

Proyecto No. 1

Desarrollo y consulta de Base de Datos

Etapas 1:

1. Descargar los archivos CSV y verificar qué información se incluye en cada uno.

Draft_Combine.csv	Ene. 26, 2022 5:15 p. m.	389 KB	Documento CSV
Draft.csv	Ene. 26, 2022 5:14 p. m.	1.1 MB	Documento CSV
Game_Inactive_Players.csv	Ene. 26, 2022 5:19 p. m.	6.8 MB	Documento CSV
Game_Officials.csv	Ene. 26, 2022 5:19 p. m.	2.2 MB	Documento CSV
Game.csv	Ene. 26, 2022 5:16 p. m.	42.2 MB	Documento CSV
News_Missing.csv	Ene. 26, 2022 5:21 p. m.	22 KB	Documento CSV
News.csv	Ene. 26, 2022 5:20 p. m.	772.4 MB	Documento CSV
Player_Attributes.csv	Ene. 26, 2022 5:23 p. m.	1.2 MB	Documento CSV
Player_Bios.csv	Ene. 26, 2022 5:23 p. m.	11.5 MB	Documento CSV
Player_Salary.csv	Ene. 26, 2022 5:23 p. m.	123 KB	Documento CSV
Player.csv	Ene. 26, 2022 5:22 p. m.	171 KB	Documento CSV
Team_Attributes.csv	Ene. 26, 2022 5:25 p. m.	7 KB	Documento CSV
Team_History.csv	Ene. 26, 2022 5:25 p. m.	2 KB	Documento CSV
Team_Salary.csv	Ene. 26, 2022 5:25 p. m.	4 KB	Documento CSV
Team.csv	Ene. 26, 2022 5:24 p. m.	2 KB	Documento CSV

2. Crear la base de datos y las tablas en donde almacenará los datos.

De las tablas elegidas anteriormente, se seleccionan solamente las que ayudarán para la investigación. Entre estas se incluyen: Draft, Game_officials, Game, Player_Attributes, Player_salary, Player, Team_attributes, Team_salary, Team.

3. Desarrollar un script de Python (o lenguaje a su elección) que pueda leer la información de los archivos CSV, conectarse a su base de datos y alimentar la información en las tablas creadas.

*El script de Python está en el zip, al igual que el diagrama Entidad Relación.

Link al repositorio de GitHub: <https://github.com/sebastianf232/Proyecto-Bases-de-Datos>

Etapas 2:

1. ¿Quién es el jugador activo más alto? ¿Y el más bajo?

Sebastián Franco -21484
Mario Puente -21290

The first screenshot shows a SQL query in pgAdmin 4 that selects player attributes where height is not null and roster status is 'Active', ordered by height in ascending order. The result shows two players: Tremont Waters and Jared Harper, both with a height of 70.

```
1 SELECT "FIRST_NAME", "LAST_NAME", "HEIGHT"
2 FROM "Player_Attributes"
3 WHERE "HEIGHT" Is Not NULL and "ROSTERSTATUS" = 'Active'
4 GROUP BY "FIRST_NAME", "LAST_NAME", "HEIGHT"
5 ORDER BY Max("HEIGHT") ASC
6
7 Select *
8 From "Player_Attributes"
9
```

	FIRST_NAME text	LAST_NAME text	HEIGHT double precision
1	Tremont	Waters	70
2	Jared	Harper	70

The second screenshot shows a similar query but ordered by height in descending order. The result shows one player: Tacko Fall, with a height of 89.

```
1 SELECT "FIRST_NAME", "LAST_NAME", "HEIGHT"
2 FROM "Player_Attributes"
3 WHERE "HEIGHT" Is Not NULL and "ROSTERSTATUS" = 'Active'
4 GROUP BY "FIRST_NAME", "LAST_NAME", "HEIGHT"
5 ORDER BY Max("HEIGHT") Desc
6
7 Select *
8 From "Player_Attributes"
9
```

	FIRST_NAME text	LAST_NAME text	HEIGHT double precision
1	Tacko	Fall	89

El Jugador más alto es Tacko Fall, y en el más bajo es un empate entre Tremont Waters y Jared Harper.

2. ¿Cuál fue el promedio de puntos anotados y recibidos por cada equipo en cada una de las temporadas relevantes?

Queries Generales:

Query que indica promedio de puntos anotados, se cambia la temporada para averiguar por cada temporada.

Sebastián Franco -21484
Mario Puente -21290

The screenshot shows the pgAdmin 4 interface with a SQL query in the 'Query' tab. The query is as follows:

```
88
89 CREATE TABLE Pregunta2 As
90 SELECT "TEAM_NAME_HOME", AVG("PTS_HOME") pts
91 FROM "Game" g
92 WHERE g."SEASON_ID" = 22020
93 GROUP BY "TEAM_NAME_HOME"
94 UNION
95 SELECT "TEAM_NAME_AWAY", AVG("PTS_AWAY") pts
96 FROM "Game" g
97 WHERE g."SEASON_ID" = 22020
98 GROUP BY "TEAM_NAME_AWAY"
99
100 SELECT "TEAM_NAME_HOME", CAST(AVG(pts) As Decimal(10,2)) promedio_ anotados
101 From Pregunta2
102 GROUP BY "TEAM_NAME_HOME"
103 Order by promedio_ anotados desc
```

The 'Data output' tab shows the following table:

	TEAM_NAME_HOME text	promedio_ anotados numeric (10,2)
1	Milwaukee Bucks	120.13
2	Brooklyn Nets	117.56
3	Washington Wizards	116.64
4	New Orleans Pelicans	114.53
5	Indiana Pacers	113.96

Query que indica promedio de puntos recibidos, se cambia la temporada.

