

```
In [6]: import pandas as pd
```

```
In [5]: import seaborn as sns
```

```
In [ ]: # Load The Data
```

```
In [7]: df = pd.read_csv("A:/Python/Extra source/TWO_CENTURIES_OF_UM_RACES.csv", low_memory=False)
```

```
In [8]: df.head(10)
```

Out[8]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Athlete country	Athlete age
0	2018	06.01.2018	Selva Costera (CHI)	50km	22	4:51:39 h	Tnfr	CHI	19
1	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:15:45 h	Roberto Echeverría	CHI	19
2	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:16:44 h	Puro Trail Osorno	CHI	19
3	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:34:13 h	Columbia	ARG	19
4	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:54:14 h	Baguales Trail	CHI	19
5	2018	06.01.2018	Selva Costera (CHI)	50km	22	6:25:01 h	NaN	ARG	19
6	2018	06.01.2018	Selva Costera (CHI)	50km	22	6:28:00 h	Los Patagones	ARG	19
7	2018	06.01.2018	Selva Costera (CHI)	50km	22	6:32:24 h	Reactiva Chile	CHI	19
8	2018	06.01.2018	Selva Costera (CHI)	50km	22	6:39:08 h	Puro Trail Osorno	CHI	19
9	2018	06.01.2018	Selva Costera (CHI)	50km	22	6:45:11 h	Marlene Flores Team	CHI	19

```
In [ ]: # Complete Data Summary
```

In [9]: df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7461195 entries, 0 to 7461194
Data columns (total 13 columns):
 #   Column                                  Dtype
---  -
 0   Year of event                          int64
 1   Event dates                           object
 2   Event name                            object
 3   Event distance/length                 object
 4   Event number of finishers            int64
 5   Athlete performance                   object
 6   Athlete club                         object
 7   Athlete country                      object
 8   Athlete year of birth                 float64
 9   Athlete gender                       object
10   Athlete age category                  object
11   Athlete average speed                 object
12   Athlete ID                           int64
dtypes: float64(1), int64(3), object(9)
memory usage: 740.0+ MB
```

In [ ]: *# Check Data Types*

In [10]: df.dtypes

```
Out[10]: Year of event                int64
Event dates                          object
Event name                          object
Event distance/length                object
Event number of finishers            int64
Athlete performance                  object
Athlete club                        object
Athlete country                     object
Athlete year of birth                float64
Athlete gender                      object
Athlete age category                 object
Athlete average speed                object
Athlete ID                          int64
dtype: object
```

In [11]: df.shape

Out[11]: (7461195, 13)

```
In [14]: df[df['Event distance/length'] == '50km']
```

```
Out[14]:
```

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Atl col
0	2018	06.01.2018	Selva Costera (CHI)	50km	22	4:51:39 h	Tnfr	
1	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:15:45 h	Roberto Echeverría	
2	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:16:44 h	Puro Trail Osorno	
3	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:34:13 h	Columbia	
4	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:54:14 h	Baguales Trail	
...	...	...	...	...	...	...	...	...
7461089	1995	07.01.1995	Centenary Lakes 50 Km Track Run (AUS)	50km	6	4:19:56 h	*QLD	
7461090	1995	07.01.1995	Centenary Lakes 50 Km Track Run (AUS)	50km	6	4:28:57 h	*QLD	
7461091	1995	07.01.1995	Centenary Lakes 50 Km Track Run (AUS)	50km	6	4:46:39 h	*QLD	
7461092	1995	07.01.1995	Centenary Lakes 50 Km Track Run (AUS)	50km	6	4:47:39 h	*QLD	
7461093	1995	07.01.1995	Centenary Lakes 50 Km Track Run (AUS)	50km	6	5:58:16 h	*QLD	

1522609 rows × 13 columns



```
In [15]: df[df['Event distance/length'] == '50mi']
```

Out[15]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Al co
55	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	9:53:05 h	*Middleville, MI	
56	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:09:35 h	*Waterloo, ON	
57	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:33:00 h	*Kitchener, ON	
58	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:38:17 h	*Utica, MI	
59	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:56:35 h	*Grass Lake, MI	
...	...	...	...	...	...	...	...	...
7461181	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	11:59:37 h	NaN	
7461182	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:01:41 h	NaN	
7461183	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	
7461184	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	
7461185	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:05:59 h	NaN	

352181 rows × 13 columns

In [16]: `df[df['Event distance/length'].isin(['50km', '50mi'])]`

Out[16]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Athle count
0	2018	06.01.2018	Selva Costera (CHI)	50km	22	4:51:39 h	Tnfr	C
1	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:15:45 h	Roberto Echeverría	C
2	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:16:44 h	Puro Trail Osorno	C
3	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:34:13 h	Columbia	AF
4	2018	06.01.2018	Selva Costera (CHI)	50km	22	5:54:14 h	Baguales Trail	C
...	...	...	...	...	...	...	...	...
7461181	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	11:59:37 h	NaN	US
7461182	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:01:41 h	NaN	US
7461183	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	US
7461184	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	US
7461185	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:05:59 h	NaN	US

1874790 rows × 13 columns



