Examining Pay Disparities among Tennessee State Employees: A Data Science Approach

eCOTS 2024 Regional Conference, Belmont University
June 12, 1:00PM CDT

by
Jayden Cruz

Advisor:
Dr. Rudolph Bedeley, Ph.D.

- Rising Junior @ Belmont University
- Computer Science and Data Science Major
- Minor in Business Administration and Mathematics

Presentation Outline

- Background & Motivation
- Research Question
- Definition of Data Science
  - The Data Science Process
- Dataset & Source
- Results
- Discussion
  - Implication
  - Limitation
- Conclusion

Background & Motivation

- Pay disparities based on ethnic group and gender have been a long standing issue.

- This project explores the trends in pay disparity among TN state-employee workers, within a one year salary dataset

- To address the research question, data science methodologies have be employed, leveraging the power of statistical analysis, data preprocessing, and data visualization techniques

Research Question

“Are there significant difference in pay between ethnic groups and genders among TN state employees, and what are the trends in pay disparities within a one year salary dataset”

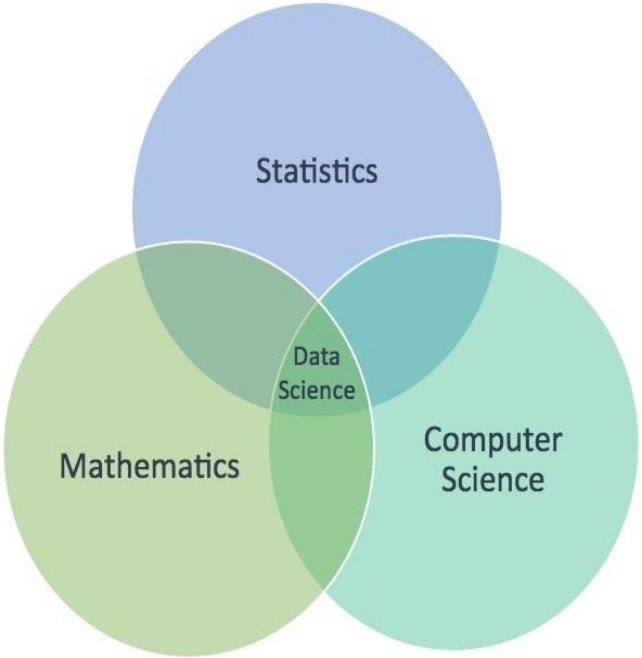
Map of Tennessee and its counties

Background & Motivation - cont’d

- The significance of this research lies in its potential to inform policy decisions and foster inclusive work environments.

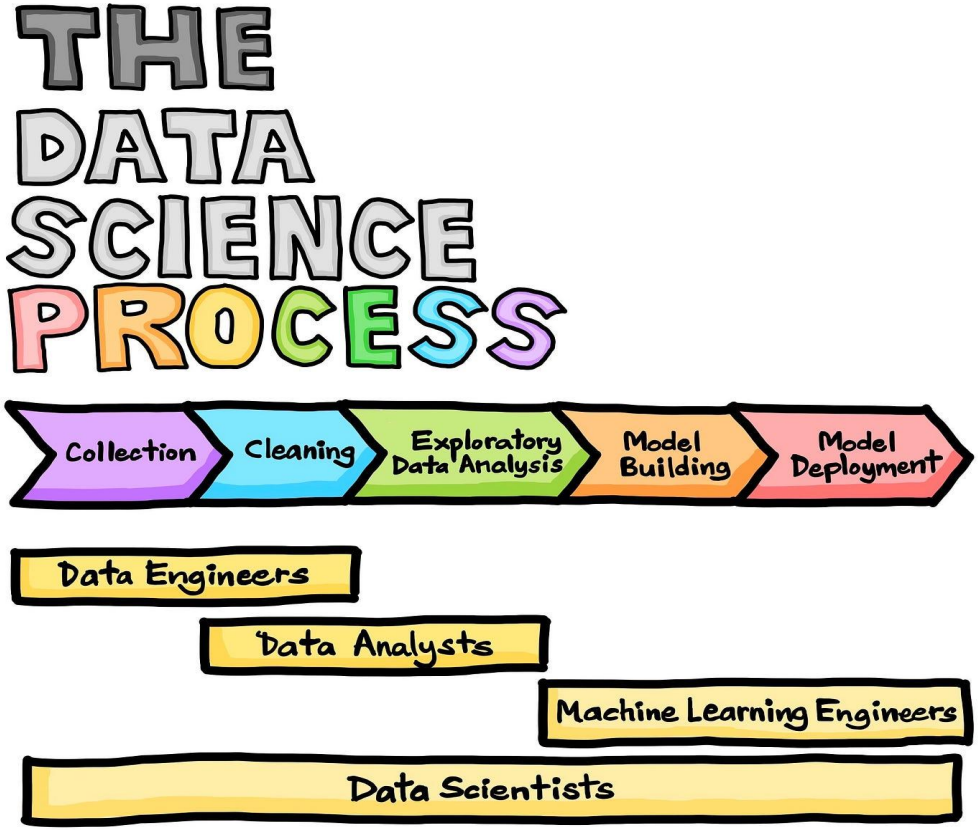
- Results and interpretation will be presented followed by a discussion of the implications, limitations, and recommendation for future research.

