

Фінальний проект

Частина 1. Building an End-to-End Streaming Pipeline

Зчитані дані фізичних показників атлетів (155861 записів):

athlete_id	name	sex	born	height	weight	country	country_noc	description	special_notes
65649	IvankaBonova	Female	4April1949	166.0	55	Bulgaria	BUL	PersonalBest40053...	nan
112510	NataliyaUryadova	Female	15March1977	184.0	70	RussianFederation	RUS	nan	ListedinOlympians...
114973	EssaIsmailRashed	Male	14December1986	165.0	55	Qatar	QAT	PersonalBest10000...	ListedinOlympians...
30359	PterBoros	Male	12January1908		nan	Hungary	HUN	Between1927and193...	nan
50557	RudolfPiowat	Male	28April1900		nan	Czechoslovakia	TCH	RudolfPiowatyjoin...	nan
146111	SvetlanaKholomina	Female	9November1997		nan	ROC	ROC	nan	nan
133041	VincentRiendeau	Male	13December1996	178.0	68	Canada	CAN	nan	ListedinOlympians...
110425	TanjaMorel	Female	4October1975	164.0	58	Switzerland	SUI	nan	ListedinOlympians...
110705	MaksimShabalin	Male	25January1982	183.0	76	RussianFederation	RUS	nan	ListedinOlympians...
54541	GREgter	Male	6March1916		nan	Netherlands	NED	nan	nan
22721	AristidePontenani	Male		nan	nan	Italy	ITA	nan	nan
56266	GoYeongChang	Male	21March1926	167.0	75	RepublicofKorea	KOR	nan	nan
82227	MarliesRostock	Female	20April1960		nan	EastGermany	GDR	MarliesRostockwon...	Wifedivorcedoflwe...
93334	CraigHutchison	Male	26May1975	198.0	97	Canada	CAN	nan	ListedinOlympians...
146013	RaquelQueirs	Female	4March2000	167.0	56	Portugal	POR	nan	nan
109912	VyacheslavKurginyan	Male	22December1986	170.0	65	RussianFederation	RUS	nan	ListedinOlympians...
37019	PhilippeLot	Male	13June1967	189.0	85	France	FRA	nan	ListedinOlympians...
22885	RudyKugeler	Male	11August1928	187.0	86	Luxembourg	LUX	nan	nan
95497	YoshihiroMiyazaki	Male	10May1930	173.0	70	Japan	JPN	nan	nan
76	RoperBarrett	Male	24November1873		nan	GreatBritain	GBR	RoperBarrettlearn...	nan
47023	LornaFrampton	Female	2April1920		nan	GreatBritain	GBR	LornaFramptonwasj...	ListedinOlympians...
42069	IsabelleHberl	Female	11May1959	167.0	67	France	FRA	nan	SisterofPhilippeH...
39033	SjoerdWartena	Male	1May1939	189.0	76	Netherlands	NED	nan	nan
115157	MariyaYakovenko	Female	6January1982	174.0	81	RussianFederation	RUS	PersonalBestJT62...	nan
207	VirgCsurg	Female	10November1972	172.0	63	Hungary	HUN	VirgCsurgwasaprof...	ListedinOlympians...
99106	RonnyYeager	Male	16August1952	181.0	70	UnitedStates	USA	RonnyYeagerskiedf...	ListedinOlympians...
126257	ShaneSmeltz	Male	29September1981	184.0	79	NewZealand	NZL	nan	ListedinOlympians...
102554	HwangSeonHong	Male	14July1968	183.0	79	RepublicofKorea	KOR	nan	ListedinOlympians...
24775	FathiKhorshid	Male	11December1937		nan	EgyptUnitedArabRe...	EGY	Growingupintheagr...	ListedinOlympians...
69614	HarryWood	Male	28November1902		nan	GreatBritain	GBR	Finishing11thath...	nan
52387	Ulfttken	Male	31May1953	196.0	86	WestGermany	FRG	UlfttkenwasawestG...	nan
100169	GalitChait	Female	29January1975	153.0	45	Israel	ISR	Dance1994WorldCha...	ListedinOlympians...
142869	JulienMartine	Male	25June1988		nan	France	FRA	nan	nan
63052	JimmyMooney	Male	25February1920	182.0	71	Ireland	IRL	Deceased.	nan
901207	AnneMetteChristensen	Female	23December1963		nan	Denmark	DEN	nan	nan
136267	ShadrackKipchirchir	Male	22February1989	173.0	54	UnitedStates	USA	nan	nan
1005805	JosHansen	Male	8June1932		nan	Luxembourg	LUX	nan	nan
95761	AlainNavillod	Male	16August1955	169.0	68	France	FRA	nan	nan
62527	AndreasJosenhans	Male	12September1950	183.0	81	Canada	CAN	nan	nan
112346	PaulEtiaNdoumb	Male	20April1984	182.0	78	Cameroon	CMR	nan	nan

only showing top 40 rows

Зчитуємо дані атлетів з бази.

Відфільтровані дані атлетів (105090 записів):

athlete_id	name	sex	born	height	weight	country	country_noc	description	special_notes
65649	IvankaBonova	Female	4April1949	166.0	55.0	Bulgaria	BUL	PersonalBest40053...	nan
112510	NataliyaUryadova	Female	15March1977	184.0	70.0	RussianFederation	RUS	nan	ListedinOlympians...
114973	EssaIsmailRashed	Male	14December1986	165.0	55.0	Qatar	QAT	PersonalBest10000...	ListedinOlympians...
133041	VincentRiendeau	Male	13December1996	178.0	68.0	Canada	CAN	nan	ListedinOlympians...
110425	TanjaMorel	Female	4October1975	164.0	58.0	Switzerland	SUI	nan	ListedinOlympians...
110705	MaksimShabalin	Male	25January1982	183.0	76.0	RussianFederation	RUS	nan	ListedinOlympians...
56266	GoYeongChang	Male	21March1926	167.0	75.0	RepublicofKorea	KOR	nan	nan
93334	CraigHutchison	Male	26May1975	198.0	97.0	Canada	CAN	nan	ListedinOlympians...
146013	RaquelQueirs	Female	4March2000	167.0	56.0	Portugal	POR	nan	nan
109912	VyacheslavKurginyan	Male	22December1986	170.0	65.0	RussianFederation	RUS	nan	ListedinOlympians...
37019	PhilippeLot	Male	13June1967	189.0	85.0	France	FRA	nan	ListedinOlympians...
22885	RudyKugeler	Male	11August1928	187.0	86.0	Luxembourg	LUX	nan	nan
95497	YoshihiroMiyazaki	Male	10May1930	173.0	70.0	Japan	JPN	nan	nan
42069	IsabelleHberl	Female	11May1959	167.0	67.0	France	FRA	nan	SisterofPhilippeH...
39033	SjoerdWartena	Male	1May1939	189.0	76.0	Netherlands	NED	nan	nan
115157	MariyaYakovenko	Female	6January1982	174.0	81.0	RussianFederation	RUS	PersonalBestJT62...	nan
207	VirgCsurg	Female	10November1972	172.0	63.0	Hungary	HUN	VirgCsurgwasaprof...	ListedinOlympians...
99106	RonnyYeager	Male	16August1952	181.0	70.0	UnitedStates	USA	RonnyYeagerskiedf...	ListedinOlympians...
126257	ShaneSmeltz	Male	29September1981	184.0	79.0	NewZealand	NZL	nan	ListedinOlympians...
102554	HwangSeonHong	Male	14July1968	183.0	79.0	RepublicofKorea	KOR	nan	ListedinOlympians...
52387	Ulfttken	Male	31May1953	196.0	86.0	WestGermany	FRG	UlfttkenwasawestG...	nan
100169	GalitChait	Female	29January1975	153.0	45.0	Israel	ISR	Dance1994WorldCha...	ListedinOlympians...
63052	JimmyMooney	Male	25February1920	182.0	71.0	Ireland	IRL	Deceased.	nan
136267	ShadrackKipchirchir	Male	22February1989	173.0	54.0	UnitedStates	USA	nan	nan
95761	AlainNavillod	Male	16August1955	169.0	68.0	France	FRA	nan	nan
62527	AndreasJosenhans	Male	12September1950	183.0	81.0	Canada	CAN	nan	nan
112346	PaulEtiaNdoumb	Male	20April1984	182.0	78.0	Cameroon	CMR	nan	nan
96506	PhilippePoirot	Male	30October1958	173.0	66.0	France	FRA	nan	nan
8923	OscarChongo	Male	25January1965	155.0	51.0	Zambia	ZAM	nan	nan
64560	EricaNixon	Female	15December1953	175.0	67.0	Australia	AUS	PersonalBestLJ6.5...	MotherofSteveHook...
115463	RyanHall	Male	14October1982	180.0	61.0	UnitedStates	USA	Hiswife,SaraHall...	nan
125355	AugustoSoares	Male	22August1986	165.0	58.0	TimorLeste	TLS	nan	nan
137459	KeijiTanaka	Male	22November1994	172.0	70.0	Japan	JPN	nan	ListedinOlympians...
58561	JzsefFarkas	Male	17March1952	192.0	100.0	Hungary	HUN	JzsefFarkashadthe...	nan
2821	IlijaLupulesku	Male	30October1967	177.0	79.0	IndependentOlympi...	IOA	nan	Husbanddivorcedof...
77525	BorisKuznetsov	Male	12June1947	175.0	63.0	SovietUnion	URS	PersonalBest50001...	ListedinOlympians...
127330	BenjaminMaier	Male	19April1994	182.0	93.0	Austria	AUT	nan	BrotherofRaphaelM...
37731	BirteSiech	Female	19March1967	180.0	75.0	EastGermanyGermany	GDR	RowerBirteSiechwo...	nan
90206	IhamKrimov	Male	2July1976	180.0	81.0	Azerbaijan	AZE	nan	nan
27348	FrantiekKunzo	Male	17September1954	175.0	70.0	Czechoslovakia	TCH	nan	nan

only showing top 40 rows

Видаляємо пусті значення

Зчитані дані результатів (316834 записів):											
edition	edition_id	country_noc	sport	event	result_id	athlete	athlete_id	pos	medal	isTeamSport	
1908SummerOlympics	5	ANZ	Athletics	100metres,Men	56265	ErnestHutcheon	64710	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	400metres,Men	56313	HenryMurray	64756	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	800metres,Men	56338	HarveySutton	64808	3h8r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	800metres,Men	56338	GuyHaskins	922519	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	800metres,Men	56338	JosephLynch	64735	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	800metres,Men	56338	HenryMurray	64756	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	1,500metres,Men	56349	JosephLynch	64735	5h2r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	1,500metres,Men	56349	CharlesSwain	79576	ACH3r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	1,500metres,Men	56349	GuyHaskins	922519	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	1,500metres,Men	56349	GeorgeBlake	64619	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	5miles,Men	56360	GeorgeBlake	64619	3h1r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	5miles,Men	56360	JosephLynch	64735	ACH5r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	5miles,Men	56360	GuyHaskins	922519	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	5miles,Men	56360	VictorAitken	64594	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	Marathon,Men	56369	VictorAitken	64594	DNF	nan	False	
1908SummerOlympics	5	ANZ	Athletics	Marathon,Men	56369	GeorgeBlake	64619	DNF	nan	False	
1908SummerOlympics	5	ANZ	Athletics	Marathon,Men	56369	JosephLynch	64735	DNF	nan	False	
1908SummerOlympics	5	ANZ	Athletics	110metresHurdles,Men	56370	HenryMurray	64756	2h1r13	nan	False	
1908SummerOlympics	5	ANZ	Athletics	400metresHurdles,Men	56392	HenryMurray	64756	2h3r13	nan	False	
1908SummerOlympics	5	ANZ	Athletics	3,200metresSteepl...	56412	GuyHaskins	922519	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	3,500metresRaceWa...	56421	HarryKerr	64719	3	Bronze	False	
1908SummerOlympics	5	ANZ	Athletics	3,500metresRaceWa...	56421	AlbertRowland	64788	5	nan	False	
1908SummerOlympics	5	ANZ	Athletics	10milesRaceWalk,Men	56427	HarryKerr	64719	DNSr22	nan	False	
1908SummerOlympics	5	ANZ	Athletics	10milesRaceWalk,Men	56427	AlbertRowland	64788	5h1r12	nan	False	
1908SummerOlympics	5	ANZ	Athletics	10milesRaceWalk,Men	56427	VictorAitken	64594	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	StandingHighJump,Men	56454	ErnestHutcheon	64710	AC	nan	False	
1908SummerOlympics	5	ANZ	Athletics	LongJump,Men	56467	ErnestHutcheon	64710	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	StandingLongJump,Men	56475	ErnestHutcheon	64710	DNS	nan	False	
1908SummerOlympics	5	ANZ	Athletics	TripleJump,Men	56482	ErnestHutcheon	64710	DNS	nan	False	
1908SummerOlympics	5	ANZ	Boxing	Lightweight,Men	21247	GeorgeBlake	64619	DNS	nan	False	
1908SummerOlympics	5	ANZ	Boxing	Middleweight,Men	21263	SnowyBaker	45153	2	Silver	False	
1908SummerOlympics	5	ANZ	Diving	Springboard,Men	54013	SnowyBaker	45153	6p4r13	nan	False	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	JohnBarnett	11237	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	PhilCarmichael	11239	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	DanCarroll	11240	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	BobCraig	11241	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	TomGriffin	11242	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	JackHickey	11243	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	MannieMcArthur	11244	1	Gold	True	
1908SummerOlympics	5	ANZ	Rugby	Rugby,Men	31505	ArthurMcCabe	11245	1	Gold	True	

only showing top 40 rows

Записуємо в кафка-топiк та зчитуємо дані по результатам змагань з перетворенням даних у датафрейм.

Batch: 2

sport	medal	sex	country_noc	avg_height	avg_weight	timestamp
Athletics	nan	Male	CAF	181.8	72.4	2024-12-21 04:55:...
Athletics	Silver	Male	NULL	183.2	78.2	2024-12-21 04:55:...
IceHockey	nan	Female	NULL	167.0	63.8	2024-12-21 04:55:...
Judo	nan	Male	ALB	175.0	81.0	2024-12-21 04:55:...
EquestrianEventing	nan	Female	NULL	166.9	95.4	2024-12-21 04:55:...
Shooting	Silver	Male	NULL	176.3	80.3	2024-12-21 04:55:...
Taekwondo	nan	Male	ARM	185.0	80.0	2024-12-21 04:55:...
ArtisticGymnastics	nan	Female	BOL	150.0	40.0	2024-12-21 04:55:...
FreestyleSkiing	nan	Female	NULL	165.4	59.0	2024-12-21 04:55:...
CrossCountrySkiing	nan	Male	ARM	170.5	67.1	2024-12-21 04:55:...
CyclingTrack	nan	Male	NULL	178.6	135.9	2024-12-21 04:55:...
CyclingRoad	nan	Male	BER	179.7	72.7	2024-12-21 04:55:...
Swimming	nan	Male	CMR	179.5	80.0	2024-12-21 04:55:...
Athletics	nan	Female	ALB	166.4	53.9	2024-12-21 04:55:...
Shooting	Silver	Male	BLR	180.0	97.0	2024-12-21 04:55:...
Swimming	nan	Male	DJI	180.0	70.0	2024-12-21 04:55:...
Shooting	Bronze	Male	BLR	175.3	75.7	2024-12-21 04:55:...
CanoeSlalom	Bronze	Male	NULL	177.8	73.1	2024-12-21 04:55:...
Rowing	nan	Female	NULL	176.4	83.6	2024-12-21 04:55:...
Swimming	nan	Female	BEN	167.0	57.0	2024-12-21 04:55:...

only showing top 20 rows

Batch: 3

sport	medal	sex	country_noc	avg_height	avg_weight	timestamp
Athletics	Silver	Male	NULL	183.2	78.2	2024-12-21 04:55:...
IceHockey	nan	Female	NULL	167.0	63.8	2024-12-21 04:55:...
CanoeSlalom	nan	Female	GDR	164.0	60.5	2024-12-21 04:55:...
EquestrianEventing	nan	Female	NULL	166.9	95.4	2024-12-21 04:55:...
Badminton	Silver	Male	INA	174.2	70.7	2024-12-21 04:55:...
Taekwondo	nan	Male	ARM	185.0	80.0	2024-12-21 04:55:...
ArtisticGymnastics	nan	Female	BOL	150.0	40.0	2024-12-21 04:55:...
ArtisticGymnastics	nan	Male	EGY	165.8	65.9	2024-12-21 04:55:...
FreestyleSkiing	nan	Female	NULL	165.4	59.0	2024-12-21 04:55:...
Hockey	nan	Female	BEL	169.6	62.0	2024-12-21 04:55:...
CrossCountrySkiing	nan	Male	ARM	170.5	67.1	2024-12-21 04:55:...
CyclingTrack	nan	Male	NULL	178.6	135.9	2024-12-21 04:55:...
Athletics	nan	Male	CAF	181.8	72.4	2024-12-21 04:55:...
Judo	nan	Male	ALB	175.0	81.0	2024-12-21 04:55:...
Shooting	Silver	Male	NULL	176.3	80.3	2024-12-21 04:55:...
Sailing	nan	Female	CZE	172.0	63.2	2024-12-21 04:55:...
Sailing	Silver	Male	GDR	176.8	80.6	2024-12-21 04:55:...
Luge	Silver	Female	GDR	172.4	72.4	2024-12-21 04:55:...
CyclingRoad	nan	Male	BER	179.7	72.7	2024-12-21 04:55:...
Shooting	Bronze	Male	BLR	175.3	75.7	2024-12-21 04:55:...

only showing top 20 rows

Результат потокового обчислення середніх ваги та росту атлетів для кожного виду спорту, типу медалі або її відсутності, статі, країни.

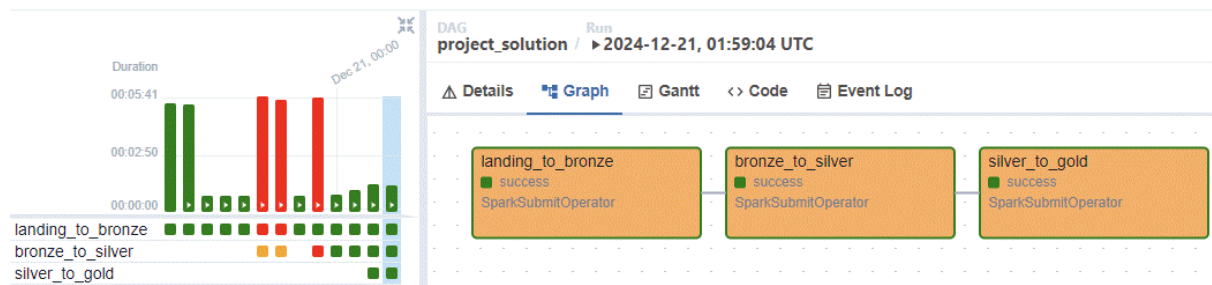
А ось п.6 з першої частини завдання я не зміг побороти.

Частина 2. Building an End-to-End Batch Data Lake

landing_to_bronze.py:

```
[2024-12-21, 01:59:08 UTC] {spark_submit.py:488} INFO - Script landing_to_bronze started...
[2024-12-21, 01:59:08 UTC] {spark_submit.py:488} INFO - Downloading from https://ftp.goit.study/neoversity/athlete\_bio.csv
[2024-12-21, 01:59:12 UTC] {spark_submit.py:488} INFO - File downloaded successfully and saved as dags/vwd/athlete_bio.csv
[2024-12-21, 01:59:12 UTC] {spark_submit.py:488} INFO - Downloading from https://ftp.goit.study/neoversity/athlete\_event\_results.csv
[2024-12-21, 01:59:14 UTC] {spark_submit.py:488} INFO - File downloaded successfully and saved as dags/vwd/athlete_event_results.csv
```

Зчитуємо таблиці



Скрін відпрацьованого DAG-файлу

[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[edition]	[edition_id]	[country_noc]	[sport]	[event]	[result_id]	[athlete]	[athlete_id]	[pos]	[medal]	[isTeamSport]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----											
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[300 metres, Men]	[56265]	[Ernest Hutchence]	[64710]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[400 metres, Men]	[56313]	[Henry Murray]	[64756]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[800 metres, Men]	[56338]	[Harvey Sutton]	[64808]	[3 h 8 r/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[800 metres, Men]	[56338]	[Guy Haskins]	[922519]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[800 metres, Men]	[56338]	[Joseph Lynch]	[64735]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[800 metres, Men]	[56338]	[Henry Murray]	[64756]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[1,500 metres, Men]	[56349]	[Joseph Lynch]	[64735]	[5 h 2 r/3/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[1,500 metres, Men]	[56349]	[Charles Swain]	[70976]	[AC h 3 r/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[1,500 metres, Men]	[56349]	[Guy Haskins]	[922519]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[1,500 metres, Men]	[56349]	[George Blake]	[64619]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[5 miles, Men]	[56368]	[George Blake]	[64619]	[3 h 1 r/3/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[5 miles, Men]	[56368]	[Joseph Lynch]	[64735]	[AC h 5 r/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[5 miles, Men]	[56368]	[Guy Haskins]	[922519]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[5 miles, Men]	[56368]	[Victor Aitken]	[64594]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Marathon, Men]	[56369]	[Victor Aitken]	[64594]	[DNF]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Marathon, Men]	[56369]	[George Blake]	[64619]	[DNF]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Marathon, Men]	[56369]	[Joseph Lynch]	[64735]	[DNF]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[310 metres Hurdles, Men]	[56378]	[Henry Murray]	[64756]	[2 h 1 r/3/1]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[400 metres Hurdles, Men]	[56392]	[Henry Murray]	[64756]	[2 h 3 r/3/3]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[3,200 metres Steeplechase, Men]	[56412]	[Guy Haskins]	[922519]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[3,500 metres Race Walk, Men]	[56421]	[Harry Kerr]	[64713]	[3 h 1 Bronze]	[false]	
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[3,500 metres Race Walk, Men]	[56421]	[Albert Rowland]	[64788]	[5 h 1]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[10 miles Race Walk, Men]	[56427]	[Harry Kerr]	[64719]	[DNS r/2/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[10 miles Race Walk, Men]	[56427]	[Albert Rowland]	[64788]	[5 h 1 r/2/2]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[10 miles Race Walk, Men]	[56427]	[Victor Aitken]	[64594]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Standing High Jump, Men]	[56454]	[Ernest Hutchence]	[64710]	[AC]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Long Jump, Men]	[56467]	[Ernest Hutchence]	[64710]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Long Jump, Men]	[56467]	[Ernest Hutchence]	[64710]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Athletics]	[Triple Jump, Men]	[56462]	[Ernest Hutchence]	[64710]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[1908 Summer Olympics]	5	[ANZ]	[Boxing]	[Lightweight, Men]	[21247]	[George Blake]	[64619]	[DNS]	[NULL]	[false]
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	only showing top 30 rows										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	root										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- edition: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- edition_id: integer (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- country_noc: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- sport: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- event: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- result_id: integer (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- athlete: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- athlete_id: integer (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- pos: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- medal: string (nullable = true)										
[2024-12-21, 01:59:29 UTC]	[spark_submit.py:488]	INFO	----	[- isTeamSport: boolean (nullable = true)										

silver_to_gold.py:

Очищений текст з дедуплікацією.

silver_to_gold.py:

