-COVID gains in living standards (Jordà & Nechio, 2023).

Balancing Wage Policies with Economic Stability

Policymakers are engaged in a delicate balancing act of supporting wage growth while ensuring the overall economic stability. They have tried to find means of increasing workers' incomes and safeguarding vulnerable populations without risking spurring an inflationary spiral or overloading businesses. One of the ways of addressing this challenge is by focusing the wage policy on lower-paid workers who most require the raise but have a proportionally modest contribution towards aggregate labor costs. For example, a minimum wage hike or the expansion of tax credits such as the Earned Income Tax Credit (EITC) would raise the take-home wages of the low-paid workers without significantly rising employers' wage costs (Gould & deCourcy, 2023). Economists have called for such actions where they argue that stronger wage floors and labor supports are key to countering the rising inequality while ensuring that inflation stays at bay (Gould & deCourcy, 2023).

Another means of addressing this challenge involves investment in productivity gains such that the wages ascend a notch as the firms improve. Through the financing of job training, education, and the installation of technology, policymakers make the workers more productive such that the firms are able to meet higher compensation without compressing the profit margin. Besides that, the government may adopt measures to mitigate the labor deficits by promoting increased labor participation. Subsidized childcare, employment-based family leave, and flex-working arrangements enable more individuals (particularly parents and carers) to join the labor market or stay within it. Raising the supply of labor in this manner relieves pressures on employers to make hires and can cut the spiral of wage-push inflation without short-changing employment.Monetary policy represents a concurrent role in stability. When inflation rose, the Federal Reserve intervened by hiking interest rates in order to slow the economy and put the brakes on price expansion. The higher interest rates have cooled off inflation but have also stemmed the job market, thus curbing wage expansion. Policymakers throughout this period grappled with the task of tuning the interest rates hikes to dampen inflation without smothering the economic rebound or unraveling hard-earned wage gains (Jordà & Nechio, 2023). Finding the balance is key towards ensuring that wage expansion ought to be sustained over the long term without fueling inflation.

LITERATURE REVIEW SUMMARY

Wage Growth and Labor Market Tightness (Heise, Pearce, & Weber, 2024)

Purpose of the Study: The analysis investigates how wage growth is related to two measurements indicators of labor market tightness which include the quits rate alongside V/ES (vacancies per effective searcher). The study aims to develop better economic frameworks for enhanced wage inflation projections and monetary policy adjustments.

Key Findings and Contributions:

Traditional metrics like the unemployment rate or vacancy-to-unemployment (V/U) ratio are not as good indicators of wage increase as the quits rate and V/ES.

Labor market tightness shows a constant relationship with wage growth rates while temporary productivity changes and some inflation theories do not affect pay increases.

A new pay forecasting indicator built from quits and V/ES data should replace the current labor market tightness measurement.

The results of the study are consistent with a New Keynesian DSGE model that takes into consideration the dynamics of on-the-job searches and the wage-setting behavior of employers.

Methodology:

The research utilizes ordinary least squares (OLS) regressions together with local projections methods to forecast wage growth.

Uses data from the Bureau of Labor Statistics (BLS), the Job Openings and Labor Turnover Survey (JOLTS), and other sources.

A "horse race" analysis detects predictive labor market tightness indicators by assessing multiple measurement methods.

Researchers investigate wage changes that occur during productivity shocks while verifying various wage-Phillips curve relationships.

Low-Wage Workers Have Seen Historically Fast Real Wage Growth in the Pandemic Business Cycle (Gould & deCourcy, 2023)

Purpose of the Study: This study examines the effects of labor market factors and economic policies on the increase of low-paid workers' wages in the US from 2019 to 2022. Real paygrowth rates for lowest-paid employees reached their highest speeds according to the study while analyzing governmental policy impacts on these trends.

Key Findings and Contributions:

Exceptional Wage Growth: Real income rates for low-wage employees (10th percentile) increased by 9.0% since 2019, which represents the largest income growth in any business cycle dating back to 1979.

Wage Compression: The wage rise at the lower paid positions created a smaller gap between income groups of workers. Despite the situation, the wealthiest 1% maintained a larger share of the total profits.

Policy Impact: Government initiatives that included stimulus checks and child tax credits along with enhanced unemployment benefits provided major advantages to people earning minimum wages.

Tight Labor Market Influence: Employers had no choice but to raise employee compensation because their workforce became increasingly scarce in essential positions.

State Minimum Wage Increases: The wages of low-paid workers expanded at a quicker rate when states increased their minimum wage requirements. However, because of more general economic causes, overall wage rises were robust in every state.

Persistent Inequality: Low-wage workers primarily made up of women along with Black and Hispanic earn insufficient pay at $12.57 per hour and year to year they do not benefit from salary advancement despite their minority status.

Policy Recommendations: The report advocates raising minimum wage rates along with better standards that safeguard workers who must rely on payroll checks.

Methodology:

Data Sources: The study analyzes pay pattern modifications through the documentation of information from Current Population Survey (CPS) Outgoing Rotation Group and supplemental data resources.

Comparative Analysis: Wage growth patterns for 2019–2022 stand apart from all other occupational cycles from 1979 to date.

Wage Distribution: The evaluation breaks workers into five income categories to study wage effects: low-, lower-, middle, upper and high wages.

Policy Impact Assessment: The research examines how both workforce conditions and regulatory modifications impact salary changes primarily for employees who need to receive minimum wages.

The Impact of COVID-19 on Labor Markets and Inequality (Piacentini et al., 2022)

Purpose of the Study: The analysis of American labor markets and economic inequality during the The COVID-19 epidemic concentrates on employment patterns together with wage disparities and adapting workforce practices. The authors analyze how different jobs and population sectors particularly service workers with lower incomes have been impacted.

Key Findings and Contributions:

Members of minority groups along with women struggled worse economically because these groups predominated in businesses that bore the worst consequences of the pandemic.

Income losses were diminished by government stimulus programs and unemployment benefits yet remote work shifted available jobs toward higher-paid employees thereby worsening permanent wage inequality.

The pandemic extracted its greatest economic impact on employees who earned lower incomes while working in service industries and resulted in higher job losses and income disparities between groups.

The study results demonstrate the critical need to protect minimum wage workers when financial emergencies occur in future.

Methodology:

Statistics regarding employment alongside wages from various industries along with demographic classifications are analyzed in the study through economic literature from 2020 and 2021 to evaluate labor market shocks and income inequality.

The research evaluates both future structural economic changes including automation growth alongside telework development along with their possible effect on labor market results.

Fastest Wage Growth Over the Last Four Years Among Historically Disadvantaged Groups (Gould & deCourcy, 2024)

Purpose of the Study: The research examines wage pattern changes from 2019 to 2023 for minimum wage earners with an analysis of the impact from state-level wage increases and market conditions.

Key Findings and Contributions:

Average wages for low-paid employees increased by 13.2%, which represents the biggest surge we have seen in decades because minimum wage hikes came from individual states combined with a competitive job market.

The pay improvements implemented for low-income workers fail to meet their basic living expense requirements.

The wage difference across different groups has shown a slight reduction but major inequalities persist particularly among female and minority employees.

The study specifies that maintaining wage growth and reducing income inequality depends on continuous legislative measures.

Methodology:

Theory-based adjustments of wages are assessed through information gathered from Current Population Survey (CPS) data.

The researcher analyzes percentage wage changes between states which raised their minimum wages and those that maintained no such policy adjustments.

Real buying power assessment relies on wage increases adjusted for inflation levels.

Inflation and Wage Growth Since the Pandemic (Jordà & Nechio, 2023)

Purpose of the Study: The research analyzes how pandemic government transfer programs affected inflation rates alongside wages through their impact on employer pay decisions and their ability to inflate prices.

Key Findings and Contributions:

Inflation and wage growth increased by 3 percentage points for every 5 percentage points when direct fiscal transfers were increased.

Employee compensation adjustments through variables were determined by inflation-related discussions that occurred between employers and their staff when negotiating payments.

Experimental data investigated by Border shows that substantial financial measures lead to sustained inflation according to research results.

The study enhances overall discussions about wage-price phenomena along with economic recovery following the pandemic.

Methodology:

A dynamic difference-in-differences model allows the researcher to establish the causal link between fiscal transfers and wage levels and inflation rates.

The evaluation analyzes price patterns together with wage modifications through data analysis from the pandemic period.

The Inflation Expectations-Models Link provides a method to evaluate the feedback system which occurs between wage determination methods and infl ation expectations.

RESEARCH GAPS

Analysis of Literature Gaps and Opportunities for Further Research

Wage movements between 2020 and 2023 reveal the relationship between inflation and government measures and restricted labor markets according to the reviewed papers. The significant research gaps make extensive investigations of a subject possible.

Identified Gaps in Existing Research

Lack of Sector-Specific Wage and Employment Trends

These investigations analyze overall market employment data while not discussing how individual industries suffered during this time. Several studies omit the specific employment and wage patterns across industries although they discuss wage differences within the low-income group.

Opportunity: Executive decision-makers need to analyze the complete effect of the pandemic on healthcare together with manufacturing and technology industries alongside the hospitality sector when making recovery strategies.

Long-Term Sustainability of Wage Gains

The examined studies focus primarily on short-term salary changes in low-income worker pay increases because of labor market shortages along with policy adjustments. Researchers infrequently analyze whether wage hikes triggered by the pandemic will persist after the recovery period.

Opportunity: Future academic research needs to assess if increased pay levels will persist as labor markets stabilize after the pandemic.

Geographic Disparities in Wage Growth

Studies show no analysis of salary dynamics between different geographical settings which include urban and rural distinctions or statewide differences. When researchers fail to account for geographical differences in their research the results become less applicable because economic conditions vary by location.

Opportunity: The research focuses on one geographical area to find which locations achieved the highest wage growth rates alongside their root causes.

Public vs. Private Sector Wage Growth

The studies only provide partial sectoral comparisons although they recognize salary trends. It is important to evaluate how the pandemic affected salary differences between public and private sectors, as public sector employment tends to be more stable but offers slower wage growth.

Opportunity: Research investigating how stabilizing wages and maintaining employment rates looks between public sector workers and private sector workers should be conducted.

Impact of Automation and Remote Work on Wages and Employment

Remote work receives mention in research documents, but the study omits how it impacts pay systems and job security and career advancement. Studies have not offered insights into how automation affects either wage patterns during pay compression or workers' job movement.

Opportunity: The analysis between automation and distant work affects how new workplace skill acquisition and employment safety policies should be developed across different business sectors and employee earnings.

Inflation's Varying Effects Across Different Wage Groups

Jordà & Nechio (2023) studied the relationship between fiscal stimulus and inflation and wages but failed to analyze wages based on varying categories. Wage adjustments that fight inflation brought better benefits for wealthy people than for individuals with lower earnings.

Opportunity: The measurement of differential inflation effects on wage workers depends on a detailed examination of their income levels.

HOW THIS RESEARCH CAN CONTRIBUTE TO FILLING THESE GAPS

Industry-Specific Analysis of Wage Trends

The research examines the complete reaction pattern between economic disruptions and recovery measures through industry-level employment and pay patterns.

Longitudinal Study on Wage Growth Sustainability

The investigation should continue beyond 2022 to determine whether pandemic wages will remain permanent changes in labor markets or return to pre-pandemic levels.

Regional Wage Disparity Analysis

Labor market results depend on local economic factors together with policies because this analysis will illustrate wage development throughout different locations.

Comparing Public vs. Private Sector Wage Growth

This investigation analyzes employment stability alongside the effectiveness of public job policy through wage and employment data examination of different working sectors.

Exploring the Effects of Automation and Remote Work

The study examines employment patterns and earnings changes and workforce composition resulting from automation and teleworking to support firms and governments adjust their labor markets.

Inflation's Impact on Different Wage Groups

Detailed understanding about economic inequality plus policy recommendations become possible through examining wage growth based on sector and different income levels.

Conclusion

Additional research becomes possible due to the gaps identified within existing literature. The research fills important knowledge gaps which lead to improved laboratory market surveys as well as wage trends and policy effects understanding. The obtained findings will support economists alongside companies and political figures when they work to build economic plans for resilience along with employment growth and wage stabilization.

RESEARCH METHODS

Data Sourcing

The data for this research is sourced from the **Quarterly Census of Employment and Wages (QCEW)**, a dataset provided by the **U.S. Bureau of Labor Statistics (BLS)**. This dataset is chosen for the following reasons:

Comprehensive Coverage

The QCEW set is one of the widest-ranging datasets available to analyze labor markets, covering nearly all employment in the U.S. subject to unemployment insurance laws. It covers private-sector employment as well as jobs with the federal, state, and local governments. Thus, it gives a holistic overview of the changes in employment dynamics.

Reliability and Accuracy

The data comes from required employer-reported unemployment insurance (UI) filings which are very accurate and complete. Since it is an administrative dataset, it avoids issues like survey biases, non-response errors, or misreporting.

Detailed Industry and Geographic Classification

The dataset follows the North American Industry Classification System (NAICS), allowing detailed sector-wise analysis. Geographic classification includes national, state, county, and metropolitan area data, enabling regional wage and employment comparisons.

Quarterly Updates for Trend Analysis

The dataset provides quarterly employment and wage data, making it suitable for time-series trend analysis. This is crucial for studying economic disruptions, recovery phases, and long-term wage growth patterns.

Publicly Available and Free Access

The dataset is open-source and easily accessible through the BLS portal, making it a cost- effective and verifiable source of information.

DATA COLLECTION METHOD

Since this research does not involve primary data collection, it relies on pre-existing administrative records from the BLS QCEW dataset. However, if additional insights were needed, supplementary data collection methods such as surveys or interviews could be used.

Primary Data Collection via QCEW

Method of Data Collection

The QCEW data set is compiled from state unemployment insurance (UI) tax reports that businesses must submit every quarter.

These reports include information on employment levels, total wages, and industry classification.

Why This Method?

Establishes the Sampling Bias: There are no worries about sampling errors since the data includes almost all wage and salaried employees.

Employers Are Legally Obligated to Submit: Due to laws mandating employers to report employment and wage data quarterly, it is very credible and complete.

Consistency Between Industries: Standardized NAICS classifications allow for comparative analysis between industries.

If Supplementary Data Was Collected (Surveys or Interviews)

Qualitative insights into employment and wage trends can be gleaned from surveying or interviewing commercial owners, HR professionals, or industry experts.

Survey Methodology (If used)

Participants: Participants shall be HR managers, business executives, and policymakers from various industries.

Sampling Strategy: Sampling Strategy-unbiased stratified sampling to get a representation of the different industries and business sizes.

Survey Format: Online questionnaire (Google Forms, Qualtrics) or structured phone interviews.

Data Collected: Insights on hiring trends, wage policies, economic disruptions, and industry- specific challenges.

Justification for This Method

Fills Contextual Gaps: Administrative data provides quantitative trends, while surveys add qualitative insights.

Explains Wage Changes: Helps understand policy decisions, automation impacts, and remote work influence on wages.

Flexible and Scalable: Can be expanded or modified based on findings from the QCEW dataset.

PARTICIPANT DESCRIPTION

The database contains 15 measurement points which provide an extensive viewpoint for studying work patterns within different sectors and regional territories. Using 798,464 records strengthens the statistical analysis power because of the extensive sample size. Researchers can study distinct economic patterns using these variables because they evaluate both industrial sectors and workforce statistics and salary distributions. The database enables diverse research questions to be answered so scientists can study workforce dynamics together with regional economic results and industry-specific labor patterns.

Area Type	Area Name	Year
Time Period	Ownership	NAICS Level
NAICS Code	Industry Name	Establishments
Average Monthly Employment	1st Month Emp	2nd Month Emp
3rd Month Emp	Total Wages (All Workers)	Average Weekly Wages

Although the study's primary focus will be quantitative analysis, the dataset structure also allows for potential qualitative insights if additional contextual data were collected through interviews or surveys. This qualitative aspect would complement the statistical analysis by refining the understanding of economic behaviors and labor market shifts, offering a more comprehensive perspective on workforce trends beyond numerical patterns.

Procedures for Conducting the Survey: Consent Forms, Communication, and Data Collection

If supplementary survey or interview-based data collection is required to complement the QCEW dataset, the following structured procedures will be followed to ensure ethical compliance, efficient communication, and accurate data collection.