```
In [17]: df[(df['Event distance/length'].isin(['50km', '50mi'])) & (df['Year of event'] == 2020)]
```

Out[17]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club
2538571	2020	07.-09.02.2020	Taipei 48hr Ultra Marathon - 50mi (TPE)	50mi	38	7:34:19 h	日本隊
2538572	2020	07.-09.02.2020	Taipei 48hr Ultra Marathon - 50mi (TPE)	50mi	38	7:43:50 h	NaN
2538573	2020	07.-09.02.2020	Taipei 48hr Ultra Marathon - 50mi (TPE)	50mi	38	8:04:40 h	NaN
2538574	2020	07.-09.02.2020	Taipei 48hr Ultra Marathon - 50mi (TPE)	50mi	38	8:30:49 h	台灣大腳丫長 跑協會
2538575	2020	07.-09.02.2020	Taipei 48hr Ultra Marathon - 50mi (TPE)	50mi	38	8:34:47 h	NaN
...	...	...	...	...	...	...	..
2762404	2020	03.10.2020	Bison Ultra- Trail 50 (POL)	50km	271	7:36:25 h	AKS Polonia Warszawa
2762405	2020	03.10.2020	Bison Ultra- Trail 50 (POL)	50km	271	7:36:27 h	*Warszawa
2762406	2020	03.10.2020	Bison Ultra- Trail 50 (POL)	50km	271	7:44:18 h	Outdoo Training
2762407	2020	03.10.2020	Bison Ultra- Trail 50 (POL)	50km	271	8:04:50 h	PH Bysewc Gdańsk
2762408	2020	03.10.2020	Bison Ultra- Trail 50 (POL)	50km	271	8:11:43 h	*Nowe Aleksandrowc

63489 rows × 13 columns





```
In [19]: df[df['Event name'] == 'Everglades 50 Mile Ultra Run (USA)']['Event name'].  
         str.split('(').str.get(1).str.split(')').str.get(0)
```

```
Out[19]: 51923      USA  
         51924      USA  
         51925      USA  
         51926      USA  
         51927      USA  
         ...  
         6417091    USA  
         6417092    USA  
         6417093    USA  
         6417094    USA  
         6417095    USA  
         Name: Event name, Length: 338, dtype: object
```

```
In [21]: df[df['Event name'].str.split('(').str.get(1).str.split(')').str.get(0) ==  
          'USA']
```

Out[21]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Al co
55	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	9:53:05 h	*Middleville, MI	
56	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:09:35 h	*Waterloo, ON	
57	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:33:00 h	*Kitchener, ON	
58	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:38:17 h	*Utica, MI	
59	2018	06.01.2018	Yankee Springs 50 Mile Winter Challenge (USA)	50mi	9	11:56:35 h	*Grass Lake, MI	
...	...	...	...	...	...	...	...	...
7461181	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	11:59:37 h	NaN	
7461182	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:01:41 h	NaN	
7461183	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	
7461184	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:03:26 h	NaN	
7461185	1995	07.01.1995	Avalon Benefit 50-Mile Run (USA)	50mi	92	12:05:59 h	NaN	

1398540 rows × 13 columns

