



SEMESTER PROJECT

EXPLORING THE IMPACT OF COVID-19 ON CA WIC DATASET: ANALYSIS AND VISUALIZATION

YIJIN ZHU

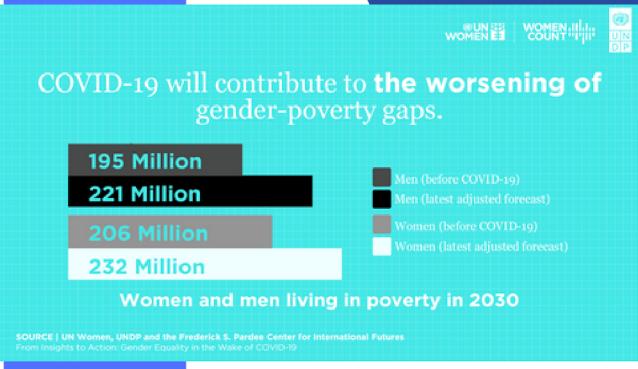
Hello everyone!

My name is Yijin Zhu.

In this presentation, I will be presenting my project, using Big Data for exploring the impact of Covid-19 on the WIC program in California.

PROJECT MOTIVATION

- Big Data for social welfare to better meet the needs of our community
- Covid-19 pandemic will worsen the gap in gender poverty



Big data is being widely used due to its applications in analysis and data-driven decision-making.

Social welfare organizations can utilize big data analysis to better understand their target audience's needs.

The economic crisis triggered by the COVID-19 pandemic has pushed many women into extreme poverty. It is crucial for governments and social welfare organizations to take steps to support financially struggling women and their families during this challenging time.



PROBLEM STATEMENT

EXAMINE THE IMPACT OF THE COVID-19 PANDEMIC ON THE WIC PROGRAM

- How has WIC redemption changed in California over the past decade?
- Did the outbreak of COVID-19 affect WIC redemption statewide?
- If so, are there any similar trends in the data from big counties in CA?

The Women, Infants and Children (WIC) Program helps California families by providing food benefits to nearly a million participants in local communities throughout the State.

In this project, I will analyze the impact of the COVID-19 pandemic on the WIC program in California. By utilizing big data analysis techniques, I want to explore the following questions:

Materials & Methodology

1 DATA COLLECTION

*California WIC Program Redemption
California Open Data Portal
<https://data.ca.gov/>*

2 DATA CLEANING AND PREPARATION

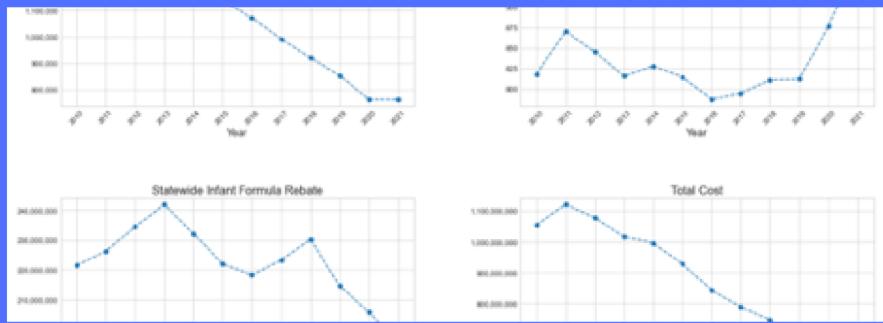
*Read CSV files, reconstruct dataframes,
convert data types and replace missing values*

3 DATA ANALYSIS

*Create graphs to evaluate the relationships
between various features*

The datasets I used in this project, were retrieved from California Open Data Portal. For data cleaning and preparation, I use Spark and Pandas to read and prep the data. Then for the data analysis part, I use Seaborn, which is built on Matplotlib to create plots and graphs to evaluate the relationships between various features.

ANALYSIS RESULTS



STATE

- GENERAL REDEMPTION TRENDS
- DETAILED REDEMPTION TRENDS

COUNTY

- NUMBER OF FAMILIES VS AVERAGE COST
- NUMBER OF FAMILIES VS TIME
- AVERAGE COST VS TIME

The analysis will be divided into two parts: state-level data analysis and county-level data analysis.

At the state level, I will examine the redemption trends using both general and detailed data, using time as a variable to investigate if there were any significant changes after the COVID-19 pandemic.

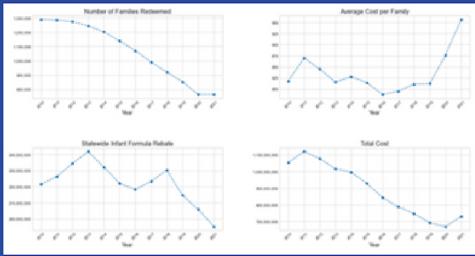
Moving on to the county level, I will break down and integrate features that cannot be shown at the state level. This will involve comparing variables such as the average costs of counties with a higher number of WIC families and evaluating whether the number of participants changed differently for each county.

In this demo, I will provide a brief overview of the analysis results. However, for a more detailed evaluation, please refer to the full report.