Ethical Considerations and Consent Forms

Before collecting any data from participants, the study ensures ethical compliance by securing informed consent and adhering to research integrity standards.

Informed Consent Process

Each participant will be required to read and sign a consent form before participation. The consent form will include:

Study Purpose: A brief overview of the research on employment and wage trends.

Data Collection Methods: Explanation of the survey or interview process.

Voluntary Participation: Participation is fully voluntary and can be withdrawn from any moment.

Confidentiality and Data Protection: This explains how responses will be anonymous and only used for the purpose of academic research.

Contact Information for Queries: Details of the researcher or institution for participants to ask questions.

Approval from Institutional Review Board (IRB)

If required, the study will be submitted for **Institutional Review Board (IRB)** approval to ensure that all human subject research follows ethical guidelines and protects participants' rights.

Participant Recruitment and Communication

To ensure a diverse and representative sample, participants will be strategically recruited from various industries and employment sectors.

Identifying Target Participants

Participants will include:

HR Managers and Business Executives – To provide insights into hiring trends and wage decisions.

Industry Professionals and Employees – To discuss their experiences with wage changes.

Government Policy Experts – To understand regulatory impacts on employment.

Economists and Labor Market Analysts – To provide expert opinions.

Recruitment Strategy

Email Invitations: Personalized email invitations will be sent to potential participants explaining the research goals and inviting participation.

Professional Networks: LinkedIn, industry-specific forums, and business associations will be used to reach relevant professionals.

Snowball Sampling: Participants will be encouraged to refer colleagues for a broader and more diverse sample.

Communication Plan

Pre-survey Briefing: This is a brief email or an introductory session where the study would be explained, participants' concerns would be answered, and expectations would be set.

Follow-up Reminders: Automated or manual reminders will be followed up to encourage participation.

Survey/Interview Scheduling: The Survey/Interview Scheduling: The participants will take interviews at the time they find most appropriate and will also have deadlines within which to fill in the surveys.

Survey Administration and Data Collection:

Depending on the research requirements, data collection will be done through online surveys or structured interviews.

Survey Format

Online Surveys (Google Forms, Qualtrics, SurveyMonkey)

Quantitative Questions: Multiple-choice, Likert-scale responses on employment and wage trends.

Qualitative Questions: Open-ended questions on industry-specific wage policies, job security, and economic impacts.

Structured Interviews (Zoom, Microsoft Teams, or phone calls)

If a more in-depth qualitative analysis is required, interviews will be conducted with business leaders and policy experts.

The interview questions will be semi-structured to allow for flexibility in responses.

Data Collection Procedures

Digital Surveys: Sent via email with a submission deadline of 2-3 weeks.

Recorded Interviews: If permission is granted, interviews will be audio-recorded and later transcribed for analysis.

Data Anonymization: All survey responses will be anonymized before analysis.

Data Security and Confidentiality

Ensuring data privacy and security is a top priority.

Data Protection Measures

Secure Storage: All collected data will be encrypted and stored in a password-protected **research folder**.

Limited Access: Only authorized researchers will have access to the raw data.

Anonymization: Personal identifiers (names, company details) will be removed before analysis to protect participant confidentiality.

Possible Alternative Procedures

If the initial survey response rate is low, the following alternative methods may be used:

Incentives for Participation – Small incentives (e.g., gift cards, recognition in research reports) may encourage more responses.

Shortened Surveys – A simplified version of the survey with only essential questions may increase response rates.

Industry-Specific Focus Groups – Instead of one-on-one interviews, participants could be gathered into small focus groups for discussion.

Secondary Data Analysis – If primary data collection is limited, the study will rely more heavily on the QCEW dataset and supplement with findings from published research reports.

These structured procedures ensure that data collection through surveys and interviews is ethically compliant, well-organized, and secure. By implementing informed consent, targeted recruitment, structured data collection, and data security measures, the research will maintain high credibility and reliability.

RESEARCH MEASURES

Dependent Variables (Outcome Variables)

Dependent variables represent the key metrics being analysed and influenced by the independent variables.

Total Wages

Description: Total remuneration paid out to all staff employed under a particular industry and location in a certain timespan.

Justification: The variable will provide a solid ground to analyze wage patterns in various industries and regions.

Unit: Measured in U.S. dollars.

Independent Variables (Predictor Variables)

Independent variables influence or explain variations in the dependent variables.

Industry Classification (NAICS Code)

Description: Categorizes business into industry sector using **North American Industry Classification System (NAICS) code**.

Justification: Industries experience different wage growth due to various economic developments and policies. Different economic conditions result in different trends in wages and employment in different industries.

Unit: Categorical variable (e.g., Manufacturing, Healthcare, Technology, Retail).

Geographic Location

Description: The area where the employment and wage data is collected (state, county, metro area).

Justification: It helps compare regional economies based on the chosen location and to assess wage difference trends over time.

Unit: Typically represented as a state, county, or metropolitan area.

Ownership Type

Description: The employer is a Private, Federal, State or Local Government.

Justification: The wages in public and private sector do not necessarily follow the same trend.

Unit: Categorical variable (Private, Federal, State, Local Government).

Time Period (Year and Quarter)

Description: The year and quarter the employment and wage data was recorded.

Justification: Essential for analyzing employment trends through time and understanding economic cycles.

Unit: Categorical variable (e.g., Q1 2020, Q2 2021).

Unemployment Rate

Description: The percentage of the labor force that is unemployed within a given region.

Justification: Unemployment rates affect wage growth and job availability.

Unit: Percentage (%).

Inflation Rate

Description: The rate at which the general price level of goods and services is rising.

Justification: Inflation affects purchasing power and wage growth.

Unit: Percentage (%).

Minimum Wage Policy

Description: The minimum remuneration allowed by law in a given state.

Justification: Affects wage increase as well as employment particularly of low wage earne.

Unit: Measured in U.S. dollars per hour.

Remote Work Adoption

Description: The percentage of employees working remotely within a sector.

Justification: Industries with high remote work adoption may see different trends in employment and wages.

Unit: The amount of workforce (% of total) working remotely.

Automation and Technological Changes

Description: The level of automation adoption within an industry.

Justification: Automation can affect employment levels by replacing or augmenting jobs.

Unit: Categorical variable (High, Medium, Low automation adoption)

Summary of Variable Relationships

Dependent Variable	**Key Influencing Independent Variables**
Total Wages	Industry Classification, Geographic Location, Minimum Wage Policy, Inflation Rate
Average Weekly Wage	Industry Classification, Time Period, Ownership Type, Inflation Rate
Employment Levels	Industry Classification, Geographic Location, Unemployment Rate, Automation
Job Growth Rate	Industry Classification, Remote Work Adoption, Technological Changes

The dependent and independent variables provide a framework to analyze trends in employment and wages using QCEW data. By observing these relationships, we can ascertain how different factors affect job growth and wage patterns.

DATA ANALYSIS

DATA CLEANING AND DATA PREPROCESING

Data Loading and dataset Analysis:

We started by mounting our Google Drive to our Google Colab notebook using drive.mount('/content/drive') so that we could access our files in Drive. Our dataset (qcew-2020-2022.csv) is quite large, so we did that to avoid having to upload it every time and to be able to work with it directly from Drive. We then read the dataset into a pandas DataFrame for analysis using pd.read_csv() and stored into 'df' .

After we loaded the dataset, we used df.head() to display the first lines. This assisted in verifying that the data had loaded properly as well as understanding its structure, like area information, industry, employment number, and wage. This is significant before performing any cleaning or analysis.

```
df.head()
```

	Area Type	Area Name	Year	Time Period	Ownership	NAICS Level	NAICS Code	Industry Name	Establishments	Average Monthly Employment	1st Month Emp	2nd Month Emp	3rd Month Emp	Total Wages (All Workers)	Average Weekly Wages
0	County	Alameda County	2020	1st Qtr	Federal Government	2	1024	Professional and Business Services	2	9	9	9	9	279730.0	2391
1	County	Alameda County	2020	1st Qtr	Federal Government	3	491	Postal Service	28	3189	3198	3181	3189	54692130.0	1319
2	County	Alameda County	2020	1st Qtr	Federal Government	3	541	Professional and Technical Services	2	9	9	9	9	279730.0	2391
3	County	Alameda County	2020	1st Qtr	Federal Government	5	54133	Engineering Services	1	2	2	2	2	31831.0	1224
4	County	Alameda County	2020	1st Qtr	Federal Government	2	62	Health Care and Social Assistance	1	376	384	376	369	8804659.0	1800

As a next step, we checked our dataset for cleanliness. First, we checked the shape of the dataset with df.shape and stored it in initial_shape. We used df.duplicated().sum() to obtain the count of duplicate rows and removed any using df.drop_duplicates(). However, duplicates was 0, so the final shape (final_shape) was the same. Then we used df.isnull().sum() to identify missing values and filtered columns with missing values using missing[missing > 0]. The result showed no missing data, confirming the dataset was clean and ready for the subsequent step of analysis

Output: Initial shape: (798464, 15)

Duplicates removed: 0

Final shape after cleaning: (798464, 15)

Columns with missing values: Series ([], dtype: int64)

Data Type check for columns: We used df.info() to view column names, data types, and non-null counts.

This helped us understand the dataset's structure and check for any needed conversions.

Output:

```
Data columns (total 15 columns):
 #   Column                      Non-Null Count   Dtype
---  ------                      --------------   -----
 0   Area Type                   798464 non-null  object
 1   Area Name                   798464 non-null  object
 2   Year                        798464 non-null  int64
 3   Time Period                 798464 non-null  object
 4   Ownership                   798464 non-null  object
 5   NAICS Level                 798464 non-null  int64
 6   NAICS Code                  798464 non-null  object
 7   Industry Name               798464 non-null  object
 8   Establishments              798464 non-null  int64
 9   Average Monthly Employment  798464 non-null  int64
 10  1st Month Emp               798464 non-null  int64
 11  2nd Month Emp               798464 non-null  int64
 12  3rd Month Emp               798464 non-null  int64
 13  Total Wages (All Workers)   798464 non-null  float64
 14  Average Weekly Wages        798464 non-null  int64
dtypes: float64(1), int64(8), object(6)
memory usage: 91.4+ MB
```

Data Filtering:

Here, we utilized personalized filters and created a new feature to prepare the data for a specific analysis, to match with the project target of investigating sectoral trends in California counties.

The Filter is performed to Primary Ownership Types:

We kept only the records of the primary employer types—Private, Federal Government, State Government, and Local Government by utilizing df[df['Ownership'].isin(valid_ownerships)]. This allows us to contrast public, private employment and wage trends, as outlined in the research goals.

Filter Out Annual Aggregates:

With df[df['Time Period'].isin([.])], we kept only quarterly data (1st Qtr to 4th Qtr) and dropped any "Annual" values. This facilitates time-series analysis at a quarterly frequency, as demanded in the technical methods (e.g., seasonal decomposition, ARIMA).

Restrict to California Counties:

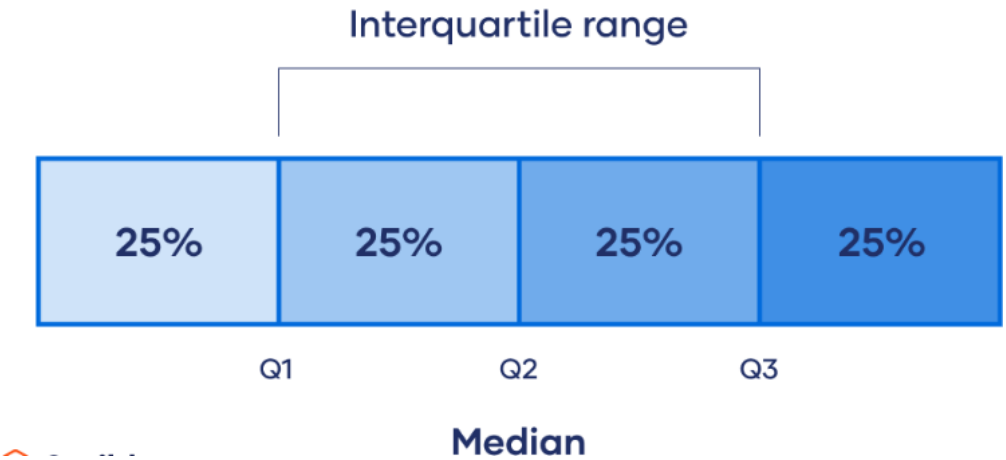
With df[df['Area Type'] == 'County'] and excluding rows where Area Name is 'California' or 'United States', we narrowed the dataset down to California county-level data so that we could do a regional analysis of

Employment and wage patterns are a major gap identified in the literature. Create Average Employment per Establishment:

We constructed a new variable, Avg_Emp_Per_Est, as Average Monthly Employment / Establishments. We also addressed the edge cases by replacing infinities and NaNs with zero with.replace([float('inf'), -float('inf')], 0).fillna(0). This indicator helps to measure workforce concentration and business size and facilitates industry-level comparisons.

Outlier Removal Using Interquartile Range (IQR) Method:

The Interquartile Range (IQR) method is a widely used statistical technique for detecting and deleting outliers from quantitative data. IQR is involved with 50% of the information center, i.e., between first quartile (Q1), the 25th percentile and third quartile (Q3), the 75th percentile.



IQR = Q3 - Q1

Outliers are values lower than Q1 - 1.5 × IQR or higher than Q3 + 1.5 × IQR

This method is useful since it is not affected by outlier values or skewed distributions, therefore it is especially well-suited to cleaning employment and wage data before analysis.

Now to improve data quality and minimize distortion in analysis, we defined a function remove_outliers_iqr() to eliminate extreme values from numeric columns using the IQR method. We applied it to 'Avg_Emp_Per_Est' and 'Average Weekly Wages', helping us reduce the impact of anomalously high or low values and ensuring more reliable statistical and regression analysis.

EXPLORATORY DATA ANALYSIS

Descriptive Statistics

We utilized the df.describe() method to generate crude statistical summaries of every numeric column in our cleaned data. The method presents us with a snapshot of the distribution of the dataset via the calculation of the following statistics: count, mean, standard deviation (std), minimum and maximum, and 25th, 50th, and 75th percentile.

Output:

To know the basic statistics of dataset

	Year	NAICS Level	Establishments	Average Monthly Employment	1st Month Emp	2nd Month Emp	3rd Month Emp	Total Wages (All Workers)	Average Weekly Wages
count	798464.000000	798464.000000	7.984640e+05	7.984640e+05	7.984640e+05	7.984640e+05	7.984640e+05	7.984640e+05	798464.000000
mean	2021.043745	4.738440	2.594658e+03	3.304981e+04	2.624615e+04	2.647355e+04	2.659946e+04	9.228472e+08	1291.407310
std	0.823588	1.273008	9.615226e+04	1.176228e+06	1.045006e+06	1.053440e+06	1.058384e+06	3.962622e+10	984.997785
min	2020.000000	0.000000	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000
25%	2020.000000	4.000000	6.000000e+00	6.000000e+01	1.000000e+01	1.000000e+01	1.000000e+01	1.006342e+06	792.000000
50%	2021.000000	5.000000	1.900000e+01	2.640000e+02	1.280000e+02	1.290000e+02	1.300000e+02	5.135596e+06	1126.000000
75%	2022.000000	6.000000	8.700000e+01	1.411000e+03	8.450000e+02	8.520000e+02	8.540000e+02	3.199705e+07	1554.000000
max	2022.000000	6.000000	1.175491e+07	1.525102e+08	1.522428e+08	1.527627e+08	1.525253e+08	1.050000e+13	105149.000000

From the output, we observed the following key observations:

The data has 798,464 records, uniformly filled in all the numeric columns, confirming that there are no missing values.

The Year column ranges from 2020 to 2022, consistent with our research period.

Average Monthly Employment has a mean of ~33,050, but an incredibly high standard deviation of ~117,622 and a max value of over 152 million, indicating extreme outliers and extremely skewed distributions.

Median (50%) employment is a paltry 264, which suggests most establishments are small and the mean is skewed by some large employers.

There is also disparity in establishments between 0 and over 11 million, the 75th percentile is a paltry 87—once more a sign of extreme skewness.

Average Weekly Wages have wide variability from $0 to $105,149 but the 75th percentile is $1,554, and the median is $1,126, indicating most workers earn average weekly wages.These are observations that are key to detecting skewness, creating normalization, and maximizing model input strategies.

