Exploratory Data Analysis Report:

Predicting Marathon Average Finishing Times: The Impact of Weather and Air Quality on Marathons

Team Lead: Zachary D’Urso

Recorder: Meghan Holden

Spokesperson: Krisha Bugajski

Link to Github:

[BugajskiSharp/capstone-project-team-c: Team C data capstone project.]

Question:

Can weather and environmental factors—such as temperature, wind speed, precipitation, dew point, visibility, sea level pressure, and air quality—predict average marathon finishing times across different performance groups (elite, competitive, average, and slow runners) and genders?

Background:

This research looks at how environmental factors like weather and air quality affect marathon finishing times for different groups of runners from 1996 to 2024, offering helpful insights that can guide organizers to plan security, medical support, and hydration stations, while allowing runners to pace themselves and adjust their expectations. By looking at multiple environmental factors together, which only a few studies have done, this work can improve race day planning, enhance overall safety, and deepen our understanding of how climate and pollution influence performance in major marathons like Boston, New York, Chicago, and Berlin.

Hypothesis:

H_0 : Weather, and air quality have no significant relationship with the average finishing times across marathon performance groups or genders.

H_1 : Poor weather conditions (e.g., higher temperatures, dew point wind speed, precipitation, sea level pressure, or worse air quality) are associated with slower average finishing times, particularly for non-elite runners.

To isolate environmental effects, we will include a control variable for the introduction of supershoes, which helps account for performance improvements due to running shoe technology.

Predictions:

- **Elite runners** will be least affected by environmental stressors due to higher fitness and experience (Ely, Martin, Cheuvront, & Montain, 2007).
- **Competitive runners** will show greater increases in average finishing times under harsher weather and environmental conditions (Vihma, 2010).
- **Slow runners** will experience the largest slowdowns under harsher weather and environmental conditions (Helou et al., 2012).
- **Female runners** will show slightly smaller performance declines under adverse weather and environmental conditions compared to males (Vihma, 2010).
- **Moderate temperatures, low humidity, and clean air** will produce the fastest finishing times across all categories (Gasparetto & Nesseler, 2020).

Methods:

Marathon results, weather, and air quality data were collected from multiple sources to analyze finisher performance in relation to environmental conditions. The marathon datasets cover Boston (1970–2019), New York City (1970–2024), Chicago (1996–2023), and Berlin (1974–2019). Weather data includes daily temperature, precipitation, dew point, wind, visibility, and sea level pressure for each marathon location. Air quality data consist of daily AQI values and specific pollutant measurements (CO, O_3, $PM_{2.5}$, PM_{10}, NO_2), with the main pollutant representing overall AQI. Berlin air quality data were supplemented from local monitoring sources since EPA data covers only U.S. locations, and pollutant concentrations were converted to AQI.

Sources:

- **Boston Marathon Data Project on GitHub** (1970–2019): https://github.com/rfordatascience/tidytuesday/tree/master/data/2020/2020-04-07
- **New York City Marathon results on Kaggle** (1970–2024): https://www.kaggle.com/datasets/runningwithrock/nyc-marathon-results-all-years
- **Chicago Marathon results on Kaggle** (1996–2023): https://www.kaggle.com/datasets/ericwan7732/chicago-marathon-2017-2019
- **Berlin Marathon results on Kaggle** (1974–2019): https://www.kaggle.com/datasets/the-dagger/berlin-marathon-results
- **Weather Underground** (U.S. marathons): https://www.wunderground.com/
- **Meteostat** (Berlin): https://meteostat.net/
- **Extreme Weather Watch** (Berlin min/max temperatures): https://www.extremeweatherwatch.com/
- **Weatherspark** (Berlin dew point and visibility, where 1996-2010 data came from Alexanderplatz and 2011-2024 data came from Tempelhof): https://weatherspark.com/
- **U.S. EPA Air Quality Index Daily Values Report**: https://aqs.epa.gov/aqsweb/airdata/download_files.html
- **Berlin Air Quality Monitoring Network**: https://luftdaten.berlin.de/

Data Preparation Summary (Merging, Joining, Aggregation):

In order to get all our data ready to start our exploratory data analysis, we had to pull together several different datasets that did not originally align with one another. Our marathon data, weather data, and air quality data were all stored separately and used different column names and formats, so our first steps were to clean and standardize everything.

We then reformatted each marathon dataset so that key variables (such as year, marathon, gender, and finishing time) were consistent across all races and filtered the data to include only the years from 1996 to 2024, noting that some datasets only extend through 2019 (Boston and Berlin) or 2023 (Chicago), depending on availability. Once the datasets all used the same structure, we combined them using a row-bind (full append), since each file represented a different marathon. We then converted finishing times into seconds and created performance subgroups using within-year quantiles. This required some aggregation, since we calculated the average finishing time for each subgroup in each marathon, year, and gender. However, because the raw Berlin 2019 data contained no female finishers, all five expected female Berlin records for that year were missing when the subgroups were created.

