Marathon EDA

Marathon events of 50km and 50 miles of 2020 in USA
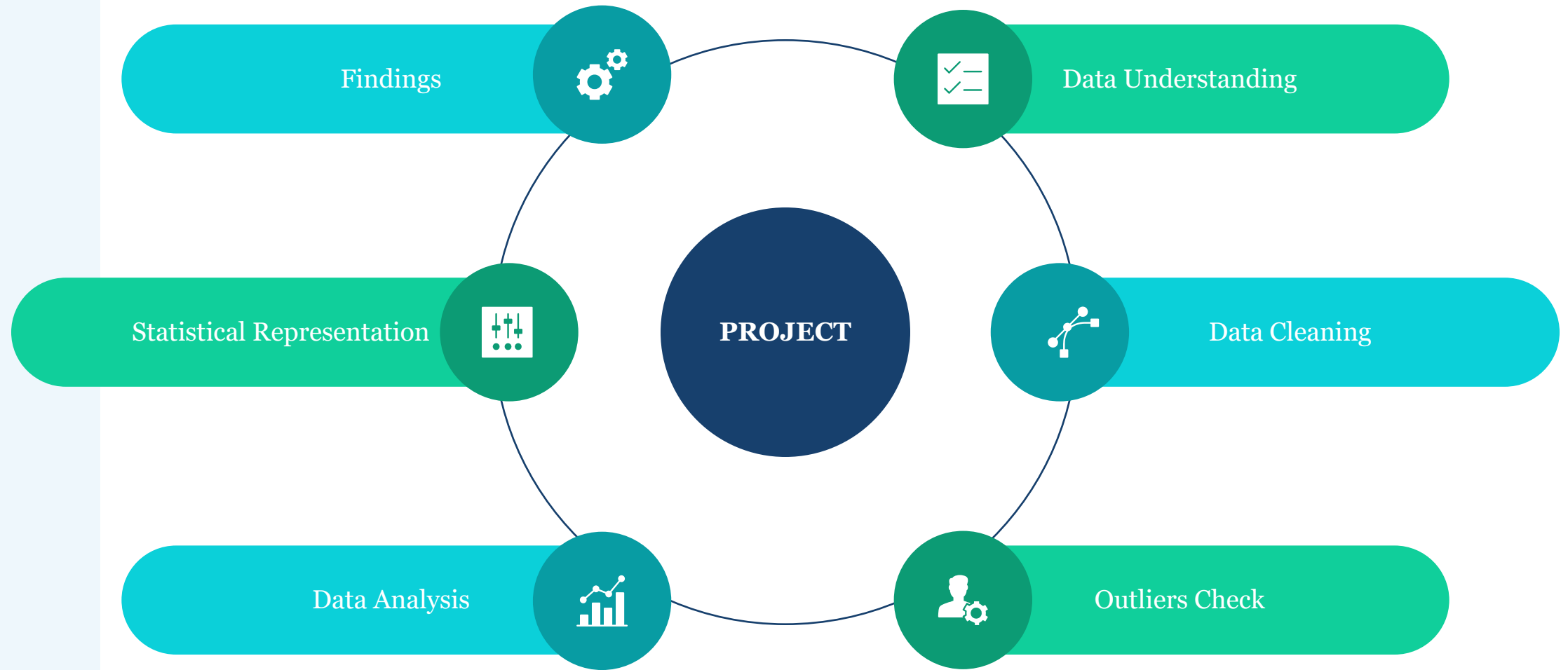
Project Statement

- To analyze the marathon events of 2020 in USA in the category of 50km and 50mi.

Hypothesis

- Age is a factor in average speed.
- Season is a factor in average speed.
- Gender wise participation.
- Gender is a factor of speed.

