CHESS API

GROUP MEMBERS

23229574-Rahim
23229565-Minh
22229598-Nilasi

GROUP 14

TABLE OF CONTENTS

- INTRODUCTION
- RESEARCH QUESTION 1
- RESEACH QUESTION 2
- CONCLUSION
- REFERNCES

- Chess.com is one of the largest online chess platforms in the world, offering players the ability to play games, solve puzzles, take lessons, and follow tournaments.

- Game Modes: players can choose from several time-control categories that significantly affect the pace and style of play:

  BULLET: VERY FAST GAMES.(3 MIN)
  BLITZ: FAST GAMES(3-5MIN)
  RAPID: MORE TIME PER PLAYER(10+ MIN)
  DAILY: MUCH SLOWER (HOURS OR DAYS)

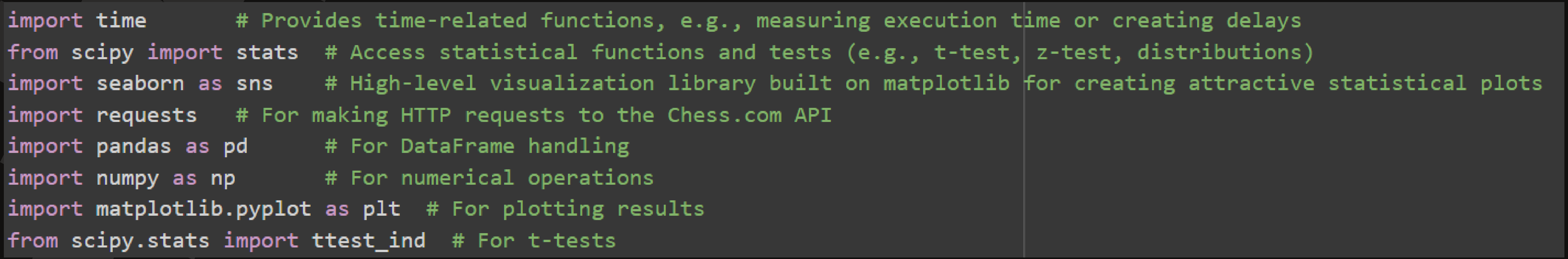
01

HOW DOES PLAYING WHITE VERSUS BLACK EFFECT THE PROBABILITY OF WINNING ACROSS DIFFERENT CHESS TIME CONTROLS (BULLET, BLITZ, RAPID, DAILY)?

- **H_0 (Null Hypothesis):** The effect of playing White on win probability does not vary by time control.
- **H_1 (Alternative Hypothesis):** The advantage of playing White differs depending on the time control (e.g., bigger in rapid/classical, smaller in blitz/bullet).

IMPORTS - GETTING THE DATA

Libraries Used:

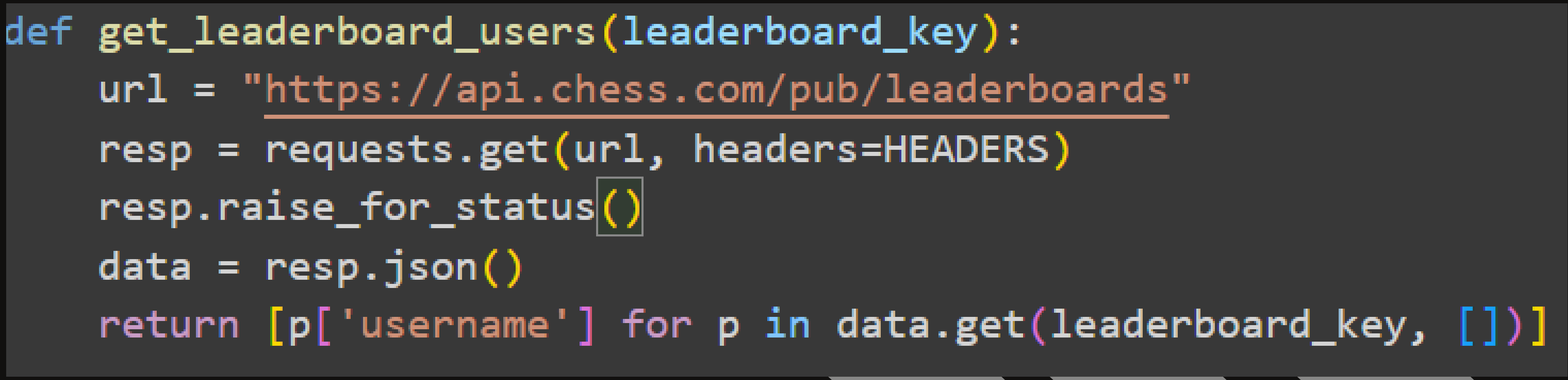


```python
import time       # Provides time-related functions, e.g., measuring execution time or creating delays
from scipy import stats  # Access statistical functions and tests (e.g., t-test, z-test, distributions)
import seaborn as sns    # High-level visualization library built on matplotlib for creating attractive statistical plots
import requests    # For making HTTP requests to the Chess.com API
import pandas as pd      # For DataFrame handling
import numpy as np       # For numerical operations
import matplotlib.pyplot as plt  # For plotting results
from scipy.stats import ttest_ind  # For t-tests
```

