

```
In [4]: import pandas as pd
```

```
In [5]: spotify_worldwide_daily_song_ranking = {'id': {0: 303651, 1: 85559, 2: 1046089, 3: 350824, 4: 776822, 5: 462814, 6: 792423, 7: 792223, 8: 793422, 9: 793622, 10: 793022}, 'position': {0: 52, 1: 160, 2: 175, 3: 25, 4: 1, 5: 2, 6: 2, 7: 2, 8: 1, 9: 1, 10: 1}, 'trackname': {0: 'Heart Won\'t Forget', 1: 'Someone In The Crowd - From "La La Land" Sound...', 2: 'The Greatest', 3: 'Unforgettable', 4: 'Bad and Boujee (feat. Lil Uzi Vert)', 5: 'DNA.', 6: 'DNA.', 7: 'HUMBLE.', 8: 'HUMBLE.', 9: 'HUMBLE.', 10: 'HUMBLE.'}, 'artist': {0: 'Matoma', 1: 'Emma Stone', 2: 'Sia', 3: 'French Montana', 4: 'Migos', 5: 'Kendrick Lamar', 6: 'Kendrick Lamar', 7: 'Kendrick Lamar', 8: 'Kendrick Lamar', 9: 'Kendrick Lamar', 10: 'Kendrick Lamar'}, 'streams': {0: 28047, 1: 17134, 2: 10060, 3: 46603, 4: 1823391, 5: 3013496, 6: 3643231, 7: 3144482, 8: 3172718, 9: 3394456}, 'url': {0: 'https://open.spotify.com/track/2of2DM5LqTh7ohm...', 1: 'https://open.spotify.com/track/7xE4vKvjUTtHyJ...', 2: 'https://open.spotify.com/track/7xHNBFm6ObGEQP...', 3: 'https://open.spotify.com/track/3B54sVLJ402zGa6...', 4: 'https://open.spotify.com/track/4Km5HrUvYTaSUfi...', 5: 'https://open.spotify.com/track/6HZILIRieu8S0iq...', 6: 'https://open.spotify.com/track/6HZILIRieu8S0iq...', 7: 'https://open.spotify.com/track/7KXjTSCq5nL1LoY...', 8: 'https://open.spotify.com/track/7KXjTSCq5nL1LoY...', 9: 'https://open.spotify.com/track/7KXjTSCq5nL1LoY...', 10: 'https://open.spotify.com/track/7KXjTSCq5nL1LoY...'}, 'date': {0: '2017-02-04 00:00:00', 1: '2017-02-26 00:00:00', 2: '2017-03-06 00:00:00', 3: '2017-10-01 00:00:00', 4: '2017-01-27 00:00:00', 5: '2017-04-15 00:00:00', 6: '2017-04-14 00:00:00', 7: '2017-04-20 00:00:00', 8: '2017-04-21 00:00:00', 9: '2017-04-18 00:00:00', 10: '2017-04-18 00:00:00'}, 'region': {0: 'no', 1: 'fr', 2: 'cl', 3: 'no', 4: 'us', 5: 'us', 6: 'us', 7: 'us', 8: 'us', 9: 'us', 10: 'us'}}
```

```
In [6]: df_spotify_worldwide_daily_song_ranking = pd.DataFrame(spotify_worldwide_daily_song_ranking)
df_spotify_worldwide_daily_song_ranking
```

	<b>id</b>	<b>position</b>	<b>trackname</b>	<b>artist</b>	<b>streams</b>	<b>url</b>	<b>date</b>	<b>region</b>
<b>0</b>	303651	52	Heart Won't Forget	Matoma	28047	https://open.spotify.com/track/2of2DM5LqTh7ohm...	2017-02-04 00:00:00	no
<b>1</b>	85559	160	Someone In The Crowd - From "La La Land" Sound...	Emma Stone	17134	https://open.spotify.com/track/7xE4vKvjUTtHyJ...	2017-02-26 00:00:00	fr
<b>2</b>	1046089	175	The Greatest	Sia	10060	https://open.spotify.com/track/7xHNBFm6ObGEQP...	2017-03-06 00:00:00	cl
<b>3</b>	350824	25	Unforgettable	French Montana	46603	https://open.spotify.com/track/3B54sVLJ402zGa6...	2017-10-01 00:00:00	no
<b>4</b>	776822	1	Bad and Boujee (feat. Lil Uzi Vert)	Migos	1823391	https://open.spotify.com/track/4Km5HrUvYTaSUfi...	2017-01-27 00:00:00	us
...	...	...	...	...	...	...	...	...
<b>95</b>	792423	2	DNA.	Kendrick Lamar	3013496	https://open.spotify.com/track/6HZILIRieu8S0iq...	2017-04-15 00:00:00	us
<b>96</b>	792223	2	DNA.	Kendrick Lamar	3643231	https://open.spotify.com/track/6HZILIRieu8S0iq...	2017-04-14 00:00:00	us
<b>97</b>	793422	1	HUMBLE.	Kendrick Lamar	3144482	https://open.spotify.com/track/7KXjTSCq5nL1LoY...	2017-04-20 00:00:00	us
<b>98</b>	793622	1	HUMBLE.	Kendrick Lamar	3172718	https://open.spotify.com/track/7KXjTSCq5nL1LoY...	2017-04-21 00:00:00	us
<b>99</b>	793022	1	HUMBLE.	Kendrick Lamar	3394456	https://open.spotify.com/track/7KXjTSCq5nL1LoY...	2017-04-18 00:00:00	us

