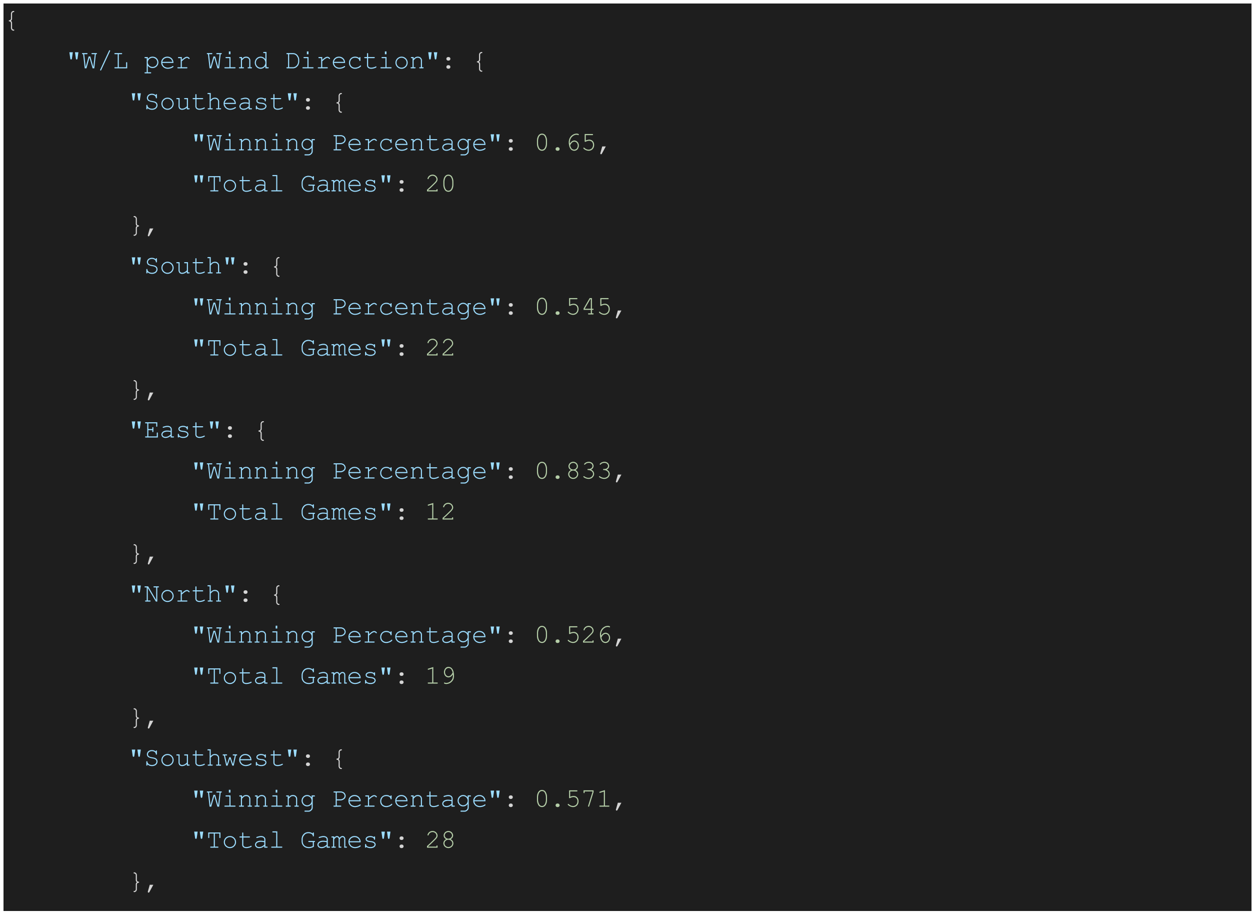
Goals: We wanted To find out if wind speed or temperature affects the amount of runs and winning percentage at Mets home games.
What we planned on using: https://weatherstack.com/ and https://www.mlb.com/mets/schedule/2022-05/list.

Achieved Goals: We compared temperature, max wind speed, precipitation, wind direction, and weather code during each baseball game to the mets runs, opponents runs, total runs, and whether or not the mets won the game.
What we actually used: https://open-meteo.com/ and https://www.baseball-reference.com/teams/NYM/2021-schedule-scores.shtml.