Once the marathon data was all ready, we merged it with the weather and air quality data. We used left joins, matching by year and marathon, to attach the two datasets to each performance group.

After doing these joins, we ended up with one clean and combined dataset, where each row represents a gender specific performance group in a given marathon year, along with the weather and air quality variables for that given race day. This final merged dataset is what we will now use for our EDA and modeling.

Methods for EDA:

Variables were categorized into categorical or continuous variables. Categorical variables include marathon, year, gender, and main pollutant. Continuous variables include average chip time, high, low, and average temperature, precipitation, dewpoint, wind speed, visibility, sea level pressure, AQI, CO, ozone, PM_{10}, $PM_{2.5}$, and NO_2. See Data Dictionary in Appendix 1 for the variable names, descriptions, and data types.

The categorical variables were assessed for frequency and proportion, stratified by performance subgroup and gender, to help identify missing data and understand variables that may have a large or disproportionate influence on the outcome variable. Table 2 shows the descriptive statistics for the continuous variables include mean, median, standard deviation, minimum, maximum, and distribution shape. After grouping the numerical variables by gender and subgroups, the variables were evaluated specifically using measurements of mean and standard deviation, which can be seen in Appendix 4 and Appendix 5. Variable correlation and collinearity were measured using a correlation matrix.

Distributions and relationships, including correlation, of the continuous variables were investigated visually with histograms and scatterplots.

Particular attention was given to relationships of the variables in the predictions, variables with a high number of missing values, as well as variables expected to have minimal or no relationship to the outcome variable, to help guide decisions for including or excluding them as predictor variables in the model.

Analyses and visualizations were done in RStudio using the dplyr, gtsummary packages to generate tables, and the ggplot2 and corrplot package for visualizations.

Results:

The final merged dataset includes 1005 rows of data, breaking each marathon down by year, gender, performance subgroup, and the corresponding average chip time. There are 21 variables in the dataset; 16 of those variables contain relevant information about each marathon on race day, including air quality data (AQI, main pollutant, CO, ozone, PM_{10}, $PM_{2.5}$, NO_2), and weather data (daily high, low and average temperature (°F), precipitation, dewpoint, wind speed, visibility, and sea level pressure).

Table 1 shows the categorical variables from the dataset. Appendix 2 shows the same data broken down by performance subgroup and gender to understand where the missing values come into play. Appendix 3 includes the breakdown of available marathon performance subgroup data by years and gender. The year and marathon variables will not be used as predictor variables but rather identifiers, and will be included in the table below to assess completeness of the data. The year variable will be used to incorporate the control for supershoes and COVID-19 during feature engineering, if applicable.

Notably, one marathon is missing all female runner data (Berlin 2019), resulting in the female subgroups having one less instance where $PM_{2.5}$ is the main pollutant. Boston is missing 2013 data due to the Boston Marathon Bombing, and New York City is missing

2012 data due to Hurricane Sandy. From 2021 to 2023, only Chicago and New York City marathon data are available, and in 2024, only New York City marathon data is available.

Among the main pollutants, $PM_{2.5}$ and ozone occur at the highest frequency, while CO does not appear as the main pollutant in any race.

Table 1. Categorical Variables

Variable	**Category**	**N(%)**
Gender	Male	505 (50%)
	Female	500 (50%)
Subgroup	Elite	201 (20%)
	Competitive	201 (20%)
	Average	201 (20%)
	Recreational	201 (20%)
	Slow	201 (20%)
Marathon	Chicago	270 (27%)
	NYC	270 (27%)
	Berlin	235 (23%)
	Boston	230 (23%)
Main Pollutant	PM2.5	595 (59%)
	Ozone	200 (20%)
	NO2	150 (15%)
	PM10	60 (6%)

Table 2 shows the continuous data from the dataset. Appendix 4 shows the continuous outcome variable broken down by subgroup and gender. Appendix 5 includes continuous variables by gender; the weather and air quality parameters are slightly different between male and female datasets because of the missing female data in the 2019 Berlin marathon.

