

Data science competition

Flyball

Tamás Berki

Optimizing Performance for Wild Runners

This analysis dives into your team's performance data to identify key areas for improvement and provide data-driven recommendations to help you achieve podium finishes!



Background

Data source:

- Competition data provided by the Wild Runners team in Excel format
- Weather data is sourced from Weather API (OpenWeatherMap.org)

Data overview:

- Run-time and change-time for each dog in each race, including errors
- Date, time, team name and hurdle heights for each race
- Dog's data: name, breed, neuter status, birthdate, ulna length, and jump height

Objective: Leverage this data to identify key areas for improvement and provide data-driven recommendations to help Wild Runners to achieve podium finishes in future competitions



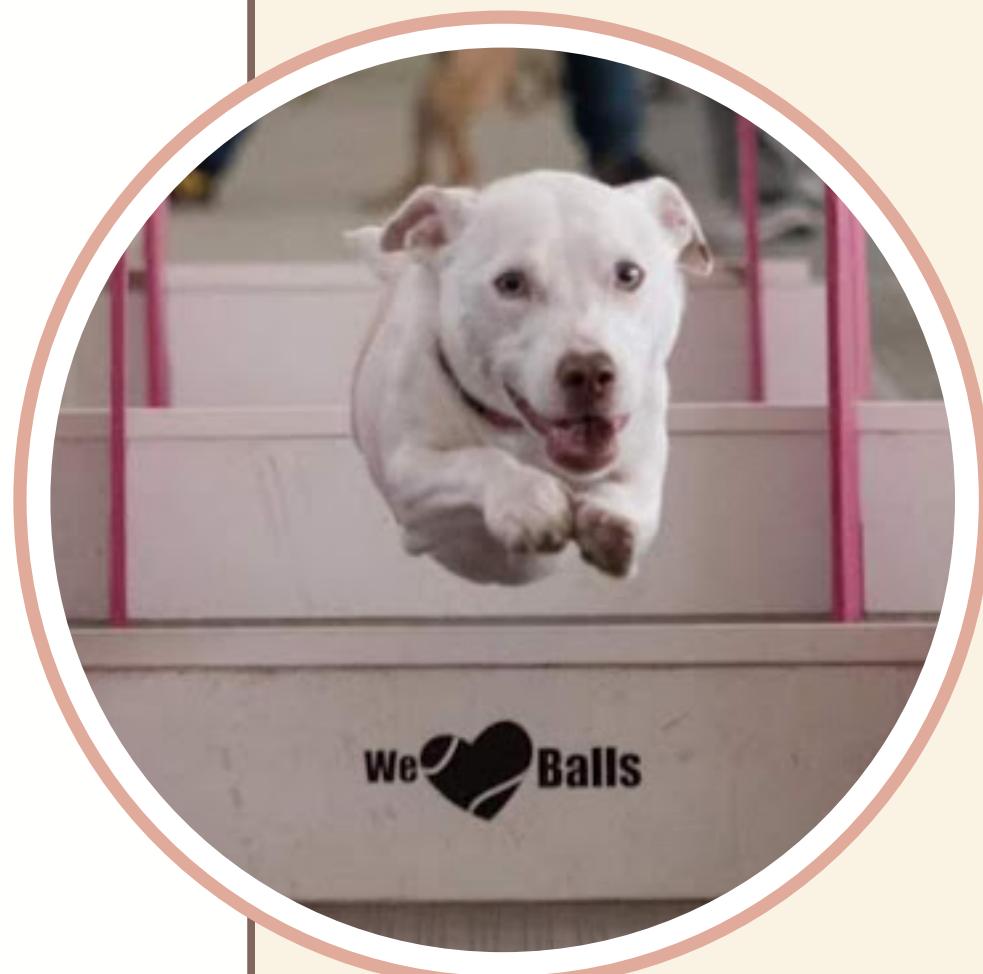
Data Cleaning

Applied libraries:

- Pandas
- Openpyxl: to read text color

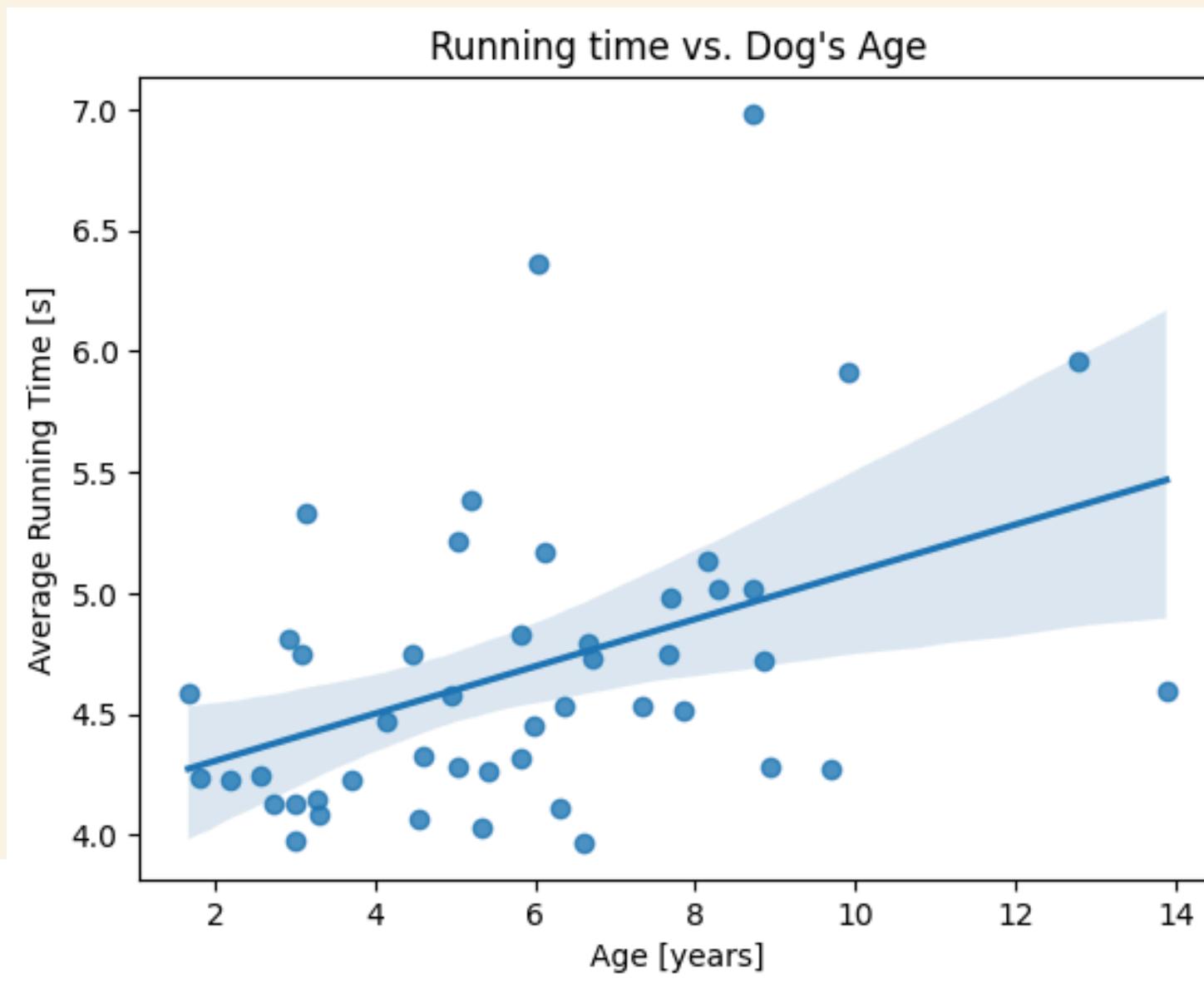
Data cleaning steps:

- Races with too many missing values are dropped
- OKs as start or change time are replaced with zero
- Errors during start and change are replaced with -1
- Data types are unified
- Missing hurdle heights are substituted
- Breed names are harmonized
- Unrealistic ulna lengths are corrected
- Data indicated with text color is converted to a new column

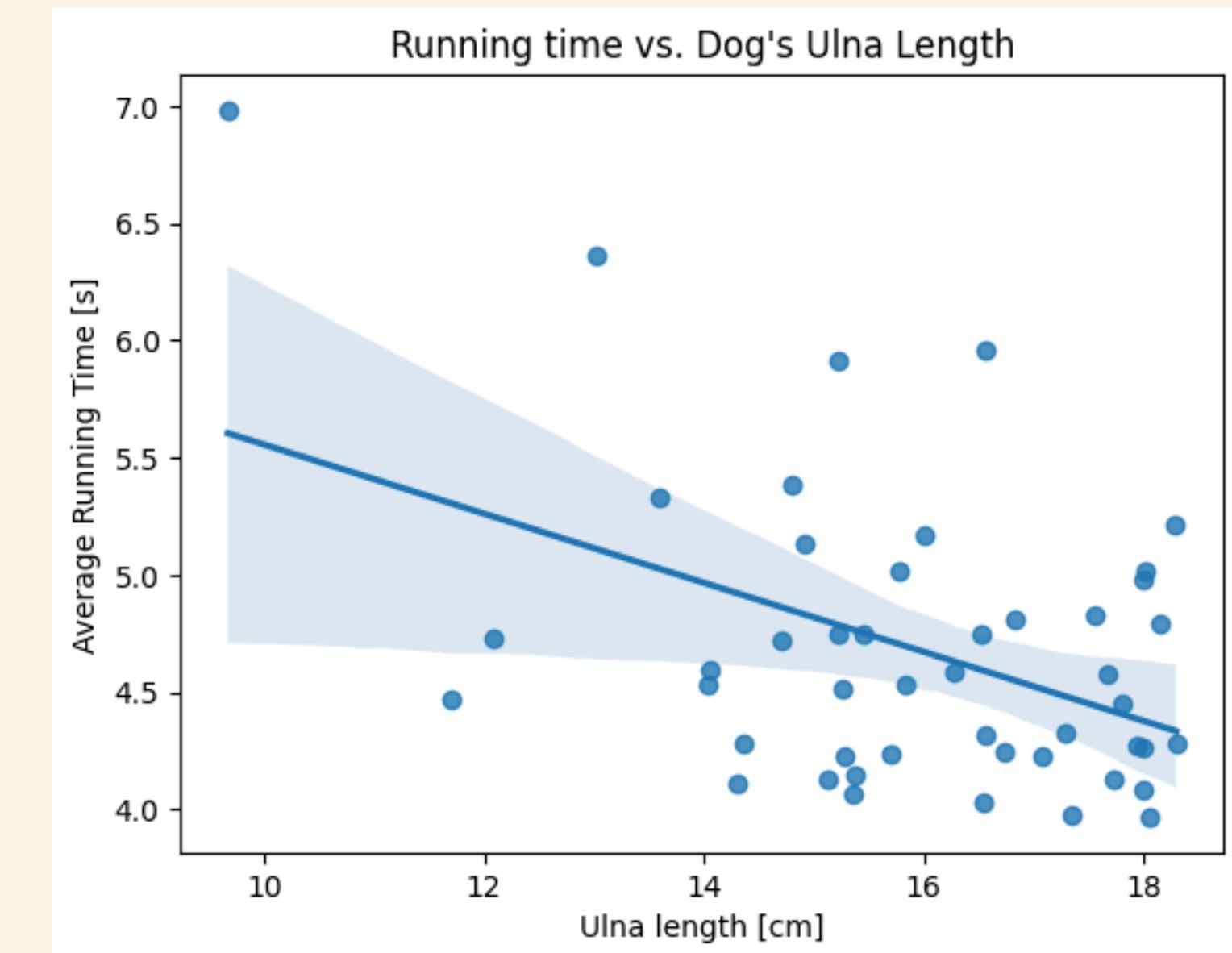


Basic Analysis 1

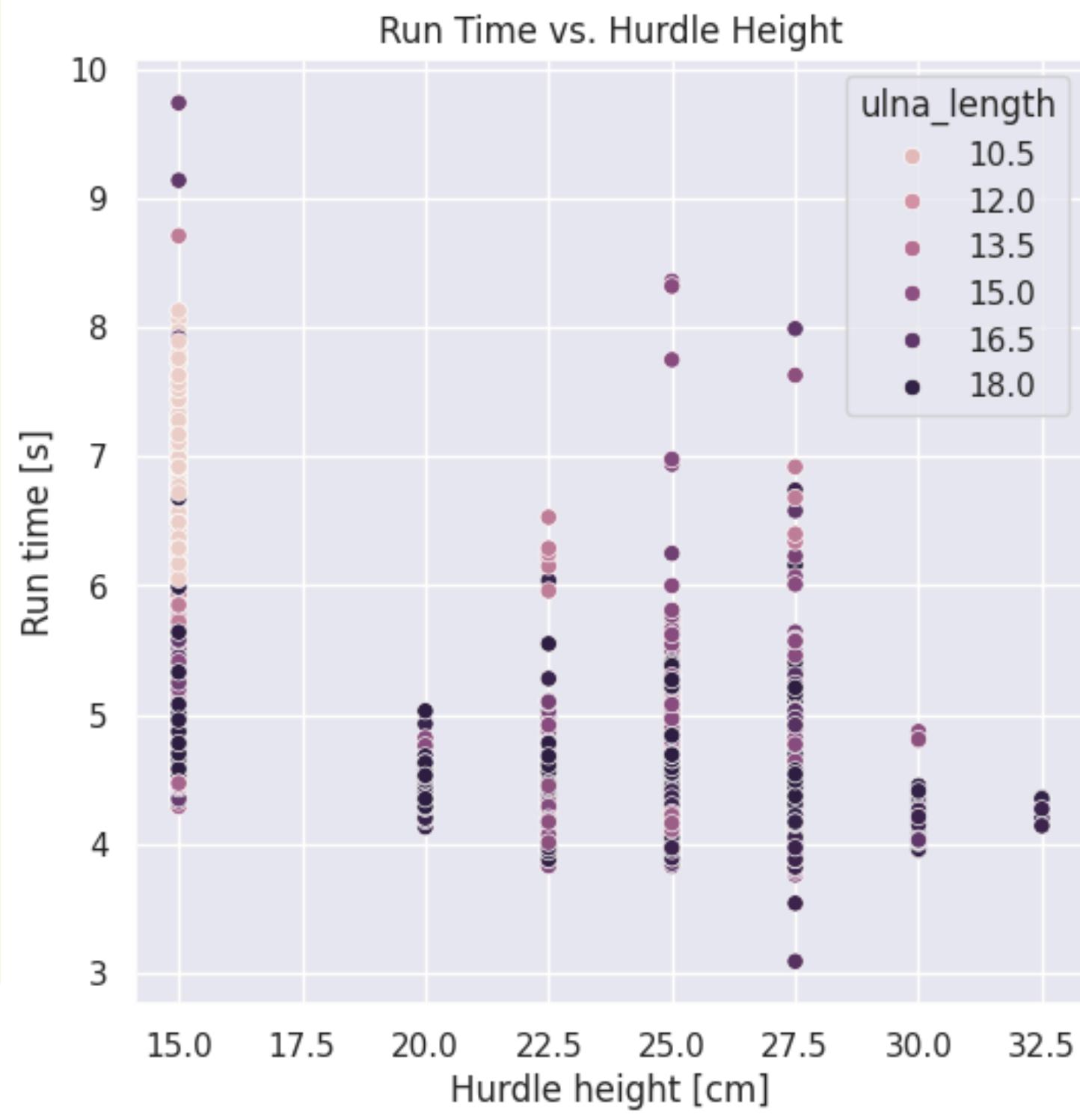
Younger dogs run faster (0.1 sec/year)



Larger dogs run faster (-0.15 sec/cm)



Basic Analysis 2

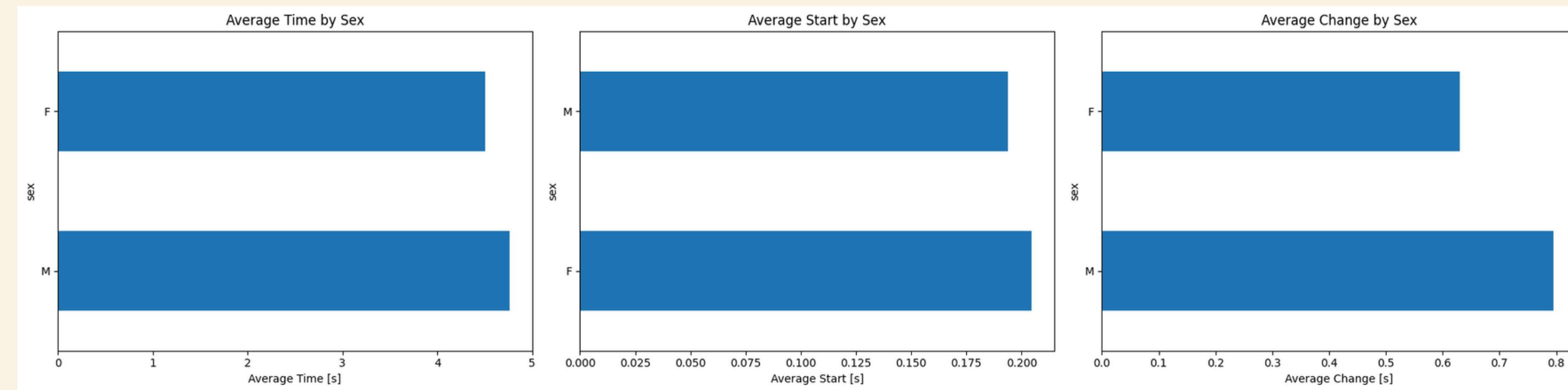


Mixing different sized dogs in the same team is a bad idea

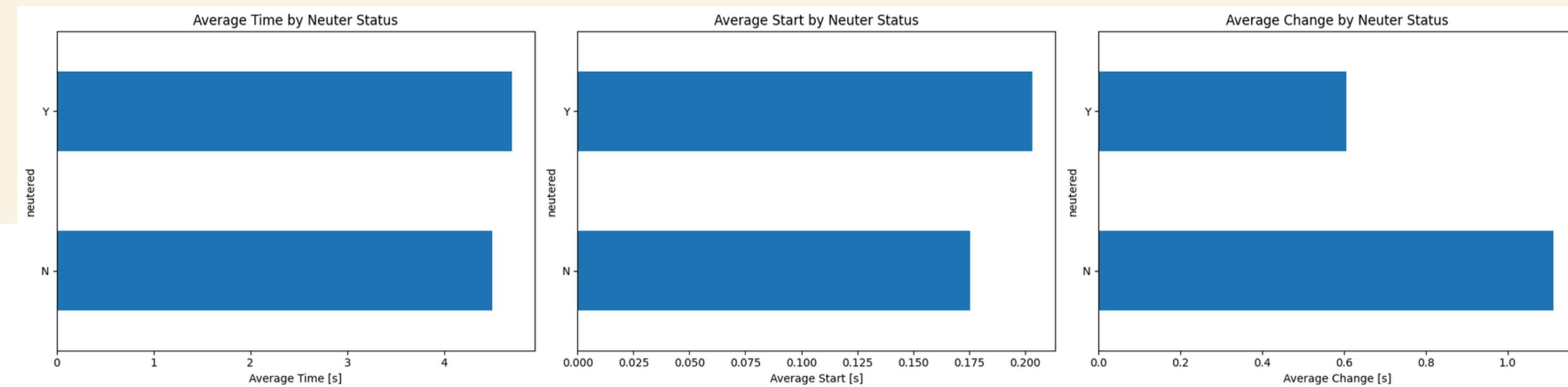
- Smaller dogs run slower, they should run in another division
- Even large dogs run faster if the hurdle is higher
- Slope -0.11 sec/cm