What is Data Science?



- If Computer Science, Mathematics, and Statistics had a baby.
- The study of data to extract meaningful insights for business.
- “Data Scientist: The Sexiest Job of the 21st Century” -Harvard Business Review

A Data Science Approach!



Dataset & Source

- Comprehensive dataset acquired from Data.Nashville.gov

- Analyze trends, changes, and implications of the data

- Data science methodologies, including data preprocessing, statistical analysis, and data visualization, are employed to uncover patterns and correlations within the dataset.

Methodology (a.k.a The Data Science life cycle)

The Step-by-Step approach you take to deliver a project

Data Collection

- Obtained a comprehensive dataset containing information on Tennessee state employees, including their pay, ethnic group, gender, and employment history.

Methodology (a.k.a The Data Science life cycle)

Data Cleaning and Preparations

- Clean the dataset by removing any impurities(irrelevant or missing data).
- Handle inconsistencies or outliers that may impact the analysis.
- Anonymize Personal information to ensure data privacy and security.

Comparative Analysis

- Perform a comparative analysis to identify andy statistically significant differences in pay.
- Use appropriate statistical test, such as t-tests, to validate the significance of the observed disparities.

Methodology (a.k.a The Data Science life cycle)

Visualization

- Create visualizations, such as bar charts, box plots, or line plots, to present the findings from the descriptive and comparative analysis.
- Use appropriate labels, titles, and legends to enhance the clarity of the visualization.

Interpretation and Reporting

- Interpret the results of the analysis, highlighting key findings and insights.
- Prepare a comprehensive report summarizing the methodology, data analysis, and key findings.

Results

- Our analysis of TN state employees' pay data revealed disparities in salaries across different ethnic groups and genders. However, it is important to note that these findings are specific to state-employed individuals and may not reflect the overall pay disparities in the entire Tennessee workforce.

- The analysis focused solely on the salaries of state employees and it is plausible that the disparities observed within this subset of the workforce could be even larger when considering all workers in Tennessee.

Results

```
male_pay = fulltime_df[(fulltime_df['Gender'] == 'M')]['Annual Salary']
female_pay = fulltime_df[(fulltime_df['Gender'] == 'F')]['Annual Salary']

print(np.var(male_pay), np.var(female_pay))
```

```
588394919.684311 629840569.8953072
```

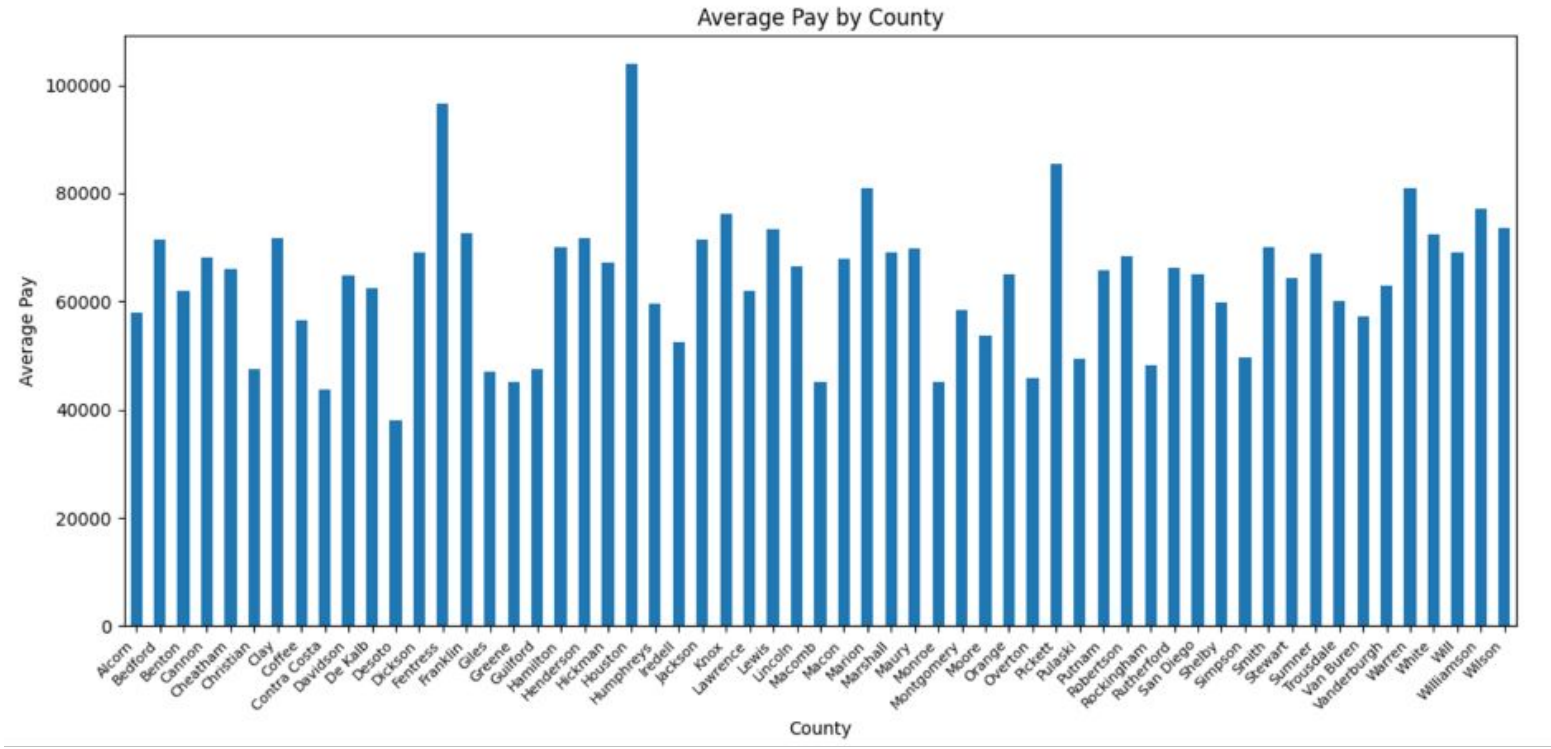
```
stats.ttest_ind(a=male_pay, b=female_pay, equal_var=True)

```

```
Ttest_indResult(statistic=10.705240905852907, pvalue=1.3919396429259966e-26)
```

