Group 4

Name: **Sakshi Patel**

Project Stage - II (Data Modeling)

Introduction:

The COVID-19 pandemic has had profound effects on public health, the economy, and society worldwide. To understand its spread and impact, it is essential to analyze not only the reported cases and deaths but also the demographic context of the affected areas. This project aims to clean and merge COVID-19 case and death data with demographic data from the ACS. By analyzing trends in different regions, particularly in Arizona, and enriching the data with demographic information, this report seeks to gain deeper insights into how factors such as population density, age, and socioeconomic status might have influenced the spread and severity of the virus.

1. Data Overview:

The dataset used for this analysis is '**super_covid19.csv**', containing 3142 rows and 2535 columns. The data covers all U.S. counties and provides daily statistics for COVID-19 cases and deaths. The analysis focused on data from **June 1, 2020 (Monday) to January 3, 2021 (Sunday)**. This period was selected to observe the COVID-19 cases and deaths.

## 2. Weekly Statistics (Arizona State):

The state of Arizona (StateFIPS = 4) was selected as a case study to analyze the weekly statistics for COVID-19 cases and deaths. The following trends were observed:

COVID Cases
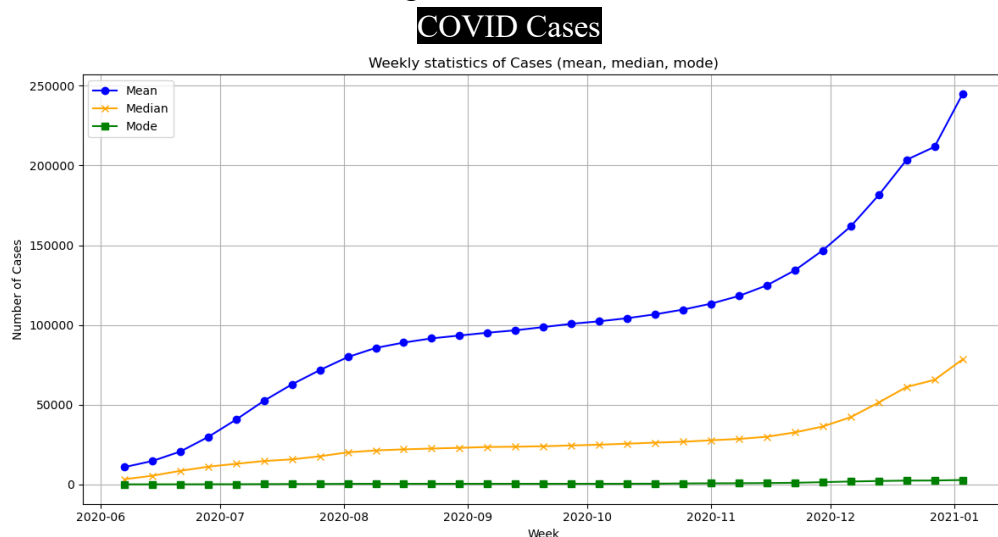
Mean: The mean, or average number of cases, shows a continuous rise over time.
Median: The median, which represents the middle value of the dataset, also increases but at a much slower rate compared to the mean.
Mode: The mode, or the most frequent number of cases reported in a week, remains constant and very low throughout the entire period.
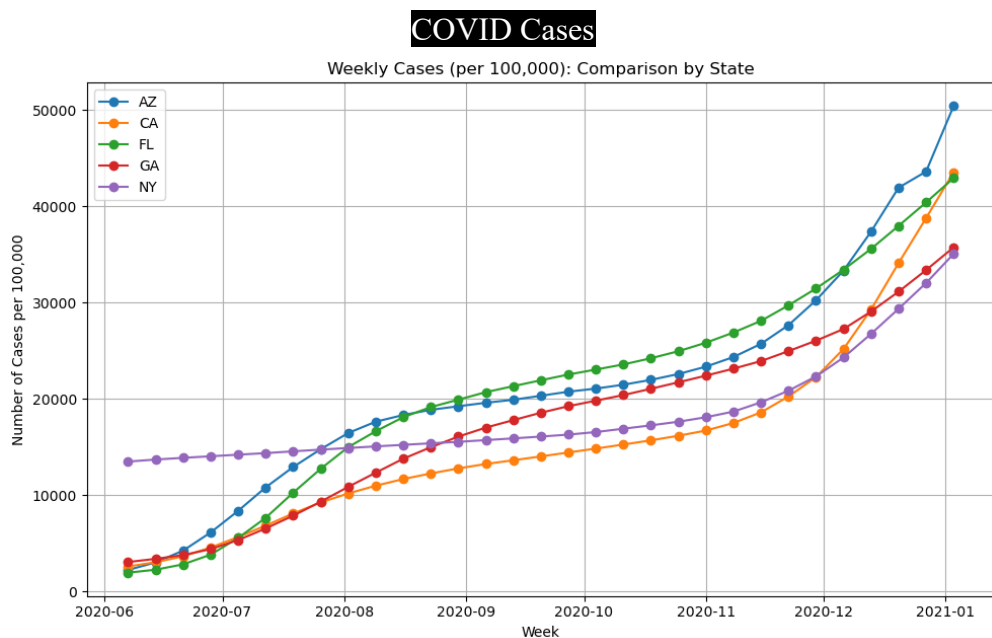
Weekly statistics of Deaths (mean, median, mode)
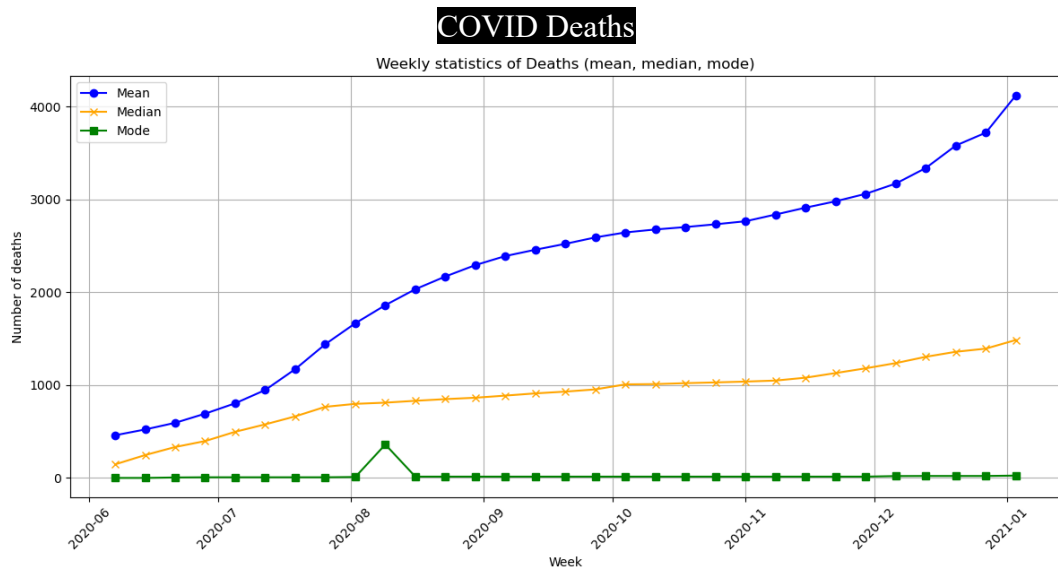
Mean: The mean continues to rise throughout the graph, with a sharp upward trend starting around November 2020.

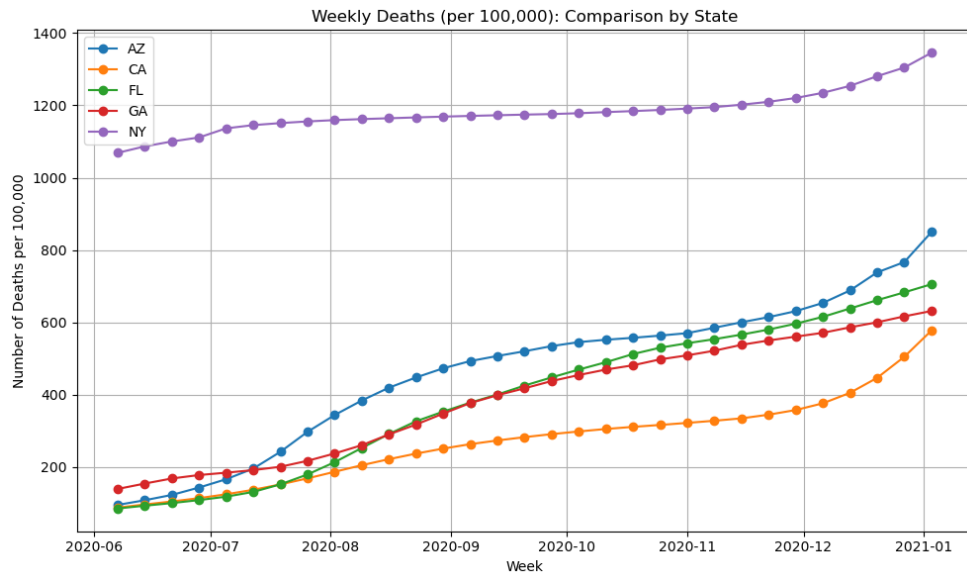**Median**: The median also rises, but at a much slower rate than the mean.

Mode: The mode remains quite low, indicating that most weeks or regions had consistently low death counts.

3. **Comparison Across State:**

Five states (Arizona, California, Florida, Georgia, and New York) were compared in terms of weekly cases and deaths. The data was normalized by population (per 100,000 people) to ensure comparability across states with varying population.

COVID Cases

Weekly Cases (per 100,000): Comparison by State

COVID Deaths

Weekly Deaths (per 100,000): Comparison by State

| COVID Cases	COVID Deaths
Differences in Case Rates Across States **Arizona**: Arizona shows the sharpest increase in cases, especially starting around November 2020. **California (CA), Florida (FL), and Georgia (GA)**: These states show similar trends but have lower peaks compared to Arizona.vember 2020. **New York (NY)**: New York has a significantly lower case rate compared to the other states.	Differences in Death Rates Across States **New York (NY)**: New York has the highest number of deaths per 100,000. **Arizona (AZ)**: Arizona shows a significant increase in death rates starting around October 2020. **California (CA), Florida (FL), and Georgia (GA)**: These states follow a similar upward trend, with California and Florida showing the next highest death rates after Arizona.
Case Peaks: The peaks in Arizona, California, Florida, and Georgia around November to January align with the national trend of a surge in COVID-19 cases during the fall and winter of 2020 due to increased social interactions and holiday gatherings.	**Death Peaks**: New York's higher death rate may reflect its earlier experiences during the first wave of the pandemic in 2020. Arizona's rapid rise in both cases and deaths.

4. County-Level Analysis in Arizona:

Weekly trends for these counties were plotted, both as raw values and using log normalization to identify patterns of exponential growth.

Weekly trends of cases for the top 5 infected counties by raw values
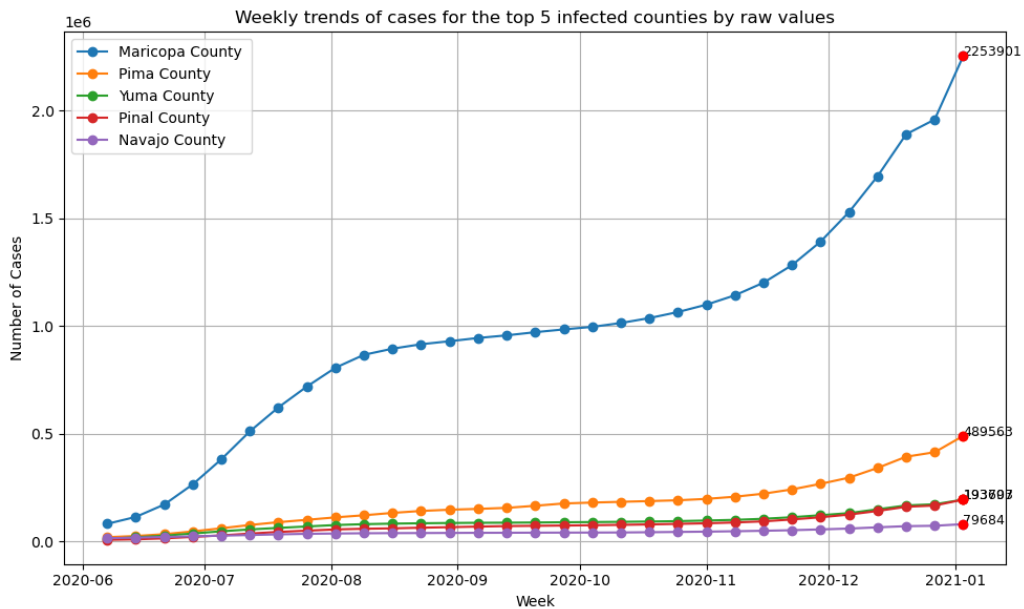
COVID Cases by log normalized values

Weekly trends of cases for the top 5 infected counties by log normalized values
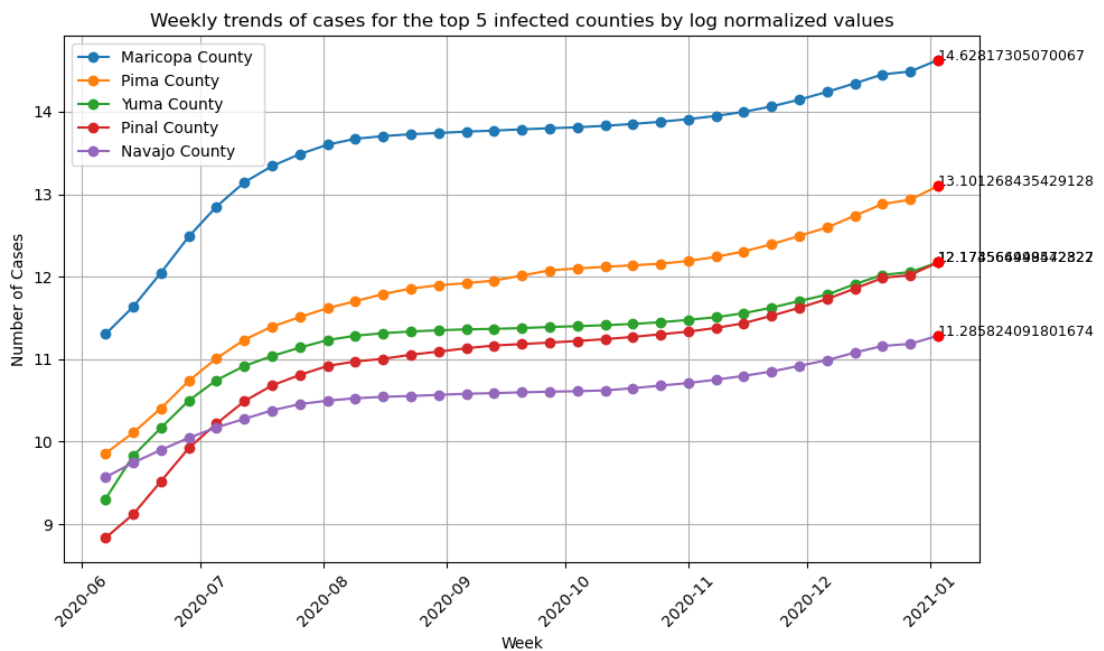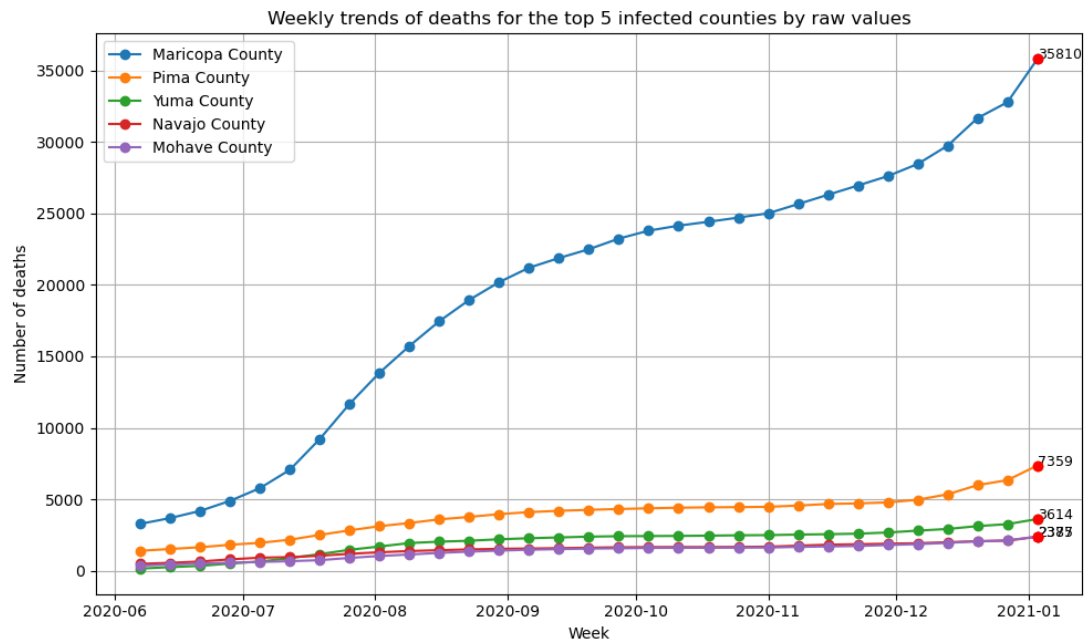

What is causing the trends? The population of Maricopa County, as Arizona's most populous county, likely contributed to its higher case numbers. Pima, Yuma, Pinal, and Navajo counties also show growth, at smaller scales. These counties are less populated, resulting in fewer cases.
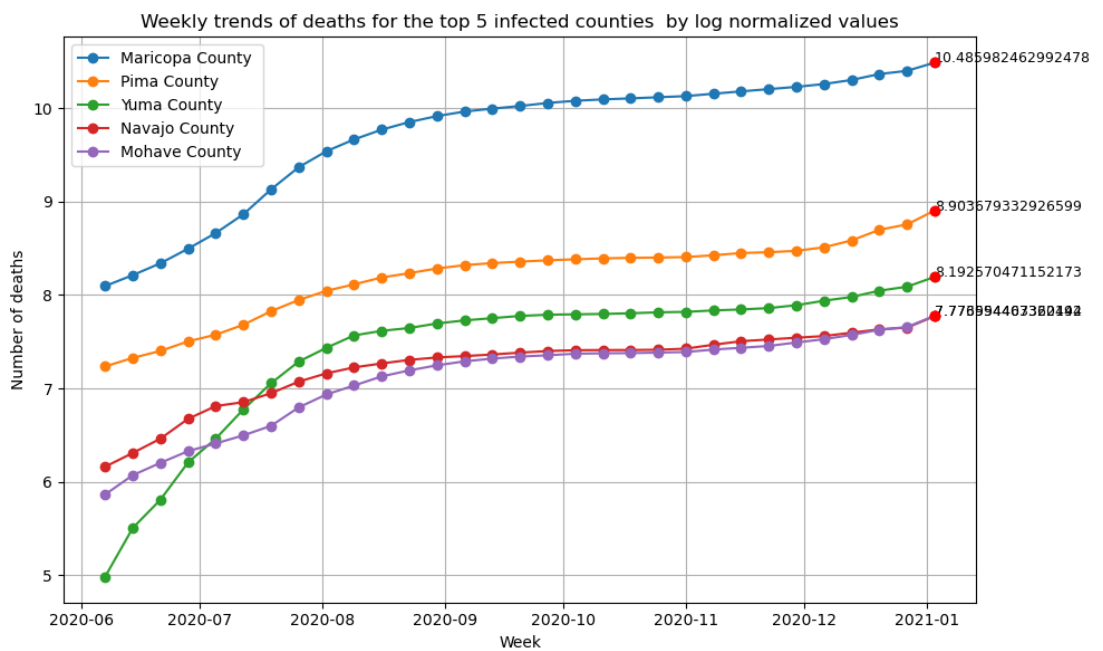
Peaks: The clear peak in Maricopa County's raw cases occurs around late December 2020 to early January 2021. This could be due to holiday travel, increased gatherings, and colder weather.

State Pattern: The trends in these top five counties likely follow the overall state pattern seen in Arizona. Like much of the U.S., Arizona experienced a surge in COVID-19 cases during the summer of 2020 and another significant peak during the winter months of 2020–2021.

Weekly trends of deaths for the top 5 infected counties by raw values

Weekly trends of deaths for the top 5 infected counties by log normalized values

What is causing the trends?: The rise in deaths corresponds with the winter wave of COVID-19 in late 2020, likely driven by increased transmission during holiday gatherings and colder weather.

Peaks: The major peak for Maricopa County was around the end of 2020, while the other counties show smaller, less distinct peaks.

State Pattern: The overall trends align with the state's COVID-19 pattern, where Maricopa County, being the most populous, shows the most significant impact.