Table 2. Descriptive stats for Continuous Variables

variable	mean	sd	median	min	max	skew	kurtosis	n
avg_chip_seconds	15631.62	2938.88	15284.20	10304.77	23679.03	0.51	-0.39	1005
high_temp	61.62	10.48	61.00	44.00	88.00	0.58	-0.36	1005
low_temp	46.01	7.97	45.00	28.00	72.00	0.77	1.02	1005
avg_temp	53.72	8.70	52.88	29.58	79.35	0.42	0.31	1005
dew_point	41.02	10.72	42.40	21.13	65.22	-0.08	-0.91	1005
wind_speed	13.12	6.77	13.00	3.00	39.00	0.79	1.09	1005
visibility	9.85	1.49	10.00	6.06	20.00	2.79	25.70	830
sea_level_pressure	29.93	0.34	29.98	29.13	30.54	-0.41	-0.75	1005
aqi	50.57	21.76	52.00	11.00	119.00	0.32	0.22	1005
co	13.95	11.78	9.00	2.00	56.00	1.59	2.16	770
ozone	36.18	18.05	34.00	7.00	100.00	1.22	2.34	1005
pm10	23.58	13.03	21.00	5.00	69.00	0.89	0.81	625
pm25	57.51	16.79	53.00	21.00	119.00	0.85	1.39	725
no2	31.92	14.93	32.00	7.00	66.00	0.16	-0.78	1005

Figure 1. Overview of the distribution of average finishing time in hours for all runners

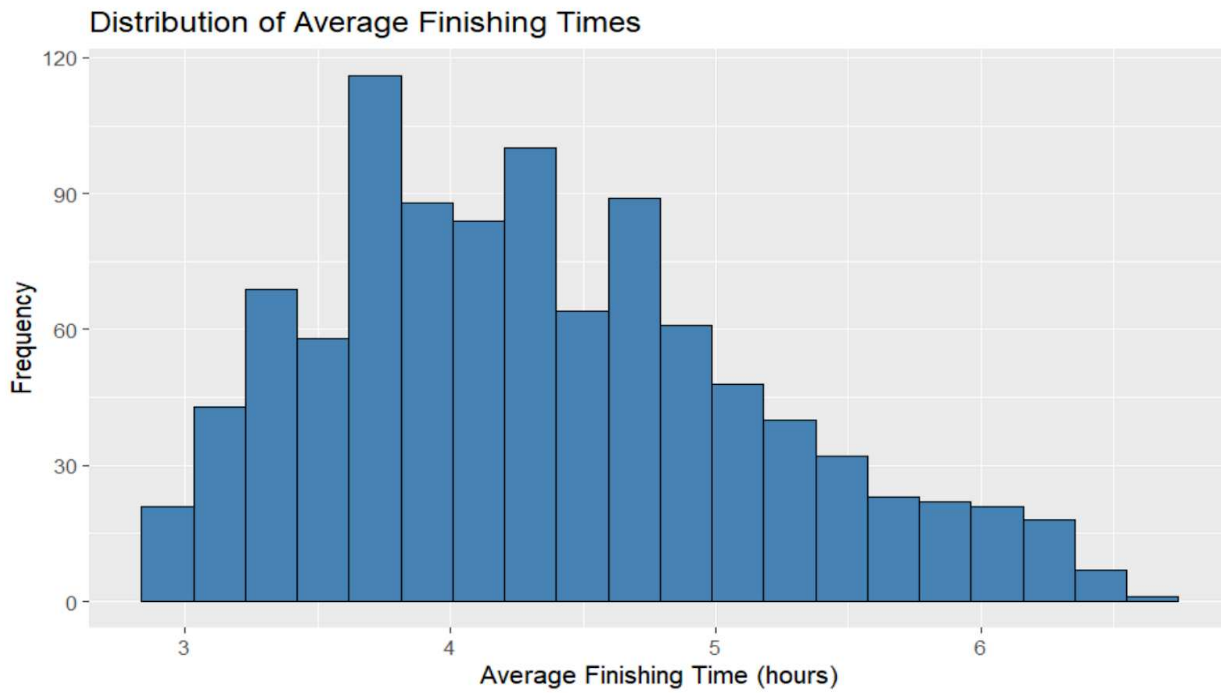


Figure 1 displays the overall distribution of average finishing times amongst all runners in all races via a histogram. The distribution is right-skewed, where a large portion of runners are clustered around the 4-hour mark (competitive runners). We can see that there are fewer runners with longer times (slow/recreational runners), extending the tail of the distribution. A strong majority of the runners have finishing times between the 3.5-hour mark and the 4.75-hour mark.

Figure 2. Relationship between average temperature and finishing time by each marathon