Basic Analysis 3

Female dogs are better in exchange

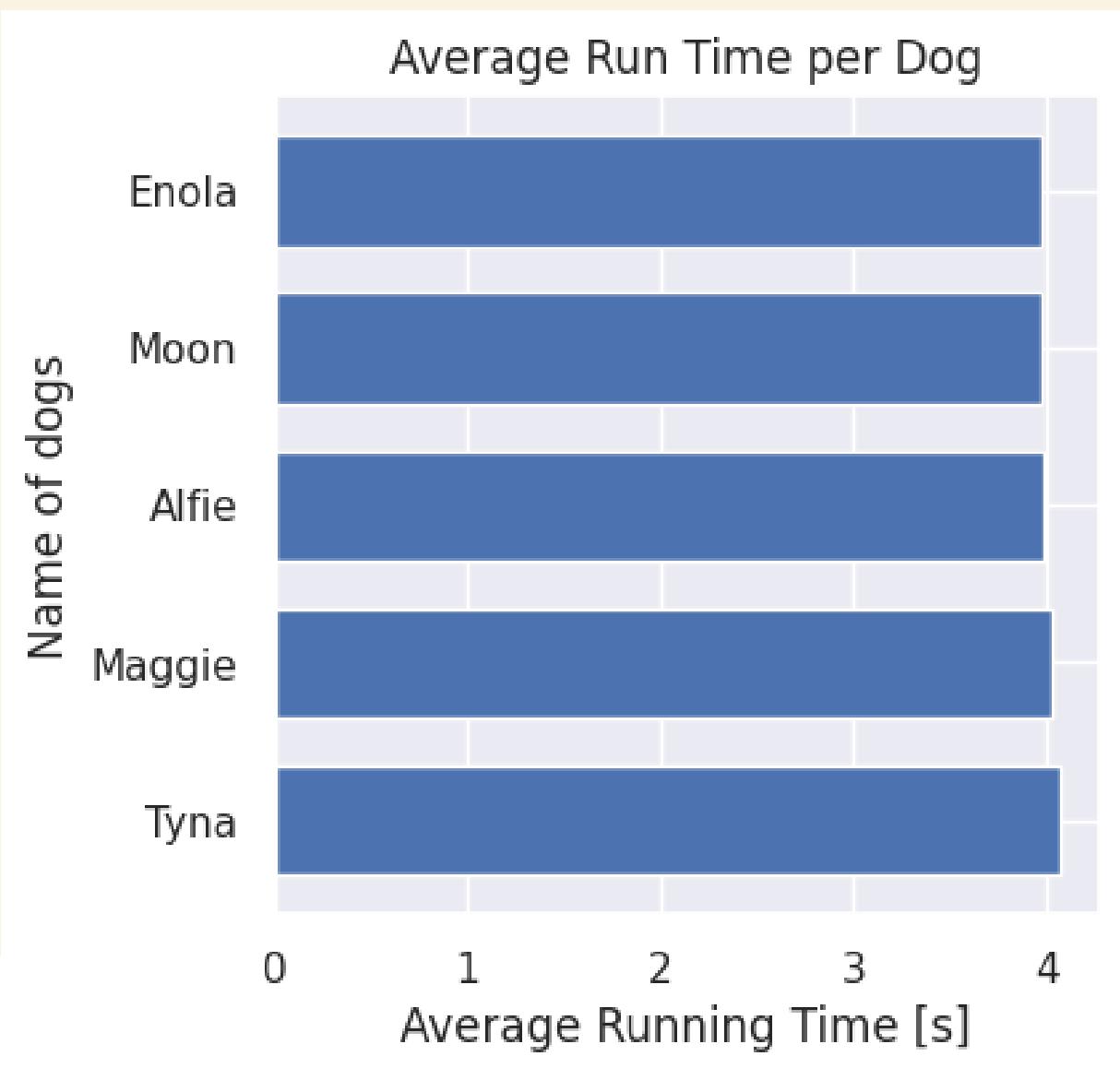


Neutered dogs are better in exchange

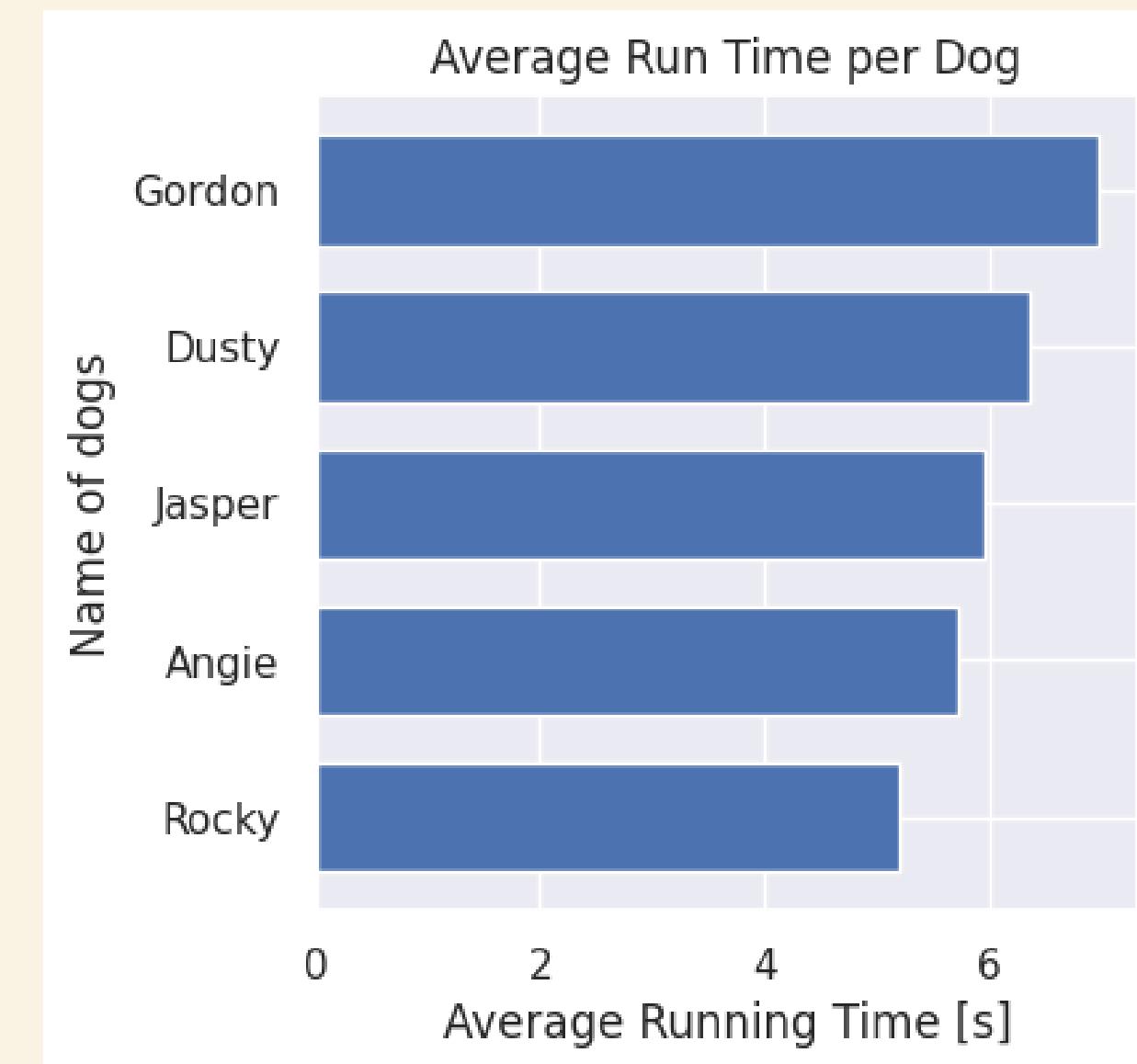


Dog Performance 1

Fastest dogs

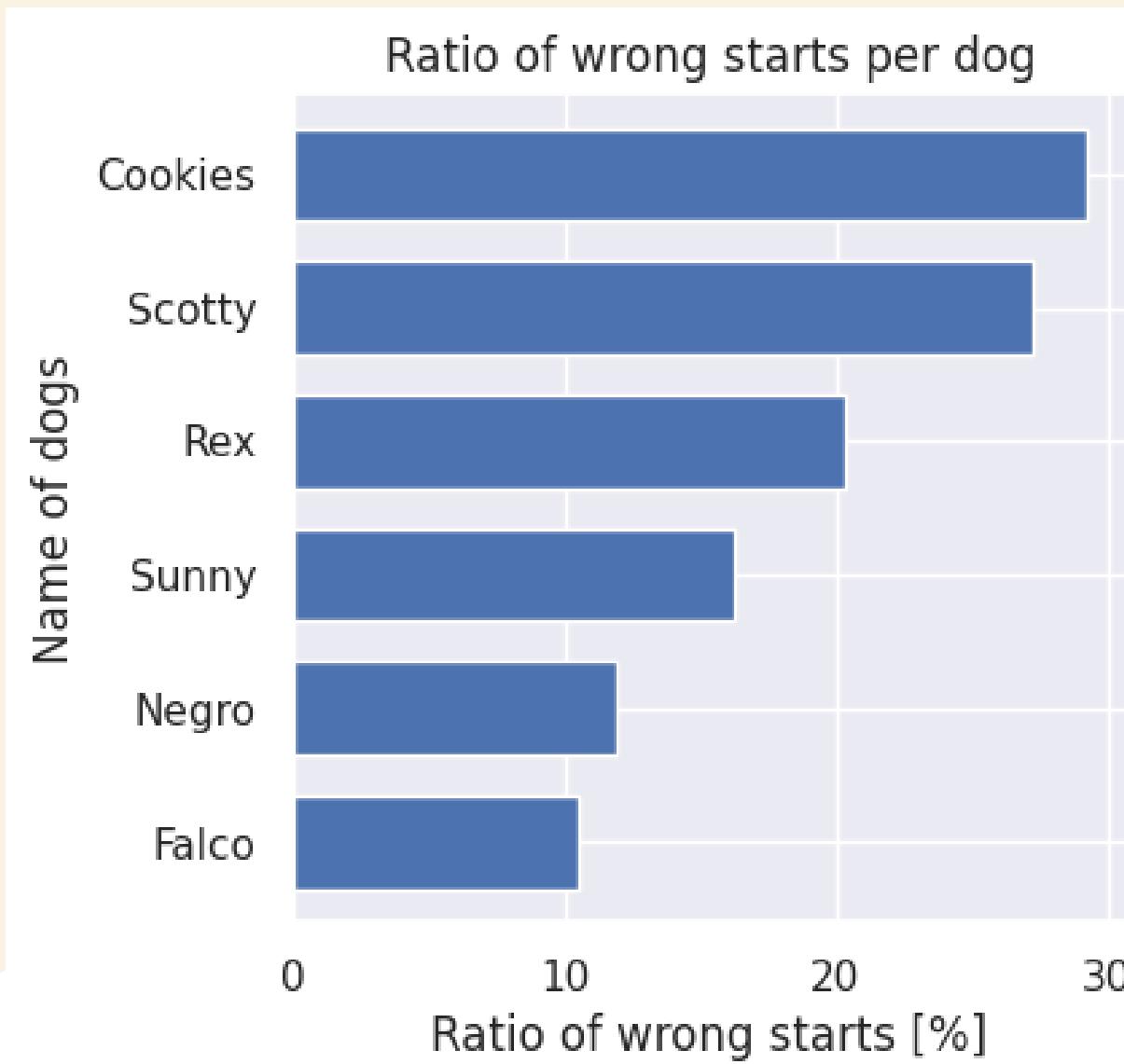


Slowest dogs

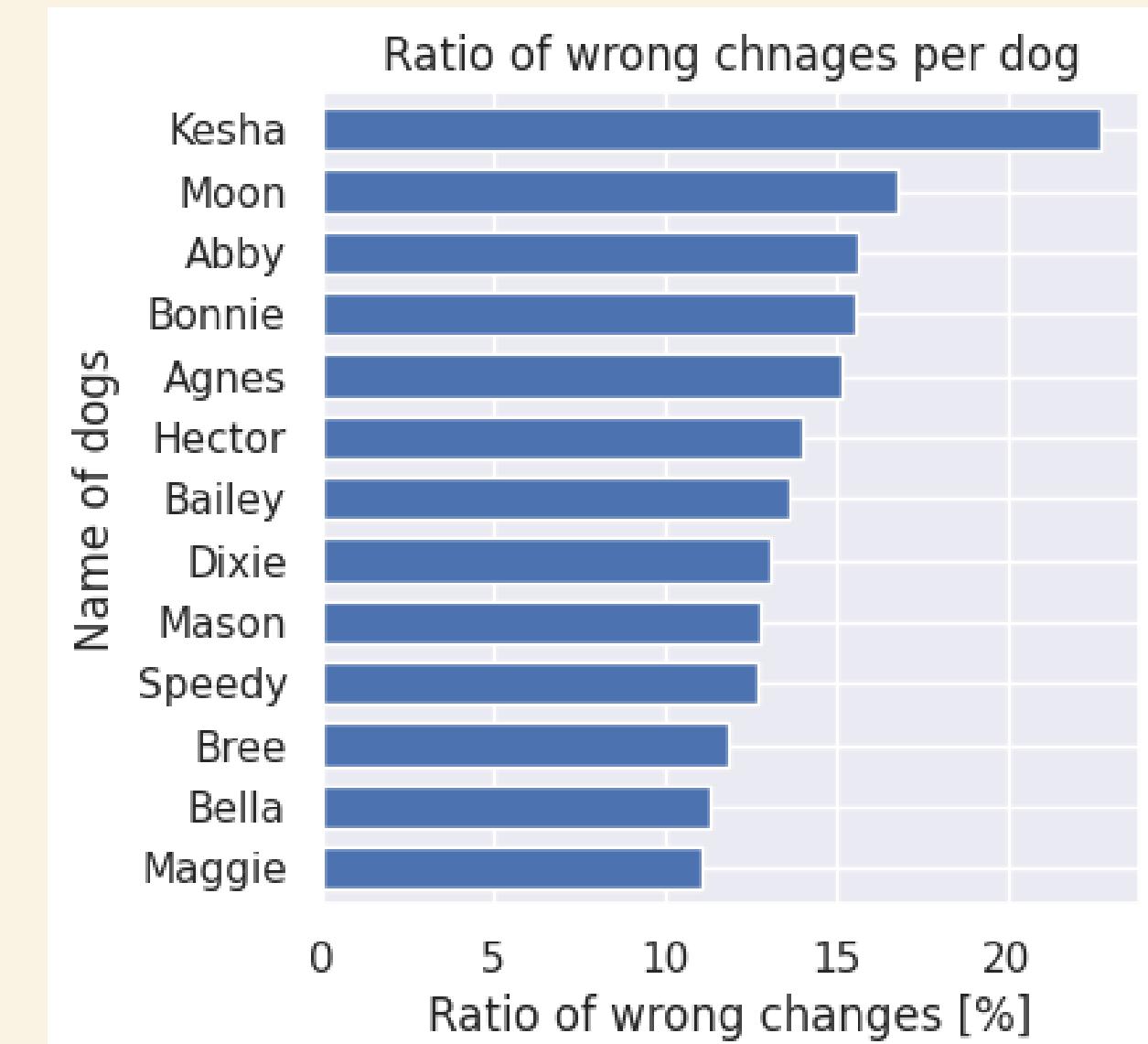


Dog Performance 2

Dogs with high wrong start ratio