STATE ANALYSIS

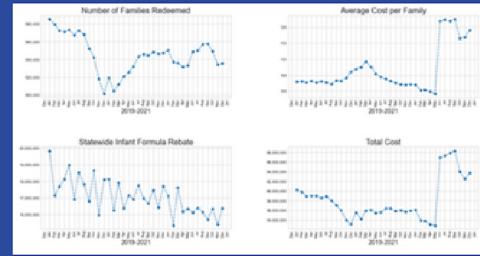
GENERAL REDEMPTION TRENDS FROM 2010-2021

- Number of Families Redeemed
- Average Cost
- Statewide Infant Formula Rebate
- Total Cost



DETAILED REDEMPTION TRENDS FROM 2019-2021

- Number of Families Redeemed
- Average Cost
- Statewide Infant Formula Rebate
- Total Cost



While analyzing the state data, we are using four variables:

Number of Families Redeemed

Average Cost

Statewide Infant Formula Rebate

Total Cost

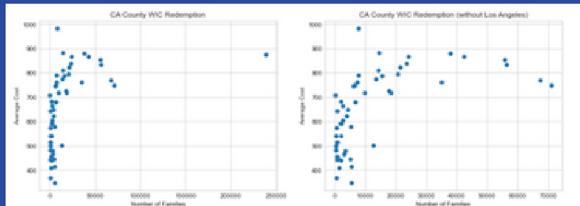
The left part is the general redemption trend for 12 years, we can see the decline in the number of participants and total cost, the increase in average cost, and the consistent drop in infant formula rebate.

Next, we take a closer look at the detailed redemption trends on the right-hand side. From 2019 to 2021, before and after the outbreak of COVID-19. We can see a sudden increase in the average cost and total cost during this period.

CONUTY ANALYSIS I

NUMBER OF FAMILIES VS AVERAGE COST (2021)

- Number of Families Redeemed
- Average Cost
- Correlation



NUMBER OF FAMILIES VS TIME (2010-2021)

- Number of Families Redeemed



Now let's have a look at the county data. The scatterplots on the left, show the relationship between the average redemption costs and the number of participants across different counties. One scatterplot includes LA County, and the other does not.

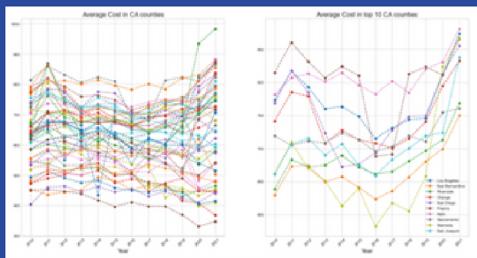
For counties with less than 10,000 WIC families, we observe a moderate correlation between the number of families and the average redemption cost. In contrast, in counties with a larger number of WIC families, the correlation is even lower.

Moving on to the right, we can see a general trend of declining numbers of WIC families from 2010 to 2021 in most counties. However, this trend changed in 2021, with most counties experiencing either a flat slope or even a positive slope, indicating an increase in the number of participants. The second graph, which excludes LA County, also shows the same trend.

CONUTY ANALYSIS II

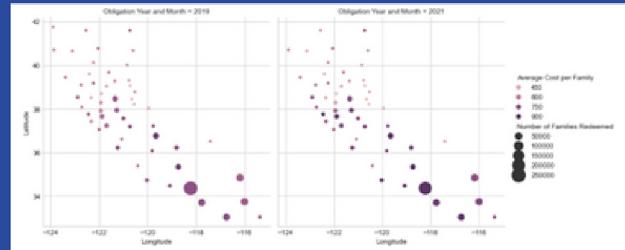
AVERAGE COST VS TIME (2010-2021)

- Average Cost



REDEMPTION VS LOCATION (2019 VS 2021)

- Number of Families Redeemed
- Average Cost
- Location



Now let's look at the average cost of WIC redemption across all counties over time. From 2010 to 2019, the majority of lines are relatively stable. However, in 2021, many counties have seen a significant increase in average cost.

The next graph shows the changes in average cost for the top 10 counties with the highest number of WIC participants. We can see rapid growth in average cost in both 2020 and 2021.

The right-hand side shows the geographic distribution of redemption costs across counties in California. Each dot represents the center of a county, with the hue of the dot indicating the average cost (darker colors indicate higher costs) and the size of the dot indicating the number of participants in each county (bigger dots indicate more participants). Comparing the two plots, we can see that although the sizes of the dots are similar, many dots in the second plot are much darker, indicating a significant increase in redemption costs in many counties.

CONCLUSION

This project

- Situation before&after the pandemic
- Differences in redemption patterns between counties
- Geography of WIC redemption

Future Works

- Data collection
- Detect correlations between various factors
- Predict future trends with machine learning

The findings suggest that while the WIC program was facing a decrease in participants and total cost before the pandemic, the situation has changed in both 2020 and 2021 with a slight increase in the number of participants and a significant rise in average cost. The analysis also highlights the differences in redemption patterns between counties. counties with more participants have higher redemption costs, especially in southern CA. After the outbreak of COVID-19, the geography of WIC redemption changed. The differences in average costs between northern and southern CA became more visible.

What we can do in the future?

Big data can be used to further improve this project in several ways. Firstly, big data provides a wider scope for data collection. We can utilize a greater variety of data sources and new data collection methods. Secondly, big data analytics can help us identify patterns and trends in the WIC program data more timely and accurately. For example, some data analytics can be applied to the data to detect correlations between various factors, such as statewide distribution of population, income levels, employment rates, and WIC participation rates. What is more, big data can be used to predict future trends in WIC participation and costs. machine learning methods can be trained to predict how the WIC program may change in the coming years. This can help policymakers and organizations to generate insights, and make informed decisions accordingly. We can also use big data to help us monitor the progress and outcomes of the programs, creating more effective social welfare programs and services.

THANK YOU!



That concludes my presentation, thank you for your time!