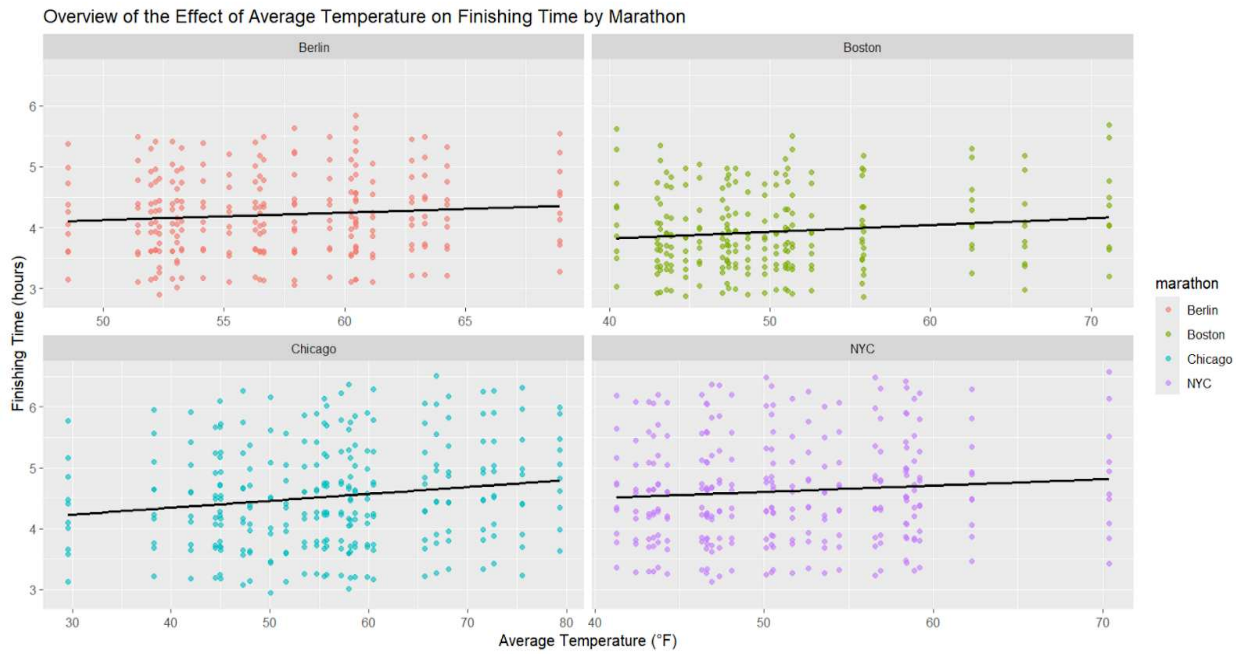


Figure 2 utilizes faceted scatterplots and picks out the effect of average temperature on the average finishing time by each individual marathon. Each marathon has its own scatterplot, with its own temperature scale on the x-axis and finishing time on the y-axis. The black lines represent fitted regression trend lines that show how the average finishing time changes alongside temperature. Runners are represented by the points on the scatterplot. All four of the trend lines have a slight upward slope, which indicates that as the temperature increases, the average finishing times tend to become slower as runners take longer to complete the race. From this, we can gather that higher temperatures likely increase the risk of dehydration and fatigue amongst runners, causing runners to slow down. It appears that Chicago specifically is most impacted by increases in temperature on race day.

Figure 3. Relationship of air quality and finishing time by marathon

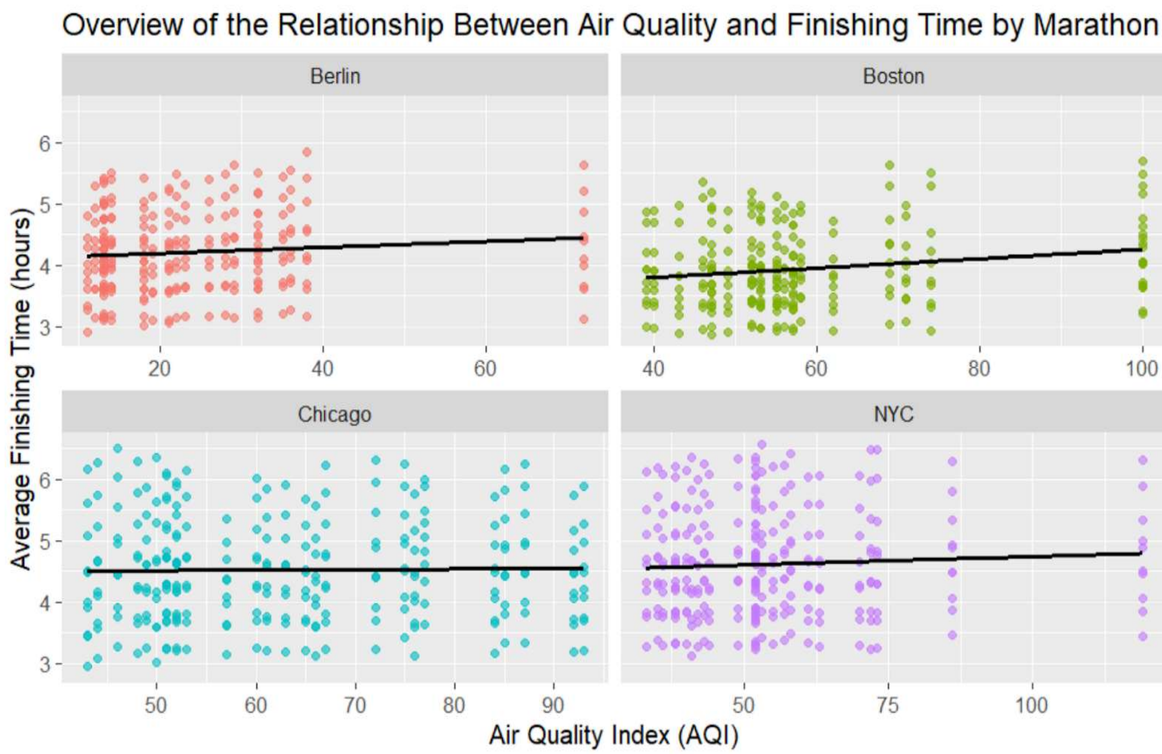


Figure 3 focuses on the relationship between air quality and the average finishing time by marathon. For Berlin, the AQI values are low, with the majority of data points within an AQI value of 40. The AQI values are low most likely because of the missing PM2.5 which is a driver for higher AQI. The trend line here shows that there is generally no relationship between air quality and finishing time. For Boston, the AQI values stretch across a wider range of values, with a slightly upward trend line, suggesting that as AQI increases, so do finishing times. For Chicago, the majority of the data falls between AQI levels of 50 and 90, with a fairly straight trend line representing no relationship between AQI and finishing time. There are many more data points scattered towards the longer finishing times for Chicago than there are for the other race locations. For NYC, the data is centered between AQI levels of 50 and 75, with again a minimal relationship between AQI and finishing times. There is no significant impact of AQI on finishing time based on this faceted plotting of the variables. The effects seem minor, but there is potential for poor AQI levels to negatively affect finishing times.

Figure 4 shows how marathon finishing times change with different pollution levels across performance groups. As pollution, especially PM2.5, increases, finish times generally get slower. Elite runners are the least affected, while slower runners show the biggest delays.

We also noticed that CO and PM10 have a lot of missing data, so we plan to remove them from the analysis. Additionally, AQI is highly correlated with PM2.5 as seen in the correlation matrix Appendix 5, so we will remove AQI as well to avoid redundancy and multicollinearity.

Figure 4: Scatterplot showing the relationship between pollutant levels and finishing times by subgroups.