100 rows × 8 columns

```
In [7]: df_No8_No10 = df_spotify_worldwide_daily_song_ranking[(df_spotify_worldwide_daily_song_ranking["position"] > 7) & (df_spotify_worldwide_daily_song_ranking["position"] < 10)]  
In [8]: df_No8_No10.sort_values(by = "position", ascending = True)
```

Out[8]:

	<b>id</b>	<b>position</b>	<b>trackname</b>	<b>artist</b>	<b>streams</b>	<b>url</b>	<b>date</b>	<b>region</b>
<b>79</b>	28007	8	Una Lady Como T\`{e}j	Manuel Turizo	14464	<a href="https://open.spotify.com/track/7MHN1aCFtLXjown...">https://open.spotify.com/track/7MHN1aCFtLXjown...</a>	2017-05-21 00:00:00	ec
<b>81</b>	306207	8	Alone	Alan Walker	92447	<a href="https://open.spotify.com/track/0JiVRyTJcJnmlwC...">https://open.spotify.com/track/0JiVRyTJcJnmlwC...</a>	2017-02-17 00:00:00	no
<b>82</b>	442607	8	Perfect Duet (Ed Sheeran & Beyonc\`{e} \u2122)	Ed Sheeran	90487	<a href="https://open.spotify.com/track/1bhUWB0zJMIKr9y...">https://open.spotify.com/track/1bhUWB0zJMIKr9y...</a>	2017-12-28 00:00:00	it
<b>84</b>	301207	8	Make Me (Cry)	Noah Cyrus	81932	<a href="https://open.spotify.com/track/2BrzlUj1u1Ctvaj...">https://open.spotify.com/track/2BrzlUj1u1Ctvaj...</a>	2017-01-23 00:00:00	no
<b>88</b>	496406	8	Attention	Charlie Puth	113556	<a href="https://open.spotify.com/track/4iLqG9SeJSnt0cS...">https://open.spotify.com/track/4iLqG9SeJSnt0cS...</a>	2017-06-25 00:00:00	ph
<b>75</b>	546207	9	Castle on the Hill	Ed Sheeran	15127	<a href="https://open.spotify.com/track/6PCUP3dWmTjcTtX...">https://open.spotify.com/track/6PCUP3dWmTjcTtX...</a>	2017-02-20 00:00:00	tw
<b>80</b>	54008	9	Una Lady Como T\`{e}j	Manuel Turizo	14966	<a href="https://open.spotify.com/track/7MHN1aCFtLXjown...">https://open.spotify.com/track/7MHN1aCFtLXjown...</a>	2017-10-01 00:00:00	ec
<b>83</b>	317408	9	Scared to Be Lonely	Martin Garrix	71259	<a href="https://open.spotify.com/track/3ebXMykcMXOcLeJ...">https://open.spotify.com/track/3ebXMykcMXOcLeJ...</a>	2017-04-14 00:00:00	no
<b>85</b>	119208	9	Mobali	Siboy	81790	<a href="https://open.spotify.com/track/6xcXAVbDuVT1pCs...">https://open.spotify.com/track/6xcXAVbDuVT1pCs...</a>	2017-08-17 00:00:00	fr
<b>87</b>	107808	9	Chocolat (feat. Awa)	Lartiste	93144	<a href="https://open.spotify.com/track/4fwtP5khM1iEoa6...">https://open.spotify.com/track/4fwtP5khM1iEoa6...</a>	2017-06-21 00:00:00	fr
<b>76</b>	145609	10	Bling Bling	Kaaris	97760	<a href="https://open.spotify.com/track/4s7OqcUEyLJ1Rkb...">https://open.spotify.com/track/4s7OqcUEyLJ1Rkb...</a>	2017-12-27 00:00:00	fr
<b>77</b>	869079	10	Ahora Dice	Chris Jeday	4543	<a href="https://open.spotify.com/track/22eADXu8DfOAUED...">https://open.spotify.com/track/22eADXu8DfOAUED...</a>	2017-08-30 00:00:00	sv
<b>78</b>	1044724	10	Vacaciones	Wisin	86922	<a href="https://open.spotify.com/track/3dQDid3IUNhZy1O...">https://open.spotify.com/track/3dQDid3IUNhZy1O...</a>	2017-02-28 00:00:00	cl
<b>86</b>	972724	10	You Don't Know Me - Radio Edit	Jax Jones	255434	<a href="https://open.spotify.com/track/00INx0OctJrS3MK...">https://open.spotify.com/track/00INx0OctJrS3MK...</a>	2017-03-11 00:00:00	de
<b>89</b>	413409	10	Ei Party	Jake La Furia	100715	<a href="https://open.spotify.com/track/5xyExld5XWZplBv...">https://open.spotify.com/track/5xyExld5XWZplBv...</a>	2017-08-04 00:00:00	it

```
In [9]: df_No8_No10[["trackname","position"]]
```

```
Out[9]:
```

