Phase 1 Report

Electricity Prices Prediction

Phase 1: Problem Definition and Design Thinking

Problem Definition:

The problem is to develop a predictive model that uses historical electricity prices and relevant factors to forecast future electricity prices.

The objective is to create a tool that assists both energy providers and consumers in making informed decisions regarding consumption and investment by predicting future electricity prices.

This project involves data preprocessing, feature engineering, model selection, training, and evaluation.

Design Thinking:

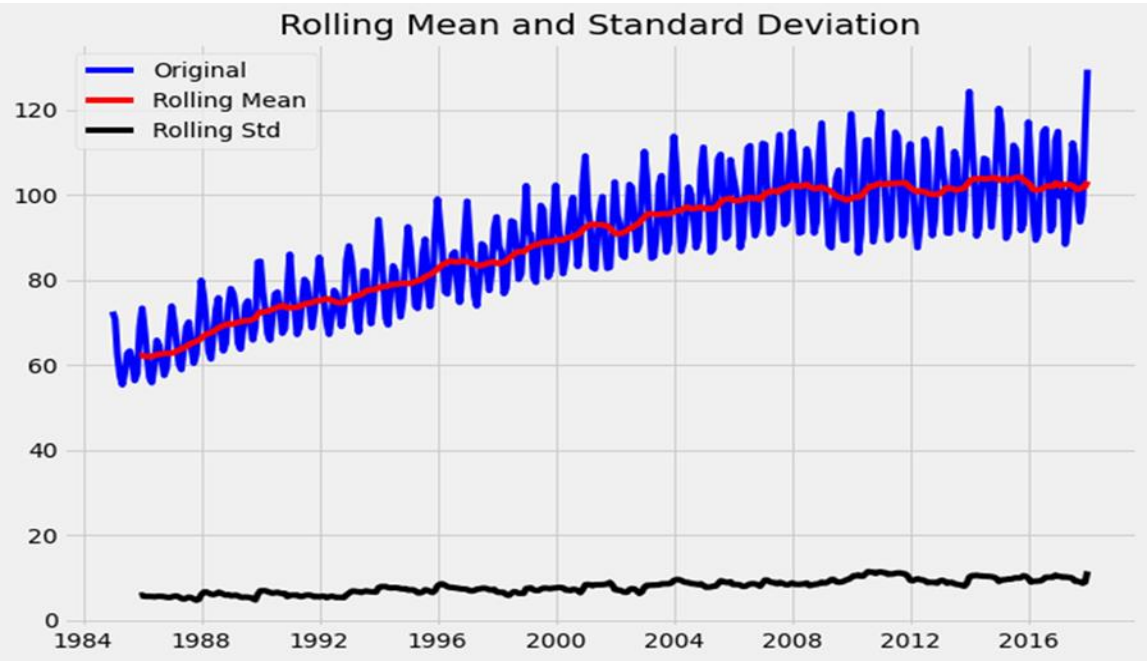
1. Data Source:

Utilize a dataset containing historical electricity prices and relevant factors like date, demand, supply, weather conditions, and economic indicators.

DateTime	Holiday	HolidayFlag	DayOfWeek	WeekOfYear	Day	Month	Year	PeriodOfDay
38014 unique values	None 96% Christmas Eve 0% Other (1392) 4%	0 1	0 6	1 52	1 31	1 12	2011 2013	0
01/11/2011 00:00	None	0	1	44	1	11	2011	0
01/11/2011 00:30	None	0	1	44	1	11	2011	1
01/11/2011 01:00	None	0	1	44	1	11	2011	2
01/11/2011 01:30	None	0	1	44	1	11	2011	3
01/11/2011 02:00	None	0	1	44	1	11	2011	4
01/11/2011 02:30	None	0	1	44	1	11	2011	5
01/11/2011 03:00	None	0	1	44	1	11	2011	6
01/11/2011 03:30	None	0	1	44	1	11	2011	7
01/11/2011 04:00	None	0	1	44	1	11	2011	8
01/11/2011 04:30	None	0	1	44	1	11	2011	9
01/11/2011 05:00	None	0	1	44	1	11	2011	10
01/11/2011 05:30	None	0	1	44	1	11	2011	11
01/11/2011 06:00	None	0	1	44	1	11	2011	12

2.Data Preprocessing:

Clean and preprocess the data, handle missing values, and convert categorical features into numerical representations.

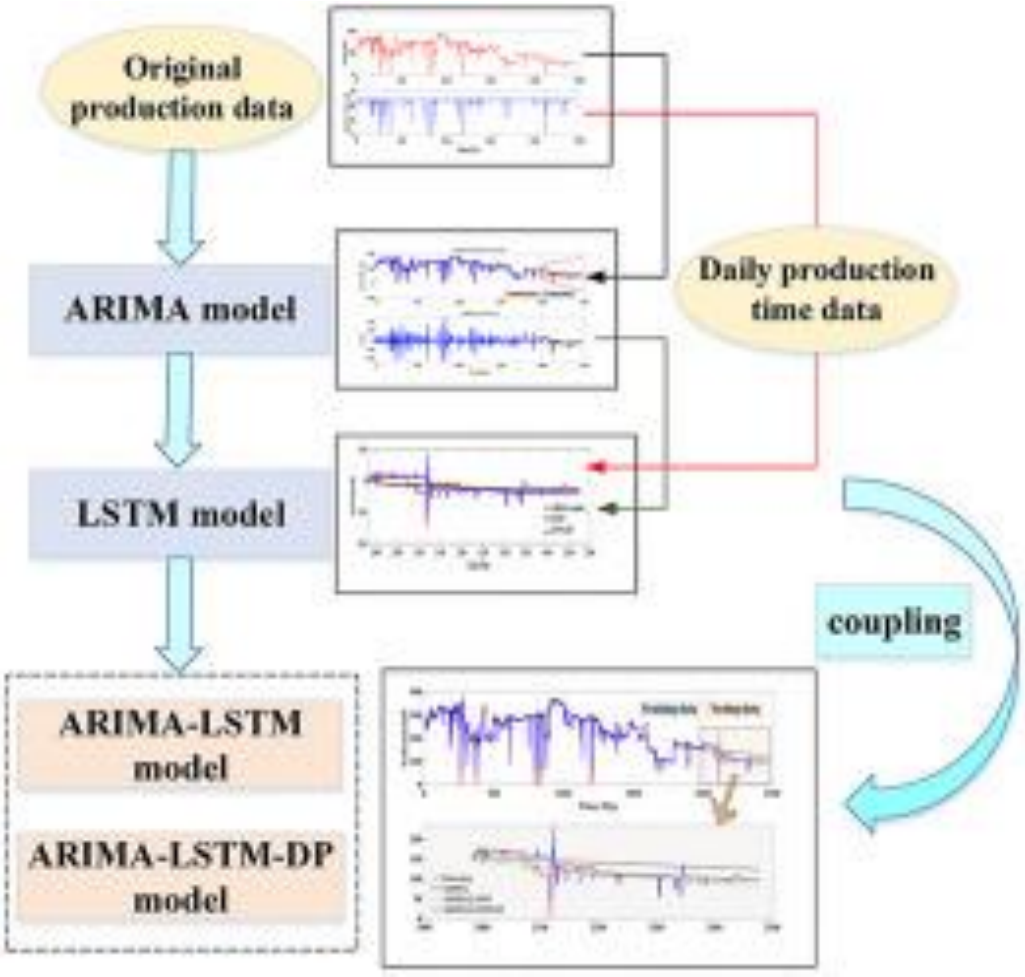


3.Feature Engineering:

Create additional features that could enhance the predictive power of the model, such as time-based features and lagged variables.

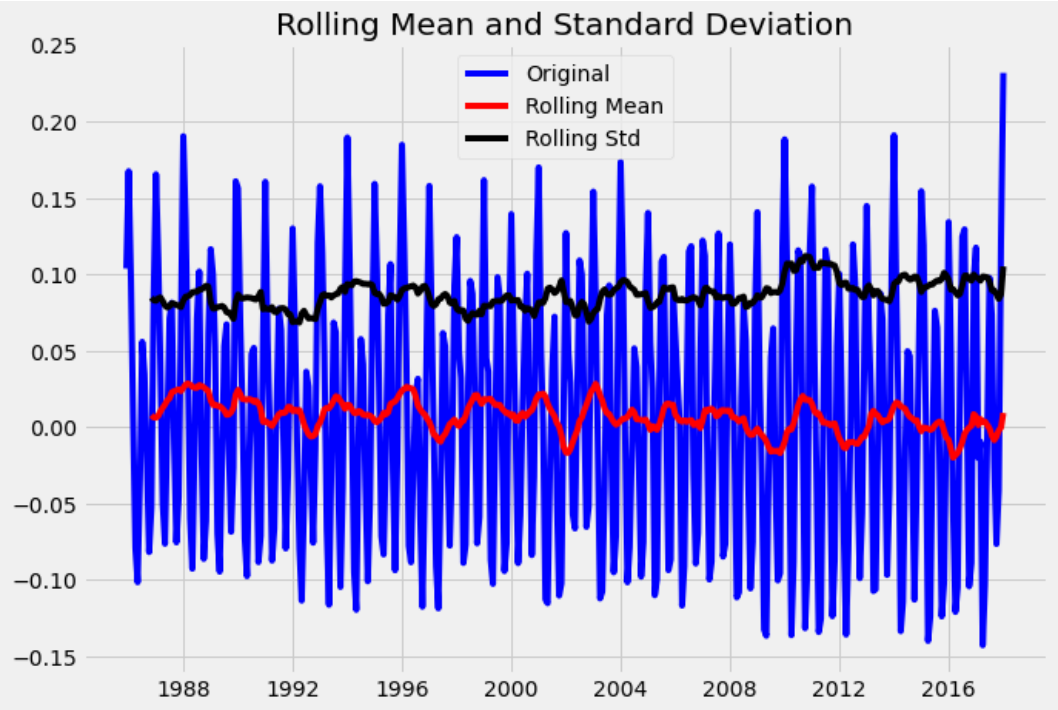
4.Model Selection:

Choose suitable time series forecasting algorithms (e.g., ARIMA, LSTM) for predicting future electricity prices.



5.Model Training:

Train the selected model using the preprocessed data.



6.Evaluation:

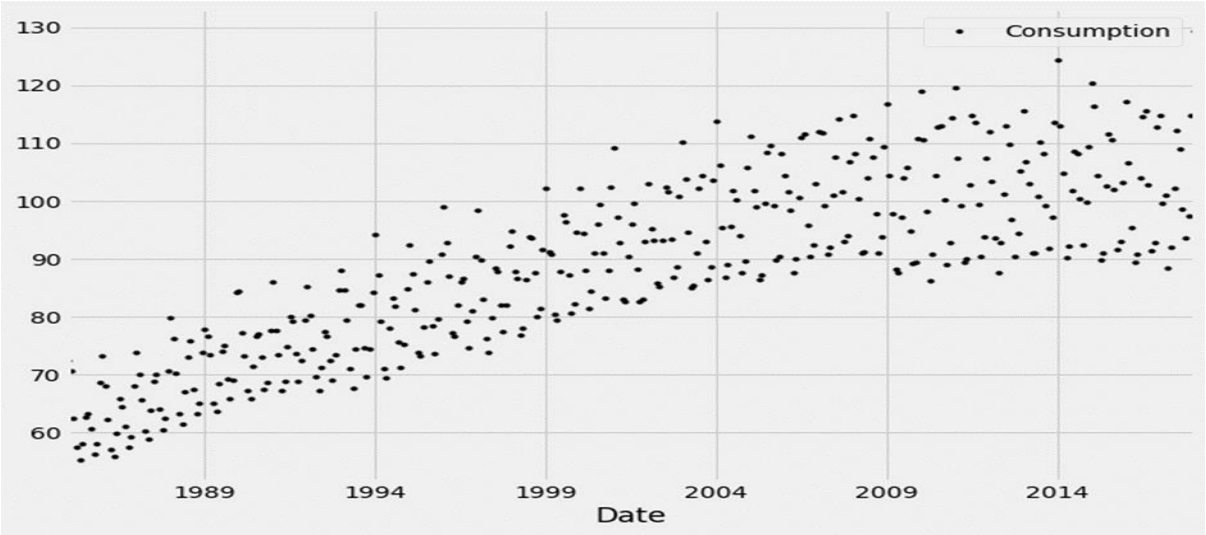
Evaluate the model's performance using appropriate time series forecasting metrics (e.g., Mean Absolute Error, Root Mean Squared Error).

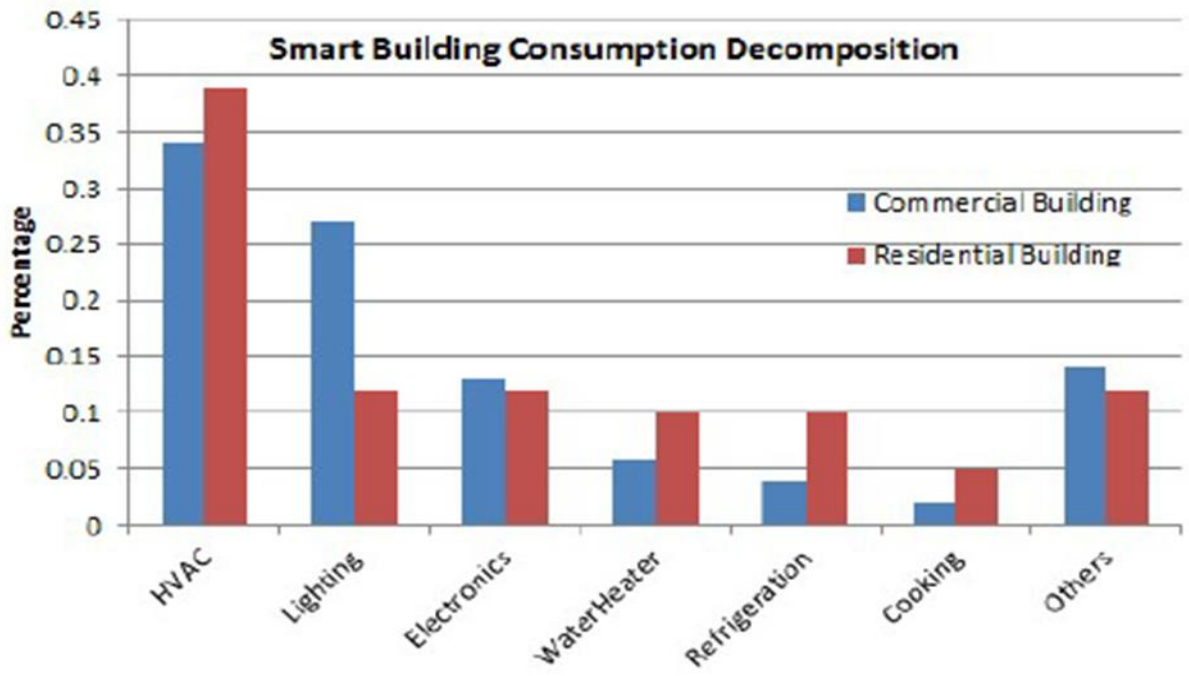
MAE = 1/n * sum(abs(actual - predicted))

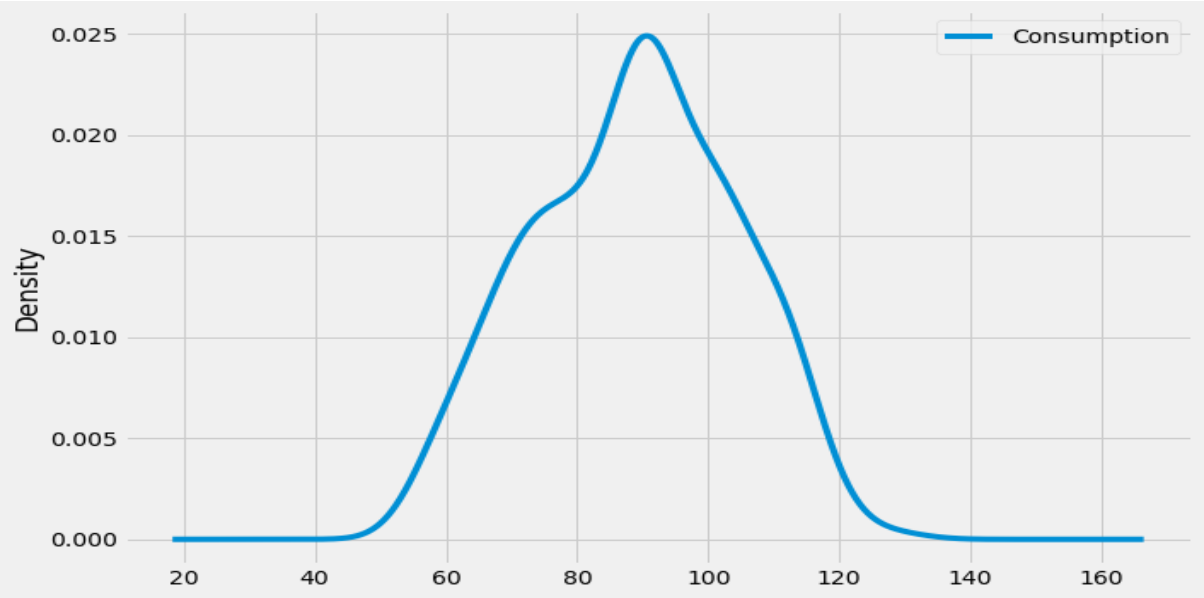
RMSE = sqrt(1/n * sum((actual - predicted)^2))

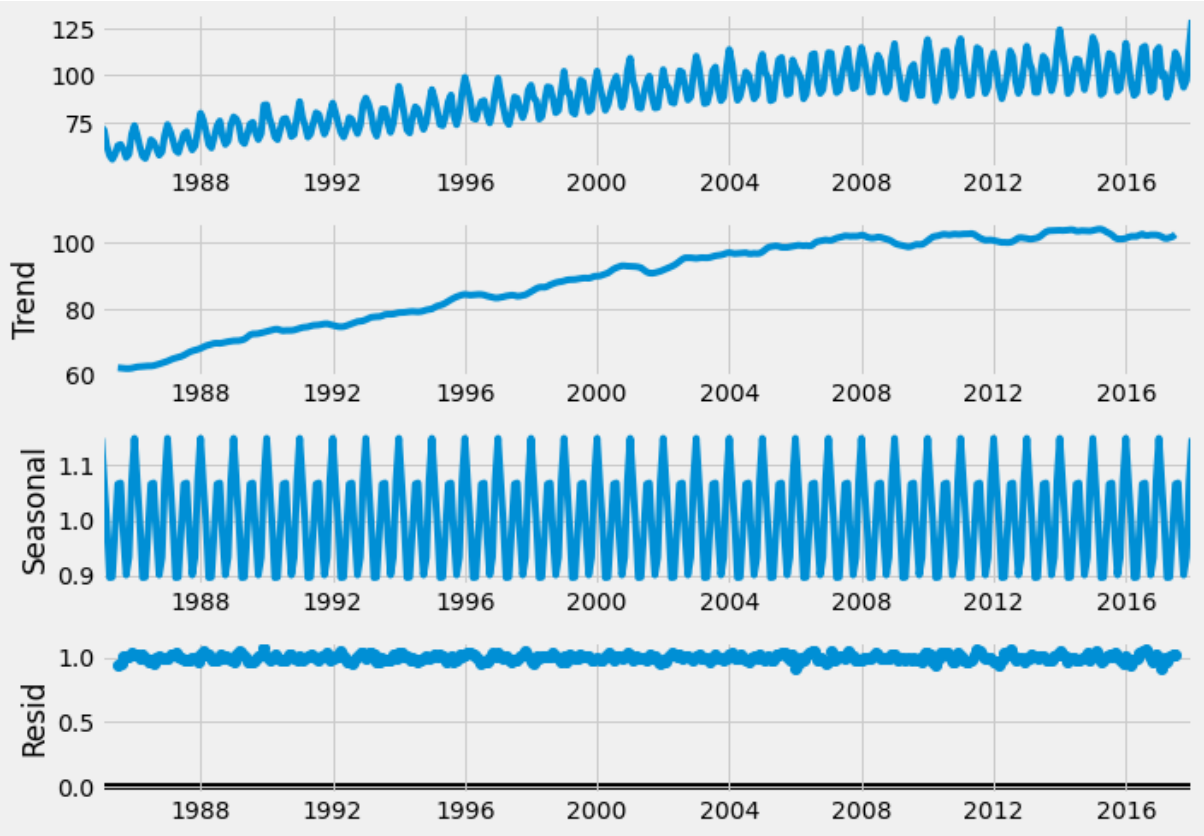
7. **Visualization Selection**:

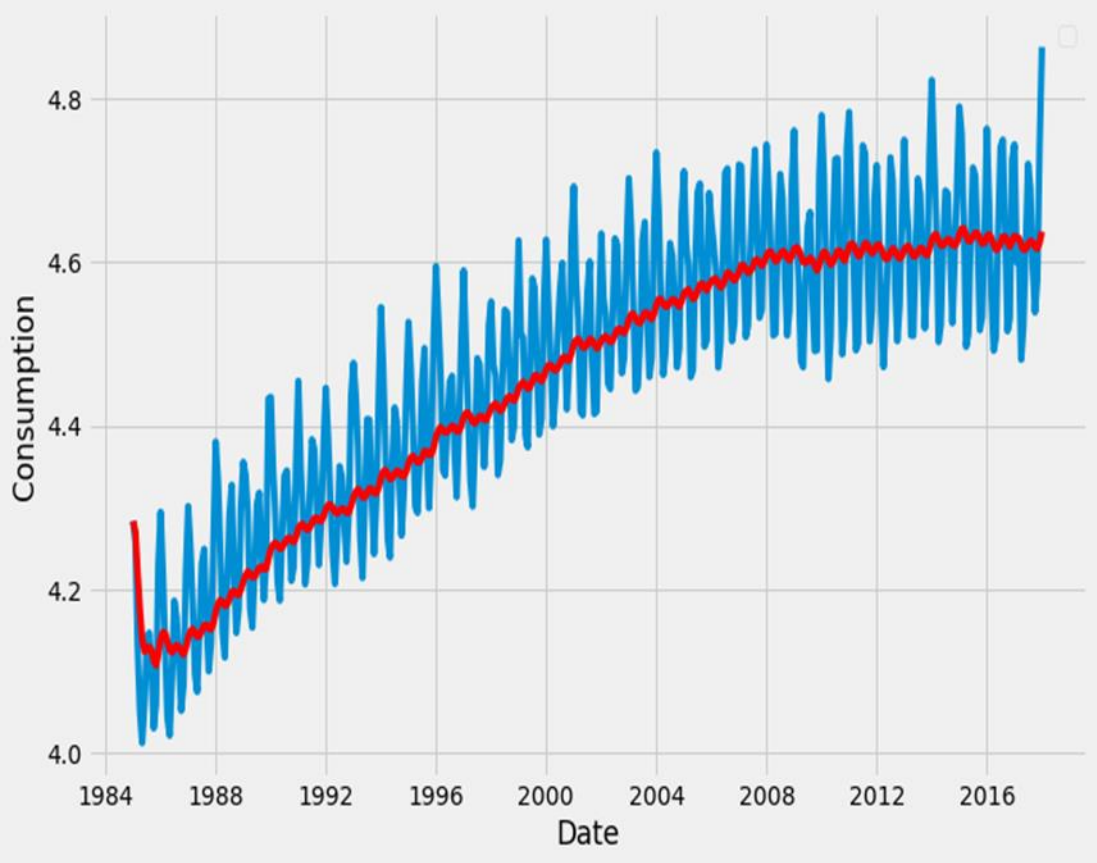
Determine suitable visualization types (e.g., bar charts, pie charts, heatmaps) to represent demographic distributions effectively.

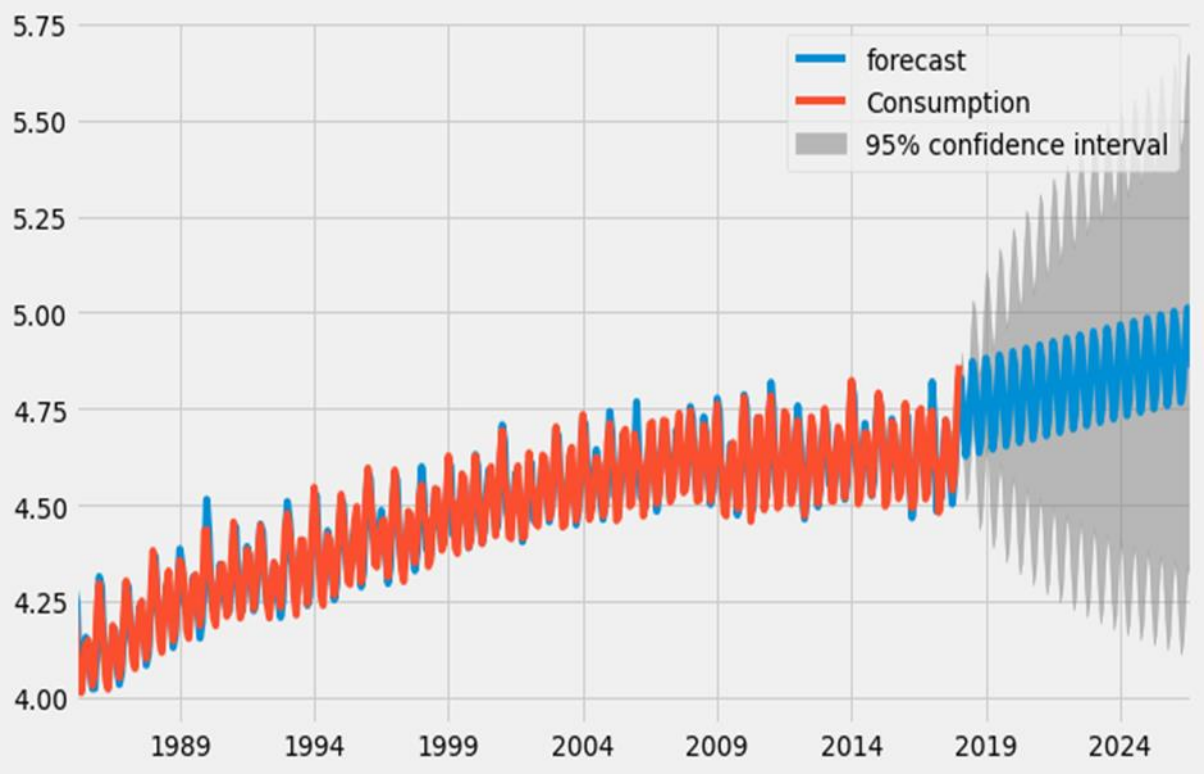





Consumption
Density
0.025
0.020
0.015
0.010
0.005
0.000
20
40
60
80
100
120
140
160


125
100
75
1988
1992
1996
2000
2004
2008
2012
2016
Trend
100
80
60
Seasonal
1.1
1.0
0.9
Resid
1.0
0.5
0.0


Consumption
4.8
4.6
4.4
4.2
4.0
1984
1988
1992
1996
2000
2004
2008
2012
2016
Date


forecast
Consumption
95% confidence interval
5.75
5.50
5.25
5.00
4.75
4.50
4.25
4.00
1989
1994
1999
2004
2009
2014
2019
2024