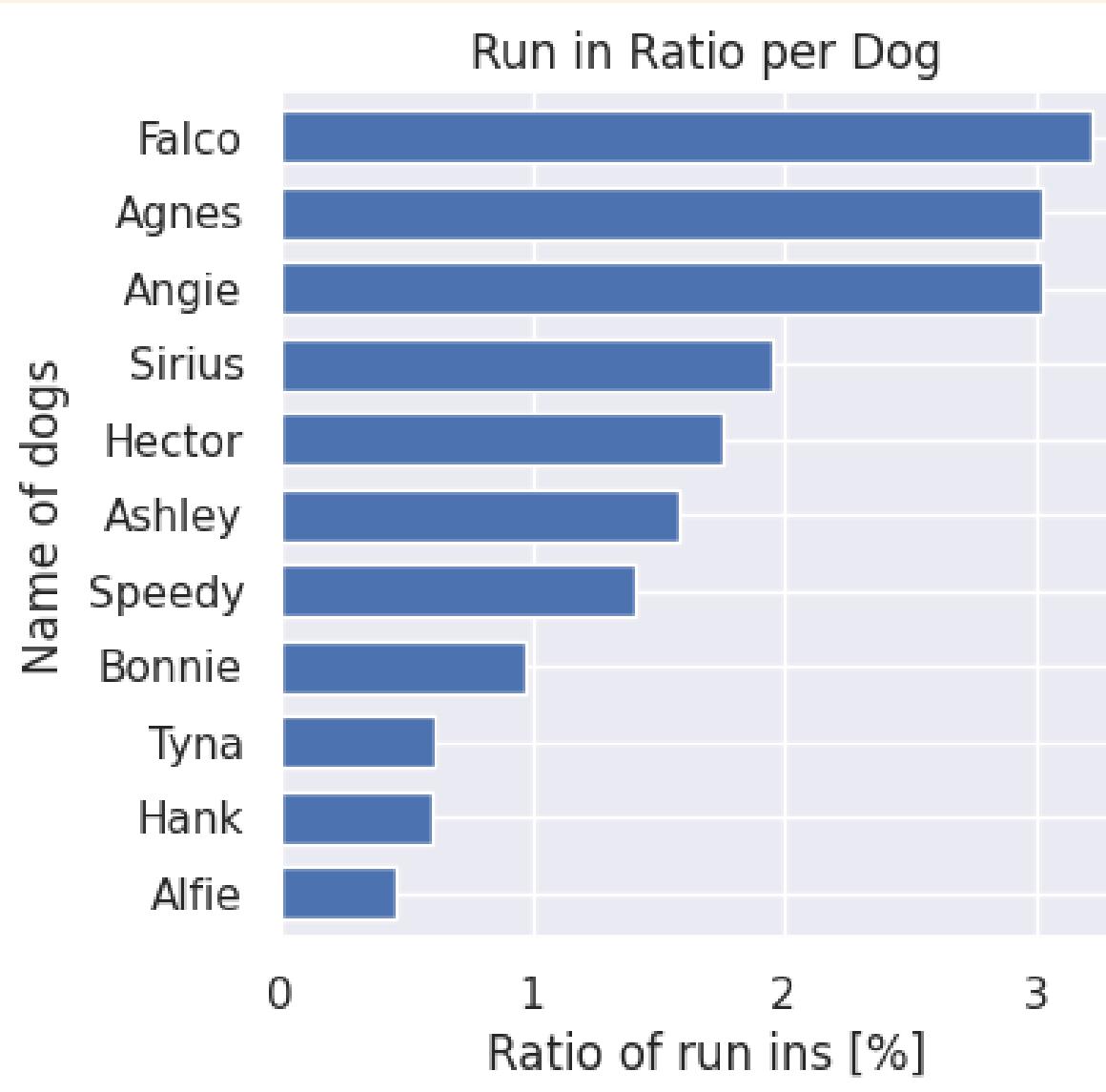
Dogs with high wrong exchange ratio



Dog Performance 3

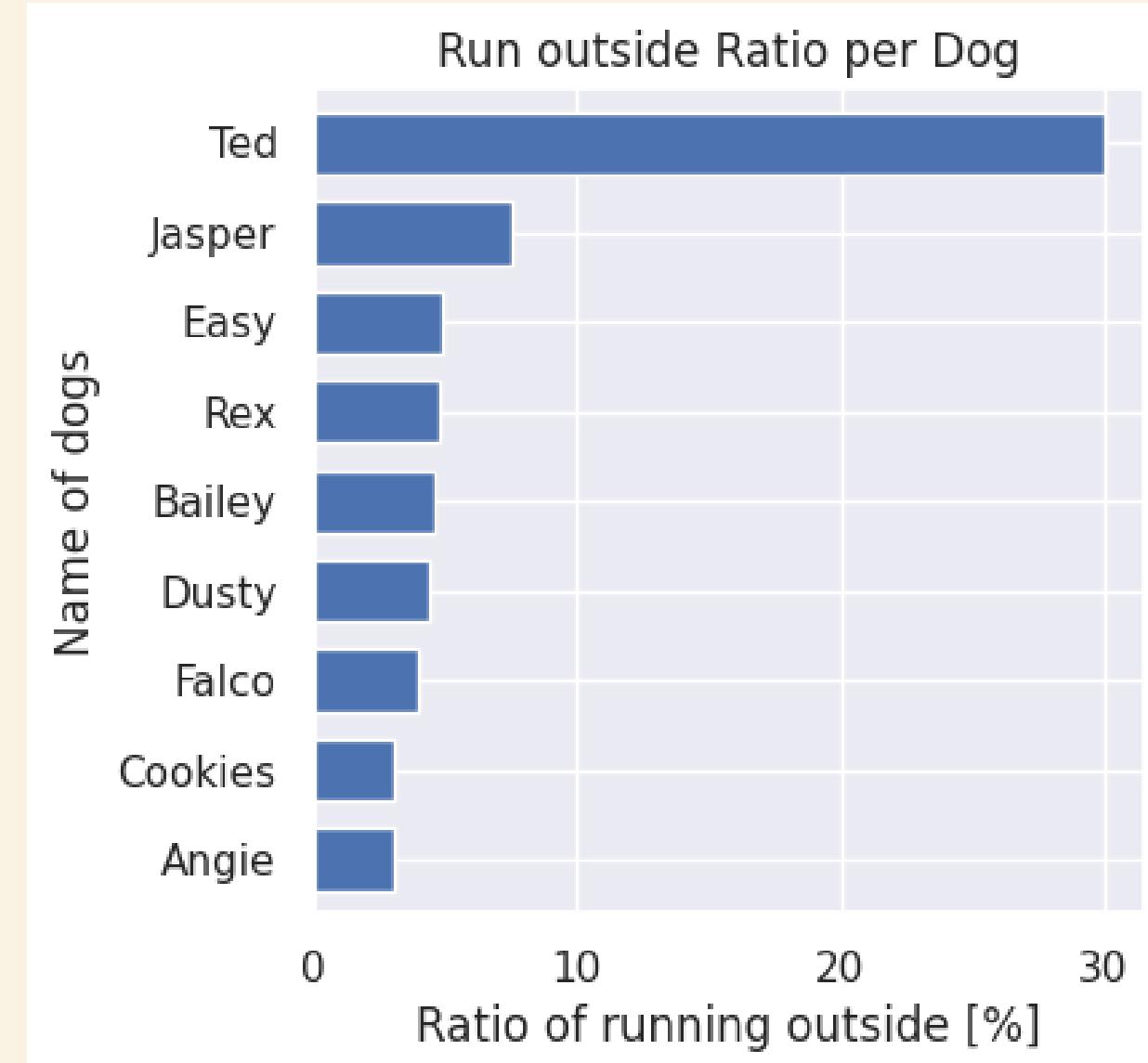
Run-in is a rare mistake

just a few dogs do that more than twice out of 100 runs



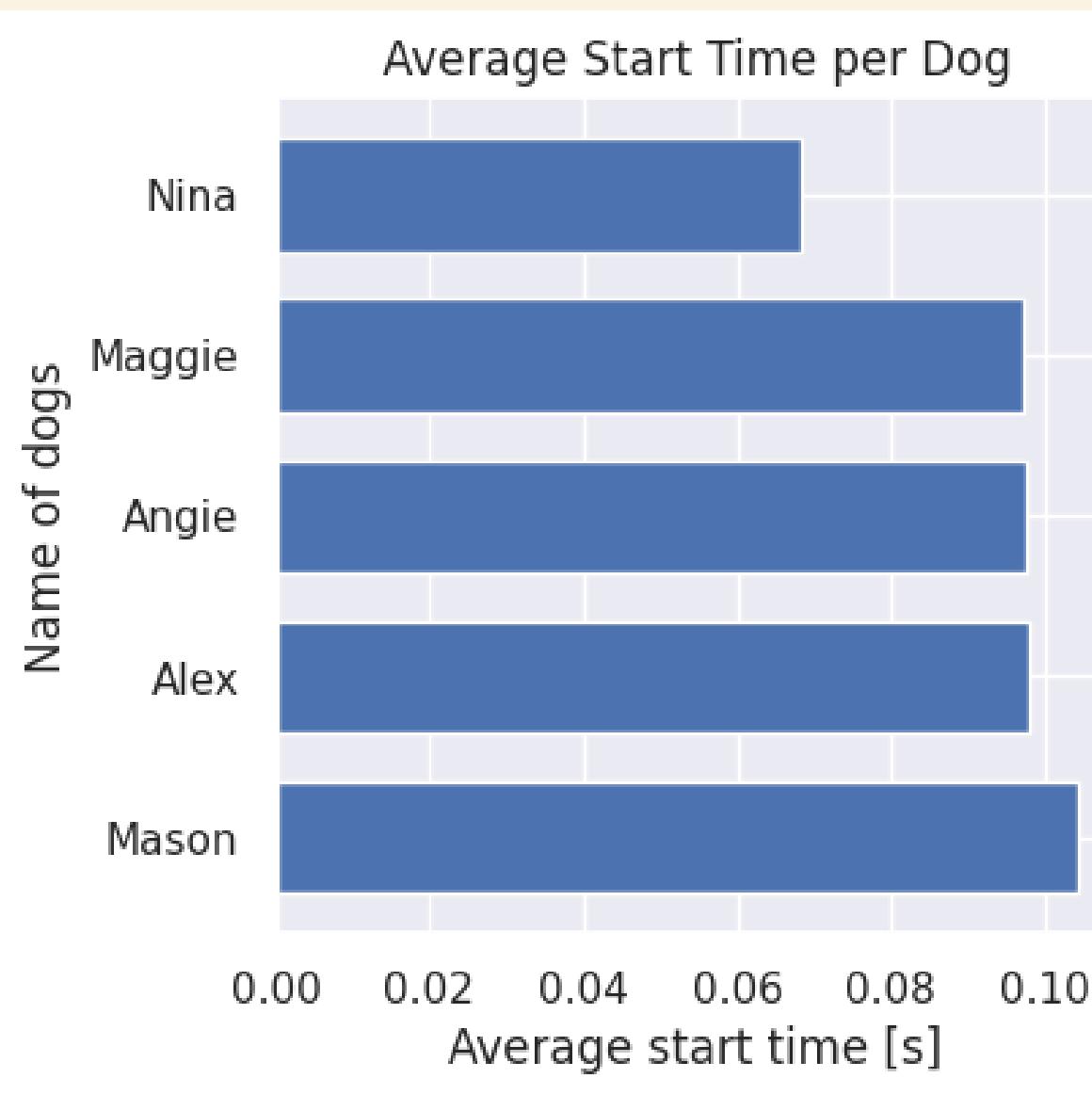
Run outside is more frequent

Ted definitely needs some training about it