	trackname	position
75	Castle on the Hill	9
76	Bling Bling	10
77	Ahora Dice	10
78	Vacaciones	10
79	Una Lady Como T/S	8
80	Una Lady Como T/S	9
81	Alone	8
82	Perfect Duet (Ed Sheeran & Beyonc/C)	8
83	Scared to Be Lonely	9
84	Make Me (Cry)	8
85	Mobali	9
86	You Don't Know Me - Radio Edit	10
87	Chocolat (feat. Awa)	9
88	Attention	8
89	El Party	10

```
In [39]: list3 = df_No8_No10[["trackname","position"]]
list4 = list3.sort_values(by = "position", ascending = True)
```

```
In [40]: import numpy as np
```

```
In [43]: list5 = list4.to_numpy().tolist()
```

```
In [45]: for i in list5:
    for j in i:
```

```
-----  
TypeError                                     Traceback (most recent call last)  
Input In [45], in <cell line: 1>()  
      1 for i in list5:  
      2     for j in i:  
----> 3         print((list5[i[0]]) + "|" + (list5[i[1]]))  
  
TypeError: list indices must be integers or slices, not str
```

## question 2

```
In [10]: employee = {'id': {0: 5, 1: 13, 2: 11, 3: 10, 4: 19, 5: 18, 6: 20, 7: 21, 8: 22, 9: 23, 10: 25, 11: 26, 12: 27, 13: 28}}  
In [11]: df_employee = pd.DataFrame(employee)  
In [12]: df_employee
```

Out[12]:		<b>id</b>	<b>first_name</b>	<b>last_name</b>	<b>age</b>	<b>sex</b>	<b>employee_title</b>	<b>department</b>	<b>salary</b>	<b>target</b>	<b>bonus</b>	<b>email</b>	<b>city</b>	<b>address</b>	
		0	5	Max	George	26	M	Sales	Sales	1300	200	150	Max@company.com	California	2638 Richards Avenue
		1	13	Katty	Bond	56	F	Manager	Management	150000	0	300	Katty@company.com	Arizona	None
		2	11	Richerd	Gear	57	M	Manager	Management	250000	0	300	Richerd@company.com	Alabama	None
		3	10	Jennifer	Dion	34	F	Sales	Sales	1000	200	150	Jennifer@company.com	Alabama	None
		4	19	George	Joe	50	M	Manager	Management	100000	0	300	George@company.com	Florida	1003 Wyatt Street
		5	18	Laila	Mark	26	F	Sales	Sales	1000	200	150	Laila@company.com	Florida	3655 Spirit Drive
		6	20	Sarrah	Bicky	31	F	Senior Sales	Sales	2000	200	150	Sarrah@company.com	Florida	1176 Tyler Avenue
		7	21	Suzan	Lee	34	F	Sales	Sales	1300	200	150	Suzan@company.com	Florida	1275 Monroe Avenue
		8	22	Mandy	John	31	F	Sales	Sales	1300	200	150	Mandy@company.com	Florida	2510 Maryland Avenue
		9	23	Britney	Berry	45	F	Sales	Sales	1200	200	100	Britney@company.com	Florida	3946 Steve Hunt Road
		10	25	Jack	Mick	29	M	Sales	Sales	1300	200	100	Jack@company.com	Hawaii	3762 Stratford Drive
		11	26	Ben	Ten	43	M	Sales	Sales	1300	150	100	Ben@company.com	Hawaii	3055 Indiana Avenue
		12	27	Tom	Fridy	32	M	Sales	Sales	1200	200	150	Tom@company.com	Hawaii	801 Stratford

	<b>id</b>	<b>first_name</b>	<b>last_name</b>	<b>age</b>	<b>sex</b>	<b>employee_title</b>	<b>department</b>	<b>salary</b>	<b>target</b>	<b>bonus</b>	<b>email</b>	<b>city</b>	<b>address</b>
													Drive
													3533
<b>13</b>	29	Antoney	Adam	34	M	Sales	Sales	1300	180	150	Antoney@company.com	Hawaii	Randall Drive
													2641
<b>14</b>	28	Morgan	Matt	25	M	Sales	Sales	1200	200	150	Morgan@company.com	Hawaii	Randall Drive
													3632
<b>15</b>	6	Molly	Sam	28	F	Sales	Sales	1400	100	150	Molly@company.com	Arizona	Polk Street
													3461
<b>16</b>	7	Nicky	Bat	33	F	Sales	Sales	1400	400	100	Molly@company.com	Arizona	Preston Street
													None
<b>17</b>	9	Monika	William	33	F	Sales	Sales	1000	200	100	Molly@company.com	Alabama	None
													None
<b>18</b>	17	Mick	Berry	44	M	Senior Sales	Sales	2200	200	150	Mick@company.com	Florida	None
													None
<b>19</b>	12	Shandler	Bing	23	M	Auditor	Audit	1100	200	150	Shandler@company.com	Arizona	None
													None
<b>20</b>	14	Jason	Tom	23	M	Auditor	Audit	1000	200	150	Jason@company.com	Arizona	None
													None
<b>21</b>	16	Celine	Anston	27	F	Auditor	Audit	1000	200	150	Celine@company.com	Colorado	None
													None
<b>22</b>	15	Michale	Jackson	44	F	Auditor	Audit	700	150	150	Michale@company.com	Colorado	None
													4541
<b>23</b>	24	Adam	Morris	30	M	Sales	Sales	1300	200	100	Adam@company.com	Alabama	Ferry Street
													2522
<b>24</b>	30	Mark	Jon	28	M	Sales	Sales	1200	200	150	Mark@company.com	Alabama	George Avenue
													4832
<b>25</b>	8	John	Ford	26	M	Senior Sales	Sales	1500	140	100	Molly@company.com	Alabama	New Creek Road
													1069
<b>26</b>	1	Allen	Wang	55	F	Manager	Management	200000	0	300	Allen@company.com	California	Ventura Drive

<b>id</b>	<b>first_name</b>	<b>last_name</b>	<b>age</b>	<b>sex</b>	<b>employee_title</b>	<b>department</b>	<b>salary</b>	<b>target</b>	<b>bonus</b>	<b>email</b>	<b>city</b>	<b>address</b>	
27	2	Joe	Jack	32	M	Sales	Sales	1000	200	150	Joe@company.com	California	995 Jim Rosa Lane
28	3	Henry	Ted	31	M	Senior Sales	Sales	2000	200	150	Henry@company.com	California	1609 Ford Street
29	4	Sam	Mark	25	M	Sales	Sales	1000	120	150	Sam@company.com	California	4869 Libby Street

```
In [13]: notRef = df_employee[df_employee["manager_id"] != 1]
```

```
In [28]: lsit1 = notRef["first_name"].tolist()

for i in list1:
    print(i, end = "\n")
```

162435

## Question3

```
In [15]: facebook_reactions = {'poster': {0: 2, 1: 2, 2: 1, 3: 1, 4: 1, 5: 1, 6: 1, 7: 2, 8: 2, 9: 1, 10: 1, 11: 1, 12: 1, 13:
```

```
In [16]: df_facebook_reactions = pd.DataFrame(facebook_reactions)
df_facebook_reactions
```

Out[16]:

	poster	friend	reaction	date_day	post_id
0	2	1	like	1	0
1	2	6	like	1	0
2	1	2	like	1	1
3	1	3	heart	1	1
4	1	4	like	1	1
5	1	5	heart	1	1
6	1	6	like	1	1
7	2	1	like	2	2
8	2	6	like	2	2
9	1	2	like	2	3
10	1	3	like	2	3
11	1	4	like	2	3
12	1	5	like	2	3
13	1	6	like	2	3
14	2	1	laugh	1	0
15	2	6	laugh	1	1
16	1	2	laugh	1	0
17	1	3	laugh	1	3
18	1	4	laugh	1	4