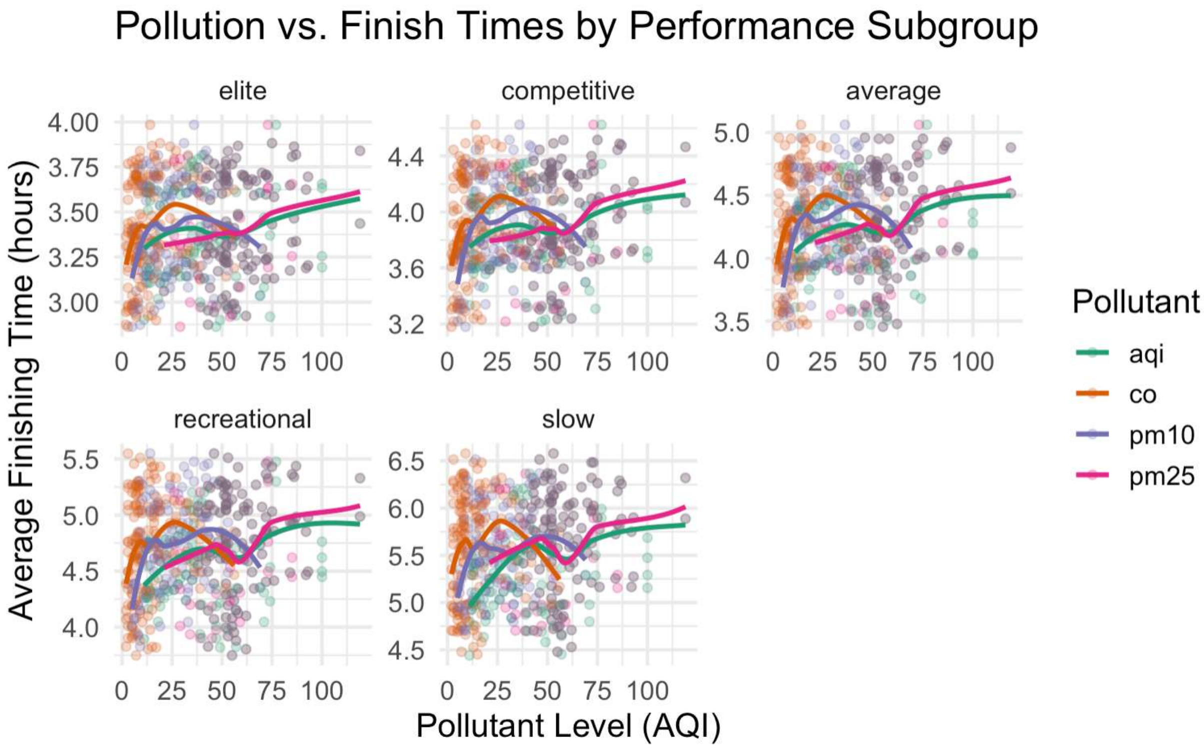
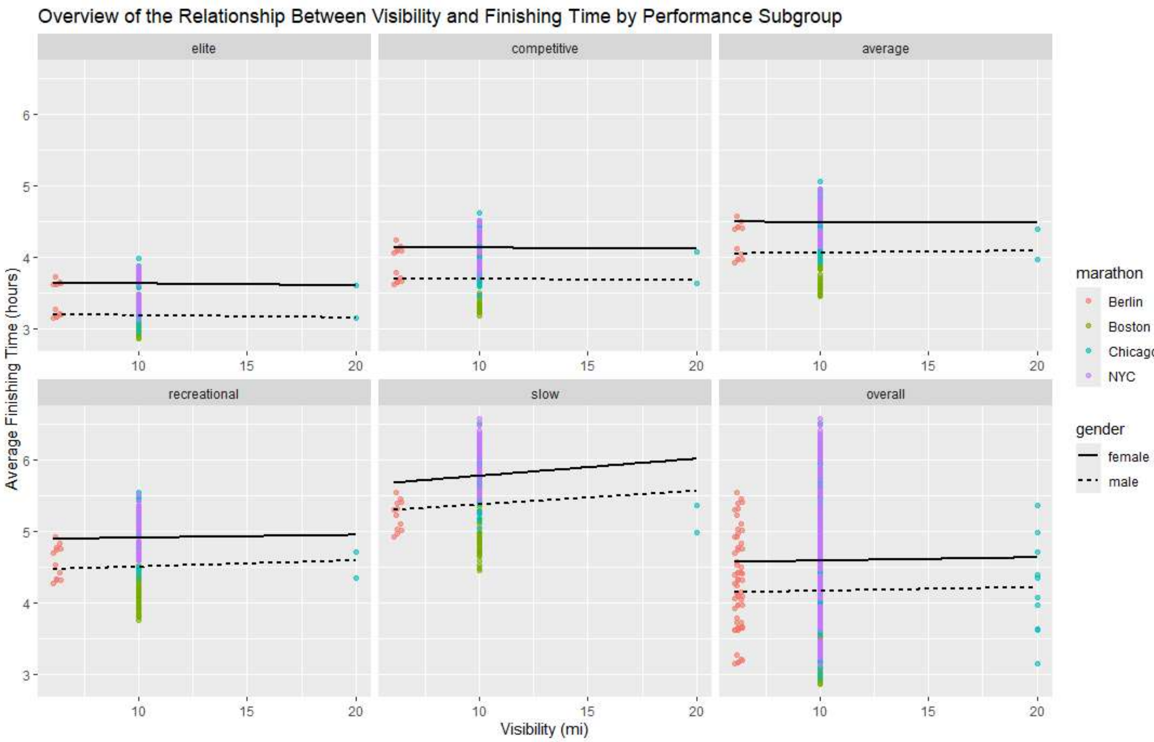


Figure 5 shows the relationship between visibility and finishing time by performance subgroup. For elite and competitive runners, there seems to be no- to a very small decrease in average finishing time as visibility increases. Average runners see almost no difference in finishing time, whereas recreational runners see a slight performance decrease as visibility increases. Slow runners see the largest change, decreasing performance as visibility increases. The lack of continuity in these changes suggests that visibility is not a driver for finishing time. Due to this and the large number of missing values, visibility will be removed as a predictor variable.

Figure 5. Relationship between visibility and finishing time by performance subgroup



Discussion & Next Steps

After completing our exploratory data analysis, we can see that pollution and temperature both significantly slow down the performance of runners in these races across all subgroups. Higher levels of pollution and higher temperature likely lead to slower finishing times across finishers.

During the EDA process, a few potential key changes to our data became clear to us. One main area of focus for our next steps in this process is to remove the data we currently have for PM10, AQI, CO, and Visibility. We believe due to the high number of missing values, and the relationships seen above, keeping these variables does not drive enough value in our future prediction making processes. As a result, we are planning to remove these variables going forward. This will allow us to focus on measuring the more valuable relationships at hand, instead of experimenting with overlapping variables or variables that are missing substantial amounts of data.

In the coming steps, our plan is to normalize the continuous variables, with the exception of avg_chip_seconds as it is the key predictor. This is because our data is looking at different marathons, in different locations, experiencing different weather patterns historically. Normalizing the data will help to create a more generalized model.

Figure 6 shows the trend of average finishing time over the years by subgroup. In consideration of the introduction of ‘supershoes’ in 2018 and how to control for a potential confounding factor, the red line denotes the year 2018 to see how race times have changed. Looking at the figure below, the opposite of the expected result is seen: a trend in slower marathons over time. This may be due to subgrouping performance groups by quintiles in each marathon instead of using cut offs. In earlier years, there were less runners overall, and possibly less charity runners as well, which may have resulted in higher performers in slower running groups and a trend in slower marathons over time.

Figure 6: Average Finishing Time by Year, Subgroup and Gender

Next Steps: Data Cleaning & Pre-processing Plan:

Over the next stage of the project, our focus will be on preparing the final combined dataset for modeling by cleaning the data. When weather or air quality values are missing, we always prioritize the actual race day measurement, and if that isn't available, we will use values from the surrounding days, short-range interpolation, or a monthly average when needed. This lets us complete the dataset without making any assumptions that might distort the results.

We will also complete several feature engineering steps, including converting all time variables to numeric seconds, scaling continuous predictors, and creating key interaction terms such as temperature × dew point and AQI × temperature. In addition, we will finalize control variables for both the introduction of “super shoes” beginning around 2018 and the onset of COVID-19-related race disruptions starting in 2020. Once cleaning is complete, we will run exploratory data analysis to examine correlations, distributions, and potential multicollinearity before fitting our first models. From there, we will compare multiple linear regression models with tree-based gradient boosting approaches to determine which best captures the relationships relevant to our research question.

APPENDIX

Appendix 1: Data Dictionary

Variable	Description	Type
n	Total number of runners that completed the marathon	Integer
marathon	Name of the Marathon Event	Categorical
year	Year of the marathon	Integer
gender	Gender category of runners	Categorical
subgroup	Running performance category: elite, competitive, average, recreational, slow	Factor w/ 5 levels
avg_chip_seconds	Average finishing chip time in seconds by marathon, subgroup, and gender	Numerical
high_temp	Daily high of race day (°F)	Integer
low_temp	Daily low of race day (°F)	Integer
avg_temp	Daily average temperature (°F)	Numerical
precipitation	Total precipitation on race day (in)	Numerical
dew_point	Average dew point temperature (°F) on race day	Numerical
wind_speed	Average wind speed on race day (mph)	Integer
visibility	Average visibility on race day (statute miles)	Numerical
sea_level_pressure	Average sea level pressure on race day (inHg)	Numerical
aqi	Average Air Quality Index (AQI) on race day	Integer
main_pollutant	Primary pollutant on race day	Categorical
co	Average carbon monoxide (AQI) on race day	Numerical
ozone	Average ozone (AQI) on race day	Numerical
pm10	Average particulate matter with diameter of less than 10 micrometres (PM10) (AQI) on race day	Numerical
pm25	Average particulate matter with diameter of less than 2.5 micrometres (PM2.5) (AQI) on race day	Numerical
no2	Average nitrogen dioxide (AQI) on race day	Numerical

Appendix 2: Categorical Variables by Subgroup and Gender

Variable	elite N=201 (100%)	competitive N=201 (100%)	average N=201 (100%)	recreational N=201 (100%)	slow N=201 (100%)	All N=1005 (100%)
gender						
female	100 (50%)	100 (50%)	100 (50%)	100 (50%)	100 (50%)	500 (50%)
male	101 (50%)	101 (50%)	101 (50%)	101 (50%)	101 (50%)	505 (50%)
marathon						
Berlin	Female: 23 (23%) Male: 24 (24%) All: 47 (23%)	Female: 23 (23%) Male: 24 (24%) All: 47 (23%)	Female: 23 (23%) Male: 24 (24%) All: 47 (23%)	Female: 23 (23%) Male: 24 (24%) All: 47 (23%)	Female: 23 (23%) Male: 24 (24%) All: 47 (23%)	Female: 115 (23%) Male: 120 (24%) All: 235 (23%)
Boston	Female: 23 (23%) Male: 23 (23%) All: 46 (23%)	Female: 23 (23%) Male: 23 (23%) All: 46 (23%)	Female: 23 (23%) Male: 23 (23%) All: 46 (23%)	Female: 23 (23%) Male: 23 (23%) All: 46 (23%)	Female: 23 (23%) Male: 23 (23%) All: 46 (23%)	Female: 115 (23%) Male: 115 (23%) All: 230 (23%)
Chicago	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 135 (27%) Male: 135 (27%) All: 270 (27%)
NYC	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 27 (27%) Male: 27 (27%) All: 54 (27%)	Female: 135 (27%) Male: 135 (27%) All: 270 (27%)
main_pollutant						
NO_2	Female: 15 (15%) Male: 15 (15%) All: 30 (15%)	Female: 15 (15%) Male: 15 (15%) All: 30 (15%)	Female: 15 (15%) Male: 15 (15%) All: 30 (15%)	Female: 15 (15%) Male: 15 (15%) All: 30 (15%)	Female: 15 (15%) Male: 15 (15%) All: 30 (15%)	Female: 75 (15%) Male: 75 (15%) All: 150 (15%)
Ozone	Female: 20 (20%) Male: 20 (20%) All: 40 (20%)	Female: 20 (20%) Male: 20 (20%) All: 40 (20%)	Female: 20 (20%) Male: 20 (20%) All: 40 (20%)	Female: 20 (20%) Male: 20 (20%) All: 40 (20%)	Female: 20 (20%) Male: 20 (20%) All: 40 (20%)	Female: 100 (20%) Male: 100 (20%) All: 200 (20%)
PM_{10}	Female: 6 (6.0%) Male: 6 (5.9%) All: 12 (6.0%)	Female: 6 (6.0%) Male: 6 (5.9%) All: 12 (6.0%)	Female: 6 (6.0%) Male: 6 (5.9%) All: 12 (6.0%)	Female: 6 (6.0%) Male: 6 (5.9%) All: 12 (6.0%)	Female: 6 (6.0%) Male: 6 (5.9%) All: 12 (6.0%)	Female: 30 (6.0%) Male: 30 (5.9%) All: 60 (6.0%)
$PM_{2.5}$	Female: 59 (59%) Male: 60 (59%) All: 119 (59%)	Female: 59 (59%) Male: 60 (59%) All: 119 (59%)	Female: 59 (59%) Male: 60 (59%) All: 119 (59%)	Female: 59 (59%) Male: 60 (59%) All: 119 (59%)	Female: 59 (59%) Male: 60 (59%) All: 119 (59%)	Female: 59 (59%) Male: 60 (59%) All: 595 (59%)

Appendix 3: Year data by subgroup

	female						male					
Variable	**elite N=100 (100%)**	**competitive N=100 (100%)**	**average N=100 (100%)**	**recreational N=100 (100%)**	**slow N=100 (100%)**	**All N=500 (100%)**	**elite N=101 (100%)**	**competitive N=101 (100%)[1]**	**average N=101 (100%)**	**recreational N=101 (100%)**	**slow N=101 (100%)**	**All N=505 (100%)**
year												
1996	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
1997	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
1998	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
1999	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2000	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2001	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2002	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2003	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2004	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2005	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2006	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2007	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2008	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2009	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2010	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)

2011	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2012	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	15 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	15 (3.0%)
2013	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	15 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	15 (3.0%)
2014	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2015	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2016	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2017	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2018	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2019	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	3 (3.0%)	15 (3.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	4 (4.0%)	20 (4.0%)
2021	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)
2022	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)
2023	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	2 (2.0%)	10 (2.0%)
2024	1 (1.0%)	1 (1.0%)	1 (1.0%)	1 (1.0%)	1 (1.0%)	5 (1.0%)	1 (1.0%)	1 (1.0%)	1 (1.0%)	1 (1.0%)	1 (1.0%)	5 (1.0%)

Appendix 4: Outcome variable by performance subgroup and gender

	Elite Mean (sd)	Competitive Mean (sd)	Average Mean (sd)	Recreational Mean (sd)	Slow Mean (sd)
Average Chip Seconds	Female: 13029.11 (552.0283) Male: 11425.49 (545.3133)	Female: 14812.50 (903.8537) Male: 13235.19 (846.8188)	Female: 16086.47 (1175.2357) Male: 14528.82 (1057.5026)	Female: 17545.20 (1433.1433) Male: 16074.19 (1256.1329)	Female: 20515.17 (1903.6196) Male: 19101.97 (1576.7362)

Appendix 5: Continuous variables by gender

	Female Mean (sd)	Male Mean (sd)
High Temp	61.62 (10.55412)	61.62376 (10.50129)
Low Temp	45.98 (8.015112)	46.0396 (7.997401)
Average Temp	53.6972 (8.753746)	53.73901 (8.719997)
Precipitation	0.0317 (0.1206134)	0.03356436 (0.1214626)
Dew Point	40.9614 (10.7576)	41.08564 (10.77626)
Wind Speed	13.13 (6.815987)	13.10891 (6.785132)
Visibility	9.848072 (1.495357)	9.848072 (1.495357)
Sea Level Pressure	29.9273 (0.3410111)	29.92465 (0.3403427)
AQI	50.72 (21.81422)	50.42574 (21.90541)
PM10	23.70968 (13.09284)	23.44444 (13.15635)
PM2.5	57.76389 (16.6729)	57.26027 (17.10671)
NO2	32.04 (14.92535)	31.79208 (15.0581)
Ozone	36.27 (18.1341)	36.08911 (18.13455)
CO	13.94805 (11.84871)	13.94805 (11.84871)

Appendix 6: Correlation Matrix (Complete)