```
In [1]:  import pandas as pd
         import numpy as np
         import seaborn as sns
         import matplotlib.pyplot as plt

         import scipy.stats as stats

         def t_test(df1, df2, target="SOG"):
             t_stat, p_value = stats.ttest_ind(df1[target].dropna(), df2[target].dropna())
             print(f"T-statistic: {t_stat:.3f}, p-value: {p_value:.15f}")

             # If p-value is less than 0.05, the difference is statistically significant
             if p_value < 0.05:
                 print("The difference is statistically significant, keeping data split.")
             else:
                 print("The difference is not statistically significant, keeping data combined.")
```

```
In [2]:  df = pd.read_csv("all_data.csv")
```

```
In [3]:  data_9juin = df[df["ISODateTimeUTC"].str.startswith("2025-06-09")]
```

T test on the TWS between runs 5,6,7 where Karl holds the weights and runs 8,9,10,11 where Gian holds 6kgs

```
In [4]:  first_runs = ["09_06_Run5","09_06_Run6","09_06_Run7"]
         data_9juin_first_runs = data_9juin[data_9juin["run"].isin(first_runs) ]
```

```
In [5]:  last_runs = ["09_06_Run8","09_06_Run9","09_06_Run10","09_06_Run11"]
         data_9juin_last_runs = data_9juin[data_9juin["run"].isin(last_runs) ]
```

```
In [6]:  t_test(data_9juin_first_runs,data_9juin_last_runs, target="TWS")
         print(data_9juin_first_runs["TWS"].mean(),data_9juin_last_runs["TWS"].mean())
```

```
T-statistic: -20.004, p-value: 0.000000000000000
The difference is statistically significant, keeping data split.
7.2590247128437175 7.541569498486814
```

t test karl heavy vs karl not heavy

```
In [7]:  only_karl_first_runs_heavy = data_9juin_first_runs[
             (data_9juin_first_runs["boat_name"] == "Karl Maeder") |
             ((data_9juin_first_runs["boat_name"] == "Senseboard") &
              (data_9juin_first_runs["opponent_name"] == "Gian Stragiotti"))
         ]
         only_karl_first_runs_heavy.sample(5)
```

Out[7]:

	ISODateTimeUTC	SecondsSince1970	Heel_Abs	Heel_Lwd	Lat	LatBow	LatCenter	LatStern	Leg	Line_C	...	run	interval_id	boat_name	opponent_name	boat_role	boat_weight
47074	2025-06-09T12:54:10.452Z	1.749474e+09	52.5	52.5	43.507987	43.507985	43.507991	43.507997	NaN	127.100	...	09_06_Run7	1	Karl Maeder	SenseBoard	master	106.975
43822	2025-06-09T12:38:33.359Z	1.749473e+09	55.6	55.6	43.503502	43.503500	43.503506	43.503513	NaN	126.500	...	09_06_Run5	1	Karl Maeder	SenseBoard	master	106.975
46355	2025-06-09T12:49:03.561Z	1.749473e+09	56.0	56.0	43.505760	43.505762	43.505756	43.505750	1.0	105.057	...	09_06_Run6	2	Karl Maeder	SenseBoard	slave	106.975
46249	2025-06-09T12:48:52.960Z	1.749473e+09	47.4	47.4	43.504547	43.504549	43.504543	43.504538	1.0	89.000	...	09_06_Run6	2	Karl Maeder	SenseBoard	slave	106.975
46272	2025-06-09T12:48:55.252Z	1.749473e+09	43.8	43.8	43.504807	43.504809	43.504803	43.504797	1.0	100.458	...	09_06_Run6	2	Karl Maeder	SenseBoard	slave	106.975

5 rows × 48 columns

```
In [8]:  only_karl_last_runs_light = data_9juin_last_runs[
             (data_9juin_last_runs["boat_name"] == "Karl Maeder") |
             ((data_9juin_last_runs["boat_name"] == "SenseBoard") &
              (data_9juin_last_runs["opponent_name"] == "Gian Stragiotti"))
         ]
         only_karl_last_runs_light.sample(5)
```

Out[8]:

	ISODateTimeUTC	SecondsSince1970	Heel_Abs	Heel_Lwd	Lat	LatBow	LatCenter	LatStern	Leg	Line_C	...	run	interval_id	boat_name	opponent_name	boat_role	boat_weight
53781	2025-06-09T13:24:44.654Z	1.749475e+09	55.4	55.4	43.507835	43.507833	43.507839	43.507846	1.0	84.5	...	09_06_Run10	1	Karl Maeder	SenseBoard	slave	100.975
49316	2025-06-09T13:05:50.561Z	1.749474e+09	52.9	52.9	43.505969	43.505967	43.505973	43.505979	NaN	126.5	...	09_06_Run8	1	Karl Maeder	SenseBoard	slave	100.975
56260	2025-06-09T13:34:45.061Z	1.749476e+09	51.6	51.6	43.505294	43.505292	43.505299	43.505305	NaN	115.1	...	09_06_Run11	1	Karl Maeder	SenseBoard	master	100.975
54094	2025-06-09T13:25:15.960Z	1.749476e+09	47.8	47.8	43.504496	43.504494	43.504500	43.504507	1.0	113.6	...	09_06_Run10	1	Karl Maeder	SenseBoard	slave	100.975
54041	2025-06-09T13:25:10.655Z	1.749476e+09	61.8	61.8	43.505074	43.505072	43.505078	43.505084	1.0	129.0	...	09_06_Run10	1	Karl Maeder	SenseBoard	slave	100.975

5 rows × 48 columns

```
In [9]:  t_test(only_karl_first_runs_heavy,only_karl_last_runs_light) #general
```

T-statistic: -7.879, p-value: 0.000000000000004
The difference is statistically significant, keeping data split.

```
In [10]: t_test(only_karl_first_runs_heavy[only_karl_first_runs_heavy["TWA"]>0],only_karl_last_runs_light[only_karl_last_runs_light["TWA"]>0]) #upwind
```

T-statistic: -1.138, p-value: 0.255372507601836
The difference is not statistically significant, keeping data combined.

```
In [11]: t_test(only_karl_first_runs_heavy[only_karl_first_runs_heavy["TWA"]<=0],only_karl_last_runs_light[only_karl_last_runs_light["TWA"]<=0]) #downwind
```

```
T-statistic: 18.894, p-value: 0.000000000000000
The difference is statistically significant, keeping data split.
```

t test karl heavy vs karl not heavy

In [13]:
```python
only_gian_first_runs_light = data_9juin_first_runs[
    (data_9juin_first_runs["boat_name"] == "Gian Stragiotti") |
    ((data_9juin_first_runs["boat_name"] == "SenseBoard") &
     (data_9juin_first_runs["opponent_name"] == "Karl Maeder"))
]
only_gian_first_runs_light.sample(5)
```

Out[13]:

	ISODateTimeUTC	SecondsSince1970	Heel_Abs	Heel_Lwd	Lat	LatBow	LatCenter	LatStern	Leg	Line_C	...	run	interval_id	boat_name	opponent_name	boat_role	boat_weight
44662	2025-06-09T12:38:48.560Z	1.749473e+09	41.3	41.3	43.502049	43.502047	43.502053	43.502059	1.0	4.3	...	09_06_Run5	1	SenseBoard	Karl Maeder	slave	109.09
47709	2025-06-09T12:54:05.072Z	1.749474e+09	58.2	58.2	43.508599	43.508597	43.508603	43.508609	1.0	3.5	...	09_06_Run7	1	SenseBoard	Karl Maeder	slave	109.09
47783	2025-06-09T12:54:12.454Z	1.749474e+09	51.3	51.3	43.507862	43.507860	43.507866	43.507872	1.0	4.1	...	09_06_Run7	1	SenseBoard	Karl Maeder	slave	109.09
48063	2025-06-09T12:54:40.454Z	1.749474e+09	53.1	53.1	43.505044	43.505042	43.505048	43.505054	1.0	6.3	...	09_06_Run7	1	SenseBoard	Karl Maeder	slave	109.09
48978	2025-06-09T12:56:57.558Z	1.749474e+09	49.8	49.8	43.505180	43.505182	43.505176	43.505171	NaN	3.7	...	09_06_Run7	2	SenseBoard	Karl Maeder	slave	109.09

5 rows × 48 columns

In [14]:
```python
only_gian_last_runs_heavy = data_9juin_last_runs[
    (data_9juin_last_runs["boat_name"] == "Gian Stragiotti") |
    ((data_9juin_last_runs["boat_name"] == "SenseBoard") &
     (data_9juin_last_runs["opponent_name"] == "Karl Maeder"))
]
only_gian_last_runs_heavy.sample(5)
```

Out[14]:

	ISODateTimeUTC	SecondsSince1970	Heel_Abs	Heel_Lwd	Lat	LatBow	LatCenter	LatStern	Leg	Line_C	...	run	interval_id	boat_name	opponent_name	boat_role	boat_weight
55882	2025-06-09T13:28:19.048Z	1.749476e+09	56.2	56.2	43.507671	43.507673	43.507667	43.507661	NaN	7.070	...	09_06_Run10	2	SenseBoard	Karl Maeder	master	115.09
54790	2025-06-09T13:25:24.257Z	1.749476e+09	57.7	57.7	43.503416	43.503414	43.503420	43.503426	NaN	7.500	...	09_06_Run10	1	SenseBoard	Karl Maeder	master	115.09
57004	2025-06-09T13:34:52.358Z	1.749476e+09	45.0	45.0	43.504591	43.504589	43.504596	43.504602	1.0	4.800	...	09_06_Run11	1	SenseBoard	Karl Maeder	slave	115.09
53580	2025-06-09T13:19:21.053Z	1.749475e+09	54.3	54.3	43.507223	43.507225	43.507219	43.507213	NaN	6.800	...	09_06_Run9	2	SenseBoard	Karl Maeder	slave	115.09
58266	2025-06-09T13:38:02.359Z	1.749476e+09	25.9	25.9	43.507805	43.507806	43.507801	43.507795	NaN	2.052	...	09_06_Run11	2	SenseBoard	Karl Maeder	slave	115.09

5 rows × 48 columns

In [15]: `t_test(only_gian_first_runs_light,only_gian_last_runs_heavy) #general`

```
T-statistic: -11.650, p-value: 0.000000000000000
The difference is statistically significant, keeping data split.
```

In [16]: `t_test(only_gian_first_runs_light[only_gian_first_runs_light["TWA"]>0],only_gian_last_runs_heavy[only_gian_last_runs_heavy["TWA"]>0]) #upwind`

```
T-statistic: -8.572, p-value: 0.000000000000000
The difference is statistically significant, keeping data split.
```

In [17]: `t_test(only_gian_first_runs_light[only_gian_first_runs_light["TWA"]<=0],only_gian_last_runs_heavy[only_gian_last_runs_heavy["TWA"]<=0]) #downwind`

```
T-statistic: 10.288, p-value: 0.000000000000000
The difference is statistically significant, keeping data split.

The difference is statistically significant, keeping data split.
```