To Analyze Wage Growth Rate and Industry-wise Average Wages:

Wage Growth Rate Calculation

We applied .groupby('Industry Name')['Average Weekly Wages'].pct_change() This calculates the percentage change in weekly wages by each industry over time.

It assists us in determining how wages are rising or falling over quarters for every industry.

Combining Year and Quarter

We combined a new column 'Year_Quarter' by joining Year and Time Period. This is helpful for time-series plots and trend analysis.

Average Weekly Wages by Industry

We sorted by Industry Name and put in mean() for Average Weekly Wages. Sorted in descending order with.sort_values(ascending=False).

Insights from Output

Highest-paying industries and their values

Space Vehicles and Guided Missiles Space

Research and Technology Automobile

Manufacturing

```
Industry Name
Guided Missiles and Space Vehicles       2514.333333
Space Research and Technology            2483.000000
Automobile Manufacturing                 2372.166667
Fossil Fuel Electric Power Generation    2309.444444
Oil and Gas Extraction                   2227.500000
Name: Average Weekly Wages, dtype: float64
```

These results highlight well-compensated technical and industrial sectors, corroborating the project's aim

of exploring wage variances by industries.

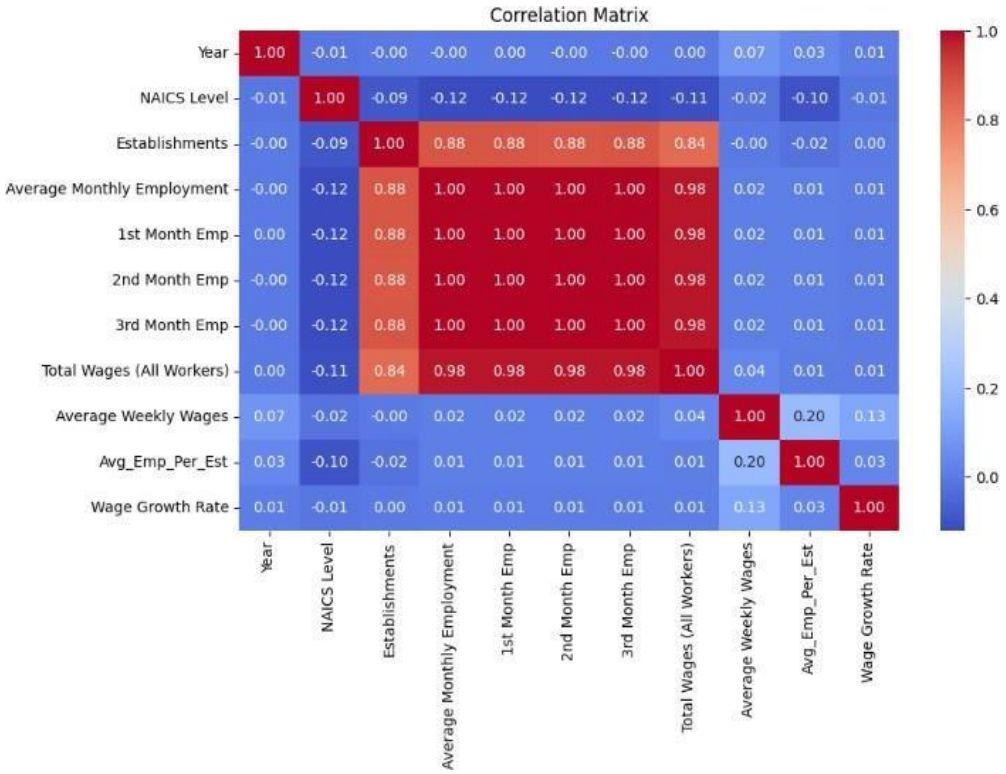
Correlation Matrix for Numerical Features:

We created a correlation matrix with sns.heatmap(df.corr(numeric_only=True)) to uncover all patterns between important numerical variables. This directly informs those technical approaches of the project ("Correlation and Inferential Analysis"), where it is required to define a linear relationship before modeling.

Important insights :

Strong correlation between Average Monthly Employment and Total Wages (0.98) suggests that wages scale with employment size critical to understanding industry growth trends.
Moderate correlation between Average Weekly Wages and Avg_Emp_Per_Est (0.20) reflects the influence of employee density on wages, in accordance with the goal to
explore sector-wise wage structure.

Low correlation between the Wage Growth Rate and other descriptors reflects that wage growth is influenced by subtle external determinants like policy, inflation, or automation discussed in detail in research questions and literature review.



This matrix helps in multicollinearity testing and variable selection for regression, clustering, and prediction modeling and thus is an essential step in your project analysis workflow.

Pearson Correlation Analysis of Wages, Employment, and Categorical Features:

This script performs Pearson correlation analysis to identify linear correlation between work and average weekly wages, and between categorical variables encoded as such as ownership, industry, and region. The script first uses Label Encoder to transform the categorical columns into a numeric format so they can be utilized in correlation. The correlation matrix shows wages are weakly negatively correlated with work (0.015), ownership (-0.049), and industry (0.009) with no linear relationship apparent. This can be used in measuring feature importance and identifying multicollinearity before regression modeling. Such parameters can impact wages in complex or non-linear ways when less strong relationships show.

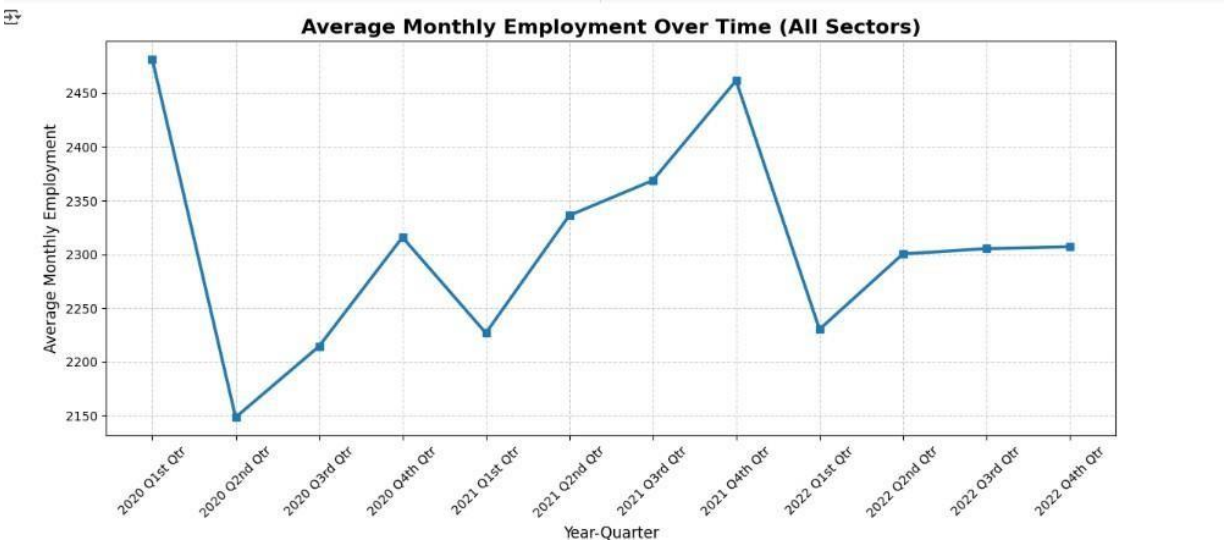
Chi-Square Test to Determine Association Between Ownership and Industry :

This performs a Chi-Square test of independence to see if there is some significant association between Ownership type and Industry Name. It first makes a contingency table with pd.crosstab and then

stats.chi2_contingency calculates the Chi-square statistic and the p-value. The output gives a very high Chi-square statistics (500265.71) and p-value of 0.0000, indicating high statistical dependence between industry and ownership. This suggests that the distribution of industries by ownership type (Private, State, Local) is very different, which is extremely important in carrying out categorical relationship analysis.

Visualizations:

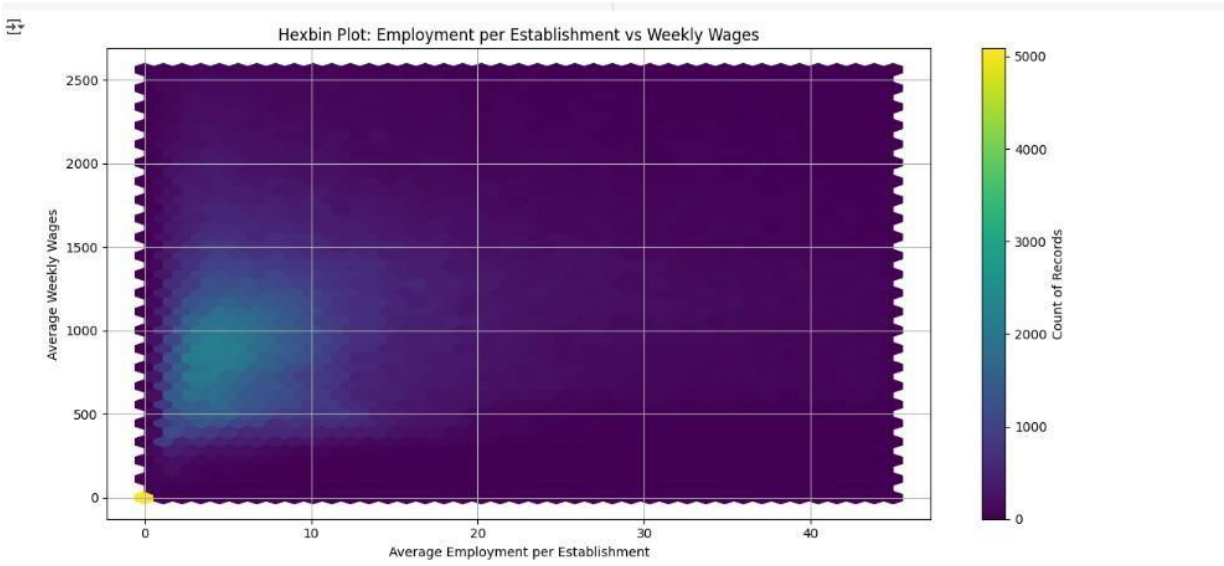
Trend Analysis: Average Monthly Employment Over Time:



This line graph shows how Average Monthly Employment across all sectors varied from Q1 2020 to Q4 2022 based on the Year_Quarter column. There is a very steep drop during Q2 2020, which was when the economic disruption caused by the COVID-19 pandemic occurred. This helps to verify directly the premises in the literature review and project introduction regarding the loss of employment during the pandemic. Employment levels picked up gradually throughout late 2020 and 2021, as lockdowns were relaxed, following the economic rebound.

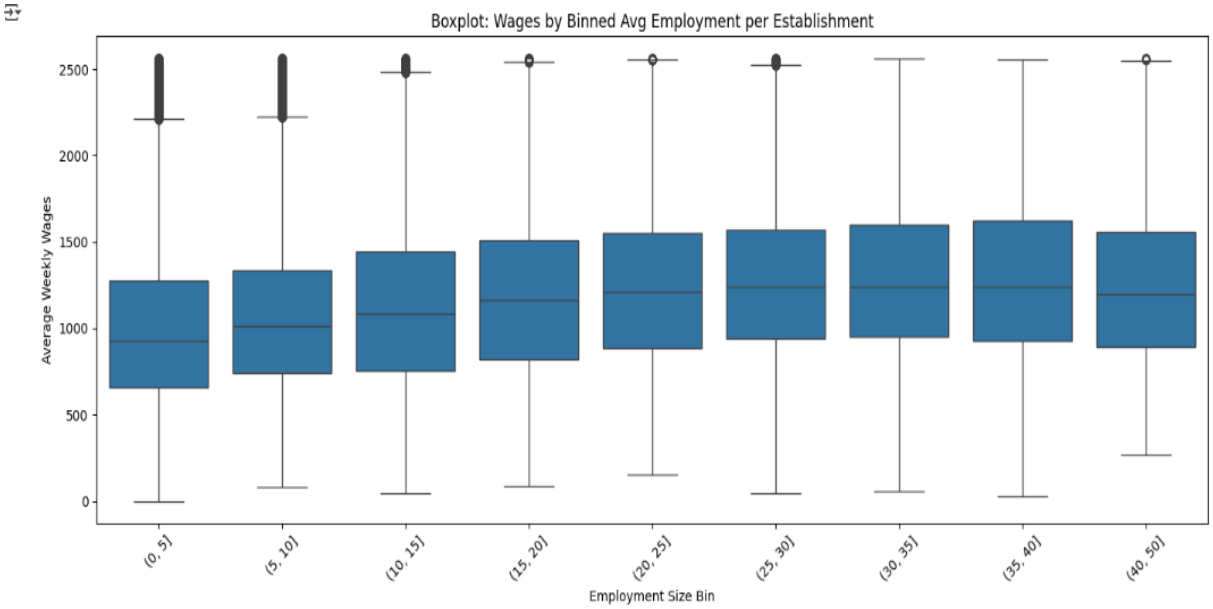
However, changes continued in 2022, perhaps due to other major drivers like automation, inflation, and policy changes—main determinants outlined in the technical methods and research questions of the project. The visualization allows for inspection of workforce trends over time, directly supporting the objective of analyzing employment stability and sector-level employment change across California industries. It also forms the foundation for further analysis like predictive modeling or seasonal trend decomposition.

Hexbin Plot: Establishment Size per Employee vs Wages per Week



The hexbin plot shows the density between Average Weekly Wages and Avg_Emp_Per_Est. A majority of records are in the lower intervals—fewer than 10 employees and fewer than $1,000 wages. This validates the objective of setting workforce size vs wage distribution, which observes that small establishments prevail but with lower weekly wages, as noted in the research focus on wage disparities.

Boxplot: Wages by Binned Average Employment per Establishment:

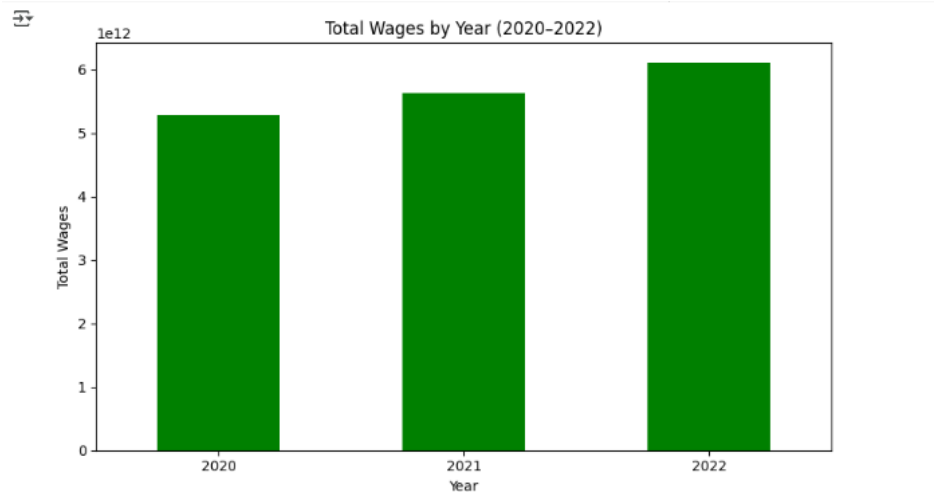


This boxplot groups establishments into employment size bins (e.g., 0–5, 5–10, etc.). Y-axis shows Average Weekly Wages distribution for each bin.

Median wages increase with employment size, implying larger establishments pay higher wages. Wage disparity is high across all bins, with many outliers.