In [17]: `like = df_facebook_reactions[df_facebook_reactions["reaction"] == "like"]`

In [18]: `import numpy as np`

In [19]: `unique_friends = like["friend"].unique()`

```
In [24]: list1 = unique_friends.tolist()
```

```
In [25]: for i in list1:  
    print(i)
```

```
1  
6  
2  
4  
3  
5
```

## Question 4

```
In [48]: twitch = {'user_id': {0: 0, 1: 2, 2: 3, 3: 1, 4: 2, 5: 0, 6: 0, 7: 3, 8: 1, 9: 2}, 'session_start': {0: '2020-08-11 05:51:31', 1: '2020-07-11 03:36:54', 2: '2020-11-26 11:41:47', 3: '2020-11-19 06:24:24', 4: '2020-11-14 03:36:05', 5: '2020-03-11 03:01:40', 6: '2020-08-11 03:50:45', 7: '2020-10-11 22:15:14', 8: '2020-11-20 06:59:57', 9: '2020-07-11 14:32:19'}, 'session_end': {0: '2020-08-11 05:54:45', 1: '2020-07-11 03:37:08', 2: '2020-11-26 11:52:01', 3: '2020-11-19 07:24:38', 4: '2020-11-14 03:39:19', 5: '2020-03-11 03:01:59', 6: '2020-08-11 03:55:59', 7: '2020-10-11 22:18:28', 8: '2020-11-20 07:20:11', 9: '2020-07-11 14:42:33'}, 'session_id': {0: 539, 1: 840, 2: 848, 3: 515, 4: 646, 5: 782, 6: 815, 7: 630, 8: 907, 9: 949}, 'session_type': {0: 'streamer', 1: 'streamer', 2: 'streamer', 3: 'viewer', 4: 'viewer', 5: 'streamer', 6: 'viewer', 7: 'viewer', 8: 'streamer', 9: 'viewer'}}
```

```
In [119... df_twitch = pd.DataFrame(twitch)
```

```
In [120... df_twitch
```

```
Out[120]:
```

	user_id	session_start	session_end	session_id	session_type
0	0	2020-08-11 05:51:31	2020-08-11 05:54:45	539	streamer
1	2	2020-07-11 03:36:54	2020-07-11 03:37:08	840	streamer
2	3	2020-11-26 11:41:47	2020-11-26 11:52:01	848	streamer
3	1	2020-11-19 06:24:24	2020-11-19 07:24:38	515	viewer
4	2	2020-11-14 03:36:05	2020-11-14 03:39:19	646	viewer
5	0	2020-03-11 03:01:40	2020-03-11 03:01:59	782	streamer
6	0	2020-08-11 03:50:45	2020-08-11 03:55:59	815	viewer
7	3	2020-10-11 22:15:14	2020-10-11 22:18:28	630	viewer
8	1	2020-11-20 06:59:57	2020-11-20 07:20:11	907	streamer
9	2	2020-07-11 14:32:19	2020-07-11 14:42:33	949	viewer

```
In [121]: df_twitch["is_streamer"] = np.where(df_twitch["session_type"] == "streamer", 1, 0)
```

```
In [122]: df_twitch["is_viewer"] = np.where(df_twitch["session_type"] == "viewer", 1, 0)
```

```
In [123]: df_twitch[["user_id","is_streamer","is_viewer"]]
```

Out[123]:	user_id	is_streamer	is_viewer
<b>0</b>	0	1	0
<b>1</b>	2	1	0
<b>2</b>	3	1	0
<b>3</b>	1	0	1
<b>4</b>	2	0	1
<b>5</b>	0	1	0
<b>6</b>	0	0	1
<b>7</b>	3	0	1
<b>8</b>	1	1	0
<b>9</b>	2	0	1

```
In [129]: streamer = df.twitch.groupby("is streamer")["user id"].apply(list).reset_index().iloc[1,1]
```

```
In [130]: viewer = df.twitch.groupby("is viewer")["user id"].apply(list).reset_index().iloc[1,1]
```

```
In [131]: set(streamer).intersection(set(viewer))
```

```
Out[131]: {0, 1, 2, 3}
```