```
In [22]: df2 = df[(df['Event distance/length'].isin(['50km','50mi'])) & (df['Year o  
f event']==2020) & (df['Event name'].str.split('(').str.get(1).str.split  
(')).str.get(0) == 'USA')]
```

In [23]: `df2.head(10)`

Out[23]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	Atf cou
2539945	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	3:17:55 h	*Normandy Park, WA	
2539946	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:02:32 h	*Gold Bar, WA	
2539947	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:07:57 h	*Vashon, WA	
2539948	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:22:02 h	*Gig Harbor, WA	
2539949	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:27:34 h	*Bainbridge Island, WA	
2539950	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:42:06 h	*Seattle, WA	
2539951	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:49:20 h	*Camano Island, WA	
2539952	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:49:50 h	*Clinton, WA	

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club	At cou
2539953	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	4:51:00 h	*Seattle, WA	
2539954	2020	02.02.2020	West Seattle Beach Run - Winter Edition (USA)	50km	20	5:02:35 h	*Sammamish, WA	

In [24]: df2.shape

Out[24]: (26090, 13)

In [26]: df2['Event name'].str.split('(').str.get(0)

Out[26]: 2539945 West Seattle Beach Run - Winter Edition  
 2539946 West Seattle Beach Run - Winter Edition  
 2539947 West Seattle Beach Run - Winter Edition  
 2539948 West Seattle Beach Run - Winter Edition  
 2539949 West Seattle Beach Run - Winter Edition  
 ...  
 2760957 Yankee Springs Fall Trail Run Festival  
 2760958 Yankee Springs Fall Trail Run Festival  
 2760959 Yankee Springs Fall Trail Run Festival  
 2760960 Yankee Springs Fall Trail Run Festival  
 2760961 Yankee Springs Fall Trail Run Festival  
 Name: Event name, Length: 26090, dtype: object

In [28]: df2.loc[:, 'Event name'] = df2['Event name'].str.split('(').str.get(0)

In [29]: df2.head(5)

Out[29]:

	Year of event	Event dates	Event name	distance/length	Event number of finishers	Athlete performance	Athlete club	Athle count
2539945	2020	02.02.2020	West Seattle Beach Run - Winter Edition	50km	20	3:17:55 h	*Normandy Park, WA	US
2539946	2020	02.02.2020	West Seattle Beach Run - Winter Edition	50km	20	4:02:32 h	*Gold Bar, WA	US
2539947	2020	02.02.2020	West Seattle Beach Run - Winter Edition	50km	20	4:07:57 h	*Vashon, WA	US
2539948	2020	02.02.2020	West Seattle Beach Run - Winter Edition	50km	20	4:22:02 h	*Gig Harbor, WA	US
2539949	2020	02.02.2020	West Seattle Beach Run - Winter Edition	50km	20	4:27:34 h	*Bainbridge Island, WA	US



In [ ]: # clean up Athlete age

In [ ]: # Remove h from athlete performance

In [35]: df2['Athlete performance'].str.split(' ').str.get(0)

Out[35]:

2746390	10:40:56
2577434	7:49:40
2589048	3:27:48
2745828	8:46:37
2592144	12:59:01
...	
2746543	8:31:26
2749869	11:55:05
2755985	9:03:25
2755994	10:37:00
2755997	12:30:00

Name: Athlete performance, Length: 26090, dtype: object



In [39]: `df2.head(5)`

Out[39]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete club
<b>2746390</b>	2020	17.10.2020	Pumpkin Holler 50 Km Race	50km	105	10:40:56	*Stilwe O
<b>2577434</b>	2020	12.07.2020	Cremator Ultra 50 Mile Endurance Race	50mi	7	7:49:40	*Sout Carolin S
<b>2589048</b>	2020	14.03.2020	Peyton's Wild and Wacky 10x5Km Ultra	50km	75	3:27:48	*Sout Carolin S
<b>2745828</b>	2020	17.-18.10.2020	Cloudsplitter 100 - 50 Km Run	50km	66	8:46:37	*Niota, T
<b>2592144</b>	2020	14.03.2020	Antelope Canyon 50 Mile	50mi	266	12:59:01	*Antigoni N

In [ ]: *# drop columns : athlete club , athlete country , athlete year of birth , athlete age category*

In [40]: `df2 = df2.drop(['Athlete club', 'Athlete country', 'Athlete year of birth', 'Athlete age category'], axis = 1)`

In [41]: `df2.head(5)`

Out[41]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete gender	ε
<b>2746390</b>	2020	17.10.2020	Pumpkin Holler 50 Km Race	50km	105	10:40:56	F	
<b>2577434</b>	2020	12.07.2020	Cremator Ultra 50 Mile Endurance Race	50mi	7	7:49:40	M	
<b>2589048</b>	2020	14.03.2020	Peyton's Wild and Wacky 10x5Km Ultra	50km	75	3:27:48	M	
<b>2745828</b>	2020	17.-18.10.2020	Cloudsplitter 100 - 50 Km Run	50km	66	8:46:37	M	
<b>2592144</b>	2020	14.03.2020	Antelope Canyon 50 Mile	50mi	266	12:59:01	F	

In [ ]: `# clean up null values`

In [42]: `df2.isna().sum()`

Out[42]:

Year of event	0
Event dates	0
Event name	0
Event distance/length	0
Event number of finishers	0
Athlete performance	0
Athlete gender	0
Athlete average speed	0
Athlete ID	0
Athlete age	233
dtype:	int64

In [45]: `df2 = df2.dropna()`

```
In [46]: df2.isnull().sum()
```

```
Out[46]: Year of event          0  
Event dates                    0  
Event name                     0  
Event distance/length          0  
Event number of finishers      0  
Athlete performance            0  
Athlete gender                 0  
Athlete average speed          0  
Athlete ID                     0  
Athlete age                    0  
dtype: int64
```

```
In [ ]: # checking the Duplicates
```

```
In [47]: df2.duplicated().sum()
```

```
Out[47]: 0
```

```
In [ ]: # reset index
```

In [48]: `df2.reset_index(drop=True)`

Out[48]:

	Year of event	Event dates	Event name	Event distance/length	Event number of finishers	Athlete performance	Athlete gender	A av s
0	2020	17.10.2020	Pumpkin Holler 50 Km Race	50km	105	10:40:56	F	
1	2020	12.07.2020	Cremator Ultra 50 Mile Endurance Race	50mi	7	7:49:40	M	
2	2020	14.03.2020	Peyton's Wild and Wacky 10x5Km Ultra	50km	75	3:27:48	M	1
3	2020	17.-18.10.2020	Cloudsplitter 100 - 50 Km Run	50km	66	8:46:37	M	
4	2020	14.03.2020	Antelope Canyon 50 Mile	50mi	266	12:59:01	F	
...	...	...	...	...	...	...	...	
25852	2020	11.01.2020	Avalon Benefit 50 Km Run	50km	293	12:01:46	M	
25853	2020	08.-09.02.2020	Grandmaster Ultras 50 Km Run	50km	9	12:02:10	M	
25854	2020	14.03.2020	Everglades 50 km Ultra Run	50km	63	8:50:11	M	
25855	2020	08.02.2020	Rocky 50 Km Trail Run	50km	158	13:34:44	M	
25856	2020	25.01.2020	Angela Ivory 50K Memorial Run	50km	23	9:52:00	M	

25857 rows × 10 columns



In [ ]: `# fix Types`

```
In [49]: df2.dtypes
```

```
Out[49]: Year of event          int64
Event dates          object
Event name           object
Event distance/length object
Event number of finishers int64
Athlete performance  object
Athlete gender        object
Athlete average speed object
Athlete ID            int64
Athlete age           float64
dtype: object
```

```
In [51]: df2['Athlete age'] = df2['Athlete age'].astype(int,inplace = True)
```

```
-----
-
TypeError                                Traceback (most recent call last)
Cell In[51], line 1
----> 1 df2['Athlete age'] = df2['Athlete age'].astype(int , inplace = True)

TypeError: NDFrame.astype() got an unexpected keyword argument 'inplace'
```

```
In [52]: df2['Athlete age'].astype(int)
```

```
Out[52]: 2746390      9
2577434      12
2589048      12
2745828      12
2592144      13
..
2563553      81
2627971      81
2588736      82
2629941      82
2547499      85
Name: Athlete age, Length: 25857, dtype: int32
```

```
In [58]: df2['Athlete average speed'].astype(float)
```

```
Out[58]: 2746390      4.681
2577434     10.280
2589048     14.437
2745828      5.697
2592144      6.198
...
2563553      4.156
2627971      4.154
2588736      5.658
2629941      3.682
2547499      5.068
Name: Athlete average speed, Length: 25857, dtype: float64
```

```
In [ ]: df2['Athlete average speed'] = df2['Athlete average speed'].astype(float)
```

In [59]: df2.dtypes

```
Out[59]: Year of event          int64
Event dates                   object
Event name                    object
Event distance/length         object
Event number of finishers     int64
Athlete performance           object
Athlete gender                object
Athlete average speed         float64
Athlete ID                    int64
Athlete age                   int32
dtype: object
```

# Rename the columns Year of event Year of event Event name Event distance/length Event number of finishers  
Athlete performance Athlete gender Athlete average speed Athlete ID Athlete age

```
In [63]: df2 = df2.rename(columns = { 'Year of event': 'Year',
                                     'Year of event': 'Race_Day',
                                     'Event name ': 'Race_Name',
                                     'Event distance/length': 'Race_Length',
                                     'Event number of finishers ': 'Race_NO_of_Finisher'
                                     })
```

In [64]: df2.columns

```
Out[64]: Index(['Race_Day', 'Event dates', 'Event name', 'Race_Length',
                'Event number of finishers', 'Athlete performance', 'Athlete gender',
                'Athlete average speed', 'Athlete ID', 'Athlete age'],
              dtype='object')
```

In [65]: `df2.head()`

Out[65]:

	Race_Day	Event dates	Event name	Race_Length	Event number of finishers	Athlete performance	Athlete gender	
<b>2746390</b>	2020	17.10.2020	Pumpkin Holler 50 Km Race	50km	105	10:40:56	F	
<b>2577434</b>	2020	12.07.2020	Cremator Ultra 50 Mile Endurance Race	50mi	7	7:49:40	M	
<b>2589048</b>	2020	14.03.2020	Peyton's Wild and Wacky 10x5Km Ultra	50km	75	3:27:48	M	
<b>2745828</b>	2020	17.-18.10.2020	Cloudsplitter 100 - 50 Km Run	50km	66	8:46:37	M	
<b>2592144</b>	2020	14.03.2020	Antelope Canyon 50 Mile	50mi	266	12:59:01	F	

In [68]: `df2[df2['Athlete ID'] == 222509]`

Out[68]:

	Race_Day	Event dates	Event name	Race_Length	Event number of finishers	Athlete performance	Athlete gender	Athlete ID
<b>2616900</b>	2020	22.02.2020	Manasota Track Club 50K	50km	36	4:49:16	M	222509
<b>2591482</b>	2020	14.03.2020	Everglades 50 Mile Ultra Run	50mi	40	9:19:10	M	222509

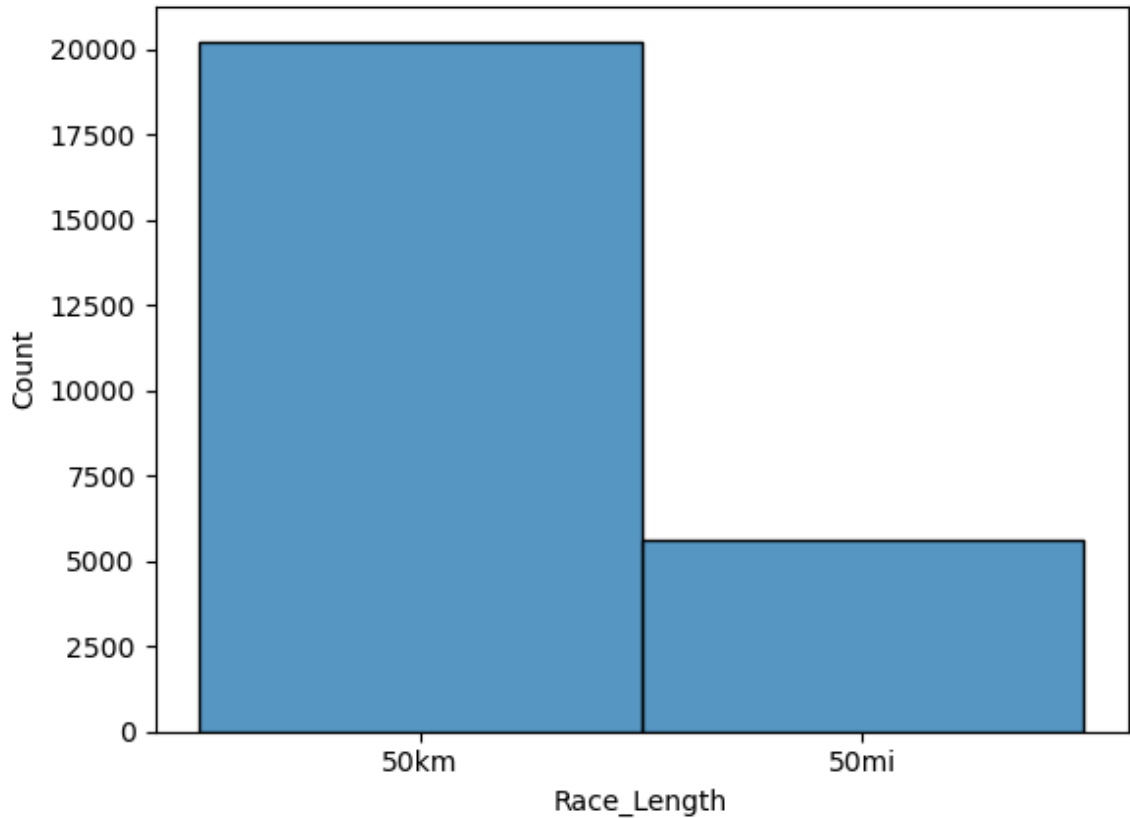
In [ ]: `# charts and graph`

In [72]: `import pandas as pd`

```
In [73]: sns.histplot(df2['Race_Length'])
```

```
C:\Anaconda\Lib\site-packages\seaborn\_oldcore.py:1119: FutureWarning: use
_inf_as_na option is deprecated and will be removed in a future version. C
onvert inf values to NaN before operating instead.
  with pd.option_context('mode.use_inf_as_na', True):
```

```
Out[73]: <Axes: xlabel='Race_Length', ylabel='Count'>
```





```
In [77]: sns.histplot(df2, x = df = df.replace([np.inf, -np.inf], np.nan), hue = 'Athlete gender')
```

C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use \_inf\_as\_na option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

```
with pd.option_context('mode.use_inf_as_na', True):
```

C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get\_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get\_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

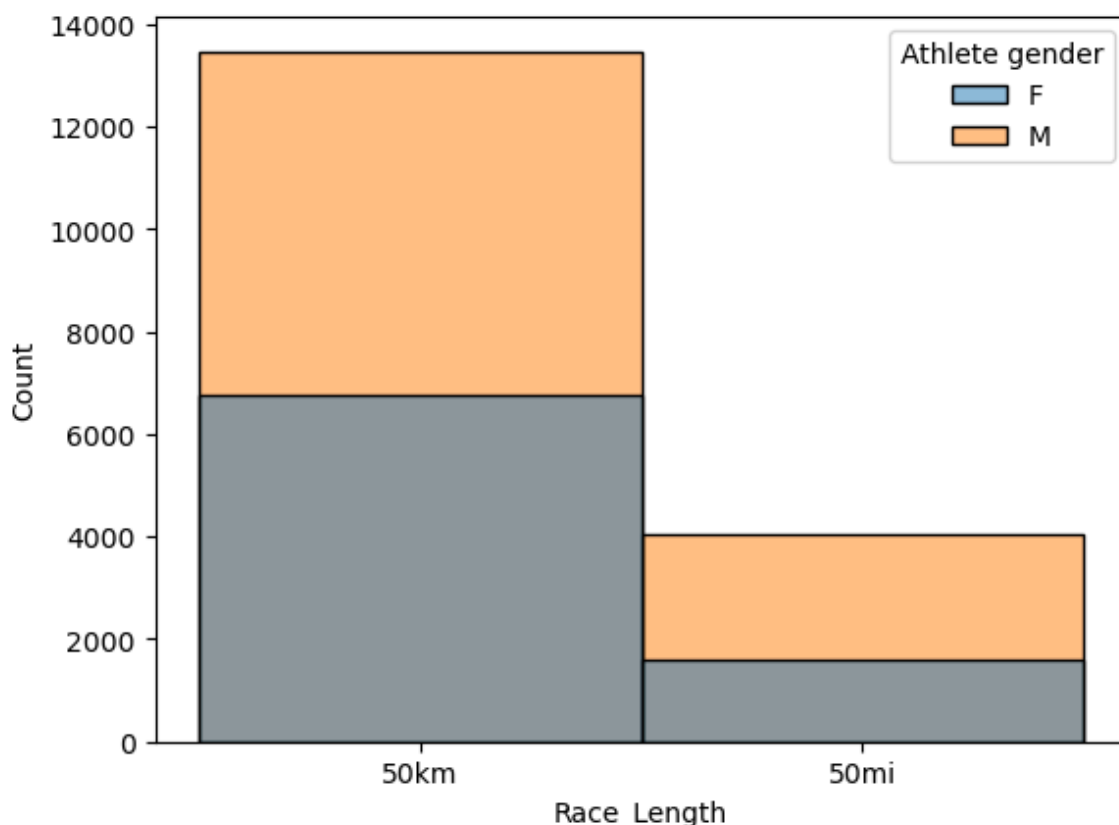
```
data_subset = grouped_data.get_group(pd_key)
```

C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get\_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```

C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1075: FutureWarning: When grouping with a length-1 list-like, you will need to pass a length-1 tuple to get\_group in a future version of pandas. Pass `(name,)` instead of `name` to silence this warning.