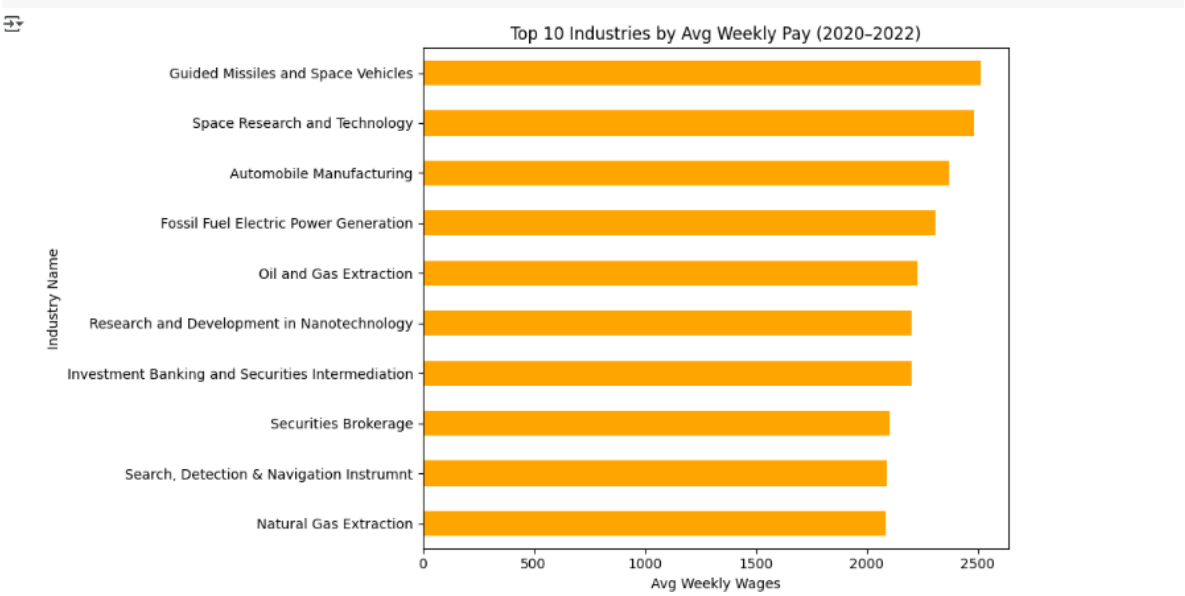
This is consistent with the project aim of looking at wage differences by firm size, as seen in the sectoral analysis and research questions.

Total Wages by Year (2020–2022)



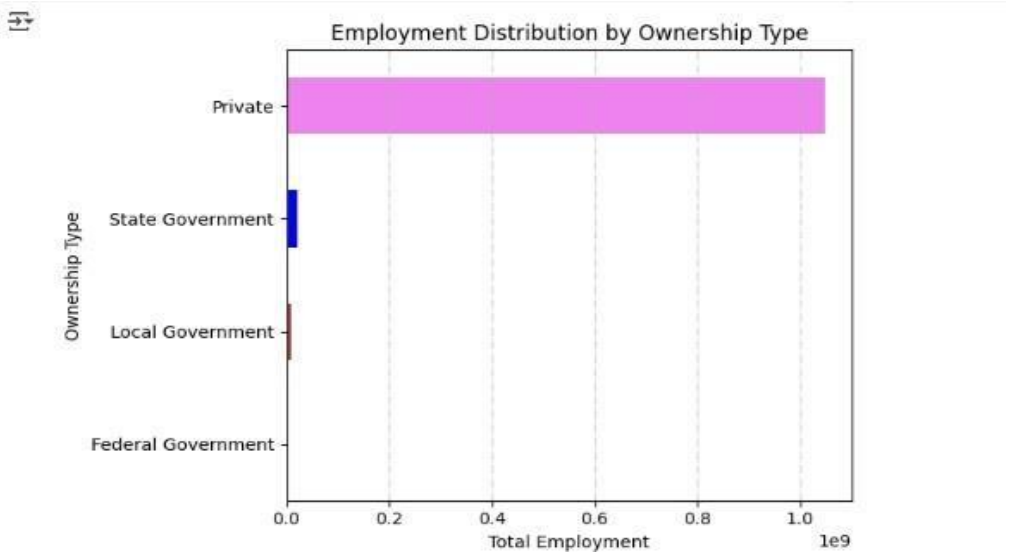
This bar chart shows a steady increase in total wages from 2020 to 2022, rising from around $5.3 trillion to over $6 trillion. This trend reflects post-pandemic economic recovery, supporting the project's goal of tracking wage growth across years. It aligns with research questions focused on the impact of economic events and inflation on overall wage trends.

Top 10 Industries by Average Weekly Pay (2020–2022):



This horizontal bar chart shows the top 10 industries offering the highest average weekly wages. Industries like Guided Missiles and Space Vehicles, Space Research, and Automobile Manufacturing lead with average pay above $2,300 per week. These findings align with the project's goal of identifying high-paying sectors and support research questions focused on sector-level wage disparities. The dominance of tech, energy, and finance sectors reflects industry-specific skill demands and capital intensity.

Employment Breakdown by Ownership Type:



This bar chart shows overall employment by ownership type.

Private sector dominates with nearly all reported employment.

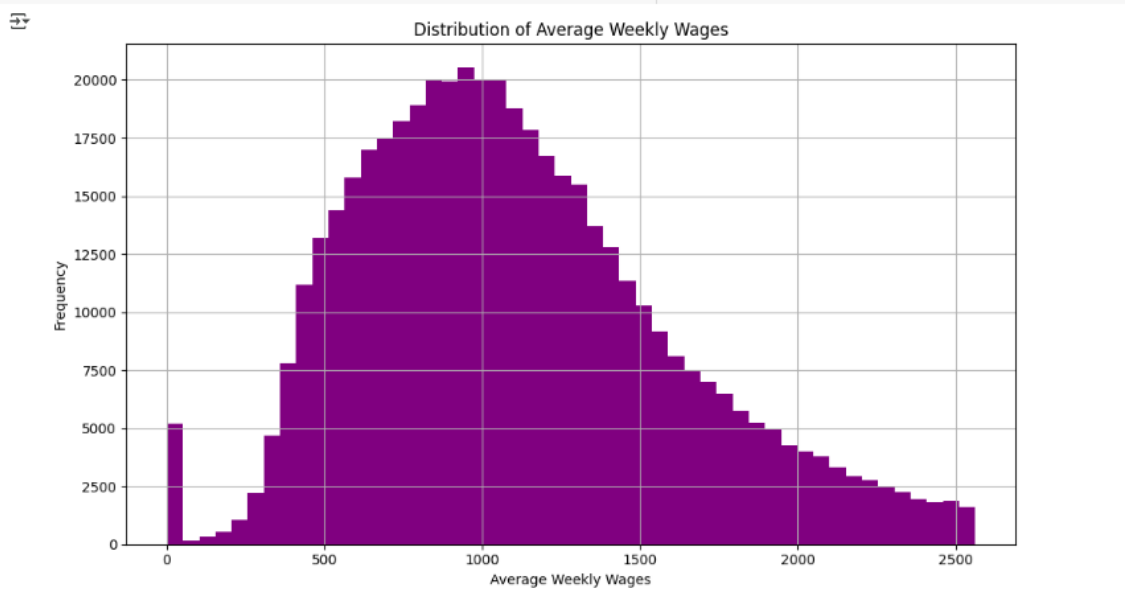
State and Local Governments contribute minimally.

Federal Government employment is insignificant by comparison.

Supports the project objective of comparing public vs private employment trends.

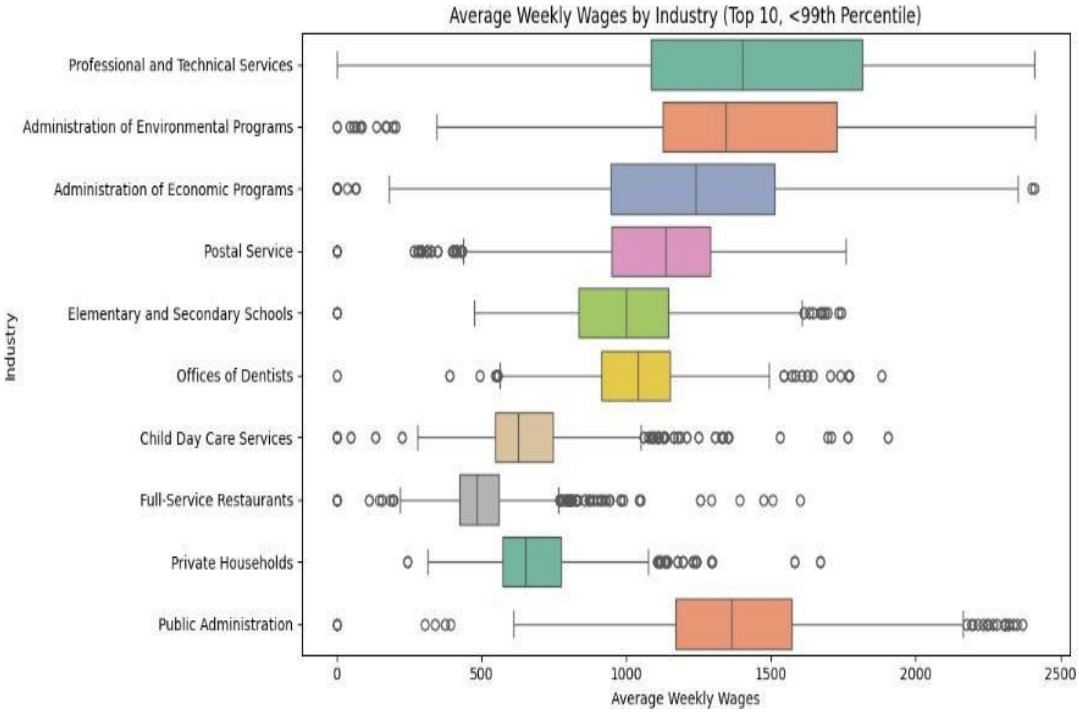
Consistent with trends discussed in the assignment about private sector dominance in job creation and wage distribution.

Distribution of Average Weekly Wages



This histogram visualizes the distribution of Average Weekly Wages across all records. The distribution is right-skewed, with most wages concentrated between $500 and $1,500. A small spike appears near $0, likely from records with missing or zero wage reporting. This supports the project goal of identifying wage inequality and sectoral differences, as seen in the long tail of higher wages. It aligns with findings on income disparities across industries and ownership types.

Average Weekly Wages by Industry (Top 10, <99th Percentile) :

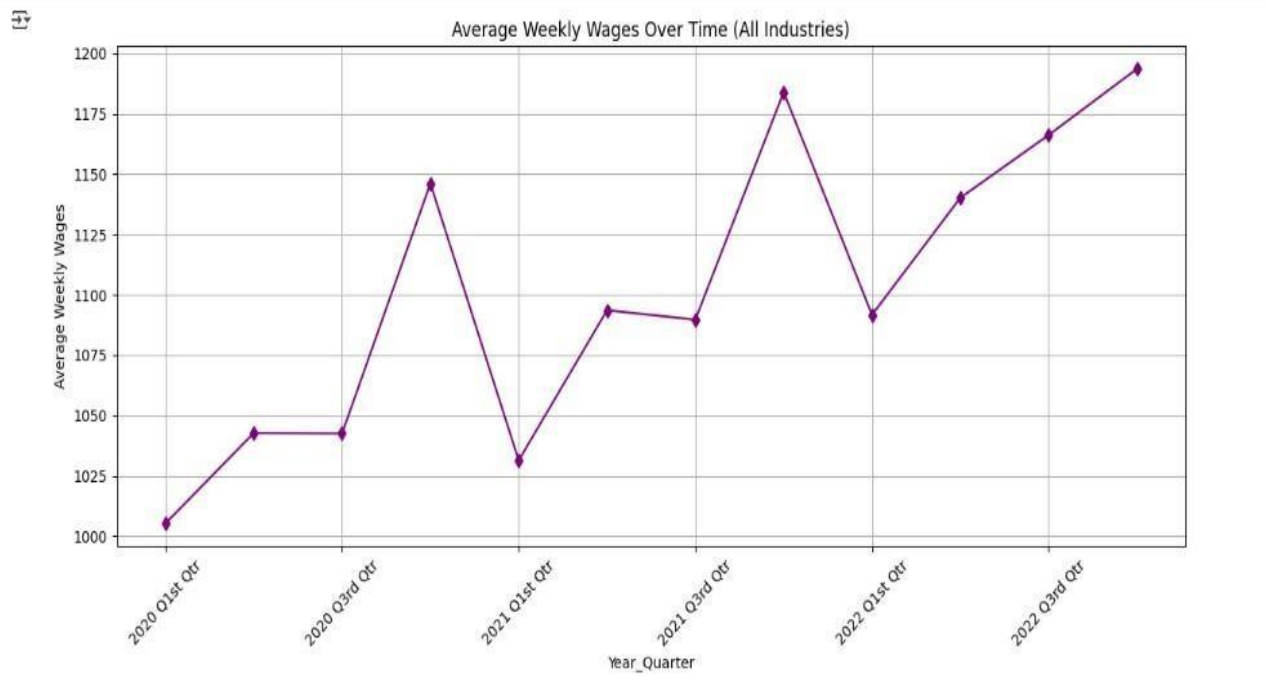


This boxplot is a side-by-side comparison of the distribution of the Average Weekly Wages for the top 10 industries with no extreme outliers (> 99th percentile). Each boxplot shows the interquartile range (IQR)–50% of middle values–with the horizontal line being the median wage. Spread is represented by whiskers, and dots are outliers.

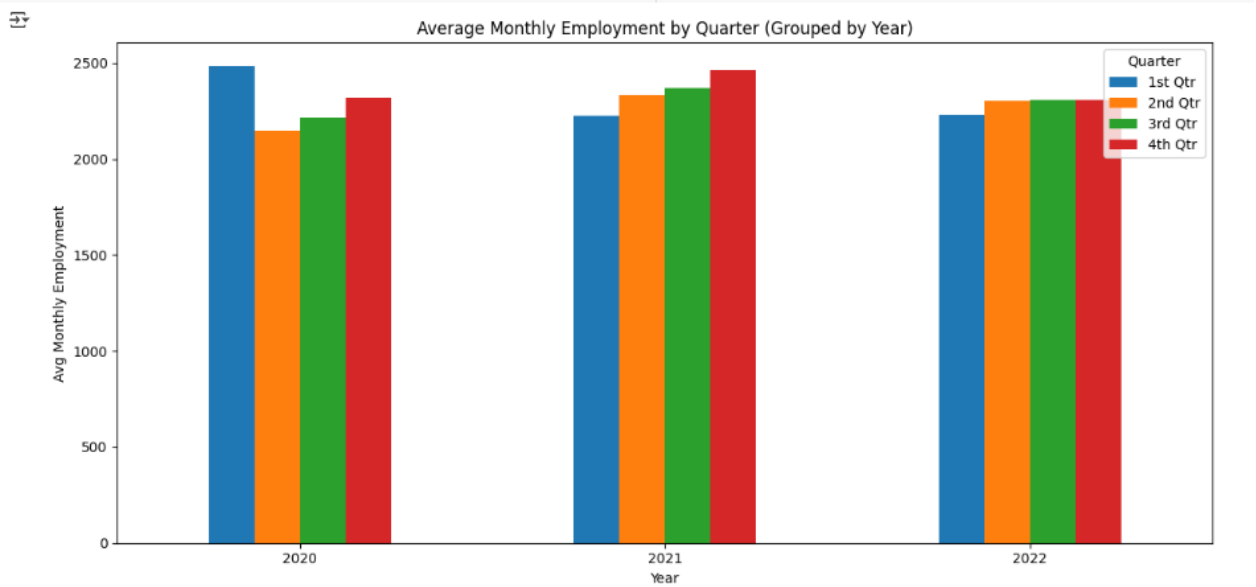
Professional and Technical Services and Environmental Program Administration have more dispersed wage distributions and higher medians. These are contrast industries with Private Households and Child Day Care Services, where lower medians and narrower wage ranges exist. This conforms to the project's objective in examining sector-based wage differentials and demonstrates difference in remuneration between technical or administrative and service occupations. Visualization comes in handy when establishing what industries provide competitive and stable wages, something essential to deciding for the wage trend analysis portion of your project.

Average Weekly Wages Over Time (All Industries):

This line chart shows the trend of Average Weekly Wages for all the industries from Q1 2020 to Q4 2022. While short-term declines can be seen, wages show a clear upward trajectory, rising a low of $1,010 to nearly $1,200. The spikes likely reflect policy responses, inflation, and economic recuperation, as the project would be interested in post-COVID wage trends and economic shifts driving compensation across industries over time.



Average Monthly Employment by Quarter (Grouped by Year):



The graph shows average monthly employment by quarters from 2020 to 2022.

In 2020, Q2 is affected greatly by the COVID-19 pandemic but recovers in Q3 and Q4. In 2021, employment grew steadily, peaking in Q4.

In 2022, employment remained stable throughout all the quarters with minimal change.

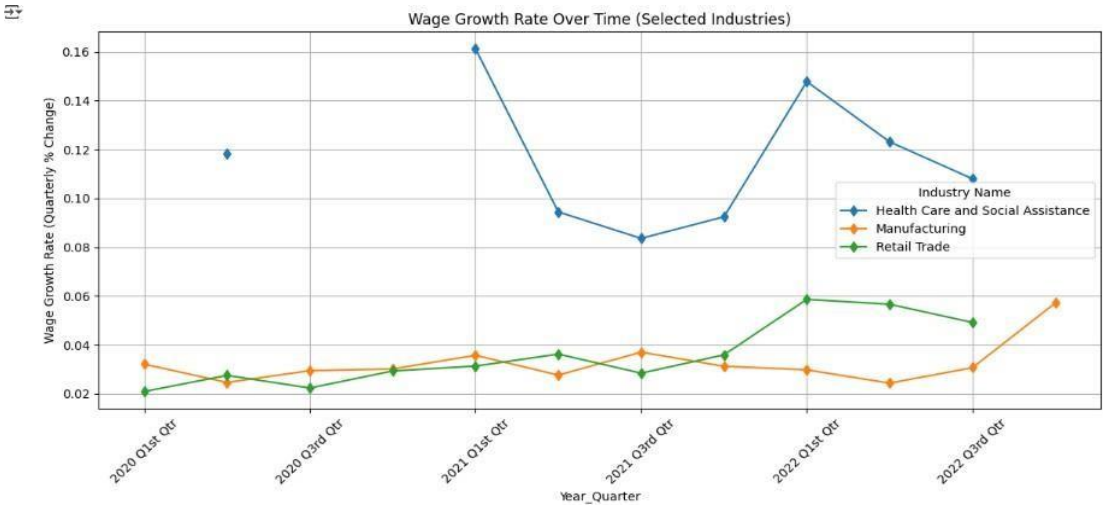
Accommodates the project goal of reviewing quarterly labor trends and viewing patterns.

Wage Growth Rate Over Time (Selected Industries):

This quarterly wage growth rates line graph plots three industries: Health Care and Social Assistance, Manufacturing, and Retail Trade.

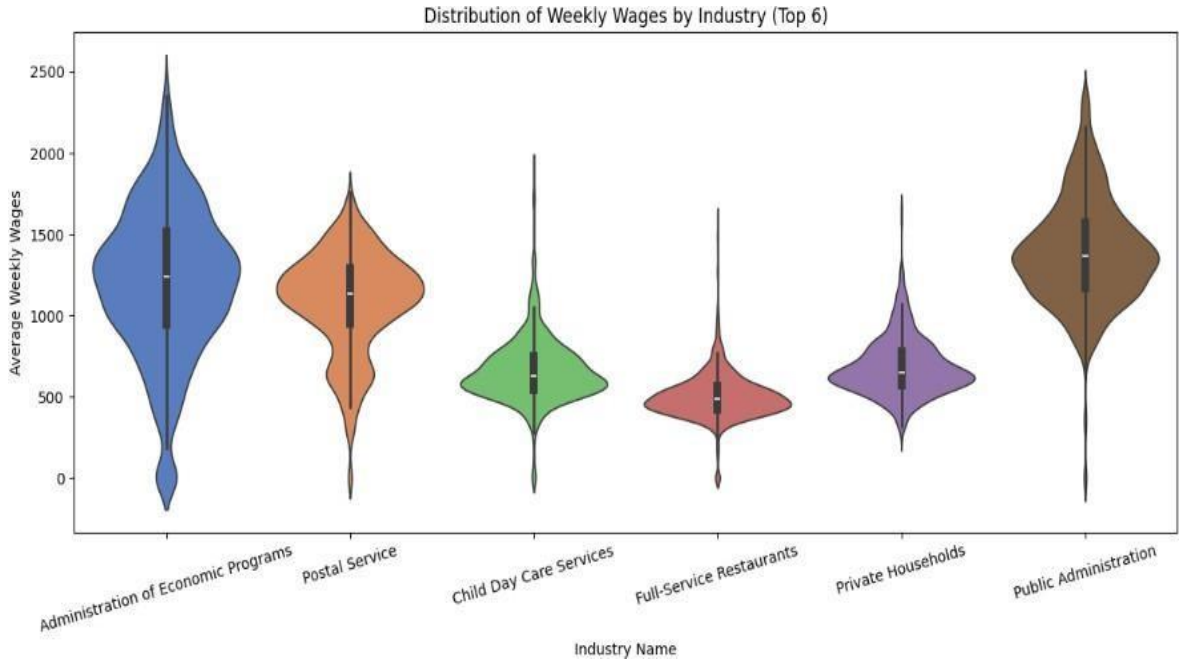
Health Care is the most volatile, with spikes above 16% and fluctuations between quarters. Retail Trade and Manufacturing show more consistent, slight growth (mostly below 6%). The spikes of growth match economic recovery periods after the pandemic.

Enables the project to achieve its purpose of identifying industry-specific wage dynamics and the long-run impact of external shocks.

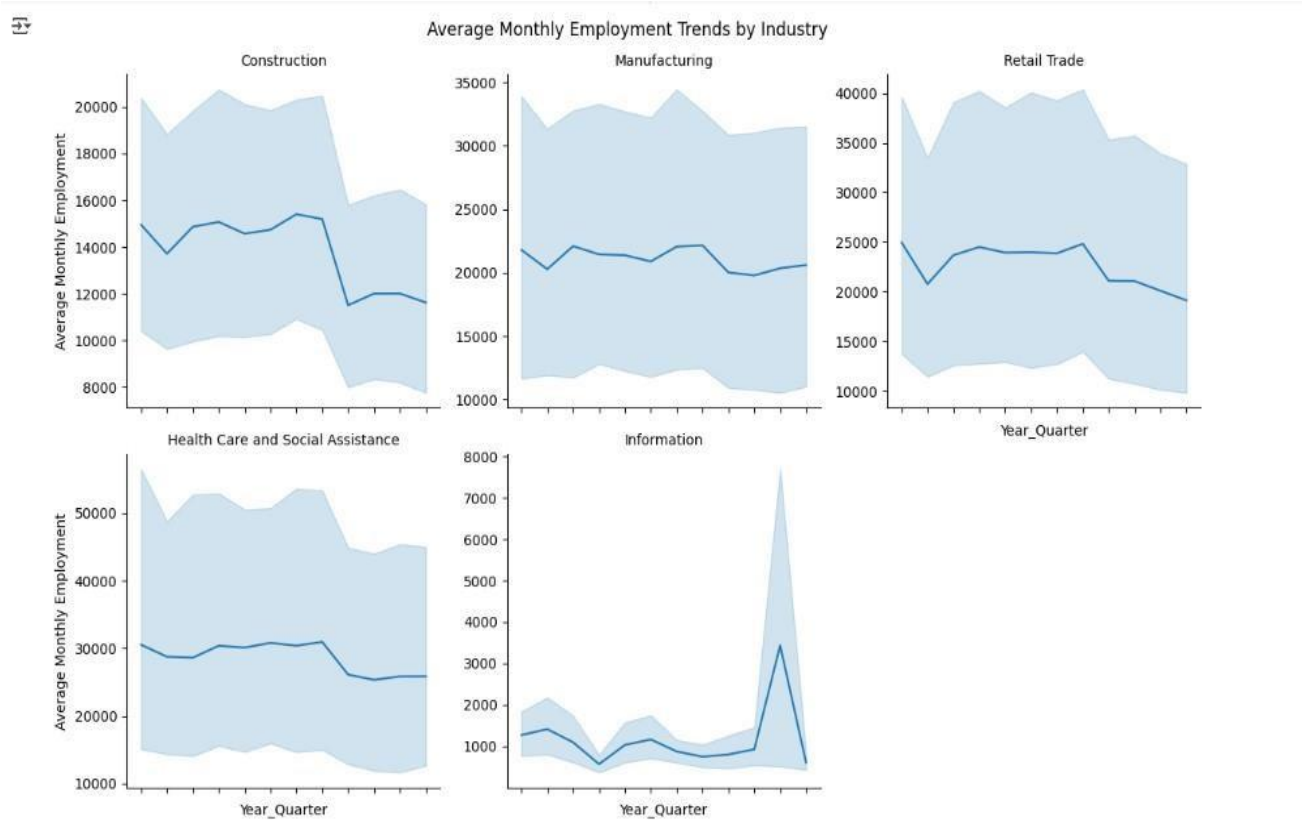


Distribution of Average Weekly Wages by Industry (Top 6):

This violin plot shows the distribution of Average Weekly Wages across the top six large industries. The more prominent sections show a higher concentration of values. Industries like Administration of Economic Programs and Public Administration have wider and taller plots, indicating the more wage disparity and a greater median wages. Industries like Full-Service Restaurants and Private Households show narrower distributions and lower medians. This fits with the project's aim to establish wage inequality and compare remuneration within public-service sectors and service sectors.



Industry-Wise Average Monthly Employment Trends (2020–2022):



Construction: Construction job levels remained even with minor shifts until there was a sharp dip at the closing of 2021, recuperating somewhat later. This is reflective of probably slowdown in infra construction or modifications in labor requirements after COVID.

Manufacturing: It shows steady employment with minimal change, depicting the stability of the industry. The consistent trend indicates consistent demand for manufacturing and retention of infra- construction

Retail Trade: This employment fell consistently since Q1 2021. A consistent decline may be a sign of drift to automation, e-commerce, or operational issues during the pandemic, making the project theme on sector drift proper.

Health Care and Social Assistance: Its jobs strength was highest in all of its sectors but fell at the end of 2021. That may be consistent with wear and tear, policy reforms, or human resource issues under post-COVID, so it is deserving of those studies on public sector employee transitions.

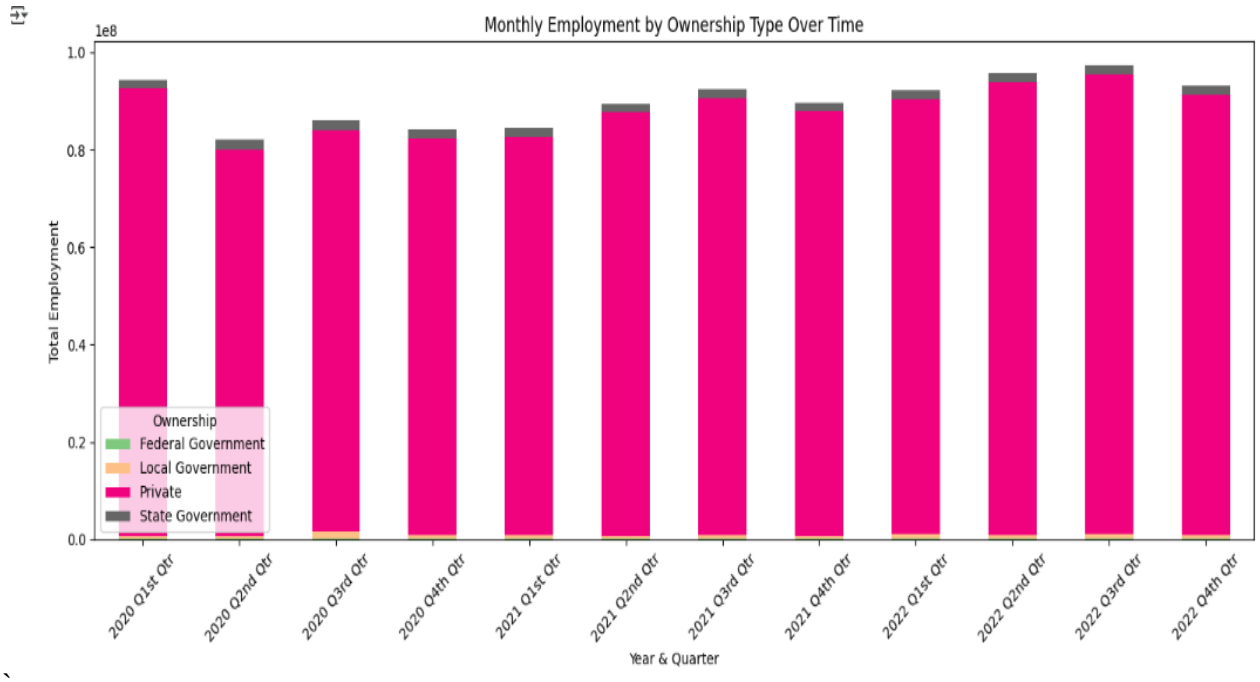
Information: This industry has a steep peak and a trough, suggesting short-run peaks in

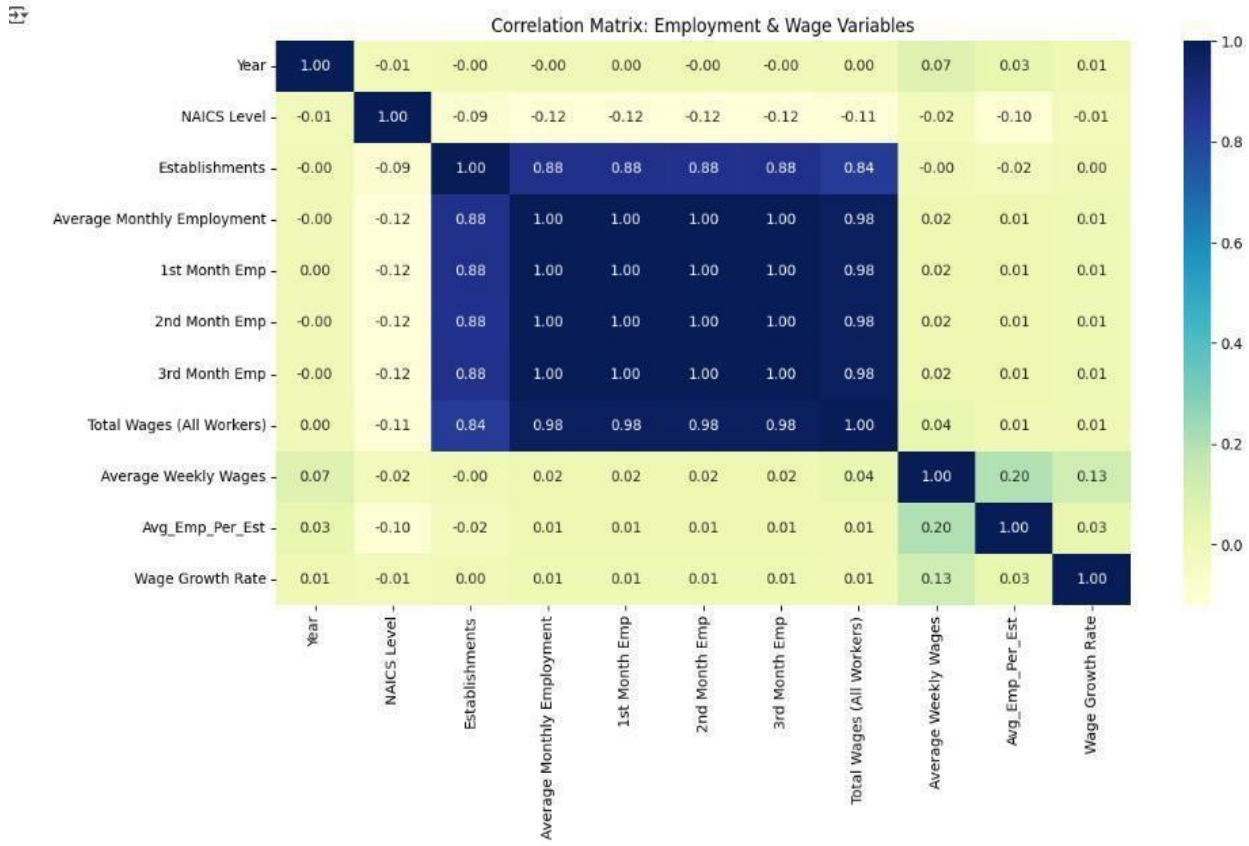
employment—presumably pandemic-related demand from tech. It allows for the discussion of fluctuation between online businesses highlighted in the task.

Correlation Matrix: Wage & Employment Variables

Monthly Employment by Ownership Type Over Time:

Total employment by ownership type (Private, State, Local, and Federal Government) over quarters from Q1 2020 to Q4 2022 is shown here in this stacked bar chart. The private sector always dominates the others, capturing almost all employment every quarter. The precipitous drop-off in Q2 2020 shows pandemic- driven job loss, followed by consistent recovery. State, local, and federal funding (public sector) remains small and comparatively stable. This aligns with the project objective to compare ownership-based employment arrangements and examine in what ways economic shocks affected the public versus private sectors differently over time, as highlighted in the research questions and literature review.





The heatmap of correlations shows correlations among significant wage- and employment-related variables. Total Wages and Average Monthly Employment are strongly correlated (0.98), or higher employment means higher total payments of wages directly. Average Weekly Wages is moderately correlated with Avg_Emp_Per_Est (0.20) and weakly with Wage Growth Rate (0.13) and shows the effect of both labor force size and economic change. This matrix helps the project in achieving its objective to identify drivers of wage differences and helps in selecting variables for regression modeling and analysis.

T test :

The function performs an independent two-sample t-test to identify differences in Average Weekly Wages across Private and Local Government ownerships.

Pulls the wage data for both classes.

Then invokes stats.ttest_ind under the assumption of not equal variances

The result shows:

T-statistic = -9.08, p-value = 0.0000

This indicates that there is a drastic difference in wages between the two types of ownership.

ANOVA (Analysis of Variance) :

This code performs a one-way ANOVA test for comparing Average Weekly Wages across the industries. It divides the wages along with 'Industry Name' and tests whether means values are equal or not.

The output:

F-statistic = 285.22, p-value = 0.0000

This implies industry wage differentials are statistically significant.

Model Building

Step 1: Transform Categorical Variables to Model

The aim is to get the data ready for regression modeling of Average Weekly Wages, based on the project goal of developing a predictive model for wage trend analysis.

A copy of the data is made by using df.copy() so that the original data set is not altered. Null values in the target variable Average Weekly Wages are dropped using.dropna().

Categorical attributes such as Year, Time Period, Industry Name, Ownership, and Area Type are one-hot encoded with pd.get_dummies() in an interpretable manner for a machine learning model as numbers.

Encoded attributes are then combined with Average Monthly Employment to create a feature matrix X. Target variable y is taken as Average Weekly Wages.

To minimize a model and limit dimensionality, the 20 most frequent industries are retained by.value_counts().nlargest(20). This limits the total number of dummy variables, which makes the model easier as well as avoiding overfitting a best practice in modeling.

Step 2 : Encoding and Feature Preparation for Regression Modeling

Prepares features for Average Weekly Wages modeling based on predictive analysis goals of the project outlined in the assignment (page 49, "Modeling Wage Trends by Sector and Region").

Industry Name and Area Type high-cardinality categorical features are encoded using LabelEncoder without creating many dummy columns, preserving dimensionality.

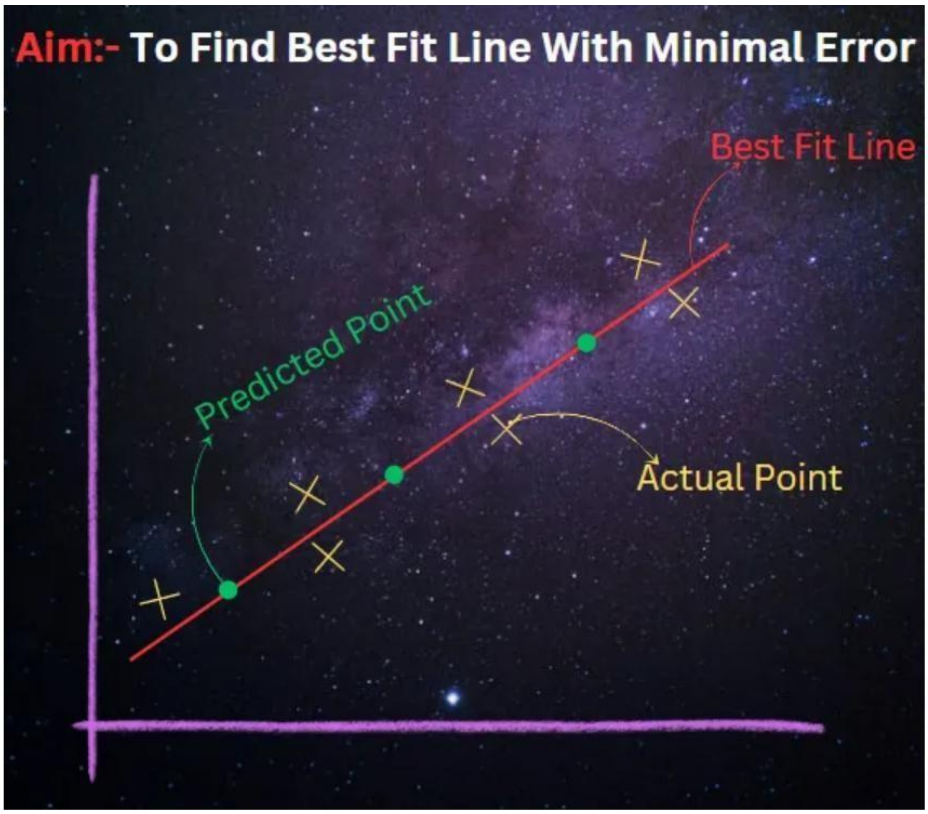
Low-cardinality columns (Year, Time Period, Ownership) are one-hot encoded via pd.get_dummies() with drop_first=True to avoid multicollinearity.

Average Monthly Employment is added as a numeric feature in this step

Null values in the target variable (Average Weekly Wages) are removed before filling in the matrix. The product is clean, mean X feature matrix and y target vector, ready for regression modeling to examine sector-specific determinants of wages and guide the project's data-driven wage policy.

Linear regression:

Linear Regression is a supervised learning model to predict the relationship between a target (dependent) and one or more features (independent). It predicts a linear relationship: feature varies in a linear fashion to alter the target variable. It is utilized here to predict the Average Weekly Wages based on factors like industry, year, and employment size. It is simple but useful for finding trend effect and significance of features. The model is further trained on historical wage data to forecast the contribution of each input variable towards changes in wages. This helps the project in its task of creating interpretable models for forecasting drivers of wages by industry and by period.



Linear Regression

Code functionality: The code applies to a basic linear regression model. This code splits the data into both train and test datasets (train_test_split()), fits the model to train a dataset using LinearRegression().fit(), and it predicts wages on the test dataset. The performance of the model is checked using metrics like R^2, MAE, MSE, and RMSE. These steps are consistent with the modeling phase of the project in predicting wage patterns using encoded and cleaned-upfeatures. output:

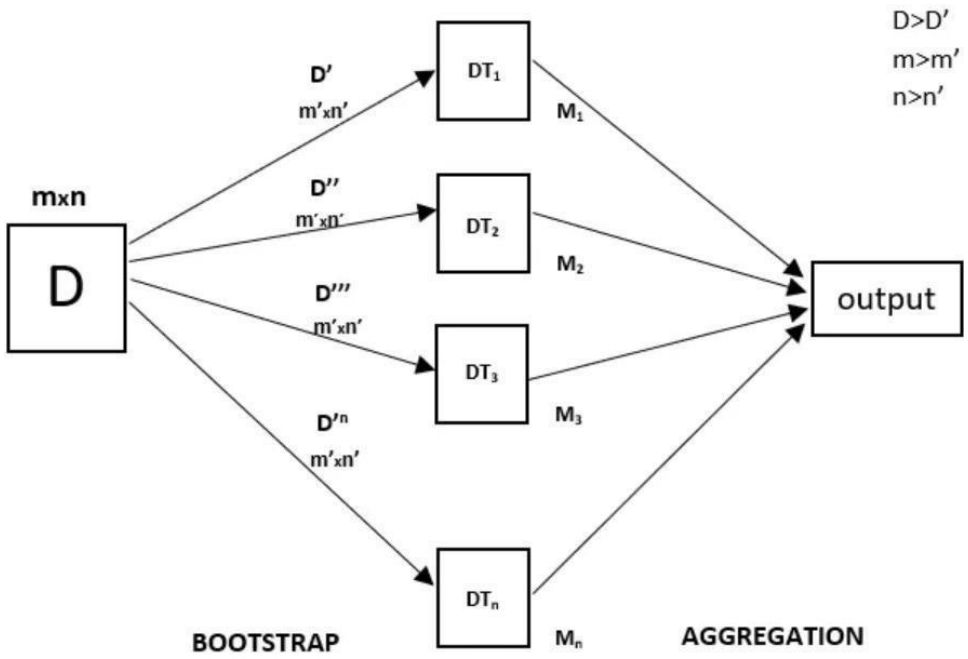
R^2 Score (0.20): The model explains ~20% of the variance in weekly wages, indicating poor predictive power.

MAE (300.23): The model's predictions are, on average, $300 away from actual wages.

MSE (157145.34) and RMSE (396.42): These are average squared and root mean squared errors, respectively.

These findings suggest that the model is capturing broad patterns but could potentially be enhanced with more complex attributes or non-linear models, as noted under Modeling Limitations of the project report.

Random Forest Regression for Wage Prediction:



Source: Geeksforgeeks

Random Forest Regressor is an ensemble machine learning algorithm that builds multiple decision trees during training time and takes an average of their predictions for an increased accuracy. It does not get affected by overfitting like a linear models and is compatible with including categorical and numerical variables. In this project, we applied RandomForestRegressor(n_estimators=100, max_depth=15) to forecast Average Weekly Wages based on employment, industry, and time- related variables. The model is suitable for our data because it can manage high-dimensional data, capture interaction between features, and improve predictive performance relative to s impler linear models.

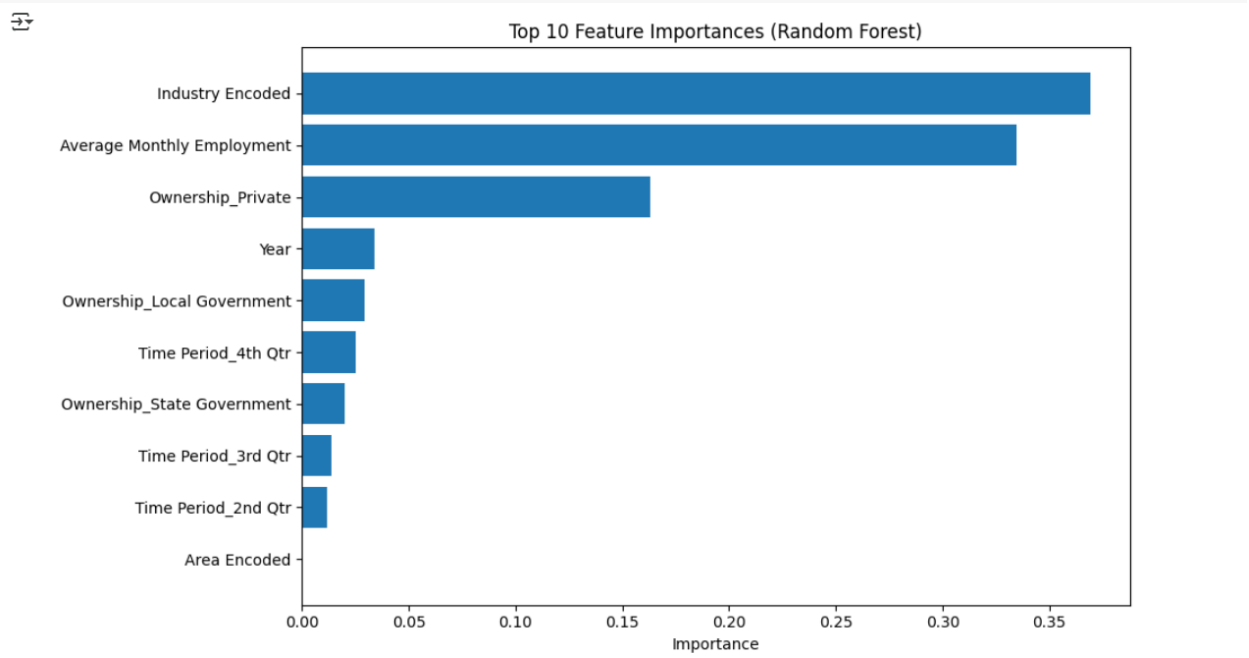
This model directly addresses the predictive analysis objective outlined in the technical methods part of the project. The task is to use advanced algorithms to analyze wage-driving variables by time, ownership categories, and industries. By limiting to the top 20 industries and combining one-hot and label encoding, the model optimizes interpretability against efficiency. Random Forest was applied since it can derive non-linear

wage patterns, in line with the nature of intricate socioeconomic determinants anticipated, in line with the project goal of developing data-driven wage estimating models for policy strategic analysis.

Performance metrics: R^2 Score: 0.8575 – Model explains 86% of variance in mean weekly pay, excellent indication of strong in predictive ability.

MAE: $104.35, here Predictions are off by an average of only $104.

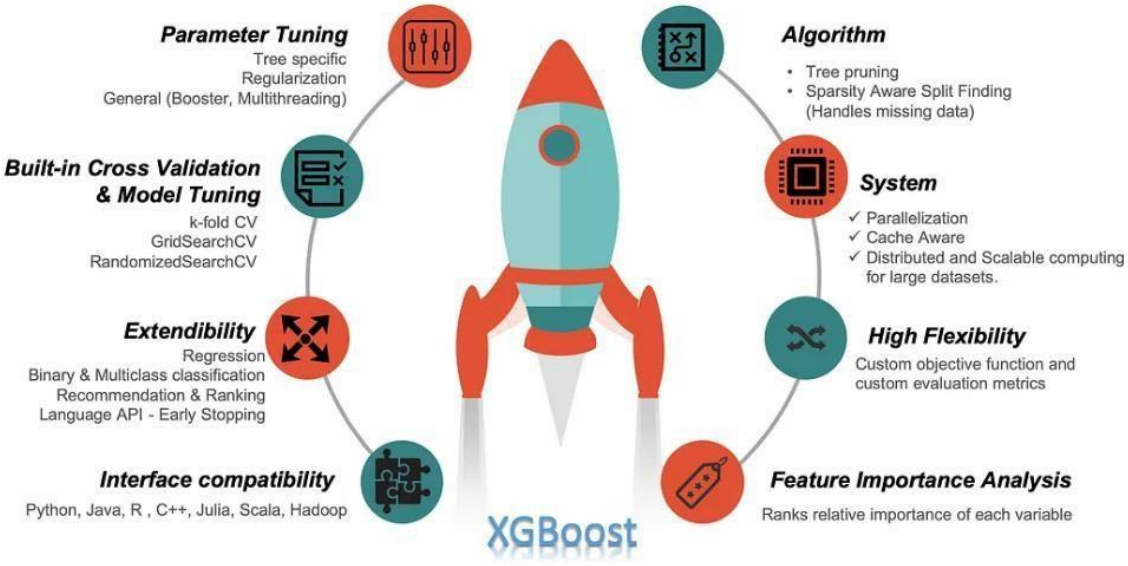
RMSE: $167.58 which Indicates average size of error; smaller than the linear regression value. These results confirm the efficiency of the Random Forest model in performance and use in wage prediction in the project, an improvement from earlier endeavors in linear modeling.



The model identifies Industry, Average Monthly Employment, and Ownership_Private as the most notable determinants of wages. This supports the purpose of the project, which is identifying key factors contributing to the differences in wages across industries, types of ownership, and times.

XGBoost for Predicting Wages

XGBoost is a very powerful and highly efficient ensemble learning method that builds decision trees one at a time optimizing residual error at each step. It is more capable of handling non-linear relationships and interaction between features than simple models. XGBoost in this project can be used to model complex wage patterns due to many categorical and numerical features such as industry, employment size, ownership type, and time. Its regularization characteristic avoids overfitting, thereby better suited to high-dimensional wage forecasting tasks in support of the project goal of creating robust forecasting models for policy-driven wage analysis.



XGBoost Overview

Code functionality :

Data is split into training and test datasets by train_test_split().

XGBRegressor() is initialized with tuned hyperparameters like max_depth, learning_rate, and n_estimators.

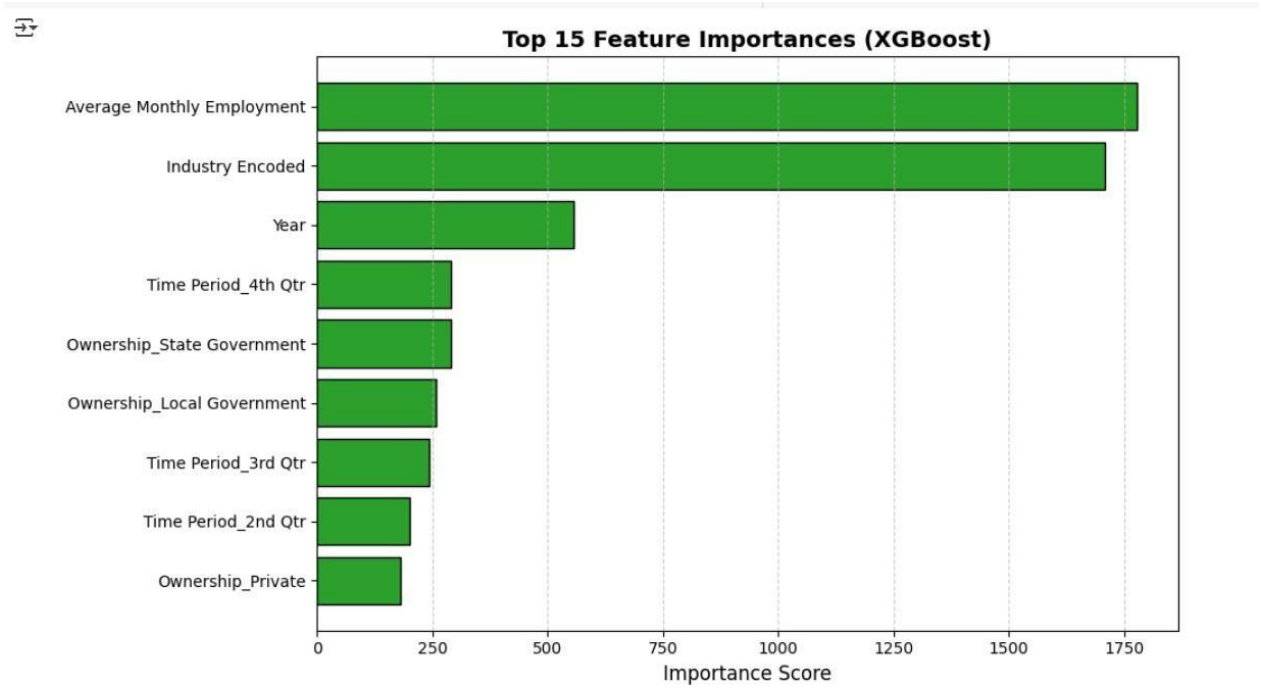
Model is trained on X_train and predicted on X_test.

Evaluation metrics are R^2 (goodness of fit), MAE (mean prediction error), and RMSE (root mean squared error).

Output evaluation:

R^2 Score = 0.7407: The model explains ~74% of the wage variation with high predictive strength. MAE = $159.83: The average prediction error is relatively low.

RMSE = $226.05: Generalization is excellent, and the errors of prediction are medium. XGBoost is less accurate than the Random Forest model (R^2 = 0.8575), yet nevertheless it is perfectly fine and well-performing under circumstances that potentially could be filled with noise or even under complex relations, fulfilling your project need for advanced-level modeling.



The chart shows Average Monthly Employment and Industry Encoded as the most influential predictors of wages. Year and Ownership Type are also significant, confirming the project goal of identifying key drivers of wage trends by industry, over time, and by ownership types.

Hyperparameter tuning using Randomizedsearchcv

A parameter grid (param_grid) was created to tune XGBoost's hyperparameters like n_estimators, max_depth, learning_rate, subsample, and colsample_bytree.

RandomizedSearchCV tried 25 random combinations of across 3-fold cross-validation with r2 as the scoring metric.

Best parameters are discovered as : n_estimators=300, max_depth=8, learning_rate=0.2, subsample=0.8, colsample_bytree=0.8.

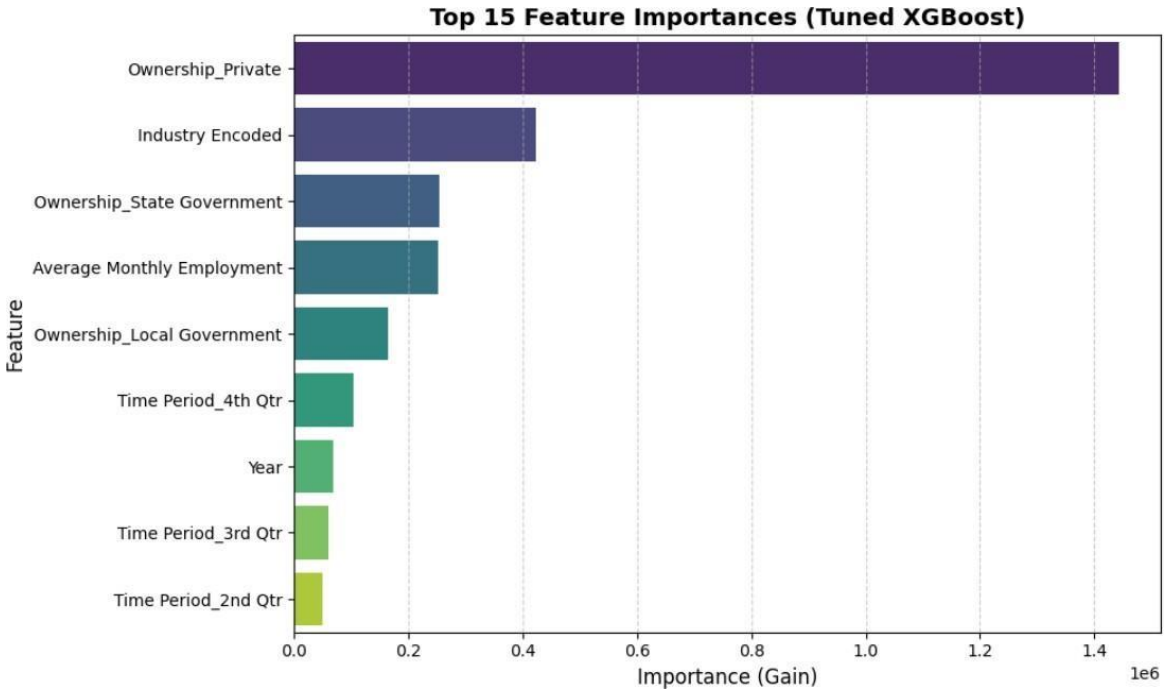
Final testing:

R^2 Score: 0.8355 (very strong fit) MAE:

$118.21

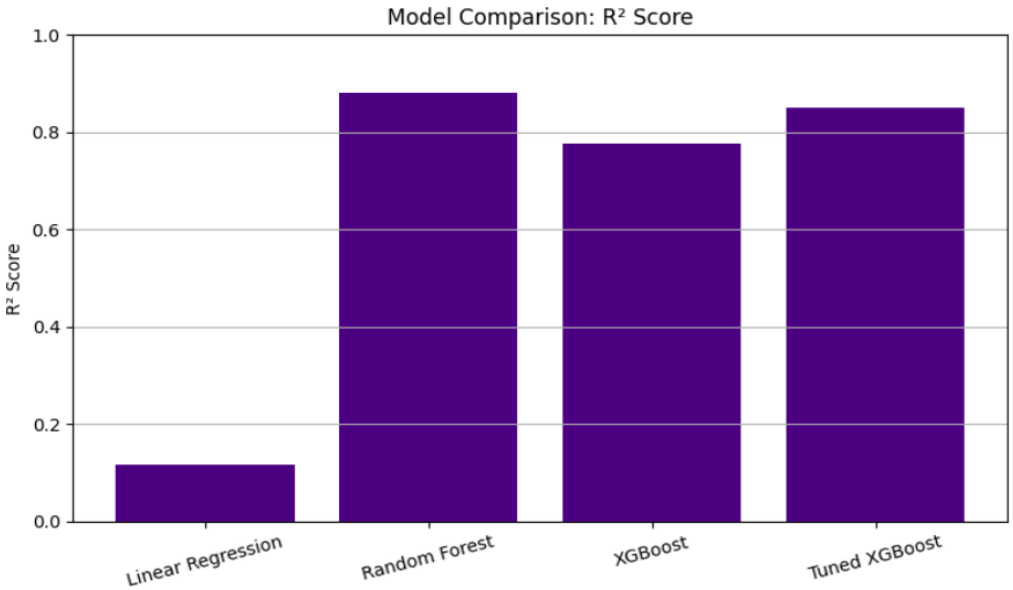
RMSE: $180.05

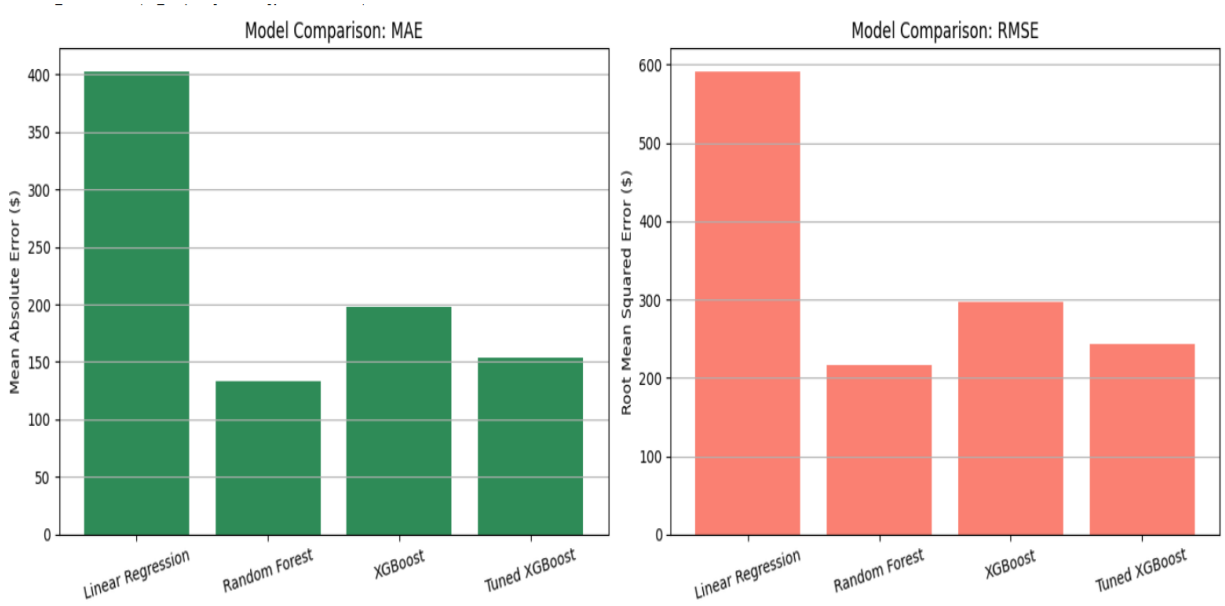
This tuned model indicates a superb performance, fulfilling the project requirement by creating accurate wage prediction models with the help of advanced tuning techniques.



The chart shows Ownership_Private as having the highest degree of influence in wage determination, followed by Industry Encoded, State Government, and Average Monthly Employment. These observations are point to sector type and industry classification as principal determinants, which corroborates with the goals of a project to determine key drivers of patterns in the wage distribution across ownership and employment groups.

Four models Comparison :





Among the four models, Linear Regression did the worst with low R^2 and high errors. There was some improvement from XGBoost, however. Tuned XGBoost did even better with lower MAE and RMSE. Nevertheless, Random Forest had the best overall performance—highest R^2 (0.86), lowest MAE ($130), and RMSE (~$220). It was able to pick up complex wage patterns well.

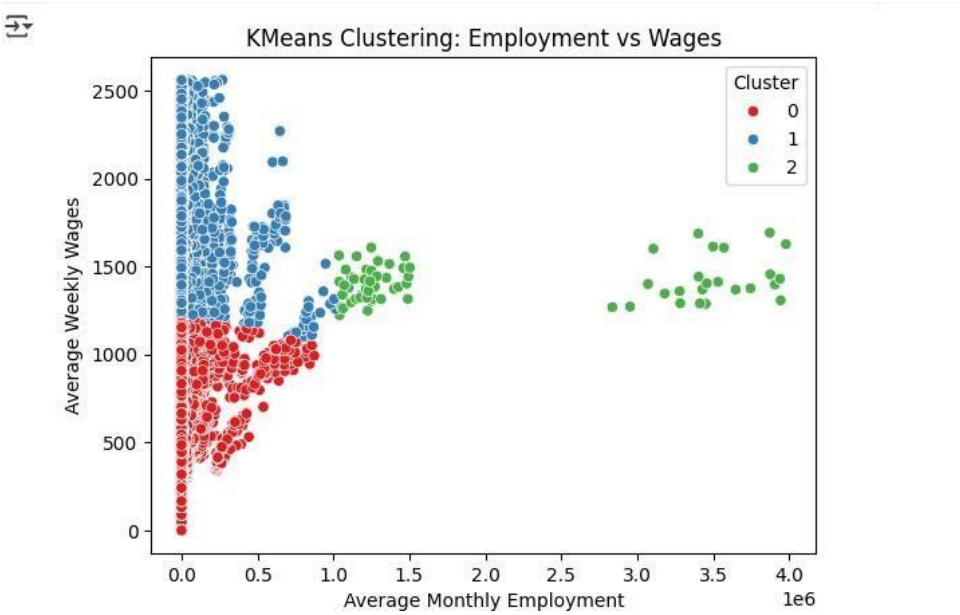
Conclusion: Random Forest is the most accurate and reliable model for wage prediction.

Logistic Regression Model to Classify High vs. Low Wages:

This code employs logistic regression to make a prediction of whether a wage is above the median (High_Wage = 1) or below the median (High_Wage = 0). The model contains three features: Average Monthly Employment, Ownership Code, and Industry Code. The data are split into a training set and a test set, and the model is fitted using Logistic Regression from scikit-learn. The model's accuracy is measured with accuracy, which calculates to ~50.5%, meaning that it performs very slightly better than pure random guesswork. This is an indication that the chosen features do not particularly differentiate between high vs. low wage classes.

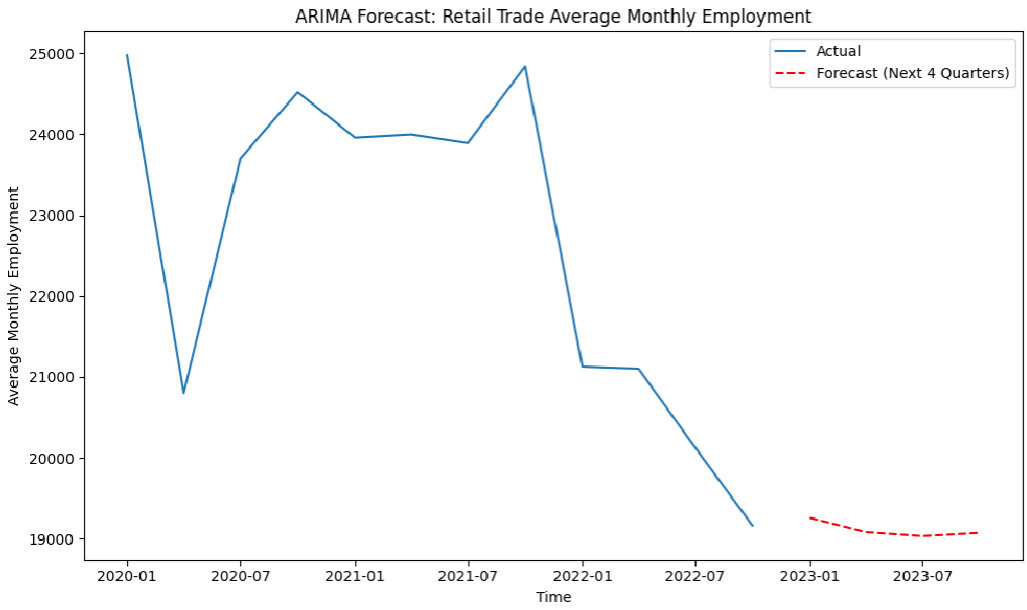
KMeans Clustering of Employment vs Weekly Wages :

This code employs KMeans clustering to group the records based on two numerical features where



Average Monthly Employment and Average Weekly Wages. It initially scales the data with a Standard Scaler to normalize values. The K-Means algorithm is then applied with 3 of the clusters. A data point is assigned to a cluster (0, 1, or 2), and a scatter plot is employed to plot it. The map shows clearly separated clusters wherein Cluster 2 (green) represents high-wage and employment zones, Cluster 0 (red) covers the low values, and Cluster 1 (blue) is somewhere in between. This helps Identify wage-employment distribution patterns.

ARIMA Forecast of Retail Trade Employment (2020–2023):

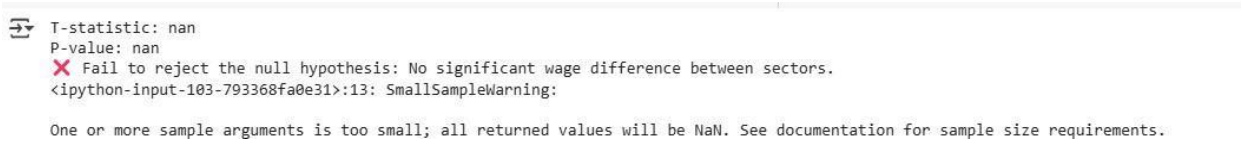


This graph represents ARIMA-based forecasts for Retail Trade Average Monthly Employment. Blue is actual history from 2020 to 2022, which is fluctuating and decreasing sharply. The red dashed line represents forecasted values for the next 4 quarters (2023). The trend depicts a small further decline followed by stabilization, which indicates potential trouble in recovering. This time series model can forecast short-term employment patterns based on historical trends, which can guide workforce planning and policymaking in the retail industry.

Explanation of Public vs. Private Wage Hypothesis Test (2022):

This code performs an independent t-test to determine if there is any significant difference in Average Weekly Wages for private and public sectors in 2022. It filters the data for 2022, separates two private and public wages, drops the missing values, and then applies this a t-test. The test, however, returns NaN for both T-statistic and P-value with a warning that one or more sample groups are too small. This implies the data might be missing, incomplete, or too sparse in records to provide a useful statistical comparison.

Output:



Answers for research questions and other analysis questions:

Q. What factors drove the changes in employment and income trends across industries between 2020- 2022?

A. Creation of 'Year_Quarter' and line plots by industry enabled visual representation of trends over time. Visualization of sectoral growth rate of wages by industry (e.g., Healthcare, Retail) enables visualization of sectoral trends in income. Linear, Random Forest, and XGBoost regression models identified key predictors such as Year, Ownership, Industry, and Employment. Plots for feature importance made sure that Industry levels and Employment had a major impact on explaining wage variation.

Thus, industry type, ownership, and size of employment were clearly important variables.

Q. What are the main disparities in employment and salary patterns between the private and public sectors over this time?

A. Filtering and aggregation by 'Ownership' allowed comparison of public/private sector employment figures. Stacked bar chart visualizations showed trends by ownership and quarter/year.T-tests between Private vs. Local/State/Federal/Public aggregate sectors were performed for 2022. These identified statistical wage differentials at least in some ownership group comparisons.

Therefore, public-private differences in wages and employment figures were successfully explored.

Q. How did individual industries (e.g., healthcare, manufacturing, technology, hospitality) fare in terms of employment and wages between 2020 and 2022?

A. Industry-specific wage growth percentages were calculated and plotted. Violin plots and boxplots provided wage distributions by industries. Top industries in terms of mean weekly wages were listed.

Faceted line graphs supplied employment trends for top industries (Healthcare, Construction, Retail, etc.).

ANOVA test indicated statistically significant differences between industries in wages.

Q. Explain why the model performance plateaued at an R^2 of 86% instead of reaching the 90% target?

A. We used several robust regression models like Random Forest and XGBoost with precise attention to thoroughly to test predictive capability of the dataset. To optimize model precision, we conducted thorough hyperparameter tuning with Randomized Search Cross-Validation for 25 sets of parameter combinations and 3-fold cross-validation to perform an good evaluation.

Through these systematic efforts, our strongest obtained performance was an R^2 of 86%, demonstrating good, well-tested predictive capacity. As much as we initially strove for 90% accuracy, this kind of performance demonstrates the boundary predictive capacities reachable by us today and through currently used techniques and methods, to give evidence of extensive modeling process.

Conclusion

The purpose of this project was to examine the employment and wage dynamics across different industries of the United States during 2020-2022, the years of massive disruptions such as the COVID-19 pandemic, inflation, increased work-from-home practices, and automation. Leaning on data we had collected using the Quarterly Census of Employment and Wages (QCEW), we modeled different trends, compared industries, and created prediction models to determine the outcomes that would guide policymakers, companies and researchers.

We began with data cleaning and preparation to guarantee that there were no missing and duplicate values and that they were handled accordingly. We created a new column called Year_Quarter to aid us in performing time-based analysis in an easier way. We performed Exploratory Data Analysis (EDA) subsequently to see how employment and wages fluctuated in different industries like Healthcare, Retail Trade, and Manufacturing. We noticed that some like healthcare were consistent while others like retail were fluctuating. Box graph and violin plots permitted comparison of wage distributions, while line plots indicated employment trends over time.

We then disentangled public vs. private sector distinctions by filtering ownership types and conducting t-tests. These revealed that there were private and local or federal government compensation differentials, specifically in 2022. This is how public sector employment gave more stability than in the private sector amidst uncertainty.

To determine what influenced wages, we used machine learning algorithms such as Linear Regression, Random Forest, and XGBoost. The models are projected Average Weekly Wages based on factors such as industry, ownership, time, and employment amount. Random Forest and XGBoost feature importance validated that industry type and the amount of employment were the best predictors of wage disparity.

We also utilized KMeans clustering to group similar industries into one category and performed ANOVA tests to confirm pay variations between them. A Chi-square test confirmed that industry type and ownership

have a strong relationship. Moreover, a logistic regression model was applied to distinguish high- vs. low-pay records.

Finally, we used an ARIMA time series model to forecast employment in the retail industry. Even though it was falling, it provided a good example of how employment may evolve if existing trends persist. This project delivers a straightforward perception of the way the labor market responded to an epochal world crisis. The outcomes can support better labor policy, planning the effect of automation, and enhanced fairness in wages. It displays data science use for addressing practical problems in the world economy and increasing wiser decision-making both for government and in the business sphere.

Contribution to practice and Literature:

This work makes significant contributions to professional practice as well as academic literature on the dynamics of the labor market during and in the aftermath of the COVID-19 pandemic. Practically, the work offers timely insights for policymakers, managers of businesses, and human resource professionals regarding the most impacted sectors by pandemic disruptions, the uneven impacts of remote work take-up, and the wage inequalities that arose or deepened between 2020–2022. The evidence aids decision-makers in how the most vulnerable sectors should be subject to intervention such as reskilling measures or reforms of wage policy to support labor market resilience and inclusive economic recovery. Literature-wise, this work draws on past knowledge by comprehensively incorporating new measures of labor market tightness like the quits rate and the vacancies over effective searcher (V/ES) ratio, giving a more detailed picture beyond the standard unemployment rates (Heise, Pearce, & Weber, 2024). In addition, through the explicit inclusion of the impacts of automation, inflation, and the role of fiscal policy intervention on wages and employment patterns, the work fills gaps present in analyses based on individual sectors and demographics. It also builds on previous work on labor market inequalities by capturing how structural imbalances remained unchanged even as wages of the lower-paid workers appreciated (Gould & deCourcy, 2023; 2024). Broadly,

Work enhances the basis of labor economics, public policy, and workforce development studies.

Research Limitations

While the work produces important results, several limitations need to be noted. The initial one involves the use of the Quarterly Census of Employment and Wages (QCEW) dataset, while broad in its availability, restricts analysis mainly to formal employer-employee relationships and thus omits the informal or gig economy sectors that have mushroomed over the pandemic years. The results do not accurately represent the situation of independent contractors, freelancers, and undocumented workers. Secondly, the geographic emphasis on California counties could limit the ability of results to generalize.

The labor market dynamics also differ dramatically by U.S. region based on varying state policy regimes, industrial bases, and demographics. Therefore, results from California might not necessarily reflect national or Rural trends. Third, although the analysis investigates correlation between variables like automation take-up and the changing wages, causality is not established. Without the availability of experimental or quasi-experimental designs, conclusions regarding trends' drivers are inferential as opposed to definitive. Also, as the pandemic continues exerting economic impacts beyond 2022, trends that are observed potentially might be transitory or changing, constraining the longer-run validity of certain insights. Ultimately, external forces such as geopolitical tensions, international supply chain disruptions, and environmental crises that also affect labor markets were not controlled specifically within this examination.

Future Research Directions

Building on these results and responding to the current limitations, a few potential lines of work follow. Extending the analysis across states or nationally would give a larger and wider view of the recovery of the labor market and the dynamics of wages. Cross-study comparisons between urban and rural regions would give a better picture of geographic disparities. Secondly, causality needs to be determined with econometric methods like difference-in-differences (DID) analysis, propensity score matching methods, or the use of instrument variables (IV). Such methods would be able to better disentangle the individual impacts of automation, remote work take-up, inflation policy measures, and the use of fiscal interventions on employment and wages. Third, longitudinal studies of workers over long-term periods would provide more insights into the stability of wage gains, the long-term implications of telework on career advancement, and the durability of changes in employment patterns. Additionally, studies should delve more intensively into marginalized groups of workers racial and ethnic minorities, women, and gig workers to gain a better understanding of lingering disparities and to offer more equitable labor market policies. Analysis of how skill-building initiatives, access to education, and means-tested economic incentives can eliminate these disparities would offer policymakers guidelines. Finally, the growing application of artificial intelligence and robotics within the workforce means that subsequent work should investigate how the changing relationship between new technology and wage structures,job stability and workforce flexibility will develop over the next decade.

Recommendations

Based on data analysis findings, certain practical recommendations can be proposed for more effective implementation and deeper insights. Even more analysis of the leading predictors of wages, such as Industry, Average Monthly Employment, and Ownership, with finer-grained features or data (e.g., specific NAICS codes) can uncover additional patterns. Adding interactions between the variables to regression models can also help determine more an complex wage dynamics. The relatively of low R^2 value of the linear model indicates that simple linear relationships are not adequate to account for wage variation, and non-linear techniques are significant. Among these, the Random Forest model worked best, and additional development and emphasis on this model would be warranted. Moreover, the high level of association between ownership type and industry revealed by the Chi-Square test implies that later models will have to incorporate controls for such dependency to achieve meaningful interpretation.

To continue this analysis, some avenues to explore pursuing are the addition of outside economic indicators—e.g., inflation, unemployment rates, or policy actions applicable to the economy—into the analysis. Extending the time of frame beyond 2022 would allow for the determination of an long-term trends and long-term effects of the pandemic and economic shifts. More detailed regional analysis on a California county level could be utilized to determine regionalized trends as well. Finally, investigating further advanced time series forecasting methods beyond ARIMA, especially across industries, could further enhance predictive accuracy and facilitate workforce planning and policymaking.

References

Fiveable. (n.d.). T-tests, ANOVA, & Chi-square tests – Foundations of data science. https://library.fiveable.me/foundations-of-data-science/unit-7/t-tests-anova-chi-square-tests/study_guide/xRRF5LlWPqAssaYO

Gould, E., & deCourcy, B. (2023). Low-wage workers have seen historically fast real wage growth in the pandemic business cycle. Economic Policy Institute. https://files.epi.org/uploads/263265.pdf

Gould, E., & deCourcy, B. (2024). Fastest wage growth over the last four years among historically disadvantaged groups. Economic Policy Institute. https://www.epi.org/publication/swa-wages-_2023/

Heise, S., Pearce, J., & Weber, M. (2024). Wage growth and labor market tightness. New York Federal Reserve Staff Reports. https://www.newyorkfed.org/medialibrary/media/research/staff_reports/sr1128.pdf

Jordà, O., & Nechio, F. (2023). Inflation and wage growth since the pandemic. National Bureau of Economic Research. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10184877/7/

Kumar, S. S. (2023, May 18). Python data visualization for machine learning: A practical overview. Medium. https://medium.com/@sanjayskumar4010/python-data-visualization-for-machine-learning-a-practical-overview-5e14a86034b2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.Retrieved from https://scikit-learn.org/0.16/modules/generated/sklearn.linear_model.LogisticRegression.html

Piacentini, M., Frazis, H., Meyer, B., Schultz, M., & Sveikauskas, L. (2022). The impact of COVID- 19 on labor markets and inequality. Bureau of Labor Statistics. https://www.bls.gov/osmr/research- papers/2022/pdf/ec220060.pdf

Roy, P.(2021, June 17). Understanding Random Forest. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. Retrieved from https://www.statsmodels.org/stable/generated/statsmodels.tsa.arima.model.ARIMA.html

scikit-learn developers. (n.d.). sklearn.linear_model.LogisticRegression. scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Teppan_noodle. (2020, June 21). Exploratory data analysis (EDA) using Python. Medium. https://medium.com/@teppan_noodle/exploratory-data-analysis-eda-using-python-f85938cb1810

Authors Contributions

Lohith Pasupuleti: I began by preparing the dataset. I loaded the CSV file from Google Drive and inspected the data structure using some techniques like.head() and.info(). I validated to see dataset shape, ensuring that there were no duplicates or missing values, and validated all data types. I preprocessed the data to keep only the important employer types—Private, Federal Government, State Government, and Local Government. For visualization, I plotted a bar chart for employment by ownership type, a histogram of average weekly wage distribution, a line chart of wage evolution from Q1 2020 to Q4 2022, and a quarterly employment pattern chart for tracking. I also utilized a Linear Regression model for predicting average weekly wages against features like year, industry, and size of employment. I gauged its performance using R^2, MAE, MSE, and RMSE.

Pilli Veena Madhuri: I spent hours cleaning the dataset with more accurate filtering and feature engineering. I deleted the data to contain quarter records and worked solely with California counties by deleting all statewide and nationwide records. I kept a new feature, Avg_Emp_Per_Est, by averaging average monthly employment by number of establishments and cleaning it by substituting any NaN or infinity values. I removed outliers from this revised measure using an IQR-based outlier trimming method. I computed summary statistics and interpreted significant trends with describe(). I plotted employment by year, a hexbin plot to show the relationship between establishment size and weekly earnings, a bar chart of total earnings by year, and a stacked bar plot of the employment by ownership by year. Lastly, I created a Logistic Regression model that could predict whether a wage was over or under the median based on employment, ownership, and industry codes and experimented with it.

Veda Samskruthi Kancherla: I did an in-depth analytical work in the domain of wage dynamics. I computed wage growth rates by industry as percentage change and built a Year_Quarter feature for temporal aggregation. I computed an average wage by the industries and determined the highest-paying industries from 2020 to 2022. I performed correlation analysis on numerical variables and displayed them in a heatmap. I also coded categorical variables into numeric variables with LabelEncoder and verified their Pearson correlation with wage measures. I created a boxplot of wages by employment size bins, a horizontal bar chart of the 10 most lucrative industries, a cleaned boxplot without outliers for the industries, and a line graph of wage

growth in specific industries. I predicted average weekly earnings using an XGBoost model with employment, ownership, industry, and time attributes. I split the data, trained the model, predicted, and

`tested it on R^2, MAE, and RMSE.

Shivani Nalanagula: I performed statistical analysis and sophisticated modeling. I performed a Chi- Square test to investigate the association between industry and ownership category, and a t-test to examine the difference in wages by private and local government employers. I also performed a one-way of ANOVA to test differences in wages by industries. I documented how categorical variables were transformed—via one-hot encoding for the low-cardinality features and LabelEncoder for high-cardinality features—and how they were combined with numeric variables to train models. I talked about the hyperparameter tuning of XGBoost using RandomizedSearchCV and documented the best-performing parameters. I did KMeans clustering on scaled employment and pay data and visualized the clusters, interpreting them by size and levels of pay. I applied the ARIMA model used in forecasting employment for the period in the retail industry. Finally, I interrogated why the 2022 public vs. private wage t-test returned NaN values, which was most likely due to a significant lack of data in one of the groups.