Marathon EDA

Findings

Data Understanding

Statistical Representation

PROJECT

Data Cleaning

Data Analysis

Outliers Check

Data Understanding

- Importing libraries

- Loading the data

- Reading the data

- Checking the structure of the data
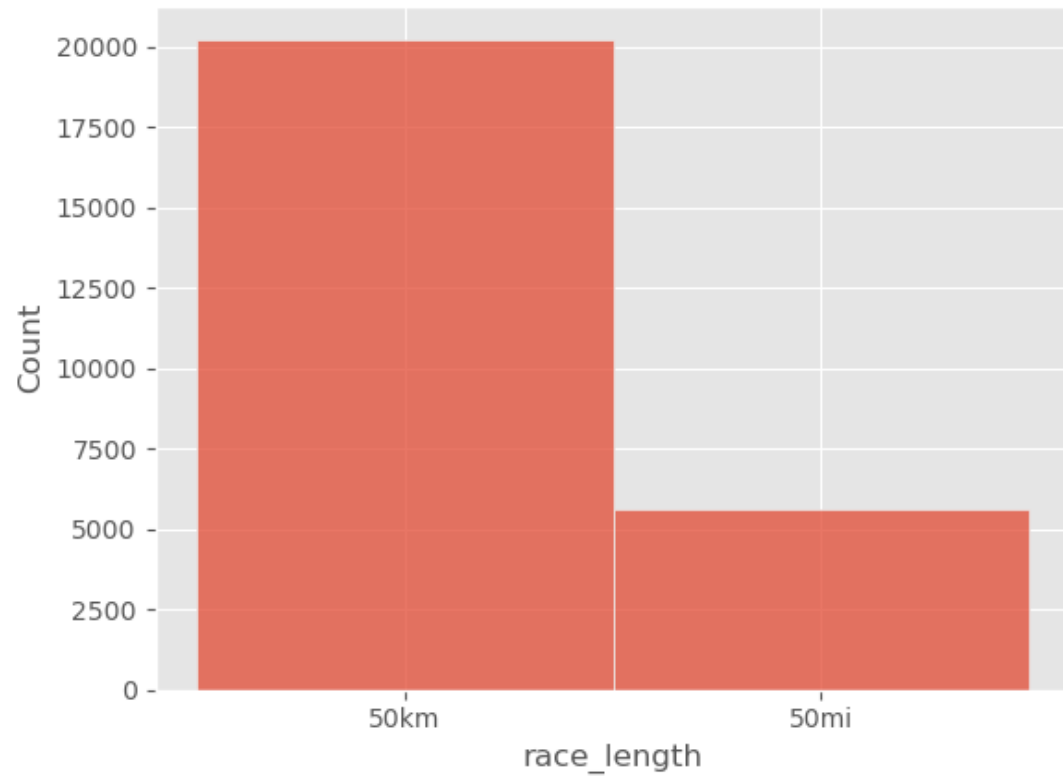
- Checking data types and null values

Data Cleaning

- Taking USA events from 2020

- Taking events of 50km and 50mi

- Age

- Removed 'h' from athlete performance

- Dropped columns: Year of birth and Age category

- Fixing data types

- Renaming and re-ordering of columns

Outliers

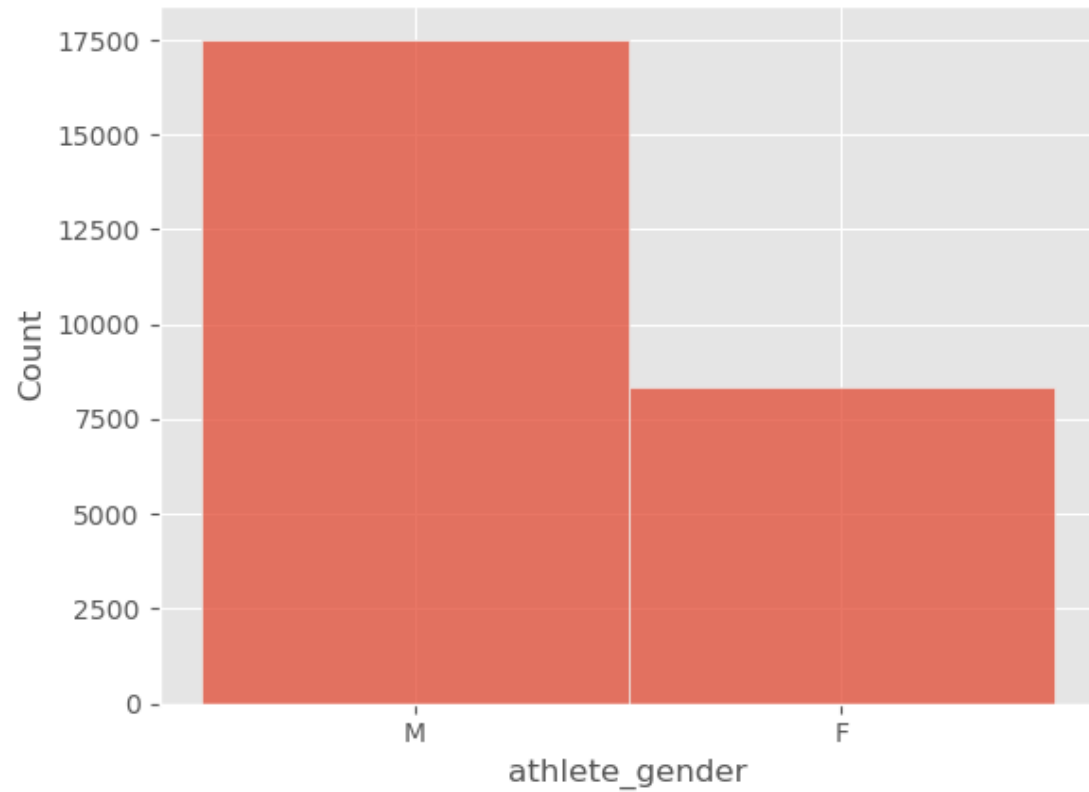- In the marathon case-study the outliers are over-performers who are in the lead.

