SQL Showcase | MySQL Soccer Players Stats Database

Introduction

This was from a freelance project for a local soccer Facebook group, where I was tasked to:
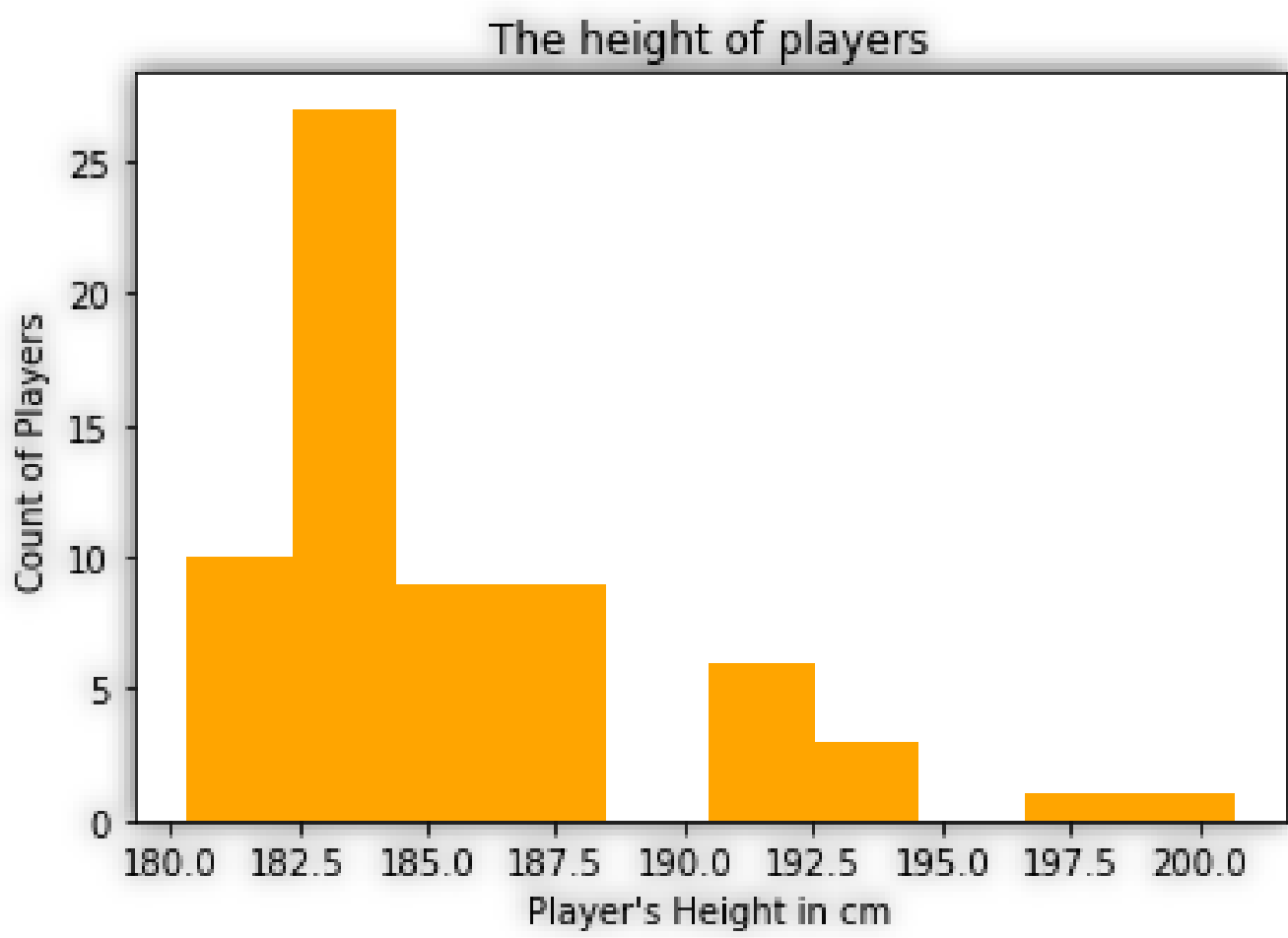
- Look at the overall structure of the MySQL database and convert it to Data Frames
- Perform exploratory data analysis on any interesting features
- Use MySQL functions, such as INNER JOIN and GROUP BY
- Create data visualizations of comparing different player's stats