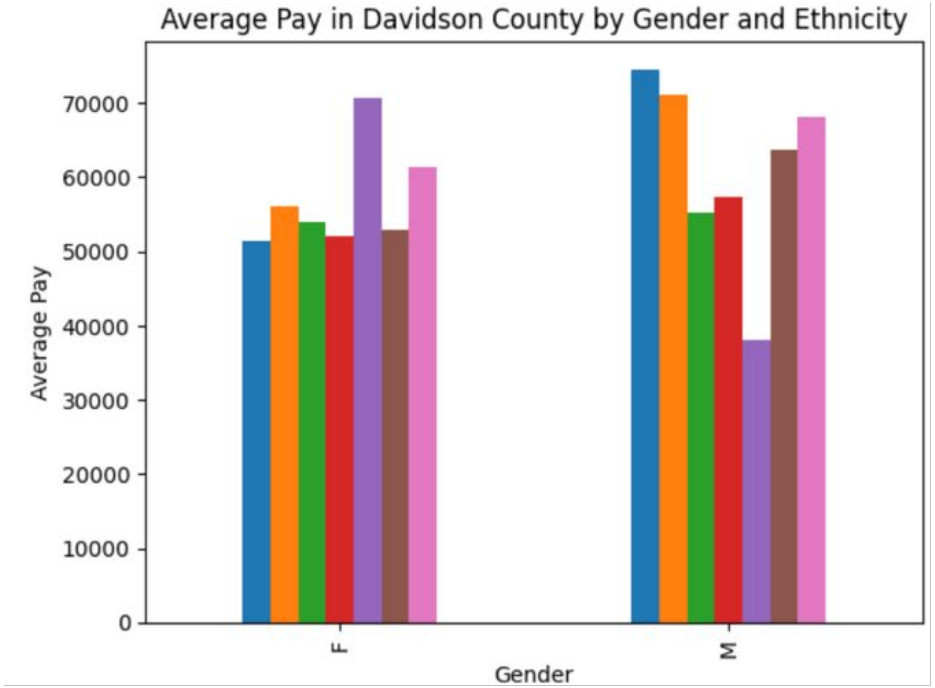
Python code snippet executing a t-test to analyze the disparity in pay between male and female employees. The t-test results provide statistical evidence confirming the existence of a significant difference in pay based on gender.

Results



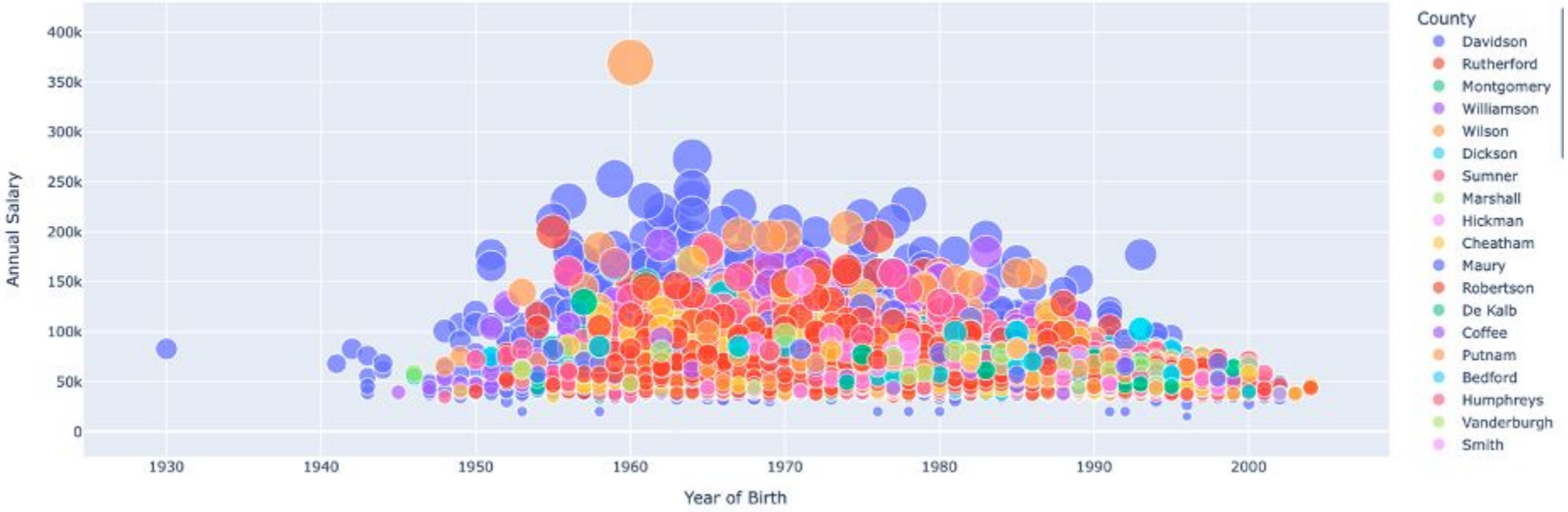
Histogram illustrating the distribution of average pay across various counties in Tennessee (TN). The x-axis represents different counties, while the y-axis represents the average pay. Each bar corresponds to the average pay in a specific county. Data analysis conducted using python and python library matplotlib.pyplot.