Short code

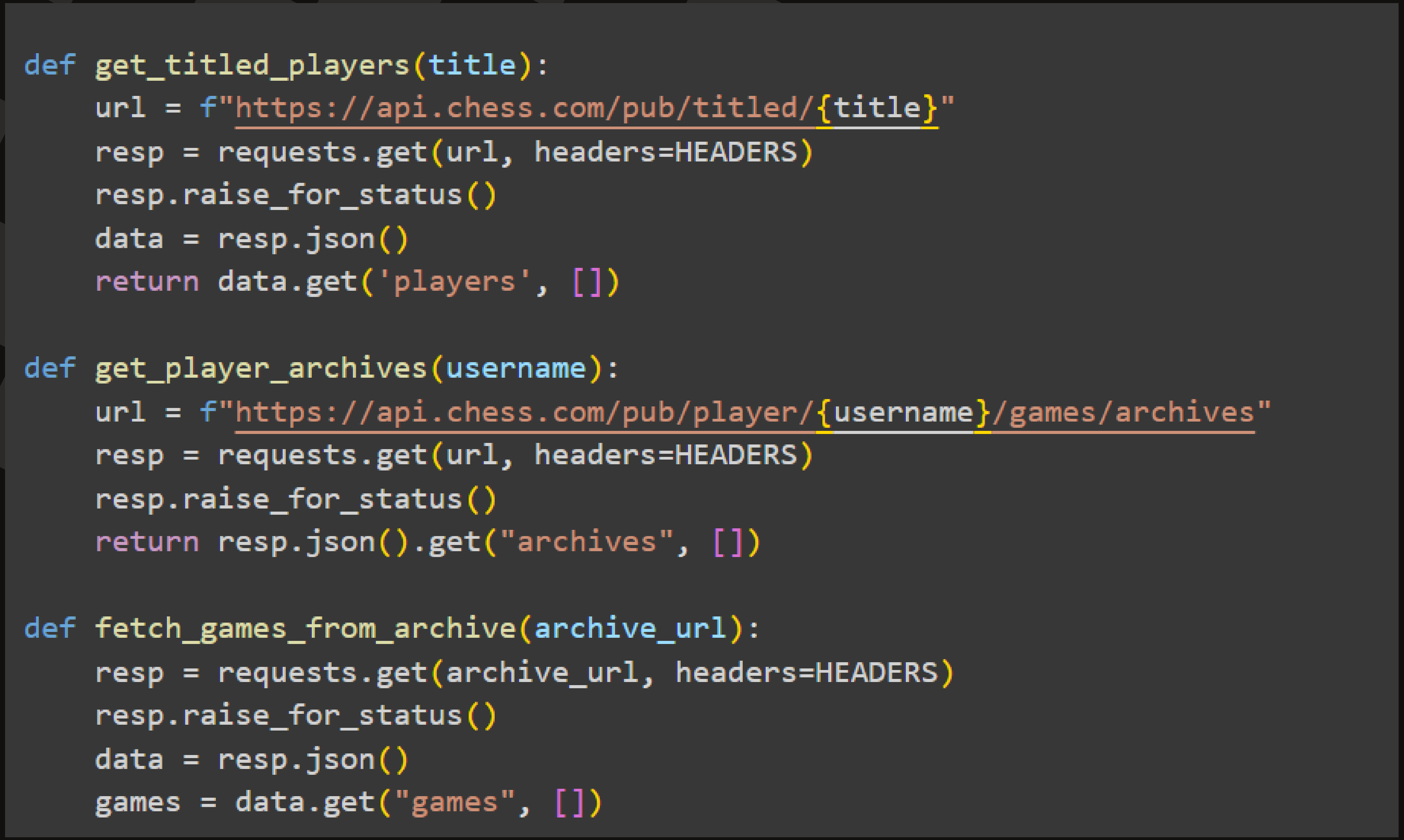
HEADERS = {"User-Agent":"MyChessApp/1.0"} → “Identifies our app to the API.”
get_player_archives(username) → “Gets monthly archive URLs for that user.”
fetch_games_from_archive(url) → “Downloads that month’s games and returns structured records.”
df["win"] = np.where(df["result"] == "win", 1, 0) → “Creates a binary win column: 1 = win, 0 = else.”
df["color"] = np.where(df["color"]=="white", 1, 0) → “Encodes color numerically for analysis.”

Pulling the Data

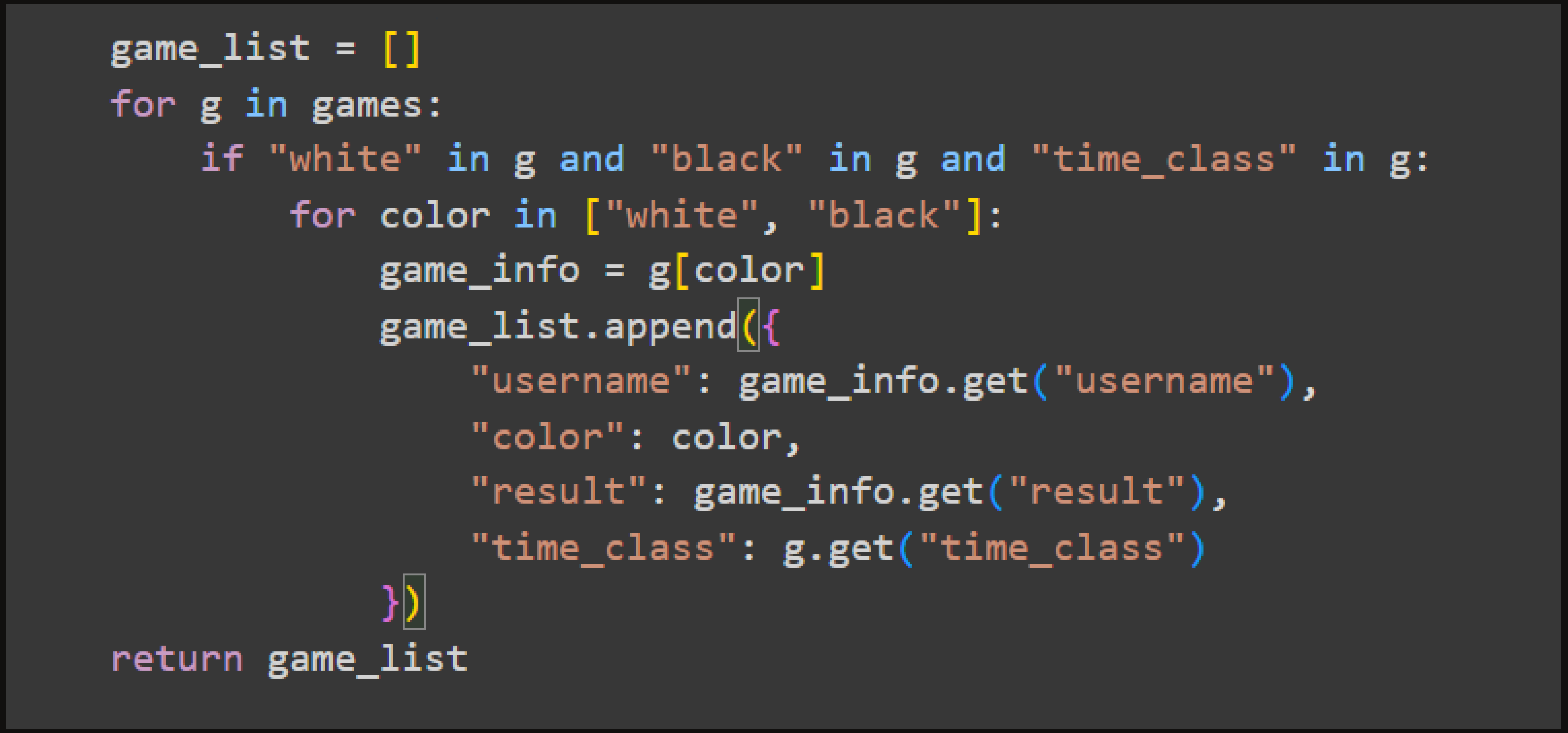


```python
def get_leaderboard_users(leaderboard_key):
    url = "https://api.chess.com/pub/leaderboards"
    resp = requests.get(url, headers=HEADERS)
    resp.raise_for_status()
    data = resp.json()
    return [p['username'] for p in data.get(leaderboard_key, [])]
```

RINSE AND REPEAT:



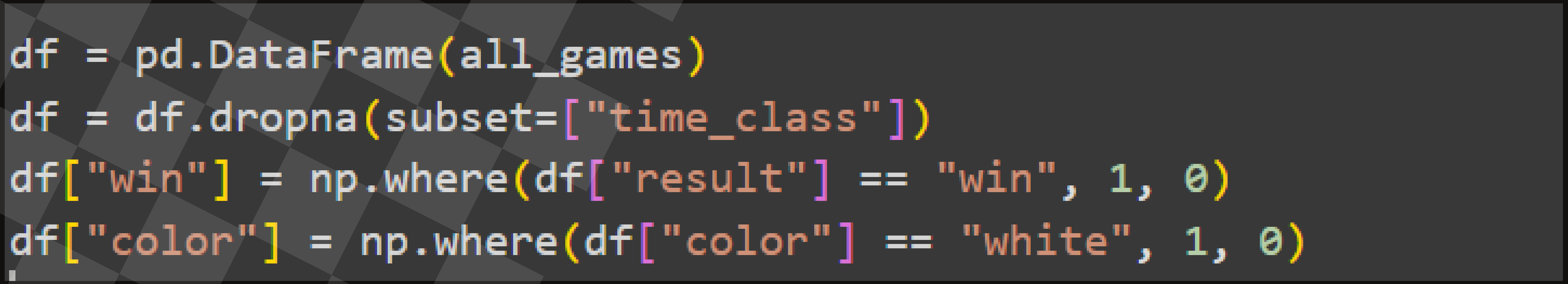
```python
def get_titled_players(title):
    url = f"https://api.chess.com/pub/titled/{title}"
    resp = requests.get(url, headers=HEADERS)
    resp.raise_for_status()
    data = resp.json()
    return data.get('players', [])

def get_player_archives(username):
    url = f"https://api.chess.com/pub/player/{username}/games/archives"
    resp = requests.get(url, headers=HEADERS)
    resp.raise_for_status()
    return resp.json().get("archives", [])

def fetch_games_from_archive(archive_url):
    resp = requests.get(archive_url, headers=HEADERS)
    resp.raise_for_status()
    data = resp.json()
    games = data.get("games", [])
```



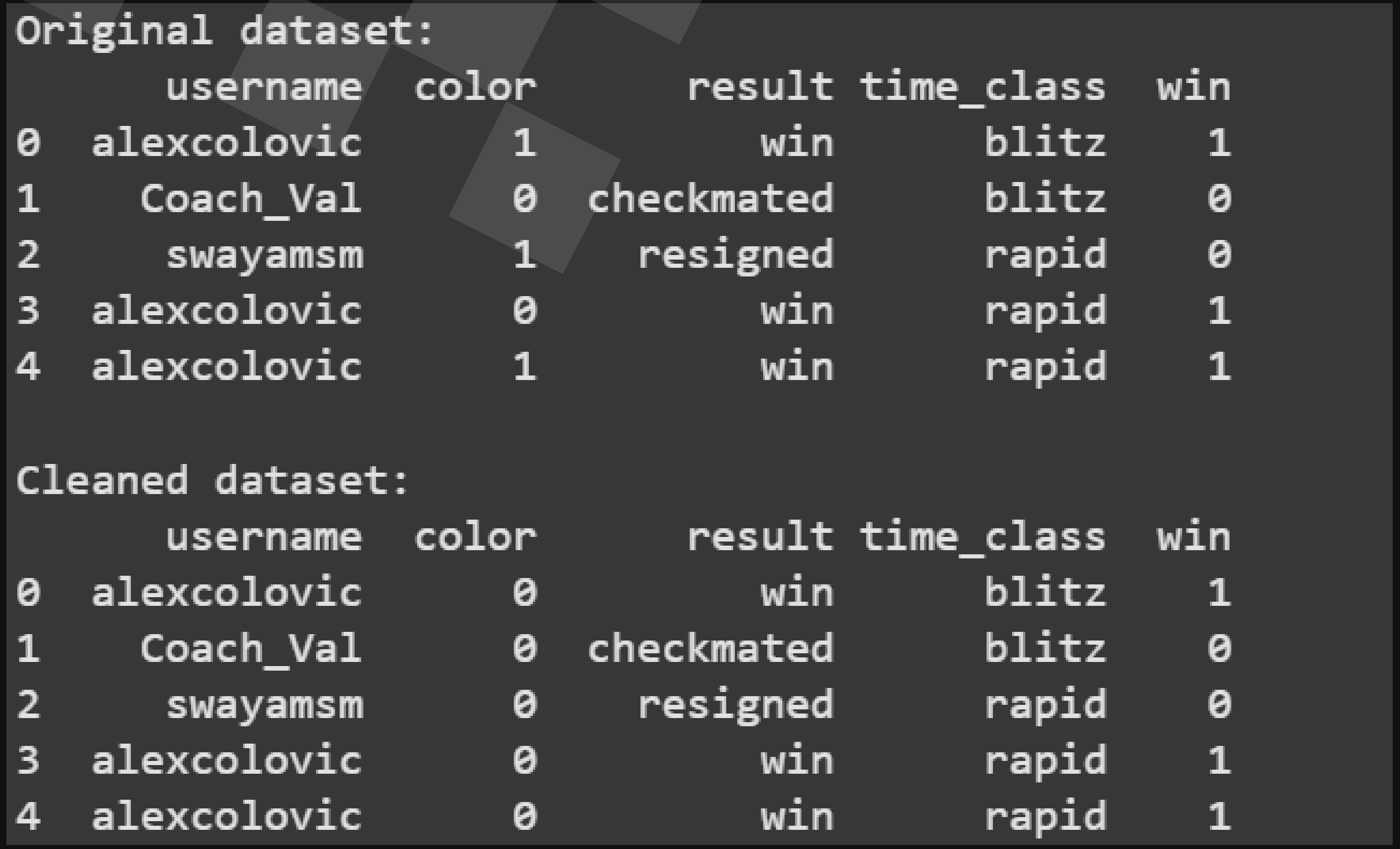
```python
    game_list = []
    for g in games:
        if "white" in g and "black" in g and "time_class" in g:
            for color in ["white", "black"]:
                game_info = g[color]
                game_list.append({
                    "username": game_info.get("username"),
                    "color": color,
                    "result": game_info.get("result"),
                    "time_class": g.get("time_class")
                })
    return game_list
```