Findings

After connecting to the remote MySQL database. I converted the database into a DataFrame for exploratory data analysis. This process allows me to play around with the data and create data visualizations. Below is a snapshot of the newly created DataFrame.
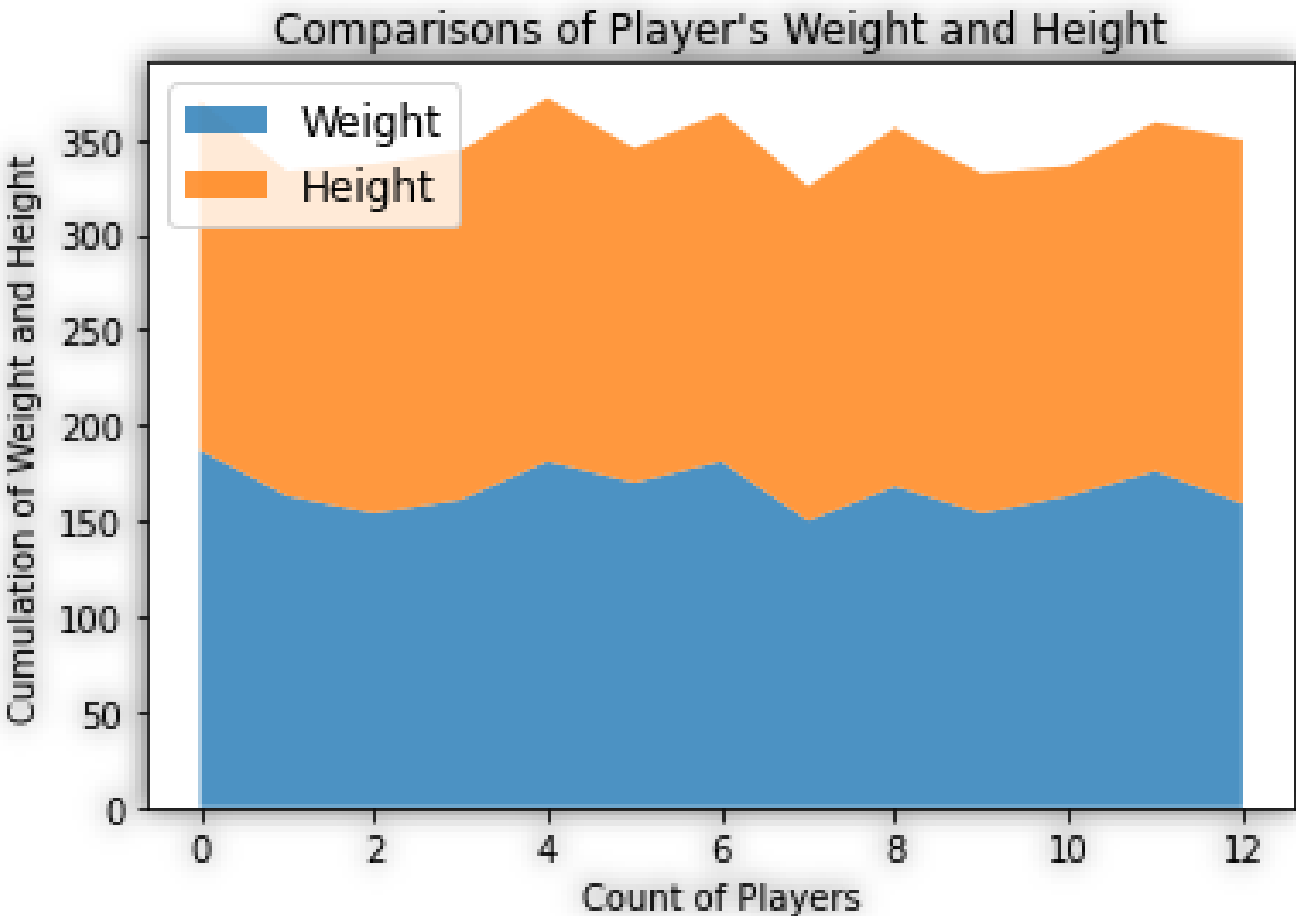
	id	player_api_id	player_name	player_fifa_api_id	birthday	height	weight
0	1	505942	Aaron Appindangoye	218353	1992-02-29 00:00:00	182.88	187
1	2	155782	Aaron Cresswell	189615	1989-12-15 00:00:00	170.18	146
2	3	162549	Aaron Doran	186170	1991-05-13 00:00:00	170.18	163
3	4	30572	Aaron Galindo	140161	1982-05-08 00:00:00	182.88	198
4	5	23780	Aaron Hughes	17725	1979-11-08 00:00:00	182.88	154
5	6	27316	Aaron Hunt	158138	1986-09-04 00:00:00	182.88	161
6	7	564793	Aaron Kuhl	221280	1996-01-30 00:00:00	172.72	146
7	8	30895	Aaron Lennon	152747	1987-04-16 00:00:00	165.10	139
8	9	528212	Aaron Lennox	206592	1993-02-19 00:00:00	190.50	181
9	10	101042	Aaron Meijers	188621	1987-10-28 00:00:00	175.26	170

After performing a few primary exploratory data analyses. I have noticed many players are taller than 180 cm (5 ft 10 inches). Performed an inner join function to combine two databases into one. Created data visualizations from the query about the player's height.

The height of players

Histogram displaying the total counts of players' heights listed in centimeters.

Performed another inner join function to combine the listed players table with another one. Compared the player's cumulative weight and height in an area chart. Below is the area chart.

Comparisons of Player's Weight and Height

Area chart displaying the cumulation of players' weight and height.

Challenges

One major challenge that I encountered for this project was looking for the right features that could be useful based on the given guidelines. Another challenge was the size of the original DataFrame. Focusing on the areas where the insights would be aligned with the given tasks.

Conclusion

I was able to complete the given tasks that were listed in the introduction section. It was a little difficult to find the right features that were based on the players' stats and seen as useful to the local soccer Facebook group. The charts will be used in an upcoming newsletter.