The screenshot shows the pgAdmin 4 interface with a SQL query in the 'Query' tab. The query is as follows:

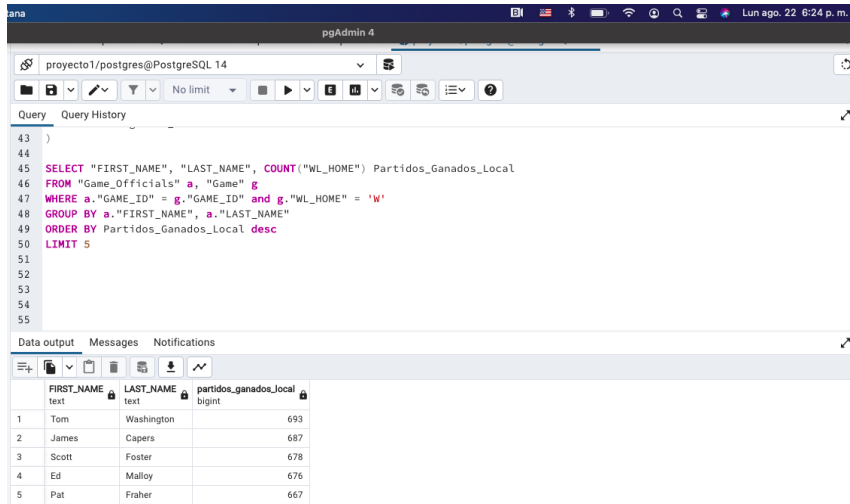
```
106 CREATE TABLE Pregunta22 As
107 SELECT "TEAM_NAME_HOME", AVG("PTS_AWAY") pts
108 FROM "Game" g
109 WHERE g."SEASON_ID" = 22020
110 GROUP BY "TEAM_NAME_HOME"
111 UNION
112 SELECT "TEAM_NAME_AWAY", AVG("PTS_HOME") pts
113 FROM "Game" g
114 WHERE g."SEASON_ID" = 22020
115 GROUP BY "TEAM_NAME_AWAY"
116
117 SELECT "TEAM_NAME_HOME", CAST(AVG(pts) As Decimal(10,2)) recibidos
118 From Pregunta22
119 GROUP BY "TEAM_NAME_HOME"
120 Order by recibidos desc
121
```

The 'Data output' tab shows the following table:

	TEAM_NAME_HOME text	recibidos numeric (10,2)
1	Washington Wizards	118.47
2	Minnesota Timberwolv...	116.82
3	New Orleans Pelicans	115.01
4	Oklahoma City Thunder	114.46
5	Milwaukee Bucks	114.15

Sebastián Franco -21484
Mario Puente -21290

3. Muestre el top 5 de árbitros en cuyos juegos pitados el equipo visitante pierde.



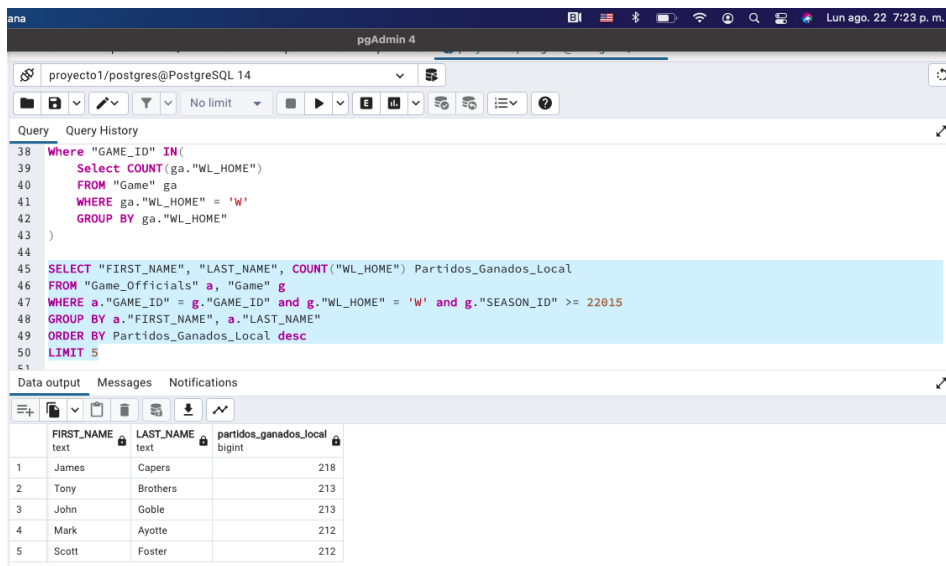
The screenshot shows the pgAdmin 4 interface. The SQL query in the editor is:

```
43 )
44
45 SELECT "FIRST_NAME", "LAST_NAME", COUNT("WL_HOME") Partidos_Ganados_Local
46 FROM "Game_Officials" a, "Game" g
47 WHERE a."GAME_ID" = g."GAME_ID" and g."WL_HOME" = 'W'
48 GROUP BY a."FIRST_NAME", a."LAST_NAME"
49 ORDER BY Partidos_Ganados_Local desc
50 LIMIT 5
51
52
53
54
55
```

The results are displayed in a table with the following data:

	FIRST_NAME	LAST_NAME	partidos_ganados_local
1	Tom	Washington	693
2	James	Capers	687
3	Scott	Foster	678
4	Ed	Malloy	676
5	Pat	Fraher	667