Results



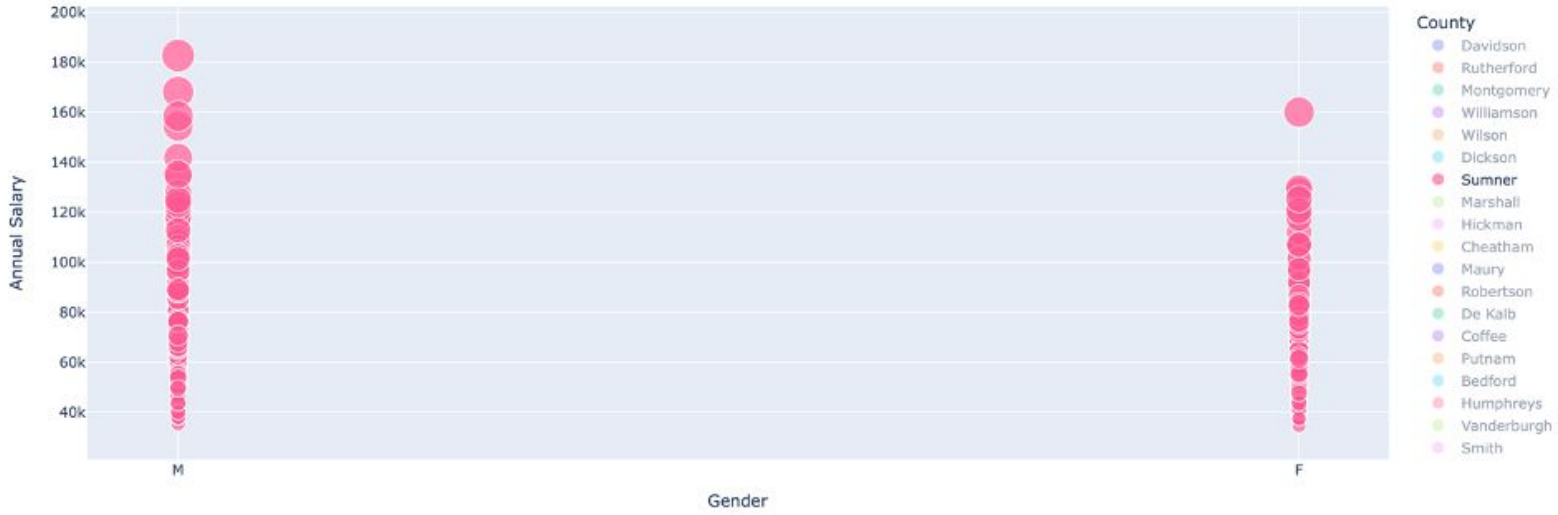
Each bar represents the average pay for a specific gender-ethnicity group. Ethnicity names were pre-assigned in the dataset and color were randomly assigned by Python. Blue represents American or Alaskan Native, Orange represents Asian, Green represents Black or African American, Red represents Hispanic or Latino of any race, Purple represents Native Hawaiian or other Pacific, Brown represents two or more race and Pink represents.

Results



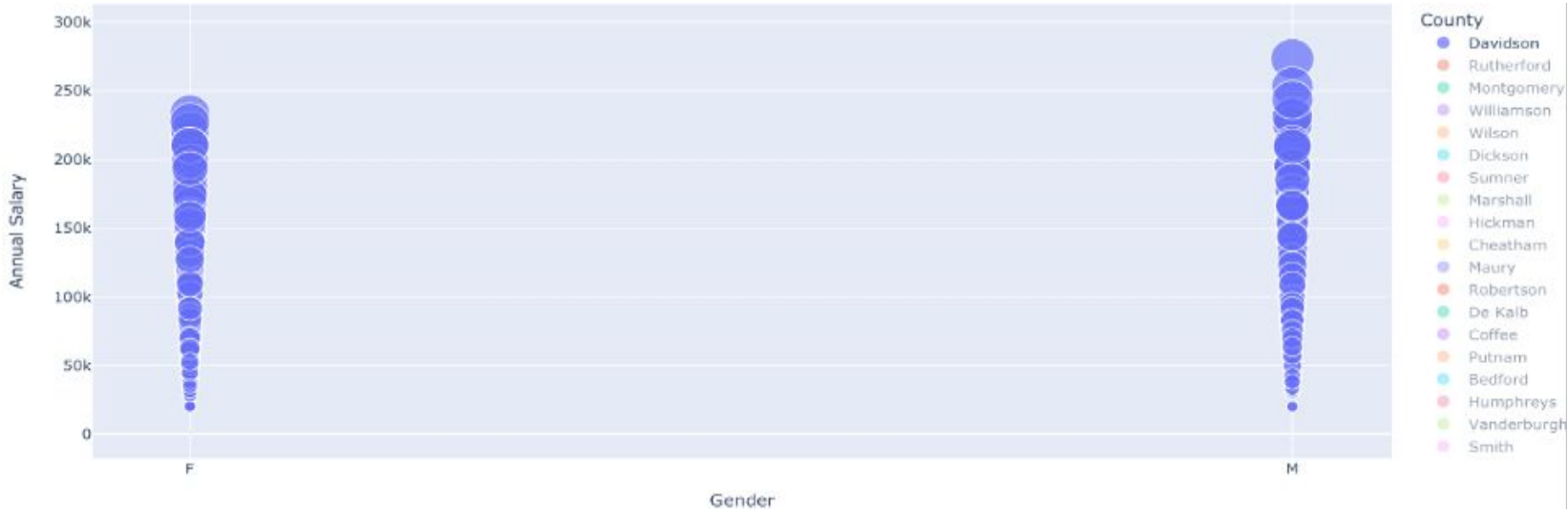
Bubble chart illustrating the relationship between annual salary and year of birth, with each bubble representing a county. The x-axis represented the year of birth, while the y-axis represents the annual salary. Each bubble's size corresponds to the count of individual within a specific county. Counties are differentiated by color, as indicated in the legend on the side of the graph.

Results



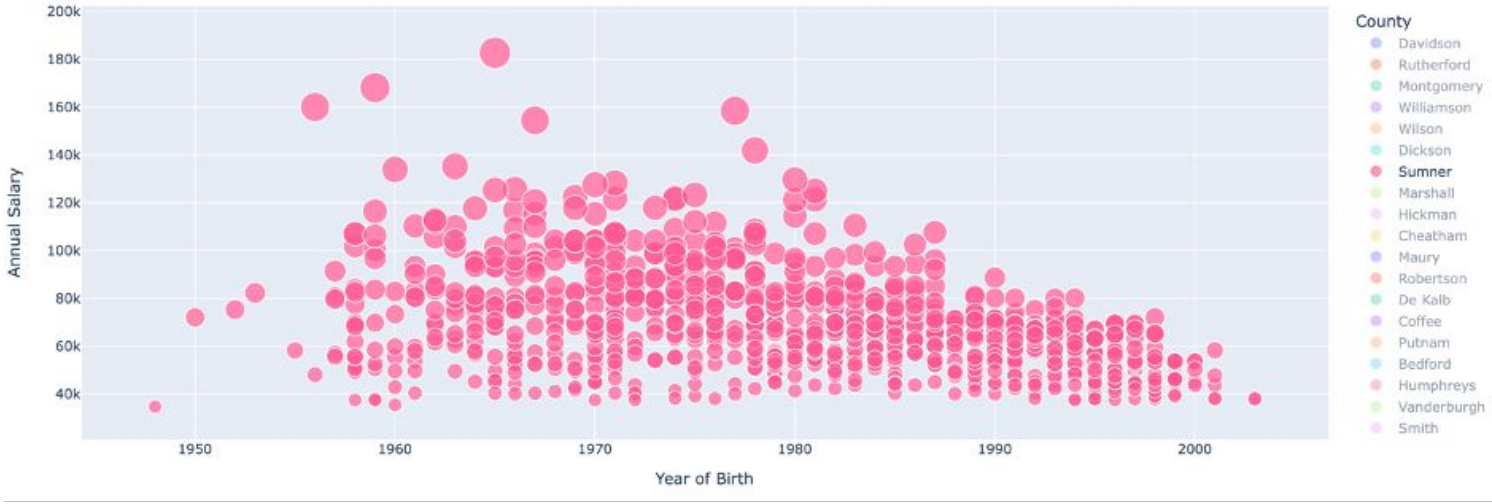
Bubble Chart illustrating annual salary by gender in Sumner County

Results



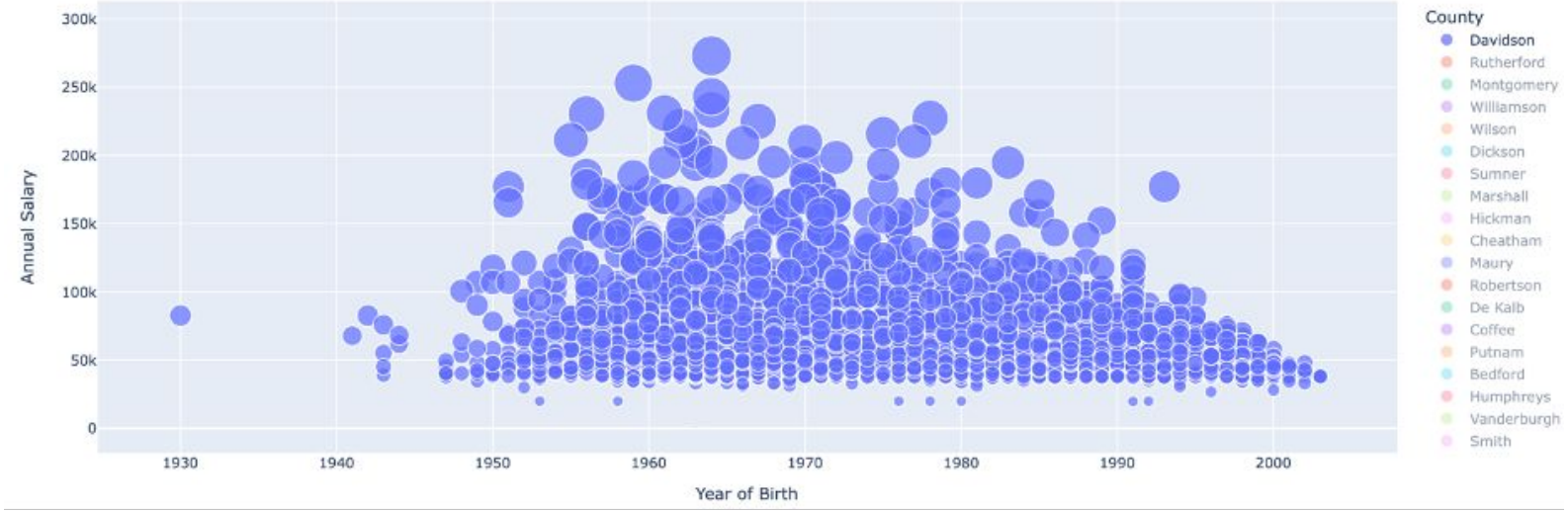
Bubble Chart illustrating annual salary by gender in Davidson County

Results



Bubble chart depicting the relationship between annual salary and year of birth for individual residing in Sumner County

Results



Bubble chart depicting the relationship between annual salary and year of birth for individual residing in Davidson County

Results

- Leveraging powerful tools and libraries such as Python's matplotlib, seaborn, and pandas, along with various data preprocessing techniques, each visualization was tailored to elucidate specific aspects of the dataset. By aggregating and analyzing vast amounts of data these visuals provide a comprehensive overview of key relationships and trends within the dataset.

- As we continue to harness the power of data science and visualization techniques, the possibilities for driving positive change and fostering inclusive growth are boundless.

Results

Overall, the findings suggest that pay disparities exist among Tennessee state employees, highlighting the importance of further examination and potential actions to address these disparities. However, it is essential to acknowledge the limitations of the analysis, which focused exclusively on state employees and did not encompass the entire Tennessee workforce. To gain a more comprehensive understanding of pay disparities, future research should consider broader datasets that encompass a representative sample of workers from different employment sectors within the state.

Discussion

Implications of Findings

- The analysis of Tennessee state employees' pay data revealed significant disparities in salaries across different ethnic groups and genders. These findings underscore the importance of addressing pay equity within the state workforce. By uncovering these disparities, we shed light on potential systemic issues that need to be addressed to foster a more inclusive and fair work environment.

Discussion

Contextual Considerations

- It is crucial to contextualize these findings within the broader landscape of pay disparities, recognizing that our analysis focused solely on state-employed individuals. While our results provide valuable insights into pay discrepancies within this subset of the workforce, it is plausible that the disparities observed here could be even larger when considering all workers in Tennessee, including those employed in the private sector.

Discussion

Statistical Validation

- The comparative analysis conducted to validate the significance of these observed disparities employed appropriate statistical tests, such as t-tests. The resulting p-values and effect sizes indicated that the observed disparities were unlikely to occur by chance alone and had meaningful practical implications. This statistical validation adds robustness to our findings and reinforces the urgency of addressing these disparities.

Discussion

Temporal Dynamics

- The time-series analysis examining the trends in pay disparities over time revealed valuable insights into the temporal dynamics of pay inequalities within the state employee population. By analyzing changes in pay disparities over time, we were able to assess if any improvements or exacerbations had occurred. This temporal perspective is essential for understanding the evolving nature of pay disparities and identifying areas for intervention.

Discussion

Limitations

- However, it's important to acknowledge the limitations of our analysis. Our focus exclusively on state employees limits the generalizability of our findings to the broader Tennessee workforce. Future research should consider broader datasets encompassing a representative sample of workers from different employment sectors within the state to gain a more comprehensive understanding of pay disparities.

Discussion

Recommendations for Future Research

- Moving forward, it is imperative to continue investigating pay disparities in Tennessee, expanding the scope of research to encompass a more diverse sample of workers. Additionally, qualitative research methods could provide deeper insights into the underlying factors contributing to these disparities, facilitating the development of targeted interventions.

Discussion

- Provide insights into the dynamics of pay disparities based on ethnic group and gender.

- Bring awareness to the potential of data science for social good. Showcasing how data-driven insights can be leveraged to tackle societal challenges such as inequality and discrimination

- Groundwork for future endeavors in data science for social good

Conclusion

- In conclusion, our analysis of Tennessee state employees' pay data revealed significant disparities in salaries across different ethnic groups and genders. However, it is important to note that these findings are specific to state-employed individuals and may not reflect the overall pay disparities in the entire Tennessee workforce.

- It is crucial to recognize the limitations of our analysis. Our findings are specific to state-employed individuals and may not represent the entire Tennessee workforce. To gain a more comprehensive understanding of pay disparities, future research should consider broader datasets that encompass a representative sample of workers from different employment sectors within the state.

Conclusion

In conclusion, our analysis serves as a starting point for addressing pay disparities among Tennessee state employees. By recognizing the existing inequities and taking proactive steps towards promoting fair compensation practices, we can foster a more inclusive and equitable workforce in the state

Questions and Discussion

Contact

/in/jaydencruz/

Email: jaydencruz.cs@gmail.com

Email: rudy.bedeley@belmont.edu

Slides

https://github.com/JaydenCruz2004/eCOTS-Conference-Slides