## Question 5

```
In [65]: airbnb_apartments = {'host_id': {0: 0, 1: 1, 2: 2, 3: 3, 4: 4, 5: 5, 6: 6, 7: 7, 8: 8, 9: 9, 10: 10, 11: 11, 12: 0, ...
```

```
In [66]: airbnb_hosts = {'host_id': {0: 0, 1: 0, 2: 0, 3: 1, 4: 1, 5: 2, 6: 3, 7: 3, 8: 4, 9: 5, 10: 5, 11: 6, 12: 7, 13: 8, ...}}
```

```
In [67]: df_airbnb_apartments = pd.DataFrame(airbnb_apartments)
        df_airbnb_hosts = pd.DataFrame(airbnb_hosts)
```

```
In [68]: df_airbnb_apartments
```

Out[68]:	host_id	nationality	gender	age
0	0	USA	M	28
1	1	USA	F	29
2	2	China	F	31
3	3	China	M	24
4	4	Mali	M	30
...	...	...	...	...
171	7	Luxembourg	F	25
172	6	Luxembourg	M	25
173	7	Luxembourg	F	25
174	6	Luxembourg	M	25

176 rows × 4 columns

```
In [69]: df.airbnb.hosts
```

Out[69]:

	host_id	apartment_id	apartment_type	n_beds	n_bedrooms	country	city
0	0	A1	Room	1	1	USA	New York
1	0	A2	Room	1	1	USA	New Jersey
2	0	A3	Room	1	1	USA	New Jersey
3	1	A4	Apartment	2	1	USA	Houston
4	1	A5	Apartment	2	1	USA	Las Vegas
5	2	A6	Yurt	3	1	Mongolia	-
6	3	A7	Penthouse	3	3	China	Tianjin
7	3	A8	Penthouse	5	5	China	Beijing
8	4	A9	Apartment	2	1	Mali	Bamako
9	5	A10	Room	3	1	Mali	Segou
10	5	A11	Room	2	1	Mali	Segou
11	6	A12	Penthouse	6	6	Luxembourg	Luxembourg
12	7	A13	Room	4	1	Luxembourg	Luxembourg
13	8	A14	Apartment	2	1	Australia	Perth
14	9	A15	Apartment	2	1	Australia	Perth
15	9	A16	Apartment	2	1	Australia	Perth
16	10	A17	Room	4	1	Brazil	Rio De Janeiro
17	10	A18	Room	4	1	Argentina	Mendoza
18	10	A19	Room	4	2	Uruguay	Mercedes
19	10	A20	Room	4	2	Brazil	Brasilia
20	11	A21	Apartment	2	2	Mexico	Mexico City

In [133...]: airbnb = pd.merge(df\_airbnb\_apartments, df\_airbnb\_hosts, on = "host\_id")

In [134...]: airbnb.groupby("nationality")["n\_bedrooms"].sum().reset\_index().sort

```
Out[134]: nationality
Australia      12
Brazil         32
China          36
Luxembourg    224
Mali            12
USA            232
Name: n_bedrooms, dtype: int64
```

```
In [ ]:
```

```
In [ ]:
```

```
In [ ]:
```