Data Analysis

Race Length
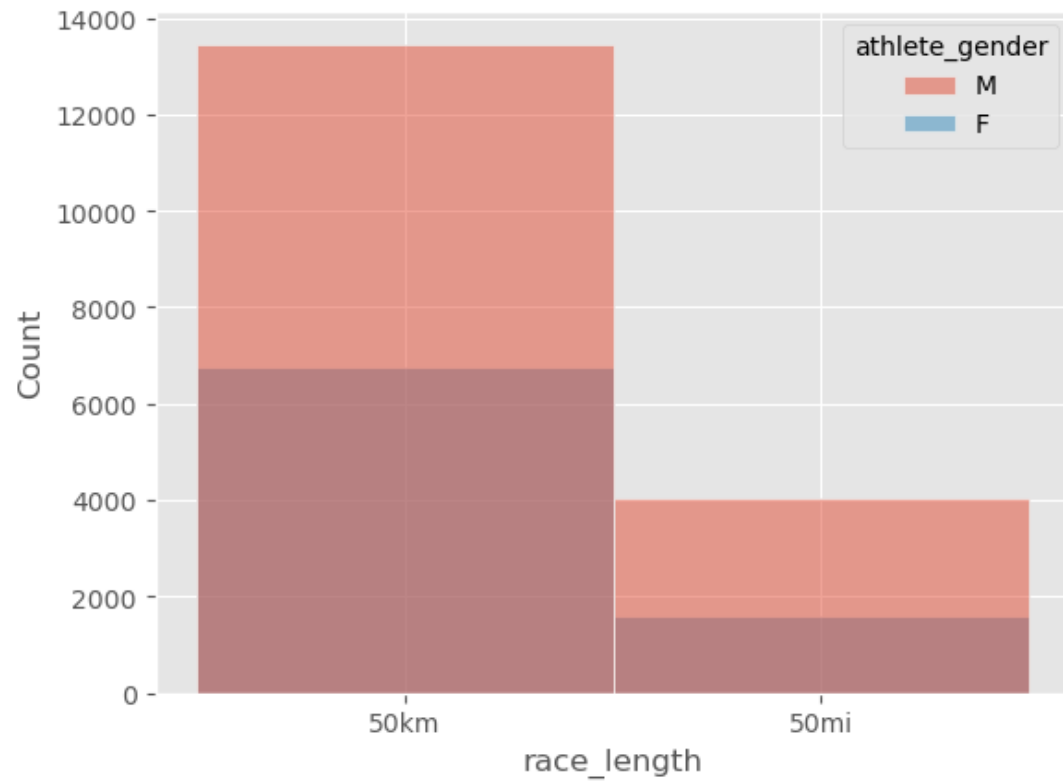

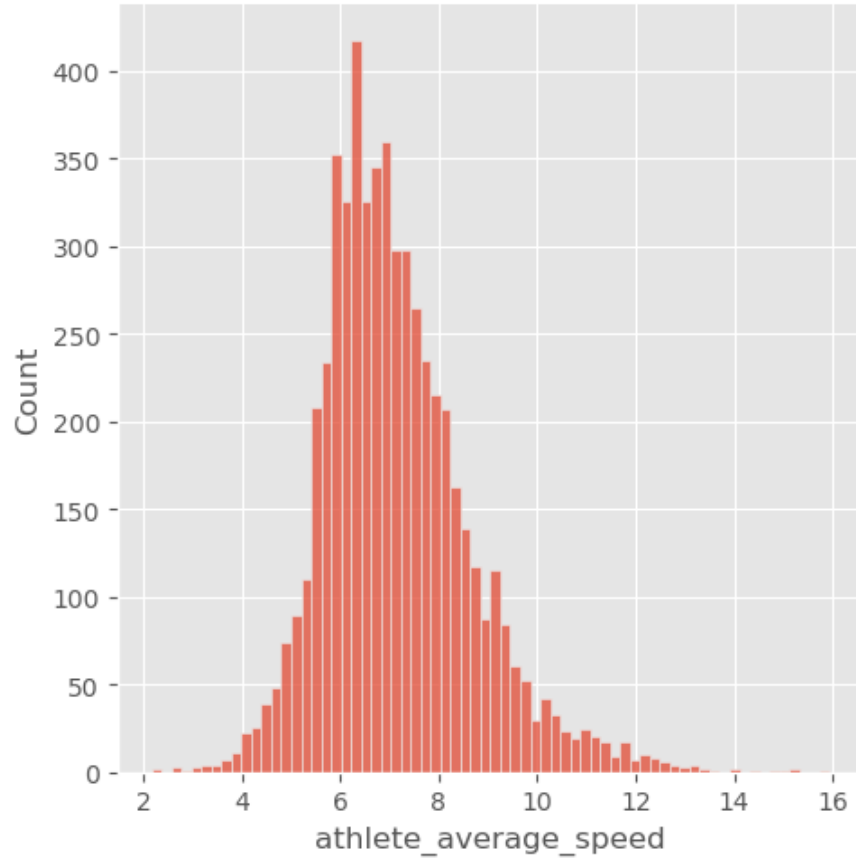- More 50km races

Gender

