Daily Calorie Intake Analysis

Andre Dutta

Background

- As an aspiring data analyst, I like to examine data and use it to make better decisions.
- I am very interested in fields related to health, wellness, and nutrition.
- I wanted to combine these two interests and create something that can be analyzed and visualized.

Background Continued

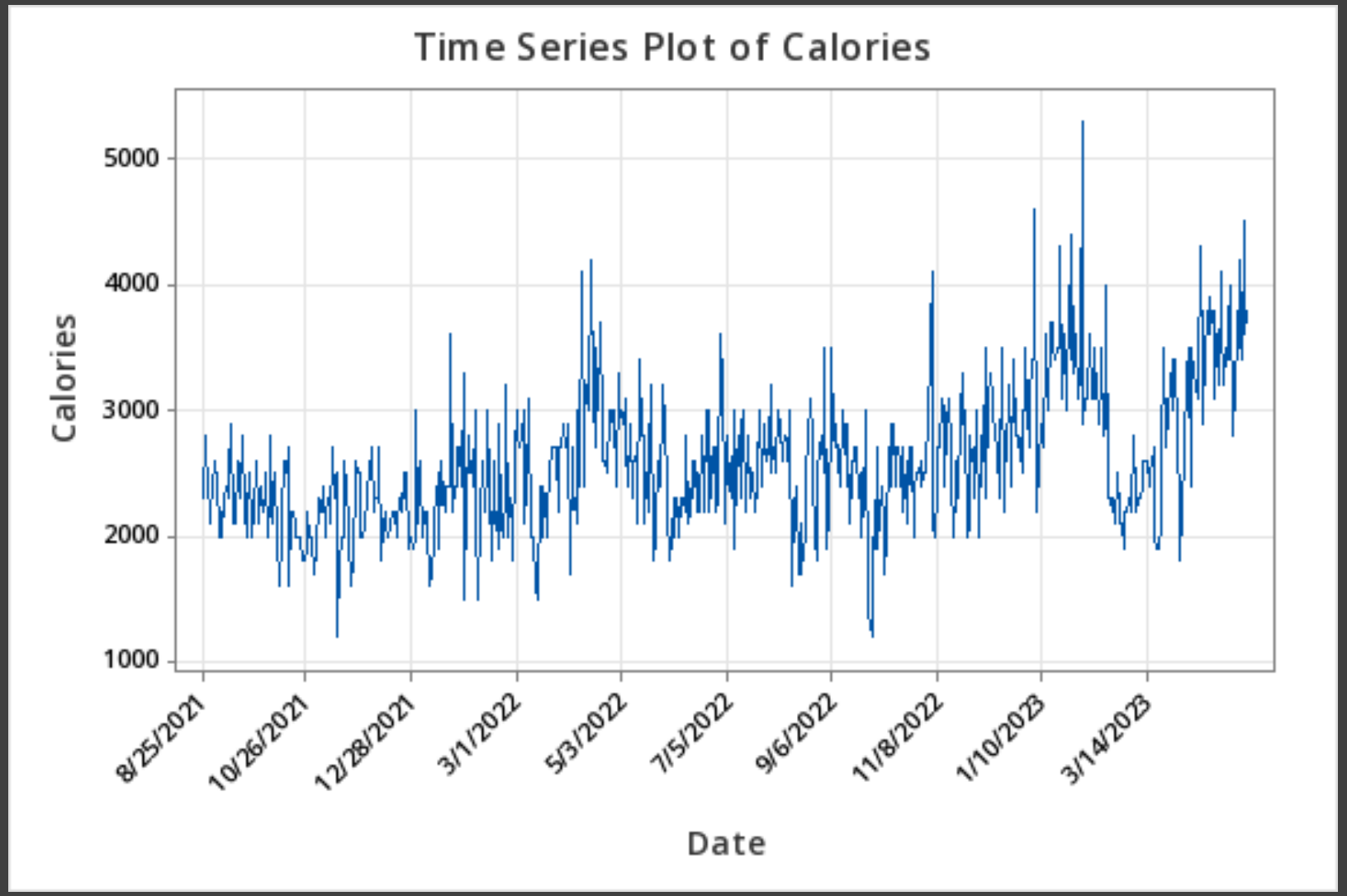
- Starting on August 25, 2021, I started tracking my caloric intake everyday
- I still track my calories today, however the data I used goes up to May 13, 2023
- I log my calories everyday in my Notes app on my phone, then I transferred this data over to Excel and Minitab to perform calculations and analysis

Objective

- The goal of my analysis is to identify trends, assess frequency and distribution, and create models to forecast future calories.

Time Series Plot

- Here is a simple time series plot of the data



Basic Statistics

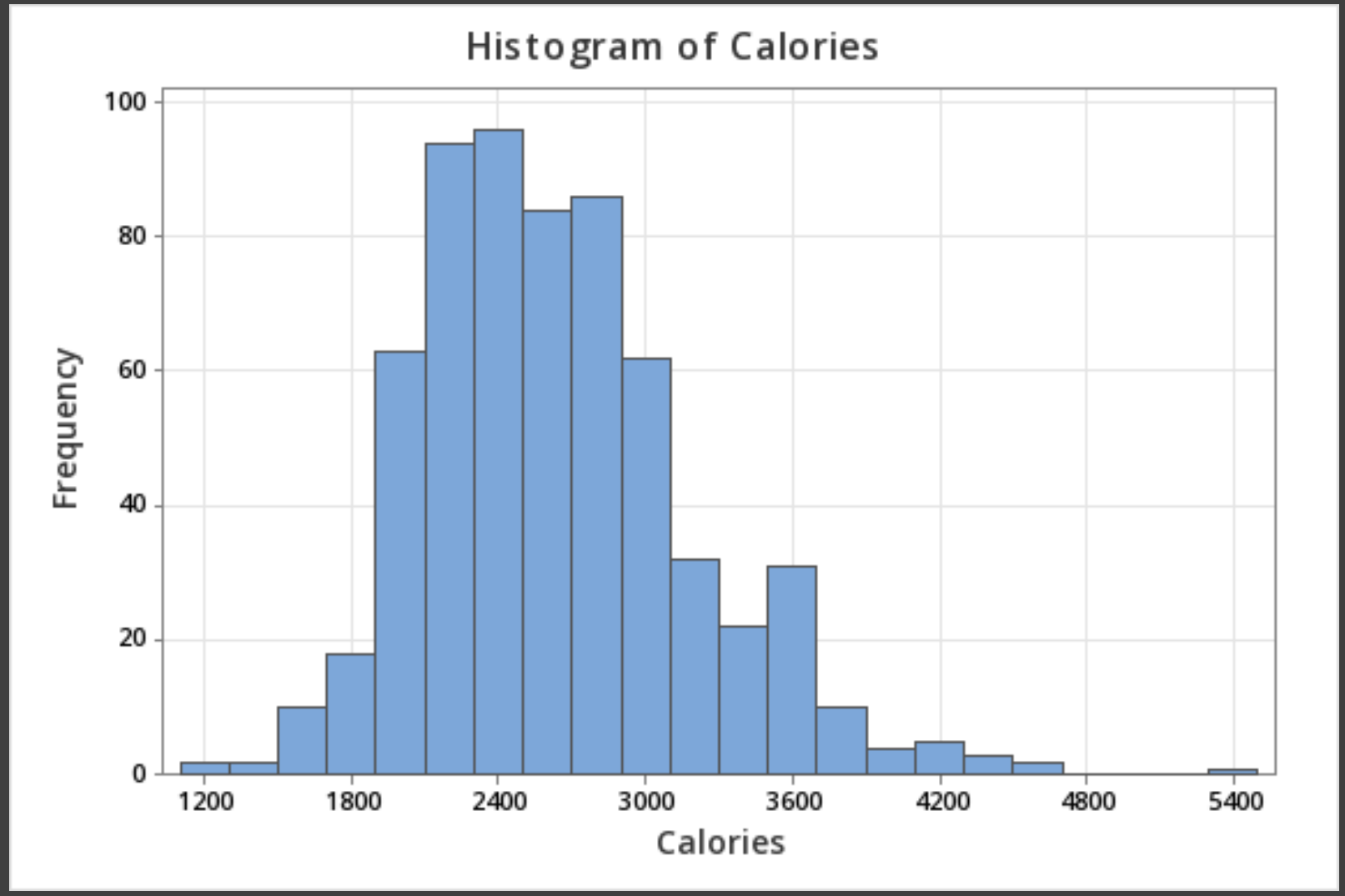
- Below are some descriptive statistics of the data, including mean, median, and standard deviation

Number of Observations	Mean	Standard Error of Mean	Standard Deviation
627	2590.3	22.3	558.9

Minimum	Q1	Median	Q3	Maximum	Range	Mode	Skewness
1200	2200	2500	2900	5300	4100	2200	0.79

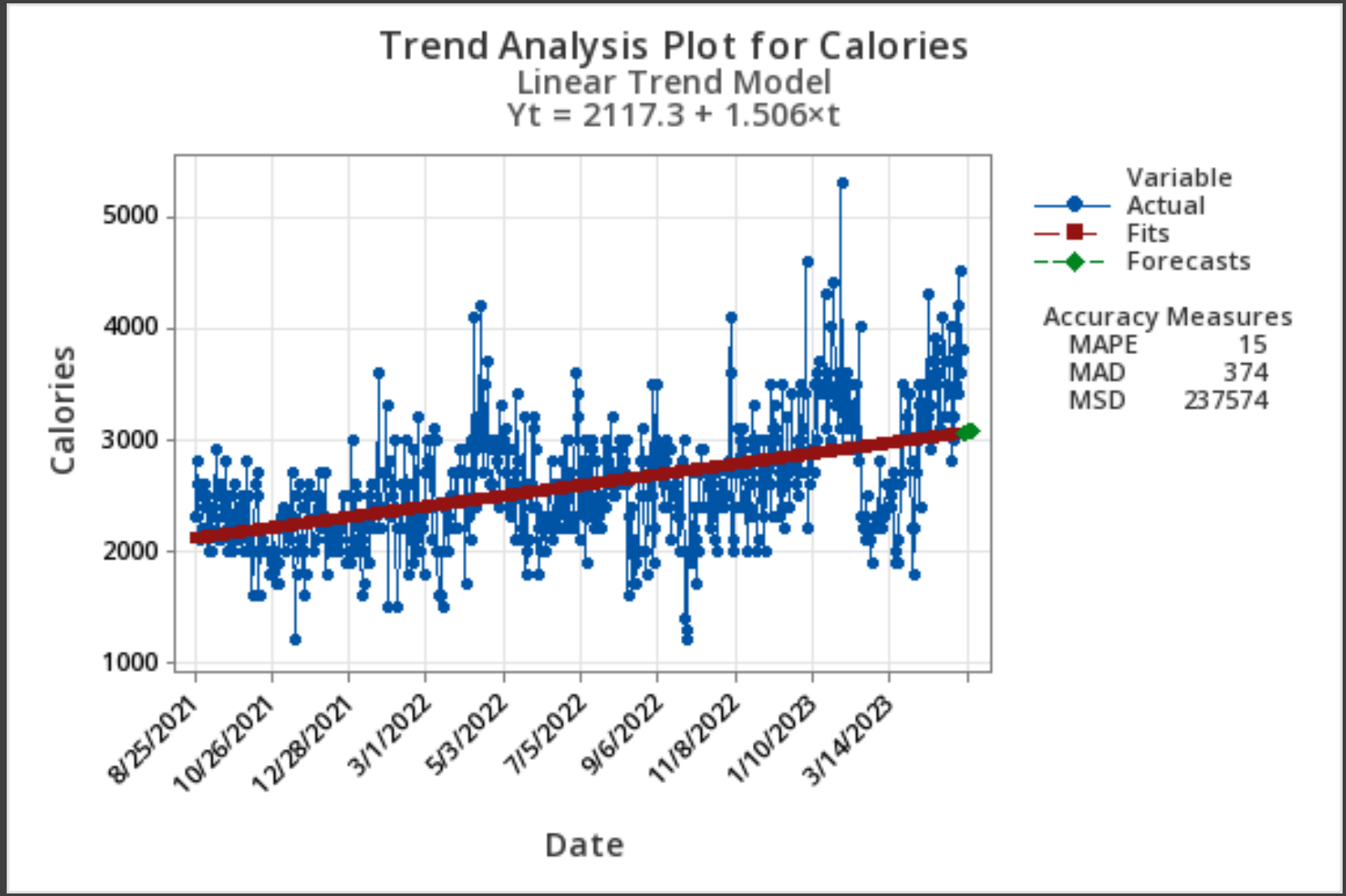
Histogram

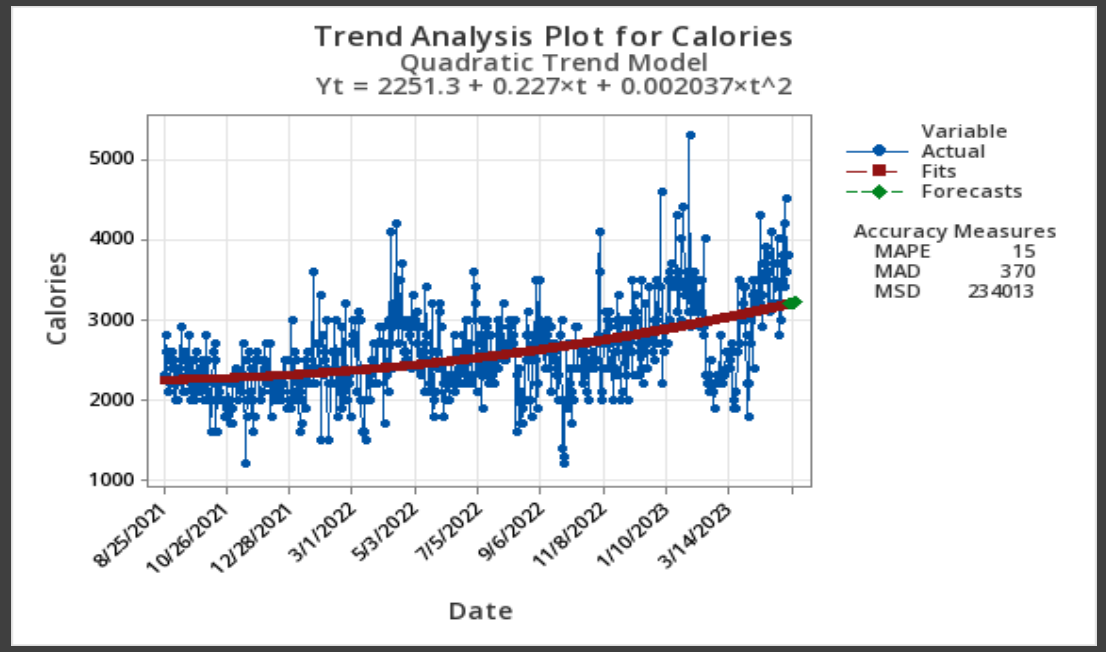
- The data takes on a unimodal distribution, with a slight positive skew

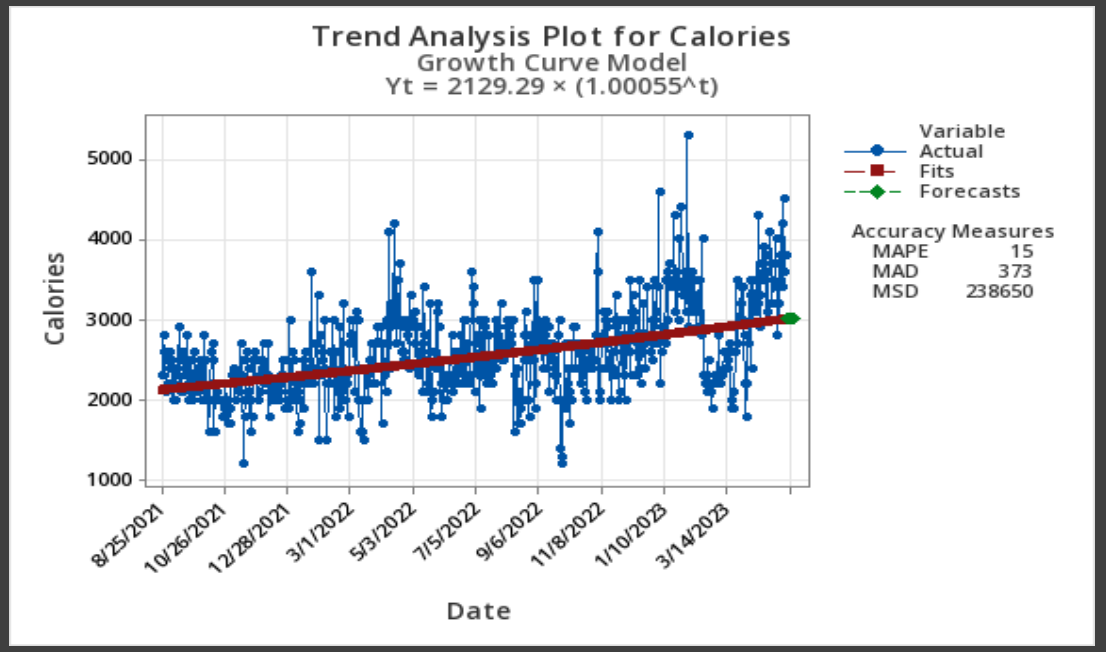


Trend Analysis: Linear Model

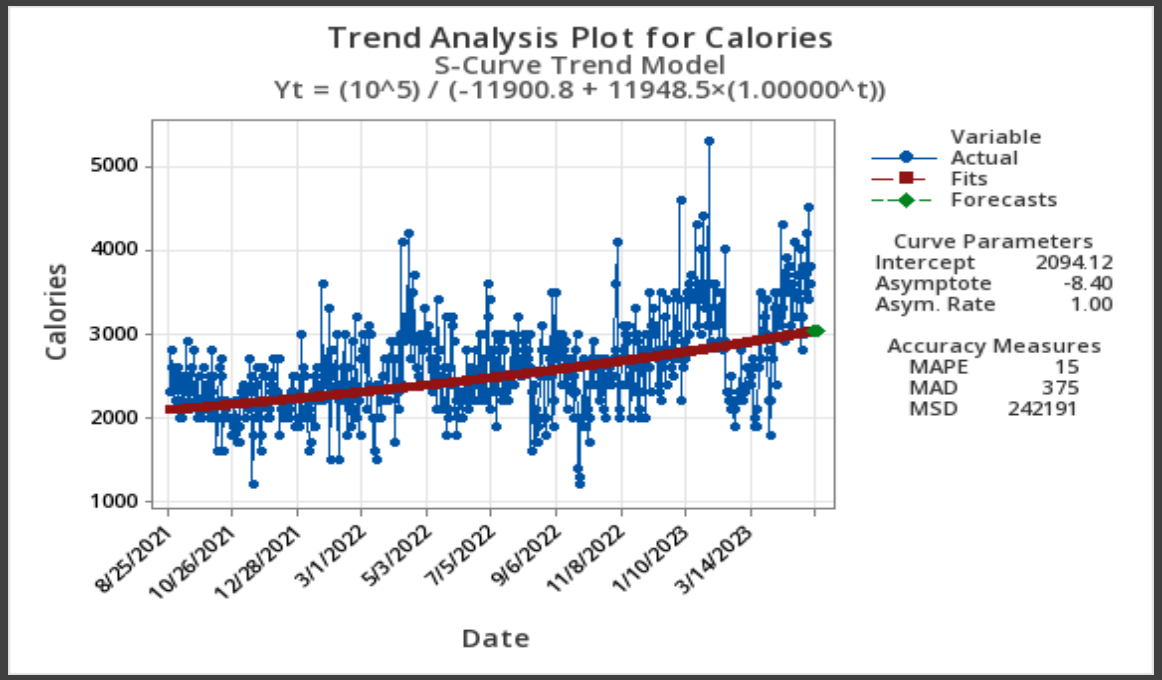
- There is an upward trend in the data, as calories increase over time on average.
- I created a 7-day forecast that shows where we expect future values to be
- Because the residuals are large, the forecasts will likely be not very accurate







- Here are more trend analysis graphs, using Quadratic, Growth, and S-Curve models



- Note that the residuals are also very large

Regression Analysis

- Now we will run a regression of calories versus index, with the index being the observation number at each date
  - For instance, the date 8/25/2021 has an index number of 1 and 5/13/2023 has an index of 627

- The equation is:

  Calories = 2117.3 + 1.506 * Index

- Note that this is the same equation as the one in the linear trend model

Regression Analysis Continued

- If we want to predict the amount of calories consumed for 5/14/2023, which has an index of 628, we just plug the index into the equation

- Calories = 2117.3 + 1.506 * 628

  =3063.3

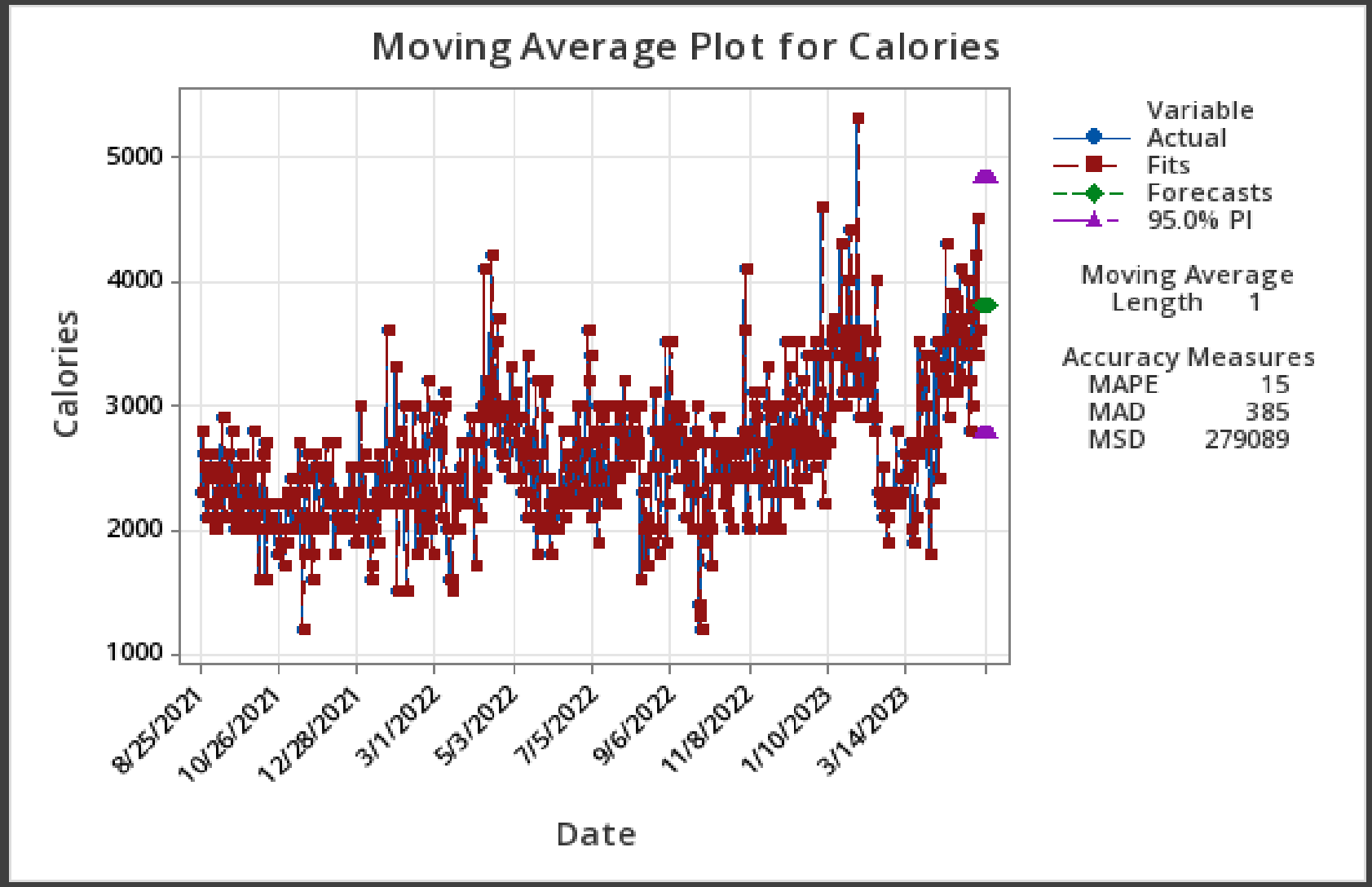
* I rounded the slope and intercept here so my answer will be slightly different than if you directly plug 628 into the equation above

Regression Analysis Continued

- Now that we have our regression model, it is worth noting that the R^2 value is 23.83%
- R^2 is the coefficient of determination and measures goodness of fit
- This means that only 23.83% of the dependent variable (calories) is explained by the independent variable (index)
- The R^2 value is very low when looking at how well the regression line fits the data
- Therefore, linear regression may not be the best method for forecasting caloric intake

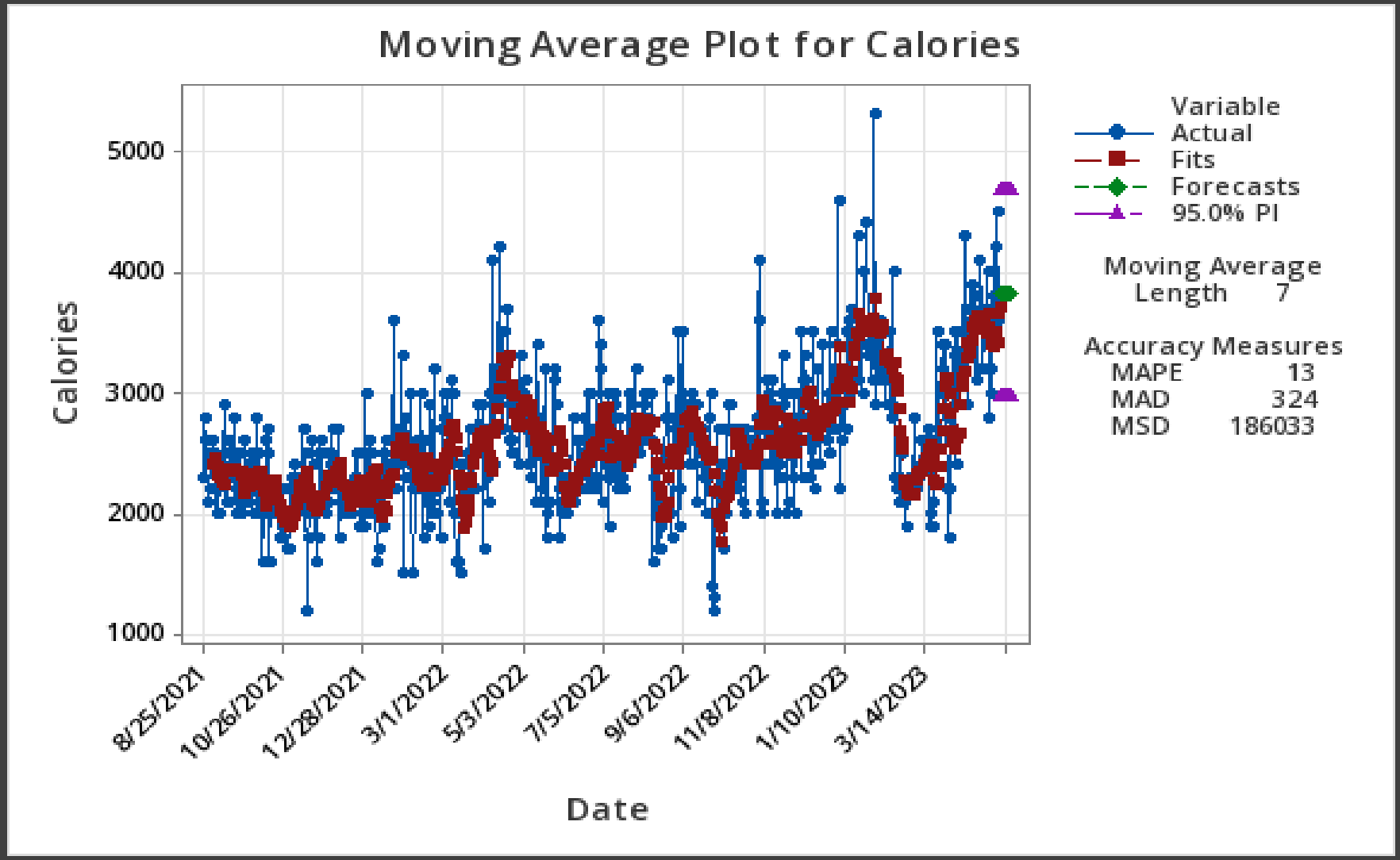
Naïve Model

- Next, we will look at some other forecasting methods
- Here is a naïve model, with a 7-day forecast and a 95% prediction interval



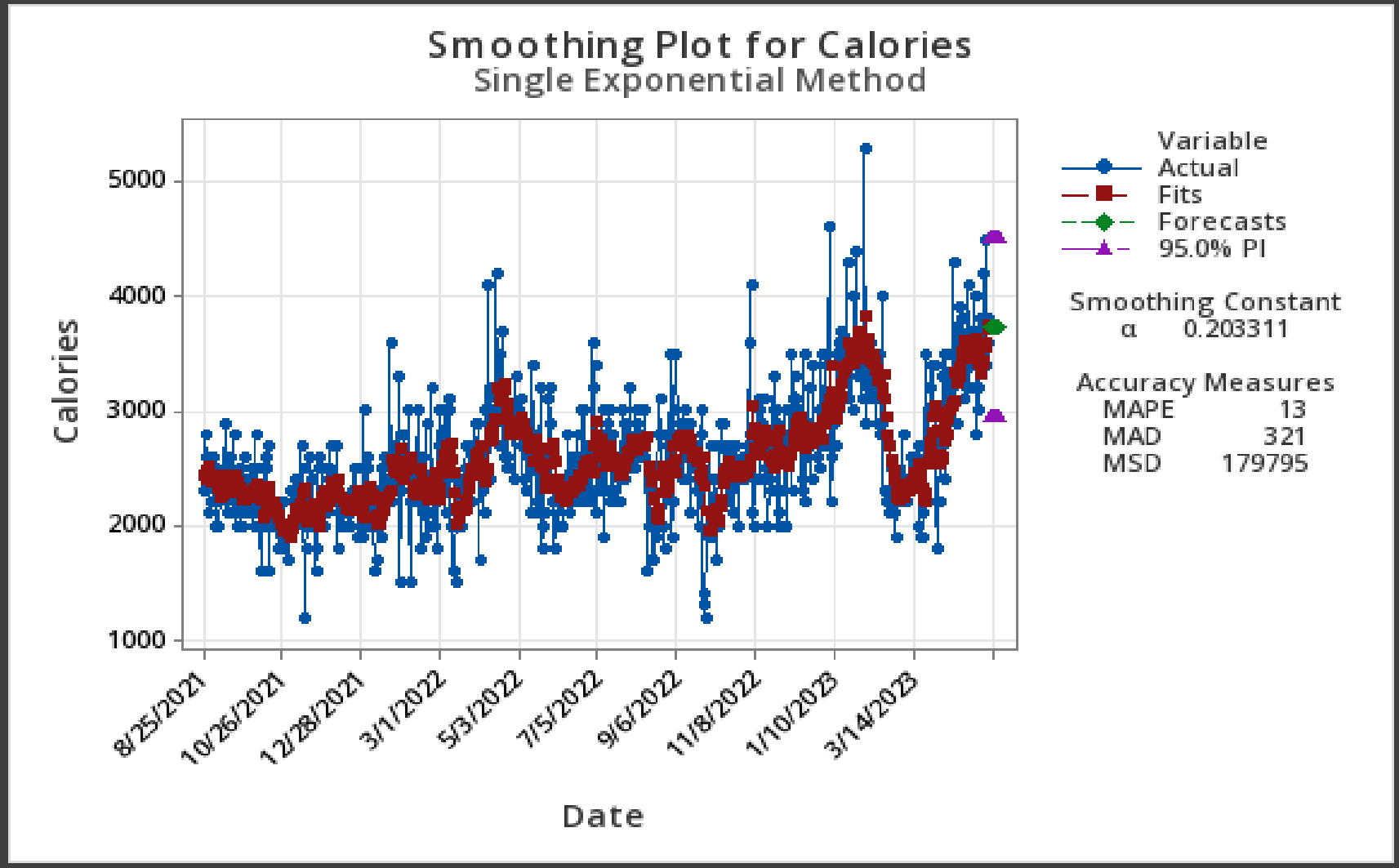
Moving Average Model

- Here is a simple moving average model with length 7
- Note that the accuracy measures all have lower values than those of the naïve model



Exponential Smoothing Method

- Lastly, here is a single exponential smoothing model with an alpha value of .2
- Note the accuracy measures compared to the other models



Choosing the Best Model

- If we look at the Mean Absolute Percentage Error (MAPE), Mean Absolute Deviation (MAD), and Mean Squared Deviation (MSD) of each model, the exponential smoothing method has the lowest values overall

- The residuals in this model will also be the smallest

- If we assume that I do not control the number of calories I eat each day, the best model out of the ones I used will be from the exponential smoothing method

Conclusion

- It is worth noting that there are other models I did not use that may give better predictions, such as Winters' Method or ARIMA modeling
- Nonetheless, it is fun to analyze and compare the different models while knowing that I can apply these methods to other real-world scenarios
- Thank you for reading!