Si se toma en cuenta solamente desde la temporada 2015/16



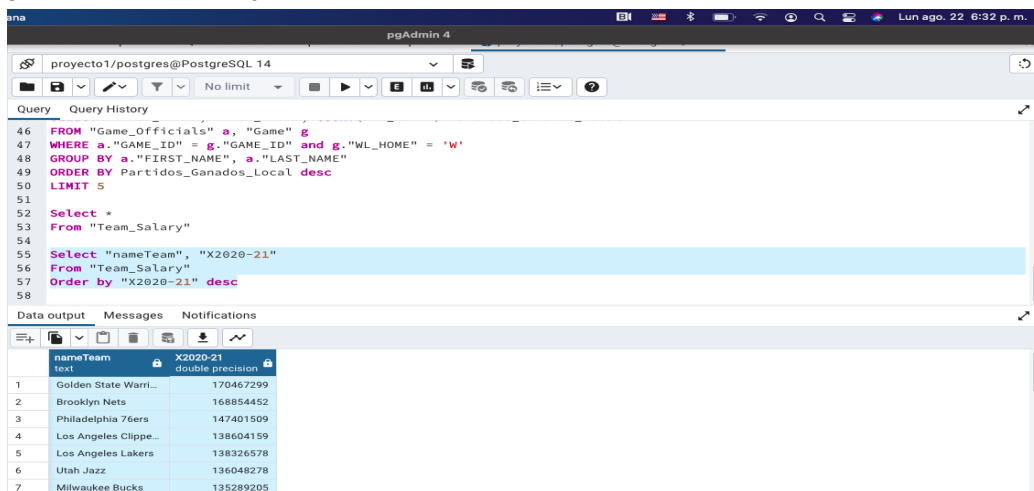
The screenshot shows the pgAdmin 4 interface. The SQL query in the editor is:

```
38 Where "GAME_ID" IN (
39     Select COUNT(ga."WL_HOME")
40 FROM "Game" ga
41 WHERE ga."WL_HOME" = 'W'
42 GROUP BY ga."WL_HOME"
43 )
44
45 SELECT "FIRST_NAME", "LAST_NAME", COUNT("WL_HOME") Partidos_Ganados_Local
46 FROM "Game_Officials" a, "Game" g
47 WHERE a."GAME_ID" = g."GAME_ID" and g."WL_HOME" = 'W' and g."SEASON_ID" >= 22015
48 GROUP BY a."FIRST_NAME", a."LAST_NAME"
49 ORDER BY Partidos_Ganados_Local desc
50 LIMIT 5
51
```

The results are displayed in a table with the following data:

	FIRST_NAME	LAST_NAME	partidos_ganados_local
1	James	Capers	218
2	Tony	Brothers	213
3	John	Goble	213
4	Mark	Ayotte	212
5	Scott	Foster	212

4. ¿Qué equipo maneja los salarios más altos actualmente?



The screenshot shows the pgAdmin 4 interface. The SQL query in the editor is:

```
46 FROM "Game_Officials" a, "Game" g
47 WHERE a."GAME_ID" = g."GAME_ID" and g."WL_HOME" = 'W'
48 GROUP BY a."FIRST_NAME", a."LAST_NAME"
49 ORDER BY Partidos_Ganados_Local desc
50 LIMIT 5
51
52 Select *
53 From "Team_Salary"
54
55 Select "nameTeam", "X2020-21"
56 From "Team_Salary"
57 Order by "X2020-21" desc
58
```

The results are displayed in a table with the following data:







	nameTeam	X2020-21
1	Golden State Warri...	170467299
2	Brooklyn Nets	168854452
3	Philadelphia 76ers	147401509
4	Los Angeles Clippe...	138604159
5	Los Angeles Lakers	138326578
6	Utah Jazz	136048278
7	Milwaukee Bucks	135289205

5. ¿Cuál fue la temporada con más partidos en la historia de la NBA? ¿Cuál fue la temporada que más se ha prolongado en fechas?

Sebastián Franco -21484
Mario Puente -21290







```
SELECT COUNT("GAME_ID") as cuenta, "SEASON"  
FROM "Game"  
GROUP BY "SEASON"  
ORDER BY cuenta desc  
LIMIT 1
```

output Messages Notifications

					
cuenta	bigint	SEASON	bigint		
1286		2013			

```
SELECT MAX("GAME_DATE")-MIN("GAME_DATE") AS DateDifference, "SEASON"  
FROM "Game"  
GROUP BY "SEASON"  
ORDER BY DateDifference desc  
LIMIT 1
```

output Messages Notifications

					
datedifference	interval	SEASON	bigint		
297 days		2019			

6. ¿Cuál es el equipo que más diferencia de puntos a favor obtuvo en promedio por partido para la temporada 2017? ¿Y para la temporada 2018?

Sebastián Franco -21484
Mario Puente -21290

```
SELECT AVG(promedio) as prom, nombre_equipo
FROM (
  SELECT SUM("PLUS_MINUS_AWAY") AS promedio, "TEAM_NAME_AWAY" AS nombre_equipo
  FROM "Game"
  WHERE "SEASON" = 2017
  GROUP BY "TEAM_NAME_AWAY"
  UNION
  SELECT SUM("PLUS_MINUS_HOME"), "TEAM_NAME_HOME"
  FROM "Game"
  WHERE "SEASON" = 2017
  GROUP BY "TEAM_NAME_HOME"
) AS subq
GROUP BY nombre_equipo
ORDER BY prom desc
LIMIT 1
```

output Messages Notifications

prom	numeric		nombre_equipo	text	
347.5000000	Houston Rockets				

```
SELECT AVG(promedio) as prom, nombre_equipo
FROM (
  SELECT SUM("PLUS_MINUS_AWAY") AS promedio, "TEAM_NAME_AWAY" AS nombre_equipo
  FROM "Game"
  WHERE "SEASON" = 2018
  GROUP BY "TEAM_NAME_AWAY"
  UNION
  SELECT SUM("PLUS_MINUS_HOME"), "TEAM_NAME_HOME"
  FROM "Game"
  WHERE "SEASON" = 2018
  GROUP BY "TEAM_NAME_HOME"
) AS subq
GROUP BY nombre_equipo
ORDER BY prom desc
LIMIT 1
```

output Messages Notifications

prom	numeric		nombre_equipo	text	
363.50000000000000000000	Milwaukee Bucks				

7. ¿Quién es el jugador más valioso del draft del 2018 hoy en día?

Sebastián Franco -21484
Mario Puente -21290

```
SELECT "DISPLAY_FIRST_LAST" AS nombre, "DRAFT_YEAR", "PIE"  
FROM "Player_Attributes"  
WHERE "DRAFT_YEAR" = '2018' AND "PIE" IS NOT NULL  
ORDER BY "PIE" desc  
LIMIT 1
```

output Messages Notifications

nombre	DRAFT_YEAR	PIE			
text	text	double precision			
Luka Doncic	2018	0.187			

8. Calcule el top 5 de los estados que más salarios pagaron durante las temporadas 2020/2021 y 2021/2022.

```
SELECT "state", SUM("X2020-21") as salario  
FROM "Team_Salary" JOIN "Team" ON "Team_Salary"."slugTeam" = "Team"."abbreviation"  
GROUP BY "state"  
ORDER BY salario DESC  
LIMIT 5
```

output Messages Notifications

state	salario				
text	double precision				
California	553350641				
Texas	388247885				
New York	266468144				
Florida	255954783				
Pennsylv...	147401509				

Sebastián Franco -21484
Mario Puente -21290

```
SELECT "state", SUM("X2021-22") as salario
FROM "Team_Salary" JOIN "Team" ON "Team_Salary"."slugTeam" = "Team"."abbreviation"
GROUP BY "state"
ORDER BY salario DESC
LIMIT 5
```

output Messages Notifications



state	salario
text	double precision
California	524437877
Texas	235941792
New York	213999731
Florida	179376763
Wisconsin	143531830

Etapas 3:

1. De los jugadores actuales, ¿Quién tiene el mejor PIE de la NBA?

```
223 Select "FIRST_NAME", "LAST_NAME", "TEAM_NAME", "PIE", "PTS", "AST", "REB"
224 From "Player_Attributes"
225 Where "ROSTERSTATUS" = 'Active'
226 Group By "FIRST_NAME", "LAST_NAME", "TEAM_NAME", "PIE", "PTS", "AST", "REB"
227 Order By "PIE" desc
228 LIMIT 15
```

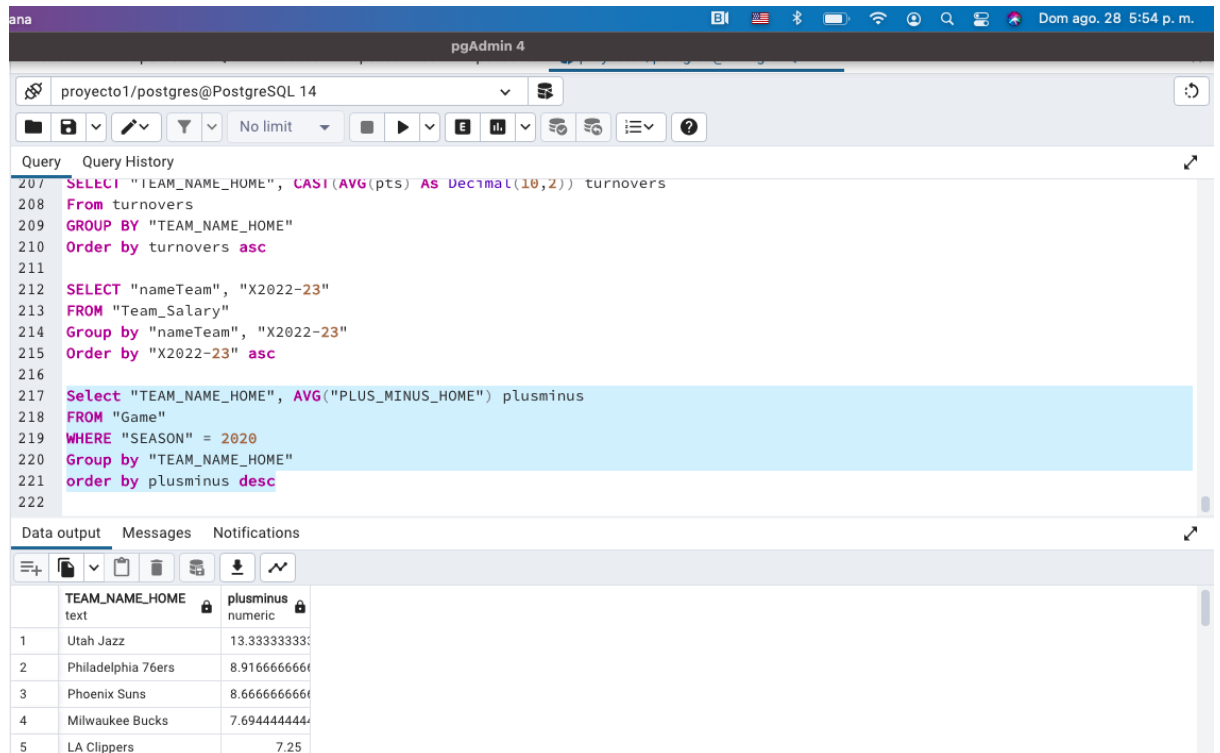
Data output Messages Notifications



	FIRST_NAME	LAST_NAME	TEAM_NAME	PIE	PTS	AST	REB
	text	text	text	double precision	double precision	double precision	double precision
1	Joel	Embiid	76ers	0.211	29.9	3.3	11.5
2	Nikola	Jokic	Nuggets	0.205	27.1	8.6	11.3
3	Giannis	Antetokounm...	Bucks	0.196	29	6.4	11.7
4	LeBron	James	Lakers	0.194	25.4	7.9	7.9
5	Jimmy	Butler	Heat	0.19	21.3	7.3	7.9

2. ¿Qué equipo tiene el mejor +/- en casa de las temporadas de interés?

Sebastián Franco -21484
Mario Puente -21290

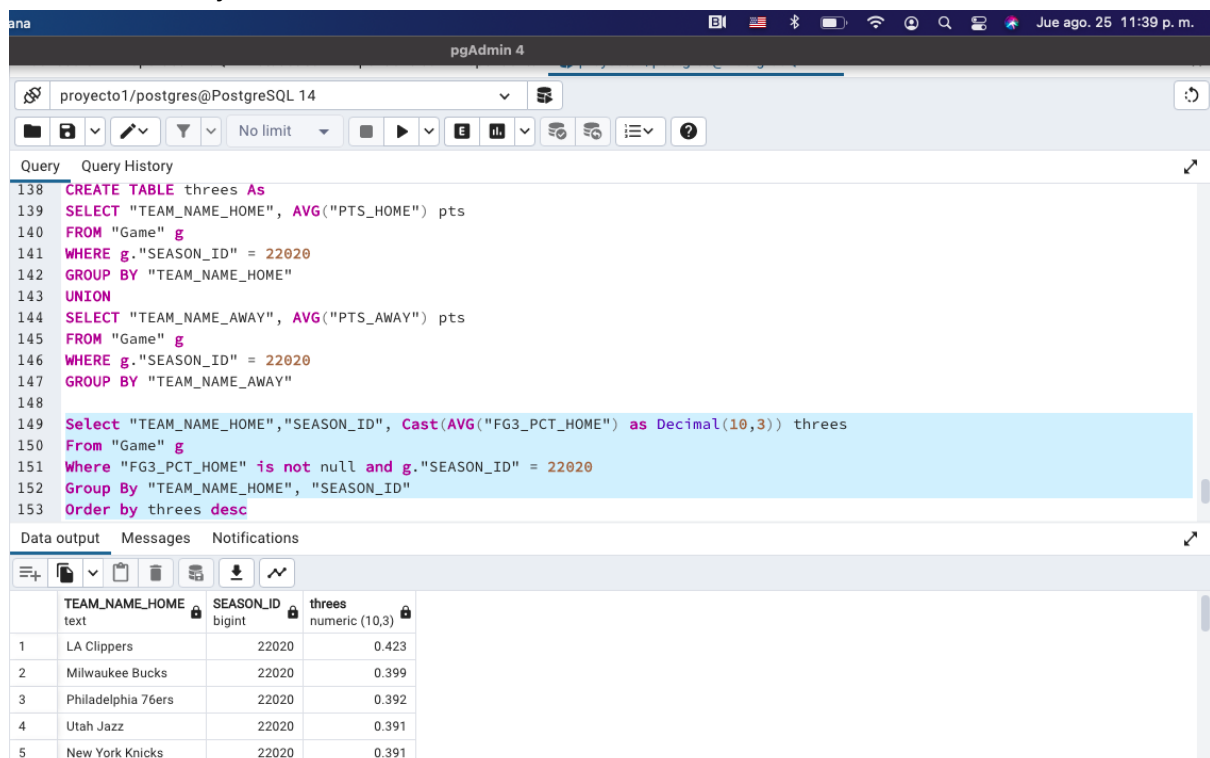


The screenshot shows the pgAdmin 4 interface with the following SQL queries and results:

```
207 SELECT "TEAM_NAME_HOME", CAST(AVG(pts) AS Decimal(10,2)) turnovers
208 FROM turnovers
209 GROUP BY "TEAM_NAME_HOME"
210 Order by turnovers asc
211
212 SELECT "nameTeam", "X2022-23"
213 FROM "Team_Salary"
214 Group by "nameTeam", "X2022-23"
215 Order by "X2022-23" asc
216
217 Select "TEAM_NAME_HOME", AVG("PLUS_MINUS_HOME") plusminus
218 FROM "Game"
219 WHERE "SEASON" = 2020
220 Group by "TEAM_NAME_HOME"
221 order by plusminus desc
222
```

	TEAM_NAME_HOME	plusminus
1	Utah Jazz	13.33333333
2	Philadelphia 76ers	8.916666666
3	Phoenix Suns	8.666666666
4	Milwaukee Bucks	7.694444444
5	LA Clippers	7.25

3. ¿Qué equipo tira mejor desde la línea de 3 puntos en la última temporada de interés? ¿Desde la línea de tiros libres?
Porcentaje de 3 en casa:



The screenshot shows the pgAdmin 4 interface with the following SQL queries and results:

```
138 CREATE TABLE threes As
139 SELECT "TEAM_NAME_HOME", AVG("PTS_HOME") pts
140 FROM "Game" g
141 WHERE g."SEASON_ID" = 2020
142 GROUP BY "TEAM_NAME_HOME"
143 UNION
144 SELECT "TEAM_NAME_AWAY", AVG("PTS_AWAY") pts
145 FROM "Game" g
146 WHERE g."SEASON_ID" = 2020
147 GROUP BY "TEAM_NAME_AWAY"
148
149 Select "TEAM_NAME_HOME", "SEASON_ID", Cast(AVG("FG3_PCT_HOME") as Decimal(10,3)) threes
150 From "Game" g
151 Where "FG3_PCT_HOME" is not null and g."SEASON_ID" = 2020
152 Group By "TEAM_NAME_HOME", "SEASON_ID"
153 Order by threes desc
```

	TEAM_NAME_HOME	SEASON_ID	threes
1	LA Clippers	22020	0.423
2	Milwaukee Bucks	22020	0.399
3	Philadelphia 76ers	22020	0.392
4	Utah Jazz	22020	0.391
5	New York Knicks	22020	0.391

Sebastián Franco -21484
Mario Puente -21290

Porcentaje de 3 de visita:

The screenshot shows the pgAdmin 4 interface with the following SQL queries and results:

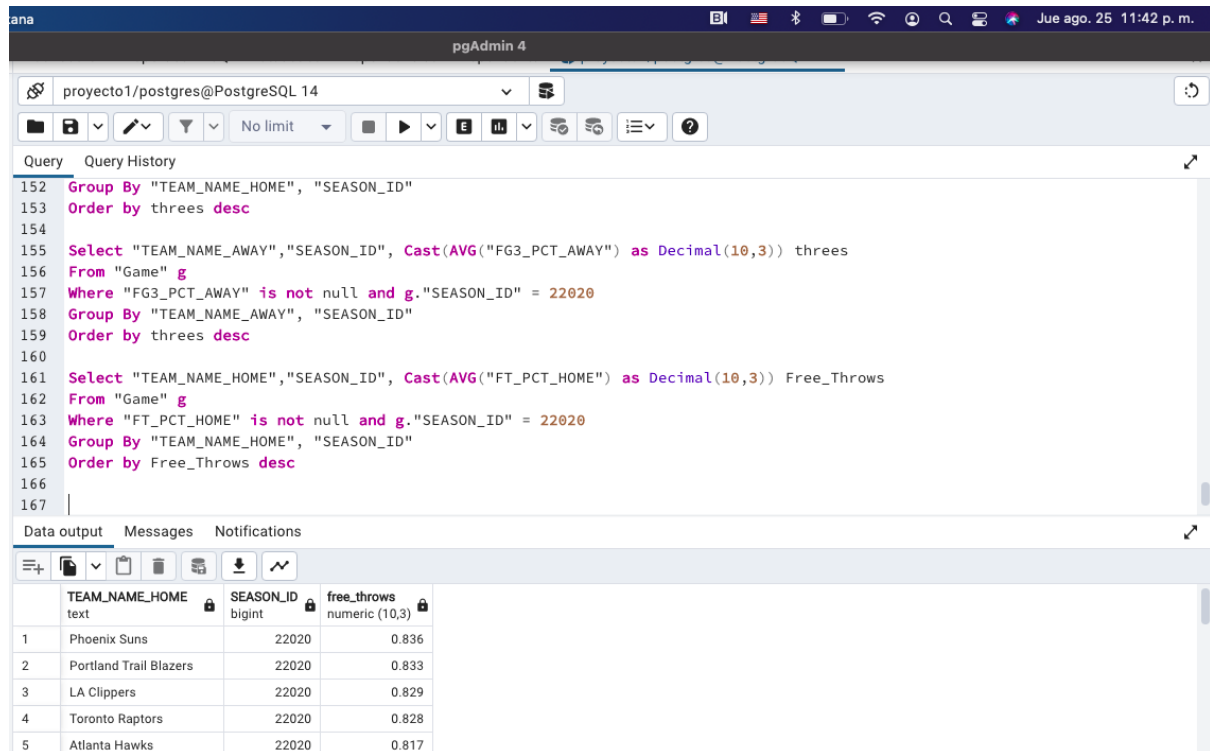
```
144 SELECT "TEAM_NAME_AWAY", AVG("PTS_AWAY") pts
145 FROM "Game" g
146 WHERE g."SEASON_ID" = 22020
147 GROUP BY "TEAM_NAME_AWAY"
148
149 Select "TEAM_NAME_HOME", "SEASON_ID", Cast(AVG("FG3_PCT_HOME") as Decimal(10,3)) threes
150 From "Game" g
151 Where "FG3_PCT_HOME" is not null and g."SEASON_ID" = 22020
152 Group By "TEAM_NAME_HOME", "SEASON_ID"
153 Order by threes desc
154
155 Select "TEAM_NAME_AWAY", "SEASON_ID", Cast(AVG("FG3_PCT_AWAY") as Decimal(10,3)) threes
156 From "Game" g
157 Where "FG3_PCT_AWAY" is not null and g."SEASON_ID" = 22020
158 Group By "TEAM_NAME_AWAY", "SEASON_ID"
159 Order by threes desc
```

The results are displayed in a table with the following columns: TEAM_NAME_AWAY, SEASON_ID, and threes. The data is as follows:

	TEAM_NAME_AWAY	SEASON_ID	threes
1	LA Clippers	22020	0.4
2	Portland Trail Blazers	22020	0.397
3	New York Knicks	22020	0.392
4	Brooklyn Nets	22020	0.387
5	Indiana Pacers	22020	0.382

Porcentaje de FT en casa:

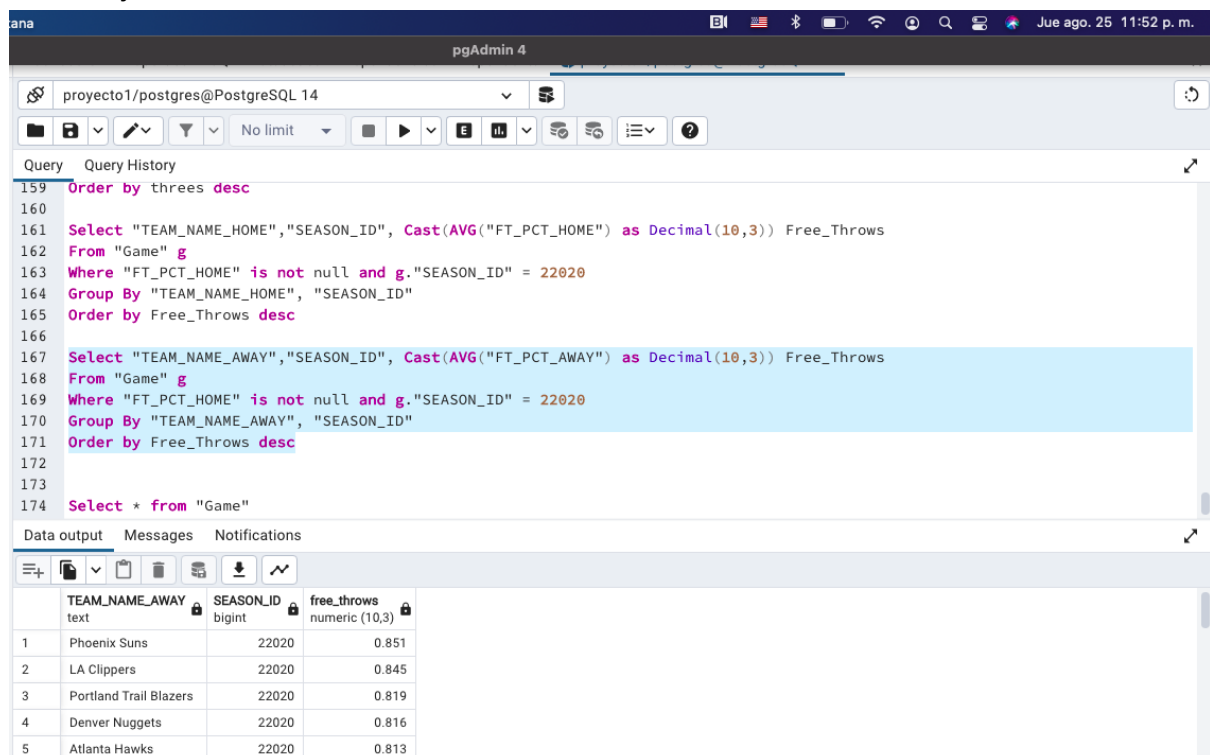
Sebastián Franco -21484
Mario Puente -21290



The screenshot shows the pgAdmin 4 interface with a SQL query executed. The query calculates the average free throw percentage for home teams in season 22020, ordered by the percentage in descending order. The results table shows the following data:

	TEAM_NAME_HOME	SEASON_ID	free_throws
1	Phoenix Suns	22020	0.836
2	Portland Trail Blazers	22020	0.833
3	LA Clippers	22020	0.829
4	Toronto Raptors	22020	0.828
5	Atlanta Hawks	22020	0.817

Porcentaje de Ft de visita:

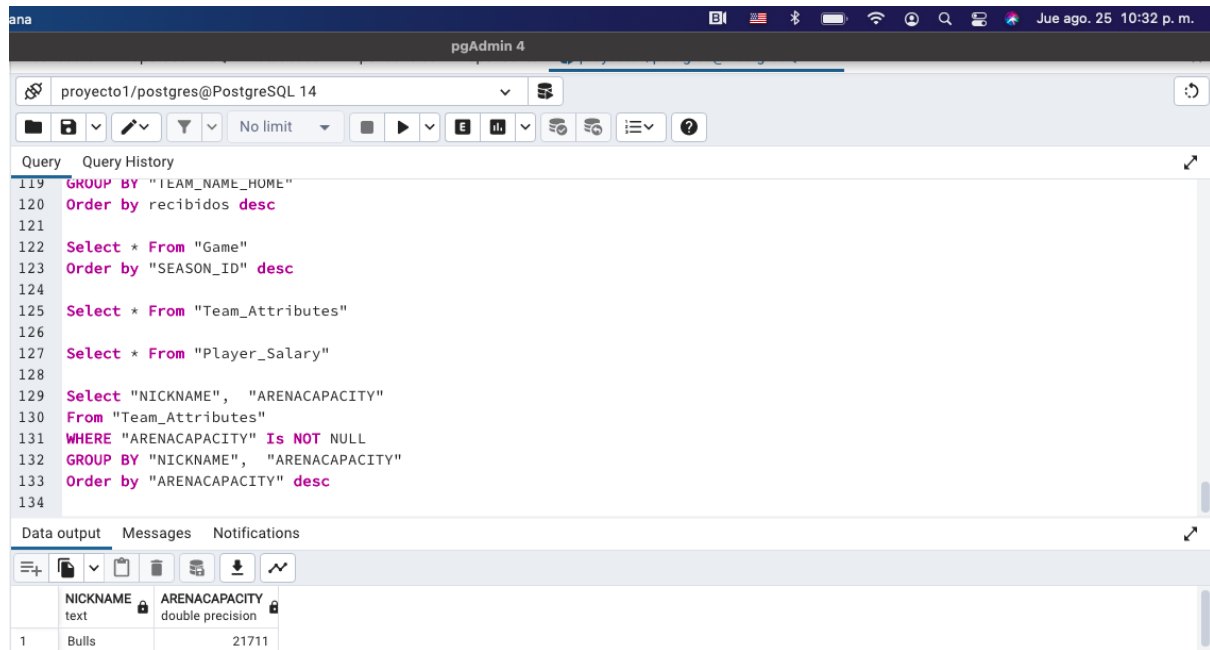


The screenshot shows the pgAdmin 4 interface with a SQL query executed. The query calculates the average free throw percentage for away teams in season 22020, ordered by the percentage in descending order. The results table shows the following data:

	TEAM_NAME_AWAY	SEASON_ID	free_throws
1	Phoenix Suns	22020	0.851
2	LA Clippers	22020	0.845
3	Portland Trail Blazers	22020	0.819
4	Denver Nuggets	22020	0.816
5	Atlanta Hawks	22020	0.813

4. ¿Qué equipo tiene la mayor capacidad en su estadio?

Sebastián Franco -21484
Mario Puente -21290



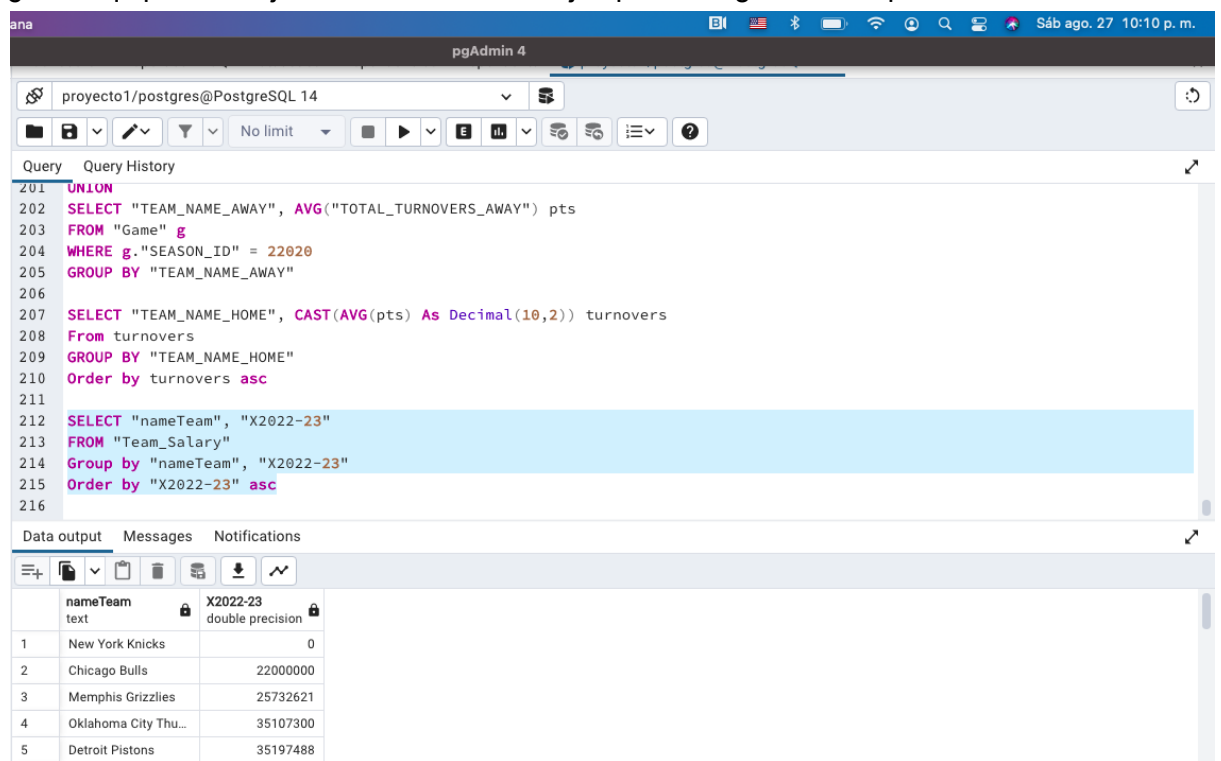
The screenshot shows the pgAdmin 4 interface. The query editor contains the following SQL code:

```
119 GROUP BY "TEAM_NAME_HOME"
120 Order by recibidos desc
121
122 Select * From "Game"
123 Order by "SEASON_ID" desc
124
125 Select * From "Team_Attributes"
126
127 Select * From "Player_Salary"
128
129 Select "NICKNAME", "ARENACAPACITY"
130 From "Team_Attributes"
131 WHERE "ARENACAPACITY" Is NOT NULL
132 GROUP BY "NICKNAME", "ARENACAPACITY"
133 Order by "ARENACAPACITY" desc
134
```

The results pane shows a table with two columns: NICKNAME (text) and ARENACAPACITY (double precision). The data is as follows:

	NICKNAME	ARENACAPACITY
1	Bulls	21711

5. ¿Qué equipos manejan los salarios más bajos para la siguiente temporada?



The screenshot shows the pgAdmin 4 interface. The query editor contains the following SQL code:

```
201 UNION
202 SELECT "TEAM_NAME_AWAY", AVG("TOTAL_TURNOVERS_AWAY") pts
203 FROM "Game" g
204 WHERE g."SEASON_ID" = 22020
205 GROUP BY "TEAM_NAME_AWAY"
206
207 SELECT "TEAM_NAME_HOME", CAST(AVG(pts) As Decimal(10,2)) turnovers
208 From turnovers
209 GROUP BY "TEAM_NAME_HOME"
210 Order by turnovers asc
211
212 SELECT "nameTeam", "X2022-23"
213 FROM "Team_Salary"
214 Group by "nameTeam", "X2022-23"
215 Order by "X2022-23" asc
216
```

The results pane shows a table with two columns: nameTeam (text) and X2022-23 (double precision). The data is as follows:

	nameTeam	X2022-23
1	New York Knicks	0
2	Chicago Bulls	22000000
3	Memphis Grizzlies	25732621
4	Oklahoma City Thu...	35107300
5	Detroit Pistons	35197488

6. ¿Qué equipo pierde menos la pelota por partido?

Sebastián Franco -21484
Mario Puente -21290

pgAdmin 4

proyecto1/postgres@PostgreSQL 14

Query Query History

```
196 CREATE TABLE turnovers As
197 SELECT "TEAM_NAME_HOME", AVG("TOTAL_TURNOVERS_HOME") pts
198 FROM "Game" g
199 WHERE g."SEASON_ID" = 22020
200 GROUP BY "TEAM_NAME_HOME"
201 UNION
202 SELECT "TEAM_NAME_AWAY", AVG("TOTAL_TURNOVERS_AWAY") pts
203 FROM "Game" g
204 WHERE g."SEASON_ID" = 22020
205 GROUP BY "TEAM_NAME_AWAY"
206
207 SELECT "TEAM_NAME_HOME", CAST(AVG(pts) As Decimal(10,2)) turnovers
208 From turnovers
209 GROUP BY "TEAM_NAME_HOME"
210 Order by turnovers asc
211
```

Data output Messages Notifications

	TEAM_NAME_HOME text	turnovers numeric (10,2)
1	Portland Trail Blazers	10.89
2	San Antonio Spurs	11.33
3	Dallas Mavericks	11.86
4	Phoenix Suns	12.26
5	Sacramento Kings	12.71