- More male athlete participants than female.
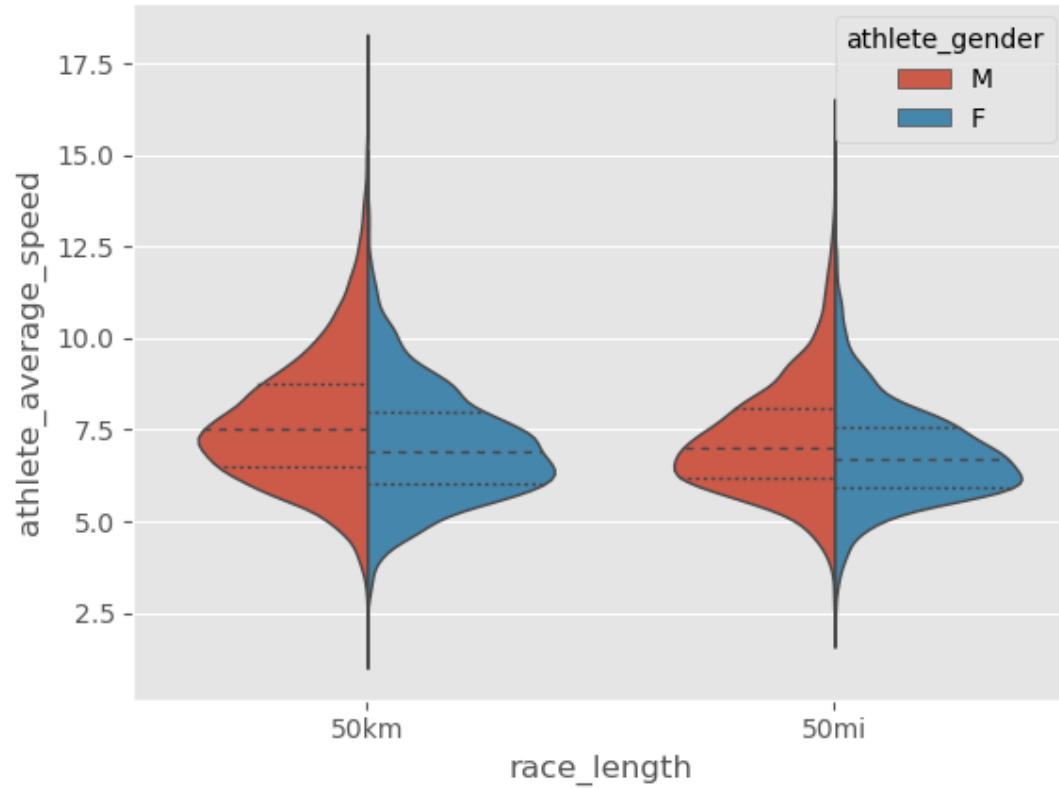
Race Length-Gender

- More male athlete than female in both event types.

