Project_2 ETL

Group_3
Dakota/ Joyce/ Sean/ Trevor



Data Source

 https://fbref.com/en/comps/12/3239/ 2019-2020-La-Liga-Stats



ETL process

• Extract (Read the data from multiple)

player_standard_stats_df

```
In [25]: # Import other cvs file
    csv_file = "sportsref_20_21.csv"
    player_standard_stats_df = pd.read_csv(csv_file, header=1, encoding='maccentraleurope')
    player_standard_stats_df.head()
```

Out[25]:

	Rk	Player	Nation	Pos	Squad	Age	Born	MP	Starts	Min	 хG	npxG	хA	npxG+xA	xG.1	xA.1	xG+xA	npxG.1	npxG+xA.1	Matche
0	1	Sabit Abdulai	gh GHA	MF	Getafe	21	1999	3	0	60	 0.0	0.0	0.0	0.0	0.00	0.00	0.00	0.00	0.00	Matche
1	2	Marcos AcuÉa	ar ARG	DF	Sevilla	28	1991	30	26	2,330	 1.1	1.1	2.8	4.0	0.04	0.11	0.15	0.04	0.15	Matche
2	3	Bobby Adekanye	nl NED	FW,MF	Cádiz	21	1999	3	0	36	 0.2	0.2	0.0	0.2	0.43	0.00	0.43	0.43	0.43	Matche
3	4	Martin Agirregabiria	es ESP	DF,MF	Alavés	24	1996	26	16	1,558	 0.2	0.2	0.9	1.1	0.01	0.05	0.06	0.01	0.06	Matche
4	5	Joseph Aidoo	gh GHA	DF	Celta Vigo	24	1995	25	14	1,289	 0.8	0.8	0.6	1.4	0.06	0.04	0.10	0.06	0.10	Matche

5 rows x 33 columns

league_table_df

In [3]:	<pre>tables = pd.read_html(url) league_table_df = tables[0] league_table_df.head()</pre>	
---------	--	--

Out[3]:

	Rk	Squad	MP	w	D	L	GF	GA	GD	Pts	хG	xGA	xGD	xGD/90	Attendance	Top Team Scorer	Goalkeeper	Notes
0	1	Atlético Madrid	38	26	8	4	67	25	42	86	52.4	32.7	19.6	0.52	NaN	Luis Suárez - 21	Jan Oblak	→ UEFA Champions League via league finish
1	2	Real Madrid	38	25	9	4	67	28	39	84	61.6	36.5	25.2	0.66	NaN	Karim Benzema - 23	Thibaut Courtois	→ UEFA Champions League via league finish
2	3	Barcelona	38	24	7	7	85	38	47	79	78.9	39.6	39.3	1.03	NaN	Lionel Messi - 30	Marc-André ter Stegen	→ UEFA Champions League via league finish
3	4	Sevilla	38	24	5	9	53	33	20	77	50.9	35.0	15.9	0.42	NaN	Youssef En- Nesyri - 18	Yassine Bounou	→ UEFA Champions League via league finish
4	5	Real Sociedad	38	17	11	10	59	38	21	62	60.4	36.6	23.8	0.63	NaN	Alexander Isak - 17	Álex Remiro	→ UEFA Europa League via league finish

ETL process

• Transform (Clean and structure data)

Source DB	Source Table	Source colum n	Source Datatyp e	Target DB	Target Table	Target colum n	Target Datatyp e	Business Rule
Sportsref_20_21.csv	player_standard_st ats	Player	string	La_liga_db	player_standard _stats	Player	string	encoding='maccentrale urope'
Sportsref_20_21.csv	player_standard_st ats	Nation	string	La_liga_db	player_standard _stats	Nation	string	Split to get last 3 nation abbreviated
Sportsref_20_21.csv	player_standard_st ats			La_liga_db	player_standard _stats			Dropped columns
Sportsref_20_21.csv	player_standard_st ats	Min	string	La_liga_db	player_standard _stats	Min	string	Replace "," with""
https://fbref.com/en/comps/12/107 31/2020-2021-La-Liga-Stats	League_table_df			La_liga_db	League_table_ta ble			Dropped columns

```
'Pts']]
new_league_table_df.head()
```

Out[5]:

	Rk	Squad	MP	w	D	L	GF	GA	GD	Pts
0	1	Atlético Madrid	38	26	8	4	67	25	42	86
1	2	Real Madrid	38	25	9	4	67	28	39	84
2	3	Barcelona	38	24	7	7	85	38	47	79
3	4	Sevilla	38	24	5	9	53	33	20	77
4	5	Real Sociedad	38	17	11	10	59	38	21	62

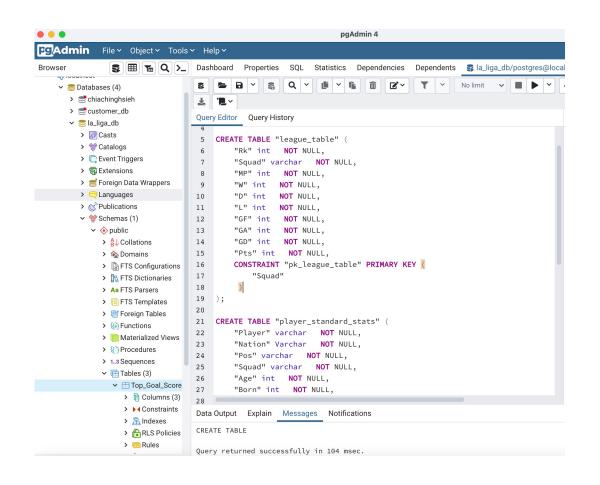
new_league_table_df

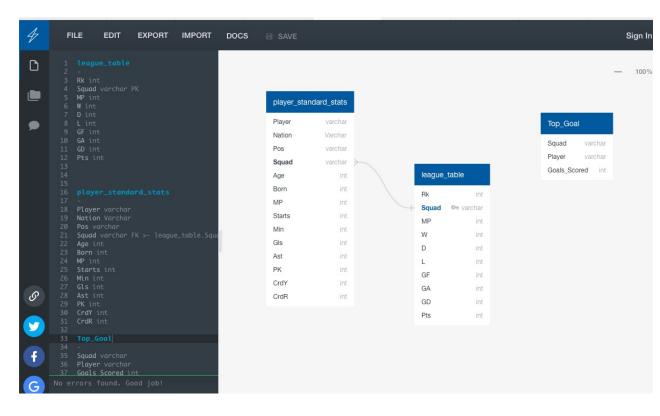
new_player_df

Out[9]:

	Player	Nation	Pos	Squad	Age	Born	MP	Starts	Min	Gls	Ast	PK	CrdY	CrdR
0	Sabit Abdulai	GHA	MF	Getafe	21	1999	3	0	60	0	0	0	1	0
1	Marcos AcuÉa	ARG	DF	Sevilla	28	1991	30	26	2330	1	2	0	5	0
2	Bobby Adekanye	NED	FW,MF	Cádiz	21	1999	3	0	36	0	0	0	0	0
3	Martin Agirregabiria	ESP	DF,MF	Alavés	24	1996	26	16	1558	0	1	0	1	0
4	Joseph Aidoo	GHA	DF	Celta Vigo	24	1995	25	14	1289	0	0	0	3	0

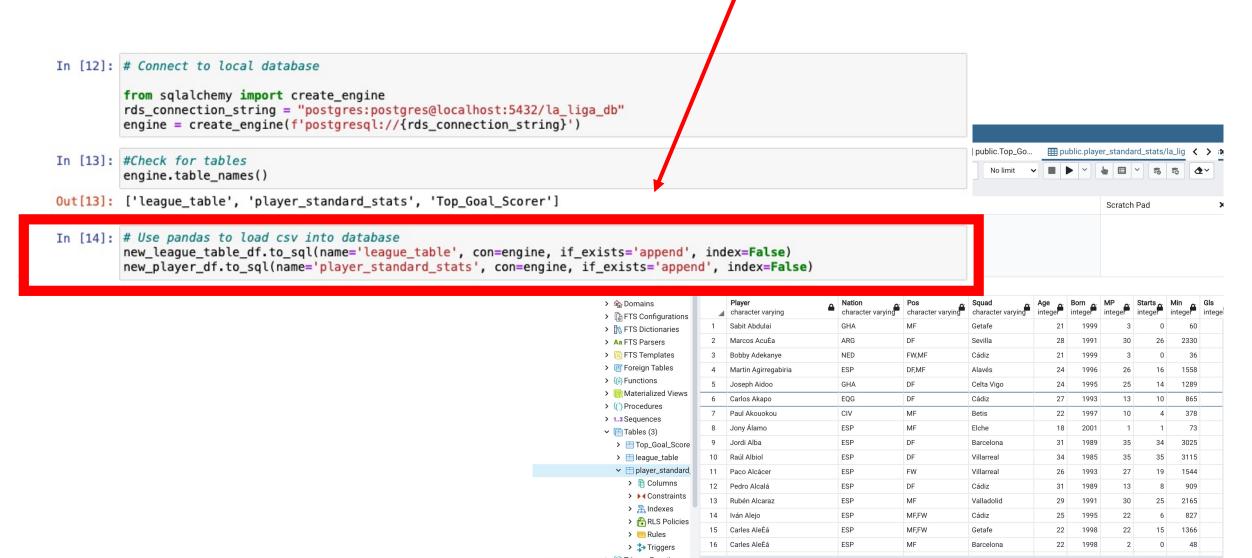
ERD & CREATE DataBase





ETL process

Load (Write the data into database for storage)



In [15]: # #querying Join the two table on Squad col
Join_df=pd.read_sql_query('select * from league_table as lt join player_standard_stats as ps on lt."Squad"=ps."Squad
Join_df.head()

Out[15]:

	Rk	Squad	MP	w	D	L	GF	GA	GD	Pts	 Age	Born	MP	Starts	Min	Gls	Ast	PK	CrdY	CrdR
0	15	Getafe	38	9	11	18	28	43	-15	38	 21	1999	3	0	60	0	0	0	1	0
1	4	Sevilla	38	24	5	9	53	33	20	77	 28	1991	30	26	2330	1	2	0	5	0
2	12	Cádiz	38	11	11	16	36	58	-22	44	 21	1999	3	0	36	0	0	0	0	0
3	16	Alavés	38	9	11	18	36	57	-21	38	 24	1996	26	16	1558	0	1	0	1	0
4	8	Celta Vigo	38	14	11	13	55	57	-2	53	 24	1995	25	14	1289	0	0	0	3	0

5 rows x 24 columns

Join Two Table

 Join_df=pd.read_sql_query('select * from league_table as lt join player_standard_stats as ps on lt."Squad"=ps."Squad"', con=engine)

Target Table

Sour DB	ce Sourc Table		Target DB		Bus Rule	iness/Target e
La_li	ga_db playeı	r_standard_stats &League_table	La_liga_db	Top_Goal_Scorer	1. 2.	Write to DataBase Groupby: "Squad" and find max goals scored then sort by max

```
df2 =Top_player.groupby(['Squad','Player'])['Gls'].agg(['max'])
df2 = df2.sort_values('max', ascending=False)
reset_df=df2.reset_index()
reset_df.drop_duplicates(['Squad'], keep='first')
TGS_df=reset_df.rename(columns={'max':'Goals Scored'})
```

#insert new datafrmae into database new table

TGS_df.to_sql(name='Top_Goal_Scorer', con=engine, if_exists='append', index=False)

Out[21]:

11 Eibar Kike 12 12 Ceita Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	Goals Scored	Player	Squad	
2 Real Madrid Karim Benzema 23 3 Sevilla Youssef En-Nesyri 18 4 Real Sociedad Alexander Isak 17 5 Celta Vigo Iago Aspas 14 6 Huesca Rafa Mir 13 7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	30	Lionel Messi	Barcelona	0
3 Sevilla Youssef En-Nesyri 18 4 Real Sociedad Alexander Isak 17 5 Celta Vigo Iago Aspas 14 6 Huesca Rafa Mir 13 7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	23	Gerard Moreno	Villarreal	1
4 Real Sociedad Alexander Isak 17 5 Celta Vigo Iago Aspas 14 6 Huesca Rafa Mir 13 7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	23	Karim Benzema	Real Madrid	2
5 Celta Vigo Iago Aspas 14 6 Huesca Rafa Mir 13 7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	18	Youssef En-Nesyri	Sevilla	3
6 Huesca Rafa Mir 13 7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	17	Alexander Isak	Real Sociedad	4
7 Barcelona Antoine Griezmann 13 8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	14	lago Aspas	Celta Vigo	5
8 Levante José Luis Morales 13 9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	13	Rafa Mir	Huesca	6
9 Levante Roger Martí 12 10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	13	Antoine Griezmann	Barcelona	7
10 Atlético Madrid Marcos Llorente 12 11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	13	José Luis Morales	Levante	8
11 Eibar Kike 12 12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	12	Roger Martí	Levante	9
12 Celta Vigo Santi Mina 12 13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	12	Marcos Llorente	Atlético Madrid	10
13 Alavés Joselu 11 14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	12	Kike	Eibar	11
14 Real Sociedad Mikel Oyarzabal 11 15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	12	Santi Mina	Celta Vigo	12
15 Osasuna Ante Budimir 11 16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	11	Joselu	Alavés	13
16 Betis Borja Iglesias 11 17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	11	Mikel Oyarzabal	Real Sociedad	14
17 Atlético Madrid Ángel Correa 9 18 Celta Vigo Brais Méndez 9	11	Ante Budimir	Osasuna	15
18 Celta Vigo Brais Méndez 9	11	Borja Iglesias	Betis	16
	9	Ángel Correa	Atlético Madrid	17
19 Cádiz Álvaro Negredo 8	9	Brais Méndez	Celta Vigo	18
	8	Álvaro Negredo	Cádiz	19

Thank you