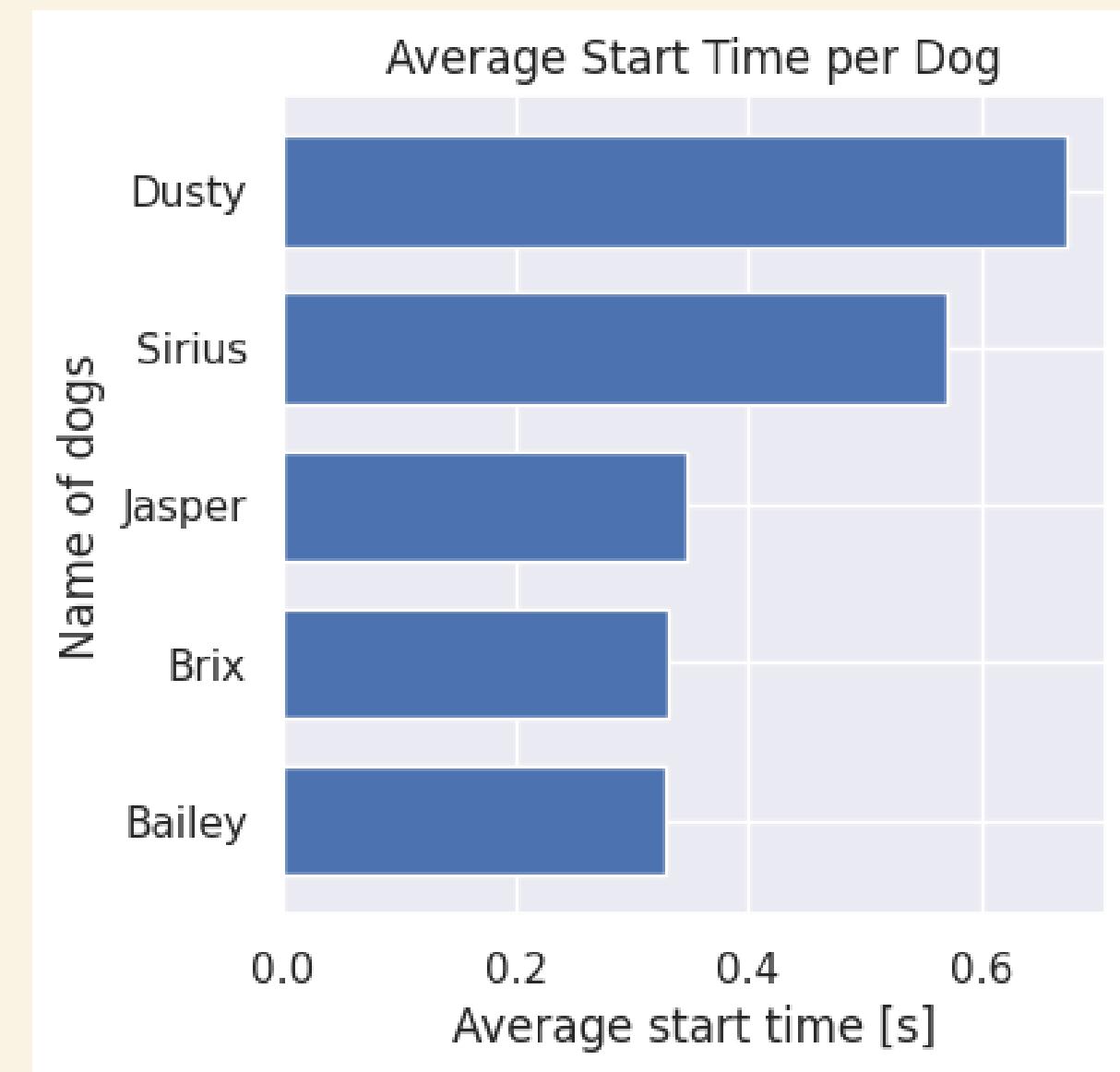


Dog Performance 4

Best starters

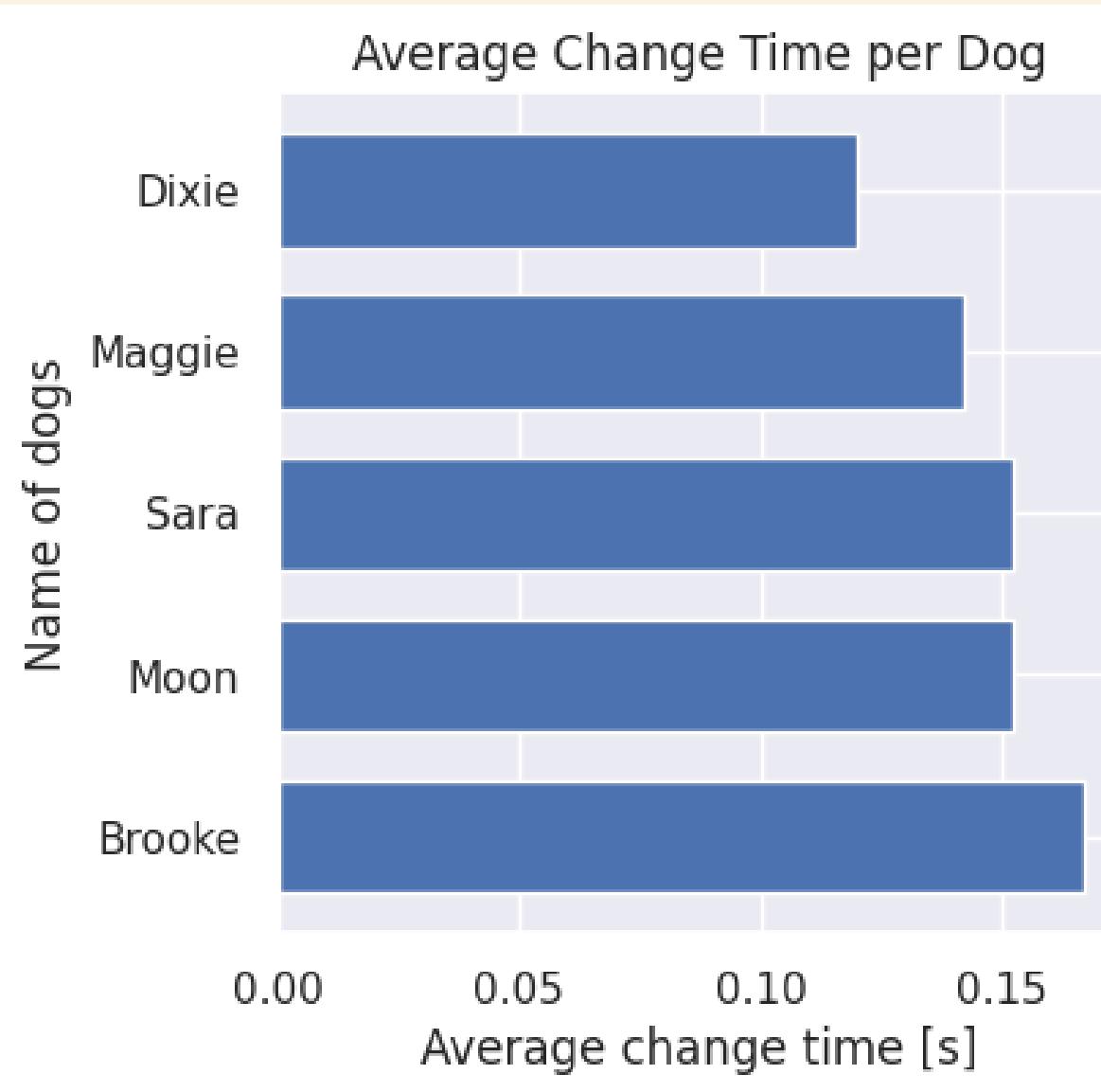


Worst starters

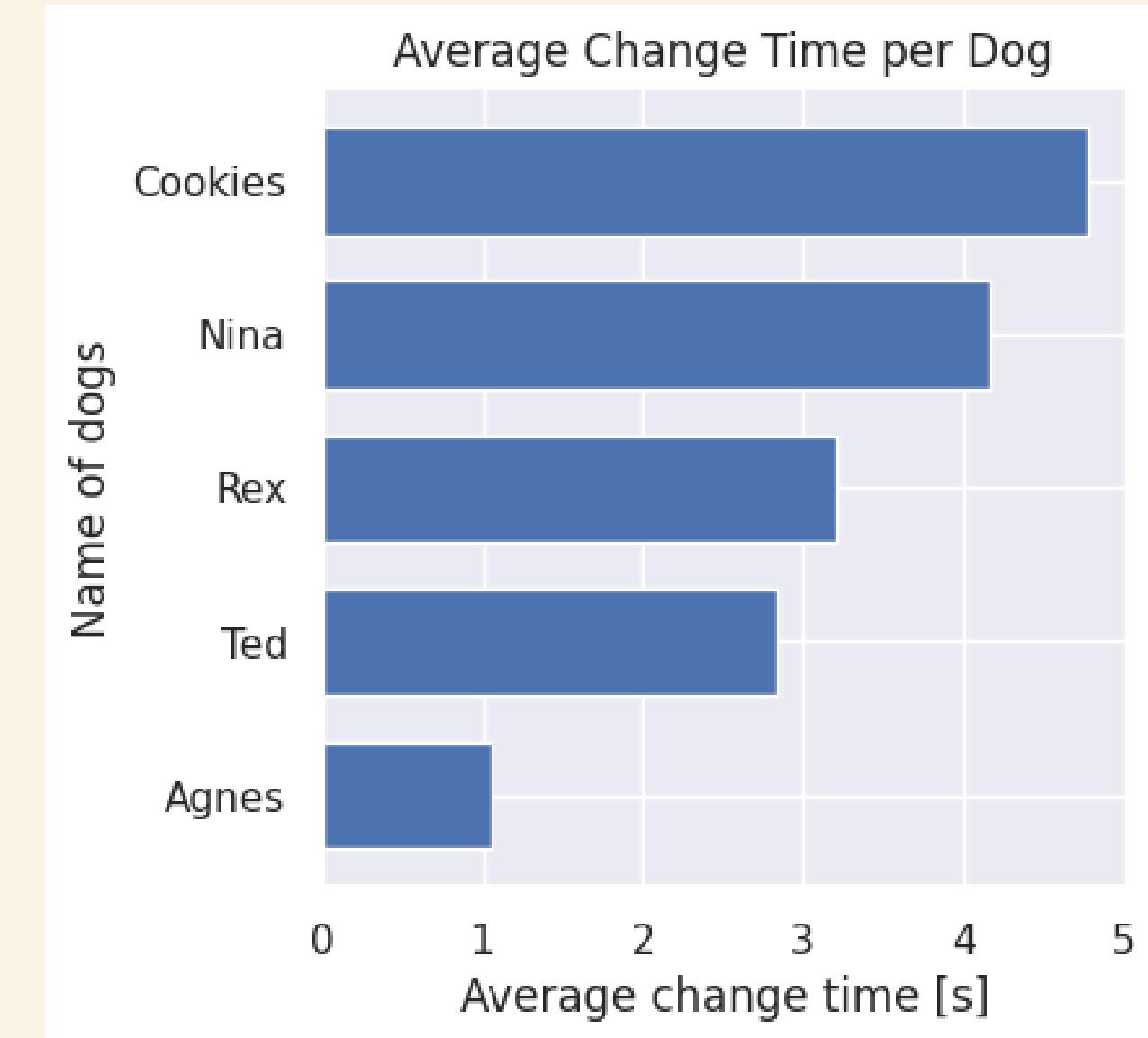


Dog Performance 5

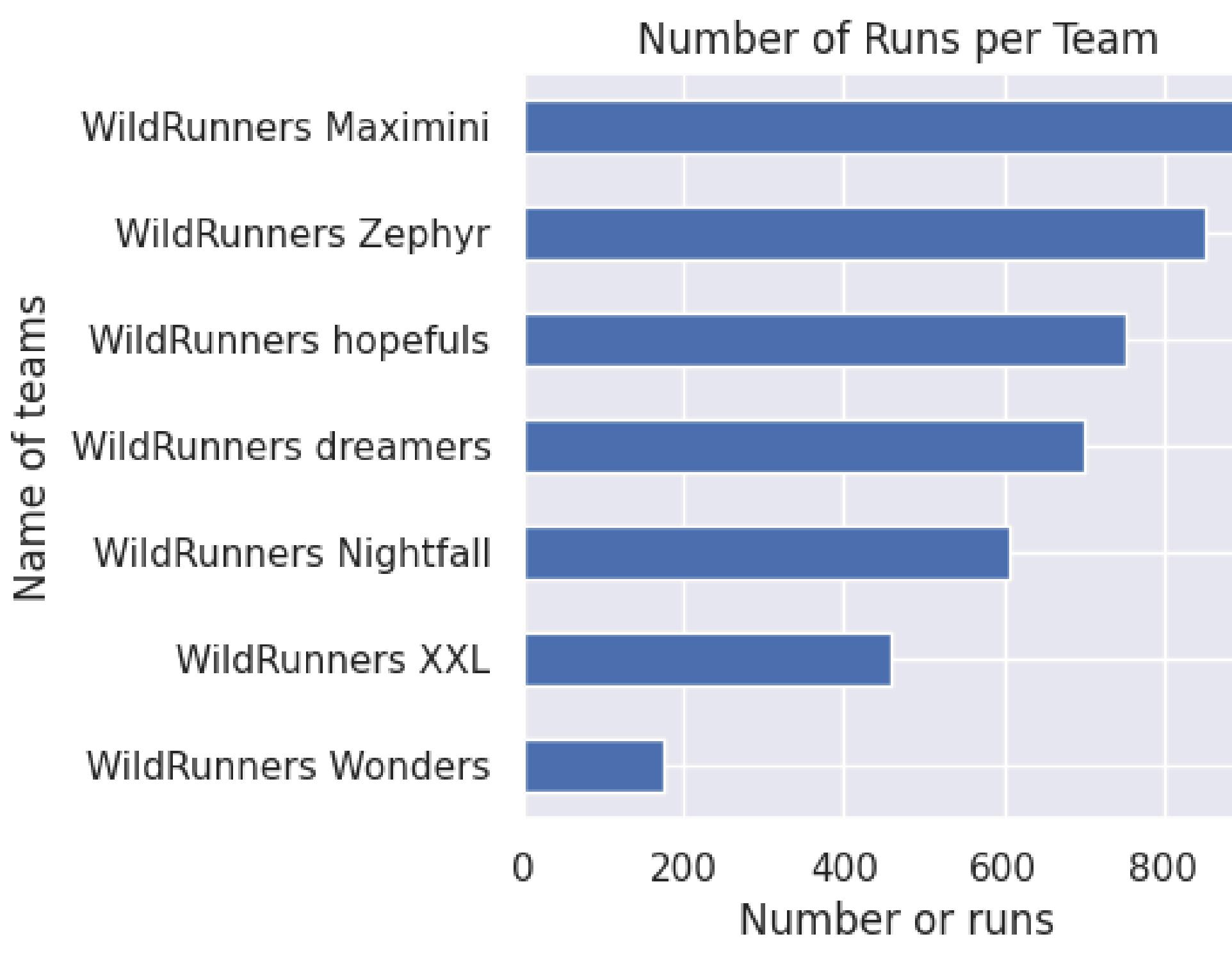
Best exchangers



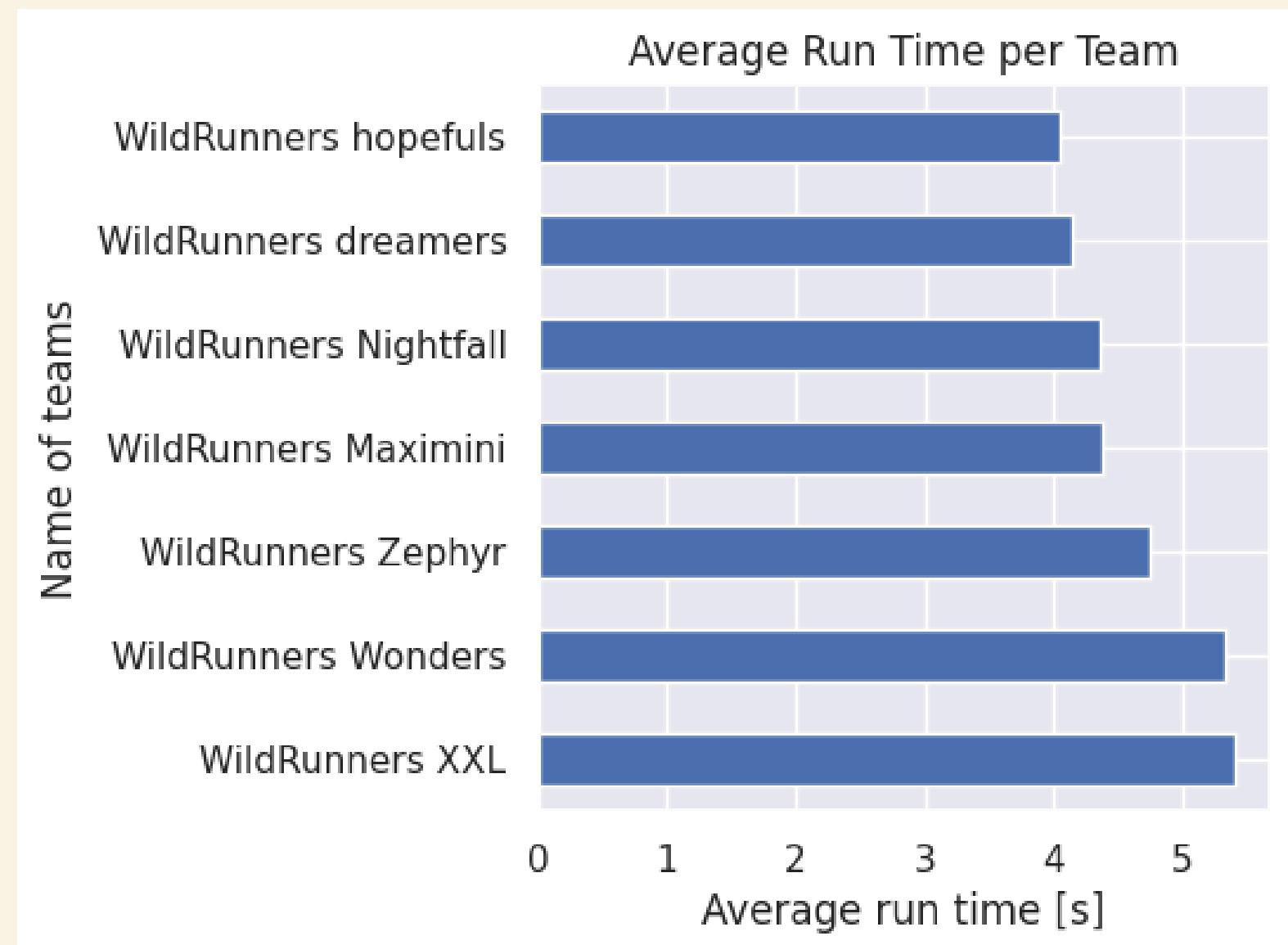
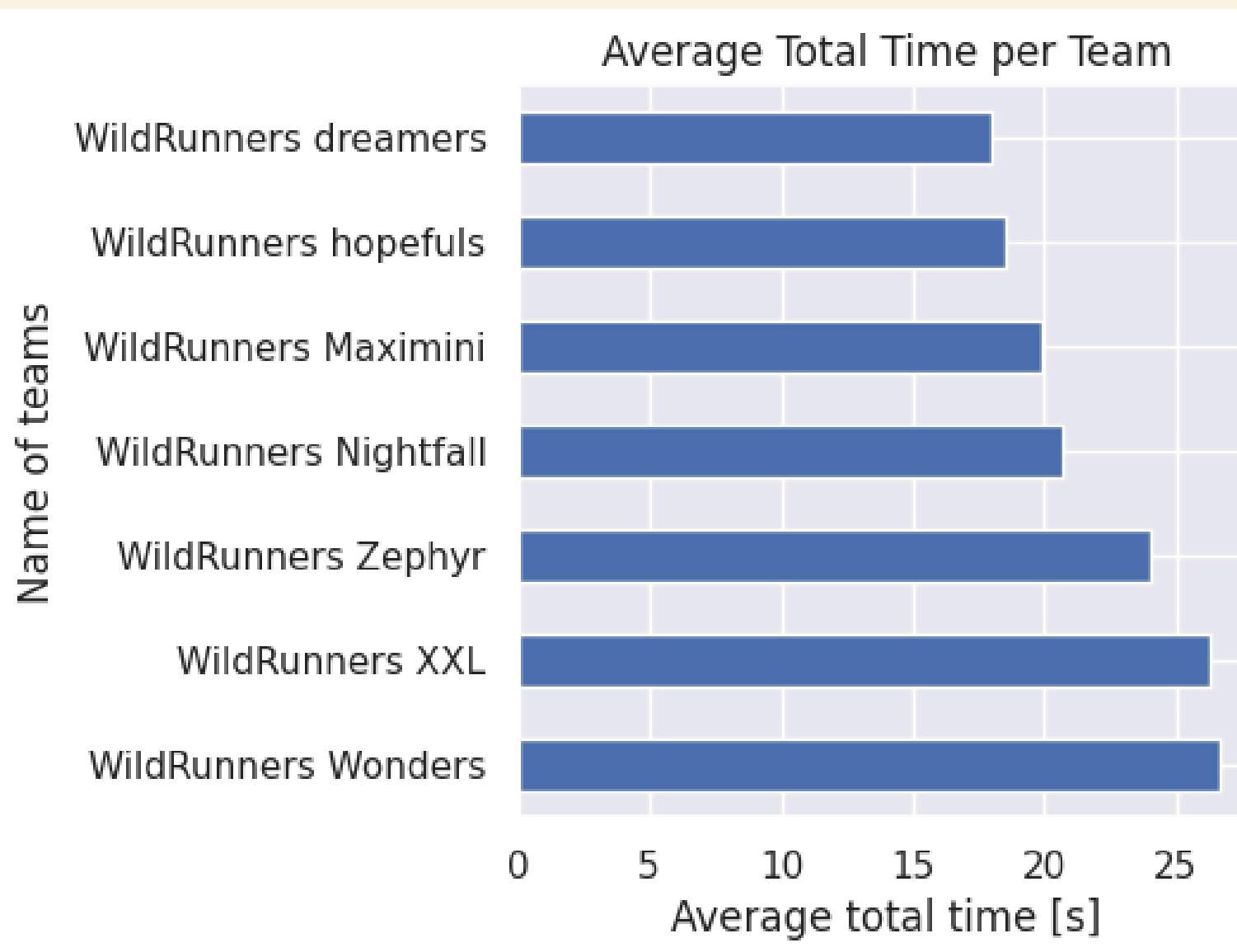
Worst exchangers