## pandas inclass

```
In [77]: player_results = {'player_id': {0: 401, 1: 401, 2: 401, 3: 401, 4: 401, 5: 401, 6: 401, 7: 401, 8: 401, 9: 401, 10: 401, 11: 401, 12: 401, 13: 401, 14: 401, 15: 401, 16: 401, 17: 401, 18: 401, 19: 401, 20: 401, 21: 401, 22: 401, 23: 401, 24: 401, 25: 401, 26: 401, 27: 401, 28: 401, 29: 401, 30: 401, 31: 401, 32: 401, 33: 401, 34: 401, 35: 401, 36: 401, 37: 401, 38: 401, 39: 401, 40: 401, 41: 401, 42: 401, 43: 401, 44: 401, 45: 401, 46: 401, 47: 401, 48: 401, 49: 401, 50: 401, 51: 401, 52: 401, 53: 401, 54: 401, 55: 401, 56: 401, 57: 401, 58: 401, 59: 401, 60: 401, 61: 401, 62: 401, 63: 401, 64: 401, 65: 401, 66: 401, 67: 401, 68: 401, 69: 401, 70: 401, 71: 401, 72: 401, 73: 401, 74: 401, 75: 401, 76: 401, 77: 401, 78: 401, 79: 401, 80: 401, 81: 401, 82: 401, 83: 401, 84: 401, 85: 401, 86: 401, 87: 401, 88: 401, 89: 401, 90: 401, 91: 401, 92: 401, 93: 401, 94: 401, 95: 401, 96: 401, 97: 401, 98: 401, 99: 401}, 'stage': {0: 'Stage 1', 1: 'Stage 2', 2: 'Stage 3', 3: 'Stage 4', 4: 'Stage 5', 5: 'Stage 6', 6: 'Stage 7', 7: 'Stage 8', 8: 'Stage 9', 9: 'Stage 10', 10: 'Stage 11', 11: 'Stage 12', 12: 'Stage 13', 13: 'Stage 14', 14: 'Stage 15', 15: 'Stage 16', 16: 'Stage 17', 17: 'Stage 18', 18: 'Stage 19', 19: 'Stage 20', 20: 'Stage 21', 21: 'Stage 22', 22: 'Stage 23', 23: 'Stage 24', 24: 'Stage 25', 25: 'Stage 26', 26: 'Stage 27', 27: 'Stage 28', 28: 'Stage 29', 29: 'Stage 30', 30: 'Stage 31', 31: 'Stage 32', 32: 'Stage 33', 33: 'Stage 34', 34: 'Stage 35', 35: 'Stage 36', 36: 'Stage 37', 37: 'Stage 38', 38: 'Stage 39', 39: 'Stage 40', 40: 'Stage 41', 41: 'Stage 42', 42: 'Stage 43', 43: 'Stage 44', 44: 'Stage 45', 45: 'Stage 46', 46: 'Stage 47', 47: 'Stage 48', 48: 'Stage 49', 49: 'Stage 50', 50: 'Stage 51', 51: 'Stage 52', 52: 'Stage 53', 53: 'Stage 54', 54: 'Stage 55', 55: 'Stage 56', 56: 'Stage 57', 57: 'Stage 58', 58: 'Stage 59', 59: 'Stage 60', 60: 'Stage 61', 61: 'Stage 62', 62: 'Stage 63', 63: 'Stage 64', 64: 'Stage 65', 65: 'Stage 66', 66: 'Stage 67', 67: 'Stage 68', 68: 'Stage 69', 69: 'Stage 70', 70: 'Stage 71', 71: 'Stage 72', 72: 'Stage 73', 73: 'Stage 74', 74: 'Stage 75', 75: 'Stage 76', 76: 'Stage 77', 77: 'Stage 78', 78: 'Stage 79', 79: 'Stage 80', 80: 'Stage 81', 81: 'Stage 82', 82: 'Stage 83', 83: 'Stage 84', 84: 'Stage 85', 85: 'Stage 86', 86: 'Stage 87', 87: 'Stage 88', 88: 'Stage 89', 89: 'Stage 90', 90: 'Stage 91', 91: 'Stage 92', 92: 'Stage 93', 93: 'Stage 94', 94: 'Stage 95', 95: 'Stage 96', 96: 'Stage 97', 97: 'Stage 98', 98: 'Stage 99', 99: 'Stage 100}'}
```

```
In [78]: df_player_results = pd.DataFrame(player_results)
```

```
In [79]: df_player_results
```

Out[79]:

	player_id	match_date	match_result
0	401	2021-05-04 00:00:00	W
1	401	2021-05-09 00:00:00	L
2	401	2021-05-16 00:00:00	L
3	401	2021-05-18 00:00:00	W
4	401	2021-05-22 00:00:00	L
5	401	2021-06-15 00:00:00	L
6	401	2021-06-16 00:00:00	W
7	401	2021-06-18 00:00:00	W
8	401	2021-07-06 00:00:00	L
9	401	2021-07-13 00:00:00	L
10	402	2021-05-14 00:00:00	L
11	402	2021-05-23 00:00:00	L
12	402	2021-05-24 00:00:00	W
13	402	2021-06-01 00:00:00	W
14	402	2021-06-02 00:00:00	W
15	402	2021-07-01 00:00:00	W
16	402	2021-07-11 00:00:00	W
17	402	2021-07-20 00:00:00	L
18	402	2021-07-26 00:00:00	L
19	402	2021-07-30 00:00:00	L
20	403	2021-05-03 00:00:00	L
21	403	2021-05-11 00:00:00	W
22	403	2021-05-12 00:00:00	W
23	403	2021-05-13 00:00:00	W

