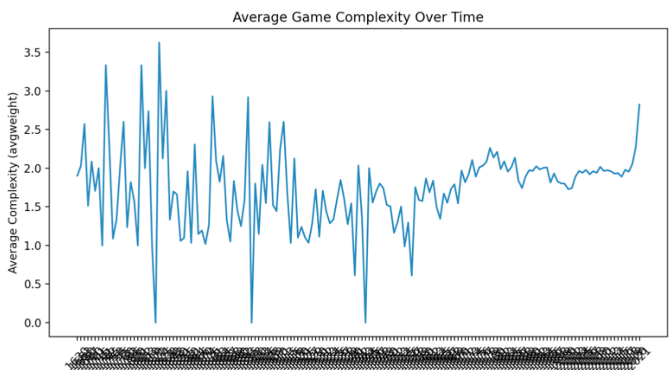
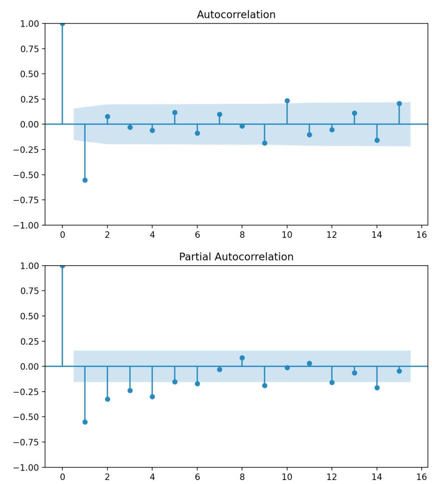
Time Series Results:

- Motivation: Analyze the trend of game complexity (avgweight) over the years (yearpublished).
- Model used: ARIMA

ADF Statistic: -0.9981974083526541
p-value: 0.7539211871495028

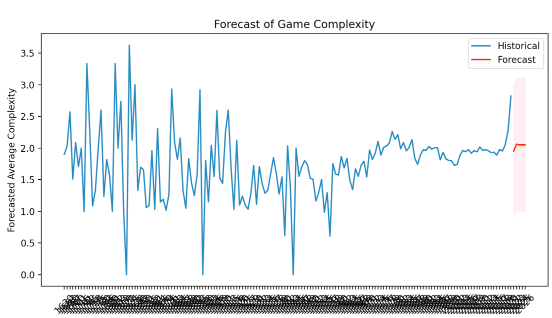
Here, A high p-value suggests that the time series of game complexity over time is likely non-stationary. This means that the statistical properties of the series, like the mean and variance, change over time. Moreover, the ADF statistic of around -1 is not significantly negative, further indicating that the time series is likely non-stationary.





The plot displays significant fluctuations in complexity over time, with notable periods of both high and low complexity. Toward the end, there's a noticeable upward trend in complexity, suggesting that recent board games might be becoming more complex.

The ACF and PACF plots suggest that an ARIMA(1,1,1) model may be appropriate for the time series, with the significant spike at lag 1 in both plots indicating the need for an AR(1) and MA(1) component. The ACF's quick drop after lag 1 and the PACF's sharp cutoff at lag 1 further support this model choice.



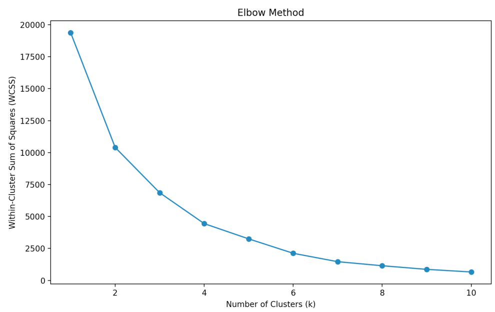

```
                               SARIMAX Results
==============================================================================
Dep. Variable:              avgweight   No. Observations:                  159
Model:                 ARIMA(1, 1, 1)   Log Likelihood                -124.872
Date:                Mon, 26 Aug 2024   AIC                            255.743
Time:                        18:08:10   BIC                            264.931
Sample:                    12-31-1630   HQIC                           259.475
                         - 12-31-2021
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1         -0.1257      0.061     -2.056      0.040      -0.246      -0.006
ma.L1         -0.9038      0.045    -20.046      0.000      -0.992      -0.815
sigma2         0.2810      0.024     11.701      0.000       0.234       0.328
===================================================================================
Ljung-Box (L1) (Q):                   0.01   Jarque-Bera (JB):                29.13
Prob(Q):                              0.92   Prob(JB):                         0.00
Heteroskedasticity (H):               0.09   Skew:                            -0.19
Prob(H) (two-sided):                  0.00   Kurtosis:                         5.07
===================================================================================
```

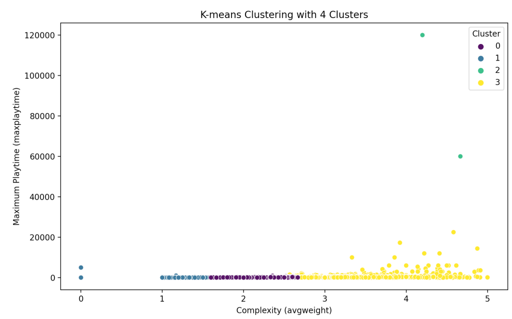
The red line represents the forecasted values of game complexity for the next few periods. The forecast shows a slight increase, suggesting that game complexity might continue to rise, although the increase is relatively modest.

The ARIMA(1,1,1) model for game complexity shows that both the AR(1) and MA(1) components are statistically significant, with the model fitting the data reasonably well as indicated by the AIC and Ljung-Box test. However, the residuals exhibit non-normality and heteroskedasticity, which suggests that while the model is useful, some caution is needed in interpreting the results, and further model refinement might be necessary.

Game Clustering:

- Motivation: Group games based on complexity and playtime
- Model: Kmeans Clustering with elbow method



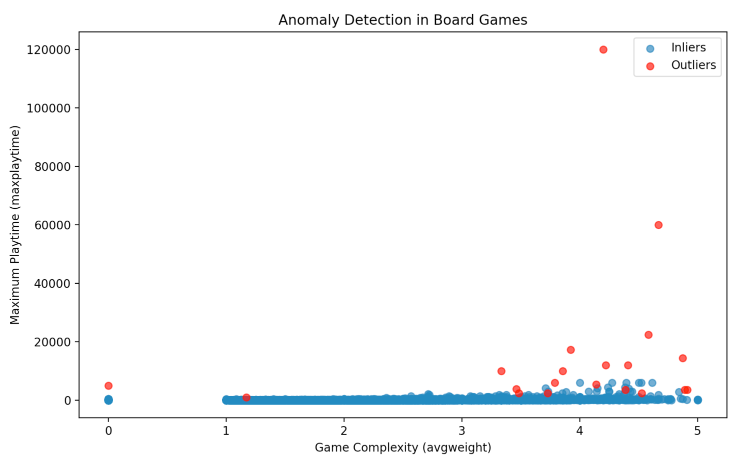


The "elbow" is most likely around **k = 4**. Therefore, **4 clusters** might be the optimal choice for the data. Most games fall into clusters with low complexity and moderate playtime, while a few outliers exhibit extremely high playtime. This clustering reveals distinct patterns, highlighting differences in game characteristics that may appeal to various player preferences.

WCSS (Within-Cluster Sum of Squares) for k=4: 4443.46675235357

Anomaly Detection:

- Motivation: Identify outlier games in terms of complexity or playtime
- Model: DBSCAN



Legacy of Dragonholt
Empires in Arms
Axis Empires Totaler Krieg!
Advanced Third Reich
Case Blue
The Longest Day
Where Eagles Dare The Battle for Hell's Highw
Europa Universalis
1914 Offensive \u00e0 outrance
La Grande Guerre 14-18
The 7th Continent Classic Edition
The Habit of Victory
La Bataille d'Espagnol Talavera
1985 Under an Iron Sky
Atlantic Wall D-Day to Falaise
Age of Muskets Volume I Tomb for an Empire
Drang Nach Osten!
Gameception
The Enigma Box
The Campaign for North Africa

The outliers are mostly games with either extremely high playtimes or unusual complexity, suggesting that these games may cater to niche audiences or have unique characteristics. And those on the columns were the outliers.

Popularity Classification:

- Motivation: Classify games into popularity categories (high, medium, low)
- Model: Naïve Bayes

```
Accuracy: 0.64825
              precision    recall  f1-score   support

        high       0.94      0.61      0.74      1327
         low       0.62      0.90      0.73      1329
      medium       0.49      0.44      0.47      1344

    accuracy                           0.65      4000
   macro avg       0.68      0.65      0.65      4000
weighted avg       0.68      0.65      0.64      4000

Data with predicted popularity categories has been saved to 'classified_boardgames.csv'.
```

Overall accuracy of 64.8%. The model performs best in predicting the "high" popularity category, with a precision of 0.94 and an F1-score of 0.74. The "low" popularity category also has good recall (0.90) but a lower precision (0.62), suggesting that while it identifies most low-popularity games, it also includes many false positives. The "medium" category is the most challenging to predict, with a lower precision (0.49) and F1-score (0.47), reflecting difficulty in distinguishing these games from others.

Market Analysis:

- Motivation: Analyze common combinations of designers and mechanics in successful games
- Model: Market basket analysis

```
Test Support: 0.05475
                          antecedents        consequents   support  confidence       lift
0                    ('Hexagon Grid')    ('Simulation'])  0.017750    0.613391  14.915281
1                    ('Hexagon Grid')  (['Dice Rolling')  0.018125    0.626350   5.884673
2                   ('Hexagon Grid'])  (['Dice Rolling')  0.011187    0.542424   5.096176
3  ('Hexagon Grid', ['Dice Rolling')    ('Simulation'])  0.013625    0.751724  18.279006
4    ('Hexagon Grid', 'Simulation'])  (['Dice Rolling')  0.013625    0.767606   7.211797
5   (['Dice Rolling', 'Simulation'])   ('Hexagon Grid')  0.013625    0.778571  26.905276
```

The combination of a "Hexagon Grid" with "Simulation" mechanics shows a high lift of 14.92, indicating that these mechanics often appear together in successful games. Additionally, "Hexagon Grid" combined with "Dice Rolling" has a lift of 5.88, showing another frequent pairing. The combination of all three mechanics—"Hexagon Grid," "Dice Rolling," and "Simulation"—has the highest lift values, particularly when "Hexagon Grid" and "Simulation" are paired with "Dice Rolling" (lift of 18.28) or when "Dice Rolling" and

"Simulation" are paired with "Hexagon Grid" (lift of 26.91). These results suggest that games incorporating these mechanics together tend to be more successful.

Game Complexity Prediction

- Motivation: Predict game complexity (avgweight) based on various features
- Model: Random Forest

  Mean Squared Error: 0.432076
  R^2 Score: 0.4678

The random forest model for predicting game complexity (avgweight) has a Mean Squared Error (MSE) of 0.4321, indicating some prediction error, and an R^2 score of 0.4671, meaning the model explains about 46.7% of the variance in game complexity