Kalyan Pothineni
07/14/2024
DSC680-T301 (2247-1)
Applied Data Science

Analysis of Powerball Winning Numbers

Kalyan Pothineni

Applied Data Science

Course Number: DSC680

07/14/2024

Analysis of Powerball winning numbers

This milestone aims to evaluate winning Powerball numbers to notice specific patterns, frequency, and popularity of the numbers and the specificity of the given period. The work's scope is, first, identifying the most popular numbers; second, comparing the data by the year drawn; and third, checking the data quality and its completeness. The conclusions obtained will give a better view of how lottery outcomes occur and are distributed and provide a statistical take on approaching picking numbers.

Business Problem

Knowledge of patterns may be informative for physical and mathematical researchers who are analyzing the lotteries' outcomes and for people who play and like Powerball. Although all the lotteries are games of chance, they may present significant trends by analyzing the data from previous draws. These insights may be applied in better ways of choosing numbers, instilling a better perception of the lottery's randomness, and further arguing on the randomness of the lottery system statistically.

Background/History

Powerball is considered one of the best and globally known lottery games used in the United States under the Multi-State Lottery Association's (MUSL) management. It was initially launched in 1992, after the earlier game, Lotto America. Through the years, Powerball has become more popular because it has massive jackpots and appeals to many players; whenever the jackpot has gigantic amounts, players all over the country get highly enthusiastic.

The game's conception is simple – the player must choose five numbers among sixty-nine white balls and one number of twenty-six red Powerballs. The draws take place on Wednesdays and Saturdays, making it convenient for most players. Over the years, changes in the rules and format of Powerball have enhanced the size of the jackpots and the likelihood of winning secondary prizes.

Of all the patterns relating to Powerball winning numbers, the analysis has always captured the interest of ordinary betters and statisticians. All people trust some chances to know patterns or trends that

might give them superiority in choosing numbers, despite the nature of the game. This work intends to investigate the data of Powerball winning numbers since 2010 to establish such patterns and trends.

Data Explanation

Data Preparation:

The dataset contains information on Powerball winning numbers from 2010 onwards. The data includes the following columns:

- **Draw Date:** The date when the Powerball numbers were drawn.
- **Winning Numbers:** The winning numbers for the draw, consisting of five white balls and one red Powerball.
- **Multiplier:** The Power Play multiplier for the draw, which is optional and not always present.

Data Dictionary:

```
# Display the first few rows of the dataset and its structure
data_head = data.head()
data_info = data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1650 entries, 0 to 1649
Data columns (total 3 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   Draw Date       1650 non-null   object 
 1   Winning Numbers  1650 non-null   object 
 2   Multiplier      1440 non-null   float64
dtypes: float64(1), object(2)
memory usage: 38.8+ KB
```

Data Cleaning and Transformation

- **Date Conversion:** Convert the Draw Date column from object type to datetime for better manipulation and analysis.
- **Splitting Winning Numbers:** Split the Winning Numbers column into six separate columns: five for the white balls and one for the red Powerball.
- **Handling Missing Values:** Address missing values if any in the Multiplier column.

- **Type Conversion:** Ensure all numerical values are of the correct data type for analysis.

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1650 entries, 0 to 1649
Data columns (total 10 columns):
 #   Column         Non-Null Count  Dtype
---  ------         --------------  -----
 0   Draw Date      1650 non-null   datetime64[ns]
 1   Multiplier     1650 non-null   float64
 2   White Ball 1   1650 non-null   int32
 3   White Ball 2   1650 non-null   int32
 4   White Ball 3   1650 non-null   int32
 5   White Ball 4   1650 non-null   int32
 6   White Ball 5   1650 non-null   int32
 7   Red Powerball  1650 non-null   int32
 8   Year           1650 non-null   int32
 9   Cluster        1650 non-null   int32
dtypes: datetime64[ns](1), float64(1), int32(8)
memory usage: 77.5 KB
```

	Draw Date	Multiplier	White Ball 1	White Ball 2	White Ball 3	White Ball 4	White Ball 5	Red Powerball	Year	Cluster
0	2020-09-26	3.0	11	21	27	36	62	24	2020	1
1	2020-09-30	2.0	14	18	36	49	67	18	2020	0
2	2020-10-03	2.0	18	31	36	43	47	20	2020	2
3	2020-10-07	2.0	6	24	30	53	56	19	2020	2
4	2020-10-10	3.0	5	18	23	40	50	18	2020	1

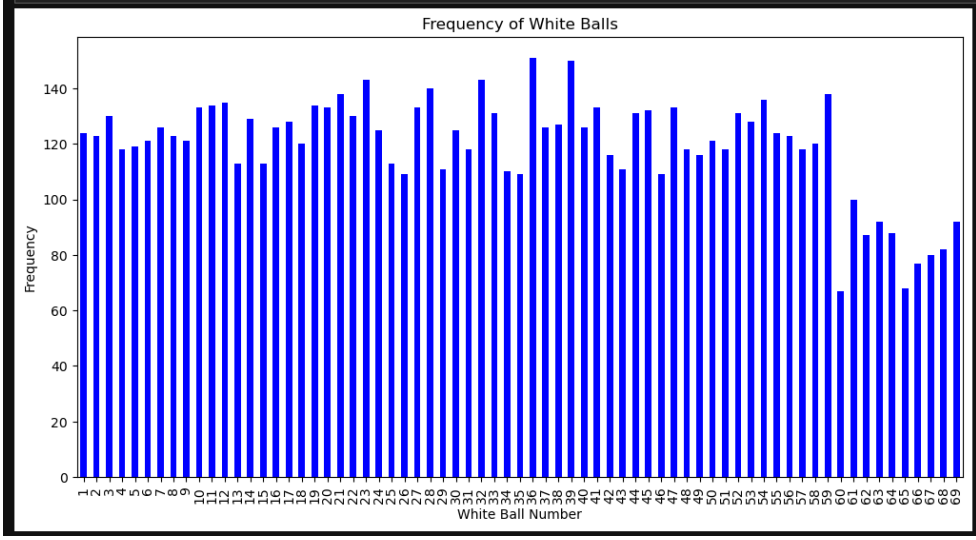
Methods

Frequency Analysis:

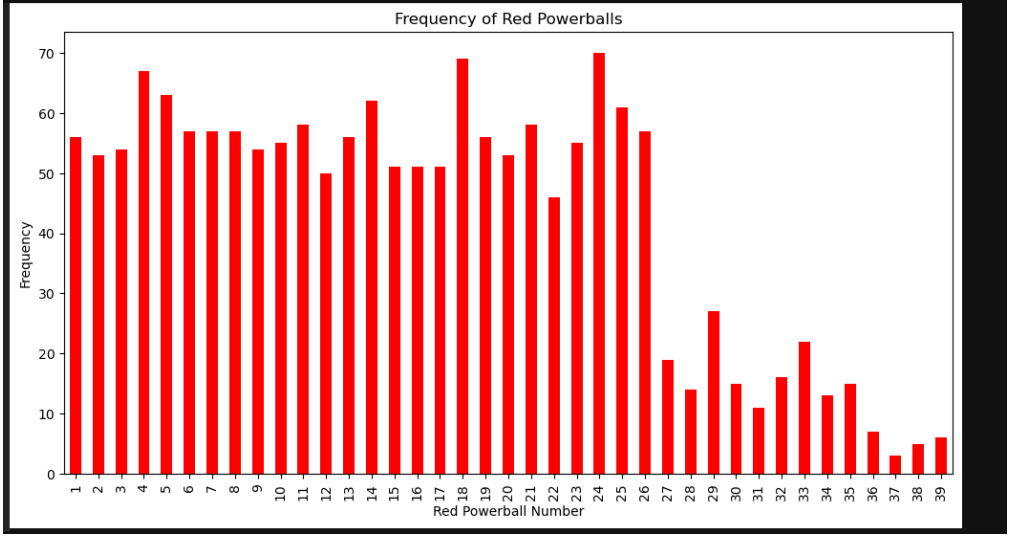
Identify and visualize the most common individual numbers, pairs, triples, quadruples, and quintuples.

```
# Visualize the frequency of each White Ball number
white_balls = data[ball_columns[:-1]].values.flatten()
white_ball_freq = pd.Series(white_balls).value_counts().sort_index()
```

```
plt.figure(figsize=(12, 6))
white_ball_freq.plot(kind='bar', color='blue')
plt.title('Frequency of White Balls')
plt.xlabel('White Ball Number')
plt.ylabel('Frequency')
plt.show()
```




```
# Visualize the frequency of Red Powerball numbers
red_ball_freq = data['Red Powerball'].value_counts().sort_index()

plt.figure(figsize=(12, 6))
red_ball_freq.plot(kind='bar', color='red')
plt.title('Frequency of Red Powerballs')
plt.xlabel('Red Powerball Number')
plt.ylabel('Frequency')
plt.show()
```



Display the Most Common numbers

```
# Display the most common numbers
print("Most common White Ball numbers:")
print(white_ball_freq.sort_values(ascending=False).head(10))

print("\nMost common Red Powerball numbers:")
print(red_ball_freq.sort_values(ascending=False).head(5))
```

```
Most common White Ball numbers:
36    151
39    150
23    143
32    143
28    140
59    138
21    138
54    136
12    135
19    134
Name: count, dtype: int64

Most common Red Powerball numbers:
Red Powerball
24    70
18    69
4     67
5     63
14    62
Name: count, dtype: int64
```

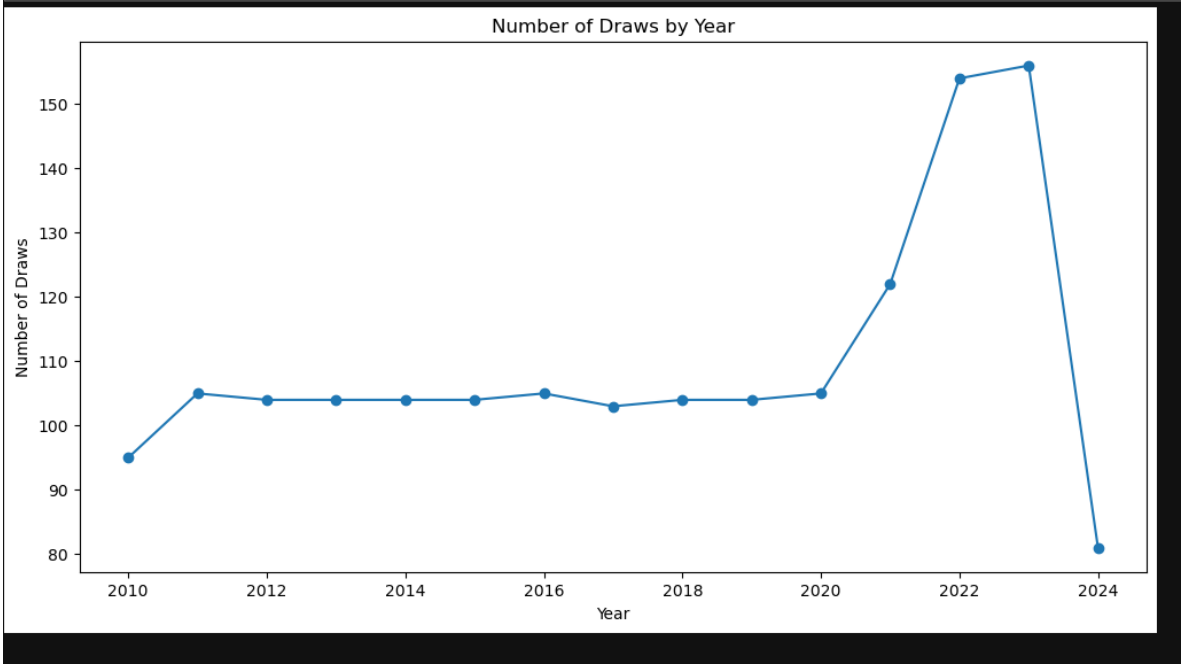
Frequency Combinations (most Common Pairs)

```
Most common pairs:
[((37, 44), 19), ((41, 59), 19), ((37, 39), 18), ((22, 32), 17), ((30, 59), 17), ((30, 48), 17), ((36, 52), 17), ((21, 32), 16), ((12, 20), 16), ((30, 5
3), 15)]

Most common triples:
[((1, 3, 13), 5), ((23, 28, 56), 5), ((18, 32, 45), 4), ((8, 44, 51), 4), ((1, 2, 7), 4), ((8, 17, 59), 4), ((12, 20, 21), 4), ((28, 53, 56), 4), ((7, 1
5, 36), 4), ((28, 40, 48), 4)]

Most common quadruples:
[((37, 52, 53, 58), 2), ((1, 3, 13, 44), 2), ((10, 24, 35, 53), 2), ((8, 31, 39, 43), 2), ((5, 18, 33, 43), 2), ((37, 44, 45, 53), 2), ((35, 41, 44, 58),
2), ((1, 2, 39, 66), 2), ((6, 8, 37, 40), 2), ((5, 23, 28, 56), 2)]

Most common quintuples:
[((11, 21, 27, 36, 62), 1), ((14, 18, 36, 49, 67), 1), ((18, 31, 36, 43, 47), 1), ((6, 24, 30, 53, 56), 1), ((5, 18, 23, 40, 50), 1), ((21, 37, 52, 53, 5
8), 1), ((6, 10, 31, 37, 44), 1), ((1, 3, 13, 44, 56), 1), ((18, 20, 27, 45, 65), 1), ((11, 28, 37, 40, 53), 1)]
```

Trend Analysis:

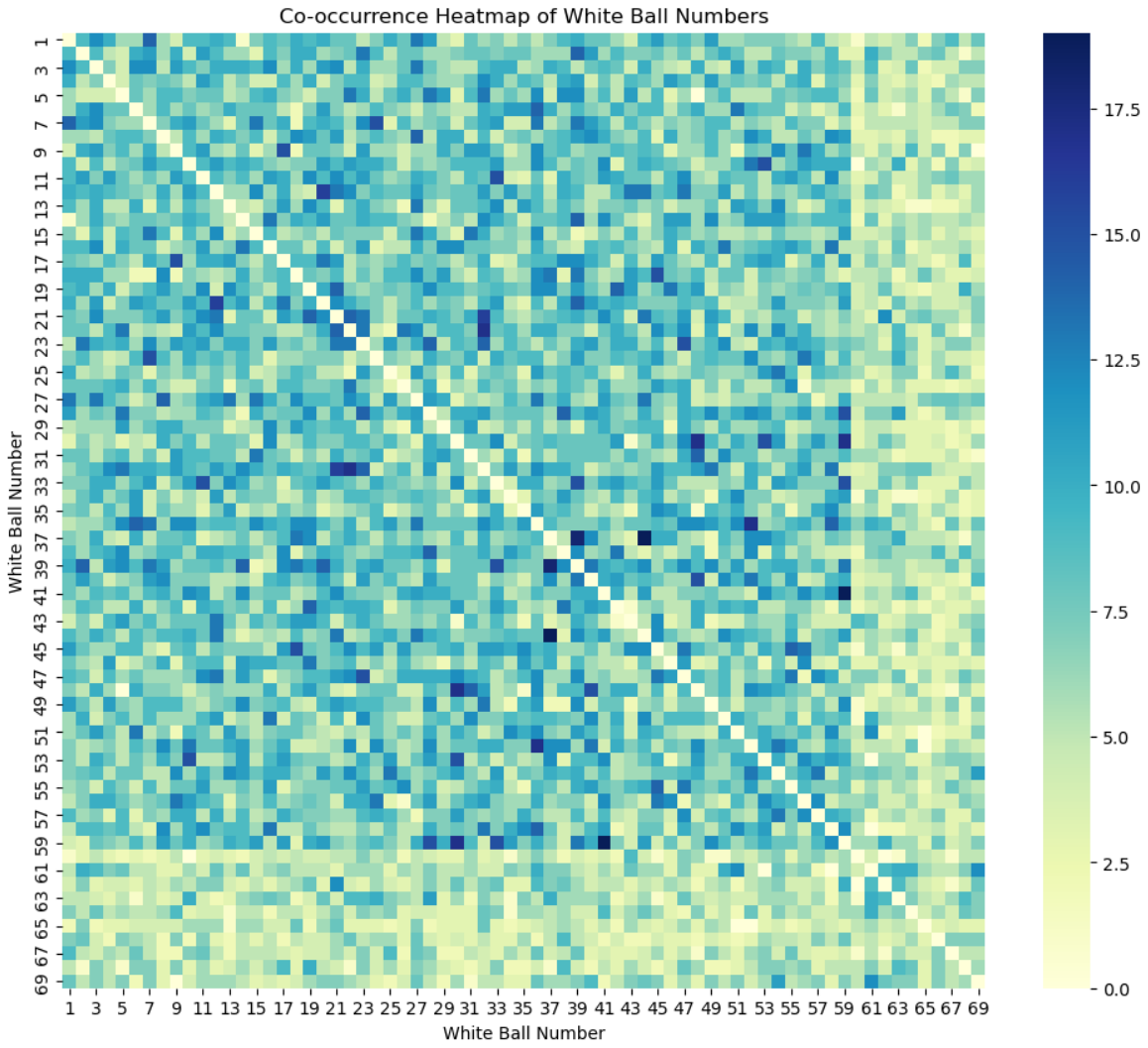
♦ Visualize the number of draws by year to identify trends

```python
# Visualize the trend of draws by year
data['Year'] = data['Draw Date'].dt.year
yearly_trends = data.groupby('Year').size()

plt.figure(figsize=(12, 6))
yearly_trends.plot(kind='line', marker='o')
plt.title('Number of Draws by Year')
plt.xlabel('Year')
plt.ylabel('Number of Draws')
plt.show()
```

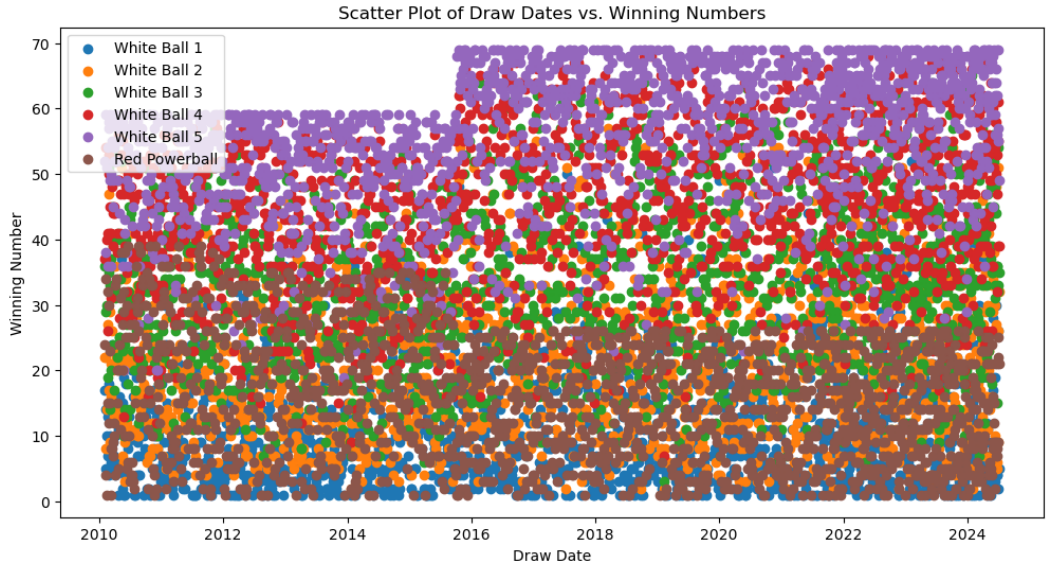


Pattern Recognition:

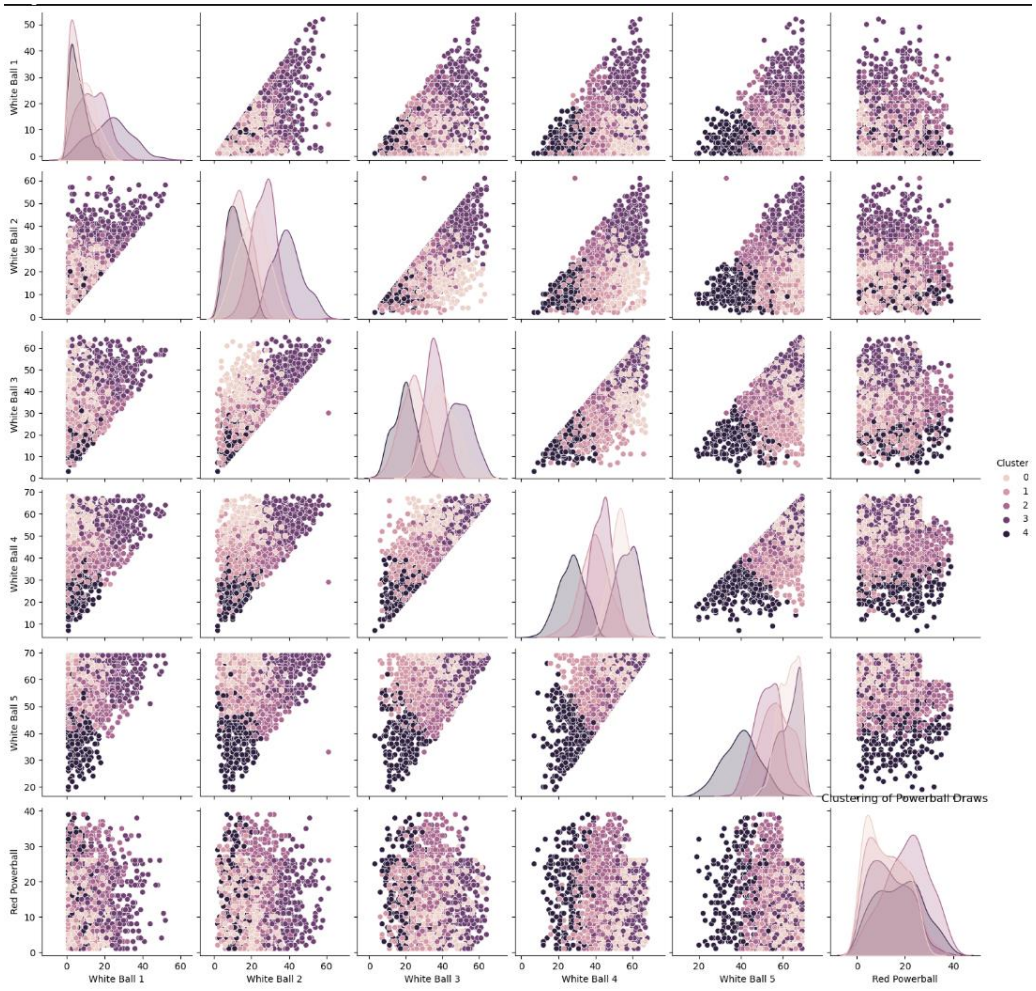
- **Heatmap**: Visualizing the co-occurrence of white ball numbers.



♦ **Scatter Plot**: Showing the relationship between drawing dates and winning numbers.

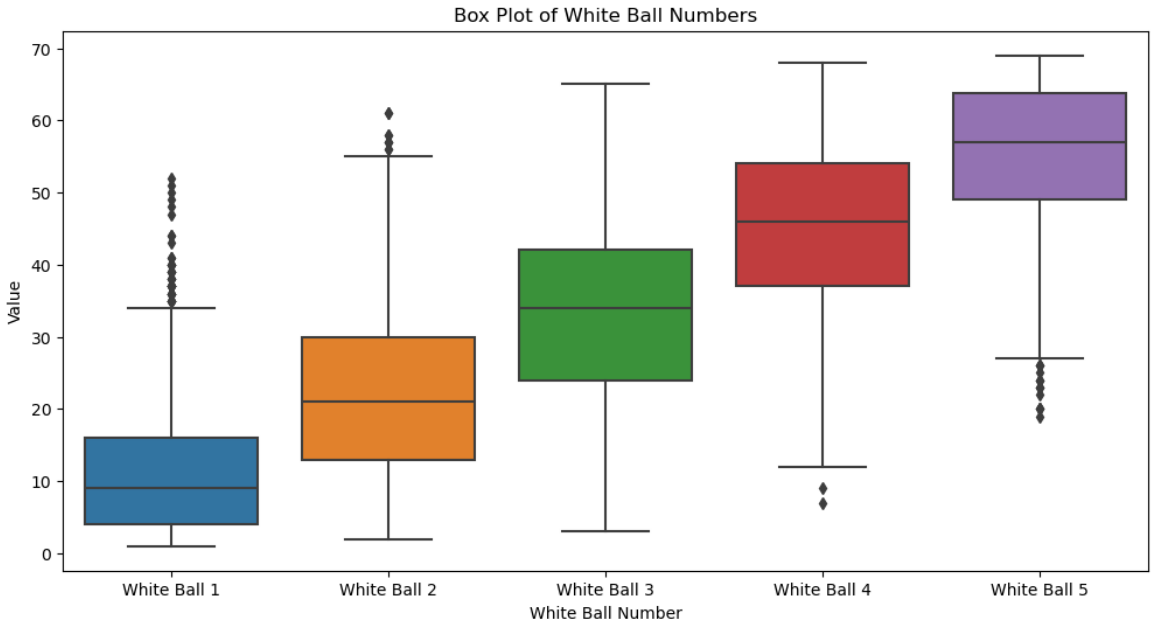


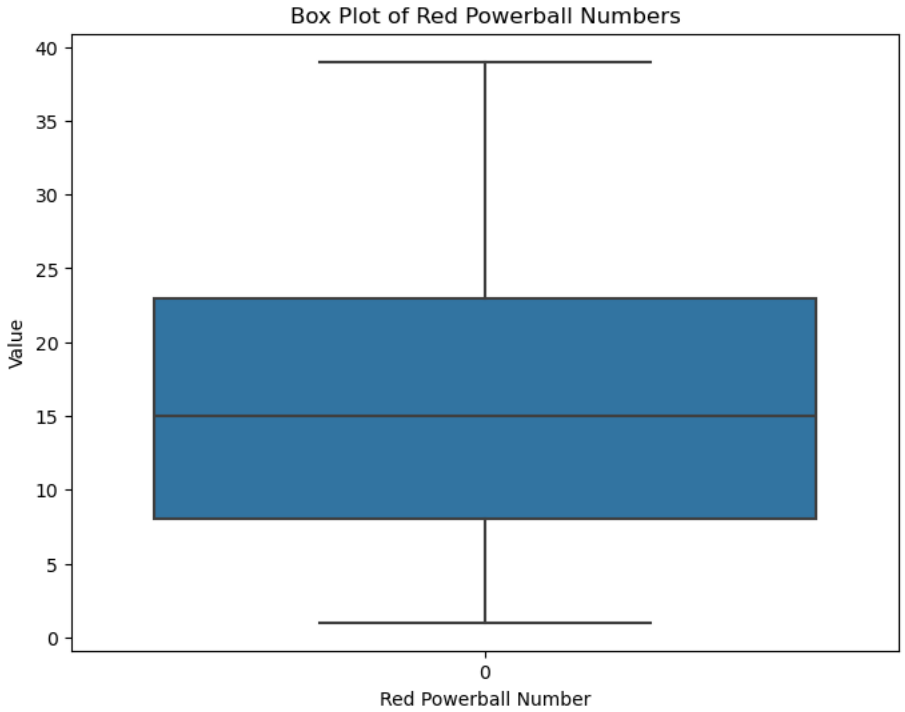
♦ **Clustering**: Using K-means clustering to group similar draws.



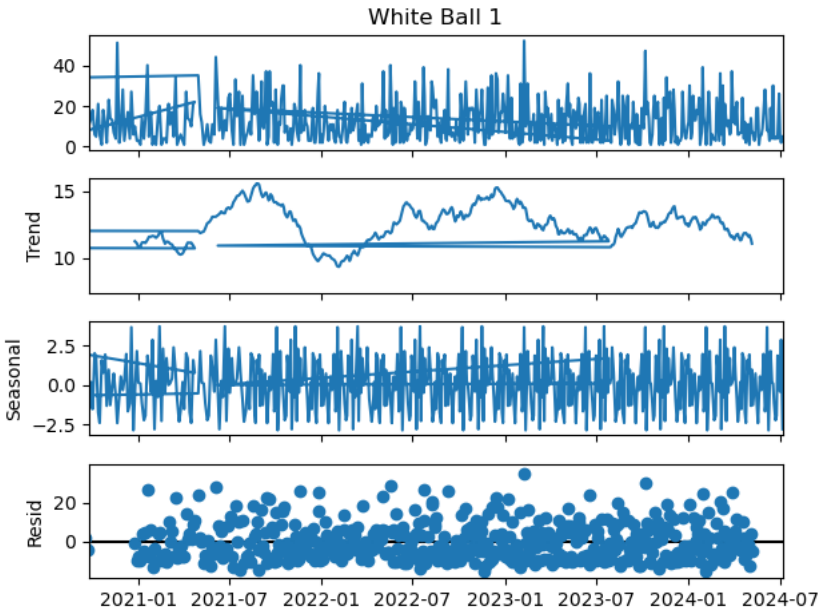
Distribution Analysis:

- **Box Plots**: Displaying the distribution of white ball numbers and the red Powerball number.





- **Time Series Decomposition**: Analyzing the trend, seasonality, and residuals of the drawn numbers.



Conclusion

Conducting Powerball-winning numbers does bring out general patterns and frequencies. Some frequently appearing numbers are 36, 39, and 23 in this category of numbers. Contains frequently observed pairs of points like (37, 44) and (41, 59). Also, common tuples of three, four, and five numbers were defined, adding more depth to the observations that the lottery numbers are random. K-means cluster analysis showed that similar drawings are grouped, which gave more depth to the drawing process analysis. The achieved results may provide significant information to all those who want to know more about the lottery statistics.

Assumptions

- The dataset from data.gov is accurate and comprehensive.
- The Powerball drawing process is random and unbiased.
- Historical patterns may provide insights, though future draws remain random.

Limitations

- Forecasting is done on historical performance, and this cannot determine the future draw.
- Includes data only from the year 2010 onwards; may differ from the previous years' data.
- Lottery drawings' randomness inherently reduces the efficacy of overall patterns that are pointed out.

Challenges

- Managing substantial amounts of data and data accuracy.
- The distinctions between trends in patterns and mere coincidence.
- Dealing with missing values as well as data transformation.

Implementation Plan

- **Data Collection and Preparation:**
  - Utilize data to pull historical and real-time data. gov and Powerball websites.
  - Clean the data by removing all the unnecessary elements and transform the raw data into a meaningful format that will be easy to analyze.
- **Exploratory Data Analysis:**
  - Conduct an EDA to learn the data characteristics and possible preliminary patterns.
- **Analysis:**
  - Carry out qualitative studies on frequency, tendency, pattern, and dispersion.
  - Also, present the results in the form of different charts and plots.
- **Documentation:**
  - Thus, keeping track of all the findings compiled in the report is an essential aspect of tracking.
  - Ensure presentation preparation and other related college materials.
- **Review and Recommendations:**
  - Evaluate the findings of the analysis and make sound recommendations.
  - Update the analysis periodically to incorporate new data

Ethical Assessment

- **Unbiased Interpretation:** Remember to be bias-free when developing the interpretations and recommendations and be fully aware that past performance does not guarantee future performance.
- **Responsible Messaging:** Mr. Argo tells the viewers that it is random and having numbers on the ticket does not necessarily mean one will get good cash.

Appendix

- **Data Source:**

  Dataset retrieved from data. gov and Powerball websites.

- **Python Code:**

  This is the complete source code for data cleaning and transformation diligently documented in the 'methods' section of the manuscript.

- **Visualization Samples:**

  Various categories of prominent charts are histograms, bar charts, line charts, heat maps, scatter plots chart, and others like box plots and grouping charts.

10 Questions an Audience Might Ask:

1. How accurate and reliable is the dataset used for the analysis?
2. Can the identified patterns and frequencies predict future Powerball draws?
3. How did you handle missing values and ensure data quality?
4. What were the most common individual white ball and red Powerball numbers?
5. Are there any observable trends over time in the winning numbers?
6. How did you identify and visualize the co-occurrence of white ball numbers?
7. What insights did clustering the draws provide?
8. How can this analysis benefit lottery players and researchers?
9. What are the limitations of this analysis in terms of predicting future outcomes?

10. How do you ensure ethical considerations, particularly data privacy and responsible messaging, in your analysis?

References

data.gov. (n.d.). Powerball Winning Numbers. Retrieved from https://catalog.data.gov/dataset/lottery-powerball-winning-numbers-beginning-2010

Powerball. (n.d.). Winning Numbers History. Retrieved from https://www.powerball.com/