BUILDING AND CLEANING THE DATA:

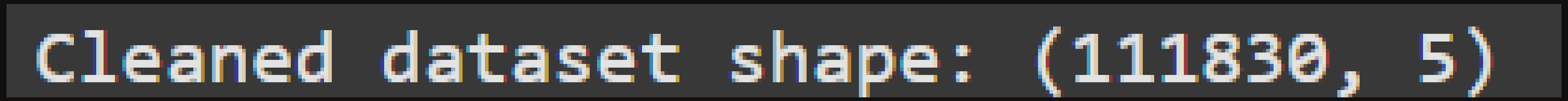
Old vs Clean



```
df = pd.DataFrame(all_games)
df = df.dropna(subset=["time_class"])
df["win"] = np.where(df["result"] == "win", 1, 0)
df["color"] = np.where(df["color"] == "white", 1, 0)
```



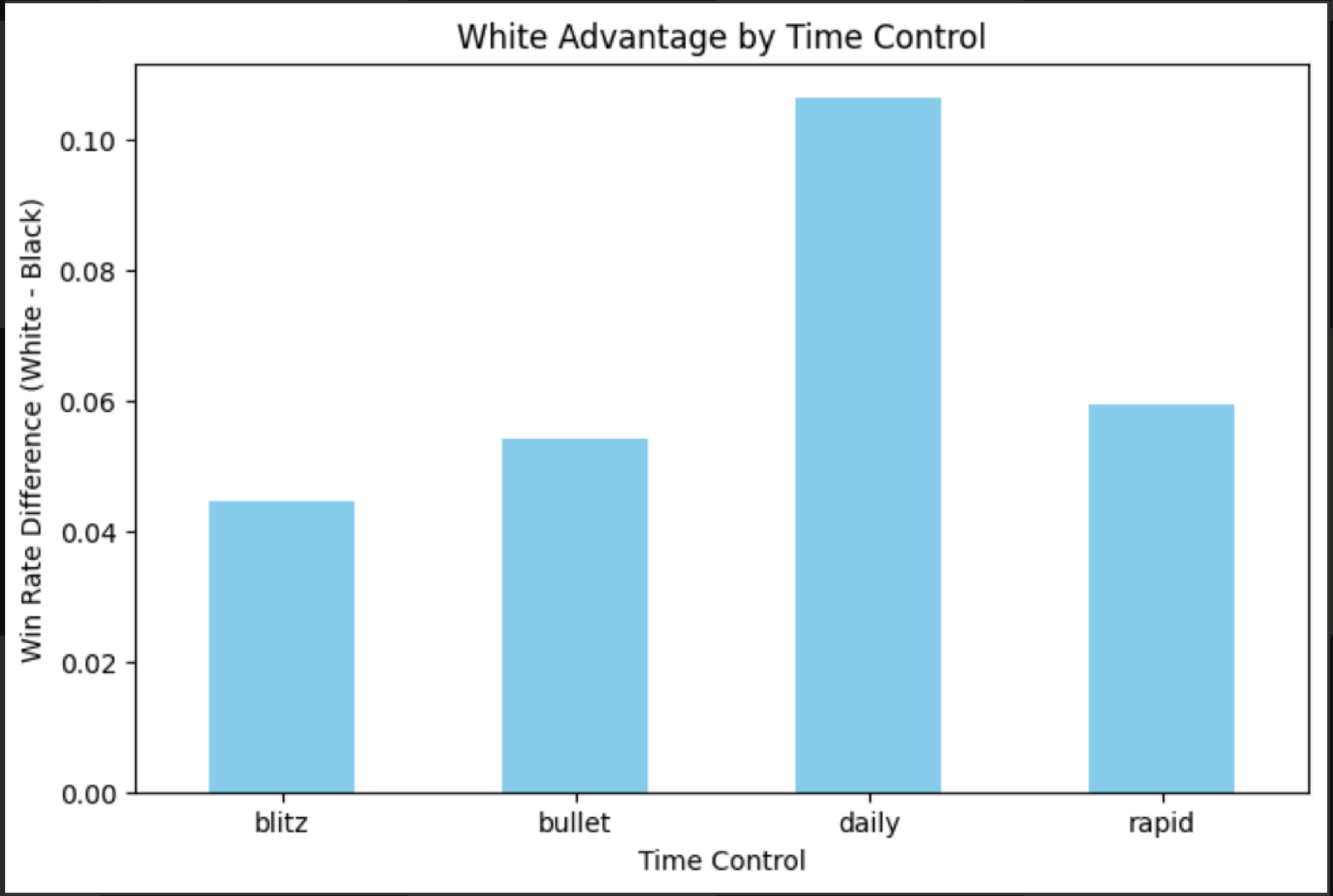
```
Original dataset:
      username  color      result time_class  win
0  alexcolovic      1         win      blitz    1
1    Coach_Val      0  checkmated      blitz    0
2     swayamsm      1    resigned      rapid    0
3  alexcolovic      0         win      rapid    1
4  alexcolovic      1         win      rapid    1

Cleaned dataset:
      username  color      result time_class  win
0  alexcolovic      0         win      blitz    1
1    Coach_Val      0  checkmated      blitz    0
2     swayamsm      0    resigned      rapid    0
3  alexcolovic      0         win      rapid    1
4  alexcolovic      0         win      rapid    1
```

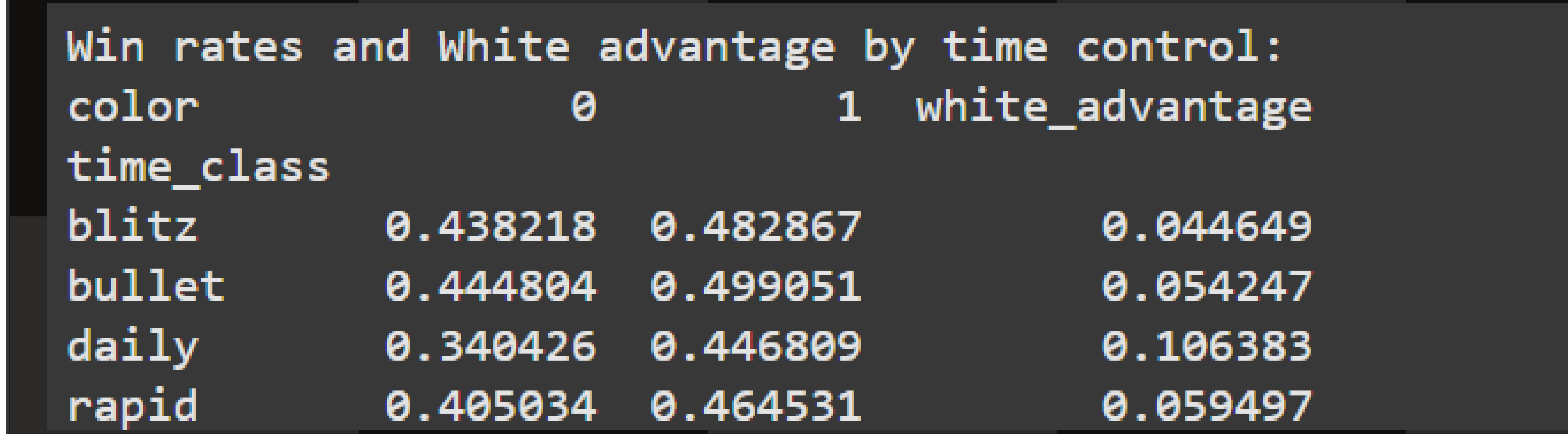


```
Cleaned dataset shape: (111830, 5)
```

WIN RATES

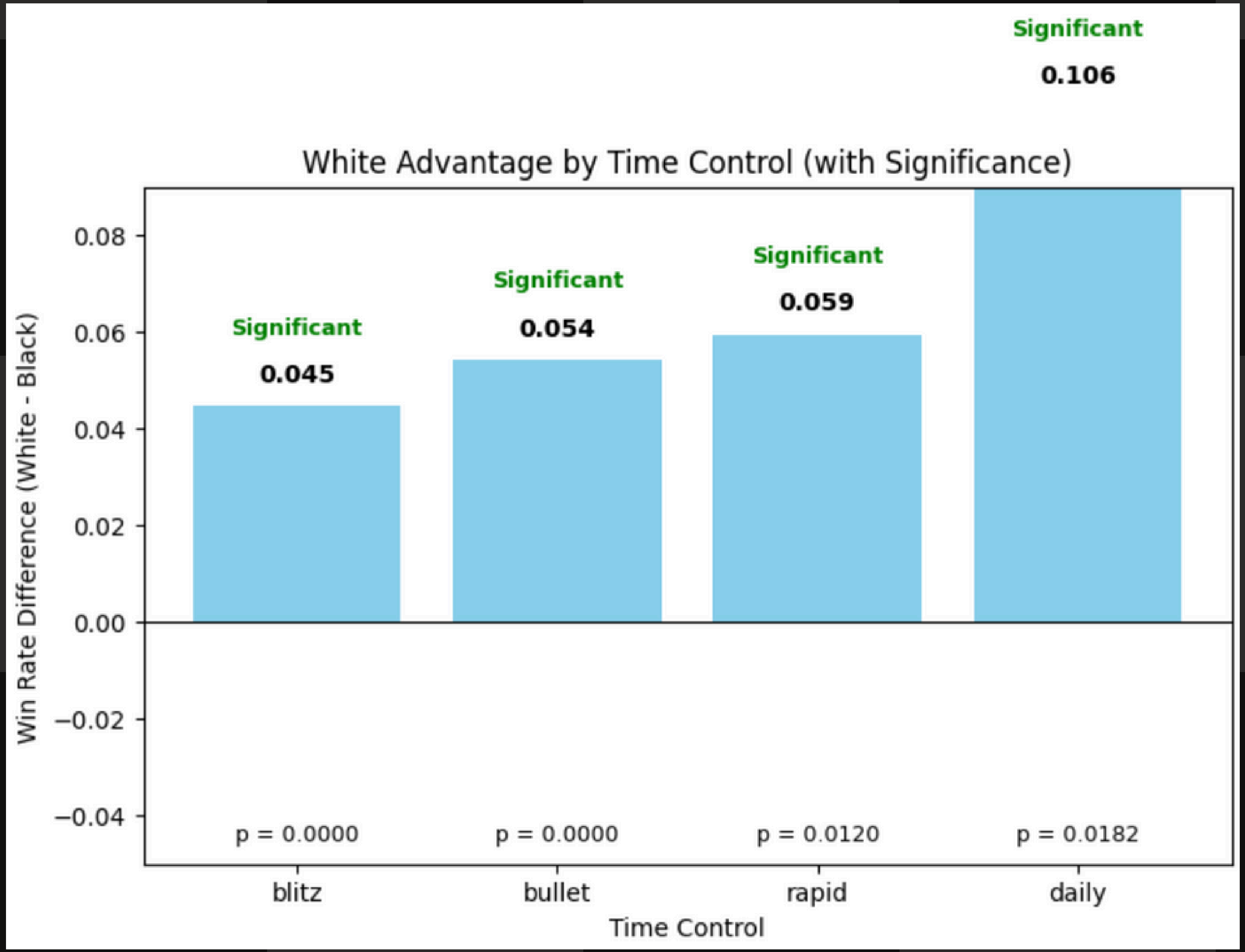
White Vs Black

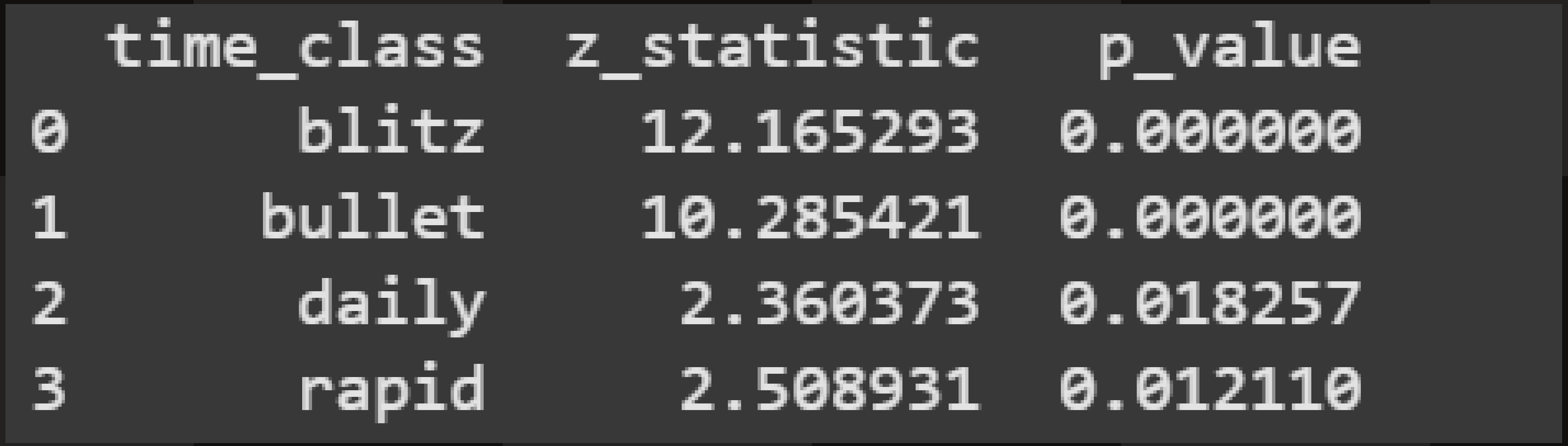




```
Win rates and White advantage by time control:
color              0         1  white_advantage
time_class
blitz       0.438218  0.482867         0.044649
bullet      0.444804  0.499051         0.054247
daily       0.340426  0.446809         0.106383
rapid       0.405034  0.464531         0.059497
```

WIN RATES


Significant
0.106
White Advantage by Time Control (with Significance)
Win Rate Difference (White - Black)
0.08
0.06
0.04
0.02
0.00
−0.02
−0.04
Significant
0.045
Significant
0.054
Significant
0.059
p = 0.0000
p = 0.0000
p = 0.0120
p = 0.0182
blitz
bullet
rapid
daily
Time Control



```
  time_class  z_statistic   p_value
0      blitz    12.165293  0.000000
1     bullet    10.285421  0.000000
2      daily     2.360373  0.018257
3      rapid     2.508931  0.012110
```

Based on the results of the two-proportion Z-test, there is a statistically significant difference between the win rates of White and Black players across most time controls. For all formats — blitz, bullet, rapid, and daily — the p-values were below the 0.05 significance level, leading us to reject the null hypothesis (H_0) that both colors have equal chances of winning.

02

ARE THERE DIFFERENCES IN DRAW RATES FOR BLITZ AND BULLET GAMES?

Null Hypothesis: There is no correlation in players having different draw rate in Blitz games and in Bullet games

Alternative Hypothesis: Players have a higher draw rate in Blitz games than in Bullet games

Alternative Hypothesis: Players have a lower draw rate in Blitz games than in Bullet games

IMPORTS - GETTING THE DATA

```
leaderboard_keys = ['live_blitz', 'live_bullet', 'daily_blitz', 'daily_bullet']
for key in leaderboard_keys:
 try:
 lb_users = get_leaderboard_users(key)
 usernames.update(lb_users)
 print(f"Added {len(lb_users)} usernames from leaderboard '{key}'")
 except Exception as e:
 print(f"Failed to get leaderboard '{key}': {e}")
titles = ['GM', 'IM', 'FM', 'WGM', 'WIM', 'WFM']
```

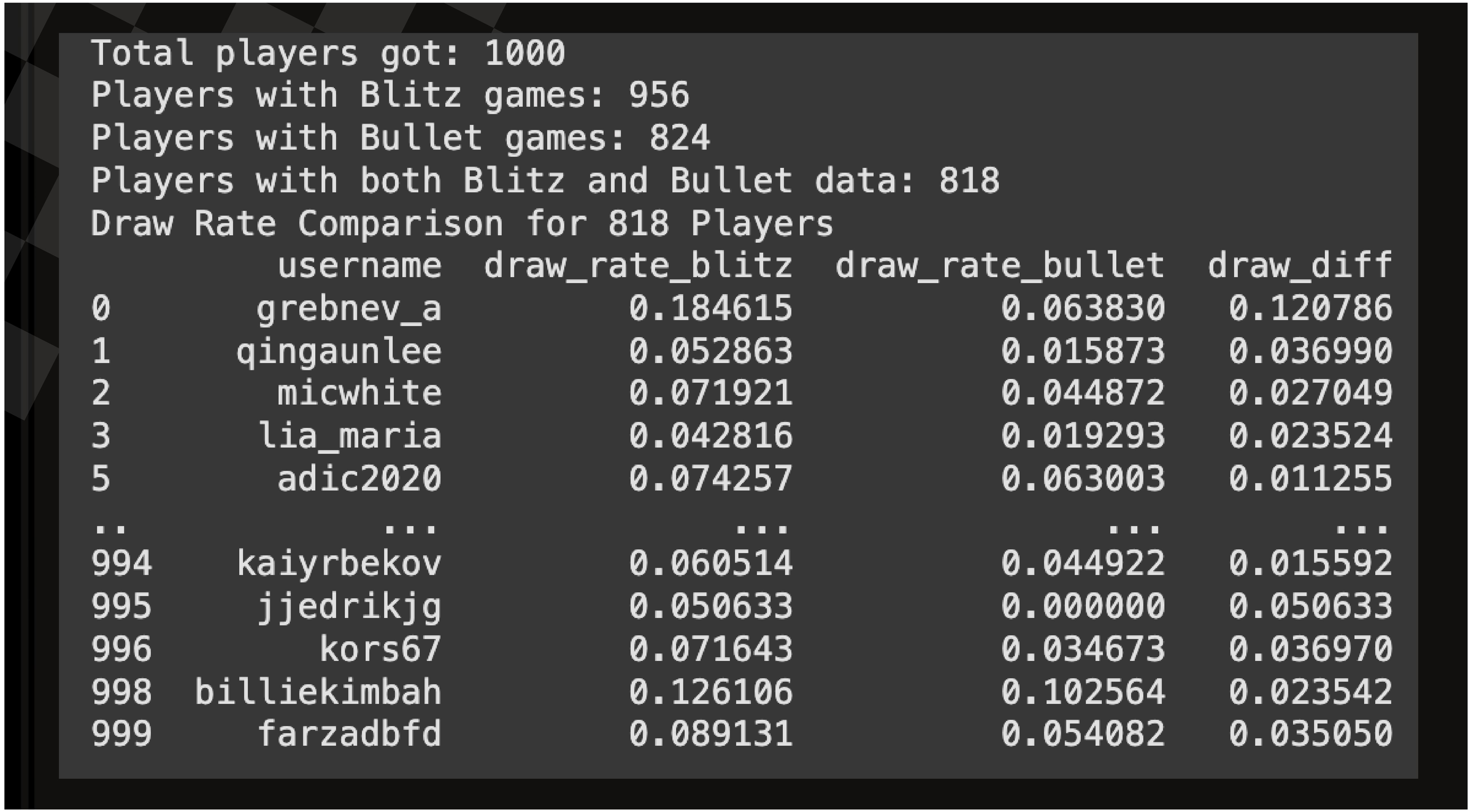
CODE SNIPPETS AND CALCULATION

```
blitz = player_stats.get("chess_blitz", {}).get("record", {})
bullet = player_stats.get("chess_bullet", {}).get("record", {})

blitz_total = blitz.get("win", 0) + blitz.get("loss", 0) + blitz.get("draw", 0)
bullet_total = bullet.get("win", 0) + bullet.get("loss", 0) + bullet.get("draw", 0)

blitz_draws = blitz.get("draw", 0)
bullet_draws = bullet.get("draw", 0)
```

CODE SNIPPETS AND CALCULATION

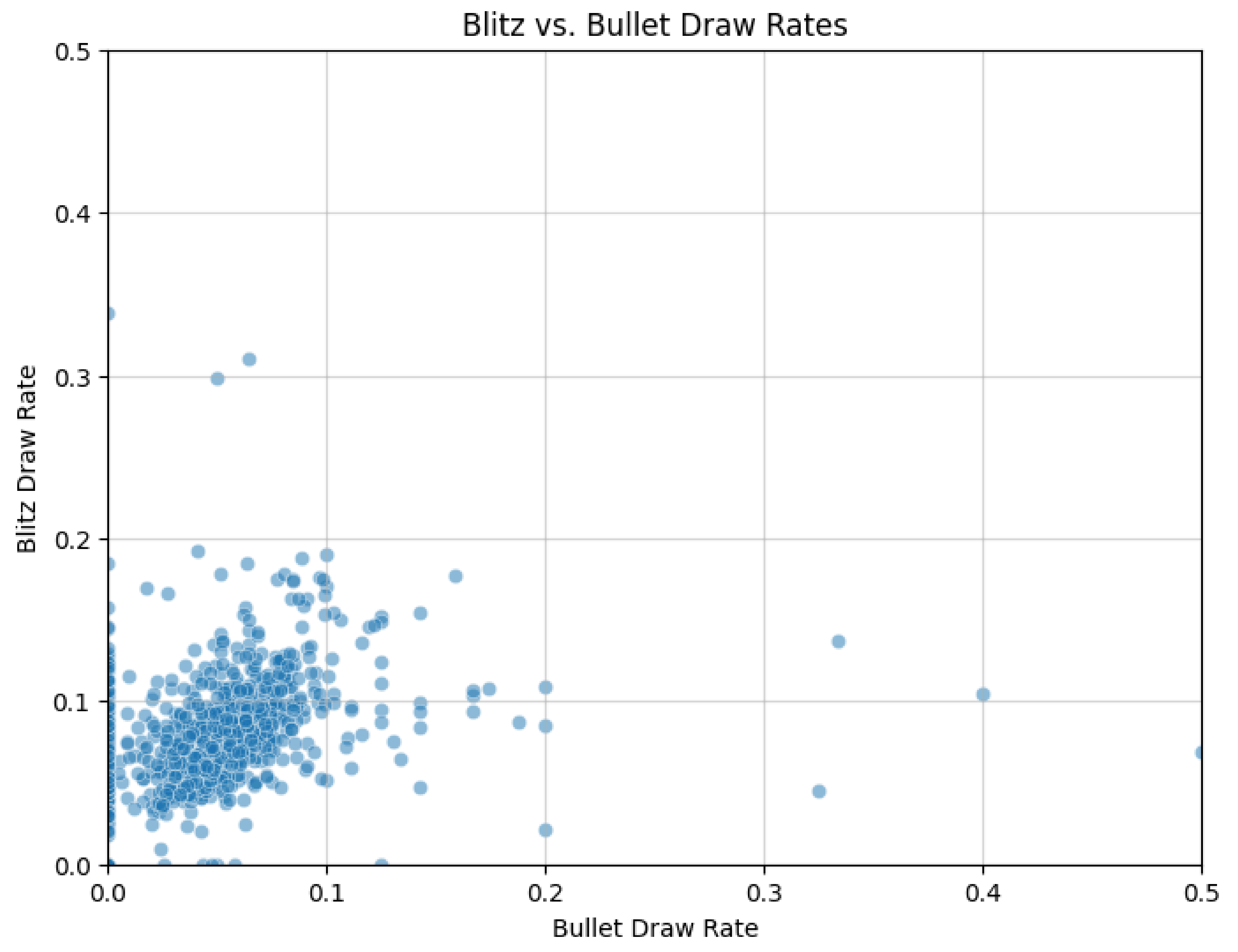
```
df["draw_rate_blitz"] = np.where(
 df["blitz_games_played"] > 0,
 df["blitz_draws"] / df["blitz_games_played"],
 np.nan,
)
df["draw_rate_bullet"] = np.where(
 df["bullet_games_played"] > 0,
 df["bullet_draws"] / df["bullet_games_played"],
 np.nan,
)
df["draw_diff"] = df["draw_rate_blitz"] - df["draw_rate_bullet"]
```

RESULTS

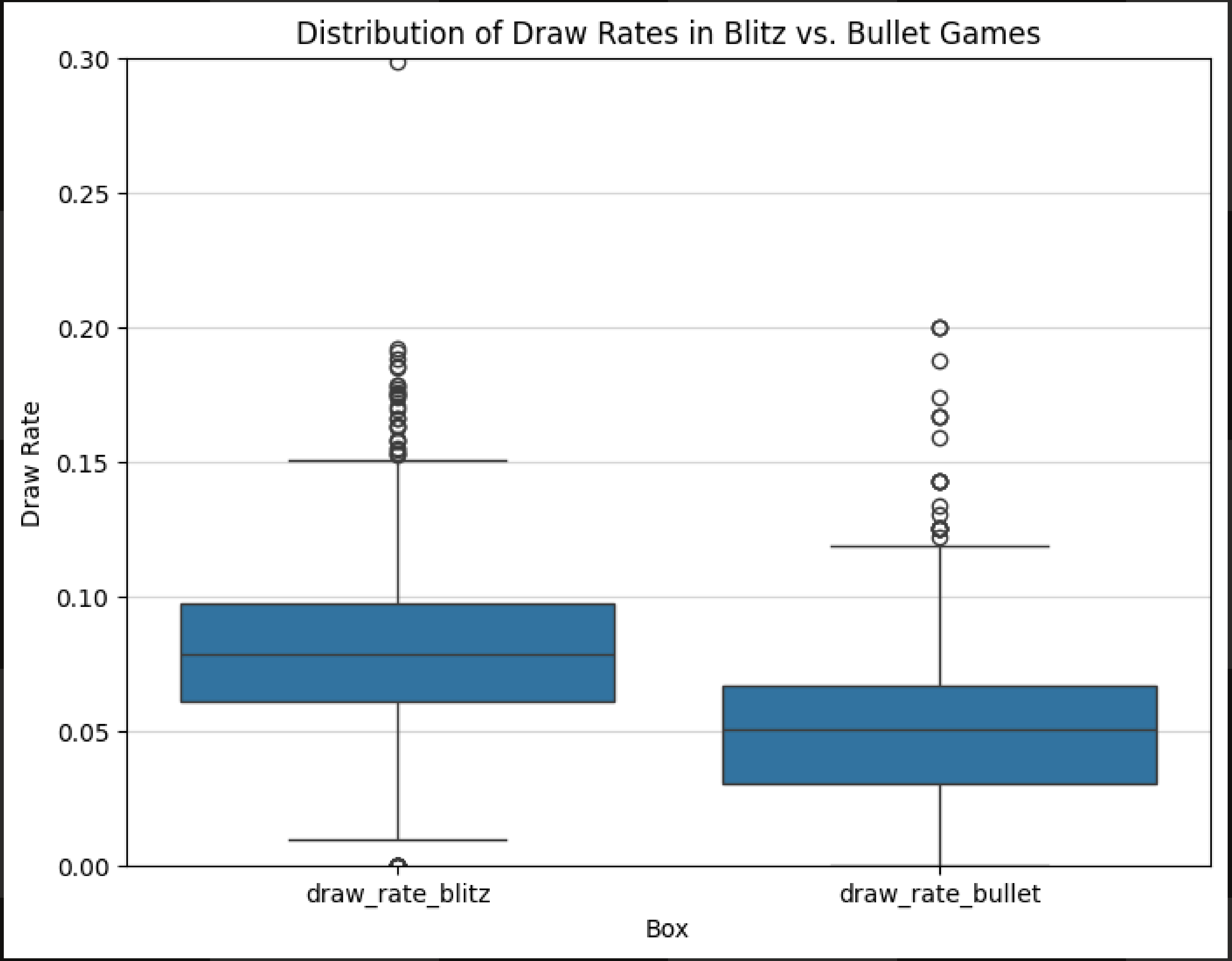


```
Total players got: 1000
Players with Blitz games: 956
Players with Bullet games: 824
Players with both Blitz and Bullet data: 818
Draw Rate Comparison for 818 Players
         username  draw_rate_blitz  draw_rate_bullet  draw_diff
0       grebnev_a         0.184615          0.063830   0.120786
1      qingaunlee         0.052863          0.015873   0.036990
2        micwhite         0.071921          0.044872   0.027049
3       lia_maria         0.042816          0.019293   0.023524
5        adic2020         0.074257          0.063003   0.011255
..            ...              ...               ...        ...
994    kaiyrbekov         0.060514          0.044922   0.015592
995      jjedrikjg        0.050633          0.000000   0.050633
996        kors67         0.071643          0.034673   0.036970
998  billiekimbah         0.126106          0.102564   0.023542
999     farzadbfd         0.089131          0.054082   0.035050
```

DATA VISUALISATION



DATA VISUALISATION



RESULTS

- T-test: 12.3649
- P-value: 0.0000000
- Reject the null hypothesis
- → Players have a HIGHER draw rate in Blitz games than in Bullet games.

CONCLUSION

THE FIRST NULL HYPOTHESIS STATED THAT THE EFFECT OF PLAYING AS WHITE ON WIN PROBABILITY DOES NOT VARY BY TIME CONTROL.HOWEVER, THE RESULTS OF THE TWO-PROPORTION Z-TEST REVEALED STATISTICALLY SIGNIFICANT DIFFERENCES BETWEEN THE WIN RATES OF WHITE AND BLACK PLAYERS ACROSS ALL TIME CONTROLS — BLITZ, BULLET, RAPID, AND DAILY. SINCE THE P-VALUES FOR ALL FORMATS WERE BELOW THE 0.05 SIGNIFICANCE LEVEL, WE REJECT THE NULL HYPOTHESIS (H_0) AND CONCLUDE THAT PLAYING AS WHITE HAS A DIFFERENT, AND GENERALLY HIGHER, IMPACT ON WIN PROBABILITY DEPENDING ON THE TIME CONTROL.

THE SECOND NULL HYPOTHESIS PROPOSED THAT THERE IS NO DIFFERENCE IN DRAW RATES BETWEEN BLITZ AND BULLET GAMES. THE T-TEST PRODUCED A STATISTIC OF 12.3649 WITH A P-VALUE OF 0.0000000, WHICH IS FAR BELOW THE 0.05 THRESHOLD. THEREFORE, WE ALSO REJECT THE NULL HYPOTHESIS AND CONCLUDE THAT THERE IS A SIGNIFICANT DIFFERENCE IN DRAW RATES BETWEEN BLITZ AND BULLET FORMATS.

REFERENCES

- TAN, S. (2024, JUNE 5). GETTING CHESS.COM GAME DATA WITH POSTMAN, USE PYTHON AND ANALYZE. MEDIUM. HTTPS://MEDIUM.COM/@STANLEYTAN_4595/GETTING-CHESS-COM-GAME-DATA-WITH-POSTMAN-USE-PYTHON-AND-ANALYZE-48891A502483

- SHEDLOAD OF CODE. (2021, JULY 10). HOW TO SCRAPE AND ANALYSE YOUR CHESS.COM DATA. SHEDLOAD OF CODE. HTTPS://WWW.SHEDLOADOFCODE.COM/BLOG/HOW-TO-SCRAPE-AND-ANALYSE-YOUR-CHESS-COM-DATA

- CHESS.COM. CHESS API. HTTPS://WWW.CHESS.COM/NEWS/VIEW/PUBLISHED-DATA-API

- BURNS, B. D. (2004). THE EFFECTS OF SPEED ON SKILLED CHESS PERFORMANCE. PSYCHOLOGICAL SCIENCE, 15(7), 442–447. HTTPS://DOI.ORG/10.1111/J.0956-7976.2004.00699.X