Exam3035: Five plots vizualizing my running stats from strava

This exam is published as an open repository at Github: https://github.com/Kristine-jo/ExamPSY3035.git

The project explores 4 hypothesis and uses five figures visualizing the results. The purpuse of the project is to gain insigth in my running acheivments and habits. Based on these plots I will present and analyze the results of my research. The datafile I use is from my own personal Strava records. The data can be found in my github under the folder data and is called "my_running.csv". Using this data I get information about my runs. The variables I use, and which are important for my project is: average graded pace, distance, time and activity date. From these variables I extract the information needed to answer these 4 hypotheses:

- H1: Time of day effects the average pace of all of my runs
- H2: Time of day effects the pace on my 4-6km runs
- H3: Which distance did I run most frequently in 2024
- H4: Which month of the year did I run the most in 2024

In [2]:
```
#All packages i need to be able to access functions easily
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

In [3]:
```
#Loaded data in to pd, because pandas handles data well and works well with matp
# Using packadges such as pandas makes it easier to handle strings, booleans and
df = pd.read_csv(r"..\Data\runlog\my_running.csv")
```

In [4]:
```
df  #showing table containing strava stats
```

Out[4]:

	Activity ID	Activity Date	Activity Name	Activity Type	Activity Description	Elapsed Time
0	11288364028	Apr 29, 2024, 5:14:04 AM	Morning Run	Run	NaN	2025
1	11327182053	May 4, 2024, 6:54:55 AM	Morning Run	Run	NaN	1759
2	11354165092	May 7, 2024, 4:31:09 PM	Evening Run	Run	NaN	827
3	11371024533	May 9, 2024, 4:22:16 PM	Løp på mølle, etter skolen 💔	Run	Løp 8 km med 1,5 stigning på mølle	2620
4	11402044804	May 13, 2024, 2:30:00 PM	Løp på mølle etter skole 🤓	Run	NaN	1630
5	11425811688	May 16, 2024, 4:58:18 PM	Bratte bakker med martin	Run	NaN	1931
6	11447612172	May 19, 2024, 12:34:48 PM	Templarheimen-telegrafbukta-sentrum	Run	Gikk som smuurt, føltes veldig bra heleveien	3163
7	11463824675	May 21, 2024, 2:31:21 PM	Afternoon Run	Run	6 grader ute og shorts 🌞 grei pace	2104
8	11478659771	May 23, 2024, 12:47:45 PM	Løp te telegrafbukta	Run	20 varmegrader ute! Så ble et lite bad når me ...	1674
9	11491349434	May 25, 2024, 8:00:08 AM	Morning Run	Run	Endelig har snøen smelta i lysløypå! 🏃 ❤️	1904
10	11504674828	May 26, 2024, 6:26:35 PM	Evening Run	Run	NaN	1498
11	11518449873	May 28, 2024,	Løp i solå	Run	Gikk veldig bra, var sykt varmt da	2188

	Activity ID	Activity Date	Activity Name	Activity Type	Activity Description	Elapsed Time
		4:55:00 PM				
12	11535082647	May 30, 2024, 6:26:47 PM	restutisjonsløp 💪 🌞	Run	NaN	1777
13	11539555466	May 31, 2024, 11:27:34 AM	Løp te polet	Run	Perf temp, sto og venta på verdens lengste rød...	3002
14	11563861067	Jun 3, 2024, 2:55:18 PM	Afternoon Run	Run	Løp innom butikken og hjem	1712
15	11597675325	Jun 7, 2024, 5:41:10 PM	tebake i stavaaaangeeeRRRRR!!!	Run	Nice å løpe i sørmarkå igjen 🌿🌳, men sattan for...	2160
16	11612344724	Jun 9, 2024, 1:28:23 PM	Dagen derpå... tungt det😮‍💨	Run	Var ganske tungt. Mem deilig at det var mye vi...	2996
17	11656959150	Jun 15, 2024, 10:28:28 AM	Ææ det var tungt🤯	Run	Veldig mye oppoverbakker for en 11k..	3510
18	11674087087	Jun 17, 2024, 2:08:23 PM	Ups æ bada	Run	Var digg å bade med fikk litt vondt under fote...	1210
19	11719328361	Jun 23, 2024, 11:14:45 AM	Løp med sabbie før sankt hans feiring 💔 ‼️	Run	NaN	2207
20	11736909938	Jun 25, 2024, 3:10:45 PM	Vaaaaarmt🤯	Run	NaN	3523
21	11752483688	Jun 27, 2024, 2:13:54 PM	Løp på mølle	Run	Oppvarming før mage økt🤯	1000
22	11772700697	Jun 30, 2024, 8:24:37 AM	Oppover og oppover og oppover...🤯	Run	Såg en fasanfamilie! Bakken opp te ullanhaugst...	1587

	Activity ID	Activity Date	Activity Name	Activity Type	Activity Description	Elapsed Time
23	11798960902	Jul 3, 2024, 2:23:25 PM	Afternoon Run	Run	NaN	1657
24	11926180574	Jul 16, 2024, 4:00:32 PM	Post covid run på mølle 😳🙈	Run	Det føltes ikkje bra 😵 tung i pusten og støl i...	1800
25	11926206972	Jul 18, 2024, 2:00:00 PM	Back at it 💥	Run	Det gikk bra! Andre løp etter korona, men hadd...	1738
26	11932610893	Jul 20, 2024, 6:35:59 AM	Dalsnuten	Run	Sykt digg! Fint vær og nesten helt tomt for fo...	3059
27	11953922104	Jul 22, 2024, 6:01:58 PM	3k	Run	Igjen, treningsenteret på uis e så\nJÆVLIG var...	960
28	12012007482	Jul 29, 2024, 5:11:08 PM	Vålandstårnet og mosvannet	Run	Tungt og godt	2422
29	12029965346	Jul 31, 2024, 6:29:48 PM	Ullanhaugstårnet og tebake	Run	NOOOOOOO eg kom borti pauseknappen under løpet...	1654
30	12036820394	Aug 1, 2024, 3:46:04 PM	Jogga te trening	Run	NaN	1320
31	12051297516	Aug 3, 2024, 12:05:16 PM	Oppvarming på mølle	Run	Kokvarmt på sis som alltid 😵 men selve løping...	1038
32	12077888082	Aug 6, 2024, 2:32:10 PM	Løp te trening	Run	NaN	1253
33	12089152285	Aug 7, 2024, 6:36:14 PM	Siste joggetur i stavanger i sommer	Run	Va nice, litt stiv i kneet 🥹	1418
34	12111709943	Aug 10, 2024,	Fuste løpetur i Tromsø dette semesteret 😎	Run	Føltes veldig bra! Litt varmt bare, og	1581

	Activity ID	Activity Date	Activity Name	Activity Type	Activity Description	Elapsed Time
		1:53:21 PM			de sist...	
35	12126628243	Aug 12, 2024, 9:19:56 AM	Perf temp 😌	Run	Veldig deilig løpetur 😊	3758
36	12156193243	Aug 15, 2024, 5:20:15 PM	Te kirkegården og tebake	Run	Rolig jogg de fusste 3 kilometrene. Ganske ras...	2144
37	12180076216	Aug 18, 2024, 1:10:25 PM	Liten luftetur	Run	Martin fikk gnagsår under foten av skonå 😣	811
38	12203349729	Aug 21, 2024, 9:34:41 AM	UURG	Run	Det var treigt i kroppen.. e så støl etter tre...	2502
39	12322004987	Sep 4, 2024, 8:07:41 AM	Snørrete løp 🥲 🤧	Run	Fusste dag med trening etter å ha vært syk i e...	1650
40	12347858622	Sep 7, 2024, 10:03:56 AM	Hølje ned 🌧️	Run	Sykt digg å løpe i regnet 😌	2249
41	12368102065	Sep 9, 2024, 3:32:42 PM	Litt bortfor fløya	Hike	Fikk pressa inn en liten fjelltur etter skolen...	7904
42	12380939837	Sep 11, 2024, 6:10:20 AM	Morning Run	Run	Blir godt med frokost nå 😌 🥹	1619
43	12416719456	Sep 15, 2024, 12:54:33 PM	Deilig på en søndag 😌	Run	Sku egentlig ikkje springe så langt men det lu...	3401
44	12435219995	Sep 17, 2024, 5:37:06 PM	Rolig jogg med pookie	Run	Veldig deilig, litt sånn halloweenstemning	1944
45	12477125422	Sep 22, 2024, 3:01:24 PM	Brrr, litt kjølig nå 🥶	Run	NaN	1668

	Activity ID	Activity Date	Activity Name	Activity Type	Activity Description	Elapsed Time
46	12523214664	Sep 28, 2024, 11:34:43 AM	1 mil i sol og 6 grader ☀️ ❄️	Run	Meget godt😉	3353
47	12588916760	Oct 6, 2024, 10:52:22 AM	Folksomt ute idag 🏃‍♂️	Run	NaN	1606
48	12607959321	Oct 8, 2024, 5:43:44 PM	Kveldstur med Martin 🏃‍♂️ 🚲	Run	Va veldig digg	2568
49	12662092835	Oct 15, 2024, 3:15:01 PM	Afternoon Run	Run	NaN	2025
50	12690090965	Oct 19, 2024, 7:35:41 AM	Morning Run	Run	NaN	1968
51	12867926253	Nov 10, 2024, 12:53:42 PM	Afternoon Run	Run	NaN	2059
52	13403019941	Jan 20, 2025, 7:23:42 AM	Til kraft	Walk	NaN	2401

53 rows × 94 columns

In [5]:
```
#making sure the file is fully loaded, i had some issues with not all the variab
print(df.shape) #all variables are included
```

```
(53, 94)
```

Converting and calculating variables

Before I could begin exploring my hypothesis, I needed to extract and convert some variables from the dataset. First, Strava only records the average graded pace and not the average pace, which is the variable I actually need. To address this, I first converted the time variable from seconds to minutes. Then, I calculated the average pace using the time in minutes and the distance in kilometers, which was already available in the data file.

Next, I needed to extract the time of day when the runs were completed, based on the activity date variable.

In [7]:
```
#convert time from seconds to minutes, and add avrage pace
df["time_minutes"] = df["Elapsed Time"] / 60
df["average_pace"] = df["time_minutes"] / df["Distance"]
print(df)
```

	Activity ID	Activity Date	Activity Name	Activity Type
0	11288364028	Apr 29, 2024, 5:14:04 AM	Morning Run	Run
1	11327182053	May 4, 2024, 6:54:55 AM	Morning Run	Run
2	11354165092	May 7, 2024, 4:31:09 PM	Evening Run	Run
3	11371024533	May 9, 2024, 4:22:16 PM	Løp på mølle, etter skolen 💔	Run
4	11402044804	May 13, 2024, 2:30:00 PM		
5	11425811688	May 16, 2024, 4:58:18 PM		
6	11447612172	May 19, 2024, 12:34:48 PM		
7	11463824675	May 21, 2024, 2:31:21 PM		
8	11478659771	May 23, 2024, 12:47:45 PM		
9	11491349434	May 25, 2024, 8:00:08 AM		
10	11504674828	May 26, 2024, 6:26:35 PM		
11	11518449873	May 28, 2024, 4:55:00 PM		
12	11535082647	May 30, 2024, 6:26:47 PM		
13	11539555466	May 31, 2024, 11:27:34 AM		
14	11563861067	Jun 3, 2024, 2:55:18 PM		
15	11597675325	Jun 7, 2024, 5:41:10 PM		
16	11612344724	Jun 9, 2024, 1:28:23 PM		
17	11656959150	Jun 15, 2024, 10:28:28 AM		
18	11674087087	Jun 17, 2024, 2:08:23 PM		
19	11719328361	Jun 23, 2024, 11:14:45 AM		
20	11736909938	Jun 25, 2024, 3:10:45 PM		
21	11752483688	Jun 27, 2024, 2:13:54 PM		
22	11772700697	Jun 30, 2024, 8:24:37 AM		
23	11798960902	Jul 3, 2024, 2:23:25 PM		
24	11926180574	Jul 16, 2024, 4:00:32 PM		
25	11926206972	Jul 18, 2024, 2:00:00 PM		
26	11932610893	Jul 20, 2024, 6:35:59 AM		
27	11953922104	Jul 22, 2024, 6:01:58 PM		
28	12012007482	Jul 29, 2024, 5:11:08 PM		
29	12029965346	Jul 31, 2024, 6:29:48 PM		
30	12036820394	Aug 1, 2024, 3:46:04 PM		
31	12051297516	Aug 3, 2024, 12:05:16 PM		
32	12077888082	Aug 6, 2024, 2:32:10 PM		
33	12089152285	Aug 7, 2024, 6:36:14 PM		
34	12111709943	Aug 10, 2024, 1:53:21 PM		
35	12126628243	Aug 12, 2024, 9:19:56 AM		
36	12156193243	Aug 15, 2024, 5:20:15 PM		
37	12180076216	Aug 18, 2024, 1:10:25 PM		
38	12203349729	Aug 21, 2024, 9:34:41 AM		
39	12322004987	Sep 4, 2024, 8:07:41 AM		
40	12347858622	Sep 7, 2024, 10:03:56 AM		
41	12368102065	Sep 9, 2024, 3:32:42 PM		
42	12380939837	Sep 11, 2024, 6:10:20 AM		
43	12416719456	Sep 15, 2024, 12:54:33 PM		
44	12435219995	Sep 17, 2024, 5:37:06 PM		
45	12477125422	Sep 22, 2024, 3:01:24 PM		
46	12523214664	Sep 28, 2024, 11:34:43 AM		
47	12588916760	Oct 6, 2024, 10:52:22 AM		
48	12607959321	Oct 8, 2024, 5:43:44 PM		
49	12662092835	Oct 15, 2024, 3:15:01 PM		
50	12690090965	Oct 19, 2024, 7:35:41 AM		
51	12867926253	Nov 10, 2024, 12:53:42 PM		
52	13403019941	Jan 20, 2025, 7:23:42 AM		

4	Løp på mølle etter skole🤓	Run
5	Bratte bakker med martin	Run
6	Templarheimen-telegrafbukta-sentrum	Run
7	Afternoon Run	Run
8	Løp te telegrafbukta	Run
9	Morning Run	Run
10	Evening Run	Run
11	Løp i solå	Run
12	restutisjonsløp💪🌞	Run
13	Løp te polet	Run
14	Afternoon Run	Run
15	tebake i stavaaaangeeeRRRRR!!!	Run
16	Dagen derpå… tungt det😮‍💨	Run
17	Æ det var tungt🤡	Run
18	Ups æ bada	Run
19	Løp med sabbie før sankt hans feiring💔‼️	Run
20	Vaaaaarmt🤡	Run
21	Løp på mølle	Run
22	Oppover og oppover og oppover…🤡	Run
23	Afternoon Run	Run
24	Post covid run på mølle😳🙈	Run
25	Back at it💥	Run
26	Dalsnuten	Run
27	3k	Run
28	Vålandstårnet og mosvannet	Run
29	Ullanhaugstårnet og tebake	Run
30	Jogga te trening	Run
31	Oppvarming på mølle	Run
32	Løp te trening	Run
33	Siste joggetur i stavanger i sommer	Run
34	Fuste løpetur i Tromsø dette semesteret😎	Run
35	Perf temp😌	Run
36	Te kirkegården og tebake	Run
37	Liten luftetur	Run
38	UURG	Run
39	Snørrete løp👱‍♂️🤧	Run
40	Hølje ned🌧️	Run
41	Litt bortfor fløya	Hike
42	Morning Run	Run
43	Deilig på en søndag😮‍💨	Run
44	Rolig jogg med pookie	Run
45	Brrr, litt kjølig nå🥶	Run
46	1 mil i sol og 6 grader🌞❄️	Run
47	Folksomt ute idag🏃	Run
48	Kveldstur med Martin🏃🦾	Run
49	Afternoon Run	Run
50	Morning Run	Run
51	Afternoon Run	Run
52	Til kraft	Walk

	Activity Description	Elapsed Time	Distance	\
0	NaN	2025	6.00	
1	NaN	1759	5.83	
2	NaN	827	3.01	
3	Løp 8 km med 1,5 stigning på mølle	2620	8.00	
4	NaN	1630	5.00	
5	NaN	1931	5.29	
6	Gikk som smuurt, føltes veldig bra heleveien	3163	10.01	
7	6 grader ute og shorts🌞 grei pace	2104	6.00	
8	20 varmegrader ute! Så ble et lite bad når me ...	1674	5.02	

9	Endelig har snøen smelta i lysløypå! ⛷️ ❤️	1904	6.00	
10		NaN	1498	5.00
11	Gikk veldig bra, var sykt varmt da	2188	6.83	
12		NaN	1777	3.86
13	Perf temp, sto og venta på verdens lengste rød...	3002	9.01	
14	Løp innom butikken og hjem	1712	4.44	
15	Nice å løpe i sørmarkå igjen 🌿 🌳, men sattan for...	2160	6.74	
16	Var ganske tungt. Mem deilig at det var mye vi...	2996	8.00	
17	Veldig mye oppoverbakker for en 11k..	3510	10.72	
18	Var digg å bade med fikk litt vondt under fote...	1210	4.01	
19		NaN	2207	5.56
20		NaN	3523	9.00
21	Oppvarming før mage økt 🤡	1000	3.00	
22	Såg en fasanfamilie! Bakken opp te ullanhaugst...	1587	3.77	
23		NaN	1657	5.16
24	Det føltes ikkje bra 😳 tung i pusten og støl i...	1800	4.50	
25	Det gikk bra! Andre løp etter korona, men hadd...	1738	5.00	
26	Sykt digg! Fint vær og nesten helt tomt for fo...	3059	3.35	
27	Igjen, treningsenteret på uis e så\nJÆVLIG var...	960	3.00	
28	Tungt og godt	2422	7.14	
29	NOOOOOOO eg kom borti pauseknappen under løpet...	1654	4.80	
30		NaN	1320	3.85
31	Kokvarmt på sis som alltid 😳 men selve løping...	1038	3.00	
32		NaN	1253	3.84
33	Va nice, litt stiv i kneet 🥹	1418	4.21	
34	Føltes veldig bra! Litt varmt bare, og de sist...	1581	5.00	
35	Veldig deilig løpetur 😊	3758	10.00	
36	Rolig jogg de fusste 3 kilometrene. Ganske ras...	2144	5.99	
37	Martin fikk gnagsår under foten av skonå 😣	811	1.82	
38	Det var treigt i kroppen.. e så støl etter tre...	2502	7.00	
39	Fusste dag med trening etter å ha vært syk i e...	1650	5.01	
40	Sykt digg å løpe i regnet 😤	2249	7.01	
41	Fikk pressa inn en liten fjelltur etter skolen...	7904	6.16	
42	Blir godt med frokost nå 😤 🥹	1619	5.00	
43	Sku egentlig ikkje springe så langt men det lu...	3401	9.40	
44	Veldig deilig, litt sånn halloweenstemning	1944	4.53	
45		NaN	1668	5.01
46	Meget godt 😓	3353	10.01	
47		NaN	1606	5.00
48	Va veldig digg	2568	7.01	
49		NaN	2025	6.01
50		NaN	1968	6.00
51		NaN	2059	6.00
52		NaN	2401	2.85

	Max Heart Rate	Relative Effort	Commute	...	Carbon Saved	Pool Length	\
0	NaN	NaN	False	...	NaN	NaN	
1	NaN	NaN	False	...	NaN	NaN	
2	NaN	NaN	False	...	NaN	NaN	
3	NaN	NaN	False	...	NaN	NaN	
4	NaN	NaN	False	...	NaN	NaN	
5	NaN	NaN	False	...	NaN	NaN	
6	NaN	NaN	False	...	NaN	NaN	
7	NaN	NaN	False	...	NaN	NaN	
8	NaN	NaN	False	...	NaN	NaN	
9	NaN	NaN	False	...	NaN	NaN	
10	NaN	NaN	False	...	NaN	NaN	
11	NaN	NaN	False	...	NaN	NaN	
12	NaN	NaN	False	...	NaN	NaN	
13	NaN	NaN	False	...	NaN	NaN	

14	NaN	NaN	False	...	NaN	NaN
15	NaN	NaN	False	...	NaN	NaN
16	NaN	NaN	False	...	NaN	NaN
17	NaN	NaN	False	...	NaN	NaN
18	NaN	NaN	False	...	NaN	NaN
19	NaN	NaN	False	...	NaN	NaN
20	NaN	NaN	False	...	NaN	NaN
21	NaN	NaN	False	...	NaN	NaN
22	NaN	NaN	False	...	NaN	NaN
23	NaN	NaN	False	...	NaN	NaN
24	NaN	NaN	False	...	NaN	NaN
25	NaN	NaN	False	...	NaN	NaN
26	NaN	NaN	False	...	NaN	NaN
27	NaN	NaN	False	...	NaN	NaN
28	NaN	NaN	False	...	NaN	NaN
29	NaN	NaN	False	...	NaN	NaN
30	NaN	NaN	False	...	NaN	NaN
31	NaN	NaN	False	...	NaN	NaN
32	NaN	NaN	False	...	NaN	NaN
33	NaN	NaN	False	...	NaN	NaN
34	NaN	NaN	False	...	NaN	NaN
35	NaN	NaN	False	...	NaN	NaN
36	NaN	NaN	False	...	NaN	NaN
37	NaN	NaN	False	...	NaN	NaN
38	NaN	NaN	False	...	NaN	NaN
39	NaN	NaN	False	...	NaN	NaN
40	NaN	NaN	False	...	NaN	NaN
41	NaN	NaN	False	...	NaN	NaN
42	NaN	NaN	False	...	NaN	NaN
43	NaN	NaN	False	...	NaN	NaN
44	NaN	NaN	False	...	NaN	NaN
45	NaN	NaN	False	...	NaN	NaN
46	NaN	NaN	False	...	NaN	NaN
47	NaN	NaN	False	...	NaN	NaN
48	NaN	NaN	False	...	NaN	NaN
49	NaN	NaN	False	...	NaN	NaN
50	NaN	NaN	False	...	NaN	NaN
51	NaN	NaN	False	...	NaN	NaN
52	NaN	NaN	False	...	NaN	NaN

	Training Load	Intensity	Average Grade Adjusted Pace	Timer Time	\
0	NaN	NaN	3.178656	NaN	
1	NaN	NaN	3.318436	NaN	
2	NaN	NaN	3.591707	NaN	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	3.073783	NaN	
6	NaN	NaN	3.167556	NaN	
7	NaN	NaN	2.999857	NaN	
8	NaN	NaN	3.047023	NaN	
9	NaN	NaN	3.245483	NaN	
10	NaN	NaN	3.405906	NaN	
11	NaN	NaN	3.293012	NaN	
12	NaN	NaN	2.243204	NaN	
13	NaN	NaN	3.071404	NaN	
14	NaN	NaN	3.096923	NaN	
15	NaN	NaN	3.301594	NaN	
16	NaN	NaN	3.131291	NaN	
17	NaN	NaN	3.159788	NaN	
18	NaN	NaN	3.341625	NaN	

19	NaN	NaN	2.663355	NaN
20	NaN	NaN	2.912657	NaN
21	NaN	NaN	NaN	NaN
22	NaN	NaN	3.043963	NaN
23	NaN	NaN	3.185167	NaN
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	NaN	NaN	1.971398	NaN
27	NaN	NaN	NaN	NaN
28	NaN	NaN	3.120135	NaN
29	NaN	NaN	3.124420	NaN
30	NaN	NaN	3.042576	NaN
31	NaN	NaN	NaN	NaN
32	NaN	NaN	3.238561	NaN
33	NaN	NaN	3.173074	NaN
34	NaN	NaN	3.229981	NaN
35	NaN	NaN	2.826743	NaN
36	NaN	NaN	2.948909	NaN
37	NaN	NaN	2.634259	NaN
38	NaN	NaN	2.907398	NaN
39	NaN	NaN	3.089793	NaN
40	NaN	NaN	3.194382	NaN
41	NaN	NaN	NaN	NaN
42	NaN	NaN	3.137430	NaN
43	NaN	NaN	2.918823	NaN
44	NaN	NaN	2.573085	NaN
45	NaN	NaN	3.191632	NaN
46	NaN	NaN	3.074574	NaN
47	NaN	NaN	3.178862	NaN
48	NaN	NaN	2.849507	NaN
49	NaN	NaN	3.077501	NaN
50	NaN	NaN	3.170600	NaN
51	NaN	NaN	3.034329	NaN
52	NaN	NaN	NaN	NaN

	Total Cycles	Media
0	NaN	media/6DD4200E-5443-4D7C-A824-B0396B6B3839.jpg...
1	NaN	media/995194E5-1508-44C3-A0F0-B2795E0C2E39.jpg
2	NaN	media/F2F78DC7-C6E4-493F-B42C-F096B761C3EB.jpg...
3	NaN	NaN
4	NaN	NaN
5	NaN	media/683F2556-A97F-4A2F-A2D1-0495EBBD815F.jpg...
6	NaN	media/B119778F-2011-4784-A434-0425CFFEF1EF.jpg...
7	NaN	media/35529E64-4AC9-44A6-A285-F92AB1BC56A3.jpg
8	NaN	media/93067FFF-C006-48D1-BC72-9D3C6641B21F.mp4...
9	NaN	NaN
10	NaN	media/46284B7B-0370-4C5C-A8FA-CEC23766BA1B.jpg
11	NaN	media/15FA89F4-91F7-45A1-8C6C-525272512726.jpg...
12	NaN	NaN
13	NaN	media/A923B9E7-3983-4473-8C5F-FC5A020C2991.jpg...
14	NaN	media/562F0A5A-0779-4737-9810-826C394AC7CD.jpg...
15	NaN	media/1C7F6E96-FF99-4238-85EE-9433D2BD6D26.jpg...
16	NaN	media/F9A7C297-147C-4542-B3B5-2F6C7D0FCDE3.jpg...
17	NaN	media/1A40EA82-B27F-4ECC-9A64-9401A067E2DD.jpg
18	NaN	media/F108A5BE-0C90-4D4F-9E6F-972944E3388E.jpg
19	NaN	media/06BDB864-69DF-44A7-AE8E-0BB9D8581AD5.jpg
20	NaN	media/5B515018-6CA7-49EE-9436-E7C09C1FC70E.jpg
21	NaN	NaN
22	NaN	media/DA0ABE14-3452-4D85-B311-6EC3DE98B3C3.jpg...
23	NaN	NaN

24	NaN	NaN
25	NaN	NaN
26	NaN	media/B62310C7-82C3-4E7C-98A1-056ECD848D7A.jpg...
27	NaN	NaN
28	NaN	media/67B058AE-A0DF-4253-9ABD-82D5D493DFCB.jpg
29	NaN	media/1486BC1C-369A-43BF-AB52-981F14278D03.jpg...
30	NaN	NaN
31	NaN	NaN
32	NaN	NaN
33	NaN	media/F14F9451-4CC2-4B89-8D18-40C357308D8D.jpg
34	NaN	media/A353707E-98B7-4BB8-91C1-BBC2806D8402.jpg...
35	NaN	media/88DD6D06-0B1D-4D34-934C-63D019BD338C.jpg...
36	NaN	media/71011CE1-111D-4DBC-988B-66016E00D835.jpg...
37	NaN	media/D44B4114-87E7-4A0C-B200-494706FFAFDB.jpg
38	NaN	media/E4ACBEFF-6F2C-450F-BCC5-239AD41C47B7.jpg
39	NaN	media/1A1D4CD3-7707-4B01-BE22-322555550482.jpg
40	NaN	NaN
41	10412.0	media/BF7842C5-11EB-4281-8BA4-69B8EDEFBB62.jpg...
42	NaN	media/E0FCC385-02C0-4FC8-8DB1-DDB97B9DADF1.jpg
43	NaN	media/0334290C-CB32-4D07-BB48-4FCDD4B6411E.jpg...
44	NaN	media/DF8ACE2E-1CCA-4BC9-85EC-4C2CDC115179.jpg
45	NaN	media/A5922A48-0C37-4CF6-B965-27578D998162.jpg
46	NaN	media/86001F29-3DC6-4C31-93C3-EA3557231D3C.jpg...
47	NaN	media/17E35EFB-7DB8-4A43-B823-63C988D8C706.jpg
48	NaN	media/F5521441-37EF-42D6-8885-724CCC62DAA5.jpg
49	NaN	media/42F310B2-0A9C-43D2-9D49-2460D8096E67.jpg
50	NaN	media/CA124C9F-DDD8-462F-820C-E158A2D86577.jpg
51	NaN	media/107691BC-07C3-4448-B401-88D9645F865C.jpg...
52	NaN	NaN

	time_minutes	average_pace
0	33.750000	5.625000
1	29.316667	5.028588
2	13.783333	4.579181
3	43.666667	5.458333
4	27.166667	5.433333
5	32.183333	6.083806
6	52.716667	5.266400
7	35.066667	5.844444
8	27.900000	5.557769
9	31.733333	5.288889
10	24.966667	4.993333
11	36.466667	5.339190
12	29.616667	7.672712
13	50.033333	5.553089
14	28.533333	6.426426
15	36.000000	5.341246
16	49.933333	6.241667
17	58.500000	5.457090
18	20.166667	5.029094
19	36.783333	6.615707
20	58.716667	6.524074
21	16.666667	5.555556
22	26.450000	7.015915
23	27.616667	5.352067
24	30.000000	6.666667
25	28.966667	5.793333
26	50.983333	15.218905
27	16.000000	5.333333
28	40.366667	5.653595

```
29     27.566667      5.743056
30     22.000000      5.714286
31     17.300000      5.766667
32     20.883333      5.438368
33     23.633333      5.613618
34     26.350000      5.270000
35     62.633333      6.263333
36     35.733333      5.965498
37     13.516667      7.426740
38     41.700000      5.957143
39     27.500000      5.489022
40     37.483333      5.347123
41    131.733333     21.385281
42     26.983333      5.396667
43     56.683333      6.030142
44     32.400000      7.152318
45     27.800000      5.548902
46     55.883333      5.582751
47     26.766667      5.353333
48     42.800000      6.105563
49     33.750000      5.615641
50     32.800000      5.466667
51     34.316667      5.719444
52     40.016667     14.040936

[53 rows x 96 columns]
```

In [8]:
```
df.dtypes #showing the diffrent datatypes of variables i use.
```

Out[8]:
```
Activity ID               int64
Activity Date            object
Activity Name            object
Activity Type            object
Activity Description     object
                          ...
Timer Time              float64
Total Cycles            float64
Media                    object
time_minutes            float64
average_pace            float64
Length: 96, dtype: object
```

In [9]:
```
# extract hour of the day form activety date
df['Activity Date'] = pd.to_datetime(df['Activity Date'])
df["hour"] = df['Activity Date'].dt.hour  # Extract hour
```

In [10]:
```
print(df.columns)  # Check column names, want to see if the added variabel "arve
```

```
Index(['Activity ID', 'Activity Date', 'Activity Name', 'Activity Type',
       'Activity Description', 'Elapsed Time', 'Distance', 'Max Heart Rate',
       'Relative Effort', 'Commute', 'Activity Private Note', 'Activity Gear',
       'Filename', 'Athlete Weight', 'Bike Weight', 'Elapsed Time.1',
       'Moving Time', 'Distance.1', 'Max Speed', 'Average Speed',
       'Elevation Gain', 'Elevation Loss', 'Elevation Low', 'Elevation High',
       'Max Grade', 'Average Grade', 'Average Positive Grade',
       'Average Negative Grade', 'Max Cadence', 'Average Cadence',
       'Max Heart Rate.1', 'Average Heart Rate', 'Max Watts', 'Average Watts',
       'Calories', 'Max Temperature', 'Average Temperature',
       'Relative Effort.1', 'Total Work', 'Number of Runs', 'Uphill Time',
       'Downhill Time', 'Other Time', 'Perceived Exertion', 'Type',
       'Start Time', 'Weighted Average Power', 'Power Count',
       'Prefer Perceived Exertion', 'Perceived Relative Effort', 'Commute.1',
       'Total Weight Lifted', 'From Upload', 'Grade Adjusted Distance',
       'Weather Observation Time', 'Weather Condition', 'Weather Temperature',
       'Apparent Temperature', 'Dewpoint', 'Humidity', 'Weather Pressure',
       'Wind Speed', 'Wind Gust', 'Wind Bearing', 'Precipitation Intensity',
       'Sunrise Time', 'Sunset Time', 'Moon Phase', 'Bike', 'Gear',
       'Precipitation Probability', 'Precipitation Type', 'Cloud Cover',
       'Weather Visibility', 'UV Index', 'Weather Ozone', 'Jump Count',
       'Total Grit', 'Average Flow', 'Flagged', 'Average Elapsed Speed',
       'Dirt Distance', 'Newly Explored Distance',
       'Newly Explored Dirt Distance', 'Activity Count', 'Total Steps',
       'Carbon Saved', 'Pool Length', 'Training Load', 'Intensity',
       'Average Grade Adjusted Pace', 'Timer Time', 'Total Cycles', 'Media',
       'time_minutes', 'average_pace', 'hour'],
      dtype='object')
```

In [11]:
```
#now i want to exclude all activetypes which are not runs, because the df contai
```

In [12]:
```
print(df["Activity Type"].unique()) #define types
```

```
['Run' 'Hike' 'Walk']
```

In [13]:
```
df_runs = df[df["Activity Type"] == "Run"] #filter dataset to only include run
```

In [14]:
```
print(df_runs)
```

	Activity ID	Activity Date	Activity Name
0	11288364028	2024-04-29 05:14:04	Morning Run
1	11327182053	2024-05-04 06:54:55	Morning Run
2	11354165092	2024-05-07 16:31:09	Evening Run
3	11371024533	2024-05-09 16:22:16	Løp på mølle, etter skolen 💖
4	11402044804	2024-05-13 14:30:00	Løp på mølle etter skole 😳
5	11425811688	2024-05-16 16:58:18	Bratte bakker med martin
6	11447612172	2024-05-19 12:34:48	Templarheimen-telegrafbukta-sentrum
7	11463824675	2024-05-21 14:31:21	Afternoon Run
8	11478659771	2024-05-23 12:47:45	Løp te telegrafbukta
9	11491349434	2024-05-25 08:00:08	Morning Run
10	11504674828	2024-05-26 18:26:35	Evening Run
11	11518449873	2024-05-28 16:55:00	Løp i solå
12	11535082647	2024-05-30 18:26:47	restutisjonsløp 💪 🌞
13	11539555466	2024-05-31 11:27:34	Løp te polet
14	11563861067	2024-06-03 14:55:18	Afternoon Run
15	11597675325	2024-06-07 17:41:10	tebake i stavaaaangeeeRRRRR!!!
16	11612344724	2024-06-09 13:28:23	Dagen derpå… tungt det 😮‍💨
17	11656959150	2024-06-15 10:28:28	Æ det var tungt 🤯
18	11674087087	2024-06-17 14:08:23	Ups æ bada
19	11719328361	2024-06-23 11:14:45	Løp med sabbie før sankt hans feiring 💖 ‼️
20	11736909938	2024-06-25 15:10:45	Vaaaaarmt 🤯
21	11752483688	2024-06-27 14:13:54	Løp på mølle
22	11772700697	2024-06-30 08:24:37	Oppover og oppover og oppover… 🤯
23	11798960902	2024-07-03 14:23:25	Afternoon Run
24	11926180574	2024-07-16 16:00:32	Post covid run på mølle 😳 🙈
25	11926206972	2024-07-18 14:00:00	Back at it 💥
26	11932610893	2024-07-20 06:35:59	Dalsnuten
27	11953922104	2024-07-22 18:01:58	3k
28	12012007482	2024-07-29 17:11:08	Vålandstårnet og mosvannet
29	12029965346	2024-07-31 18:29:48	Ullanhaugstårnet og tebake
30	12036820394	2024-08-01 15:46:04	Jogga te trening
31	12051297516	2024-08-03 12:05:16	Oppvarming på mølle
32	12077888082	2024-08-06 14:32:10	Løp te trening
33	12089152285	2024-08-07 18:36:14	Siste joggetur i stavanger i sommer
34	12111709943	2024-08-10 13:53:21	Fuste løpetur i Tromsø dette semesteret 😎
35	12126628243	2024-08-12 09:19:56	Perf temp 😌
36	12156193243	2024-08-15 17:20:15	Te kirkegården og tebake
37	12180076216	2024-08-18 13:10:25	Liten luftetur
38	12203349729	2024-08-21 09:34:41	UURG
39	12322004987	2024-09-04 08:07:41	Snørrete løp 🤧 😤
40	12347858622	2024-09-07 10:03:56	Hølje ned 🌧️
42	12380939837	2024-09-11 06:10:20	Morning Run
43	12416719456	2024-09-15 12:54:33	Deilig på en søndag 😮‍💨
44	12435219995	2024-09-17 17:37:06	Rolig jogg med pookie
45	12477125422	2024-09-22 15:01:24	Brrr, litt kjølig nå 🥶
46	12523214664	2024-09-28 11:34:43	1 mil i sol og 6 grader 🌞 ❄️
47	12588916760	2024-10-06 10:52:22	Folksomt ute idag 🏃
48	12607959321	2024-10-08 17:43:44	Kveldstur med Martin 🏃 🚴
49	12662092835	2024-10-15 15:15:01	Afternoon Run
50	12690090965	2024-10-19 07:35:41	Morning Run
51	12867926253	2024-11-10 12:53:42	Afternoon Run

	Activity Type	Activity Description
0	Run	NaN
1	Run	NaN
2	Run	NaN
3	Run	Løp 8 km med 1,5 stigning på mølle
4	Run	NaN
5	Run	NaN

6	Run	Gikk som smuurt, føltes veldig bra heleveien
7	Run	6 grader ute og shorts🌞 grei pace
8	Run	20 varmegrader ute! Så ble et lite bad når me ...
9	Run	Endelig har snøen smelta i lysløypå!🏃‍♂️❤️‍🩹
10	Run	NaN
11	Run	Gikk veldig bra, var sykt varmt da
12	Run	NaN
13	Run	Perf temp, sto og venta på verdens lengste rød...
14	Run	Løp innom butikken og hjem
15	Run	Nice å løpe i sørmarkå igjen🌿🌳, men sattan for...
16	Run	Var ganske tungt. Mem deilig at det var mye vi...
17	Run	Veldig mye oppoverbakker for en 11k..
18	Run	Var digg å bade med fikk litt vondt under fote...
19	Run	NaN
20	Run	NaN
21	Run	Oppvarming før mage økt🤡
22	Run	Såg en fasanfamilie! Bakken opp te ullanhaugst...
23	Run	NaN
24	Run	Det føltes ikkje bra😳tung i pusten og støl i...
25	Run	Det gikk bra! Andre løp etter korona, men hadd...
26	Run	Sykt digg! Fint vær og nesten helt tomt for fo...
27	Run	Igjen, treningsenteret på uis e så\nJÆVLIG var...
28	Run	Tungt og godt
29	Run	NOOOOOOO eg kom borti pauseknappen under løpet...
30	Run	NaN
31	Run	Kokvarmt på sis som alltid😳 men selve løping...
32	Run	NaN
33	Run	Va nice, litt stiv i kneet🥹
34	Run	Føltes veldig bra! Litt varmt bare, og de sist...
35	Run	Veldig deilig løpetur😊
36	Run	Rolig jogg de fusste 3 kilometrene. Ganske ras...
37	Run	Martin fikk gnagsår under foten av skonå😖
38	Run	Det var treigt i kroppen.. e så støl etter tre...
39	Run	Fusste dag med trening etter å ha vært syk i e...
40	Run	Sykt digg å løpe i regnet😤
42	Run	Blir godt med frokost nå 😤🥹
43	Run	Sku egentlig ikkje springe så langt men det lu...
44	Run	Veldig deilig, litt sånn halloweenstemning
45	Run	NaN
46	Run	Meget godt🥵
47	Run	NaN
48	Run	Va veldig digg
49	Run	NaN
50	Run	NaN
51	Run	NaN

	Elapsed Time	Distance	Max Heart Rate	Relative Effort	Commute	...	\
0	2025	6.00	NaN	NaN	False	...	
1	1759	5.83	NaN	NaN	False	...	
2	827	3.01	NaN	NaN	False	...	
3	2620	8.00	NaN	NaN	False	...	
4	1630	5.00	NaN	NaN	False	...	
5	1931	5.29	NaN	NaN	False	...	
6	3163	10.01	NaN	NaN	False	...	
7	2104	6.00	NaN	NaN	False	...	
8	1674	5.02	NaN	NaN	False	...	
9	1904	6.00	NaN	NaN	False	...	
10	1498	5.00	NaN	NaN	False	...	
11	2188	6.83	NaN	NaN	False	...	
12	1777	3.86	NaN	NaN	False	...	

13	3002	9.01	NaN	NaN	False	...
14	1712	4.44	NaN	NaN	False	...
15	2160	6.74	NaN	NaN	False	...
16	2996	8.00	NaN	NaN	False	...
17	3510	10.72	NaN	NaN	False	...
18	1210	4.01	NaN	NaN	False	...
19	2207	5.56	NaN	NaN	False	...
20	3523	9.00	NaN	NaN	False	...
21	1000	3.00	NaN	NaN	False	...
22	1587	3.77	NaN	NaN	False	...
23	1657	5.16	NaN	NaN	False	...
24	1800	4.50	NaN	NaN	False	...
25	1738	5.00	NaN	NaN	False	...
26	3059	3.35	NaN	NaN	False	...
27	960	3.00	NaN	NaN	False	...
28	2422	7.14	NaN	NaN	False	...
29	1654	4.80	NaN	NaN	False	...
30	1320	3.85	NaN	NaN	False	...
31	1038	3.00	NaN	NaN	False	...
32	1253	3.84	NaN	NaN	False	...
33	1418	4.21	NaN	NaN	False	...
34	1581	5.00	NaN	NaN	False	...
35	3758	10.00	NaN	NaN	False	...
36	2144	5.99	NaN	NaN	False	...
37	811	1.82	NaN	NaN	False	...
38	2502	7.00	NaN	NaN	False	...
39	1650	5.01	NaN	NaN	False	...
40	2249	7.01	NaN	NaN	False	...
42	1619	5.00	NaN	NaN	False	...
43	3401	9.40	NaN	NaN	False	...
44	1944	4.53	NaN	NaN	False	...
45	1668	5.01	NaN	NaN	False	...
46	3353	10.01	NaN	NaN	False	...
47	1606	5.00	NaN	NaN	False	...
48	2568	7.01	NaN	NaN	False	...
49	2025	6.01	NaN	NaN	False	...
50	1968	6.00	NaN	NaN	False	...
51	2059	6.00	NaN	NaN	False	...

	Pool Length	Training Load	Intensity	Average Grade Adjusted Pace	\
0	NaN	NaN	NaN	3.178656	
1	NaN	NaN	NaN	3.318436	
2	NaN	NaN	NaN	3.591707	
3	NaN	NaN	NaN	NaN	
4	NaN	NaN	NaN	NaN	
5	NaN	NaN	NaN	3.073783	
6	NaN	NaN	NaN	3.167556	
7	NaN	NaN	NaN	2.999857	
8	NaN	NaN	NaN	3.047023	
9	NaN	NaN	NaN	3.245483	
10	NaN	NaN	NaN	3.405906	
11	NaN	NaN	NaN	3.293012	
12	NaN	NaN	NaN	2.243204	
13	NaN	NaN	NaN	3.071404	
14	NaN	NaN	NaN	3.096923	
15	NaN	NaN	NaN	3.301594	
16	NaN	NaN	NaN	3.131291	
17	NaN	NaN	NaN	3.159788	
18	NaN	NaN	NaN	3.341625	
19	NaN	NaN	NaN	2.663355	

20	NaN	NaN	NaN	2.912657
21	NaN	NaN	NaN	NaN
22	NaN	NaN	NaN	3.043963
23	NaN	NaN	NaN	3.185167
24	NaN	NaN	NaN	NaN
25	NaN	NaN	NaN	NaN
26	NaN	NaN	NaN	1.971398
27	NaN	NaN	NaN	NaN
28	NaN	NaN	NaN	3.120135
29	NaN	NaN	NaN	3.124420
30	NaN	NaN	NaN	3.042576
31	NaN	NaN	NaN	NaN
32	NaN	NaN	NaN	3.238561
33	NaN	NaN	NaN	3.173074
34	NaN	NaN	NaN	3.229981
35	NaN	NaN	NaN	2.826743
36	NaN	NaN	NaN	2.948909
37	NaN	NaN	NaN	2.634259
38	NaN	NaN	NaN	2.907398
39	NaN	NaN	NaN	3.089793
40	NaN	NaN	NaN	3.194382
42	NaN	NaN	NaN	3.137430
43	NaN	NaN	NaN	2.918823
44	NaN	NaN	NaN	2.573085
45	NaN	NaN	NaN	3.191632
46	NaN	NaN	NaN	3.074574
47	NaN	NaN	NaN	3.178862
48	NaN	NaN	NaN	2.849507
49	NaN	NaN	NaN	3.077501
50	NaN	NaN	NaN	3.170600
51	NaN	NaN	NaN	3.034329

	Timer Time	Total Cycles	\
0	NaN	NaN	
1	NaN	NaN	
2	NaN	NaN	
3	NaN	NaN	
4	NaN	NaN	
5	NaN	NaN	
6	NaN	NaN	
7	NaN	NaN	
8	NaN	NaN	
9	NaN	NaN	
10	NaN	NaN	
11	NaN	NaN	
12	NaN	NaN	
13	NaN	NaN	
14	NaN	NaN	
15	NaN	NaN	
16	NaN	NaN	
17	NaN	NaN	
18	NaN	NaN	
19	NaN	NaN	
20	NaN	NaN	
21	NaN	NaN	
22	NaN	NaN	
23	NaN	NaN	
24	NaN	NaN	
25	NaN	NaN	
26	NaN	NaN	

```
27         NaN           NaN
28         NaN           NaN
29         NaN           NaN
30         NaN           NaN
31         NaN           NaN
32         NaN           NaN
33         NaN           NaN
34         NaN           NaN
35         NaN           NaN
36         NaN           NaN
37         NaN           NaN
38         NaN           NaN
39         NaN           NaN
40         NaN           NaN
42         NaN           NaN
43         NaN           NaN
44         NaN           NaN
45         NaN           NaN
46         NaN           NaN
47         NaN           NaN
48         NaN           NaN
49         NaN           NaN
50         NaN           NaN
51         NaN           NaN
```

	Media	time_minutes	\
0	media/6DD4200E-5443-4D7C-A824-B0396B6B3839.jpg...	33.750000	
1	media/995194E5-1508-44C3-A0F0-B2795E0C2E39.jpg	29.316667	
2	media/F2F78DC7-C6E4-493F-B42C-F096B761C3EB.jpg...	13.783333	
3	NaN	43.666667	
4	NaN	27.166667	
5	media/683F2556-A97F-4A2F-A2D1-0495EBBD815F.jpg...	32.183333	
6	media/B119778F-2011-4784-A434-0425CFFEF1EF.jpg...	52.716667	
7	media/35529E64-4AC9-44A6-A285-F92AB1BC56A3.jpg	35.066667	
8	media/93067FFF-C006-48D1-BC72-9D3C6641B21F.mp4...	27.900000	
9	NaN	31.733333	
10	media/46284B7B-0370-4C5C-A8FA-CEC23766BA1B.jpg	24.966667	
11	media/15FA89F4-91F7-45A1-8C6C-525272512726.jpg...	36.466667	
12	NaN	29.616667	
13	media/A923B9E7-3983-4473-8C5F-FC5A020C2991.jpg...	50.033333	
14	media/562F0A5A-0779-4737-9810-826C394AC7CD.jpg...	28.533333	
15	media/1C7F6E96-FF99-4238-85EE-9433D2BD6D26.jpg...	36.000000	
16	media/F9A7C297-147C-4542-B3B5-2F6C7D0FCDE3.jpg...	49.933333	
17	media/1A40EA82-B27F-4ECC-9A64-9401A067E2DD.jpg	58.500000	
18	media/F108A5BE-0C90-4D4F-9E6F-972944E3388E.jpg	20.166667	
19	media/06BDB864-69DF-44A7-AE8E-0BB9D8581AD5.jpg	36.783333	
20	media/5B515018-6CA7-49EE-9436-E7C09C1FC70E.jpg	58.716667	
21	NaN	16.666667	
22	media/DA0ABE14-3452-4D85-B311-6EC3DE98B3C3.jpg...	26.450000	
23	NaN	27.616667	
24	NaN	30.000000	
25	NaN	28.966667	
26	media/B62310C7-82C3-4E7C-98A1-056ECD848D7A.jpg...	50.983333	
27	NaN	16.000000	
28	media/67B058AE-A0DF-4253-9ABD-82D5D493DFCB.jpg	40.366667	
29	media/1486BC1C-369A-43BF-AB52-981F14278D03.jpg...	27.566667	
30	NaN	22.000000	
31	NaN	17.300000	
32	NaN	20.883333	
33	media/F14F9451-4CC2-4B89-8D18-40C357308D8D.jpg	23.633333	

34	media/A353707E-98B7-4BB8-91C1-BBC2806D8402.jpg...	26.350000
35	media/88DD6D06-0B1D-4D34-934C-63D019BD338C.jpg...	62.633333
36	media/71011CE1-111D-4DBC-988B-66016E00D835.jpg...	35.733333
37	media/D44B4114-87E7-4A0C-B200-494706FFAFDB.jpg	13.516667
38	media/E4ACBEFF-6F2C-450F-BCC5-239AD41C47B7.jpg	41.700000
39	media/1A1D4CD3-7707-4B01-BE22-322555550482.jpg	27.500000
40	NaN	37.483333
42	media/E0FCC385-02C0-4FC8-8DB1-DDB97B9DADF1.jpg	26.983333
43	media/0334290C-CB32-4D07-BB48-4FCDD4B6411E.jpg...	56.683333
44	media/DF8ACE2E-1CCA-4BC9-85EC-4C2CDC115179.jpg	32.400000
45	media/A5922A48-0C37-4CF6-B965-27578D998162.jpg	27.800000
46	media/86001F29-3DC6-4C31-93C3-EA3557231D3C.jpg...	55.883333
47	media/17E35EFB-7DB8-4A43-B823-63C988D8C706.jpg	26.766667
48	media/F5521441-37EF-42D6-8885-724CCC62DAA5.jpg	42.800000
49	media/42F310B2-0A9C-43D2-9D49-2460D8096E67.jpg	33.750000
50	media/CA124C9F-DDD8-462F-820C-E158A2D86577.jpg	32.800000
51	media/107691BC-07C3-4448-B401-88D9645F865C.jpg...	34.316667

	average_pace	hour
0	5.625000	5
1	5.028588	6
2	4.579181	16
3	5.458333	16
4	5.433333	14
5	6.083806	16
6	5.266400	12
7	5.844444	14
8	5.557769	12
9	5.288889	8
10	4.993333	18
11	5.339190	16
12	7.672712	18
13	5.553089	11
14	6.426426	14
15	5.341246	17
16	6.241667	13
17	5.457090	10
18	5.029094	14
19	6.615707	11
20	6.524074	15
21	5.555556	14
22	7.015915	8
23	5.352067	14
24	6.666667	16
25	5.793333	14
26	15.218905	6
27	5.333333	18
28	5.653595	17
29	5.743056	18
30	5.714286	15
31	5.766667	12
32	5.438368	14
33	5.613618	18
34	5.270000	13
35	6.263333	9
36	5.965498	17
37	7.426740	13
38	5.957143	9
39	5.489022	8
40	5.347123	10

```
42      5.396667     6
43      6.030142    12
44      7.152318    17
45      5.548902    15
46      5.582751    11
47      5.353333    10
48      6.105563    17
49      5.615641    15
50      5.466667     7
51      5.719444    12

[51 rows x 97 columns]
```

My first hypothesis: Time of day effects the average pace on all my runs

In my first hypothesis, I aimed to determine whether my running performance varies based on the time of day, specifically, if there are certain times when I achieve better or worse results.

It's important to note that during the colder months (January through April), I primarily ran on a treadmill, and those runs were not recorded on Strava. As a result, my Strava data mainly reflects my outdoor runs, though not exclusively.

In [17]:
```python
#Classifies whether the hour of the run is in the morning, evening or afternoon
def classify_time(hour):
    if hour < 13:
        return "Morning" # if run is before before 13.00
    elif hour >= 18:
        return "Evening" # if run is after 18.00
    else:
        return "Afternoon"  # Handle hours between 13.00 and 18.00

df["time_of_day"] = df["hour"].apply(classify_time) #converts hours in to time o
```

In [18]:
```python
print(df["time_of_day"]) #Showing the convertion
```

```
0       Morning
1       Morning
2     Afternoon
3     Afternoon
4     Afternoon
5     Afternoon
6       Morning
7     Afternoon
8       Morning
9       Morning
10      Evening
11    Afternoon
12      Evening
13      Morning
14    Afternoon
15    Afternoon
16    Afternoon
17      Morning
18    Afternoon
19      Morning
20    Afternoon
21    Afternoon
22      Morning
23    Afternoon
24    Afternoon
25    Afternoon
26      Morning
27      Evening
28    Afternoon
29      Evening
30    Afternoon
31      Morning
32    Afternoon
33      Evening
34    Afternoon
35      Morning
36    Afternoon
37    Afternoon
38      Morning
39      Morning
40      Morning
41    Afternoon
42      Morning
43      Morning
44    Afternoon
45    Afternoon
46      Morning
47      Morning
48    Afternoon
49    Afternoon
50      Morning
51      Morning
52      Morning
Name: time_of_day, dtype: object
```

In [19]:
```
# Because I had a substantial outlier, it makes the plot difficult to read. I th
df_runs = df_runs.drop(26)  # Remove the row with index 26
```

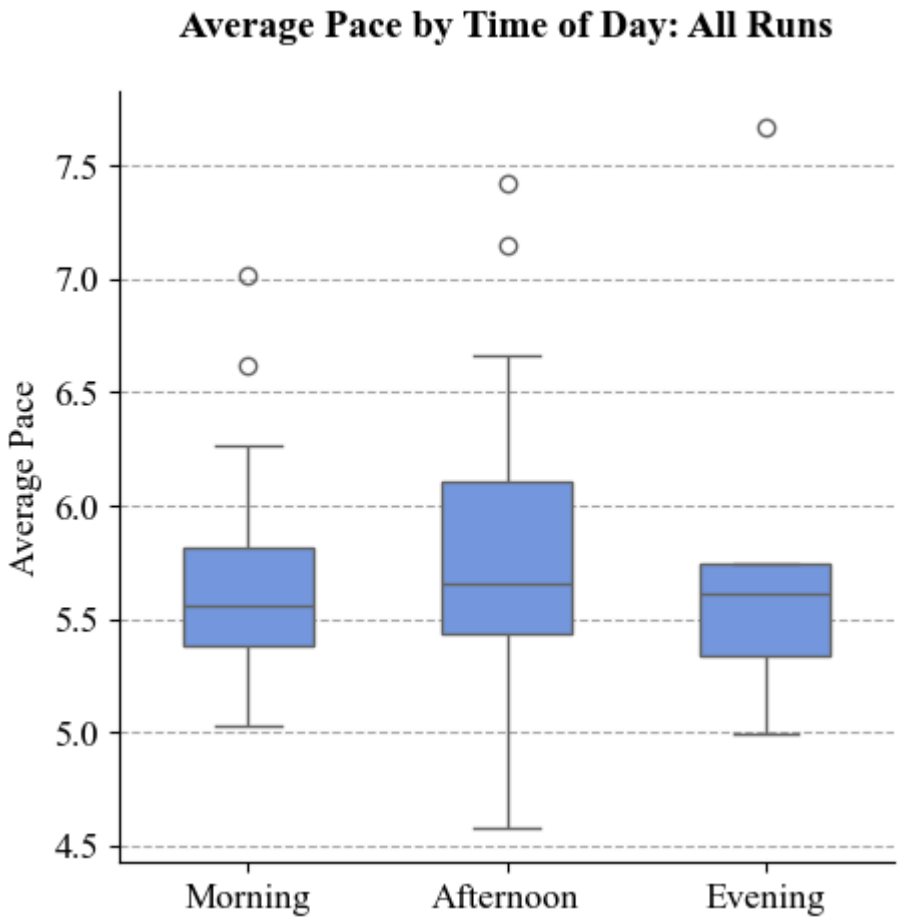
boxplot for all my runs average pace:

The boxplot shows there is no substantial change in running pace across all of my runs based on the time of day. I have an average pace just above 5.5 in the morning, afternoon, and evening. Based on the plot, I can conclude that the time of day does not affect my running results.

Boxplots are useful for comparing averages like these. They provide a clear picture of the average pace while also displaying the distribution of data within each category.

To achieve a clean, elegant look that allows for simple interpretation, I chose a calm color and a clean, easy-to-read font, with thin lines tracing the figures. To make the plot easily interpretable, I added some gray horizontal lines in the background so you can quickly see which pace on the y-axis the average pace in the boxplot aligns with. The figure size is set to 5.5, making the boxplots slimmer and easier to visually compare.

In [21]:
```
#boxplot for all runs
plt.figure(figsize=(5, 5))
plt.grid(axis='y', linestyle="--", alpha=0.7, color="gray") #adds grey lines to
sns.boxplot(
    x=df["time_of_day"],
    y=df_runs["average_pace"],
    color="cornflowerblue",
    width=0.5,  # adjust box width
    linewidth=1.0  # adjust contoring lines
)

#labels and titles
plt.xlabel("") #didn't want a lable for x axis
plt.ylabel("Average Pace", fontsize=13, fontname="Times New Roman")
plt.title("Average Pace by Time of Day: All Runs", fontsize=14, fontname="Times

plt.xticks(fontsize=13, fontname="Times New Roman")
plt.yticks(fontsize=13, fontname="Times New Roman")

sns.despine()  # Removes top and right borders for a cleaner look
plt.show()

#boxplot for evning does not have a upper whisker, this is because of seaborns s
```



Second hypothesis: Time of day effects the pace on my 4-6km runs

After finding no substantial diffrence between all my runs depending on time of day, i wanted to further investigate if a effect could be found in only the 4-6 runs.

In [24]:
```
df_4_6k = df[(df["Distance"] >= 4) & (df["Distance"] <= 6)]  # Keep runs between
```

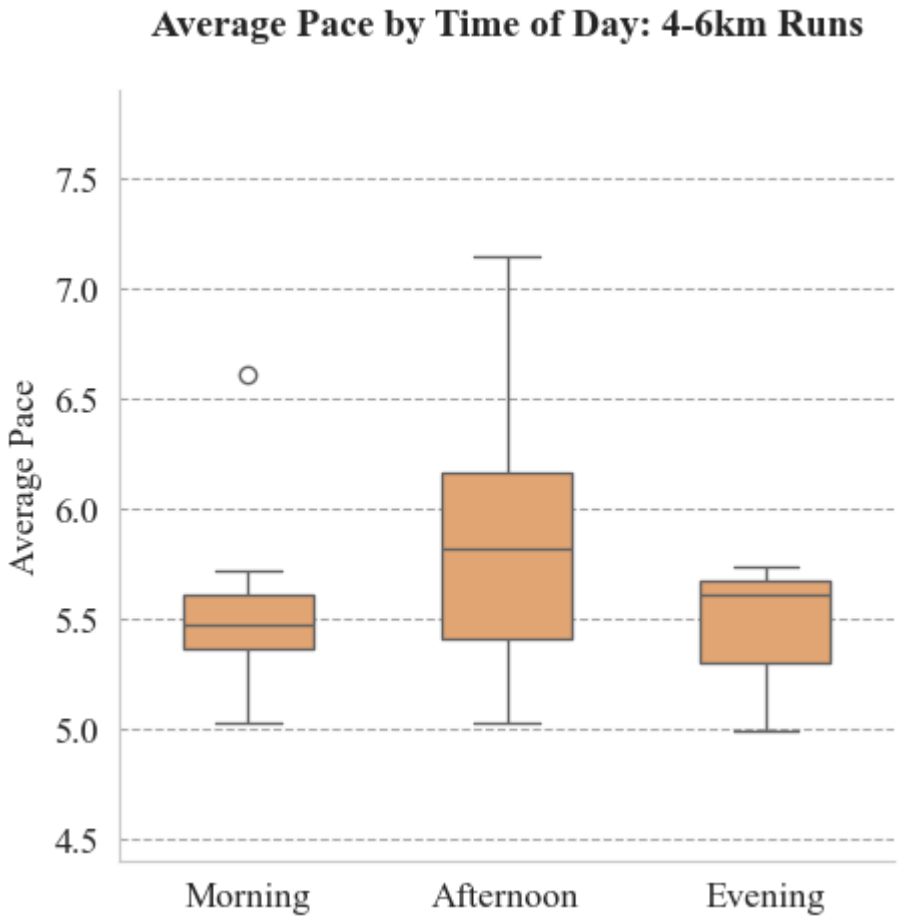
Boxplot for 4-6km runs:

The 4-6km runs show a bigger mean diffrence in pace, based on time of day. In the morning I run the fastest, with a mean pace just below 5.5 minutes per kilometer. In the evening the pace is a bit slower, with just about ca 5.6 minutes per kilometer. I run the slowest in the afternoon with a pace that is closer to 6.

To make this plot easily compareble with the boxplot in H1, i want the figures to have an identical structure. So i use the same lines in the background, i use the same figure size and i use the same values on the y-axsis

In [26]:
```
# Set Seaborn style
sns.set_style("whitegrid")

#boxplot for 4-6km runs
plt.figure(figsize=(5, 5))
```

```
plt.grid(axis='y', linestyle="--", alpha=0.7, color="gray")
sns.boxplot(
    x=df_4_6k["time_of_day"],
    y=df_4_6k["average_pace"],
    color="#F4A460",
    width=0.5,
    linewidth=1.0,
)

# Labels and title
plt.xlabel("")
plt.ylabel("Average Pace", fontsize=13, fontname="Times new roman")
plt.title("Average Pace by Time of Day: 4-6km Runs", fontsize=14, fontname="Time

plt.xticks(fontname="Times new roman", fontsize=13)
plt.yticks(fontname="Times new roman", fontsize=13)
plt.ylim(4.4, 7.9)

sns.despine()
plt.show()
```



Thrid hypothesis: Which distance did i run most frequently in 2024

Now I want to find out which distance I ran most frequently in 2024. I will categorize the runs into bins of 1-3 km, 4-6 km, and 7-11 km. I will use both a pichart and a bar plot to visualize my data.

In [28]:
```
# Define the bins for categorizing the distances
bins = [0, 3, 6, 11]  # These are the upper bounds of each range (1-3 km, 4-6 km
labels = ['1-3 km', '4-6 km', '7-11 km']  # names for the categories

# Categorize the runs in to the bins
df_runs['Distance_category'] = pd.cut(df_runs['Distance'], bins=bins, labels=lab

# Display the updated DataFrame with the new category column
print(df_runs[['Distance', 'Distance_category']])
```

	Distance	Distance_category
0	6.00	4-6 km
1	5.83	4-6 km
2	3.01	4-6 km
3	8.00	7-11 km
4	5.00	4-6 km
5	5.29	4-6 km
6	10.01	7-11 km
7	6.00	4-6 km
8	5.02	4-6 km
9	6.00	4-6 km
10	5.00	4-6 km
11	6.83	7-11 km
12	3.86	4-6 km
13	9.01	7-11 km
14	4.44	4-6 km
15	6.74	7-11 km
16	8.00	7-11 km
17	10.72	7-11 km
18	4.01	4-6 km
19	5.56	4-6 km
20	9.00	7-11 km
21	3.00	1-3 km
22	3.77	4-6 km
23	5.16	4-6 km
24	4.50	4-6 km
25	5.00	4-6 km
27	3.00	1-3 km
28	7.14	7-11 km
29	4.80	4-6 km
30	3.85	4-6 km
31	3.00	1-3 km
32	3.84	4-6 km
33	4.21	4-6 km
34	5.00	4-6 km
35	10.00	7-11 km
36	5.99	4-6 km
37	1.82	1-3 km
38	7.00	7-11 km
39	5.01	4-6 km
40	7.01	7-11 km
42	5.00	4-6 km
43	9.40	7-11 km
44	4.53	4-6 km
45	5.01	4-6 km
46	10.01	7-11 km
47	5.00	4-6 km
48	7.01	7-11 km
49	6.01	7-11 km
50	6.00	4-6 km
51	6.00	4-6 km

In [29]:
```
# Count the number of runs in each distance category
category_counts = df_runs['Distance_category'].value_counts()

# Print the counts to verify
print(category_counts)
```

```
Distance_category
4-6 km     30
7-11 km    16
1-3 km      4
Name: count, dtype: int64
```

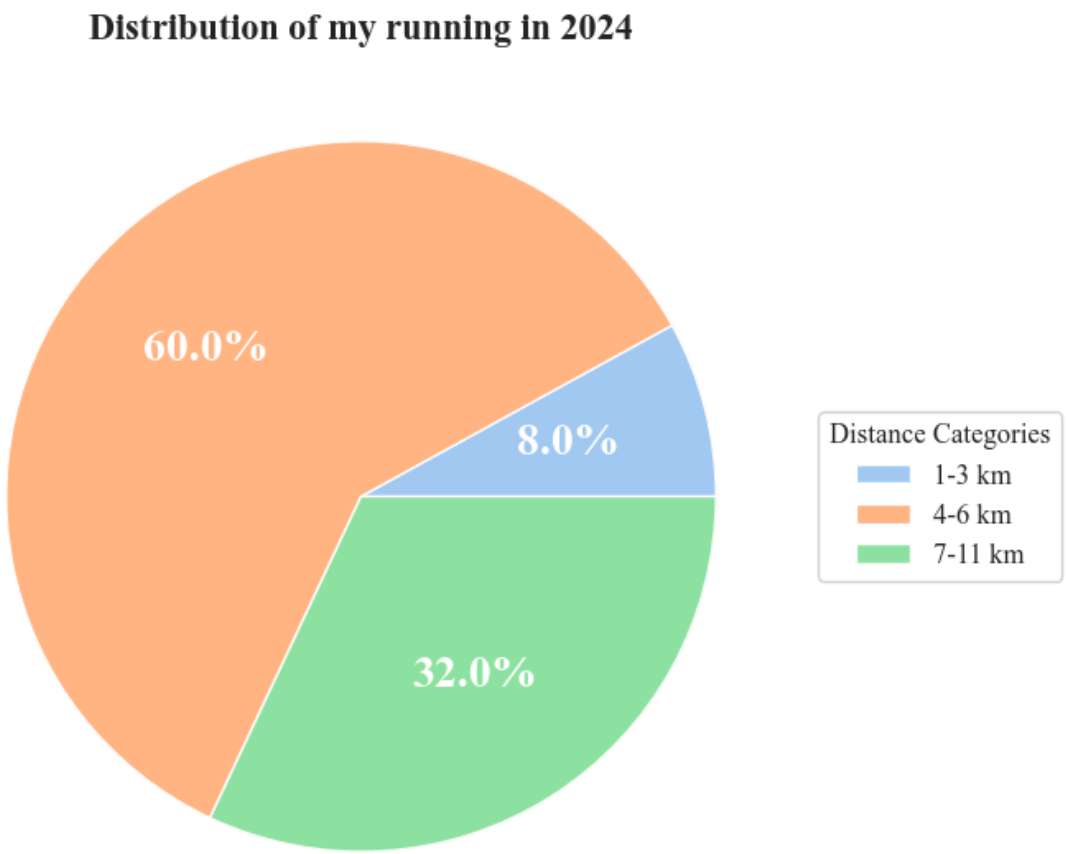
Pie chart

The pie chart clearly shows that the majority of my runs are in the 4-6 km range. In fact, 60% of all my runs in 2024 fall into this category. The next most common distance is 7-11 km, making up 32% of the chart. The shortest runs, 1-3 km, are the least frequent, accounting for just 8% of my 2024 runs.

The pie chart provides a visually intuitive representation of how each running distance contributes to my overall runs. I added a colormap in the legend to make it easy to distinguish between categories. This makes sure there isn't excessive text around the chart, which could otherwise be distracting or confusing.

For aesthetic reasons, I use white text inside the pie chart to display the percentages. I think this is the more aesthetically pleasing color, but to make sure the percentages are clearly visible, I have made their font size quite big. The colors chosen in the figure are colorblind friendly.

In [31]:
```
#The data
category_counts = pd.Series([4, 30, 16], index=["1-3 km", "4-6 km", "7-11 km"])

# Create figure
plt.figure(figsize=(7, 7))

# Plot pie chart without category labels surrounding it
wedges, _, autotexts = plt.pie(
    category_counts,
    autopct='%1.1f%%', #display prosentages
    colors=sns.color_palette("pastel", n_colors=len(category_counts)),
    textprops={"fontname": "Times New Roman", "fontweight": "bold", "fontsize":
    wedgeprops={"edgecolor": "white"}  # White line between the categories
)

#colormap outside the pie chart
plt.legend(wedges, category_counts.index, title="Distance Categories",
           loc="center left", bbox_to_anchor=(1, 0.5), #Placment of legend
           prop={"family": "Times new roman", "size": 12},
          title_fontproperties={"family": "Times new roman", "size": 12})

plt.title("Distribution of my running in 2024", fontname="Times new roman", font

plt.show()
```



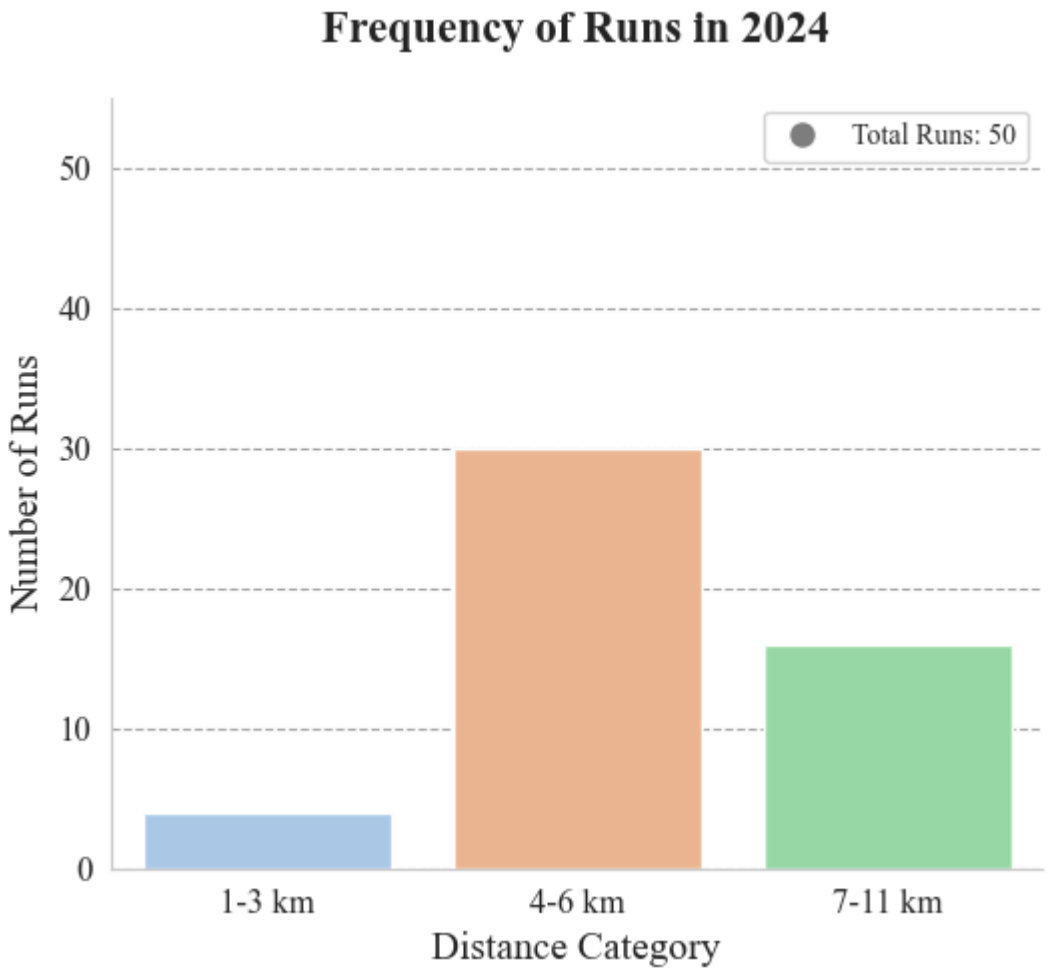
Barplot

The pie chart shows us the percentage, however, I want to see the specific number of runs, and I will do this through a bar plot. The bar plot shows that out of all my 50 runs in 2024, 30 of them were 4-6 km distances, 16 were 7-11 km, and 4 were 1-3 km.

The bar plot contains a legend giving the total number of runs, making it easier to understand just how much 30 runs are compared to the total number of runs. While each bar has its corresponding category of distance. I have used the same color as the piechart categories. I have the Y-limit a bit larger than my maximum value of 50, this is to make some space at the top, giving it a cleaner look.

In [33]:
```
# Set figure size
plt.figure(figsize=(6, 5))

# Barplot
sns.barplot(
    x=category_counts.index,
    y=category_counts.values,
    hue=category_counts.index,
    legend=False,
    palette=sns.color_palette("pastel", n_colors=len(category_counts))
)

# Calculate total runs
total_runs = category_counts.sum()
```

```
# Add the total runs as a label in the legend
handles, labels = plt.gca().get_legend_handles_labels()

# modify legend "Total Runs"
handles.append(plt.Line2D([0], [0], marker='o', color='w', label=f"Total Runs: {
                          markerfacecolor='gray', markersize=10))
labels.append(f"Total Runs: {total_runs}")

# Customize the legend to only display total runs
plt.legend(handles=handles, labels=labels, loc="upper right", title="",
           fontsize=12, title_fontsize=14, prop={'family': 'Times New Roman'})

# Adding labels and title
plt.xlabel("Distance Category", fontname="Times New Roman", fontsize= 14)
plt.ylabel("Number of Runs", fontname="Times New Roman", fontsize= 14)
plt.title("Frequency of Runs in 2024", fontname="Times New Roman", fontsize=16,

# Customizing tick labels and grid
plt.xticks(fontname="Times New Roman", fontsize= 12)
plt.yticks(fontname="Times New Roman", fontsize= 12)
plt.grid(axis='y', linestyle="--", alpha=0.7, color="gray")
plt.ylim(0, 55)
sns.despine()

# Show the plot
plt.show()
```



Third hypothesis: Which month of the year did i run the most in 2024

In [35]:
```
# previosuly i wanted the hour form activity data variable, now i want to extrac
df['Activity Date'] = pd.to_datetime(df['Activity Date'])

# Extract abbreviated month names
df['Month'] = df['Activity Date'].dt.strftime('%b')  # 'Jan', 'Feb', etc.

# Count runs per month
monthly_counts = df['Month'].value_counts()

# Define month order in abbreviations
month_order = ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun',
               'Jul', 'Aug', 'Sep', 'Oct', 'Nov', 'Dec']
monthly_counts = monthly_counts.reindex(month_order, fill_value=0)
```

Scatterplot

The scatter plot shows that in 2024, May had the highest number of runs. I chose a bar plot because it effectively presents all the months and their corresponding run counts without looking cluttered. It maintains a clean design by forming a simple shape with lines connecting each dot, making it intuitive to understand the diffrences between each month. I chose to have the abberations of the months because it makes the plot less clutterd.

In [41]:
```
# Set figure size
plt.figure(figsize=(12, 6))

# Create scatter plot
sns.scatterplot(
    x=monthly_counts.index,
    y=monthly_counts.values,
    s=150, #size of dots
    color="cornflowerblue",
    edgecolor="cornflowerblue",
    linewidth=1.5,
    zorder=3
)

# Connect points with a smooth line
plt.plot(
    monthly_counts.index,
    monthly_counts.values,
    linestyle="-",
    linewidth=1.5,
    color="gray",
    alpha=0.6, #detemrines trancparency
    zorder=2 #determines whether the items are placed in front or behind eachoth
)

# Labels and title with Times New Roman font
plt.xlabel("", fontsize=14, fontweight="bold", color="black", fontname="Times Ne
```

```
plt.ylabel("Number of Runs", fontsize=14, fontweight="bold", color="black", font
plt.ylim(-1, 15) #adds some space on the y axis so that the datapoints don't ali
plt.title("My Running through 2024", fontsize=16, fontweight="bold", color="blac

# Adjust x and y ticks
plt.xticks(rotation=0, fontsize=12, color="black", fontname="Times New Roman")
plt.yticks(fontsize=12, color="black", fontname="Times New Roman")

#background lines
plt.grid(axis="y", linestyle="--", alpha=0.5)

sns.despine()

# Show the plot
plt.show()
```