	player_id	match_date	match_result
24	403	2021-05-20 00:00:00	W
25	403	2021-05-25 00:00:00	W
26	403	2021-07-06 00:00:00	L
27	403	2021-07-15 00:00:00	L
28	403	2021-07-22 00:00:00	W
29	403	2021-07-23 00:00:00	W
30	404	2021-05-10 00:00:00	W
31	404	2021-05-16 00:00:00	W
32	404	2021-05-20 00:00:00	W
33	404	2021-05-22 00:00:00	W
34	404	2021-05-28 00:00:00	L
35	404	2021-06-06 00:00:00	L
36	404	2021-06-14 00:00:00	W
37	404	2021-07-25 00:00:00	W
38	404	2021-07-26 00:00:00	L
39	405	2021-05-07 00:00:00	L
40	405	2021-05-25 00:00:00	L
41	405	2021-06-06 00:00:00	L
42	405	2021-06-07 00:00:00	L
43	405	2021-06-14 00:00:00	L
44	405	2021-07-01 00:00:00	L
45	405	2021-07-02 00:00:00	L
46	405	2021-07-14 00:00:00	W
47	405	2021-07-16 00:00:00	L

	player_id	match_date	match_result
48	405	2021-07-30 00:00:00	L

```
In [88]: df_403 = df_player_results[df_player_results["player_id"] == 401]
```

```
In [89]: df_403["shifted_result"] = df_403["match_result"].shift()
```

C:\Users\mause\AppData\Local\Temp\ipykernel\_12740\1930227303.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row\_indexer,col\_indexer] = value instead

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_403["shifted_result"] = df_403["match_result"].shift()
```

```
In [90]: df_403
```

```
Out[90]:
```

	player_id	match_date	match_result	shifted_result
0	401	2021-05-04 00:00:00	W	NaN
1	401	2021-05-09 00:00:00	L	W
2	401	2021-05-16 00:00:00	L	L
3	401	2021-05-18 00:00:00	W	L
4	401	2021-05-22 00:00:00	L	W
5	401	2021-06-15 00:00:00	L	L
6	401	2021-06-16 00:00:00	W	L
7	401	2021-06-18 00:00:00	W	W
8	401	2021-07-06 00:00:00	L	W
9	401	2021-07-13 00:00:00	L	L

```
In [93]: df_403["start_streak"] = df_403["match_result"] != df_403["shifted_result"]
```

```
C:\Users\mause\AppData\Local\Temp\ipykernel_12740\1735739422.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_403["start_streak"] = df_403["match_result"] != df_403["shifted_result"]
```

In [94]: df\_403

Out[94]:

	player_id	match_date	match_result	shifted_result	start_streak
0	401	2021-05-04 00:00:00	W	NaN	True
1	401	2021-05-09 00:00:00	L	W	True
2	401	2021-05-16 00:00:00	L	L	False
3	401	2021-05-18 00:00:00	W	L	True
4	401	2021-05-22 00:00:00	L	W	True
5	401	2021-06-15 00:00:00	L	L	False
6	401	2021-06-16 00:00:00	W	L	True
7	401	2021-06-18 00:00:00	W	W	False
8	401	2021-07-06 00:00:00	L	W	True
9	401	2021-07-13 00:00:00	L	L	False

In [112...]

```
df_403["streak_id"] = df_403["start_streak"].cumsum()
```

```
C:\Users\mause\AppData\Local\Temp\ipykernel_12740\486196066.py:1: SettingWithCopyWarning:  
A value is trying to be set on a copy of a slice from a DataFrame.  
Try using .loc[row_indexer,col_indexer] = value instead
```

See the caveats in the documentation: [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)

```
df_403["streak_id"] = df_403["start_streak"].cumsum()
```

In [113...]

```
df_403
```

Out[113]:

	player_id	match_date	match_result	shifted_result	start_streak	start_id	streak_id
0	401	2021-05-04 00:00:00	W	NaN	1	1	1
1	401	2021-05-09 00:00:00	L	W	2	3	3
2	401	2021-05-16 00:00:00	L	L	2	5	5
3	401	2021-05-18 00:00:00	W	L	3	8	8
4	401	2021-05-22 00:00:00	L	W	4	12	12
5	401	2021-06-15 00:00:00	L	L	4	16	16
6	401	2021-06-16 00:00:00	W	L	5	21	21
7	401	2021-06-18 00:00:00	W	W	5	26	26
8	401	2021-07-06 00:00:00	L	W	6	32	32
9	401	2021-07-13 00:00:00	L	L	6	38	38

In [114...]: df\_403 = df\_403.groupby(["streak\_id","match\_result"])["player\_id"].count().reset\_index()

In [115...]

df\_403

Out[115]:

	streak_id	match_result	player_id
0	1	W	1
1	3	L	1
2	5	L	1
3	8	W	1
4	12	L	1
5	16	L	1
6	21	W	1
7	26	W	1
8	32	L	1
9	38	L	1

In [ ]: