Data Collection and Preprocessing

In [13]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# Fetch data with an extended date range
from fredapi import Fred
fred = Fred(api_key='6b6f56be8ed7dafee7f6f255d71b8f8c')

# Fetch data from FRED API
unemployment_rate = fred.get_series('UNRATE', observation_start='2000-01-01')
initial_jobless_claims = fred.get_series('ICSA', observation_start='2000-01-01')
consumer_sentiment_index = fred.get_series('UMCSENT', observation_start='2000-01-01')
housing_starts = fred.get_series('HOUST', observation_start='2000-01-01')
sp500_index = fred.get_series('SP500', observation_start='2000-01-01')
ten_year_treasury_yield = fred.get_series('GS10', observation_start='2000-01-01')

# Combine all data into a single DataFrame
combined_data = pd.DataFrame({
    'Unemployment Rate': unemployment_rate,
    'Initial Jobless Claims': initial_jobless_claims,
    'Consumer Sentiment Index': consumer_sentiment_index,
    'Housing Starts': housing_starts,
    'S&P 500 Index': sp500_index,
    '10-Year Treasury Yield': ten_year_treasury_yield
})

# Resample data to monthly, using the last observation for each month
combined_data = combined_data.resample('M').last()

# Load the Daily Treasury Par Yield Curve Rates data
yield_curve_data = pd.read_csv(r'C:\Users\salin\OneDrive\Desktop\DSC680 Applied Data Science\week
yield_curve_data['Date'] = pd.to_datetime(yield_curve_data['Date'])
yield_curve_data.set_index('Date', inplace=True)

# Load the Daily Treasury Long-Term Rates data
long_term_rates_data = pd.read_csv(r'C:\Users\salin\OneDrive\Desktop\DSC680 Applied Data Science\
long_term_rates_data['Date'] = pd.to_datetime(long_term_rates_data['Date'])
long_term_rates_data.set_index('Date', inplace=True)

# Merge yield_curve_data with combined_data
combined_data = pd.merge(combined_data, yield_curve_data, left_index=True, right_index=True, how=

# Merge long_term_rates_data with the already merged data
combined_data = pd.merge(combined_data, long_term_rates_data, left_index=True, right_index=True,

# Advanced imputation for missing values in S&P 500 Index
combined_data['S&P 500 Index'] = combined_data['S&P 500 Index'].interpolate(method='time').fillna

# Verify missing values after advanced interpolation
missing_values = combined_data.isnull().sum()
print(missing_values)

# Handle any potential NaN values (forward fill as an example)
combined_data.fillna(method='ffill', inplace=True)

# Inspect the data
print(combined_data.info())
print(combined_data.describe())

# Check for missing values
```

```
missing_values = combined_data.isnull().sum()
print(missing_values)
```

```
Unemployment Rate             1
Initial Jobless Claims        0
Consumer Sentiment Index      1
Housing Starts                1
S&P 500 Index                 0
10-Year Treasury Yield        1
1 Mo                        102
2 Mo                        248
3 Mo                         89
4 Mo                        282
6 Mo                         89
1 Yr                         89
2 Yr                         89
3 Yr                         89
5 Yr                         89
7 Yr                         89
10 Yr                        89
20 Yr                        89
30 Yr                       122
LT COMPOSITE (>10 Yrs)       89
TREASURY 20-Yr CMT           89
Extrapolation Factor        260
dtype: int64
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 293 entries, 2000-01-31 to 2024-05-31
Freq: M
Data columns (total 22 columns):
 #   Column                    Non-Null Count  Dtype
---  ------                    --------------  -----
 0   Unemployment Rate         293 non-null    float64
 1   Initial Jobless Claims    293 non-null    float64
 2   Consumer Sentiment Index  293 non-null    float64
 3   Housing Starts            293 non-null    float64
 4   S&P 500 Index             293 non-null    float64
 5   10-Year Treasury Yield    293 non-null    float64
 6   1 Mo                      275 non-null    float64
 7   2 Mo                      68 non-null     float64
 8   3 Mo                      293 non-null    float64
 9   4 Mo                      20 non-null     float64
 10  6 Mo                      293 non-null    float64
 11  1 Yr                      293 non-null    float64
 12  2 Yr                      293 non-null    float64
 13  3 Yr                      293 non-null    float64
 14  5 Yr                      293 non-null    float64
 15  7 Yr                      293 non-null    float64
 16  10 Yr                     293 non-null    float64
 17  20 Yr                     293 non-null    float64
 18  30 Yr                     293 non-null    float64
 19  LT COMPOSITE (>10 Yrs)    293 non-null    float64
 20  TREASURY 20-Yr CMT        293 non-null    float64
 21  Extrapolation Factor      268 non-null    float64
dtypes: float64(22)
memory usage: 60.8 KB
None
       Unemployment Rate  Initial Jobless Claims  Consumer Sentiment Index  \
count         293.000000            2.930000e+02                293.000000
mean            5.730717            3.862321e+05                 83.205461
std             1.970626            4.065895e+05                 13.216608
min             3.400000            1.870000e+05                 50.000000
25%             4.200000            2.610000e+05                 73.500000
50%             5.100000            3.320000e+05                 84.900000
75%             6.600000            3.990000e+05                 93.500000
max            14.800000            5.946000e+06                112.000000

       Housing Starts  S&P 500 Index  10-Year Treasury Yield        1 Mo  \
count      293.000000     293.000000              293.000000  275.000000
mean      1304.610922    2425.492833                3.263311    1.477273
```

```
std        437.032543     863.068638               1.310693    1.748011
min        478.000000    1920.030000               0.620000    0.000000
25%       1002.000000    1923.570000               2.190000    0.070000
50%       1311.000000    1923.570000               3.200000    0.900000
75%       1625.000000    2673.610000               4.260000    2.175000
max       2273.000000    5304.720000               6.660000    5.560000

           2 Mo        3 Mo       4 Mo  ...        2 Yr        3 Yr  \
count  68.000000  293.000000  20.000000  ...  293.000000  293.000000
mean    2.227353    1.777952   5.248000  ...    2.145358    2.320717
std     2.123580    1.954917   0.418941  ...    1.767449    1.658739
min     0.010000    0.000000   4.330000  ...    0.110000    0.110000
25%     0.097500    0.100000   4.970000  ...    0.580000    0.920000
50%     1.775000    1.070000   5.490000  ...    1.620000    1.900000
75%     4.347500    2.990000   5.507500  ...    3.600000    3.730000
max     5.570000    6.380000   5.610000  ...    6.690000    6.660000

             5 Yr        7 Yr       10 Yr       20 Yr       30 Yr  \
count  293.000000  293.000000  293.000000  293.000000  293.000000
mean     2.688396    2.993242    3.247065    3.761331    3.902082
std      1.498462    1.401972    1.314798    1.315270    1.254081
min      0.210000    0.390000    0.550000    0.980000    1.200000
25%      1.510000    1.920000    2.170000    2.660000    2.920000
50%      2.520000    2.900000    3.180000    3.910000    3.850000
75%      3.870000    4.090000    4.360000    4.810000    5.190000
max      6.710000    6.750000    6.680000    6.720000    6.490000

       LT COMPOSITE (>10 Yrs)  TREASURY 20-Yr CMT  Extrapolation Factor
count              293.000000          293.000000            268.000000
mean                 3.729966            3.761331             -0.036045
std                  1.265895            1.315270              0.047555
min                  1.040000            0.980000             -0.150000
25%                  2.710000            2.660000             -0.050000
50%                  3.820000            3.910000             -0.050000
75%                  4.730000            4.810000             -0.050000
max                  6.670000            6.720000              0.180000

[8 rows x 22 columns]
Unemployment Rate             0
Initial Jobless Claims        0
Consumer Sentiment Index      0
Housing Starts                0
S&P 500 Index                 0
10-Year Treasury Yield        0
1 Mo                         18
2 Mo                        225
3 Mo                          0
4 Mo                        273
6 Mo                          0
1 Yr                          0
2 Yr                          0
3 Yr                          0
5 Yr                          0
7 Yr                          0
10 Yr                         0
20 Yr                         0
30 Yr                         0
LT COMPOSITE (>10 Yrs)        0
TREASURY 20-Yr CMT            0
Extrapolation Factor         25
dtype: int64
```

Exploratory Data Analysis (EDA)

In [14]:
```
# Inspect the data
print(combined_data.info())
print(combined_data.describe())

# Check for missing values
missing_values = combined_data.isnull().sum()
print(missing_values)

# Summary statistics
print(combined_data.describe())

# Time series plots
combined_data.plot(subplots=True, figsize=(12, 18))
plt.tight_layout()
plt.show()

# Histograms
combined_data.hist(bins=50, figsize=(20, 15))
plt.tight_layout()
plt.show()

# Calculate correlations with the Unemployment Rate
correlation_matrix = combined_data.corr()
print("Correlation Matrix with Unemployment Rate:")
print(correlation_matrix['Unemployment Rate'])
```

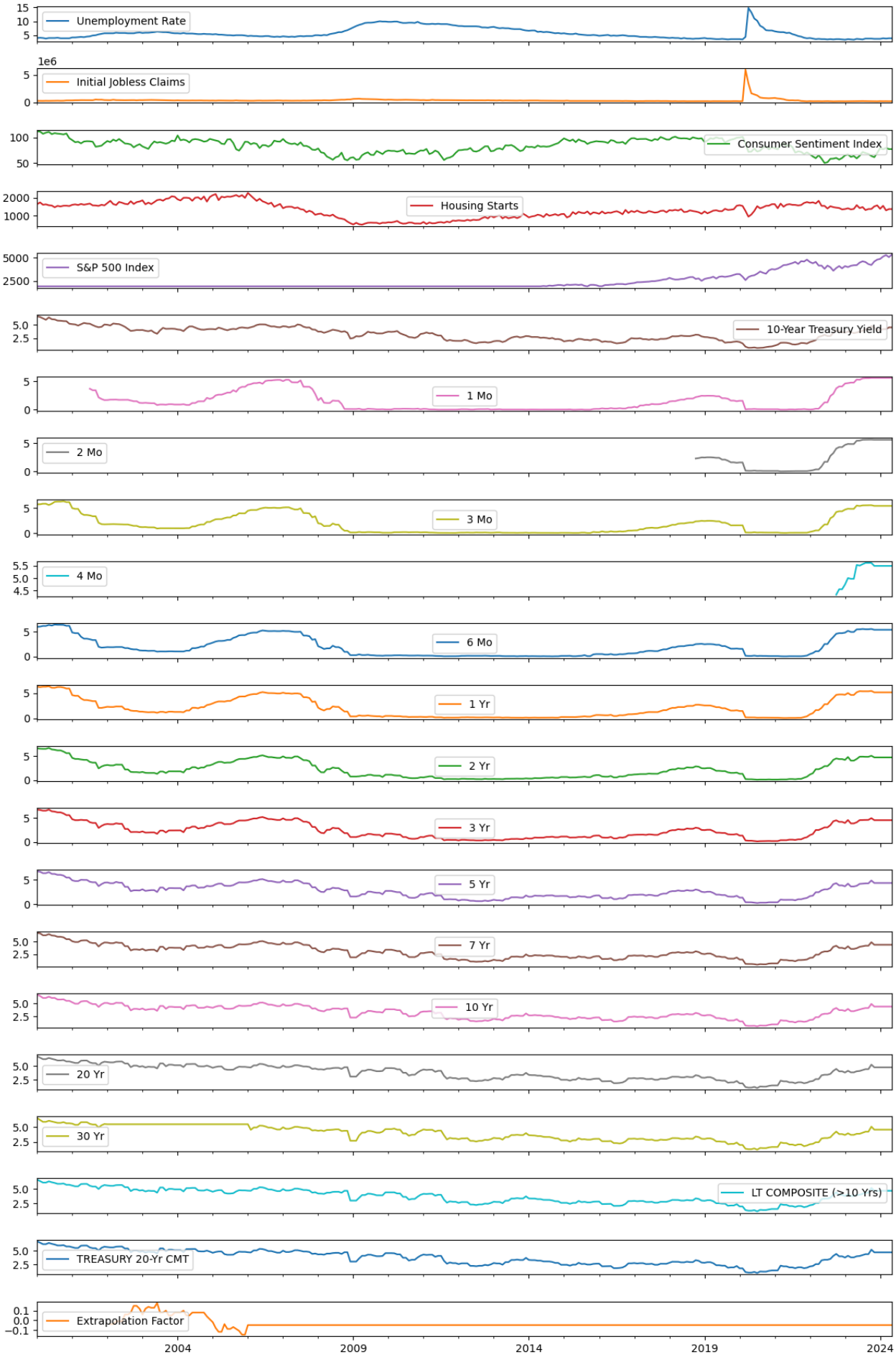
```
<class 'pandas.core.frame.DataFrame'>
DatetimeIndex: 293 entries, 2000-01-31 to 2024-05-31
Freq: M
Data columns (total 22 columns):
 #   Column                    Non-Null Count  Dtype  
---  ------                    --------------  -----  
 0   Unemployment Rate         293 non-null    float64
 1   Initial Jobless Claims    293 non-null    float64
 2   Consumer Sentiment Index  293 non-null    float64
 3   Housing Starts            293 non-null    float64
 4   S&P 500 Index             293 non-null    float64
 5   10-Year Treasury Yield    293 non-null    float64
 6   1 Mo                      275 non-null    float64
 7   2 Mo                      68 non-null     float64
 8   3 Mo                      293 non-null    float64
 9   4 Mo                      20 non-null     float64
 10  6 Mo                      293 non-null    float64
 11  1 Yr                      293 non-null    float64
 12  2 Yr                      293 non-null    float64
 13  3 Yr                      293 non-null    float64
 14  5 Yr                      293 non-null    float64
 15  7 Yr                      293 non-null    float64
 16  10 Yr                     293 non-null    float64
 17  20 Yr                     293 non-null    float64
 18  30 Yr                     293 non-null    float64
 19  LT COMPOSITE (>10 Yrs)    293 non-null    float64
 20  TREASURY 20-Yr CMT        293 non-null    float64
 21  Extrapolation Factor      268 non-null    float64
dtypes: float64(22)
memory usage: 60.8 KB
None
       Unemployment Rate  Initial Jobless Claims  Consumer Sentiment Index  \
count         293.000000            2.930000e+02                293.000000   
mean            5.730717            3.862321e+05                 83.205461   
std             1.970626            4.065895e+05                 13.216608   
min             3.400000            1.870000e+05                 50.000000   
25%             4.200000            2.610000e+05                 73.500000   
50%             5.100000            3.320000e+05                 84.900000   
75%             6.600000            3.990000e+05                 93.500000   
max            14.800000            5.946000e+06                112.000000   

       Housing Starts  S&P 500 Index  10-Year Treasury Yield        1 Mo  \
count      293.000000     293.000000              293.000000  275.000000   
mean      1304.610922    2425.492833                3.263311    1.477273   
std        437.032543     863.068638                1.310693    1.748011   
min        478.000000    1920.030000                0.620000    0.000000   
25%       1002.000000    1923.570000                2.190000    0.070000   
50%       1311.000000    1923.570000                3.200000    0.900000   
75%       1625.000000    2673.610000                4.260000    2.175000   
max       2273.000000    5304.720000                6.660000    5.560000   

            2 Mo        3 Mo       4 Mo  ...        2 Yr        3 Yr  \
count  68.000000  293.000000  20.000000  ...  293.000000  293.000000   
mean    2.227353    1.777952   5.248000  ...    2.145358    2.320717   
std     2.123580    1.954917   0.418941  ...    1.767449    1.658739   
min     0.010000    0.000000   4.330000  ...    0.110000    0.110000   
25%     0.097500    0.100000   4.970000  ...    0.580000    0.920000   
50%     1.775000    1.070000   5.490000  ...    1.620000    1.900000   
75%     4.347500    2.990000   5.507500  ...    3.600000    3.730000   
max     5.570000    6.380000   5.610000  ...    6.690000    6.660000   

             5 Yr        7 Yr       10 Yr       20 Yr       30 Yr  \
count  293.000000  293.000000  293.000000  293.000000  293.000000   
mean     2.688396    2.993242    3.247065    3.761331    3.902082   
std      1.498462    1.401972    1.314798    1.315270    1.254081   
min      0.210000    0.390000    0.550000    0.980000    1.200000   
25%      1.510000    1.920000    2.170000    2.660000    2.920000   
```

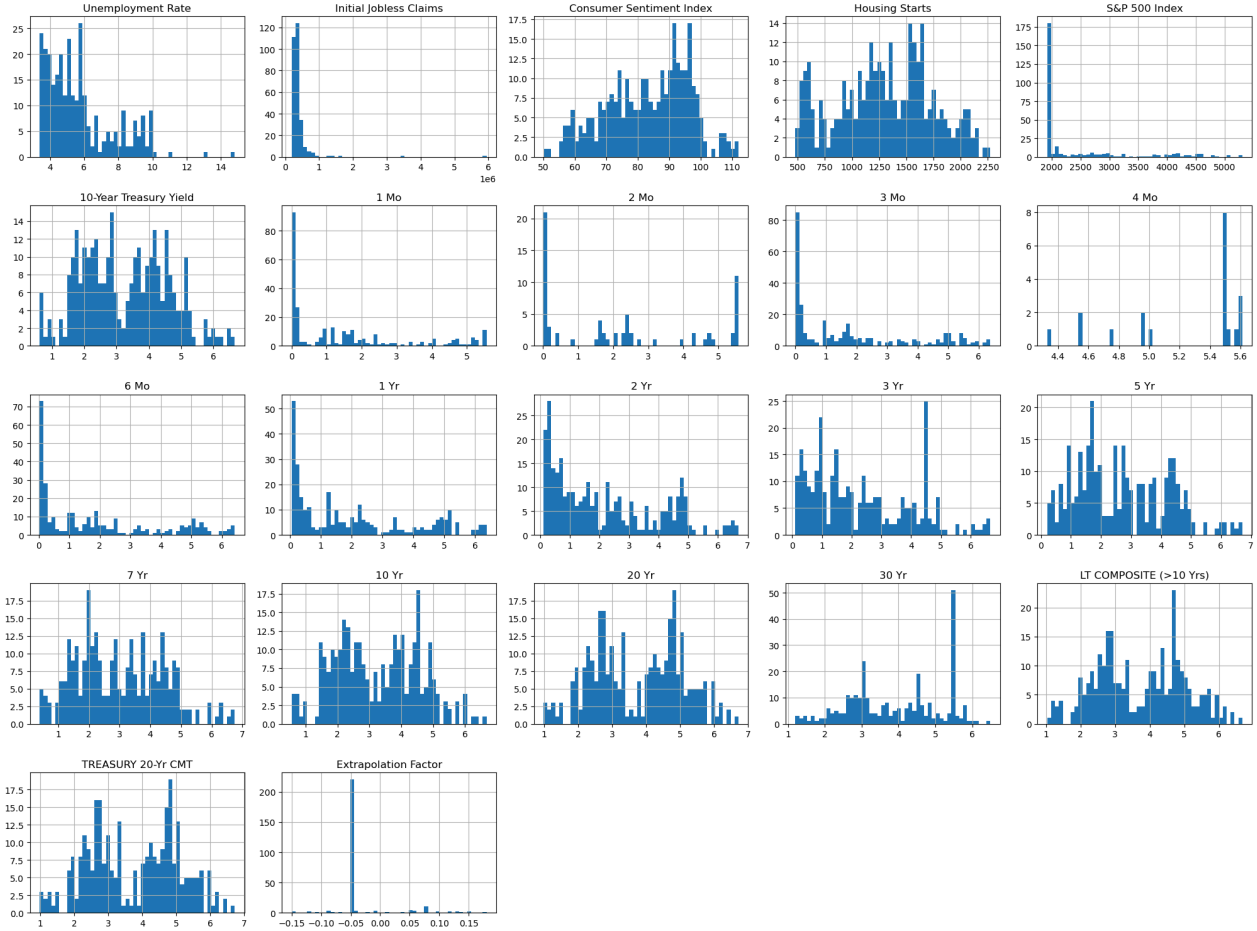
```
50%      2.520000    2.900000    3.180000    3.910000    3.850000
75%      3.870000    4.090000    4.360000    4.810000    5.190000
max      6.710000    6.750000    6.680000    6.720000    6.490000

       LT COMPOSITE (>10 Yrs)  TREASURY 20-Yr CMT  Extrapolation Factor
count              293.000000          293.000000            268.000000
mean                 3.729966            3.761331             -0.036045
std                  1.265895            1.315270              0.047555
min                  1.040000            0.980000             -0.150000
25%                  2.710000            2.660000             -0.050000
50%                  3.820000            3.910000             -0.050000
75%                  4.730000            4.810000             -0.050000
max                  6.670000            6.720000              0.180000

[8 rows x 22 columns]
Unemployment Rate             0
Initial Jobless Claims        0
Consumer Sentiment Index      0
Housing Starts                0
S&P 500 Index                 0
10-Year Treasury Yield        0
1 Mo                         18
2 Mo                        225
3 Mo                          0
4 Mo                        273
6 Mo                          0
1 Yr                          0
2 Yr                          0
3 Yr                          0
5 Yr                          0
7 Yr                          0
10 Yr                         0
20 Yr                         0
30 Yr                         0
LT COMPOSITE (>10 Yrs)        0
TREASURY 20-Yr CMT            0
Extrapolation Factor         25
dtype: int64
       Unemployment Rate  Initial Jobless Claims  Consumer Sentiment Index  \
count         293.000000            2.930000e+02                293.000000
mean            5.730717            3.862321e+05                 83.205461
std             1.970626            4.065895e+05                 13.216608
min             3.400000            1.870000e+05                 50.000000
25%             4.200000            2.610000e+05                 73.500000
50%             5.100000            3.320000e+05                 84.900000
75%             6.600000            3.990000e+05                 93.500000
max            14.800000            5.946000e+06                112.000000

       Housing Starts  S&P 500 Index  10-Year Treasury Yield        1 Mo  \
count      293.000000     293.000000               293.000000  275.000000
mean      1304.610922    2425.492833                 3.263311    1.477273
std        437.032543     863.068638                 1.310693    1.748011
min        478.000000    1920.030000                 0.620000    0.000000
25%       1002.000000    1923.570000                 2.190000    0.070000
50%       1311.000000    1923.570000                 3.200000    0.900000
75%       1625.000000    2673.610000                 4.260000    2.175000
max       2273.000000    5304.720000                 6.660000    5.560000

            2 Mo        3 Mo       4 Mo  ...        2 Yr        3 Yr  \
count  68.000000  293.000000  20.000000  ...  293.000000  293.000000
mean    2.227353    1.777952   5.248000  ...    2.145358    2.320717
std     2.123580    1.954917   0.418941  ...    1.767449    1.658739
min     0.010000    0.000000   4.330000  ...    0.110000    0.110000
25%     0.097500    0.100000   4.970000  ...    0.580000    0.920000
50%     1.775000    1.070000   5.490000  ...    1.620000    1.900000
75%     4.347500    2.990000   5.507500  ...    3.600000    3.730000
max     5.570000    6.380000   5.610000  ...    6.690000    6.660000
```

	5 Yr	7 Yr	10 Yr	20 Yr	30 Yr
count	293.000000	293.000000	293.000000	293.000000	293.000000
mean	2.688396	2.993242	3.247065	3.761331	3.902082
std	1.498462	1.401972	1.314798	1.315270	1.254081
min	0.210000	0.390000	0.550000	0.980000	1.200000
25%	1.510000	1.920000	2.170000	2.660000	2.920000
50%	2.520000	2.900000	3.180000	3.910000	3.850000
75%	3.870000	4.090000	4.360000	4.810000	5.190000
max	6.710000	6.750000	6.680000	6.720000	6.490000

	LT COMPOSITE (>10 Yrs)	TREASURY 20-Yr CMT	Extrapolation Factor
count	293.000000	293.000000	268.000000
mean	3.729966	3.761331	-0.036045
std	1.265895	1.315270	0.047555
min	1.040000	0.980000	-0.150000
25%	2.710000	2.660000	-0.050000
50%	3.820000	3.910000	-0.050000
75%	4.730000	4.810000	-0.050000
max	6.670000	6.720000	0.180000

[8 rows x 22 columns]

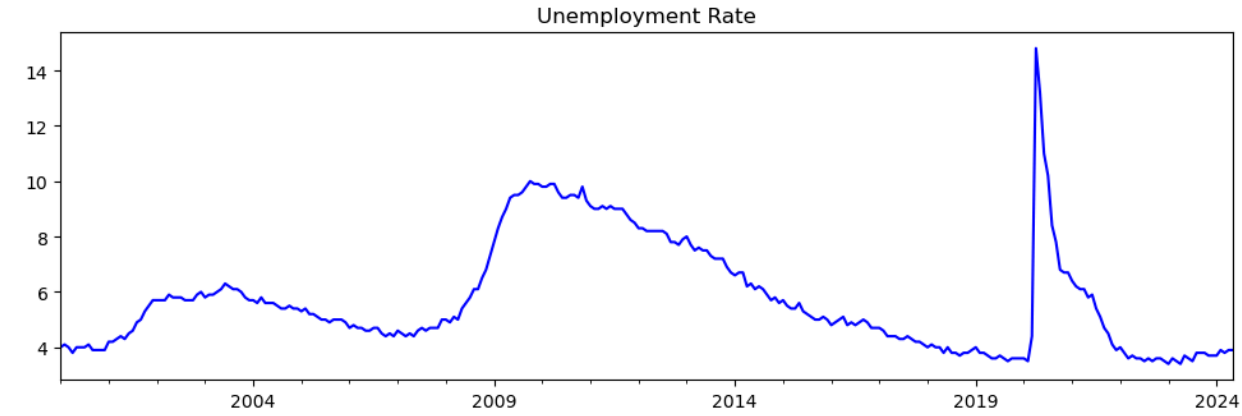

Unemployment Rate
1e6
Initial Jobless Claims
Consumer Sentiment Index
Housing Starts
S&P 500 Index
10-Year Treasury Yield
1 Mo
2 Mo
3 Mo
4 Mo
6 Mo
1 Yr
2 Yr
3 Yr
5 Yr
7 Yr
10 Yr
20 Yr
30 Yr
LT COMPOSITE (>10 Yrs)
TREASURY 20-Yr CMT
Extrapolation Factor
2004
2009
2014
2019
2024

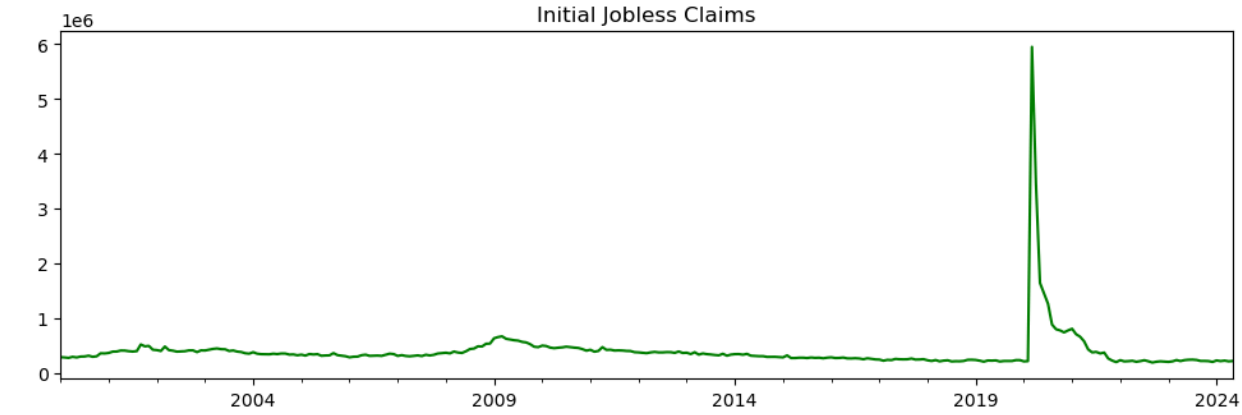


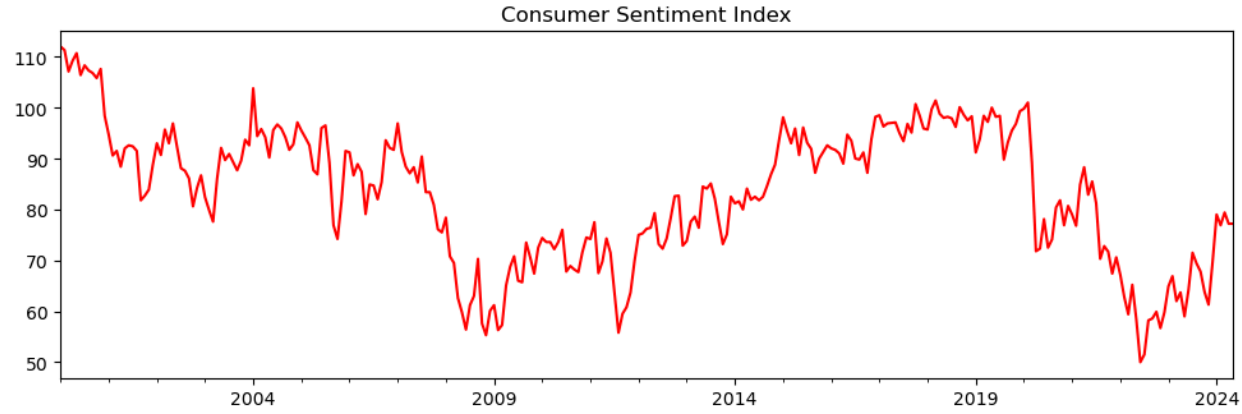

```
Correlation Matrix with Unemployment Rate:
Unemployment Rate           1.000000
Initial Jobless Claims      0.350648
Consumer Sentiment Index   -0.407165
Housing Starts             -0.597250
S&P 500 Index              -0.356714
10-Year Treasury Yield     -0.299425
1 Mo                       -0.564185
2 Mo                       -0.486816
3 Mo                       -0.596492
4 Mo                        0.637163
6 Mo                       -0.607067
1 Yr                       -0.614367
2 Yr                       -0.599557
3 Yr                       -0.574485
5 Yr                       -0.490350
7 Yr                       -0.405321
10 Yr                      -0.304152
20 Yr                      -0.178577
30 Yr                      -0.129028
LT COMPOSITE (>10 Yrs)     -0.207523
TREASURY 20-Yr CMT         -0.178577
Extrapolation Factor        0.009905
Name: Unemployment Rate, dtype: float64
```

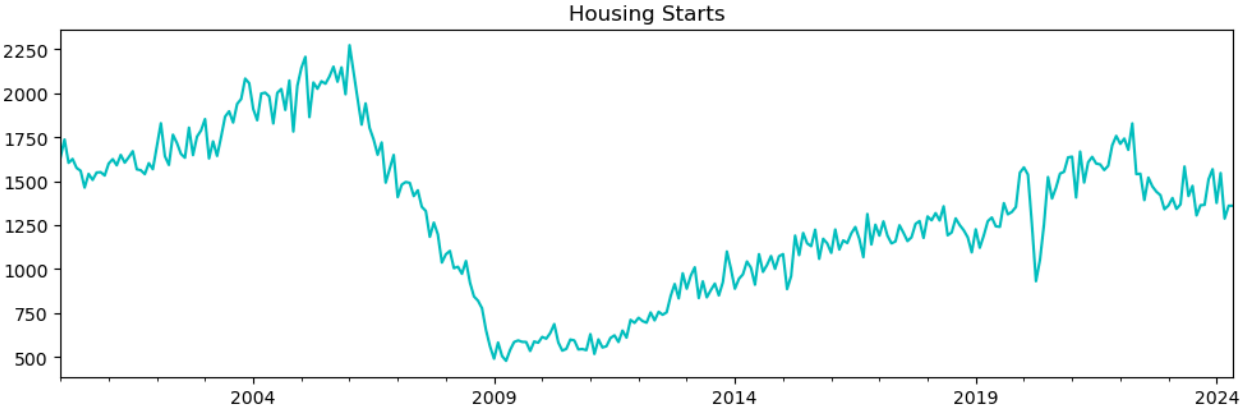
In [15]:
```python
import matplotlib.pyplot as plt

# Set figure size for Microsoft document
fig, axs = plt.subplots(6, 1, figsize=(10, 20))

# Plot each economic indicator
combined_data['Unemployment Rate'].plot(ax=axs[0], title='Unemployment Rate', color='b')
combined_data['Initial Jobless Claims'].plot(ax=axs[1], title='Initial Jobless Claims', color='g'
combined_data['Consumer Sentiment Index'].plot(ax=axs[2], title='Consumer Sentiment Index', color
```

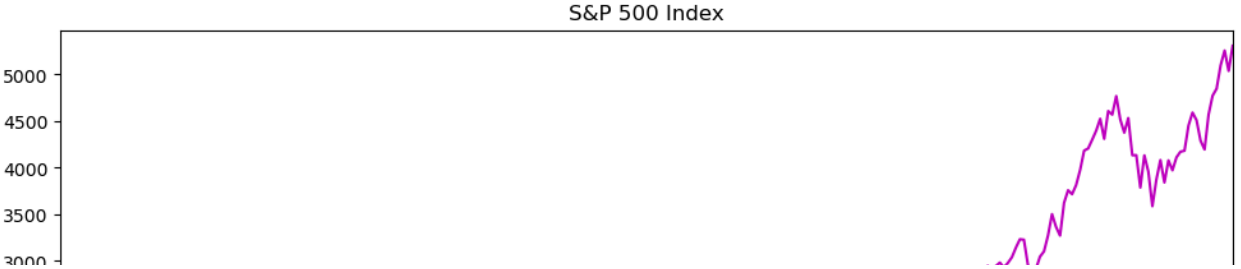
```
combined_data['Housing Starts'].plot(ax=axs[3], title='Housing Starts', color='c')
combined_data['S&P 500 Index'].plot(ax=axs[4], title='S&P 500 Index', color='m')
combined_data['10-Year Treasury Yield'].plot(ax=axs[5], title='10-Year Treasury Yield', color='y'

# Adjust layout
plt.tight_layout()
plt.savefig('time_series_plots.png')
plt.show()
```

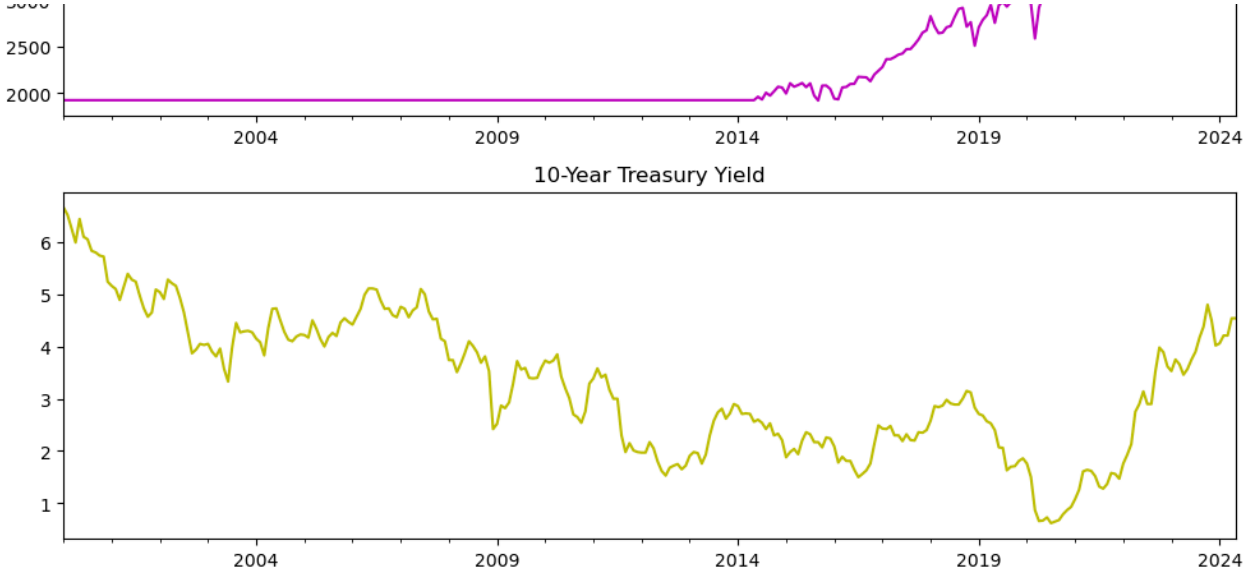

Unemployment Rate
14
12
10
8
6
4
2004
2009
2014
2019
2024


Initial Jobless Claims
1e6
6
5
4
3
2
1
0
2004
2009
2014
2019
2024

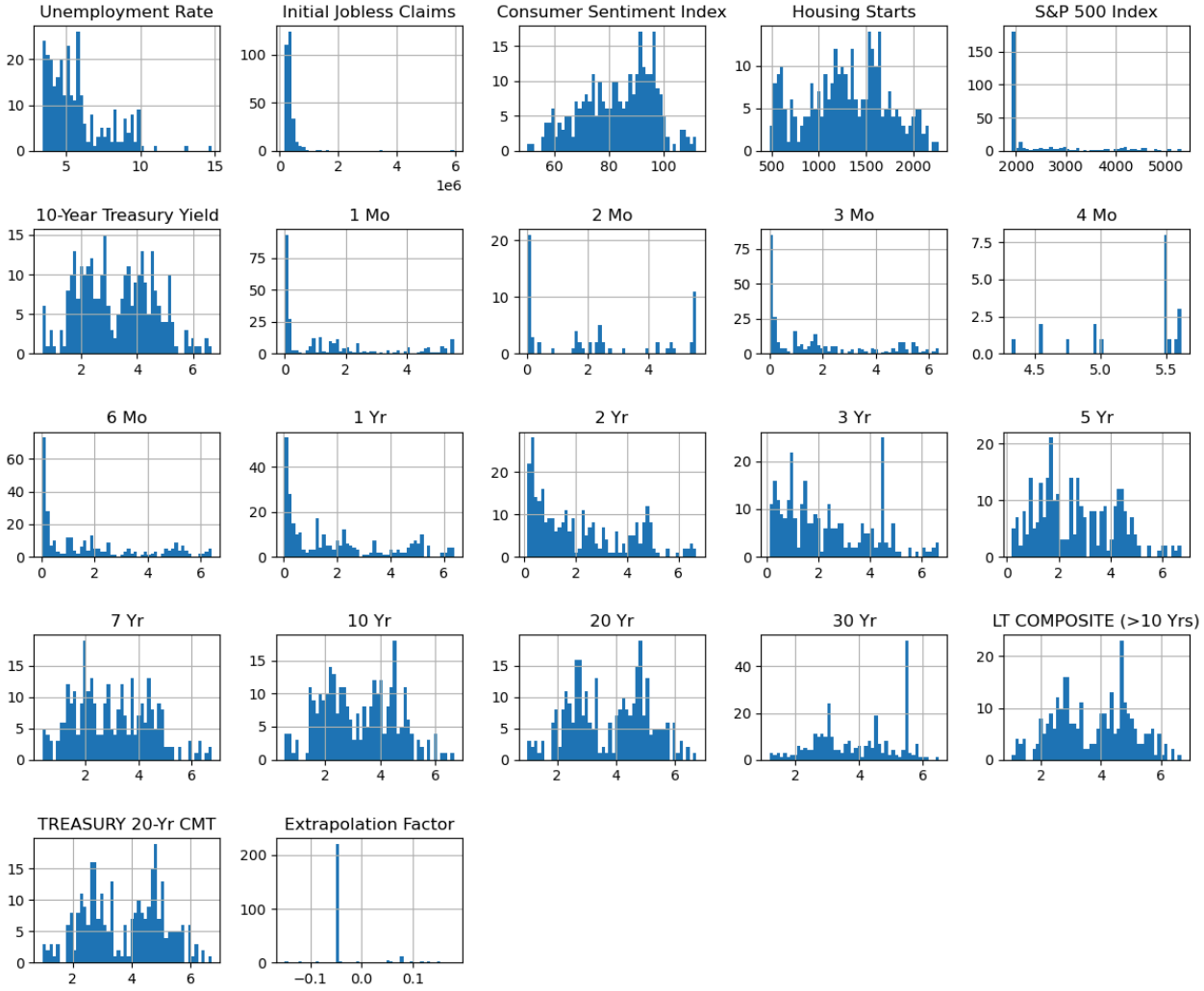

Consumer Sentiment Index
110
100
90
80
70
60
50
2004
2009
2014
2019
2024


Housing Starts
2250
2000
1750
1500
1250
1000
750
500
2004
2009
2014
2019
2024

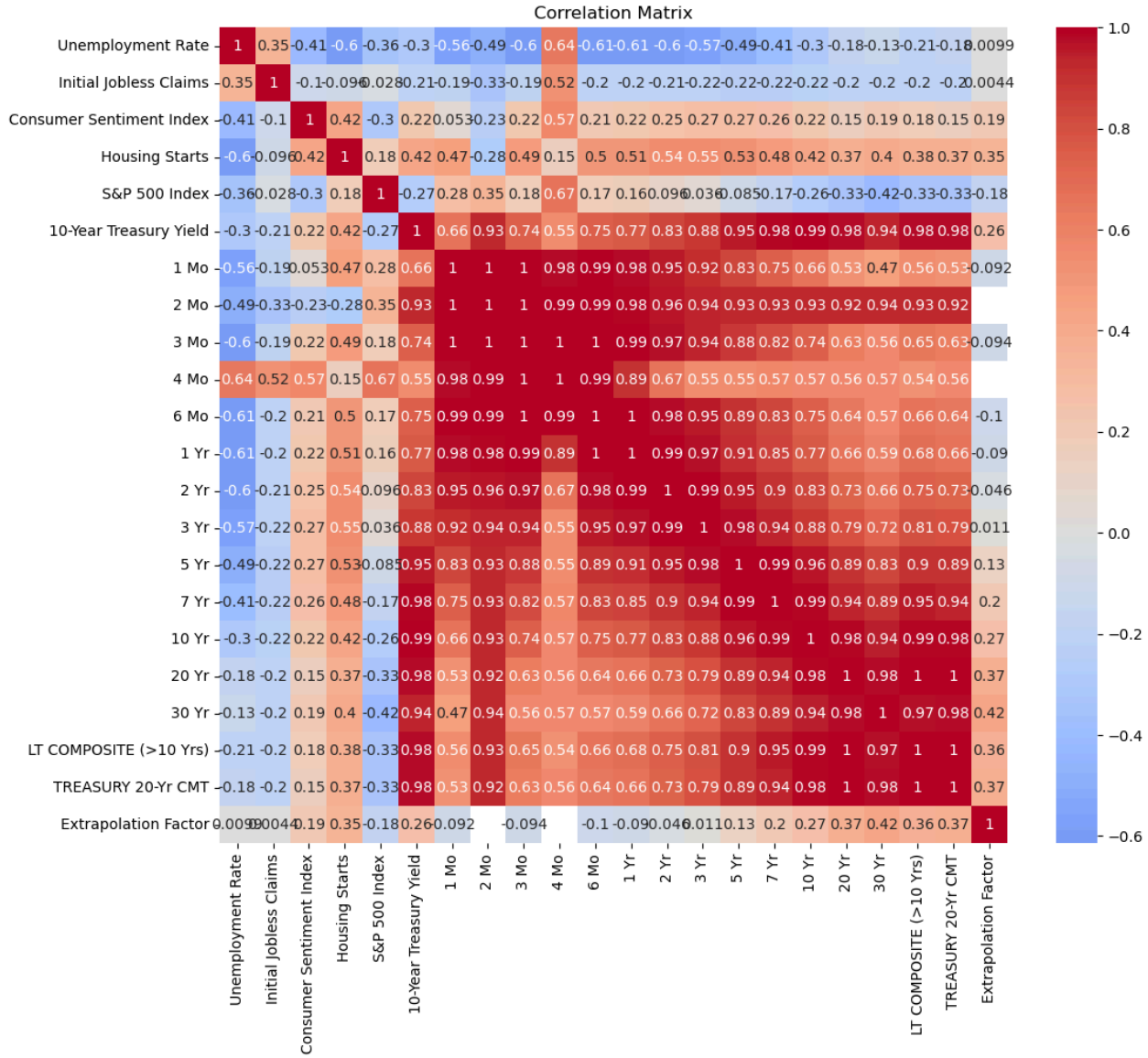

S&P 500 Index
5000
4500
4000
3500
3000



In [16]:
```
import matplotlib.pyplot as plt

# Set figure size for Microsoft document
combined_data.hist(bins=50, figsize=(12, 10))

# Adjust layout
plt.tight_layout()
plt.savefig('histograms.png')
plt.show()
```

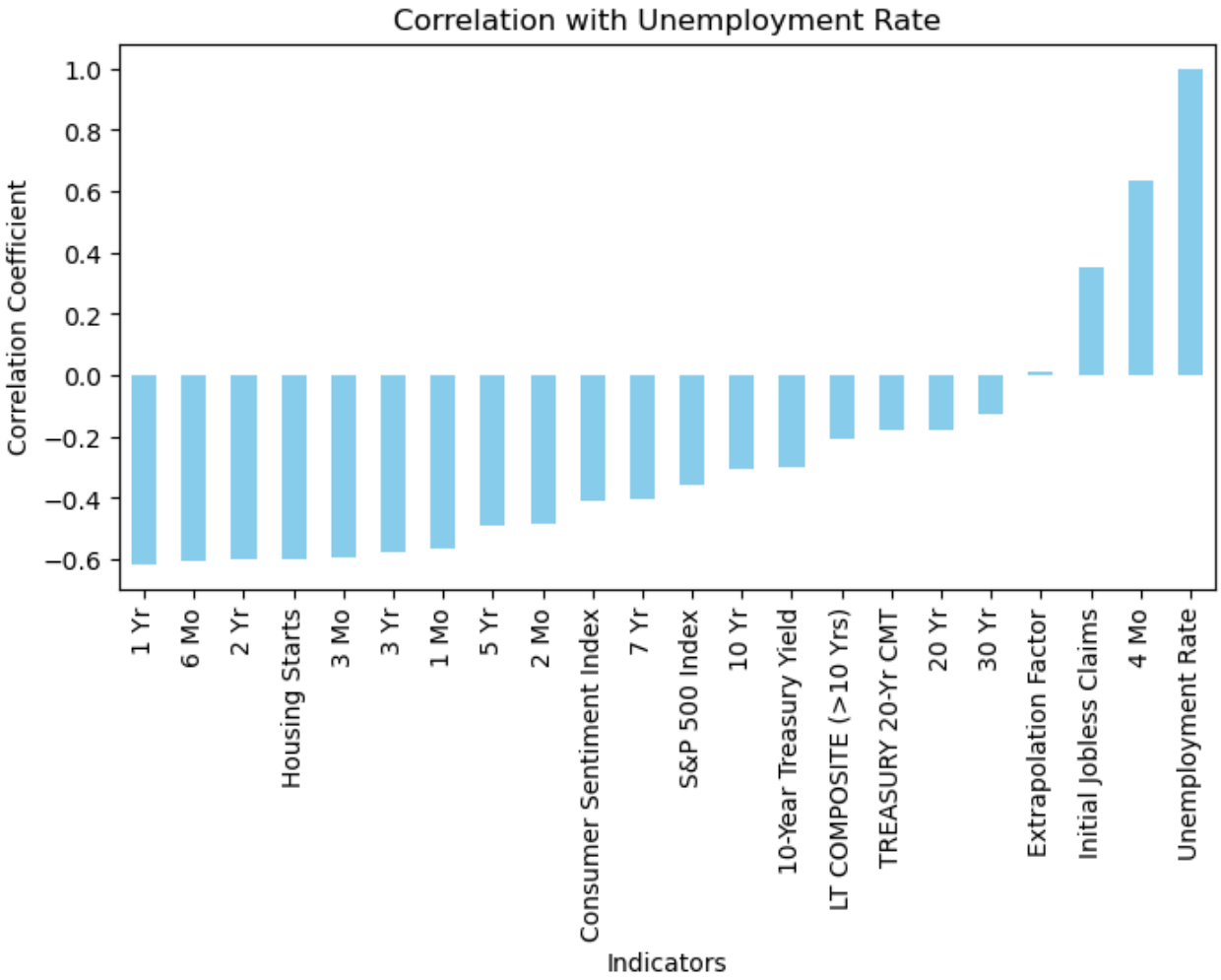


In [17]:
```
import seaborn as sns
import matplotlib.pyplot as plt

# Calculate correlation matrix
correlation_matrix = combined_data.corr()

# Set figure size for Microsoft document
plt.figure(figsize=(12, 10))

# Plot heatmap
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', center=0)

# Set title and save figure
plt.title('Correlation Matrix')
plt.savefig('correlation_heatmap.png')
plt.show()
```



In [18]:

```
import matplotlib.pyplot as plt

# Calculate correlations with the Unemployment Rate
correlation_with_unemployment = correlation_matrix['Unemployment Rate'].sort_values()

# Set figure size for Microsoft document
plt.figure(figsize=(8, 4))

# Plot bar chart
correlation_with_unemployment.plot(kind='bar', color='skyblue')

# Set title and labels
plt.title('Correlation with Unemployment Rate')
plt.xlabel('Indicators')
plt.ylabel('Correlation Coefficient')

# Save figure
plt.savefig('correlation_with_unemployment.png')
plt.show()
```



Feature Engineering

In [19]:
```
# Create lagged variables
for lag in range(1, 13):
    combined_data[f'Unemployment Rate Lag {lag}'] = combined_data['Unemployment Rate'].shift(lag)
    combined_data[f'Initial Jobless Claims Lag {lag}'] = combined_data['Initial Jobless Claims'].
    combined_data[f'Consumer Sentiment Index Lag {lag}'] = combined_data['Consumer Sentiment Inde
    combined_data[f'Housing Starts Lag {lag}'] = combined_data['Housing Starts'].shift(lag)
    combined_data[f'S&P 500 Index Lag {lag}'] = combined_data['S&P 500 Index'].shift(lag)
    combined_data[f'10-Year Treasury Yield Lag {lag}'] = combined_data['10-Year Treasury Yield'].

# Create moving averages
combined_data['Unemployment Rate MA 3'] = combined_data['Unemployment Rate'].rolling(window=3).me
combined_data['Unemployment Rate MA 6'] = combined_data['Unemployment Rate'].rolling(window=6).me
combined_data['Unemployment Rate MA 12'] = combined_data['Unemployment Rate'].rolling(window=12).

combined_data.dropna(inplace=True)
```

Model Training and Evaluation

In [20]:
```
from sklearn.model_selection import train_test_split, GridSearchCV
from sklearn.ensemble import RandomForestRegressor, GradientBoostingRegressor
from sklearn.metrics import mean_squared_error, mean_absolute_error
from statsmodels.tsa.arima.model import ARIMA

# Prepare data for modeling
X = combined_data.drop(columns=['Unemployment Rate'])
```

```
y = combined_data['Unemployment Rate']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# ARIMA model
arima_model = ARIMA(y_train, order=(5, 1, 0))
arima_model_fit = arima_model.fit()
arima_forecast = arima_model_fit.forecast(steps=len(y_test))
arima_rmse = mean_squared_error(y_test, arima_forecast, squared=False)
arima_mae = mean_absolute_error(y_test, arima_forecast)
print(f'ARIMA RMSE: {arima_rmse}, MAE: {arima_mae}')

# Random Forest model
rf_model = RandomForestRegressor()
rf_model.fit(X_train, y_train)
rf_predictions = rf_model.predict(X_test)
rf_rmse = mean_squared_error(y_test, rf_predictions, squared=False)
rf_mae = mean_absolute_error(y_test, rf_predictions)
print(f'Random Forest RMSE: {rf_rmse}, MAE: {rf_mae}')

# Gradient Boosting model
gb_model = GradientBoostingRegressor()
gb_model.fit(X_train, y_train)
gb_predictions = gb_model.predict(X_test)
gb_rmse = mean_squared_error(y_test, gb_predictions, squared=False)
gb_mae = mean_absolute_error(y_test, gb_predictions)
print(f'Gradient Boosting RMSE: {gb_rmse}, MAE: {gb_mae}')
```

```
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it has no associated frequency information and so will be ignor
ed when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it is not monotonic and so will be ignored when e.g. forecastin
g.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it has no associated frequency information and so will be ignor
ed when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it is not monotonic and so will be ignored when e.g. forecastin
g.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it has no associated frequency information and so will be ignor
ed when e.g. forecasting.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:471: ValueWarning: A
date index has been provided, but it is not monotonic and so will be ignored when e.g. forecastin
g.
  self._init_dates(dates, freq)
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:834: ValueWarning: N
o supported index is available. Prediction results will be given with an integer index beginning
at `start`.
  return get_prediction_index(
ARIMA RMSE: 0.10776656125850885, MAE: 0.07707577442511163
Random Forest RMSE: 0.06452131430775292, MAE: 0.05499999999999916
Gradient Boosting RMSE: 0.12433563564472944, MAE: 0.1192440577188234
```

Deployment and Monitoring

In [21]:

```
# Implementing a simple prediction function
def predict_unemployment_rate(model, new_data):
    prediction = model.predict(new_data)
```

```
    return prediction

# Continuous monitoring and updating models
# This part will be more extensive in a real-world scenario involving setting up pipelines and au
import joblib

# Save models
joblib.dump(arima_model_fit, 'arima_model.pkl')
joblib.dump(rf_model, 'random_forest_model.pkl')
joblib.dump(gb_model, 'gradient_boosting_model.pkl')

# Load models
arima_model_loaded = joblib.load('arima_model.pkl')
rf_model_loaded = joblib.load('random_forest_model.pkl')
gb_model_loaded = joblib.load('gradient_boosting_model.pkl')

# Example usage with new data
new_data = X_test.iloc[:1]  # Replace with new data as needed
arima_prediction = arima_model_loaded.forecast(steps=1)
rf_prediction = predict_unemployment_rate(rf_model_loaded, new_data)
gb_prediction = predict_unemployment_rate(gb_model_loaded, new_data)

print(f'ARIMA Prediction: {arima_prediction}')
print(f'Random Forest Prediction: {rf_prediction}')
print(f'Gradient Boosting Prediction: {gb_prediction}')
```

```
ARIMA Prediction: 16    3.720503
dtype: float64
Random Forest Prediction: [3.541]
Gradient Boosting Prediction: [3.52234706]
```

```
C:\Users\salin\anaconda3\lib\site-packages\statsmodels\tsa\base\tsa_model.py:834: ValueWarning: N
o supported index is available. Prediction results will be given with an integer index beginning
at `start`.
  return get_prediction_index(
```

Code to Forecast Future Economic Indicators and Predict Unemployment Rate

In [22]:

```
import pandas as pd
import numpy as np
from datetime import timedelta

# Generate future dates for the next five years
future_dates = pd.date_range(start='2024-06-01', end='2029-06-01', freq='M')

# Create a DataFrame to hold future values of the indicators
future_data = pd.DataFrame(index=future_dates, columns=combined_data.columns)

# Forecast future values for the economic indicators
# Here, we will use a simple method of extending the last known value forward
# In practice, more sophisticated methods like ARIMA could be used for each indicator

for column in combined_data.columns:
    last_value = combined_data[column].iloc[-1]
    future_data[column] = last_value

# Apply some variation to simulate realistic changes over time (this is a simplified approach)
np.random.seed(42)  # For reproducibility
for column in future_data.columns:
    future_data[column] = future_data[column] * (1 + 0.01 * np.random.randn(len(future_data)))

# Ensure the future data aligns with the features used in the model
feature_columns = X_train.columns
future_data = future_data[feature_columns]
```

```
# Use the trained Random Forest model to make predictions
future_predictions = rf_model.predict(future_data)

# Combine the future dates with the predicted unemployment rates
future_forecast = pd.DataFrame({'Date': future_dates, 'Predicted Unemployment Rate': future_predi

# Set the date as the index
future_forecast.set_index('Date', inplace=True)

# Print the predictions
print(future_forecast)
```

Date	Predicted Unemployment Rate
2024-06-30	3.817
2024-07-31	3.806
2024-08-31	3.821
2024-09-30	3.835
2024-10-31	3.797
2024-11-30	3.813
2024-12-31	3.815
2025-01-31	3.823
2025-02-28	3.820
2025-03-31	3.825
2025-04-30	3.824
2025-05-31	3.830
2025-06-30	3.823
2025-07-31	3.820
2025-08-31	3.827
2025-09-30	3.830
2025-10-31	3.825
2025-11-30	3.805
2025-12-31	3.823
2026-01-31	3.818
2026-02-28	3.838
2026-03-31	3.818
2026-04-30	3.800
2026-05-31	3.828
2026-06-30	3.824
2026-07-31	3.813
2026-08-31	3.811
2026-09-30	3.827
2026-10-31	3.812
2026-11-30	3.804
2026-12-31	3.829
2027-01-31	3.822
2027-02-28	3.802
2027-03-31	3.821
2027-04-30	3.840
2027-05-31	3.822
2027-06-30	3.818
2027-07-31	3.818
2027-08-31	3.824
2027-09-30	3.850
2027-10-31	3.806
2027-11-30	3.825
2027-12-31	3.816
2028-01-31	3.831
2028-02-29	3.789
2028-03-31	3.814
2028-04-30	3.807
2028-05-31	3.832
2028-06-30	3.848
2028-07-31	3.813
2028-08-31	3.834
2028-09-30	3.813
2028-10-31	3.827
2028-11-30	3.810
2028-12-31	3.820
2029-01-31	3.839
2029-02-28	3.811
2029-03-31	3.831
2029-04-30	3.848
2029-05-31	3.814