

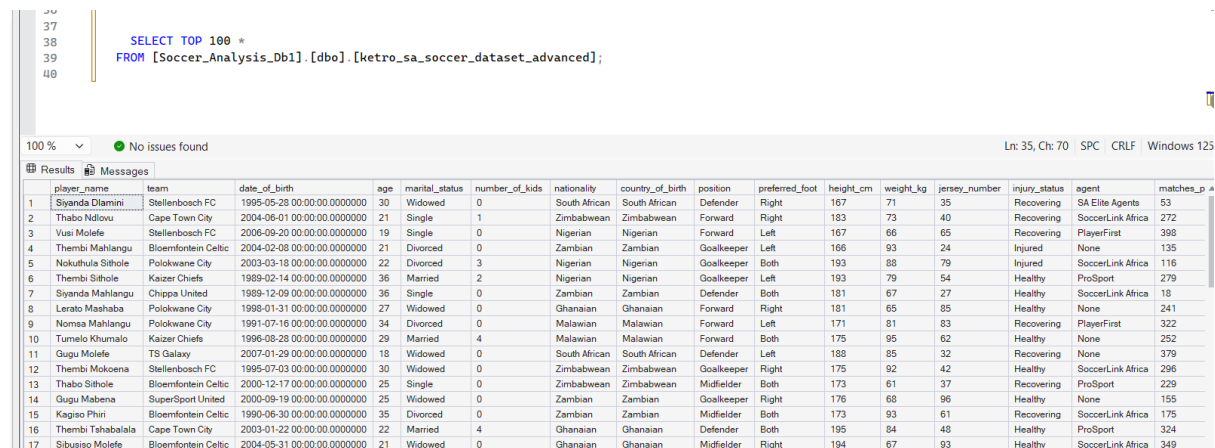
Sonali Matadin

South African Soccer Dataset – SQL

1. View the first 100 rows of the dataset to understand its structure.

Sql:

```
SELECT TOP 100 *  
FROM [Soccer_Analysis_Db1].[dbo].[ketro_sa_soccer_dataset_advanced];
```

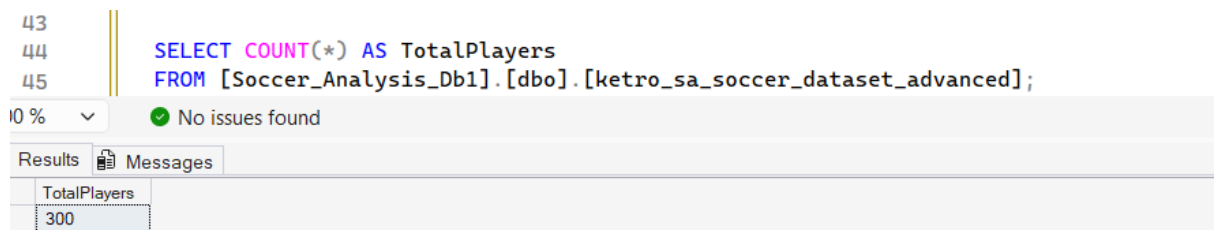


	player_name	team	date_of_birth	age	marital_status	number_of_kids	nationality	country_of_birth	position	preferred_foot	height_cm	weight_kg	jersey_number	injury_status	agent	matches_p
1	Siyanda Dlamini	Stellenbosch FC	1995-05-28 00:00:00.0000000	30	Widowed	0	South African	South African	Defender	Right	167	71	35	Recovering	SA Elite Agents	53
2	Thabo Ndlovu	Cape Town City	2004-06-01 00:00:00.0000000	21	Single	1	Zimbabwean	Zimbabwean	Forward	Right	183	73	40	Recovering	SoccerLink Africa	272
3	Vusi Molefe	Stellenbosch FC	2008-09-20 00:00:00.0000000	19	Single	0	Nigerian	Nigerian	Forward	Left	167	66	65	Recovering	PlayerFirst	398
4	Thembi Mahlangu	Bloemfontein Celtic	2004-02-08 00:00:00.0000000	21	Divorced	0	Zambian	Zambian	Goalkeeper	Left	166	93	24	Injured	None	135
5	Nokuthula Sithole	Polokwane City	2003-03-18 00:00:00.0000000	22	Divorced	3	Nigerian	Nigerian	Goalkeeper	Both	193	88	79	Injured	SoccerLink Africa	116
6	Thembi Sithole	Kaizer Chiefs	1989-02-14 00:00:00.0000000	36	Married	2	Nigerian	Nigerian	Goalkeeper	Left	193	79	54	Healthy	ProSport	279
7	Siyanda Mahlangu	Chippa United	1989-12-09 00:00:00.0000000	36	Single	0	Zambian	Zambian	Defender	Both	181	67	27	Healthy	SoccerLink Africa	18
8	Lerato Mashaba	Polokwane City	1998-01-31 00:00:00.0000000	27	Widowed	0	Ghanaian	Ghanaian	Forward	Right	181	65	85	Healthy	None	241
9	Nomosa Mahlangu	Polokwane City	1991-07-16 00:00:00.0000000	34	Divorced	0	Malawian	Malawian	Forward	Left	171	81	83	Recovering	PlayerFirst	322
10	Turnelo Khumalo	Kaizer Chiefs	1996-08-28 00:00:00.0000000	29	Married	4	Malawian	Malawian	Forward	Both	175	95	62	Healthy	None	252
11	Gugu Molefe	TS Galaxy	2007-01-29 00:00:00.0000000	18	Widowed	0	South African	South African	Defender	Left	188	85	32	Recovering	None	379
12	Thembi Mokoena	Stellenbosch FC	1995-07-03 00:00:00.0000000	30	Widowed	0	Zimbabwean	Zimbabwean	Goalkeeper	Right	175	92	42	Healthy	SoccerLink Africa	296
13	Thabo Sithole	Bloemfontein Celtic	2000-12-17 00:00:00.0000000	25	Single	0	Zimbabwean	Zimbabwean	Midfielder	Both	173	61	37	Recovering	ProSport	229
14	Gugu Mabena	SuperSport United	2000-09-19 00:00:00.0000000	25	Widowed	0	Zambian	Zambian	Goalkeeper	Right	176	68	96	Healthy	None	155
15	Kagiso Phiri	Bloemfontein Celtic	1990-06-30 00:00:00.0000000	35	Divorced	0	Zambian	Zambian	Midfielder	Both	173	93	61	Recovering	SoccerLink Africa	175
16	Thembi Tshabalala	Cape Town City	2003-01-22 00:00:00.0000000	22	Married	4	Ghanaian	Ghanaian	Defender	Both	195	84	48	Healthy	ProSport	324
17	Sibusiso Molefe	Bloemfontein Celtic	2004-05-31 00:00:00.0000000	21	Widowed	0	Ghanaian	Ghanaian	Midfielder	Right	194	67	93	Healthy	SoccerLink Africa	349

2. Count the total number of players in the dataset.

Sql:

```
SELECT COUNT(*) AS TotalPlayers  
FROM [Soccer_Analysis_Db1].[dbo].[ketro_sa_soccer_dataset_advanced];
```



TotalPlayers
300

- List all unique teams in the league.

Sql:

```
SELECT DISTINCT Team
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
```



48	SELECT DISTINCT Team
49	FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
100 %	✓ No issues found
Results	Messages
Team	
1 AmaZulu FC	
2 Bloemfontein Celtic	
3 Cape Town City	
4 Chippa United	
5 Golden Arrows	
6 Kaizer Chiefs	
7 Mamelodi Sundowns	
8 Moroka Swallows	
9 Orlando Pirates	
10 Polokwane City	
11 Richards Bay FC	
12 Royal AM	
13 Sekhukhune United	
14 Stellenbosch FC	
15 SuperSport United	
16 TS Galaxy	


- Count how many players are in each team.


Sql:

```
SELECT Team, COUNT(*) AS PlayersPerTeam
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Team;
```

52 SELECT Team, COUNT(*) AS PlayersPerTeam
53 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
54 GROUP BY Team;
55

100 %   No issues found

 Results

 Messages

	Team	PlayersPerTeam
1	AmaZulu FC	22
2	Bloemfontein Celtic	17
3	Cape Town City	15
4	Chippa United	21
5	Golden Arrows	23
6	Kaizer Chiefs	20
7	Mamelodi Sundowns	19
8	Moroka Swallows	19
9	Orlando Pirates	9
10	Polokwane City	27
11	Richards Bay FC	13
12	Royal AM	18
13	Sekhukhune United	17
14	Stellenbosch FC	20
15	SuperSport United	20
16	TS Galaxy	20

5. Identify the top 10 players with the most goals.

Sql:

```
SELECT TOP 10 player_name, Goals
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
ORDER BY Goals DESC;
```

57
58
59
60

```
SELECT TOP 10 player_name, Goals
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
ORDER BY Goals DESC;
```

100 % No issues found

Results Messages

	player_name	Goals
1	Vusi Molefe	99
2	Thabo Ndlovu	98
3	Gugu Hlongwane	98
4	Thabo Sithole	98
5	Mandla Mabena	98
6	Mpho Mahlangu	97
7	Boitumelo Nkosi	96
8	Mpho Radebe	92
9	Khanyi Nkosi	92
10	Sipho Phiri	91

6. Find the average salary for players in each team.

Sql:

```
SELECT
    Team,
    AVG(CAST(average_salary_zar AS FLOAT)) AS AvgSalary
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Team;
```

63
64
65
66
67
68

```
SELECT
    Team,
    AVG(CAST(average_salary_zar AS FLOAT)) AS AvgSalary
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Team;
```

100 % 1 0

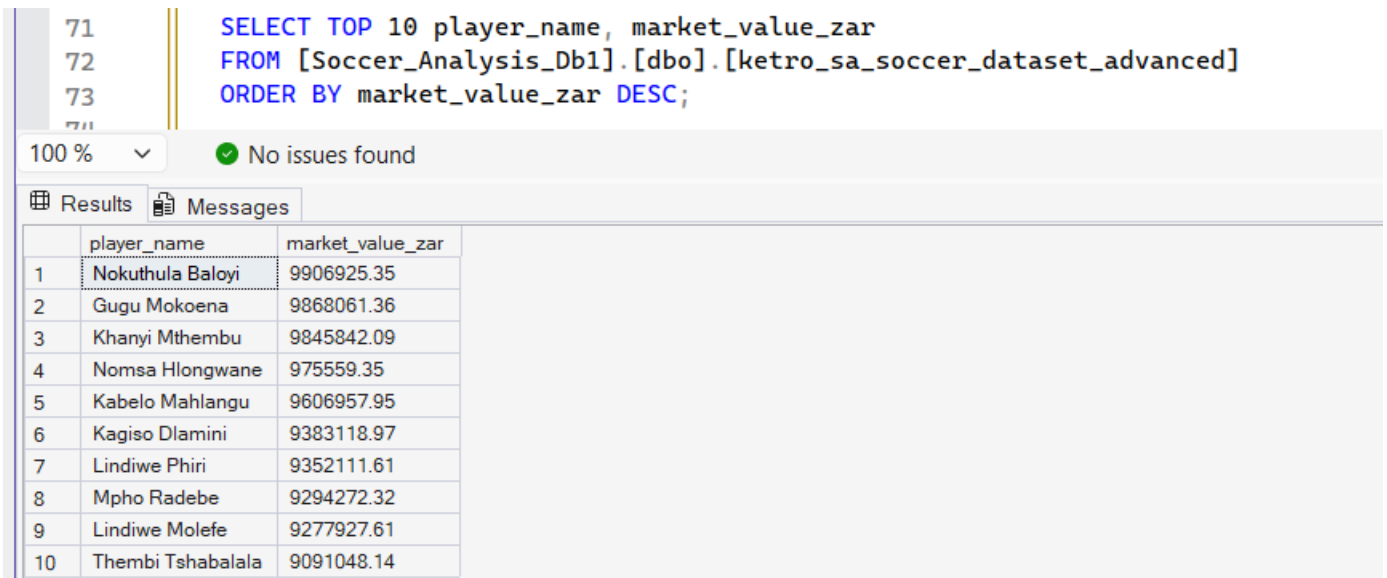
Results Messages

	Team	AvgSalary
1	AmaZulu FC	172467.569545455
2	Bloemfontein Celtic	163683.838823529
3	Cape Town City	208407.432
4	Chippa United	180947.634761905
5	Golden Arrows	199057.643913043
6	Kaizer Chiefs	188954.4615
7	Mamelodi Sundowns	194404.484736842
8	Moroka Swallows	186786.827894737
9	Orlando Pirates	178570.593333333
10	Polokwane City	171738.877777778
11	Richards Bay FC	193006.199230769
12	Royal AM	209078.331111111
13	Sekhukhune United	169945.162941176
14	Stellenbosch FC	182326.7355
15	SuperSport United	195482.9785
16	TS Galaxy	184719.574

7. Retrieve the top 10 players with the highest market value.

Sql:

```
SELECT TOP 10 player_name, market_value_zar
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
ORDER BY market_value_zar DESC;
```



```
71 SELECT TOP 10 player_name, market_value_zar
72 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
73 ORDER BY market_value_zar DESC;
```

100 % No issues found

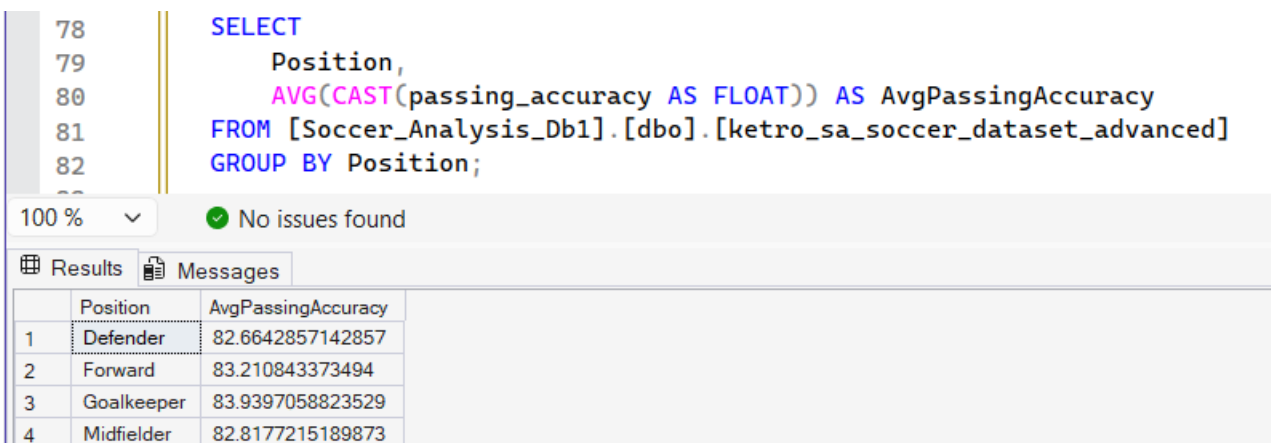
Results Messages

	player_name	market_value_zar
1	Nokuthula Baloyi	9906925.35
2	Gugu Mokoena	9868061.36
3	Khanyi Mthembu	9845842.09
4	Nomsa Hlongwane	975559.35
5	Kabelo Mahlangu	9606957.95
6	Kagiso Dlamini	9383118.97
7	Lindiwe Phiri	9352111.61
8	Mpho Radebe	9294272.32
9	Lindiwe Molefe	9277927.61
10	Thembi Tshabalala	9091048.14

8. Calculate the average passing accuracy for each position.

Sql:

```
SELECT
    Position,
    AVG(CAST(passing_accuracy AS FLOAT)) AS AvgPassingAccuracy
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Position;
```



```
78 SELECT
79     Position,
80     AVG(CAST(passing_accuracy AS FLOAT)) AS AvgPassingAccuracy
81 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
82 GROUP BY Position;
```

100 % No issues found

Results Messages

	Position	AvgPassingAccuracy
1	Defender	82.6642857142857
2	Forward	83.210843373494
3	Goalkeeper	83.9397058823529
4	Midfielder	82.8177215189873

9. Compare shot accuracy with goals to find correlations.

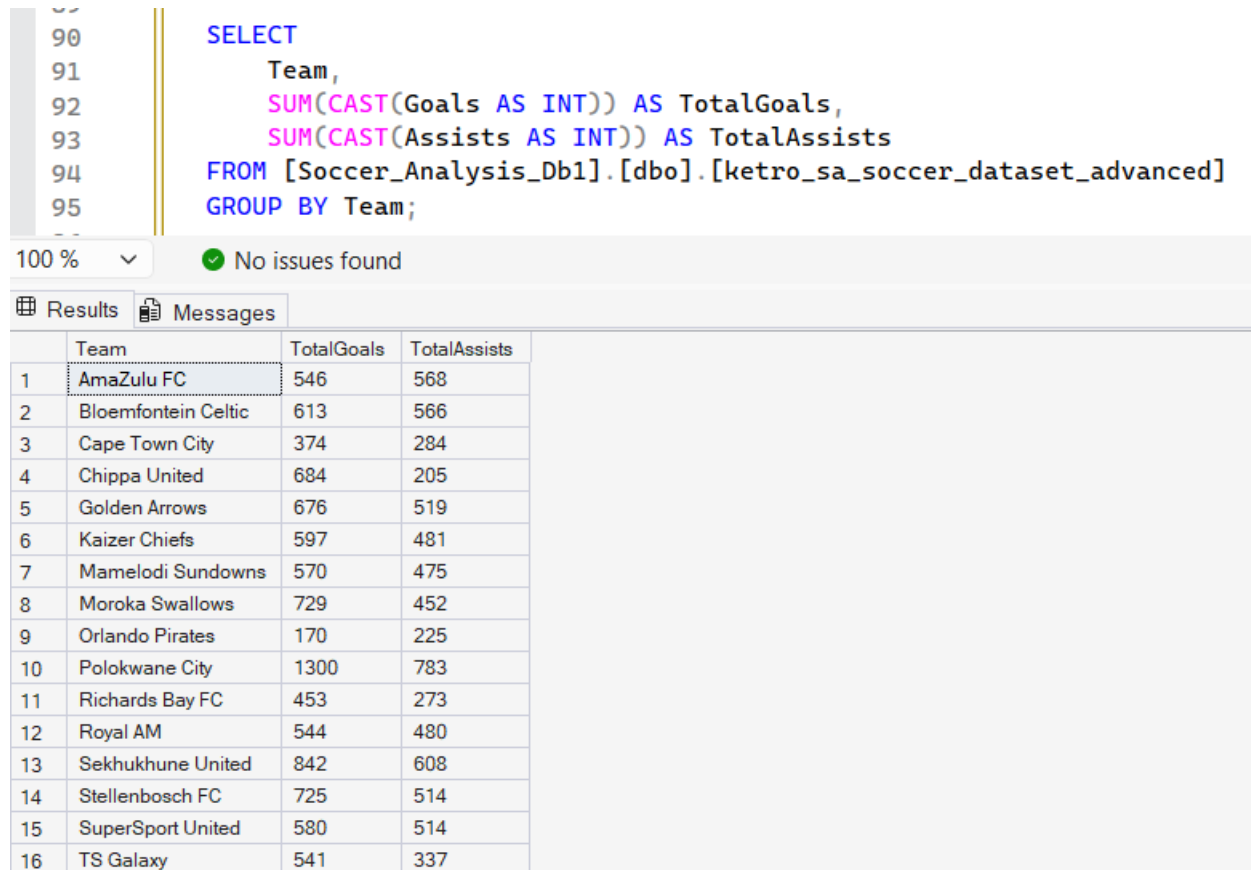
```
SELECT player_name, Goals, shot_accuracy
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
```

85		SELECT player_name, Goals, shot_accuracy
86		FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
87		
100 % No issues found		
Results Messages		
	player_name	Goals shot_accuracy
1	Siyanda Dlamini	9 83.7
2	Thabo Ndlovu	98 68.4
3	Vusi Molefe	99 79.0
4	Thembi Mahlangu	1 42.9
5	Nokuthula Sithole	2 74.0
6	Thembi Sithole	3 70.7
7	Siyanda Mahlangu	6 54.2
8	Lerato Mashaba	77 68.5
9	Nomsa Mahlangu	80 56.1
10	Tumelo Khumalo	48 46.0
11	Gugu Molefe	7 83.2
12	Thembi Mokoena	1 56.7
13	Thabo Sithole	14 80.6
14	Gugu Mabena	4 80.0
15	Kagiso Phiri	41 73.2
16	Thembi Tshabalala	7 76.7
17	Sibusiso Molefe	48 51.1
18	Lerato Tshabalala	10 56.8
19	Kagiso Tshabalala	36 64.4
20	Gugu Hlongwane	98 81.1
21	Ayanda Sithole	107 54.0
22	Mandla Baloyi	89 54.1
23	Nokuthula Baloyi	7 62.8
24	Zanele Molefe	116 61.2
25	Karabo Mthembu	5 41.5
26	Thembi Zulu	72 70.0
27	Khanyi Baloyi	2 54.8
28	Karabo Phiri	72 64.5

10. Compute total goals and assists for each team.

Sql:

```
SELECT
    Team,
    SUM(CAST(Goals AS INT)) AS TotalGoals,
    SUM(CAST(Assists AS INT)) AS TotalAssists
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Team;
```



The screenshot shows a SQL query editor with the following query:

```
SELECT
    Team,
    SUM(CAST(Goals AS INT)) AS TotalGoals,
    SUM(CAST(Assists AS INT)) AS TotalAssists
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Team;
```

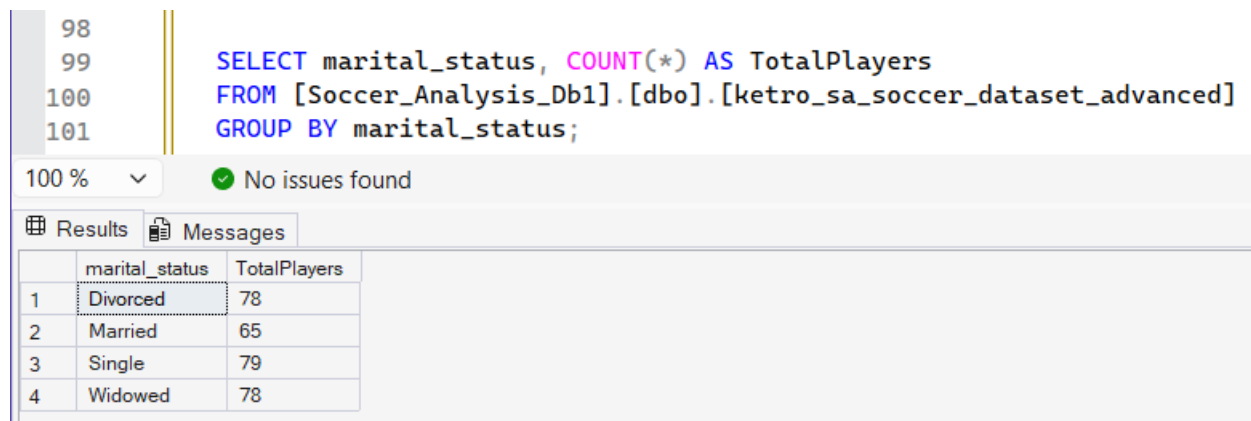
Below the query editor, the 'Results' tab is active, displaying a table with 4 columns: Team, TotalGoals, and TotalAssists. The table contains 16 rows of data, numbered 1 to 16. The status bar indicates '100 %' and 'No issues found'.

	Team	TotalGoals	TotalAssists
1	AmaZulu FC	546	568
2	Bloemfontein Celtic	613	566
3	Cape Town City	374	284
4	Chippa United	684	205
5	Golden Arrows	676	519
6	Kaizer Chiefs	597	481
7	Mamelodi Sundowns	570	475
8	Moroka Swallows	729	452
9	Orlando Pirates	170	225
10	Polokwane City	1300	783
11	Richards Bay FC	453	273
12	Royal AM	544	480
13	Sekhukhune United	842	608
14	Stellenbosch FC	725	514
15	SuperSport United	580	514
16	TS Galaxy	541	337

11. Count players by their marital status.

Sql:

```
SELECT marital_status, COUNT(*) AS TotalPlayers
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY marital_status;
```



The screenshot shows a SQL query editor with the following query:

```
SELECT marital_status, COUNT(*) AS TotalPlayers
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY marital_status;
```

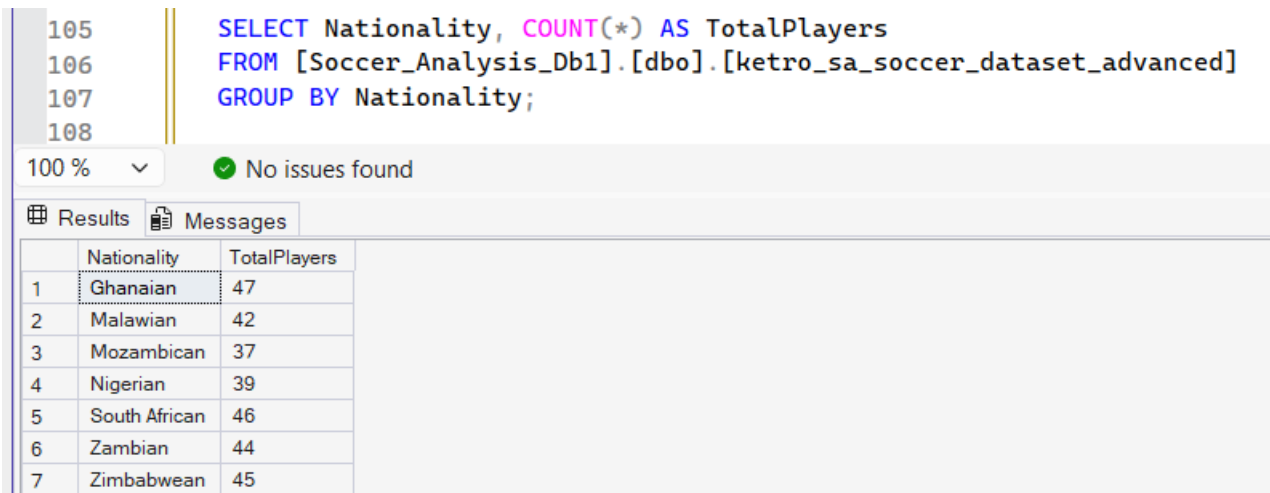
Below the query editor, the 'Results' tab is active, displaying a table with 2 columns: marital_status and TotalPlayers. The table contains 4 rows of data, numbered 1 to 4. The status bar indicates '100 %' and 'No issues found'.

	marital_status	TotalPlayers
1	Divorced	78
2	Married	65
3	Single	79
4	Widowed	78

12. Count players by nationality.

Sql:

```
SELECT Nationality, COUNT(*) AS TotalPlayers
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Nationality;
```



The screenshot shows a SQL query window with the following SQL code:

```
105 SELECT Nationality, COUNT(*) AS TotalPlayers
106 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
107 GROUP BY Nationality;
108
```

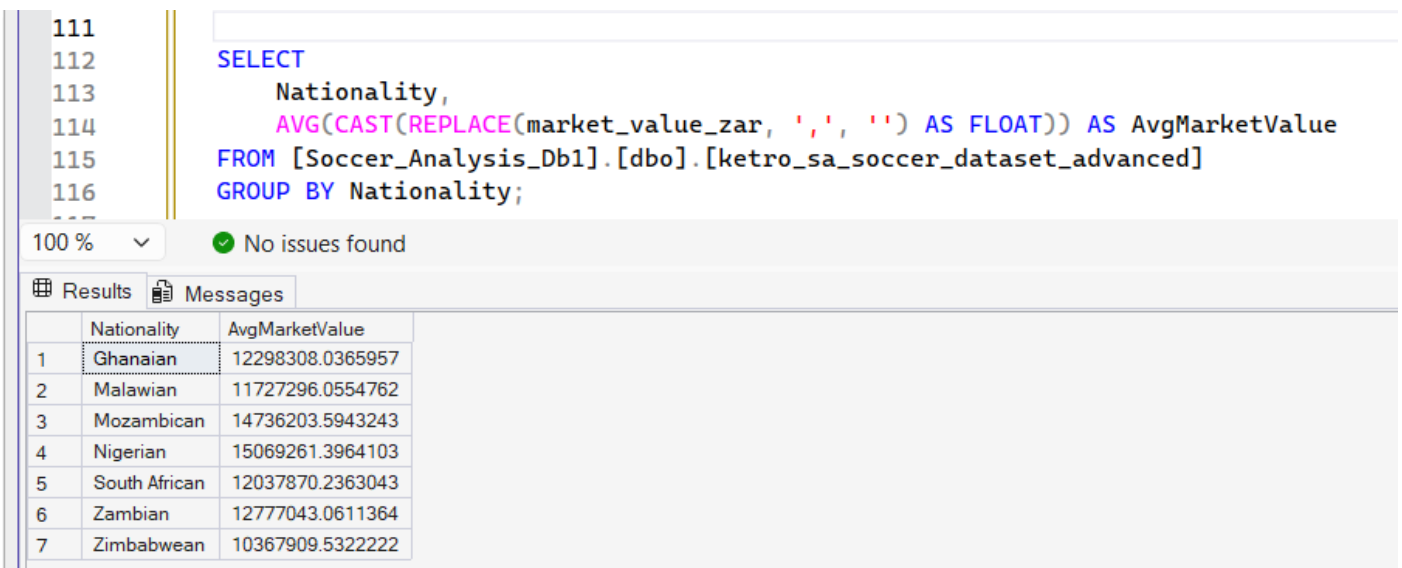
Below the query window, the 'Results' tab is active, displaying a table with 2 columns: 'Nationality' and 'TotalPlayers'. The table contains 7 rows of data.

	Nationality	TotalPlayers
1	Ghanaian	47
2	Malawian	42
3	Mozambican	37
4	Nigerian	39
5	South African	46
6	Zambian	44
7	Zimbabwean	45

13. Find average market value grouped by nationality.

Sql:

```
SELECT
    Nationality,
    AVG(CAST(REPLACE(market_value_zar, ',', '') AS FLOAT)) AS AvgMarketValue
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Nationality;
```



The screenshot shows a SQL query window with the following SQL code:

```
111 SELECT
112     Nationality,
113     AVG(CAST(REPLACE(market_value_zar, ',', '') AS FLOAT)) AS AvgMarketValue
114 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
115 GROUP BY Nationality;
116
```

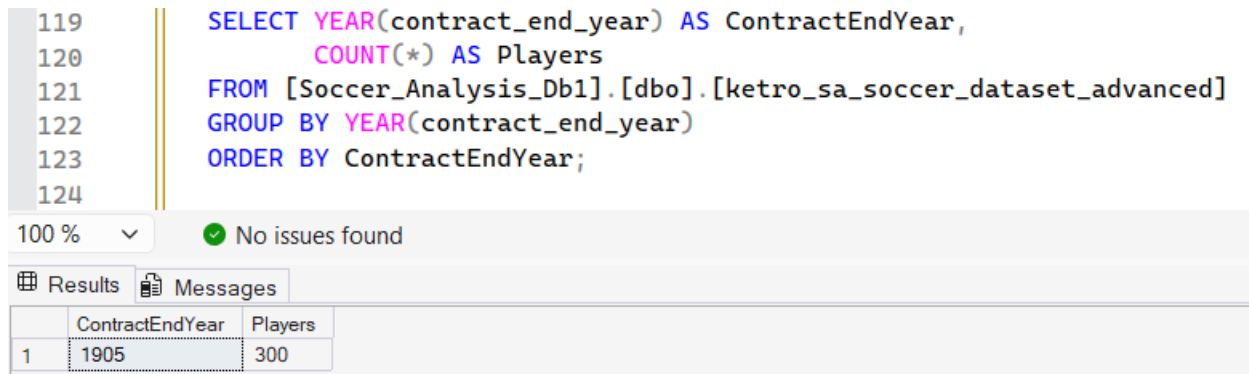
Below the query window, the 'Results' tab is active, displaying a table with 2 columns: 'Nationality' and 'AvgMarketValue'. The table contains 7 rows of data.

	Nationality	AvgMarketValue
1	Ghanaian	12298308.0365957
2	Malawian	11727296.0554762
3	Mozambican	14736203.5943243
4	Nigerian	15069261.3964103
5	South African	12037870.2363043
6	Zambian	12777043.0611364
7	Zimbabwean	10367909.5322222

14. Determine how many player contracts end in each year.

Sql:

```
SELECT YEAR(contract_end_year) AS ContractEndYear,  
       COUNT(*) AS Players  
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
GROUP BY YEAR(contract_end_year)  
ORDER BY ContractEndYear;
```



The screenshot shows a SQL query window with the following SQL code:

```
119 SELECT YEAR(contract_end_year) AS ContractEndYear,  
120       COUNT(*) AS Players  
121 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
122 GROUP BY YEAR(contract_end_year)  
123 ORDER BY ContractEndYear;  
124
```

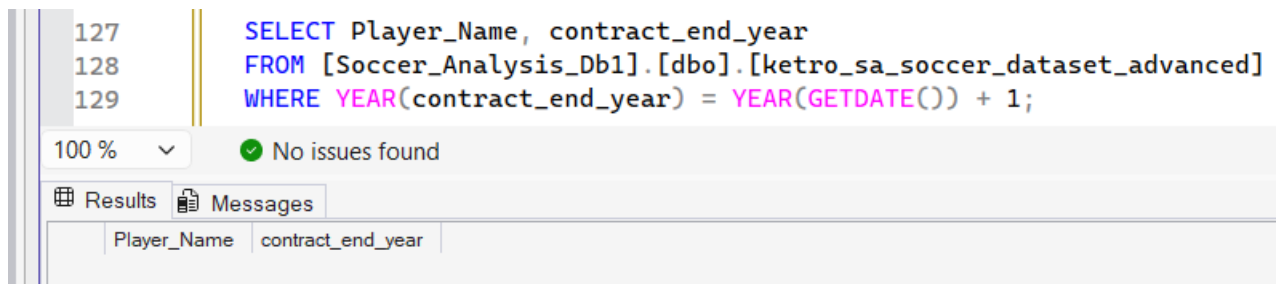
Below the query window, the 'Results' tab is active, showing a single row of data:

	ContractEndYear	Players
1	1905	300

15. Identify players whose contracts end next year.

Sql:

```
SELECT Player_Name, contract_end_year  
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
WHERE YEAR(contract_end_year) = YEAR(GETDATE()) + 1;
```



The screenshot shows a SQL query window with the following SQL code:

```
127 SELECT Player_Name, contract_end_year  
128 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
129 WHERE YEAR(contract_end_year) = YEAR(GETDATE()) + 1;
```

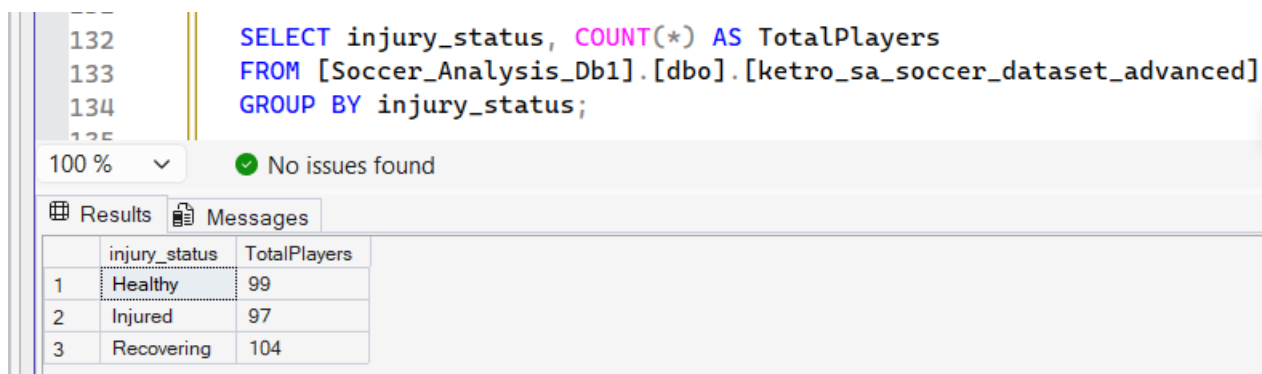
Below the query window, the 'Results' tab is active, showing a table with two columns:

Player_Name	contract_end_year
-------------	-------------------

16. Summarize the number of players by injury status.

Sql:

```
SELECT injury_status, COUNT(*) AS TotalPlayers  
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
GROUP BY injury_status;
```



The screenshot shows a SQL query window with the following SQL code:

```
132 SELECT injury_status, COUNT(*) AS TotalPlayers  
133 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
134 GROUP BY injury_status;  
135
```

Below the query window, the 'Results' tab is active, showing a table with two columns:

	injury_status	TotalPlayers
1	Healthy	99
2	Injured	97
3	Recovering	104

17. Calculate goals per match ratio for each player.

Sql:

```
SELECT player_name,
       Goals,
       matches_played,
       CASE WHEN matches_played = 0
            THEN 0
            ELSE CAST(Goals AS FLOAT) / matches_played
       END AS GoalsPerMatch
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
```

```
138      SELECT player_name,
139              Goals,
140              matches_played,
141              CASE WHEN matches_played = 0
142                   THEN 0
143                   ELSE CAST(Goals AS FLOAT) / matches_played
144              END AS GoalsPerMatch
145      FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced];
146
```

100 % No issues found

Results Messages

	player_name	Goals	matches_played	GoalsPerMatch
1	Siyanda Dlamini	9	53	0.169811320754717
2	Thabo Ndlovu	98	272	0.360294117647059
3	Vusi Molefe	99	398	0.248743718592965
4	Thembi Mahlangu	1	135	0.00740740740740741
5	Nokuthula Sithole	2	116	0.0172413793103448
6	Thembi Sithole	3	279	0.010752688172043
7	Siyanda Mahlangu	6	18	0.333333333333333
8	Lerato Mashaba	77	241	0.319502074688797
9	Nomsa Mahlangu	80	322	0.248447204968944
10	Tumelo Khumalo	48	252	0.19047619047619
11	Gugu Molefe	7	379	0.0184696569920844
12	Thembi Mokoena	1	296	0.00337837837837838
13	Thabo Sithole	14	229	0.0611353711790393
14	Gugu Mabena	4	155	0.0258064516129032
15	Kagiso Phiri	41	175	0.234285714285714
16	Thembi Tshabalala	7	324	0.0216049382716049
17	Sibusiso Molefe	48	349	0.137535816618911

18. Count how many players are managed by each agent.

Sql:

```
SELECT
    [Agent],
    COUNT(*) AS PlayersManaged
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY [Agent];
```

```
148 SELECT
149     [Agent],
150     COUNT(*) AS PlayersManaged
151 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
152 GROUP BY [Agent];
```

100 % No issues found

Results Messages

	Agent	PlayersManaged
1	None	62
2	PlayerFirst	63
3	ProSport	62
4	SA Elite Agents	51
5	SoccerLink Africa	62

19. Calculate average height and weight by player position.

Sql:

```
SELECT Position,
    AVG(height_cm) AS AvgHeight,
    AVG(weight_kg) AS AvgWeight
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
GROUP BY Position;
```

```
156 SELECT Position,
157     AVG(height_cm) AS AvgHeight,
158     AVG(weight_kg) AS AvgWeight
159 FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]
160 GROUP BY Position;
```

100 % 2 0

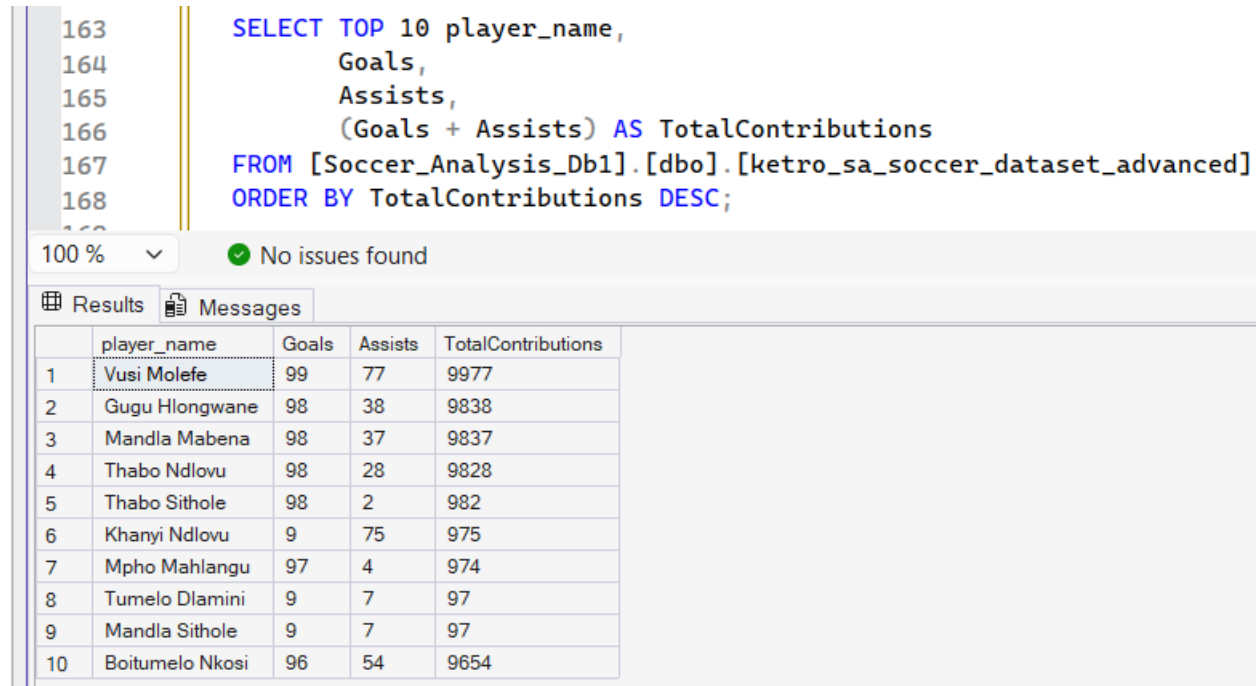
Results Messages

	Position	AvgHeight	AvgWeight
1	Defender	182	78
2	Forward	179	78
3	Goalkeeper	179	78
4	Midfielder	179	76

20. Identify players with the highest combined goals and assists.

Sql:

```
SELECT TOP 10 player_name,  
             Goals,  
             Assists,  
             (Goals + Assists) AS TotalContributions  
FROM [Soccer_Analysis_Db1].[dbo].[ketrosoccerdataset_advanced]  
ORDER BY TotalContributions DESC;
```



100 % No issues found

Results Messages

	player_name	Goals	Assists	TotalContributions
1	Vusi Molefe	99	77	9977
2	Gugu Hlongwane	98	38	9838
3	Mandla Mabena	98	37	9837
4	Thabo Ndlovu	98	28	9828
5	Thabo Sithole	98	2	982
6	Khanyi Ndlovu	9	75	975
7	Mpho Mahlangu	97	4	974
8	Tumelo Dlamini	9	7	97
9	Mandla Sithole	9	7	97
10	Boitumelo Nkosi	96	54	9654