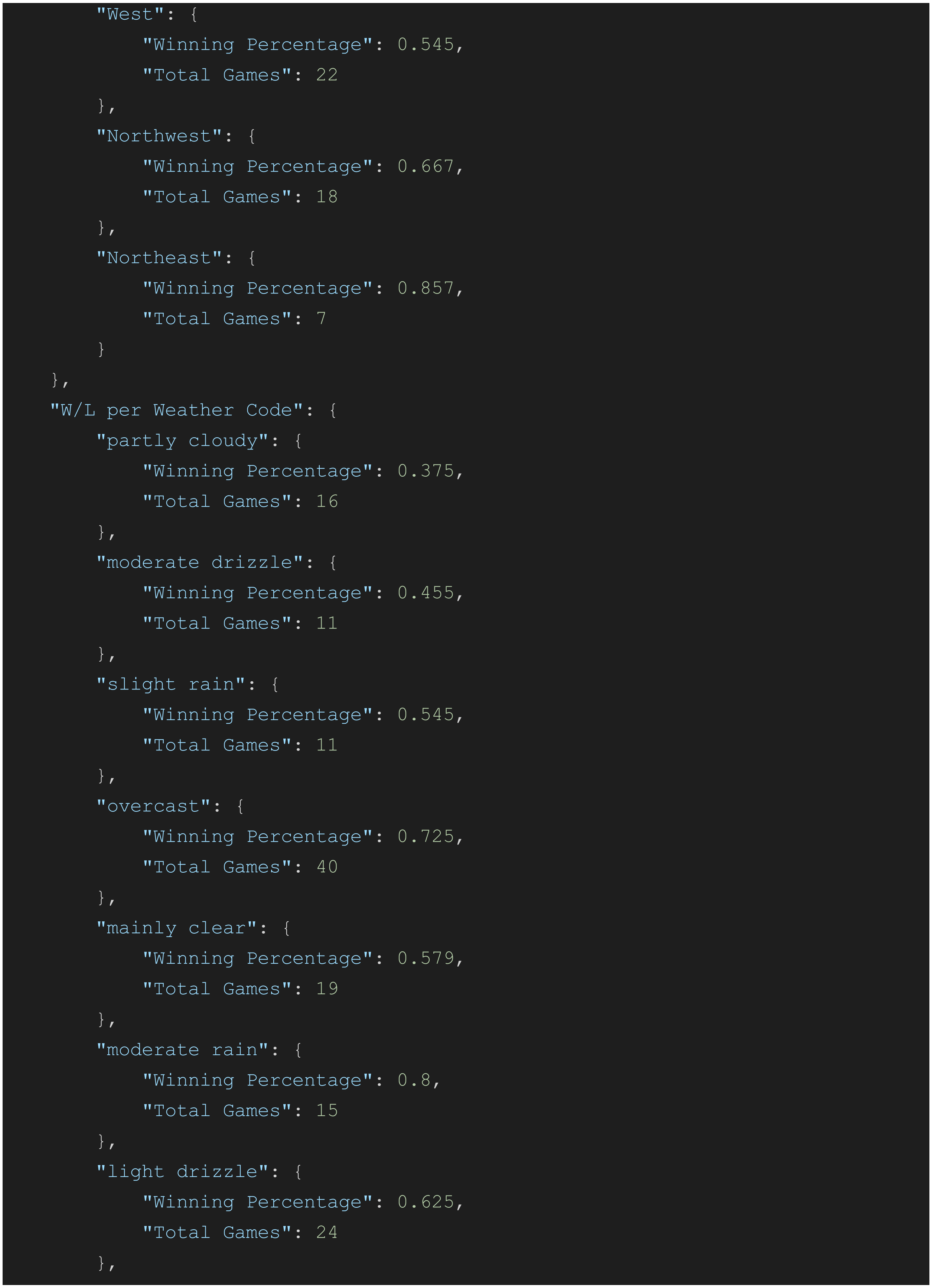
Problems:

- trying to implement a system for getting less than 25 items each run, while still grabbing all necessary information based on information from other api.
- Joining more than two tables at a time
- Combining code from each website together into one data collection python file
- Updating databases that were built during older runs of our program having partially missing information during testing.
- Making readable code

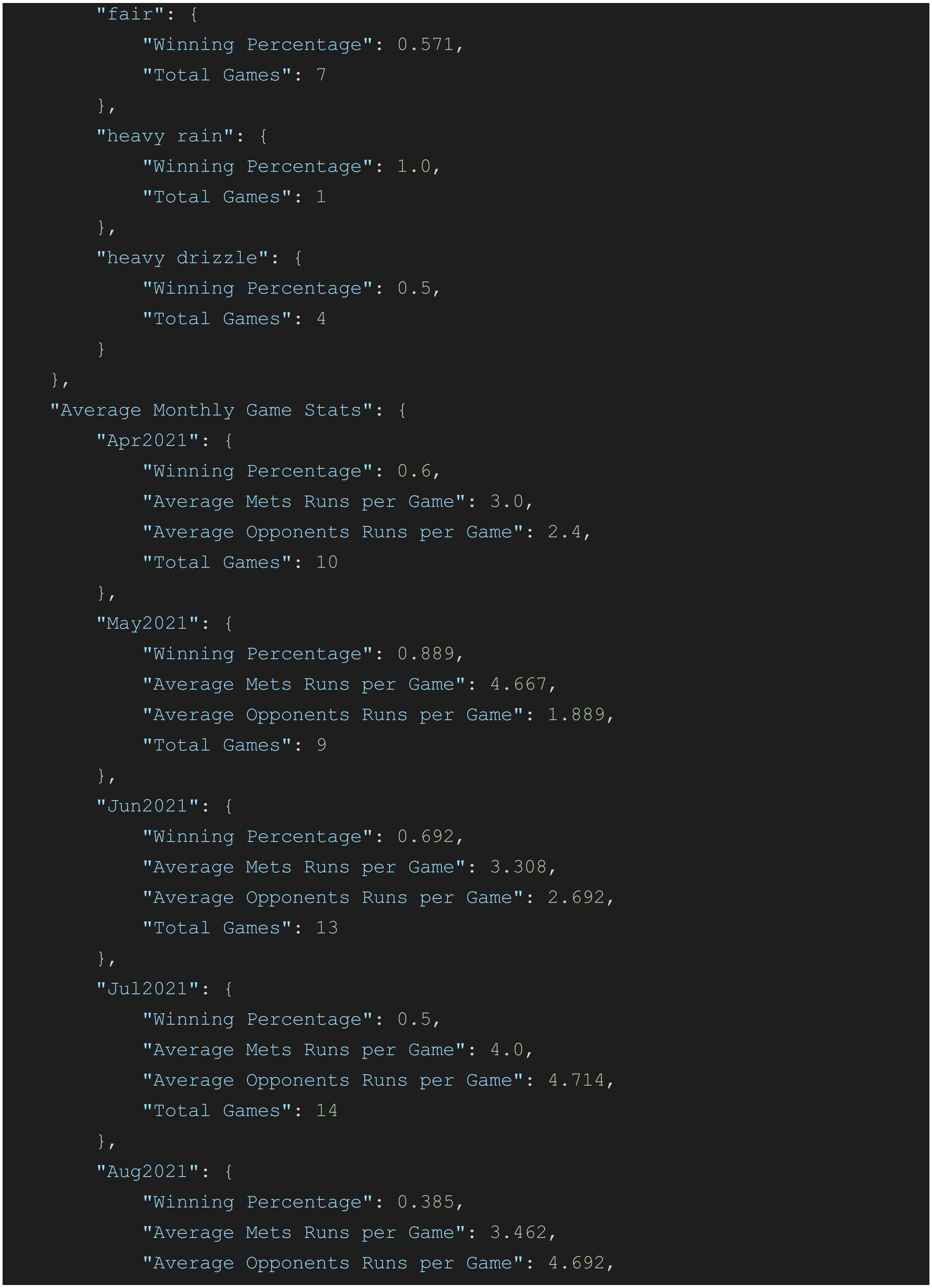
Calculations: Not sure if they want calculations or screenshot of json file?



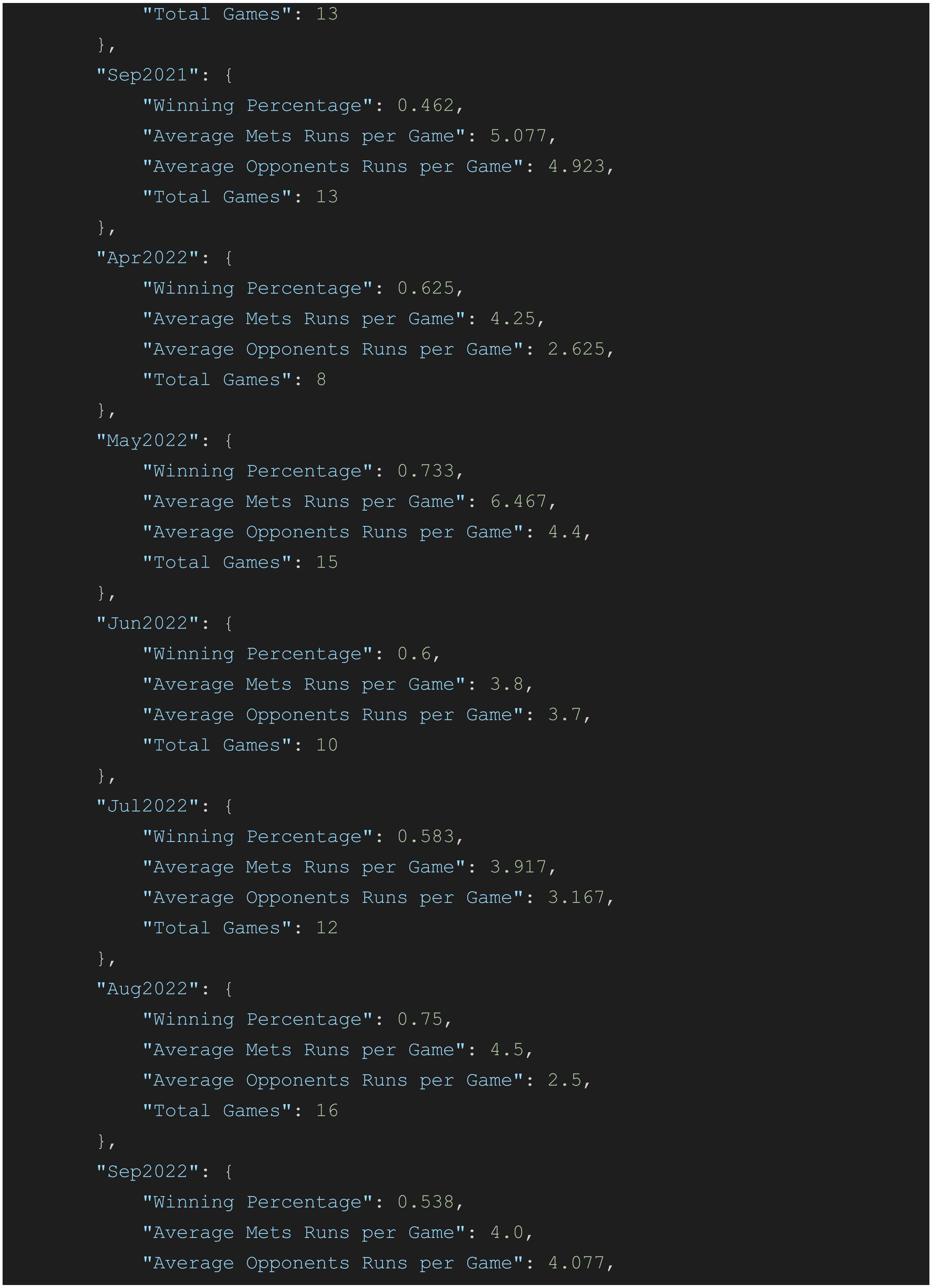
```
{
    "W/L per Wind Direction": {
        "Southeast": {
            "Winning Percentage": 0.65,
            "Total Games": 20
        },
        "South": {
            "Winning Percentage": 0.545,
            "Total Games": 22
        },
        "East": {
            "Winning Percentage": 0.833,
            "Total Games": 12
        },
        "North": {
            "Winning Percentage": 0.526,
            "Total Games": 19
        },
        "Southwest": {
            "Winning Percentage": 0.571,
            "Total Games": 28
        },
```



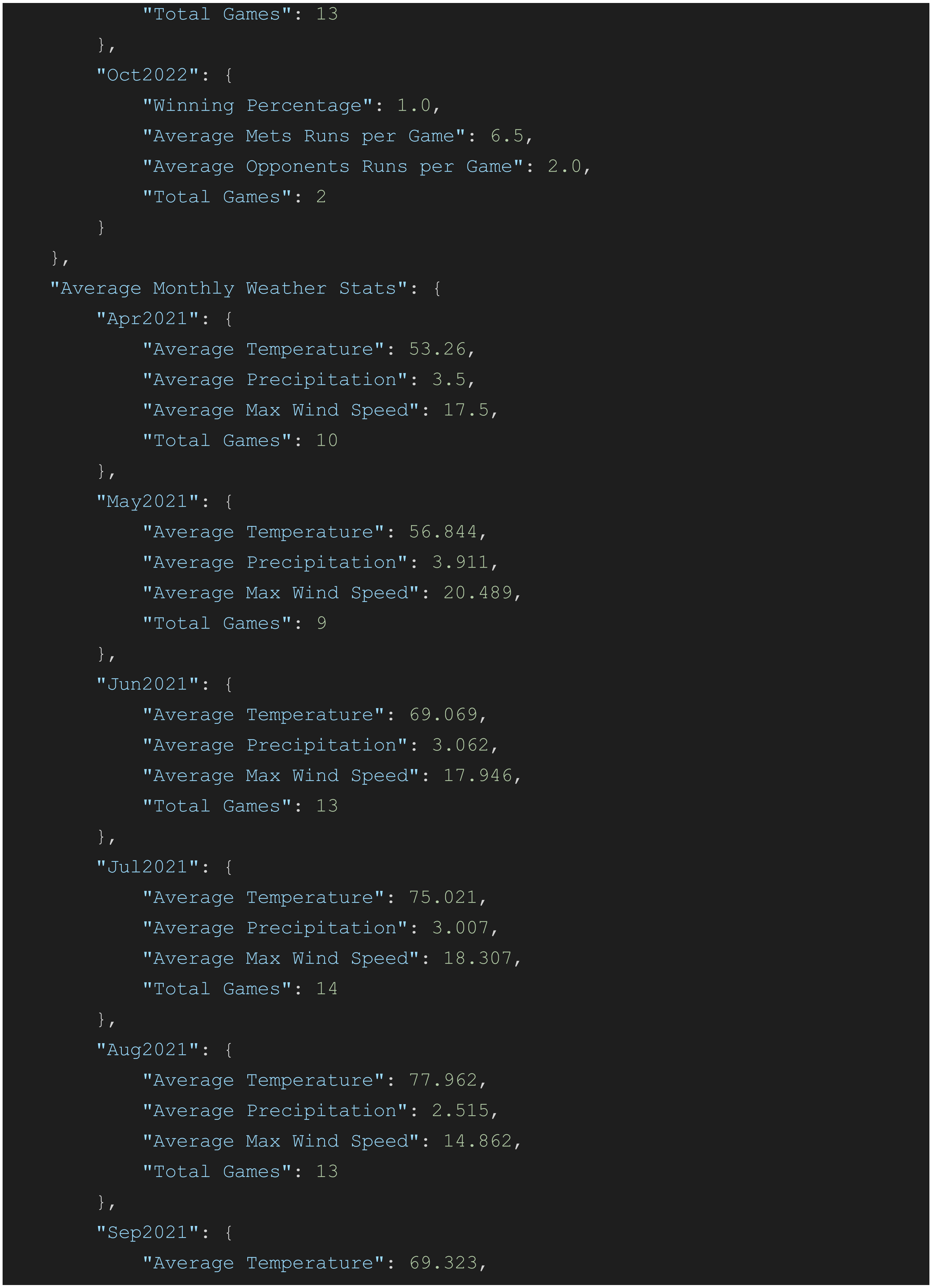
```
        "West": {
            "Winning Percentage": 0.545,
            "Total Games": 22
        },
        "Northwest": {
            "Winning Percentage": 0.667,
            "Total Games": 18
        },
        "Northeast": {
            "Winning Percentage": 0.857,
            "Total Games": 7
        }
    },
    "W/L per Weather Code": {
        "partly cloudy": {
            "Winning Percentage": 0.375,
            "Total Games": 16
        },
        "moderate drizzle": {
            "Winning Percentage": 0.455,
            "Total Games": 11
        },
        "slight rain": {
            "Winning Percentage": 0.545,
            "Total Games": 11
        },
        "overcast": {
            "Winning Percentage": 0.725,
            "Total Games": 40
        },
        "mainly clear": {
            "Winning Percentage": 0.579,
            "Total Games": 19
        },
        "moderate rain": {
            "Winning Percentage": 0.8,
            "Total Games": 15
        },
        "light drizzle": {
            "Winning Percentage": 0.625,
            "Total Games": 24
        },
```



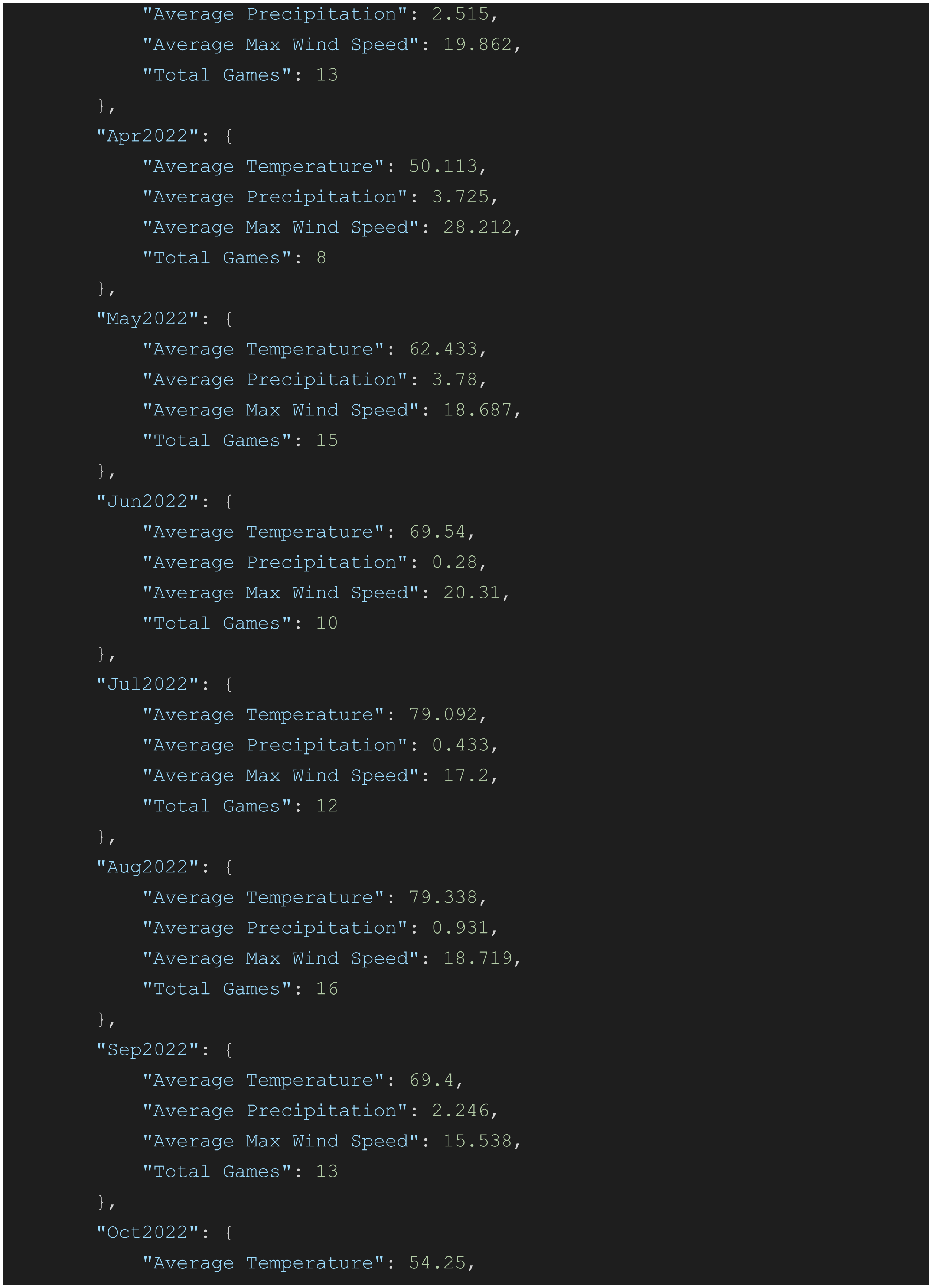
```
        "fair": {
            "Winning Percentage": 0.571,
            "Total Games": 7
        },
        "heavy rain": {
            "Winning Percentage": 1.0,
            "Total Games": 1
        },
        "heavy drizzle": {
            "Winning Percentage": 0.5,
            "Total Games": 4
        }
    },
    "Average Monthly Game Stats": {
        "Apr2021": {
            "Winning Percentage": 0.6,
            "Average Mets Runs per Game": 3.0,
            "Average Opponents Runs per Game": 2.4,
            "Total Games": 10
        },
        "May2021": {
            "Winning Percentage": 0.889,
            "Average Mets Runs per Game": 4.667,
            "Average Opponents Runs per Game": 1.889,
            "Total Games": 9
        },
        "Jun2021": {
            "Winning Percentage": 0.692,
            "Average Mets Runs per Game": 3.308,
            "Average Opponents Runs per Game": 2.692,
            "Total Games": 13
        },
        "Jul2021": {
            "Winning Percentage": 0.5,
            "Average Mets Runs per Game": 4.0,
            "Average Opponents Runs per Game": 4.714,
            "Total Games": 14
        },
        "Aug2021": {
            "Winning Percentage": 0.385,
            "Average Mets Runs per Game": 3.462,
            "Average Opponents Runs per Game": 4.692,
```



```
            "Total Games": 13
        },
        "Sep2021": {
            "Winning Percentage": 0.462,
            "Average Mets Runs per Game": 5.077,
            "Average Opponents Runs per Game": 4.923,
            "Total Games": 13
        },
        "Apr2022": {
            "Winning Percentage": 0.625,
            "Average Mets Runs per Game": 4.25,
            "Average Opponents Runs per Game": 2.625,
            "Total Games": 8
        },
        "May2022": {
            "Winning Percentage": 0.733,
            "Average Mets Runs per Game": 6.467,
            "Average Opponents Runs per Game": 4.4,
            "Total Games": 15
        },
        "Jun2022": {
            "Winning Percentage": 0.6,
            "Average Mets Runs per Game": 3.8,
            "Average Opponents Runs per Game": 3.7,
            "Total Games": 10
        },
        "Jul2022": {
            "Winning Percentage": 0.583,
            "Average Mets Runs per Game": 3.917,
            "Average Opponents Runs per Game": 3.167,
            "Total Games": 12
        },
        "Aug2022": {
            "Winning Percentage": 0.75,
            "Average Mets Runs per Game": 4.5,
            "Average Opponents Runs per Game": 2.5,
            "Total Games": 16
        },
        "Sep2022": {
            "Winning Percentage": 0.538,
            "Average Mets Runs per Game": 4.0,
            "Average Opponents Runs per Game": 4.077,
```

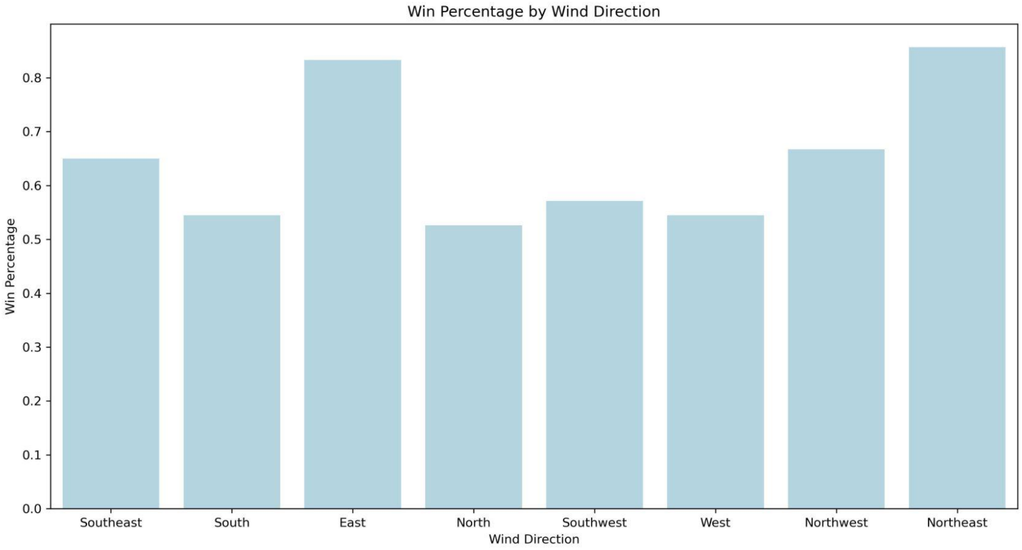


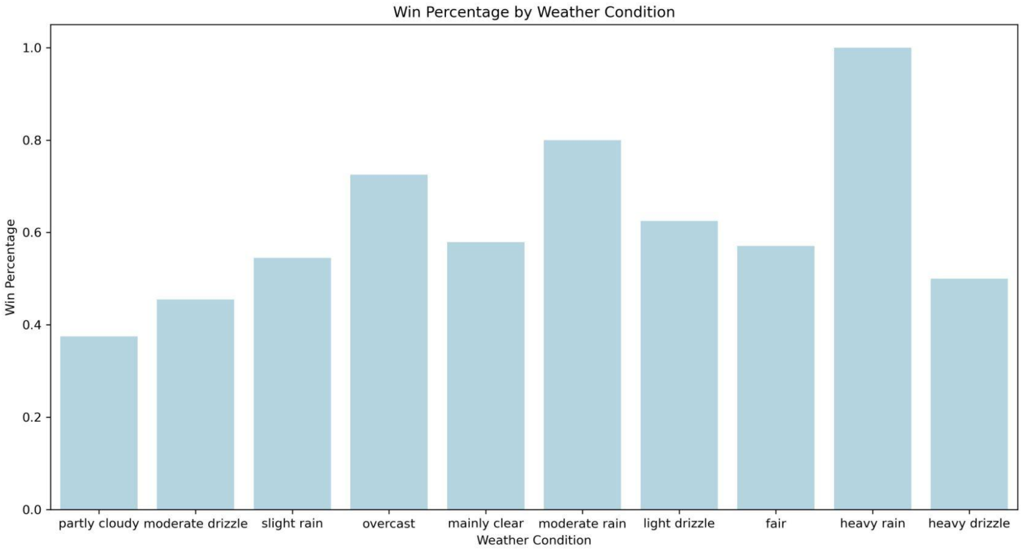
```
            "Total Games": 13
        },
        "Oct2022": {
            "Winning Percentage": 1.0,
            "Average Mets Runs per Game": 6.5,
            "Average Opponents Runs per Game": 2.0,
            "Total Games": 2
        }
    },
    "Average Monthly Weather Stats": {
        "Apr2021": {
            "Average Temperature": 53.26,
            "Average Precipitation": 3.5,
            "Average Max Wind Speed": 17.5,
            "Total Games": 10
        },
        "May2021": {
            "Average Temperature": 56.844,
            "Average Precipitation": 3.911,
            "Average Max Wind Speed": 20.489,
            "Total Games": 9
        },
        "Jun2021": {
            "Average Temperature": 69.069,
            "Average Precipitation": 3.062,
            "Average Max Wind Speed": 17.946,
            "Total Games": 13
        },
        "Jul2021": {
            "Average Temperature": 75.021,
            "Average Precipitation": 3.007,
            "Average Max Wind Speed": 18.307,
            "Total Games": 14
        },
        "Aug2021": {
            "Average Temperature": 77.962,
            "Average Precipitation": 2.515,
            "Average Max Wind Speed": 14.862,
            "Total Games": 13
        },
        "Sep2021": {
            "Average Temperature": 69.323,
```

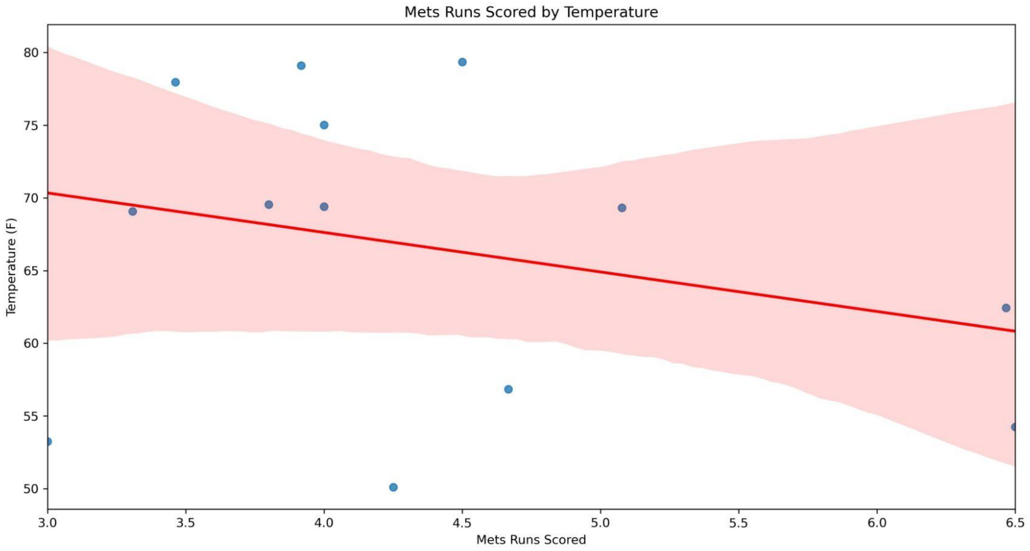


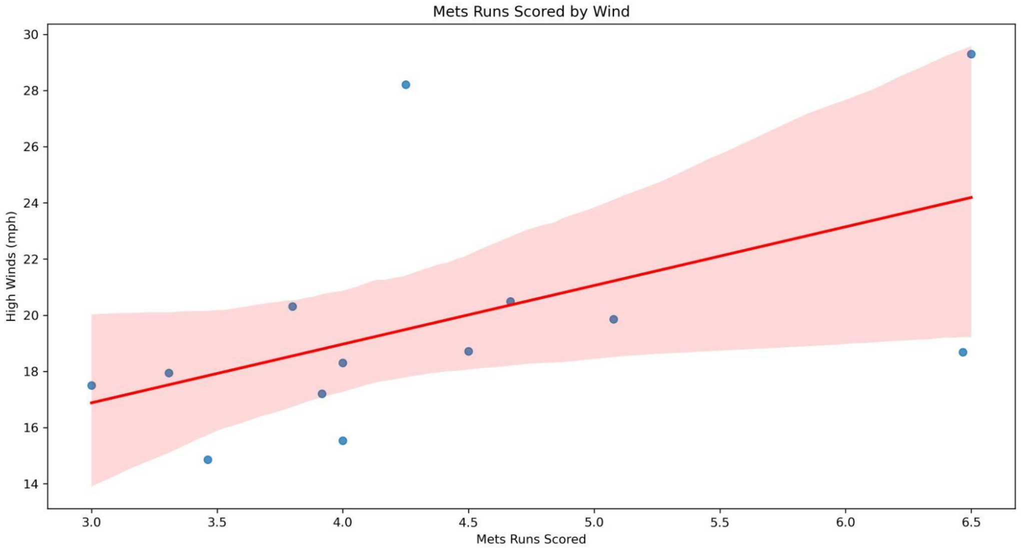
```
            "Average Precipitation": 2.515,
            "Average Max Wind Speed": 19.862,
            "Total Games": 13
        },
        "Apr2022": {
            "Average Temperature": 50.113,
            "Average Precipitation": 3.725,
            "Average Max Wind Speed": 28.212,
            "Total Games": 8
        },
        "May2022": {
            "Average Temperature": 62.433,
            "Average Precipitation": 3.78,
            "Average Max Wind Speed": 18.687,
            "Total Games": 15
        },
        "Jun2022": {
            "Average Temperature": 69.54,
            "Average Precipitation": 0.28,
            "Average Max Wind Speed": 20.31,
            "Total Games": 10
        },
        "Jul2022": {
            "Average Temperature": 79.092,
            "Average Precipitation": 0.433,
            "Average Max Wind Speed": 17.2,
            "Total Games": 12
        },
        "Aug2022": {
            "Average Temperature": 79.338,
            "Average Precipitation": 0.931,
            "Average Max Wind Speed": 18.719,
            "Total Games": 16
        },
        "Sep2022": {
            "Average Temperature": 69.4,
            "Average Precipitation": 2.246,
            "Average Max Wind Speed": 15.538,
            "Total Games": 13
        },
        "Oct2022": {
            "Average Temperature": 54.25,
            "Average Precipitation": 19.85,
            "Average Max Wind Speed": 29.3,
            "Total Games": 2
        }
    },
    "Average Mets Runs by Weather Code": {
        "partly cloudy": {
            "Average Mets Runs": 4.125,
            "Total Games": 16
        },
        "moderate drizzle": {
            "Average Mets Runs": 3.636,
            "Total Games": 11
        },
        "slight rain": {
            "Average Mets Runs": 3.909,
            "Total Games": 11
        },
        "overcast": {
            "Average Mets Runs": 4.625,
            "Total Games": 40
        },
        "mainly clear": {
            "Average Mets Runs": 3.842,
            "Total Games": 19
        },
        "moderate rain": {
            "Average Mets Runs": 5.533,
            "Total Games": 15
        },
        "light drizzle": {
            "Average Mets Runs": 4.292,
            "Total Games": 24
        },
        "fair": {
            "Average Mets Runs": 2.714,
            "Total Games": 7
        },
        "heavy rain": {
            "Average Mets Runs": 4.0,
            "Total Games": 1
        },
        "heavy drizzle": {
            "Average Mets Runs": 4.75,
            "Total Games": 4
        }
    },
    "Average Mets Runs by Wind Direction": {
        "Southeast": {
            "Average Mets Runs": 3.8,
            "Total Games": 20
        },
        "South": {
            "Average Mets Runs": 3.864,
            "Total Games": 22
        },
        "East": {
            "Average Mets Runs": 4.333,
            "Total Games": 12
        },
        "North": {
            "Average Mets Runs": 3.737,
            "Total Games": 19
        },
        "Southwest": {
            "Average Mets Runs": 4.929,
            "Total Games": 28
        },
        "West": {
            "Average Mets Runs": 3.909,
            "Total Games": 22
        },
        "Northwest": {
            "Average Mets Runs": 4.556,
            "Total Games": 18
        },
        "Northeast": {
            "Average Mets Runs": 6.429,
            "Total Games": 7
        }
    }
}
```

Visuals:


Win Percentage by Wind Direction
Win Percentage
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
Southeast
South
East
North
Southwest
West
Northwest
Northeast
Wind Direction


Win Percentage by Weather Condition
Win Percentage
0.0
0.2
0.4
0.6
0.8
1.0
partly cloudy
moderate drizzle
slight rain
overcast
mainly clear
moderate rain
light drizzle
fair
heavy rain
heavy drizzle
Weather Condition





Instructions for running code:

1. Set the text in count.txt to -1 or remove count.txt.
2. Run baseball_weather_auto.py 27 times to collect all data from both apis.
3. Run calculations.py to get some calculations
4. Run visualizations.py
5. Open data.json to view data calculations. Open the jpgs to view the graphs.

6. Great success

Code documentation:

Baseball_weather_auto.py

1. database_setup(db_name)
   This function takes in a string of the database name (db_name), and outputs the database cursor and connector path.

2. countmake()
   This function keeps track of the amount of times the program has been run, and returns that number.

3. countIncrement()
   This function increments the amount of times the program has been run.

4. soup_database(year, month_abrev, month_id, cur, conn)
   This function takes in
   year: a string containing the year
   month_abrev: a string containing the abbreviation for the month (Apr = April, May = May, etc)
   month_id: a string containing the number of the month (“04” = April, “05” = May, etc)
   And parses through the html on the website to find the table of all the games the Mets have played. It identifies the home games in the table and then extracts date, Mets score, opposing score, and W/L and writes that as a table into the database.

5. weatherApi(month, year, monthyear, cur, conn)
   This function takes in
   month: an int containing the month (4 = April, 5 = May , etc)
   year: an int containing the year
   monthyear: a list of dates that the Mets had home games in that month and year.
   Cur: database cursor
   Conn: database connector
   And uses the monthyear list to get the dates that need to be collected for the database, then Inserts these dates into the database under the Weather table along with weather codes, temperature, max wind speed, and wind direction.

6. wmodata(cur, conn)
   This function inserts the weather codes into the database under the Weather_Codes table using the database cursor (cur) and connector (conn)

7. direction(cur, conn)
   This function inserts the wind directions into the database under the Weather_Directions table using the database cursor (cur) and connector (conn)

8. get_dates(month, year, cur)
   By using

   month: an int containing the month (4 = April, 5 = May, etc)
   year: an int containing the year
   Cur: database cursor

   This function retrieves all dates in the month and year imputed that are in the Baseball table in the given database. will not work if baseball data for range is not in the database first.

Calculations.py

1. database_setup(db_name)
   Same as in baseball_weather_auto

2. json_file(filename, finaldic)
   This function takes in

   filename: string containing the json file name
   finaldic: a dictionary containing each of the calculations

   And writes the finaldic to the json file

3. dates_by_month_full_comb(cur)

Creates a dictionary that is organized by monthyear (ex: Apr2021) that contains inner keys of dates that contain all of the information about the weather and baseball game during that day.
INPUT: cur = ... (cursor to database)
OUTPUT: months: dictionary described above
ex: {"Apr2021": {'2021-04-08': {"W/L": 'W', "Mets Runs": 3, ...}, '2021-04-10': {...}, ...}, "May2021": {...}, ...}

4. avg_month_game(months)

Creates a dictionary containing the average baseball game stats of Mets winning percentage, Mets runs, Opponents runs, and total games in a month
INPUT: months: dictionary created in dates_by_month_full_comb
OUTPUT: avg_bball_month: dictionary described above
ex: {"Apr2021": {"Winning Percentage": 0.6, "Mets Runs per Game": 3.0, ...}, "May2021": {...}, ...}

5. avg_month_weather(months)

Creates a dictionary containing the average weather stats of Temperature (Fahrenheit), Precipitation (mm), Max Wind Speed, and total games in a month
INPUT: months: dictionary created in dates_by_month_full_comb
OUTPUT: avg_weather_month: dictionary described above
ex: {"Apr2021": {"Average Temperature": 53.26, "Average Precipitation": 3.5, ...}, "May2021": {...}, ...}

6. winloss_by_winddirection(months)

Creates a dictionary containing the Mets winning percentage by wind direction
INPUT: months: dictionary created in dates_by_month_full_comb
OUTPUT: average_win: dictionary described above
ex: {"Southeast": {"Winning Percentage": 0.65, "Total Games": 20}, "South": {...}, ...}

7. winloss_by_weathercode(months)

Creates a dictionary containing the Mets winning percentage by weather code
INPUT: months: dictionary created in dates_by_month_full_comb
OUTPUT: average_win: dictionary described above
ex: {"partly cloudy": {"Winning Percentage": 0.375, "Total Games": 16}, "moderate drizzle": {...}, ...}

8. avg_runs_by_weathercode(months, team)

Creates a dictionary containing the Mets winning percentage by weather code
INPUTS: months: dictionary created in dates_by_month_full_comb
team: a string containing either "Mets","Opponents", or "Total"
OUTPUT: average_win: dictionary described above
ex: given team="Mets",
{"partly cloudy": {"Average Mets Runs": 4.125, "Total Games": 16}, "moderate drizzle": {...}, ...}

9. avg_runs_by_wind_direction(months, team)

Creates a dictionary containing the Mets winning percentage by wind direction
INPUTS: months: dictionary created in dates_by_month_full_comb
team: a string containing either "Mets","Opponents", or "Total"
OUTPUT: average_win: dictionary described above
ex: given team="Mets",
{"South": {"Average Mets Runs": 4.125, "Total Games": 16}, "Southeast": {...}, ...}

Visualization.py

1. open_json(filename)
   Takes in a string with the json file name and outputs the data from the json in a dictionary called data

2. barplot(data, key, y_axis, title, x_title, y_title, outfile)

The key is the key for the dictionary to extract the right data for the dataframe. The axis titles and title of the graph are for the visualizations. The outfile is the file you want the visualization to be saved to. The function creates the barplots.

3. scatterplot(data_dict, x_var, y_var, title, x_title, y_title, outfile)

The scatterplot takes in the correctly formatted data_dict where the x_var and y_var the data keys to the dictionary to be used in the scatterplot. The title and axis titles are used to describe the plot. The outfile is a string to write the scatterplot to. The final output is the scatterplot.

Documentation of Resources used:

Date	Issue Description	Location of Resource	Result
4/16/23	Couldn't figure out how to run the program differently each time for getting info from api	https://stackoverflow.com/questions/30961310/how-can-i-count-how-many-times-this-program-has-been-executed-in-python	Created count.txt to keep track of the amount of runs to figure out what needs to be added to the database next.
4/17/23	Unsure of how date comparisons work in SQL	https://stackoverflow.com/questions/14208958/select-data-from-date-range-between-two-dates	Figured it out
4/19/23	How to manipulate the syntax in seabron to make designs	https://seaborn.pydata.org/	The visualizations worked.