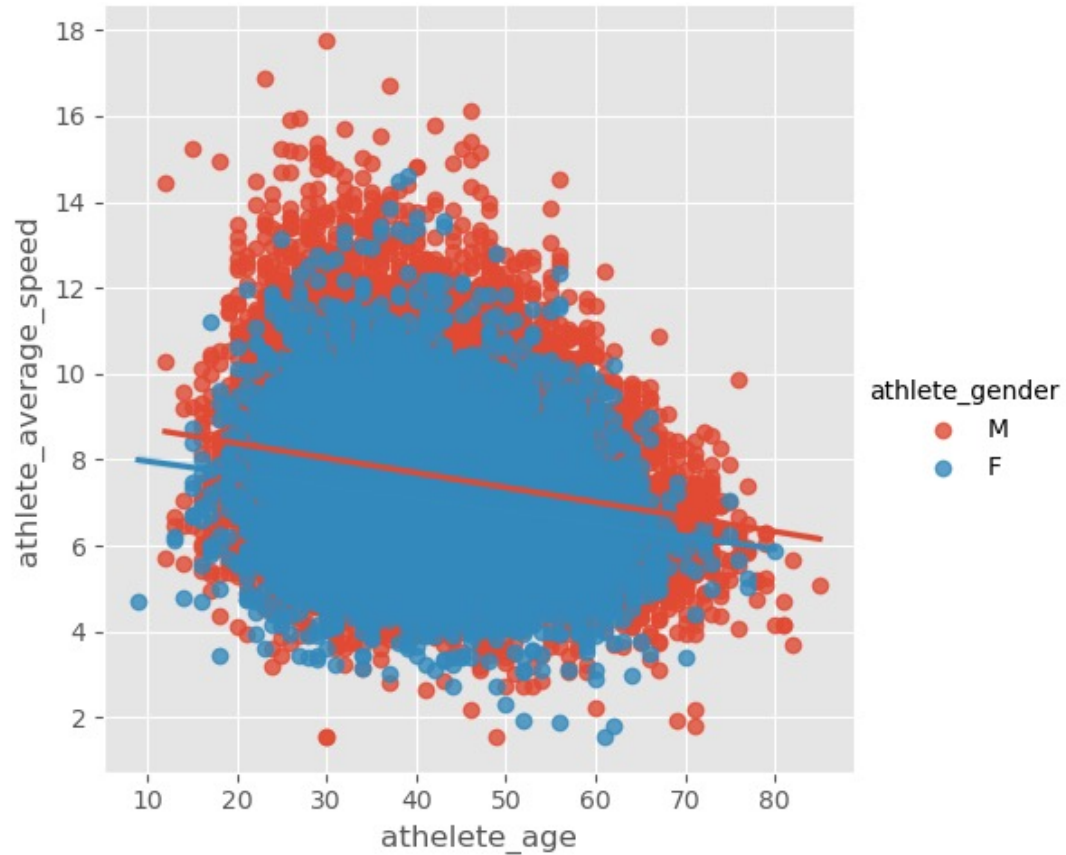# Race Length- Average Speed

- Average speed is between 6-8

Average speed on gender and event

- In the 50mi event the average speed for male and female is closer than 50km

Male-Female Average speed

- Male group has a higher average speed.

Best age- 50mil Race(20+races min)

athelete_age	mean	count
29	7.902252	135
23	7.779800	55
28	7.575252	107
30	7.569204	157
25	7.540923	91
31	7.451638	138
38	7.430022	231
35	7.422359	195
36	7.403854	185
26	7.379800	75
33	7.379188	149
22	7.367902	41
24	7.354274	73
42	7.327656	209
34	7.319011	182

- Best age is 29 with average speed of 7.9

Worst age- 50mil Race(20+races min)

athelete_age	mean	count
70	5.470667	12
65	5.934786	14
68	5.957500	10
67	6.114909	11
66	6.194300	10
60	6.261788	33
62	6.272730	37
61	6.358355	31
63	6.514806	31
58	6.582328	67
64	6.620727	22
50	6.671541	172
59	6.672072	83
48	6.696853	136
56	6.708373	67

- Worst age is 70 with average speed of 5.47

Season

	mean	count
race_season		
Spring	7.684430	3294
Winter	7.518187	11595
Fall	7.406619	8315
Summer	6.869336	2653

- Summer season has the lowest average speed.
- The maximum number of events in winter.

Summary

Summary

- Age is proved to be a factor in the average speed, with 29 being the best age group and 70 being the worst.

- Gender is proved to a factor in the average speed, with male group with higher average speed.

- Gender wise participation is also higher for male group.

- Season is proved to be a factor in the average speed, where summer season has the lowest average speed.