```
data_subset = grouped_data.get_group(pd_key)
```



```
In [78]: sns.displot(df2[df2['Race_Length'] == '50mi']['Athlete average speed'])
```

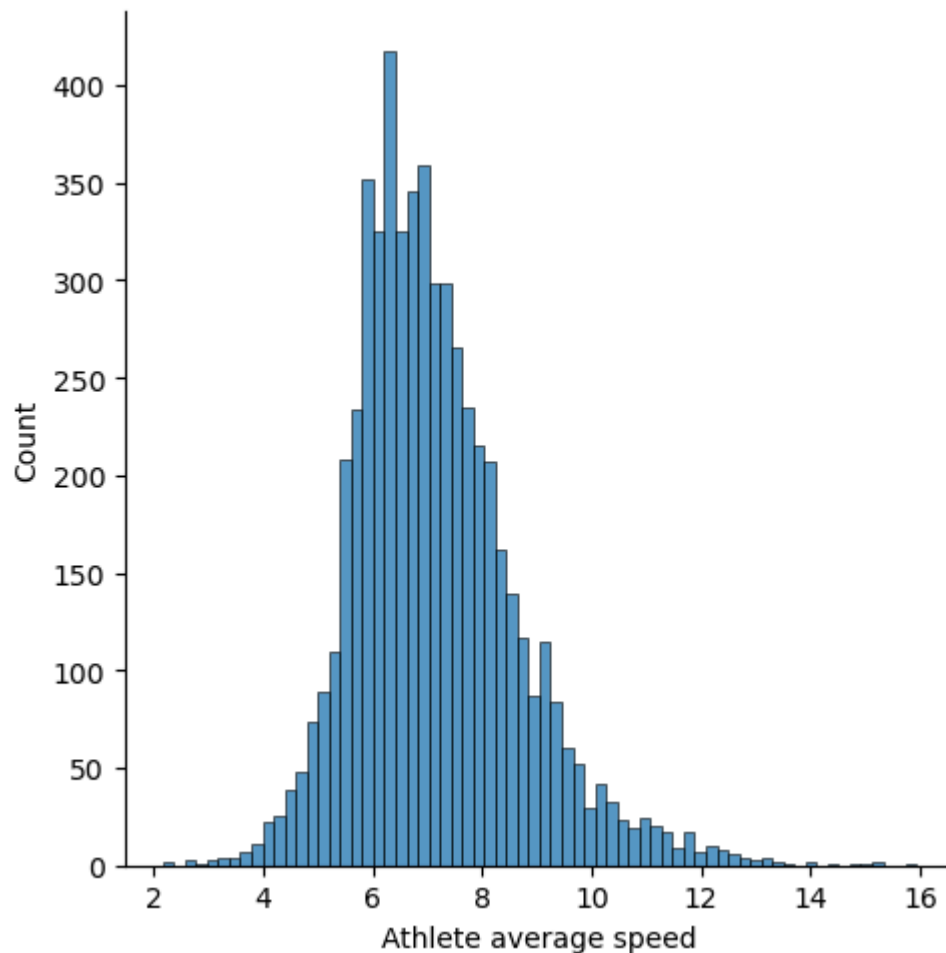
C:\Anaconda\Lib\site-packages\seaborn\\_oldcore.py:1119: FutureWarning: use `_inf_as_na` option is deprecated and will be removed in a future version. Convert inf values to NaN before operating instead.

with pd.option\_context('mode.use\_inf\_as\_na', True):

C:\Anaconda\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight

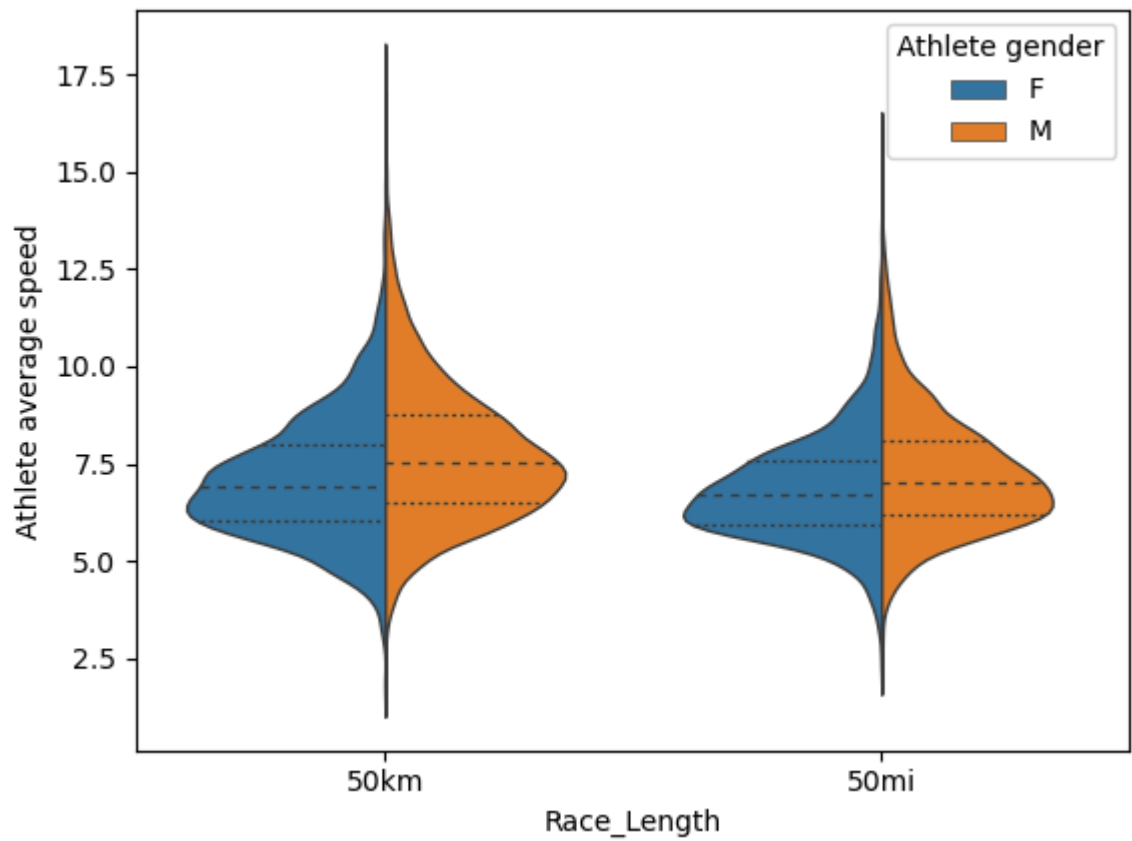
self.\_figure.tight\_layout(\*args, \*\*kwargs)

```
Out[78]: <seaborn.axisgrid.FacetGrid at 0x139a1254390>
```



```
In [85]: sns.violinplot(data=df2 , x='Race_Length', y = 'Athlete average speed', hue=
'Athlete gender', split = True , inner = 'quart', linewidth = 1 )
```

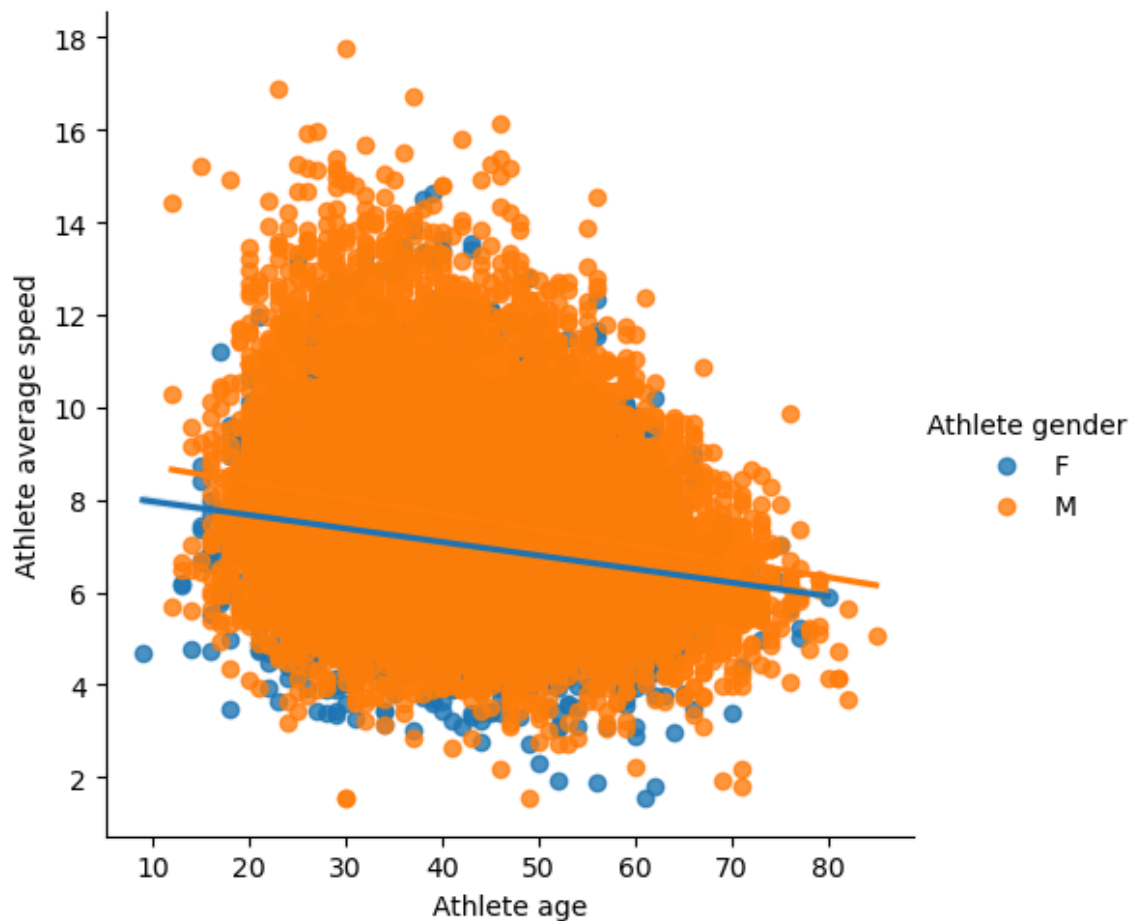
```
Out[85]: <Axes: xlabel='Race_Length', ylabel='Athlete average speed'>
```



```
In [87]: sns.lmplot(data=df2, x='Athlete age', y = 'Athlete average speed', hue = 'Athlete gender')
```

C:\Anaconda\Lib\site-packages\seaborn\axisgrid.py:118: UserWarning: The figure layout has changed to tight  
 self.\_figure.tight\_layout(\*args, \*\*kwargs)

```
Out[87]: <seaborn.axisgrid.FacetGrid at 0x1393b626790>
```



```
In [89]: df2.groupby(['Race_Length', 'Athlete gender'])['Athlete average speed'].mean()
```

```
Out[89]: Race_Length  Athlete gender
50km             F          7.083011
              M          7.738985
50mi             F          6.834371
              M          7.257633
Name: Athlete average speed, dtype: float64
```

In [90]: df2.head(10)

Out[90]:

	Race_Day	Event dates	Event name	Race_Length	Event number of finishers	Athlete performance	Athlete gender
<b>2746390</b>	2020	17.10.2020	Pumpkin Holler 50 Km Race	50km	105	10:40:56	F
<b>2577434</b>	2020	12.07.2020	Cremator Ultra 50 Mile Endurance Race	50mi	7	7:49:40	M
<b>2589048</b>	2020	14.03.2020	Peyton's Wild and Wacky 10x5Km Ultra	50km	75	3:27:48	M
<b>2745828</b>	2020	17.-18.10.2020	Cloudsplitter 100 - 50 Km Run	50km	66	8:46:37	M
<b>2592144</b>	2020	14.03.2020	Antelope Canyon 50 Mile	50mi	266	12:59:01	F
<b>2652597</b>	2020	19.09.2020	Sangre de Cristo Ultra 50 km Race	50km	68	7:30:00	M
<b>2623538</b>	2020	15.02.2020	Hagg Lake Trail Runs 50K	50km	82	8:08:12	F
<b>2678000</b>	2020	22.08.2020	Ghost Town Trail Challenge	50km	129	7:42:52	M
<b>2723047</b>	2020	08.11.2020	Lost Turkey Trail 50 Mile Ultra	50mi	17	11:25:40	M
<b>2634597</b>	2020	03.10.2020	Rock/Creek StumpJump 50K	50km	224	7:43:46	M

In [ ]: