NYCU 2023 Fall Semester - Data Visualization and Visual Analytics Final Project Report

The 22nd Group Members:
311554012 梁家瑀、311554022 劉冠慧

Introduction

- **Title:** CPBL's Cinderella Season in 2023
- **Data Source:** CPBL (中華職棒) official website (https://www.cpbl.com.tw/) [3]
- **Background:**
  The term "Cinderella season" is commonly used to describe a sports team or athlete that was initially considered less likely to succeed but performs exceptionally well during the season, achieving unexpected success or progress.

  This year, WeiChuan Dragons (味全龍) in the Chinese Professional Baseball League (CPBL) won the championship after a 24-year hiatus, marking their fifth title in franchise history.

  To explore this incredible season, we built this project for users to quickly gain insights into the performance of teams and players in CPBL during the Cinderella season.
- **Target Users:** Baseball enthusiasts and analysts (棒球愛好者、賽事分析人員).
- **Key Questions to Address:**
  Since the data on the CPBL official website are based on single games or overall results (Figure 1), it is difficult to directly observe the strength of the team. Therefore, **we designed this system to improve the data insights of their website**.

(a) CPBL dataset screen shot (single game)

球隊	出賽數	打數	得分	打點	安打	全壘打	壘打數	三振	四壞球	盜壘成功	上壘率	長打率	打擊率
味全龍	60	1970	232	217	480	34	683	419	166	36	0.313	0.347	0.244
樂天桃猿	60	2048	310	297	581	45	837	369	195	25	0.351	0.409	0.284
中信兄弟	60	1999	217	205	498	36	685	386	209	40	0.327	0.343	0.249
統一7-ELEVEn獅	60	2036	240	230	550	18	699	358	182	38	0.336	0.343	0.270
富邦悍將	60	1989	212	202	491	26	683	396	140	76	0.307	0.343	0.247

(b) CPBL dataset screen shot (overall results)

Figure 1: Data example with (a) single games (b) overall results

Besides, we also hope to use this system to reveal interesting insights into this incredible season. Therefore, our system needs to answer the following questions.

- **Q1: Pivotal moments or Turning points**
  - Were there key games, notable player performances, or strategic adjustments that played a crucial role? (是否有關鍵的比賽、球員的出色表現或策略調整發揮了關鍵作用？)

- **Q2: Teams' Performance**
  - How did different teams perform throughout the season, and what were the strengths and weaknesses of each team? (不同球隊整個賽季的表現如何，每支球隊的優勢和劣勢是什麼？)
  - What trends and changes in performance occurred during the season? (賽季中表現出現了哪些趨勢和變化？)

- **Q3: Reasons behind the Breakthrough**
  - What factors contributed to the WeiChuan Dragons's championship victory in the 2023 CPBL season? (是什麼因素促成了味全龍在2023年賽季奪冠？)

Visual Analytics

---

1. **Line chart (折線圖):**
   In this section, we use a line chart to display the "cumulative winning percentage ranking over time", which users can observe through three modes: the first half of the season, the second half of the season, and the entire season.

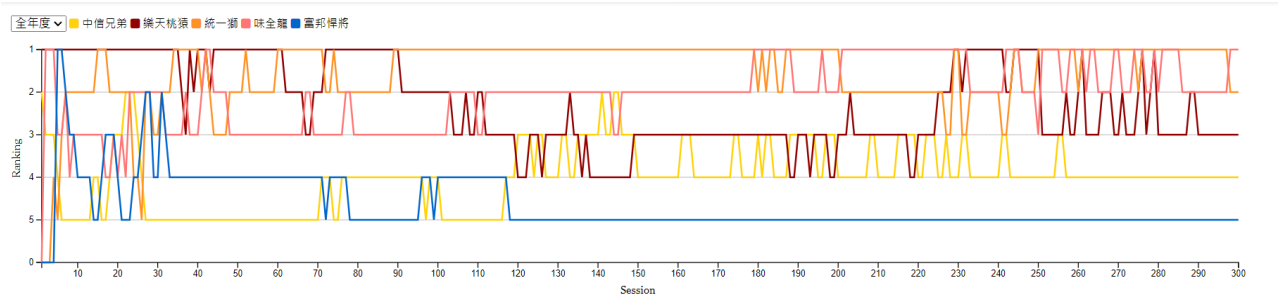


Figure 2: Interface of line chart

1.1. Advantages

Users can obtain the changes in the winning percentage of each team through the chart and can know whether a specific team performed well in a certain period.

Analysts can efficiently extract this information from the chart, enabling swift identification of games requiring analysis, as opposed to laboriously searching through data from the beginning to the end of the season.

1.2. Insights

Referring to Figure 2, the top two teams in overall winning rate are the WeiChuan Dragons (味全龍) and the Unilions (統一獅). According to the rule, the Unilions are required to engage in a best-of-four (四戰三勝) playoff series against the third-place, Rakuten Monkeys (樂天桃猿), to determine championship qualification.

While the Unilions exhibit a higher overall winning rate than the Rakuten throughout the entire season, a closer examination of the first and second halves (Figure 3 and Figure 4) reveals a noteworthy pattern.

The Unilions performed exceptionally well in the first half but experienced a significant decline in the second half. Conversely, Rakuten's performance improved over time. This indicates that the playoffs may favor Rakuten, elucidating why the first-half season champion (Unilions) is not eligible for the championship game.

Subsequently, we will delve into more detailed analysis using matrix plots.

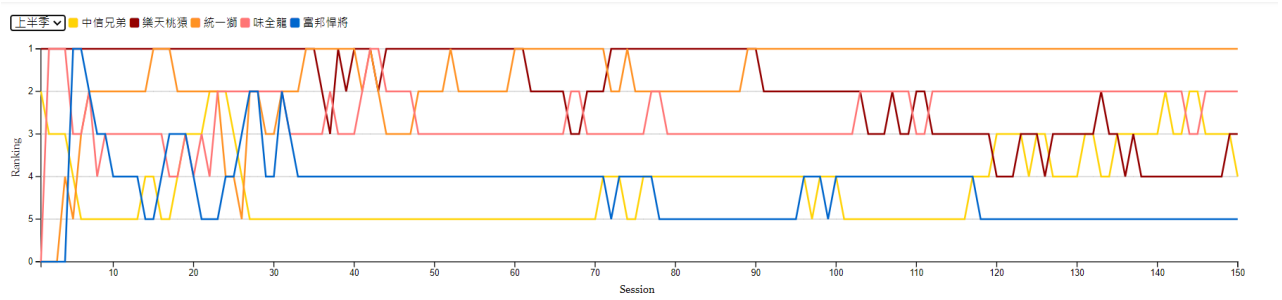


Figure 3: Line chart - first half of season

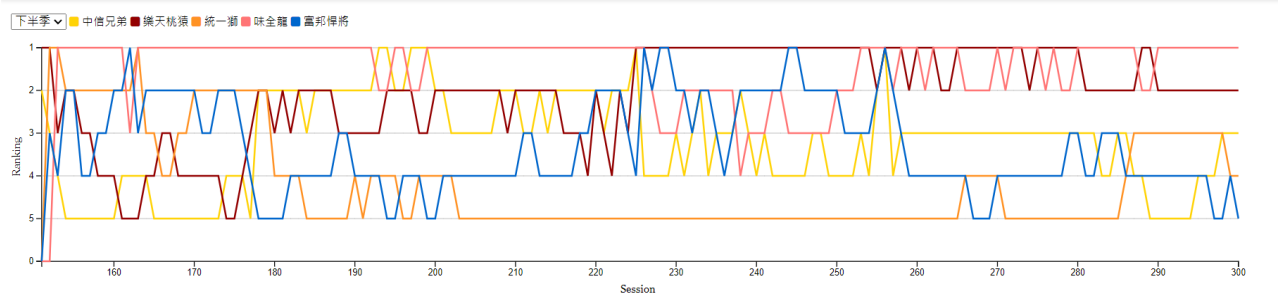


Figure 4: Line chart - second half of season

2. **Matrix plot (矩陣圖):**

The matrix chart is our innovative design, merging elements of a histogram and a pie chart. It shows the outcomes of all matches, where the home games (主場比賽) are shown vertically, and the away games (客場比賽) are shown horizontally. The final scores are in the circles (home team first).

We use cyan to indicate “the home winning rate is greater than the away winning rate”, otherwise we use dark gray. Similarly, we use orange and light gray to represent the “away win rate is greater or smaller than the home game”.

The diagram can be divided into three parts:

- Upper triangular matrix (上三角矩陣):
  The upper part focuses on the away performance. The middle circle is away games and the outer circle is home games.

- Diagonal matrix (對角線矩陣):
  The diagonal part shows the team logo of each team. When the user touches it, the corresponding home and away winning rate, draw rate, and defeat rate of the selected team will be displayed (shown in Figure 6).

- Lower triangular matrix (下三角矩陣):
  The lower part focuses on the home performance. The middle circle is home games and the outer circle is away games.

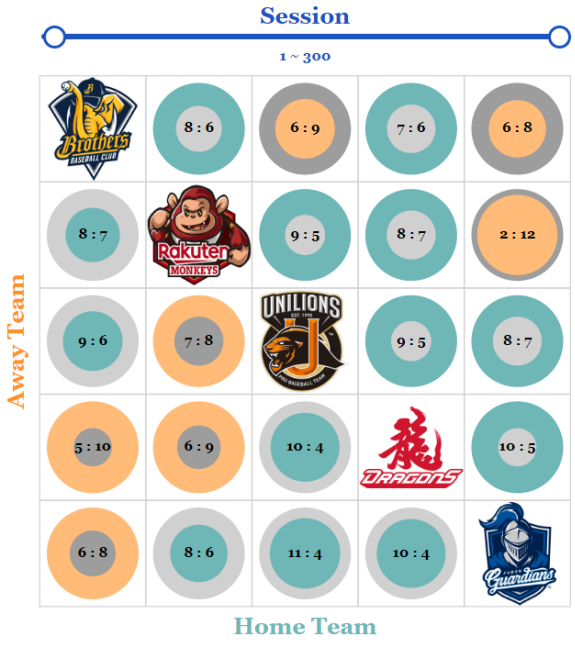


Figure 5: Interface of matrix plot

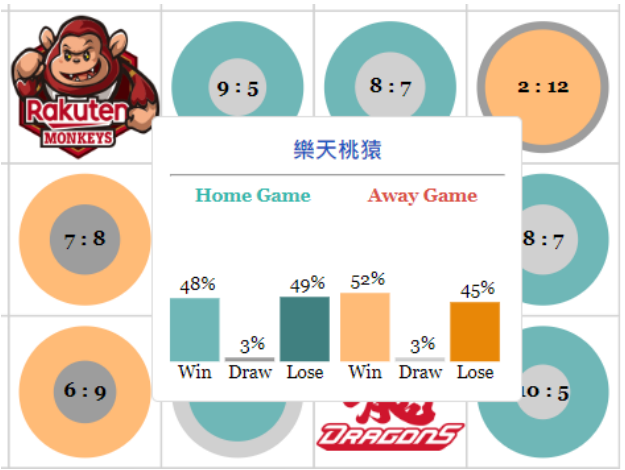


Figure 6: Matrix plot - diagonal tooltip

2.1. Advantages

The matrix chart offers a holistic overview of match outcomes, presenting winning numbers of both home and away games. The chart enhances visual contrast by using distinct colors, enabling users to easily identify trends and patterns.

Efficiently divided into upper and lower triangular matrices, the chart allows users to analyze away separately and home performances, fostering a more focused examination of team dynamics in different playing conditions. The inclusion of an interactive diagonal matrix featuring team logos adds engagement, as users can touch a team logo to access detailed information, including home and away winning rates, draw rates, and defeat rates. This feature facilitates in-depth team-specific analysis.

2.2. Insights

Continuing the analysis of the line chart, we found that in the first half of the season (Figure 7), Rakuten performed poorly at home and away, and both lost to the Unilions.

However, in the second half of the season (Figure 8), Rakuten's home winning rate improved a lot, especially against the Unilions, where its home winning rate increased from 37.5% to 57.1%. This proves why Rakuten climbed up the rankings in the second half of the season and was able to effectively defeat the Unilions at the end of the season.

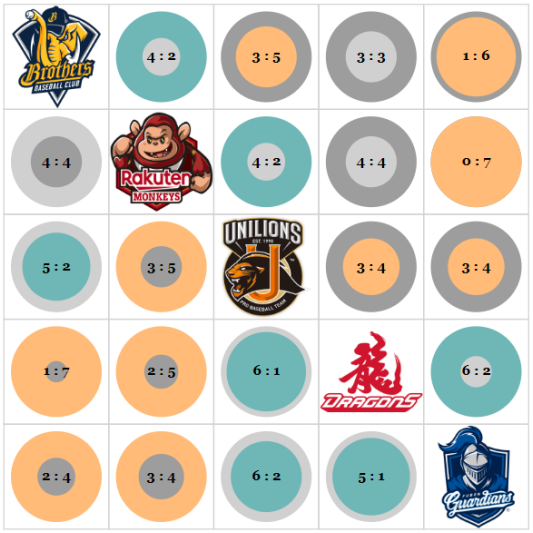


Figure 7: Matrix plot - first half of season

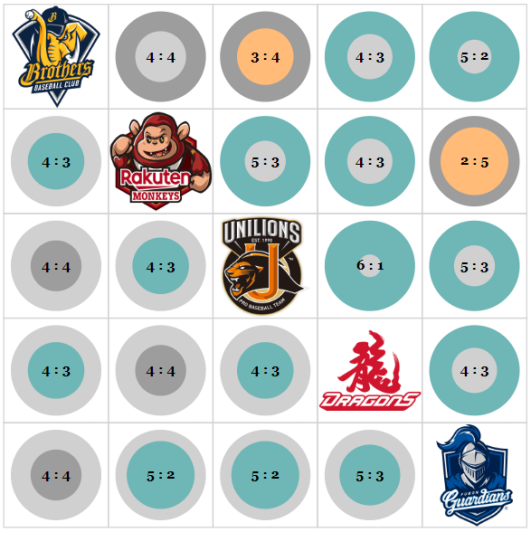


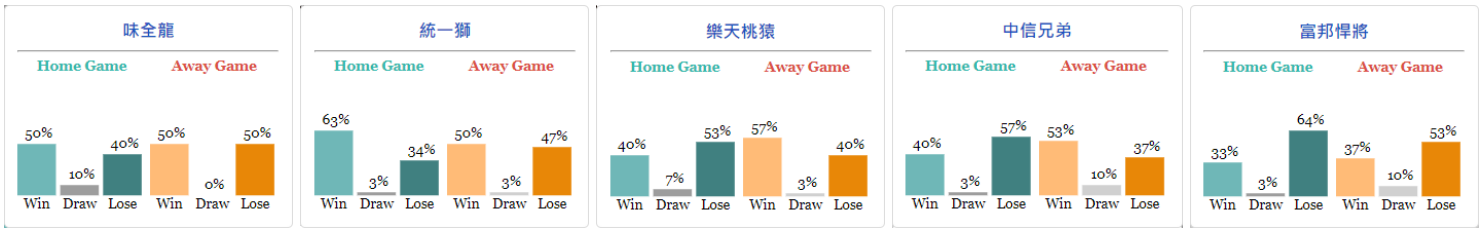
Figure 8: Matrix plot - second half of season

Figure 9 displays the tooltip in a matrix plot for each team across various seasons, organized by season ranking from left to right and by season order from top to bottom.

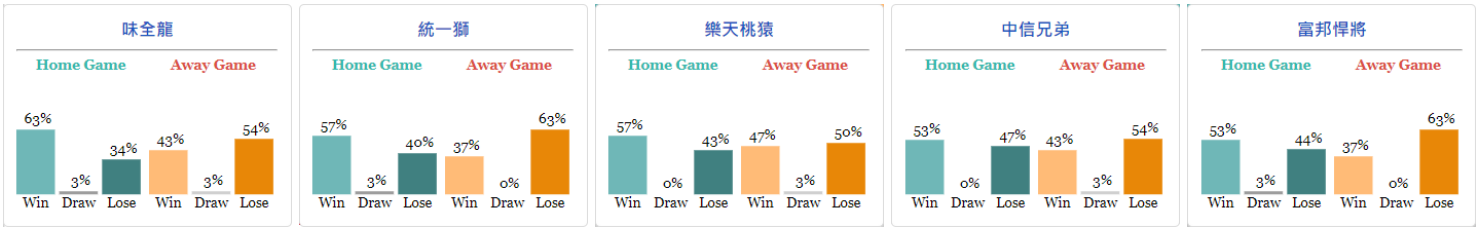
We observed an interesting phenomenon. In the first half of the season, most teams focused on their performance in away games, except for Unilions and WeiChuan Dragons, which achieved excellent results at home and away. In the second half of the

season, each team focused on home performance, so every team's home winning rate exceeded 50%.

Furthermore, during the second half of the season, while most teams concentrated on home performance, the Unilions exhibited a different trend. Upon further investigation, it was discovered that due to personnel scheduling, key pitchers Mario Sanchez (勝騎士) and Logan Ondrusek (羅昂), who played pivotal roles in the first half, were unavailable. Moreover, offensive performance did not match the standards set in the initial part of the season, leading to a decline in overall performance in both home and away games for the Unilions.



(a) first half of season



(b) second half of season

Figure 9: Matrix plot - tooltip of all teams
(a) home and away performance in first half of season (b) home and away performance in second half of season

3. Sunburst chart (旭日圖):

Sunburst charts are powerful visual tools that we used to depict the distribution of pitcher and hitter MVP awards across various months this season. The purpose of creating this chart is to offer users and analysts a comprehensive understanding of the seasonal dynamics, recognizing outstanding performances in both pitching and hitting monthly.

3.1. Advantages

This visual aid facilitates the identification of trends, patterns, and standout months for individual players or teams. Users can easily compare the distribution of MVP awards, gaining valuable insights into the performance variations over time.

3.2. Insights

We found that WeiChuan Dragons consistently secured the highest number of MVP awards throughout the season, with an even distribution between the first and second halves. This indicates the hitting and pitching stability of their team. Additionally,

Rakuten Monkeys's MVPs are concentrated primarily in the second half, signifying an improvement in their performance during this period.

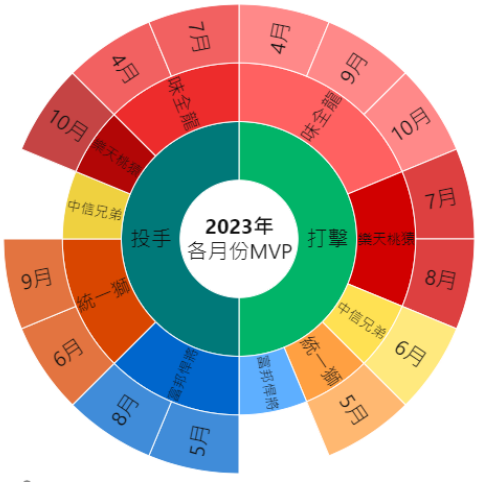


Figure 10: Interface of sunburst chart

4. **Dot plot (點狀圖):**

In this part, we use a dot plot to show the “team pitching and hitting performance in each game”, in which we use the earned run average (ERA, 防禦率) and the batting average (BA, 打擊率) to represent the pitching and hitting performance respectively.

In baseball, the batting average (BA, 打擊率) is calculated by dividing hits (H, 安打數) by at-bats (AB, 打數). Generally, professional players are deemed competent with a batting average above .280, an average exceeding .300 is considered excellent for the season, and surpassing .400 is an almost unattainable goal [1].

The other part, the earned run average (ERA, 防禦率) is calculated by dividing earned runs (ER, 失分) multiplied by nine by innings pitched (IP, 投球局數). The ERA can more clearly reflect a pitcher's performance. Generally, a pitcher with an ERA between 3 and 4 is considered a competent pitcher, while an ERA between 2 and 3 is considered excellent, and an ERA below 2 is considered elite for a pitcher [2].

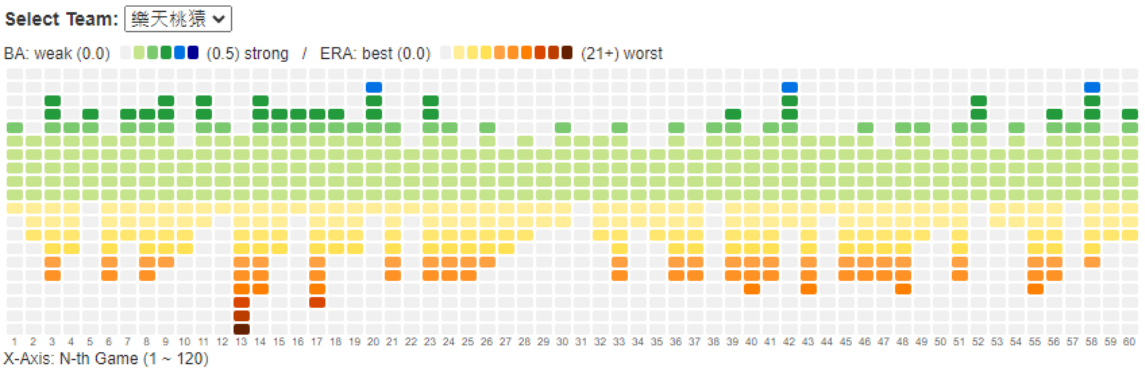


Figure 11: Interface of dot plot

4.1. Advantages

Here, we categorize both batting average (BA) and earned run average (ERA) into 10 intervals, utilizing distinct colors to represent varying degrees of quality, as shown in Table 1 and Table 2.

Table 1: Range and quality of batting average (BA)

	Level 1 (weak)	Level 2	Level 3	Level 4	Level 5 (normal)	Level 6 (competent)	Level 7 (excellent)	Level 8	Level 9 (elite)	Level 10 (best)
BA	0.0~0.05	0.05~0.1	0.1~0.15	0.15~0.2	0.2~0.25	0.25~0.3	0.3~0.35	0.35~0.4	0.4~0.45	0.45~

Table 2: Range and quality of earned run average (ERA)

	Level 1 (best)	Level 2	Level 3 (excellent)	Level 4 (competent)	Level 5 (normal)	Level 6	Level 7	Level 8	Level 9	Level 10 (worst)
ERA	0~1	1~2	2~3	3~4	4~5	5~9	9~13	13~17	17~21	21~

Given that a higher batting average is preferable and a lower earned run average is favorable, this point chart effectively integrates two indicators with contrasting trends. In summary, a higher position of the colored area generally corresponds to better performance.

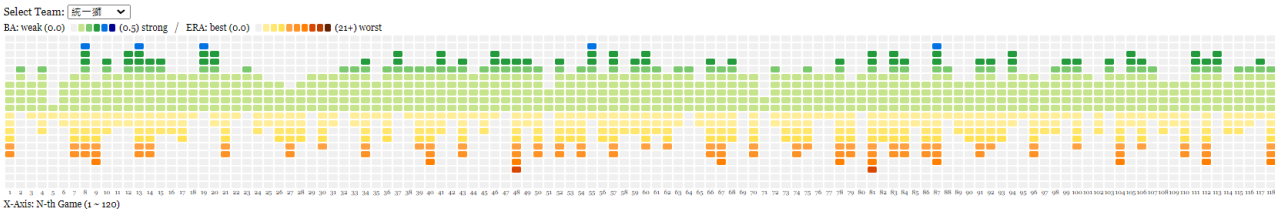
4.2. Insights

We provide examples of three teams along with corresponding news links to illustrate the observed phenomena in this chart.

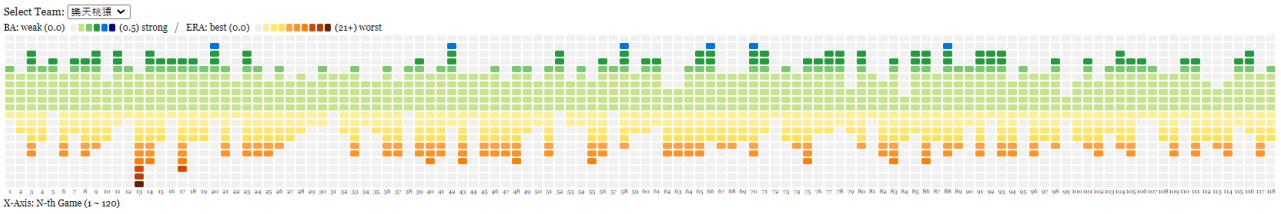
- Unilions (統一獅):
  When comparing the performance between the first half of the season (sessions 1-60) and the second half (sessions 61-120), a notable deterioration in the pitcher's earned run average (ERA) is evident. It can be attributed to the absence of the key pitchers, Mario Sanchez (勝騎士) [5] and Logan Ondrusek (羅昂) [6], in the second half of the season.

- Rakuten Monkeys (樂天桃猿):
  During the first half of the season, their Earned Run Average (ERA) was notably high. However, in the second half of the season, a significant improvement occurred after the appointment of Kawagishi Tsuyoshi [7] as the new pitching coach. With a more effective pitching arrangement, the team's ERA considerably decreased.

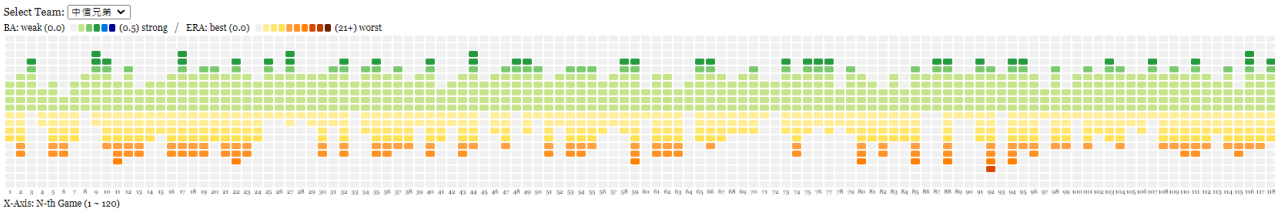
- CTBC Brothers (中信兄弟) [8]:
  Throughout the entire season, their performance in both pitching and hitting was average, raising the awareness that this might be related to player scheduling.



(a) Unilions



(b) Rakuten Monkey



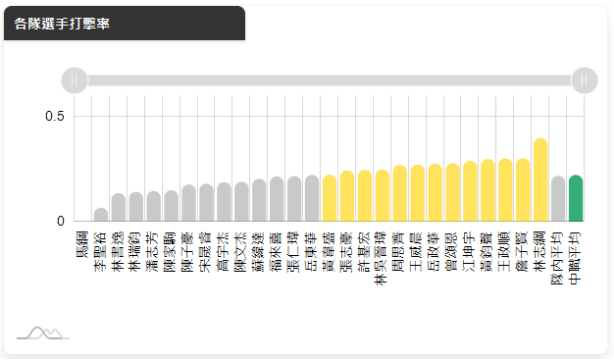
(c) CTBC Brothers

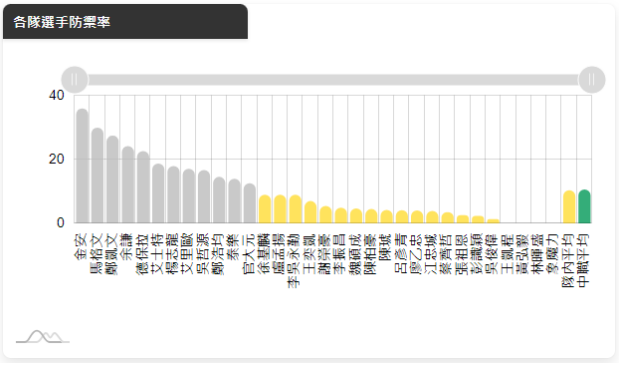
Figure 12: Dot plot with (a)Unilions (b) Rakuten Monkeys (c) CTBC Brothers

5. Bar chart (直條圖):

Combined with the results from the dot plot, our system will present the season performance of pitchers and hitters in the form of a bar graph based on the selected team.

In the bar chart, hitters with a batting average (BA) surpassing the league's season average and pitchers with an earned run average (ERA) below the league's season average will be highlighted in the team's representative color. Conversely, those not meeting these criteria will be displayed in gray (refer to Figure 13).





(a) performance of team's batters **(b) performance of team's pitchers**

Figure 13: Interface of bar chart

5.1. Advantages

From Figure 13, we can observe each team's performance in hitting and pitching. In more detail, it allows a closer look at individual player performances within each team.

This provides baseball fans and analysts with valuable insights into a comprehensive understanding of team performance and individual player contributions.

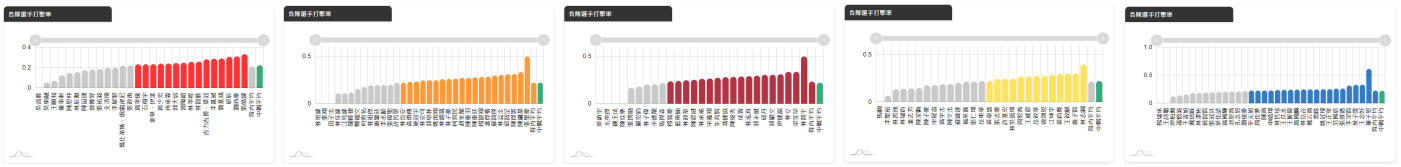
5.2. Insights

Figure 14 is a bar chart of each team, which is sorted by season ranking, from left to right: WeiChuan Dragons (味全龍), Unilions (統一獅), Rakuten Monkeys (樂天桃猿), CTBC Brothers (中信兄弟), and Fubon Guardians (富邦悍將).

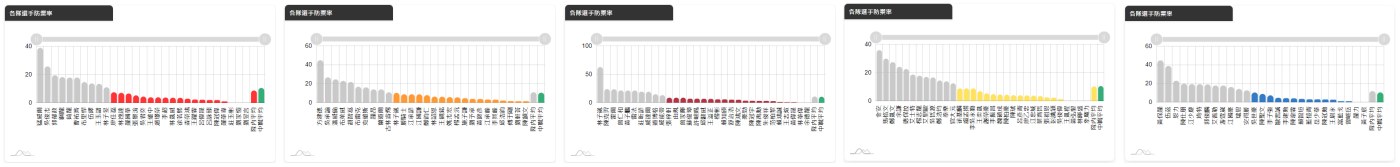
The figure illustrates that WeiChuan Dragons (味全龍) and CTBC Brothers (中信兄弟) are defensive-oriented (防守型隊伍) (i.e. the pitchers outperform hitters), whereas the other three teams are offensive-oriented (進攻型隊伍).

We've noticed that the WeiChuan Dragons have an average defensive rating of 8.8, while the CTBC Brothers' is 10.36. Although the other three teams all exceed the league average, the difference is approximately 1 point.

Consequently, the CTBC Brothers, with their defense and standard hitting, exhibit a notable gap in ranking compared to the WeiChuan Dragons, both are defensive teams. This underscores the substantial impact of pitcher performance on a team's overall winning rate.



(a) batting average
(WeiChuan Dragons / Unilions / Rakuten Monkey / CTBC Brothers / Fubon Guardians)



(b) earned run average
(WeiChuan Dragons / Unilions / Rakuten Monkey / CTBC Brothers / Fubon Guardians)

Figure 14: Bar chart - all teams with (a) batting performance (b) pitching performance

Conclusion

In light of the remarkable Cinderella season in the CPBL in 2023, notably marked by the Wei Chuan Dragons' historic fifth championship after a 24-year hiatus, we've developed a system to enhance the visualization aspect of the data from CPBL official website. This system enables baseball enthusiasts and analysts to efficiently explore the season's events, including team performances, pitching, and hitting analyses, all presented in graphical formats.

We also discovered a lot of interesting insights from our system. Next, we summarize the report and answer the key questions raised.

Q1: Pivotal moments or Turning points

The analysis suggests a few potential pivotal moments or turning points in the season:

- The decline in Unilions's performance in the second half, possibly due to the absence of key pitchers Mario Sanchez and Logan Ondrusek.
- Rakuten's improved performance on earned run average (ERA) in the second half after appointing Kawagishi Tsuyoshi as the new pitching coach.

Q2-1: Teams' Performance

- Unilions started strong but experienced a decline in the second half, possibly due to the absence of key pitchers.
- Rakuten Monkeys improved their performance in the second half, especially in home games.
- WeiChuan Dragons consistently performed well, securing the highest number of MVP awards, indicating a balanced team with stable hitting and pitching.
- CTBC Brothers and Fubon Guardians had average performances, with CTBC Brothers being defensive-oriented and Fubon Guardians showing offensive strength.

Q2-2: Trends and Changes

Most teams focused on away games in the first half, but in the second half, there was a shift towards focusing on home performance.

Unilions had a different trend in the second half, with a decline in overall performance due to scheduling issues and key player unavailability.

Q3: Reasons behind the Breakthrough

The factors contributing to WeiChuan Dragons's championship victory could include their consistent performance throughout the season, a balanced team with both defensive and offensive capabilities, and an effective pitching rotation. The team's ability to secure MVP awards throughout the season suggests stability in both hitting and pitching, which likely played a crucial role in their success.

Further analysis of specific game statistics, player contributions, and strategic decisions during key matches could provide more insights into the reasons behind their breakthrough.

Reference

[1] Batting Average - Wikipedia (https://shorturl.at/eklGV)
[2] Earned Run Average - Wikipedia (https://shorturl.at/xzBQR)
[3] Game dataset - CPBL (https://www.cpbl.com.tw/schedule)
[4] Player dataset - github CBPL open dataset (https://github.com/ldkrsi/cpbl-opendata)
[5] News for Unilions pitcher - Mario Sanchez (https://shorturl.at/ejJV8)
[6] News for Unilions pitcher - Logan Ondrusek (https://shorturl.at/jsAS2)
[7] News for Rakuten Monkey (https://udn.com/news/story/7001/7501179)
[8] News for CTBC Brothers (https://udn.com/news/story/122629/7511975)