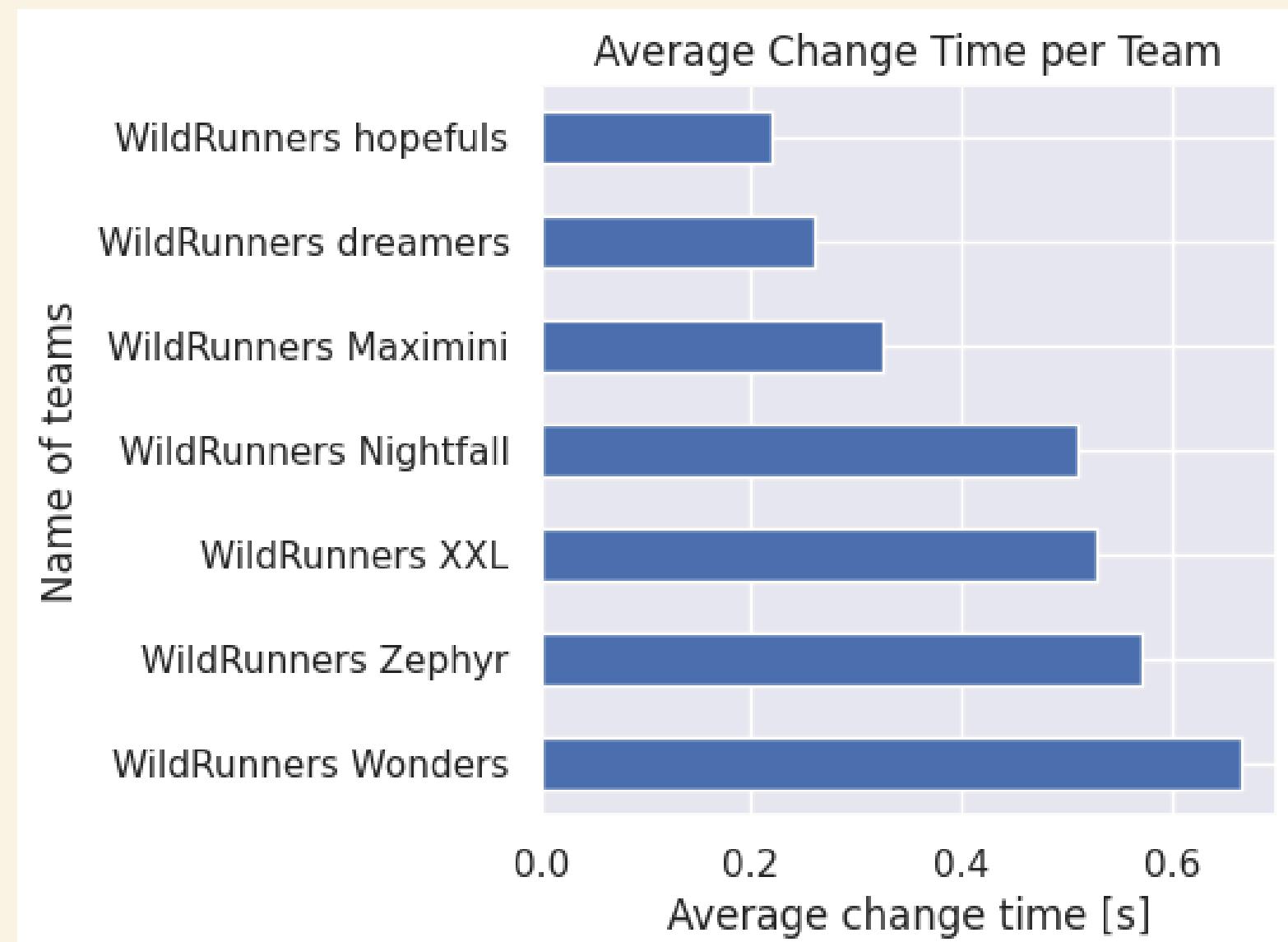
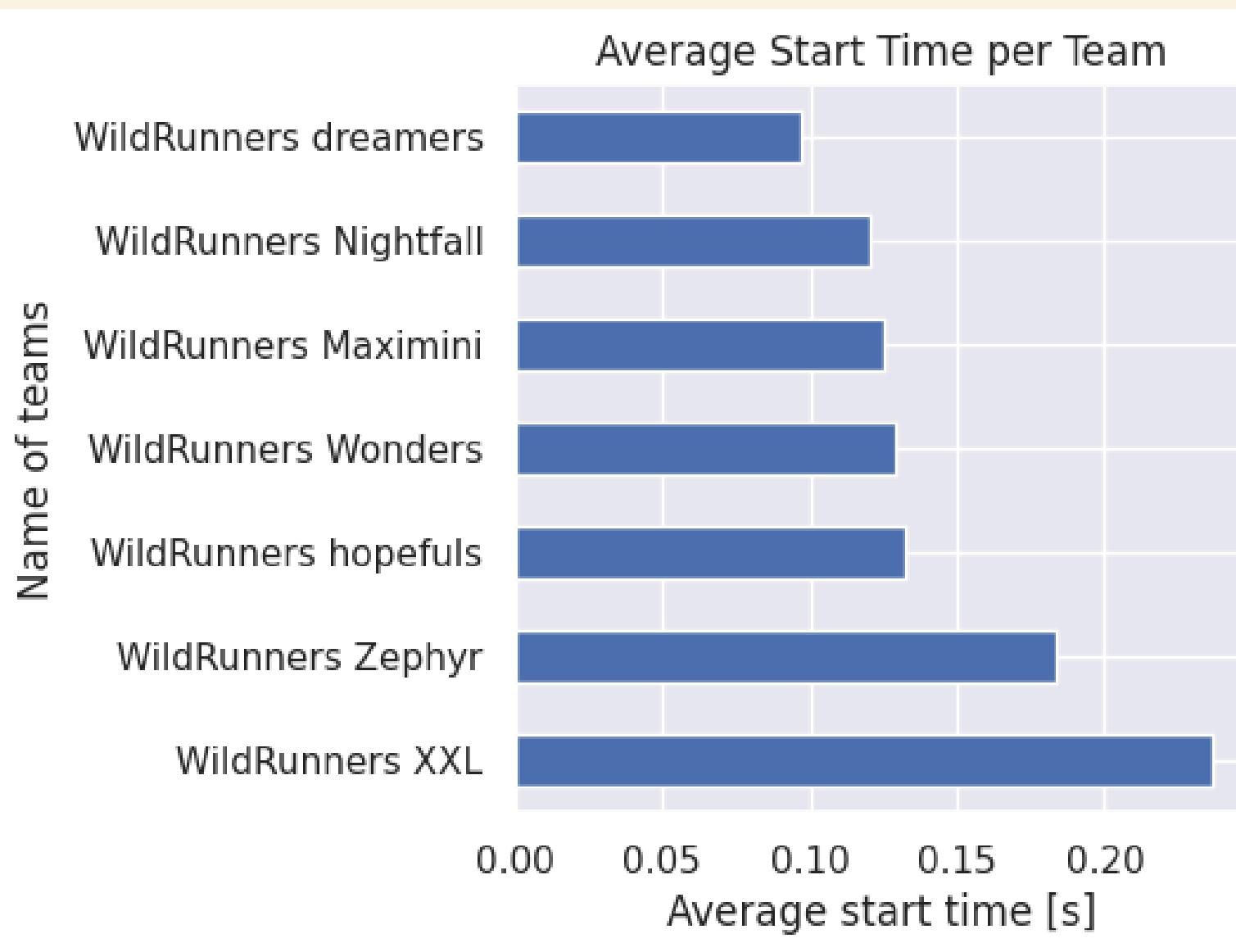
Team Analysis 1



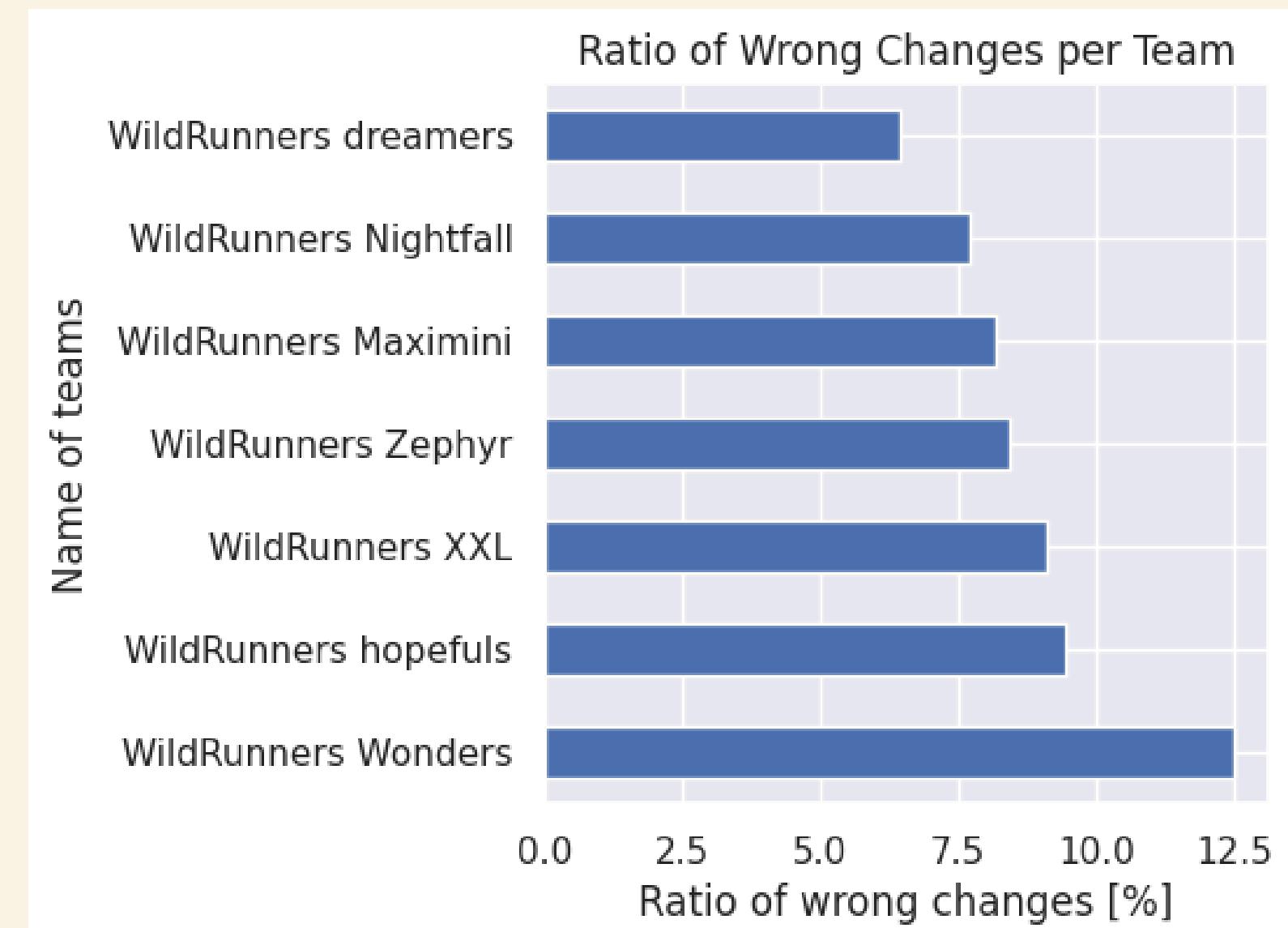
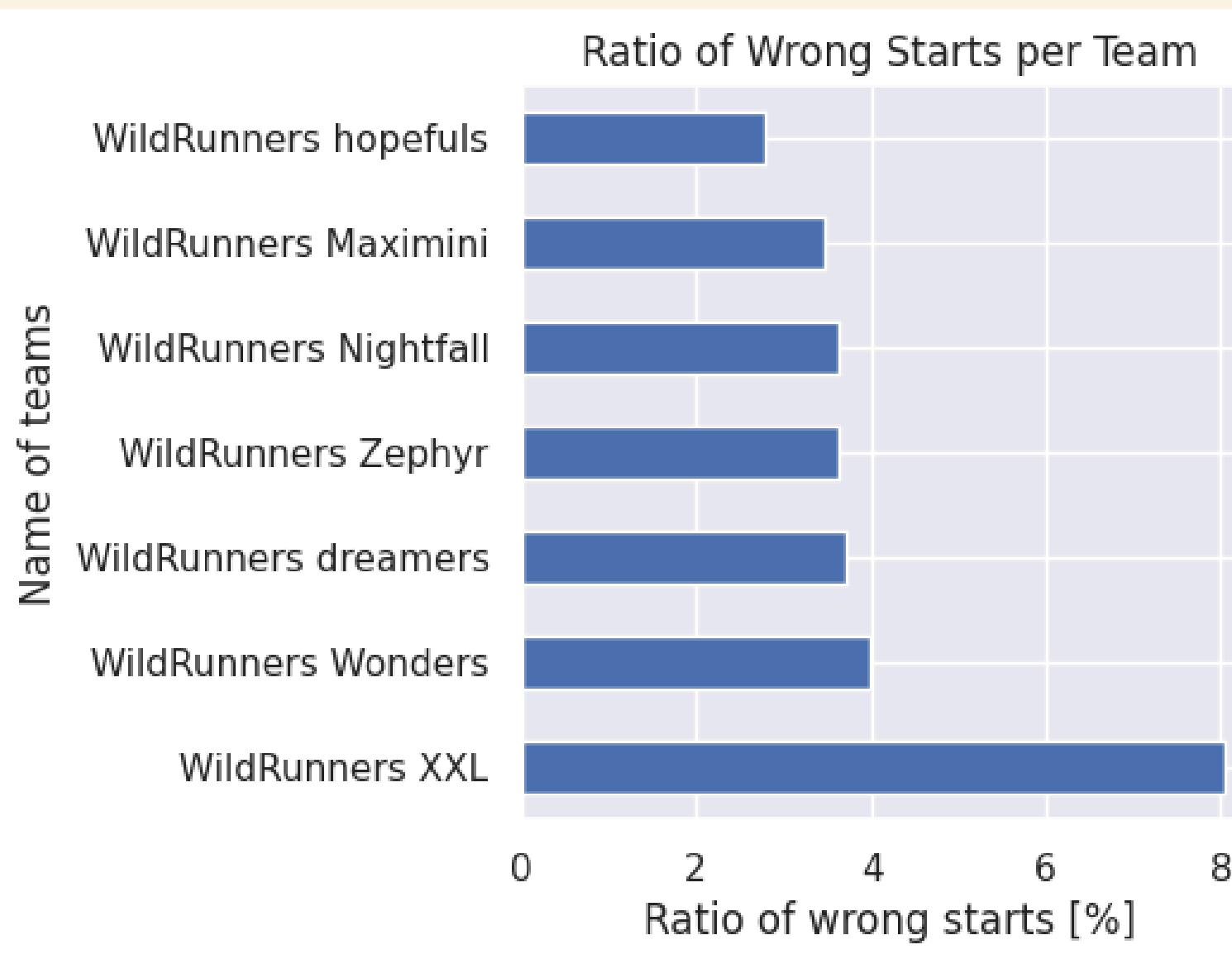
Team Analysis 2



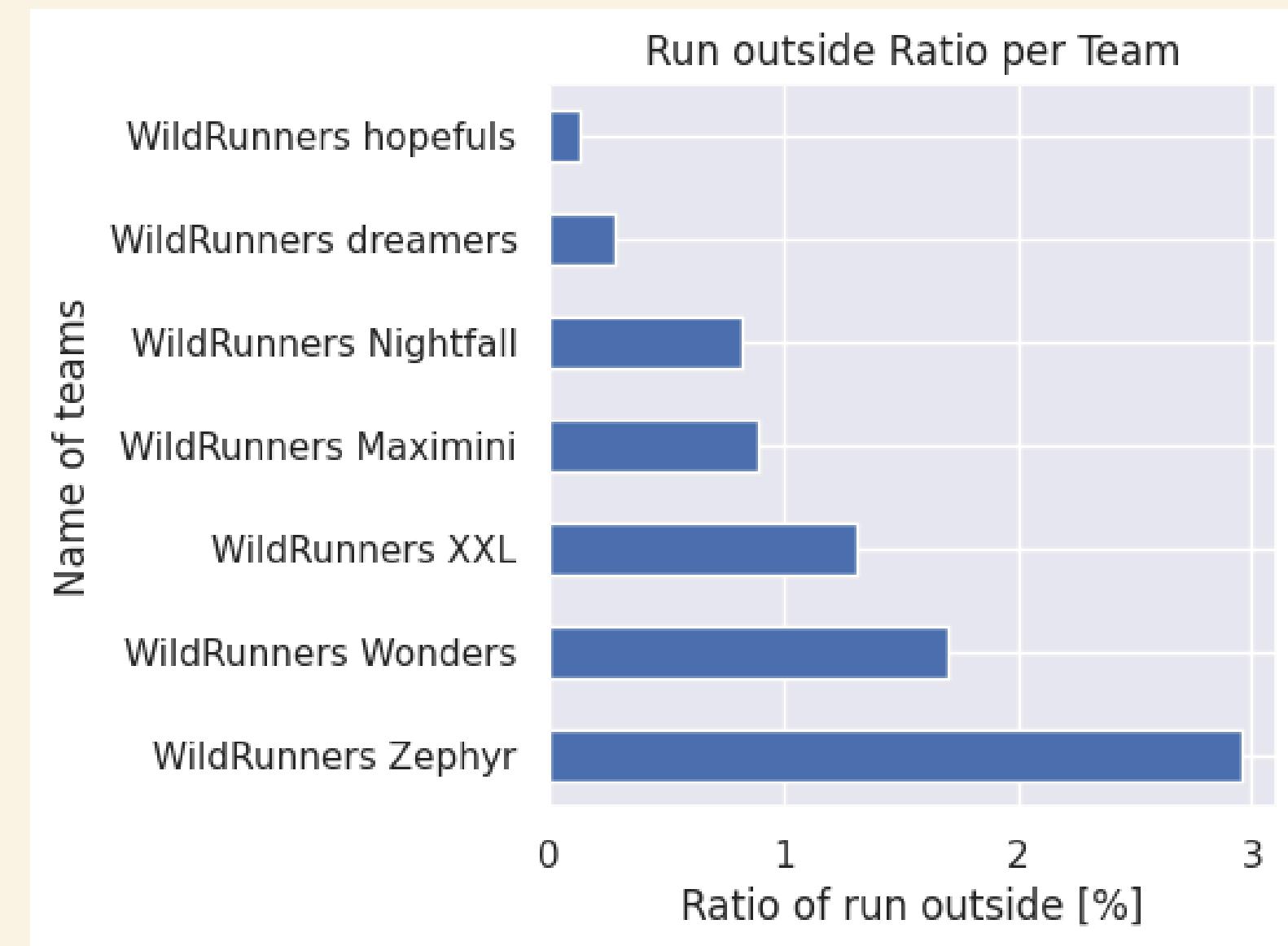
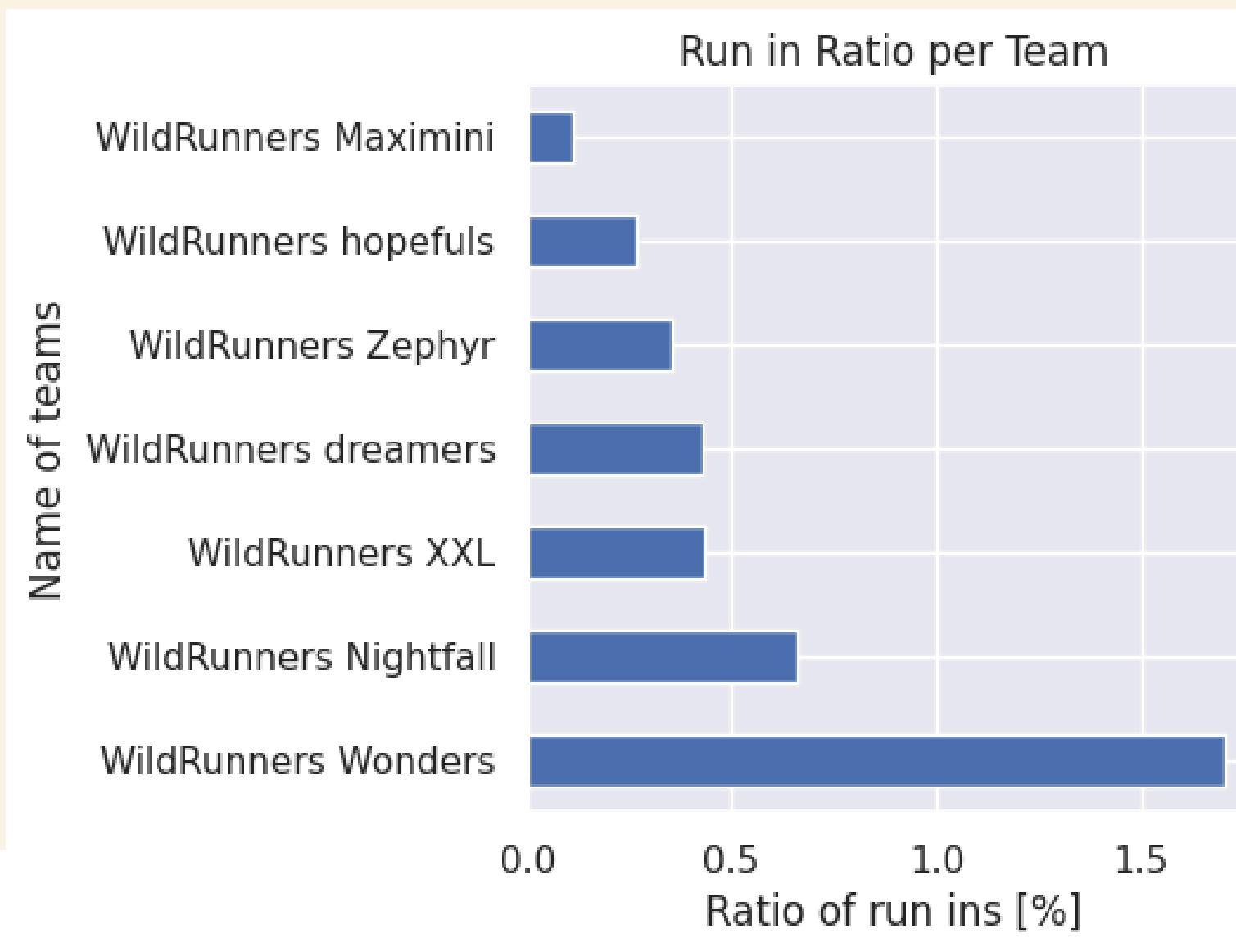
Team Analysis 3



Team Analysis 4



Team Analysis 5



Course Color

There is no significant difference in the running performance on red and blue courses

Average run time:

Red: 4.39 s

Blue: 4.40 s

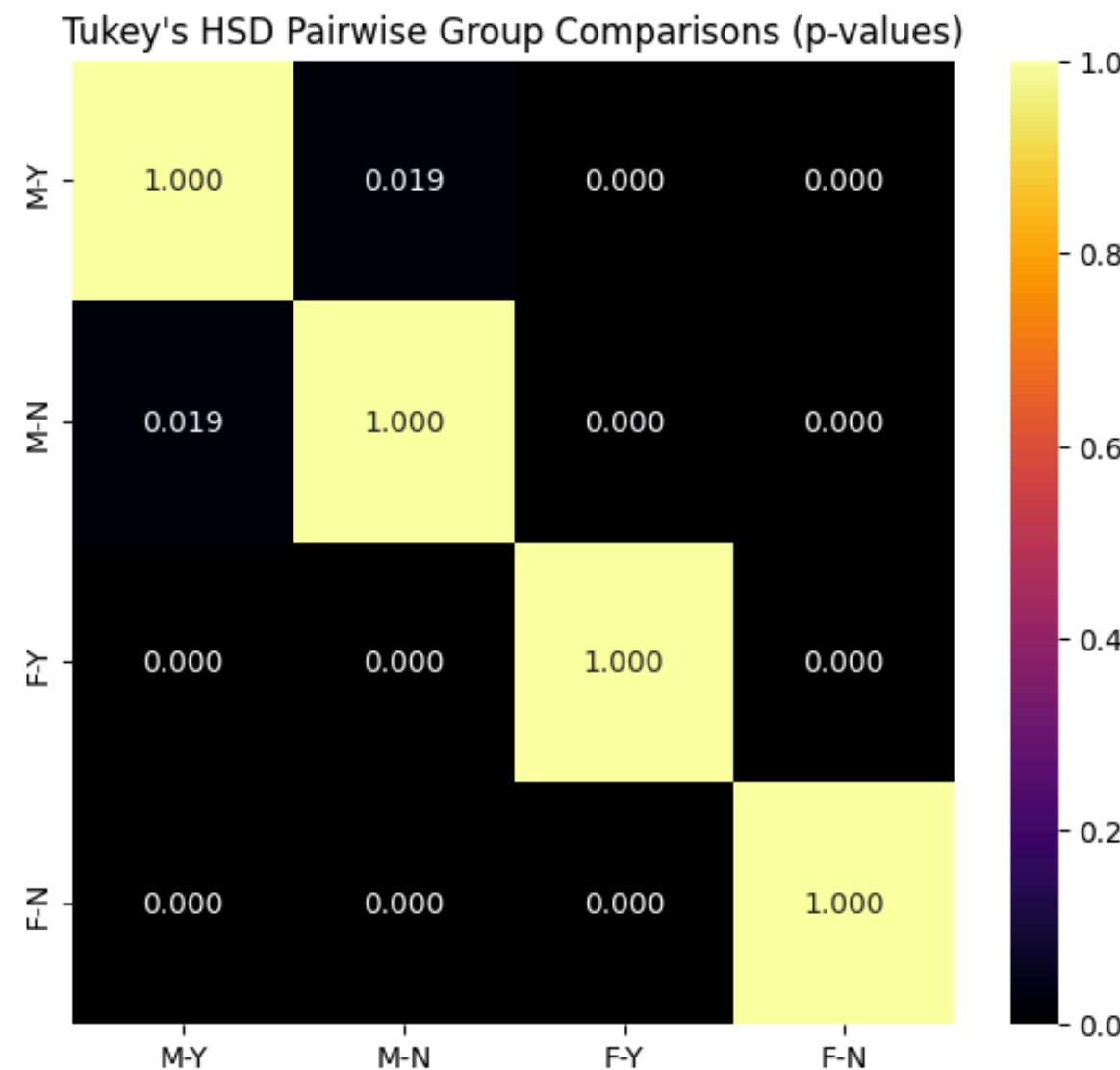
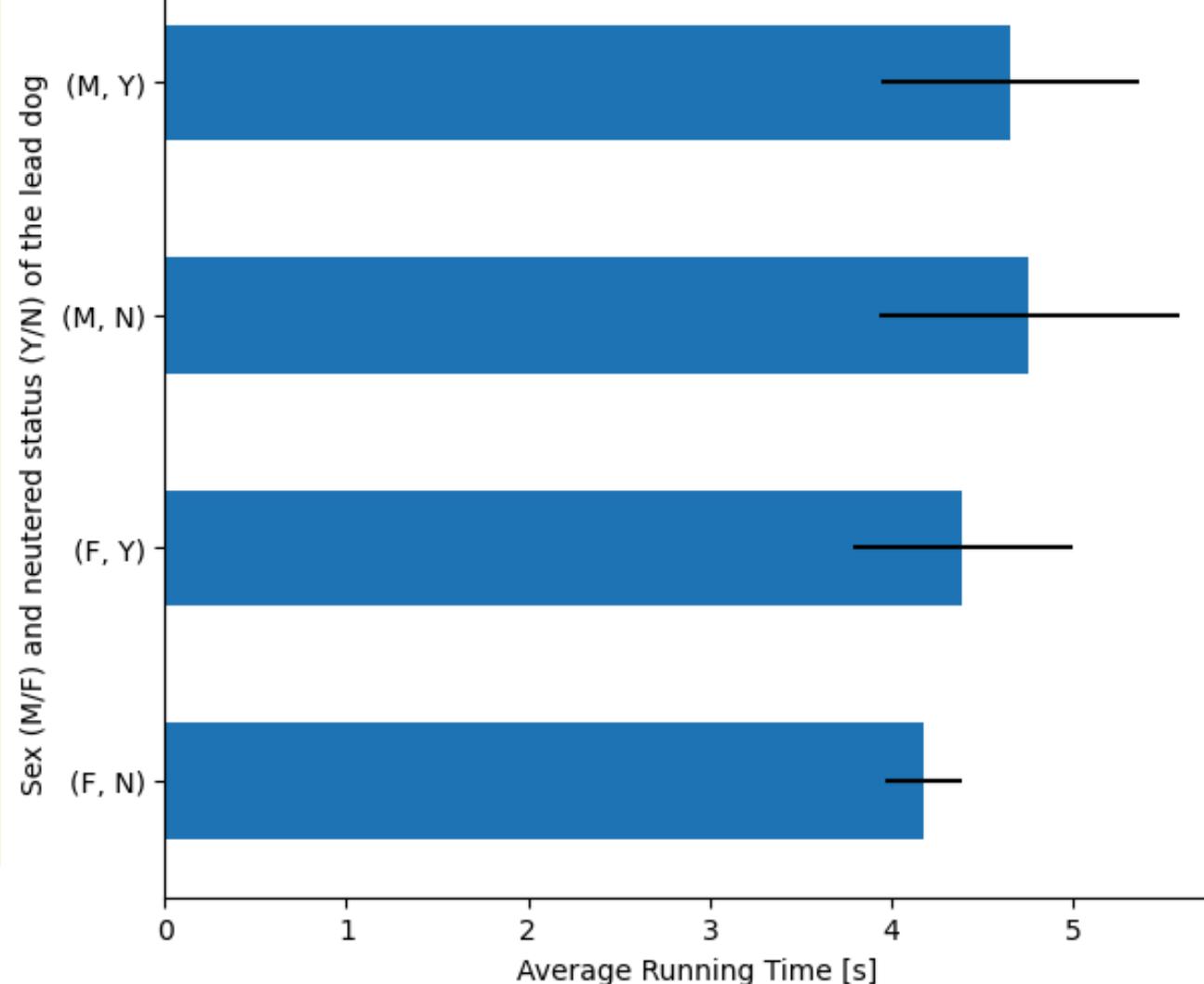
P-value from T-test:

0.33 – we failed to reject the null-hypothesis that there is no significant difference between the two groups

Note: Start and exchange times were not analyzed



Running performance vs. lead dog sex and neuter status all dogs included



ANOVA (Analysis of Variance):
The overall ANOVA p-value of 8E-43 indicates that some of the groups are significantly different from others.

Tukey's HSD (Honestly Significant Difference)
The p-values for all pairs are below 0.05, indicating that all groups differ significantly from each other.

Conclusion:
Dogs that exchange with a female dog, especially one that is not neutered, run significantly faster than otherwise.

Further question:
Is this true for both males and females?

Running performance vs. lead dog sex and neuter status

uneutered males - neutered males

Unneutered male dogs (MN)

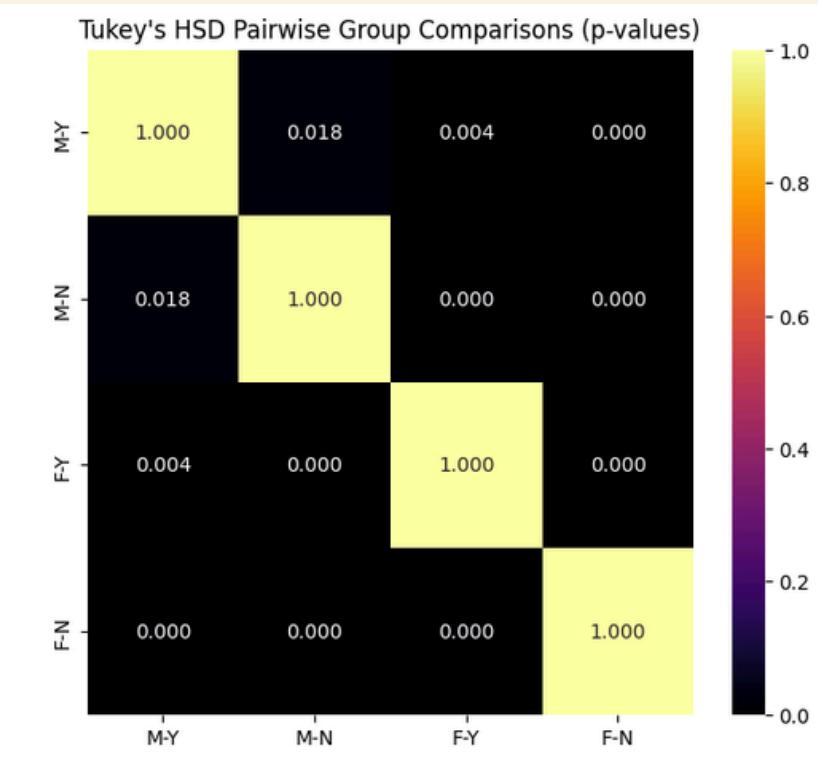
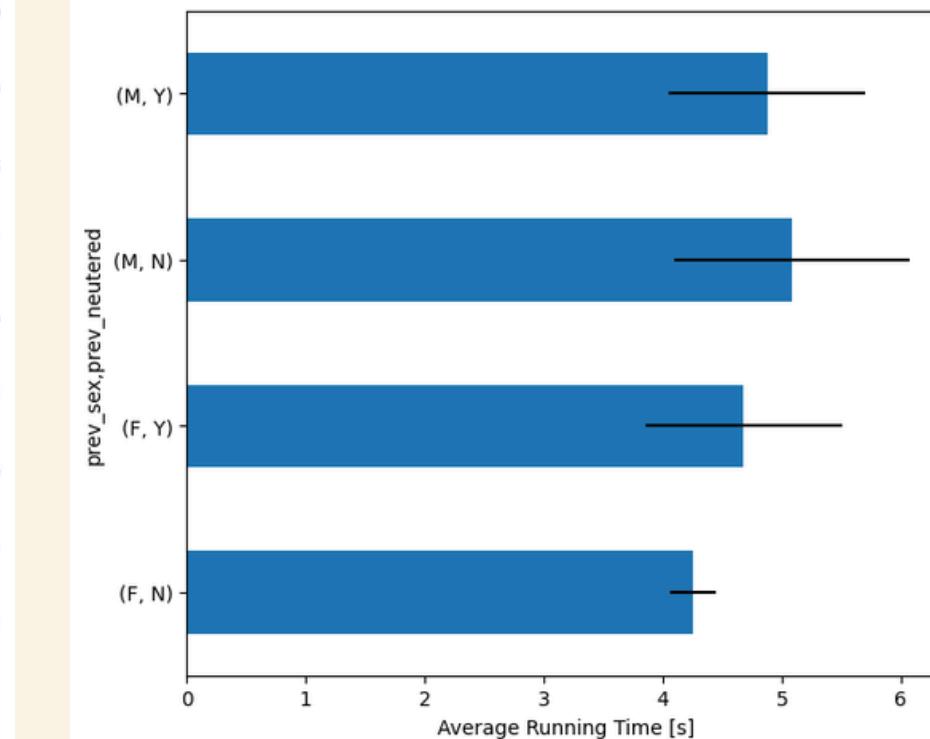
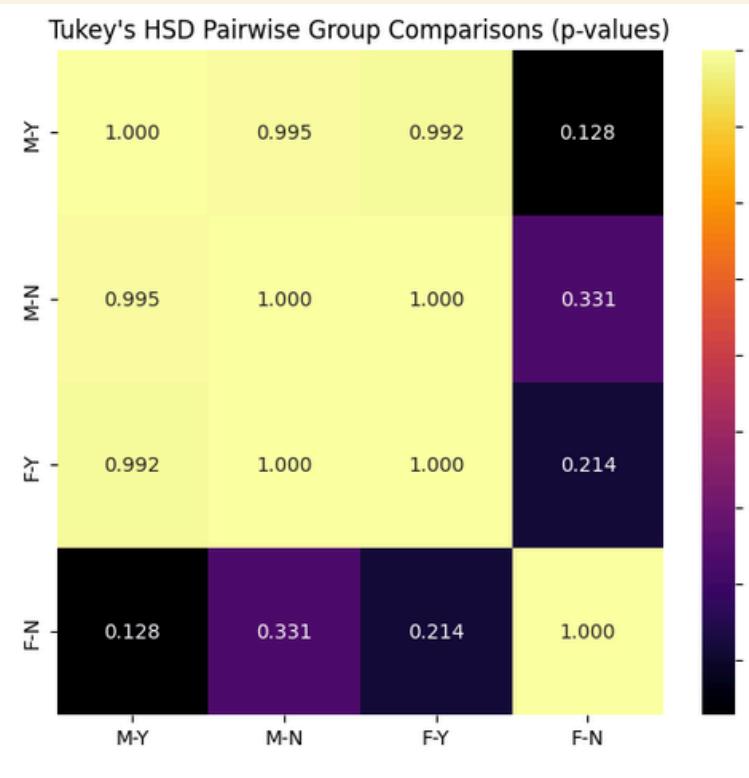
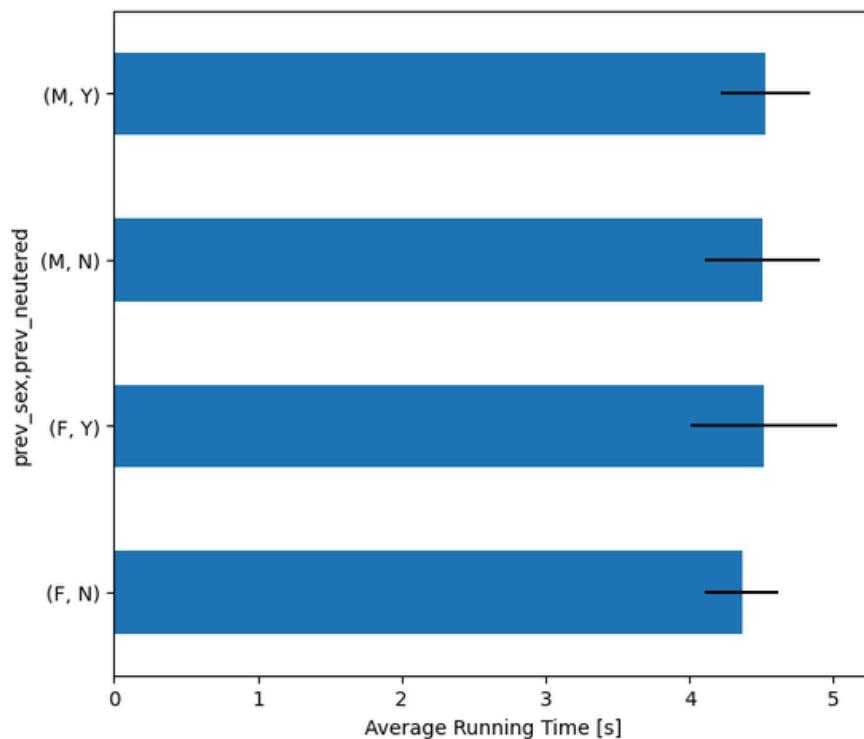
The overall ANOVA p-value of 0.18 indicates that there is no significant difference between the four groups

Surprising results. Note that the group sizes are little.

Neutered male dogs (MY)

The results are similar to the overall population. Each pair of groups differs from all the others.

Exchanging with female unneutered dogs gives the best results.



Running performance vs. lead dog

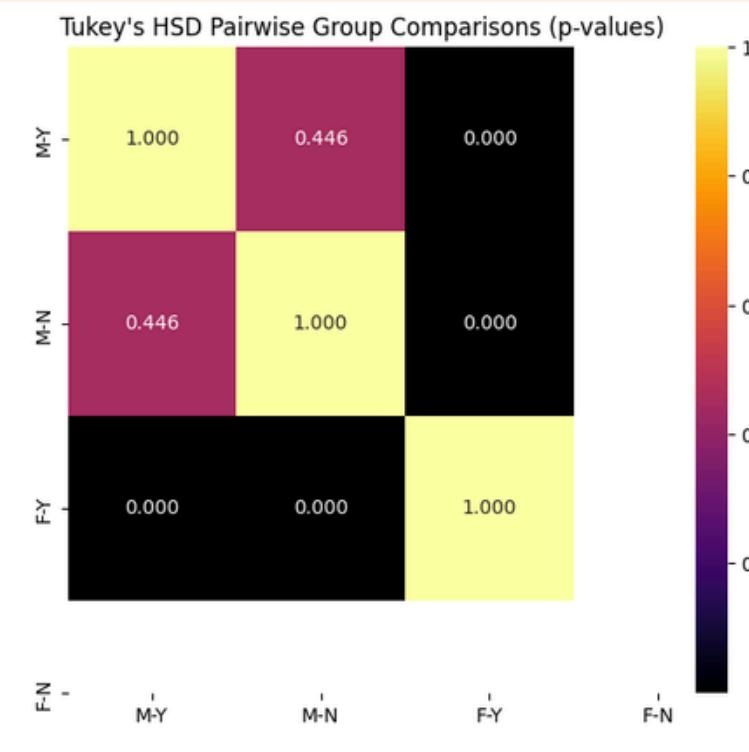
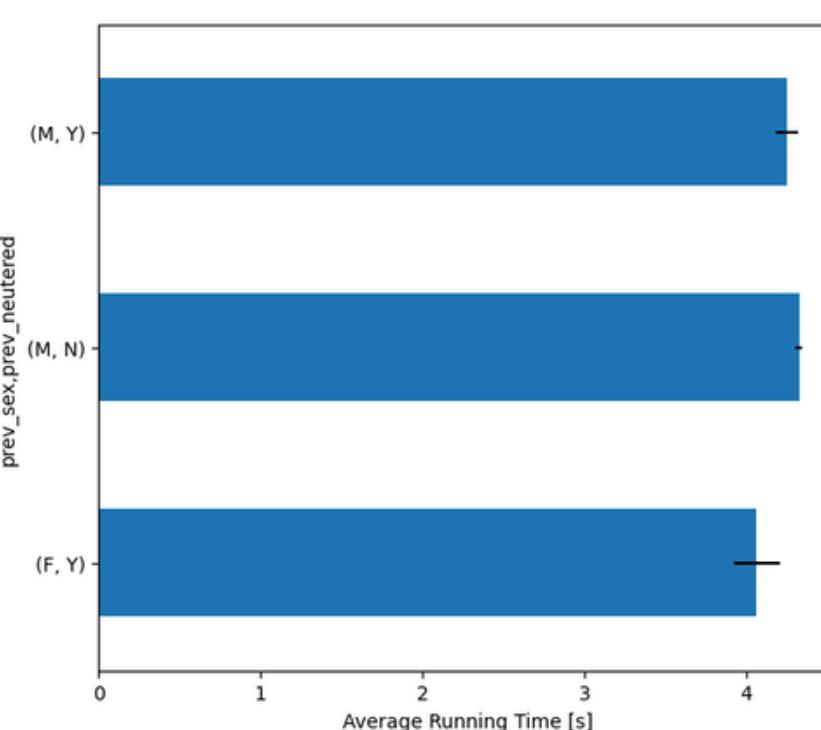
sex and neuter status

uneutered females - neutered females

Unneutered female dogs (FN)

Very little group sizes! One group is completely missing: there is no such run when an FN dog exchanges with another FN dog.

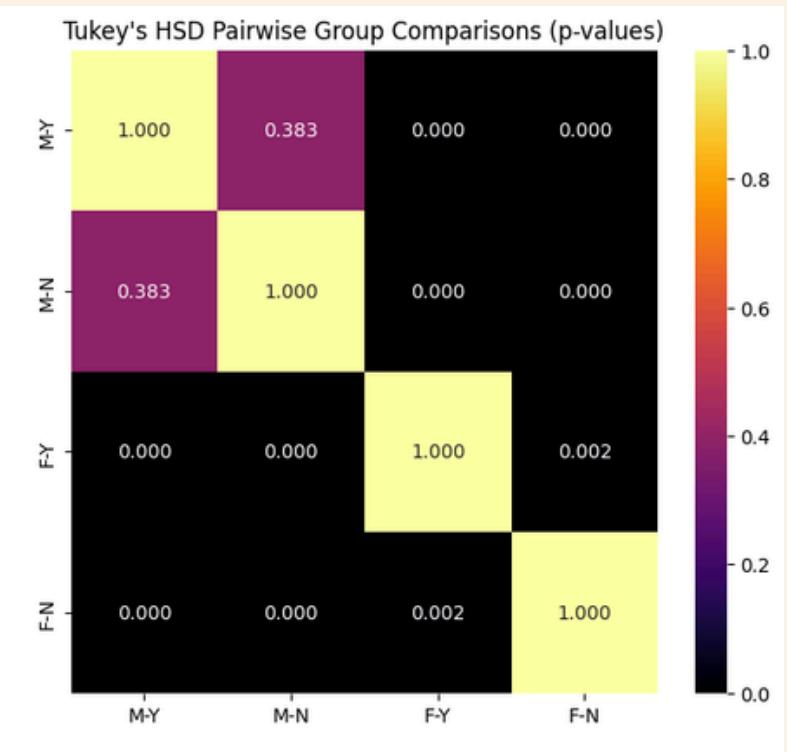
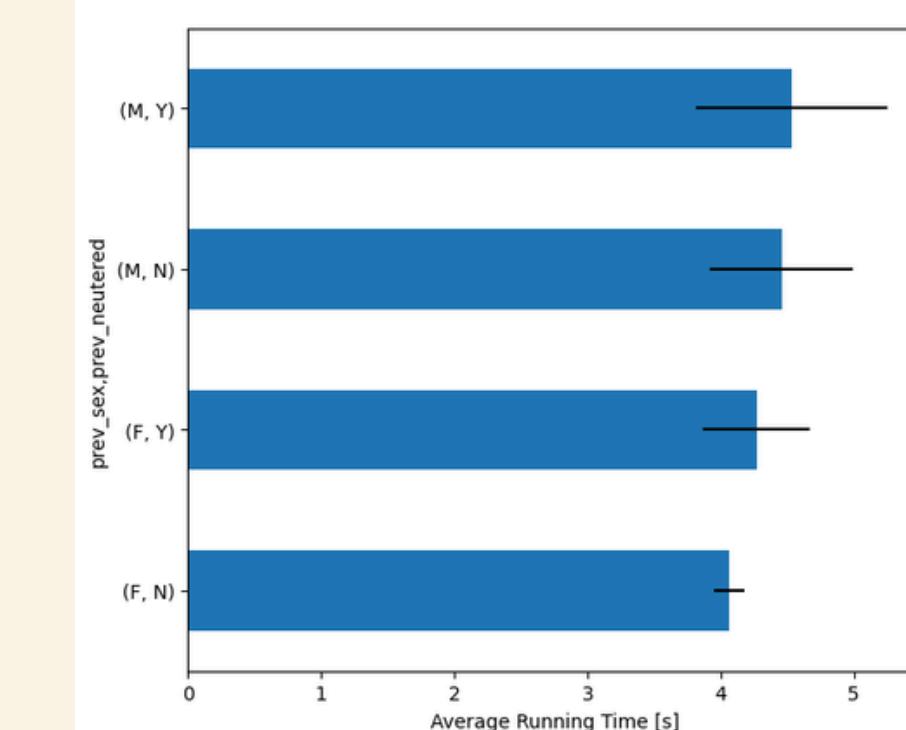
It appears that even FN dogs are faster when they exchange with another female.



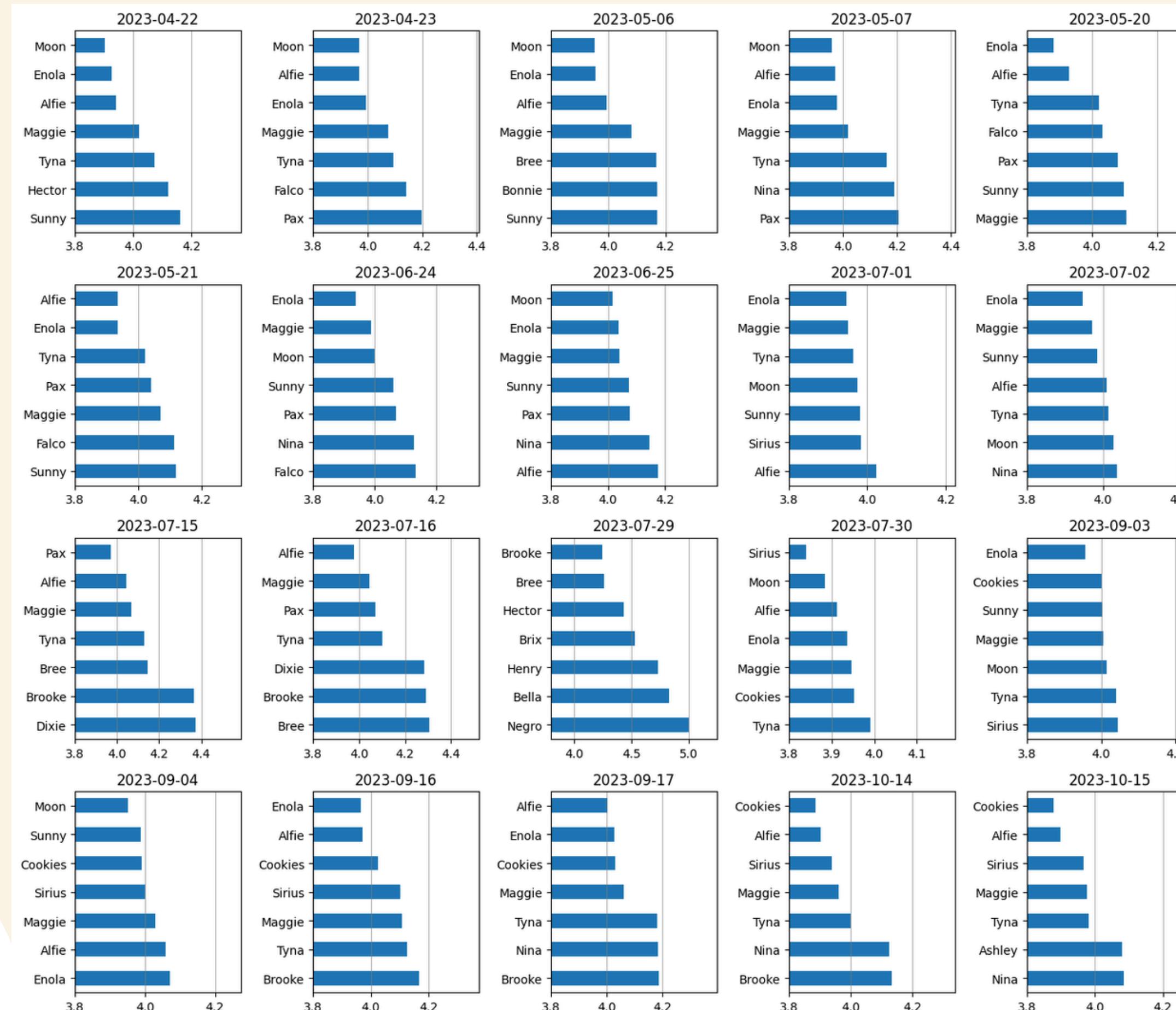
Neutered female dogs (FY)

The results are comparable to the overall population, but there is no difference between exchanging with an MY or an MN dog.

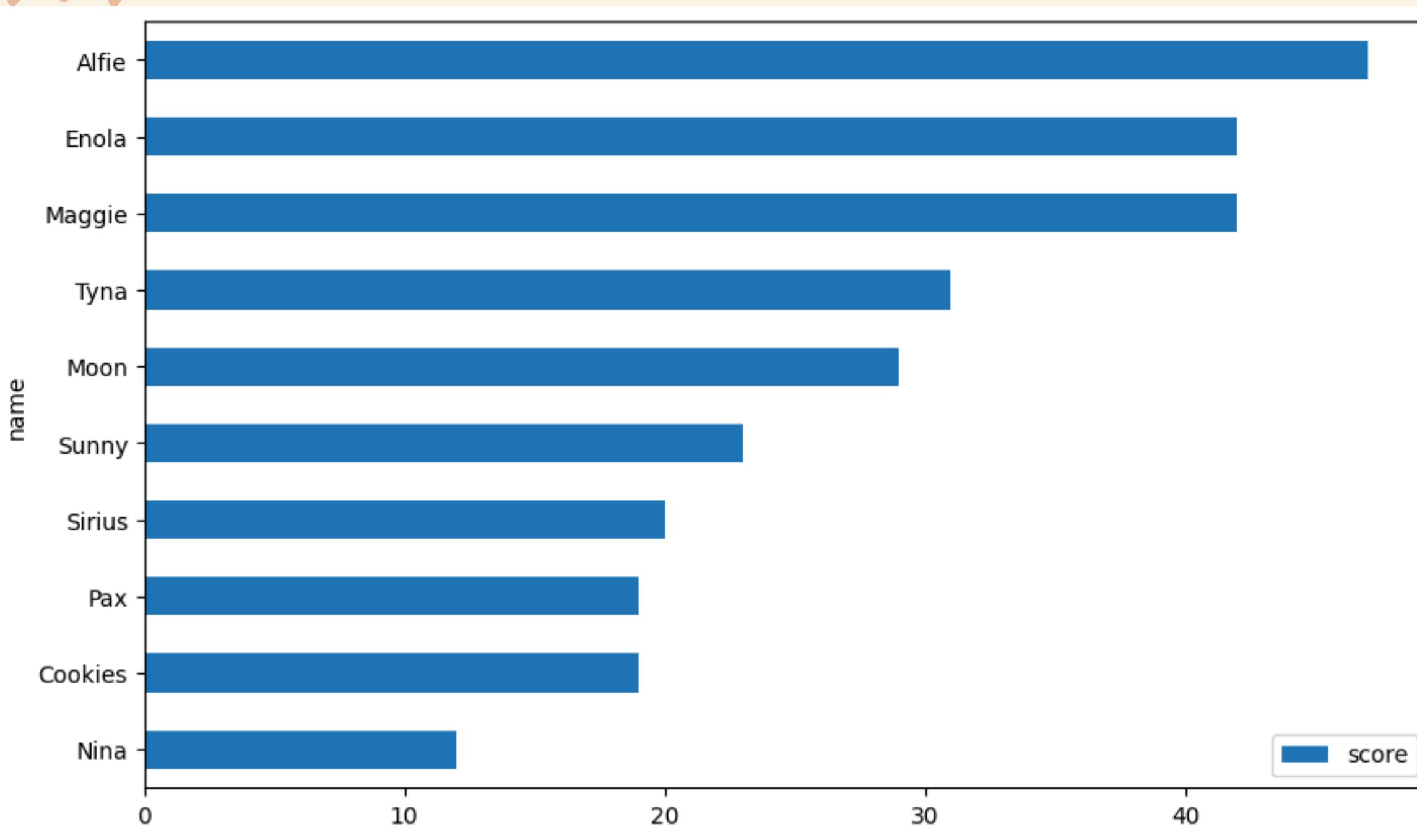
Exchanging with female unneutered dogs gives the best results.



Daily Performance



Dog Ranking



Dog ranking based on daily performance

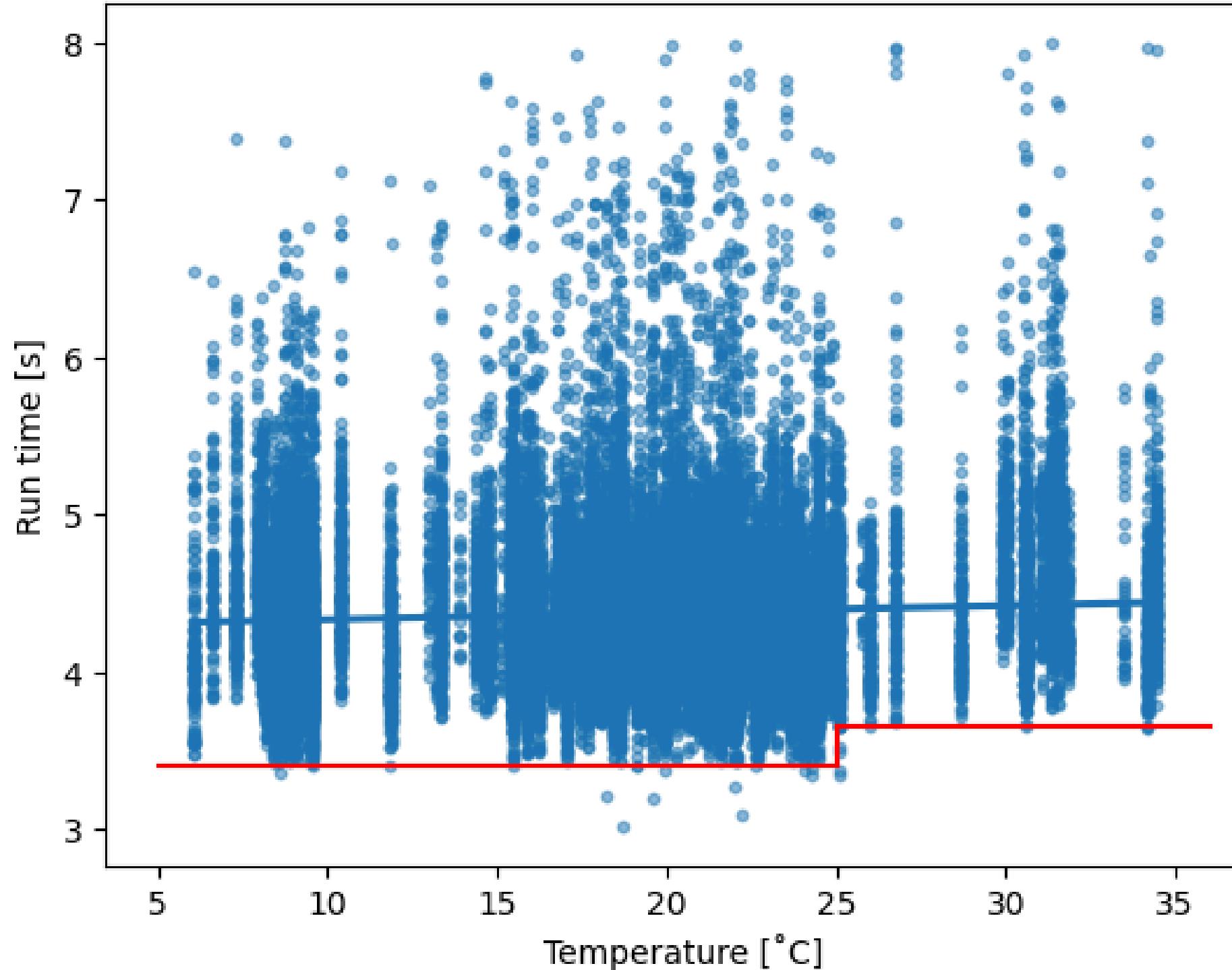
Each dog every day can collect points.

For a daily average in total time
(including run time + start/change time):

- avg. total time < 4.0 secs: 3 points
- avg. total time < 4.1 secs: 2 points
- avg. total time < 4.2 secs: 1 point

Temperature

Run Time vs. Temperature



Running time vs. temperature

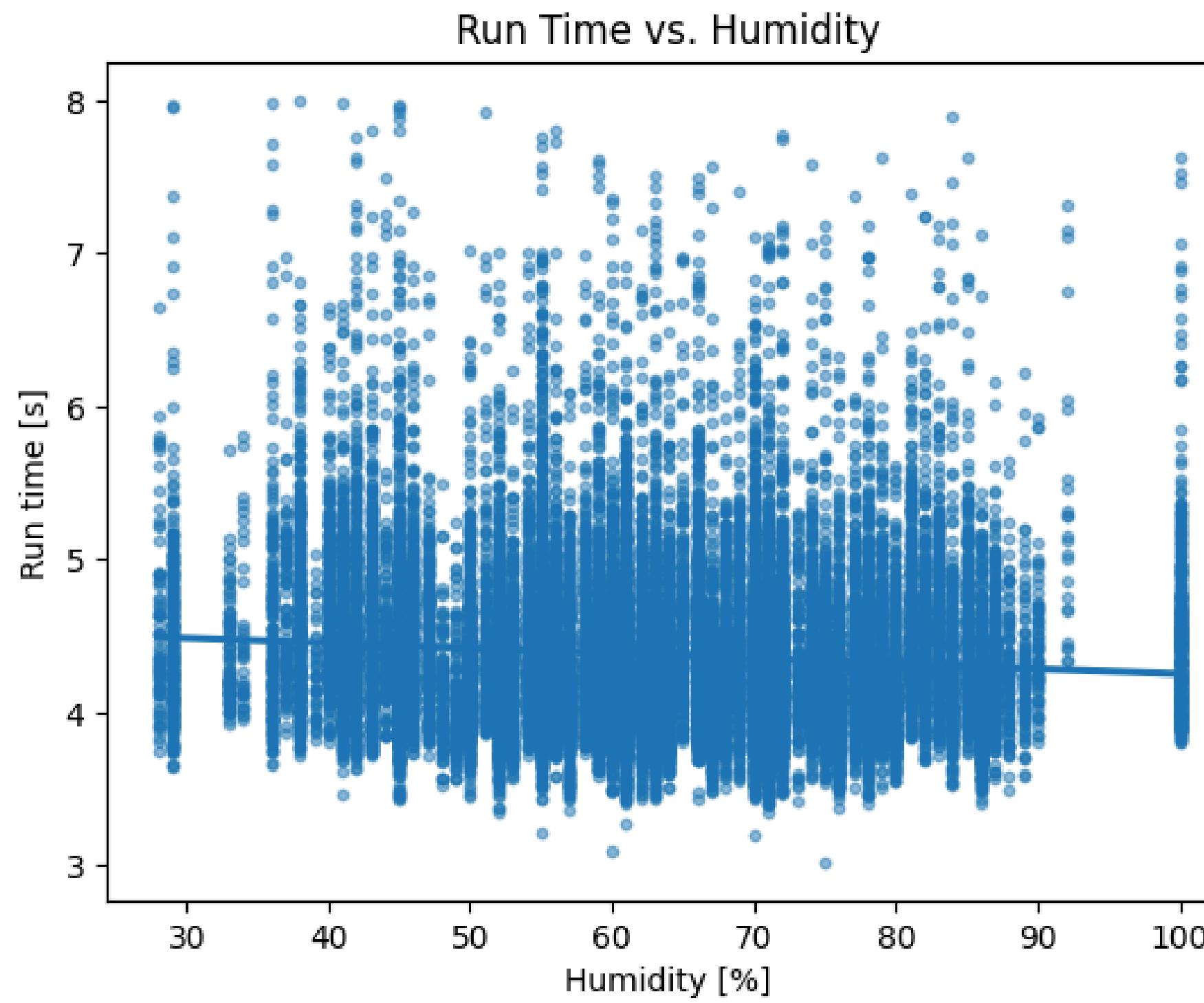
The low p-value indicates a significant correlation between temperature and running time. However, even though the correlation is significant, the slope is very low. A 1 °C increase in temperature corresponds to only a 0.004 second increase in running time.

Dog performance may improve by approximately 0.1 seconds on a cold day (< 10°C) compared to a hot day (> 30°C).

Visually examining the data suggests a stepwise performance degradation above 25°C. The run time of the best performing dogs abruptly changes from around 3.4 seconds to around 3.65 seconds. It's important to note, however, that lower performing days or events also occur at lower temperatures. Additionally, there are very few slow runs (> 6.5 seconds) on cold days (< 10°C) compared to many slower runs on warm days.

Overall, it seems reasonable to conclude that dog performance increases with lower temperatures.

Humidity

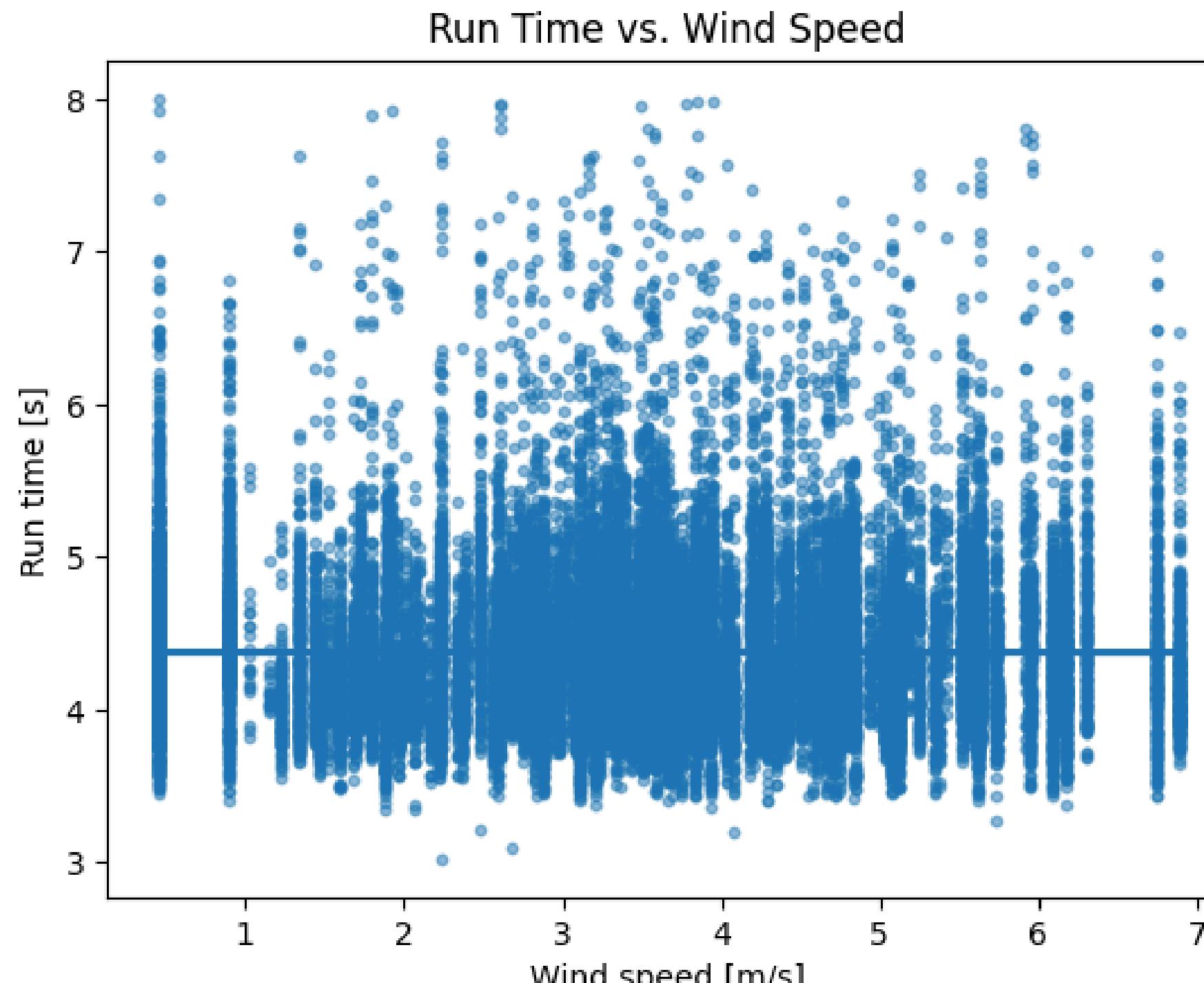


Running time vs. humidity

The low p-value indicates a significant correlation between humidity and running time. A 1% increase in humidity leads to dogs running slightly faster, by approximately 0.003 seconds. Comparing a very dry day (humidity around 30%) with a wet, likely rainy day (humidity 90–100%), the difference in average running time could be more than 0.2 seconds.

However, correlation does not equal causation. It's possible, that other factors associated with high humidity, such as lower temperatures or rain, are the actual reason improving dog performance.

Wind speed

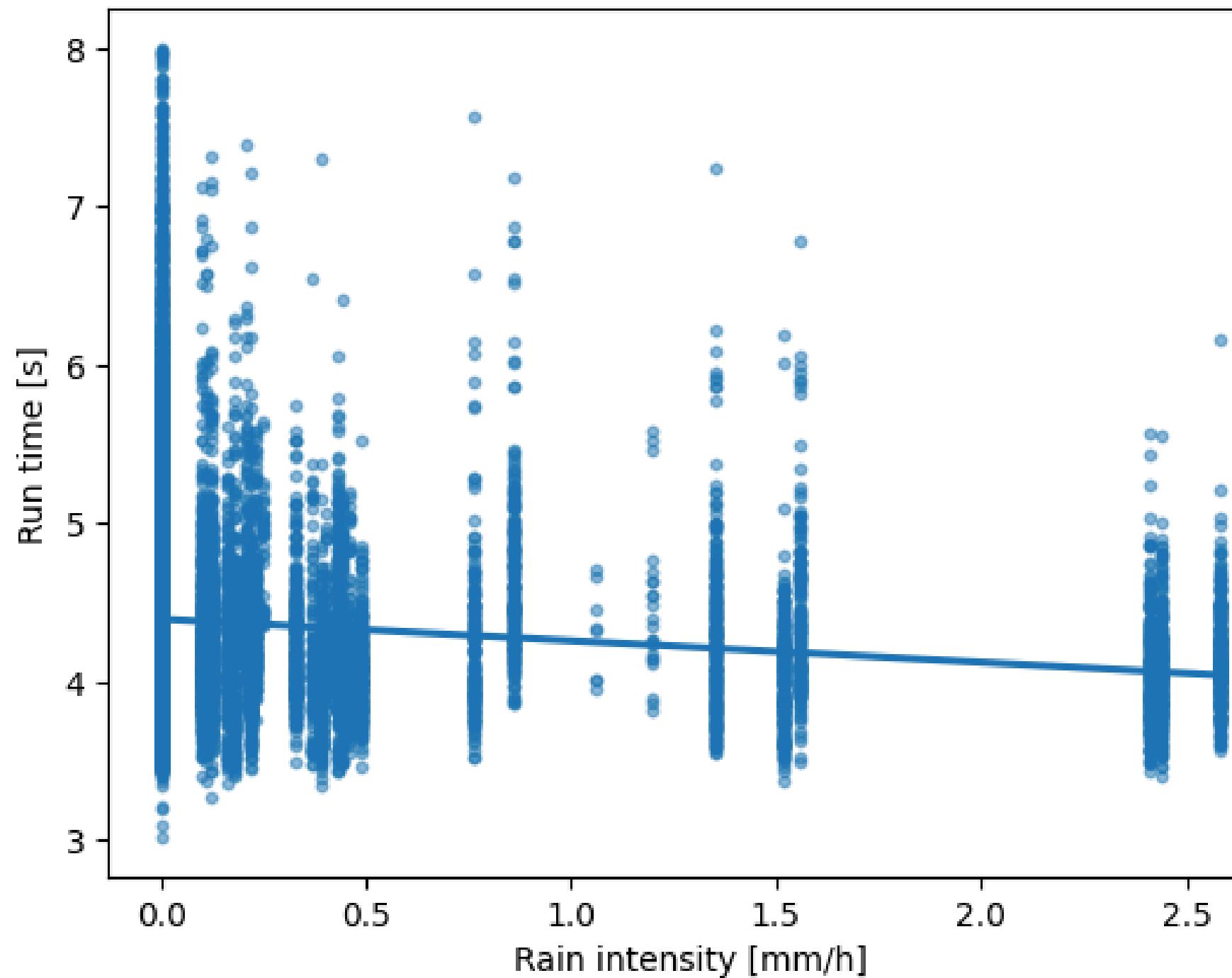


Running time vs. wind speed

The high p-value of 0.94 indicates that there is no significant correlation between wind speed and run time.

Rain Intensity

Run Time vs. Rain Intensity



Running time vs. rain intensity

The low p-value indicates that there is a significant correlation between rain intensity and running time. It appears that dogs run faster on a rainy day with a 1 mm/h increase in rain intensity corresponding to a decrease in running time of about 0.14 seconds.

Visually examining the data suggests rain may not affect peak performance. However, it seems to eliminate slow performances. There are many runs exceeding 5.5–6 seconds on dry days, but these longer runs become much rarer during rain.

Note that the real reason for slower runs on a dry day may be heat instead of the lack of rain.

Conclusion

Flyball competition data were analyzed for

- Individual dog performance
- Team performance
- Team composition
- Daily performance
- Course color
 - Course color has no impact
- Sex and neuter status of exchange partner
- Influence of weather
 - Hot weather decreases running performance



Recommendations

Do not mix different sized dogs in the same team

- Even large dogs run slower when the hurdles are lower

Use large young dogs whenever possible

Use more female dogs

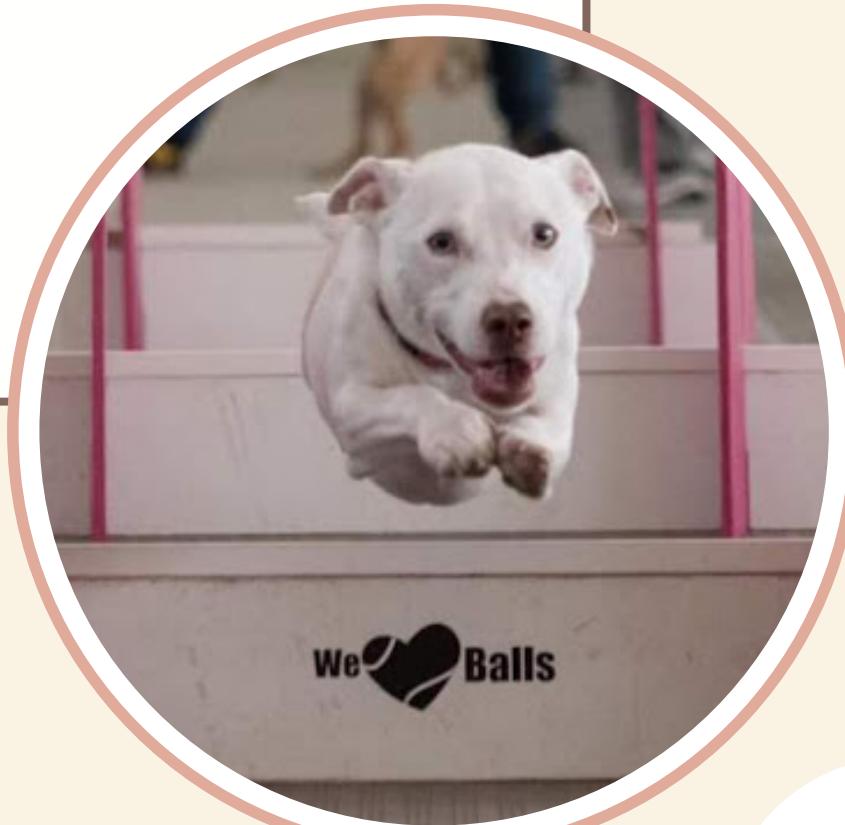
- Female dogs are better in change
- Running performance is improved when exchanging with a female, especially with an unneutered female



Further analyses

Further area of analysis:

- How much rest a dog needs between runs to keep up top performance?
- Hot and cold weather performance of different breeds
- More detailed breed analysis (current dataset is heavily biased to Border Collies)



We  Balls



Thank You
For Attention

tamas.berki.datascience@gmail.com