Data Sources:
American Presidency Project (Approval Rating Data): https://www.presidency.ucsb.edu/
Federal Reserve Data (Unemployment rate, median household income, GDP per capita): https://fred.stlouisfed.org/series/
Yahoo Finance (S&P500): https://finance.yahoo.com/

Github Repository (with cleaned/modified datasets and scripts): All data and code is accessible via this link
https://github.com/ArjunaBazaz/Presidential_Approval_Rating_Prediction

Preprocessing:

1. Dataset Compilation
    a. In order to conduct analysis both on individual president's data and on long term data including multiple presidents, we needed one csv file for every president
    b. Combined all candidate datasets into one final dataset by adding a candidate column to specify the candidate, and then merging each original csv into a approval_rating.csv dataset.
    c. Since some data points have a general increase over time (stock market, median income, and GDP per capita) it could lead to issues. As a result, a column was added called "Change" to each of those datasets. It was calculated by doing value[x]/value[x-1] - 1, where x is the date and x-1 is the previous recorded date.
2. Data cleaning: Each dataset is structured as a date, but not every poll is taken at the same rate. Some are taken weekly, some are taken biweekly, and some are taken a couple days fin a row. To fix this, we set a week as the main measuring point. In order to normalize this, data points that are more common than that will be removed. Data points that are more common than that will have points filled in as the mean between the previous point and the next one.

Unit of Observation: Each row of data corresponds to data taken of a specific date or range of dates. For the approval rating table, it refers to all of the data taken within the range of the dates given. For the economic metrics columns, it refers to the data on that specific date.

Data Dictionary:

Column Name	Type	Description
Candidate	String	The name of the candidate
Start Time	Datetime	Start date of polling in dd/mm/yyyy. Usually a week long, but can vary.
End Time	Datetime	End date of polling in dd/mm/yyyy

Approving	int	Percent of polled that answered yes to the question, “Do you approve of _____’s job performance”
Disapproving	int	Percent of polled that answered no
Unsure/No Data	int	Percent of polled that left the question blank or said “unsure”

Table 1 - approval_rating.csv

Column Name	Type	Description
Observation Date	Datetime	The date of the data being gathered. One data point per year
MEHOINUSA672N	int	Median household monthly income in dollars
Change	float	Percent yearly change in median household income

Table 2 - median_household_income_daily_change.csv

Column Name	Type	Description
Observation Date	Datetime	The date of the data being gathered. One data point per 3 months
A939RX0Q048SBEA	int	GDP per capita adjusted for inflation
Change	float	Percent monthly change in GDP per capita

Table 3 - real_GDP_per_capita_daily_change.csv

Column Name	Type	Description
Observation Date	Datetime	The date of the data being gathered. One data point per month
UNRATE	int	GDP per capita adjusted for inflation

Table 4 - unemployment_rate.csv

Column Name	Type	Description
Observation Date	Datetime	The date of the data being gathered. One data point per day
Close	float	The value of the S&P500 in dollars at closing on that day
Open	float	The value of the S&P500 in dollars at opening time

High	float	The high value of the S&P500 in dollars that day
Close_Change	float	The change in value of the S&P500 per day compared to previous opening day

Table 5 - sp500_daily_change.csv

Figures:

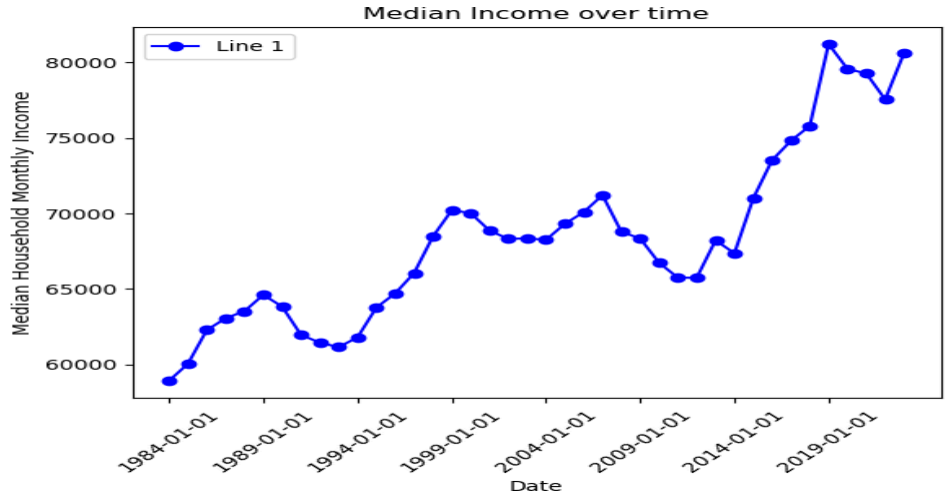

Figure 1 - Median Household Income per year

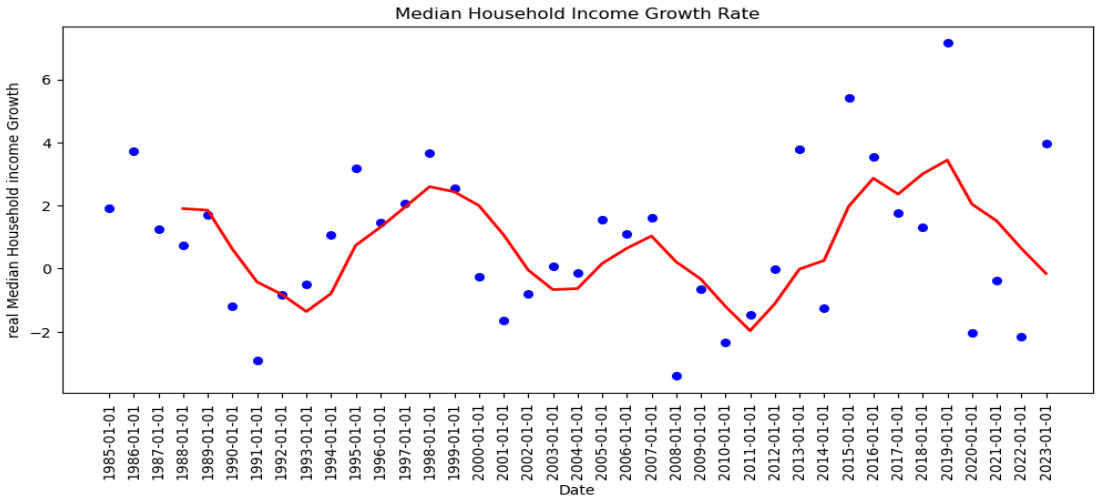

Figure 2 - Yearly Percent Growth of Median Household Monthly Income (with 4 year rolling average)

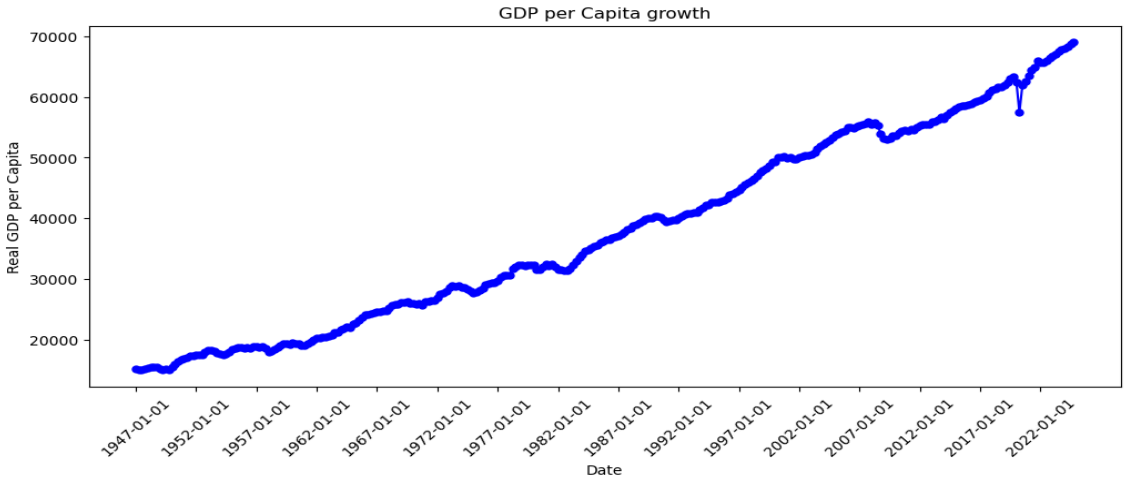

Figure 3 - real GDP per capita over time

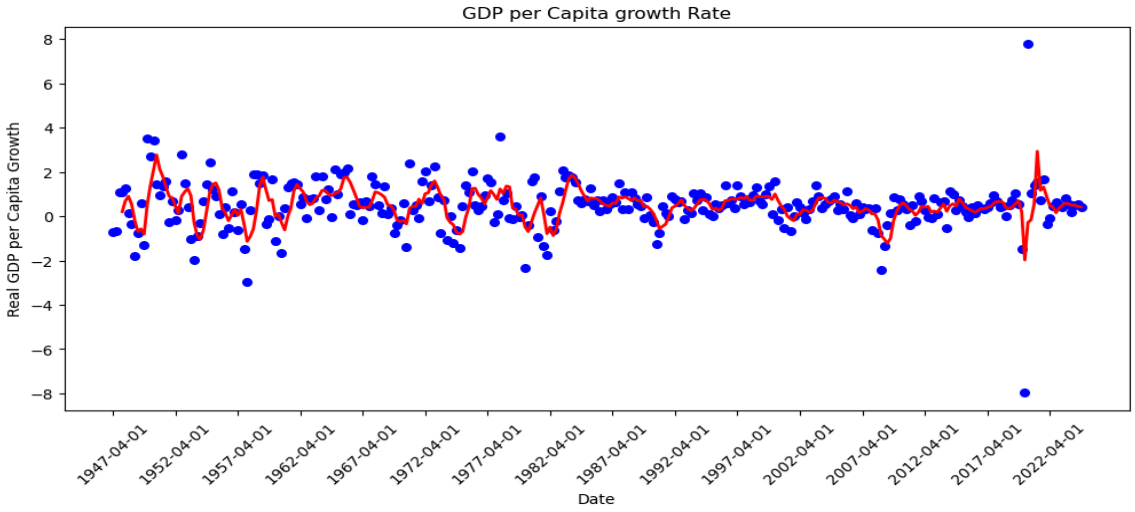

Figure 4 - real GDP per capita growth rate (%) over time

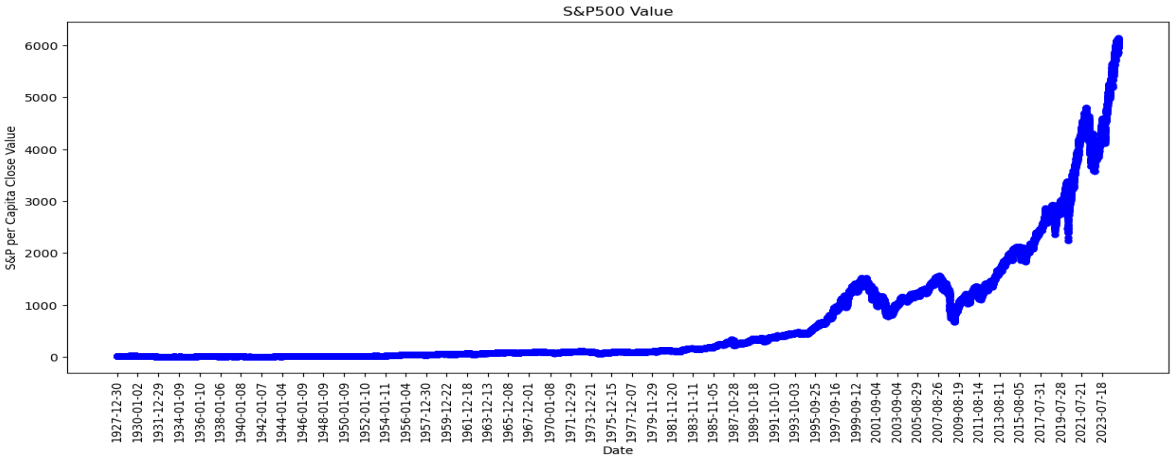

Figure 5 - S&P500 daily close valuation over time

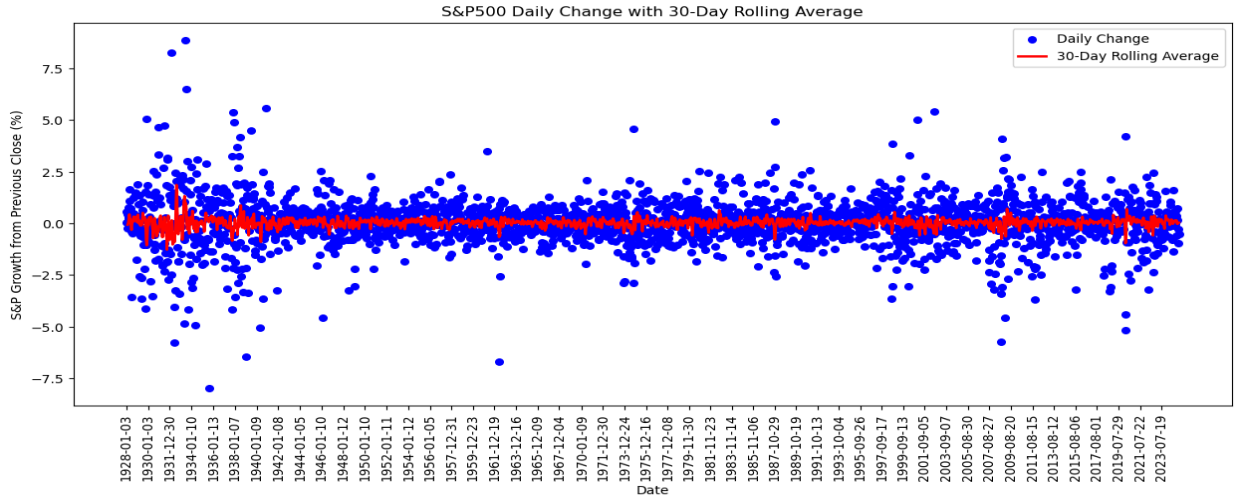

Figure 6 - S&P500 Daily change in close valuation over time

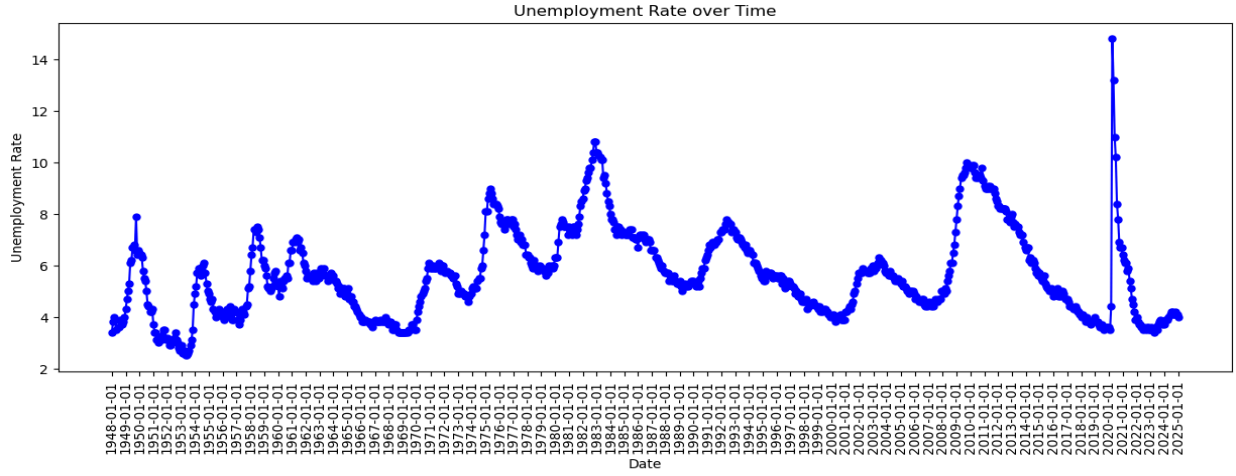

Figure 7 - Unemployment rate over time (%)

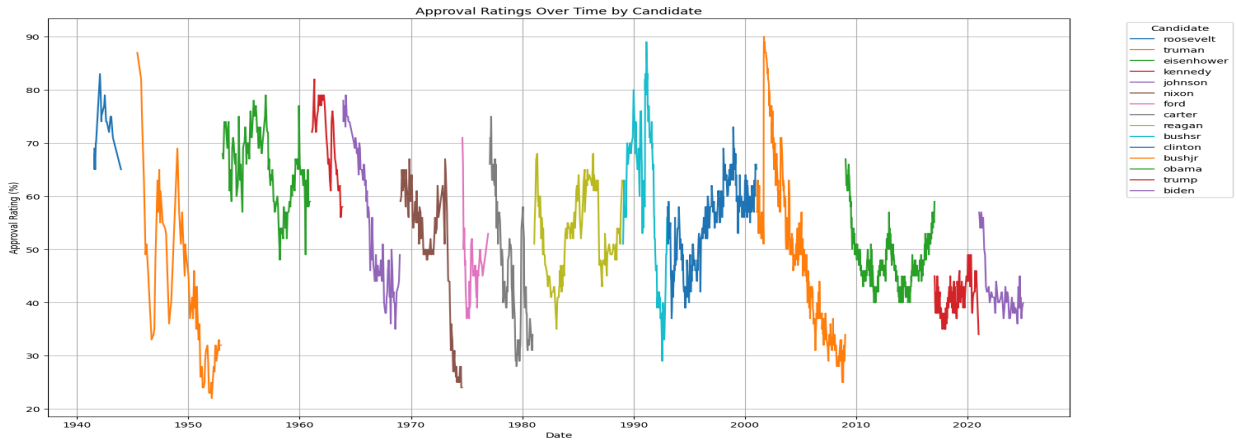

Figure 8 - Approval Ratings Over Time by Candidate (%)