# EDA ON FIFA\_SPORT DATASAT USING MYSQL

Exploratory Data Analysis (aka EDA) is the process of Descriptive Data Analytics that helps to know what has happened based on the given historical data.

This is the initial Descriptive Analysis part of the Data Analysis (aka DA) that involves cleaning and transforming data for the next step of DA which can be Data Modelling (determining and defining the relationships between the tables).

**Why EDA is done?**

To get a grasp on the organization’s data assets and extract meaningful information from their historical patterns that further helps the business team to perform required actions to meet business targets or enhance KPIs (Key Performance Indicators - to measure the performance of the tasks/targets).

Here I have considered the sports dataset on FIFA, to perform common and essential data tasks in EDA especially if you are beginning to analyze data, these are essential steps whether DA is performed with MySQL or Python, or R. There could be subtle changes in the syntaxes for other relational databases.

In short, DA helps to extract valuable information from the given dataset.

Dataset: The sample dataset that can be available from the Tableau website, under [FIFA 18](https://public.tableau.com/en-us/s/resources) then click on ‘Sample Data’ at the top, scroll down to the ‘download’ button for .xlsx or .csv formats - <https://public.tableau.com/en-us/s/resources>

With relational database - we can compute descriptive statistics using aggregation functions such as min(), max(), avg(), sum(), stddev() on a numerical fields/column. It’s similar to using describe() function from Pandas if you are familiar with Data Analysis with Python.

**What’s in EDA here:**

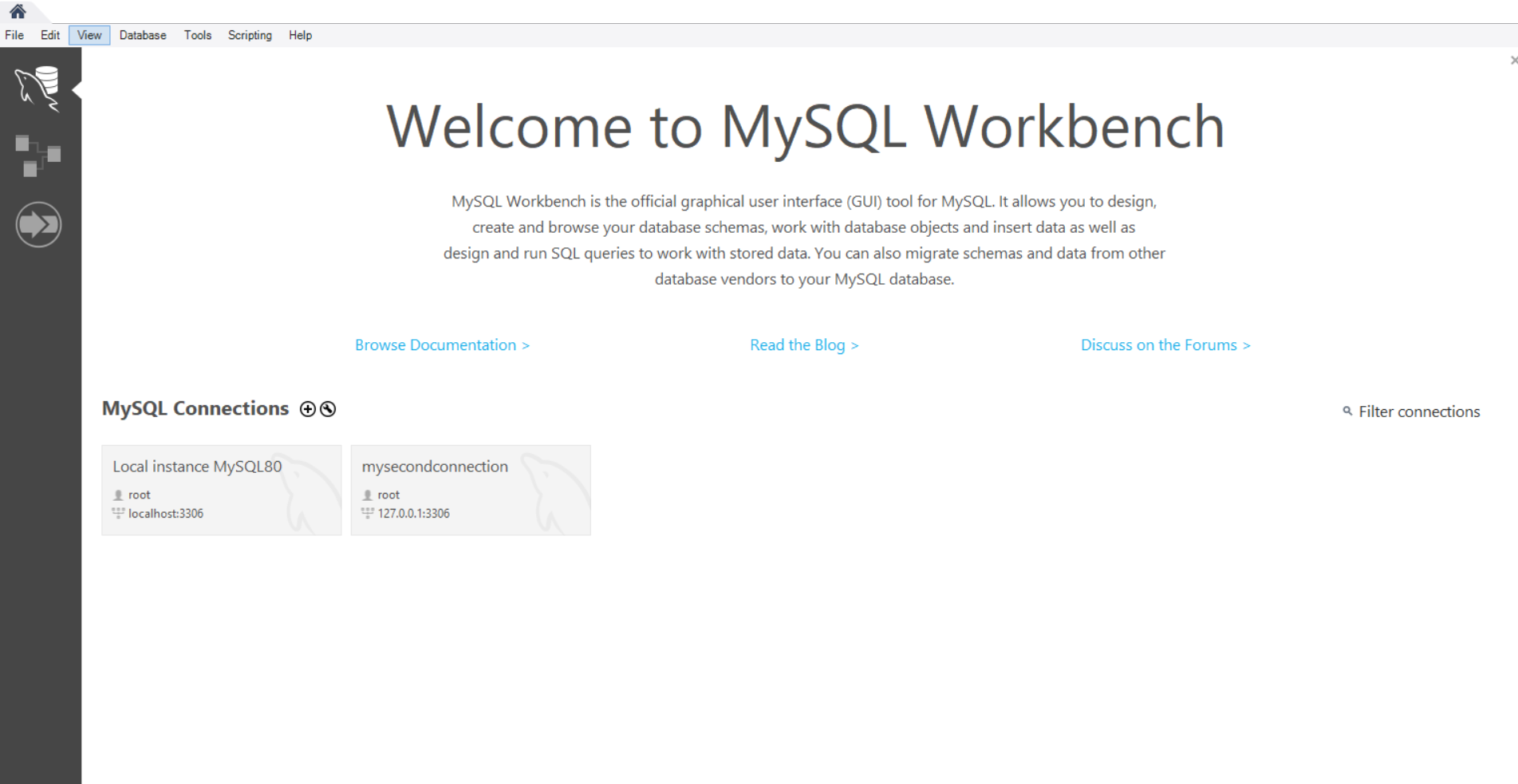
Exploratory Data Analysis, done on multiple scenarios. Ex: analysis of the company’s data assets or it’s used for building Machine Learning predictive models. EDA is performed in and for nearly all business types - to analyze KPIs and for companies adopting data-driven decision making.

Note: Before you begin to write & execute SQL statements whether in MySQL Shell or Workbench, remember to keep the database server turned on, sometimes this step can be missed which happens with me! so as a note to turn, the server on.

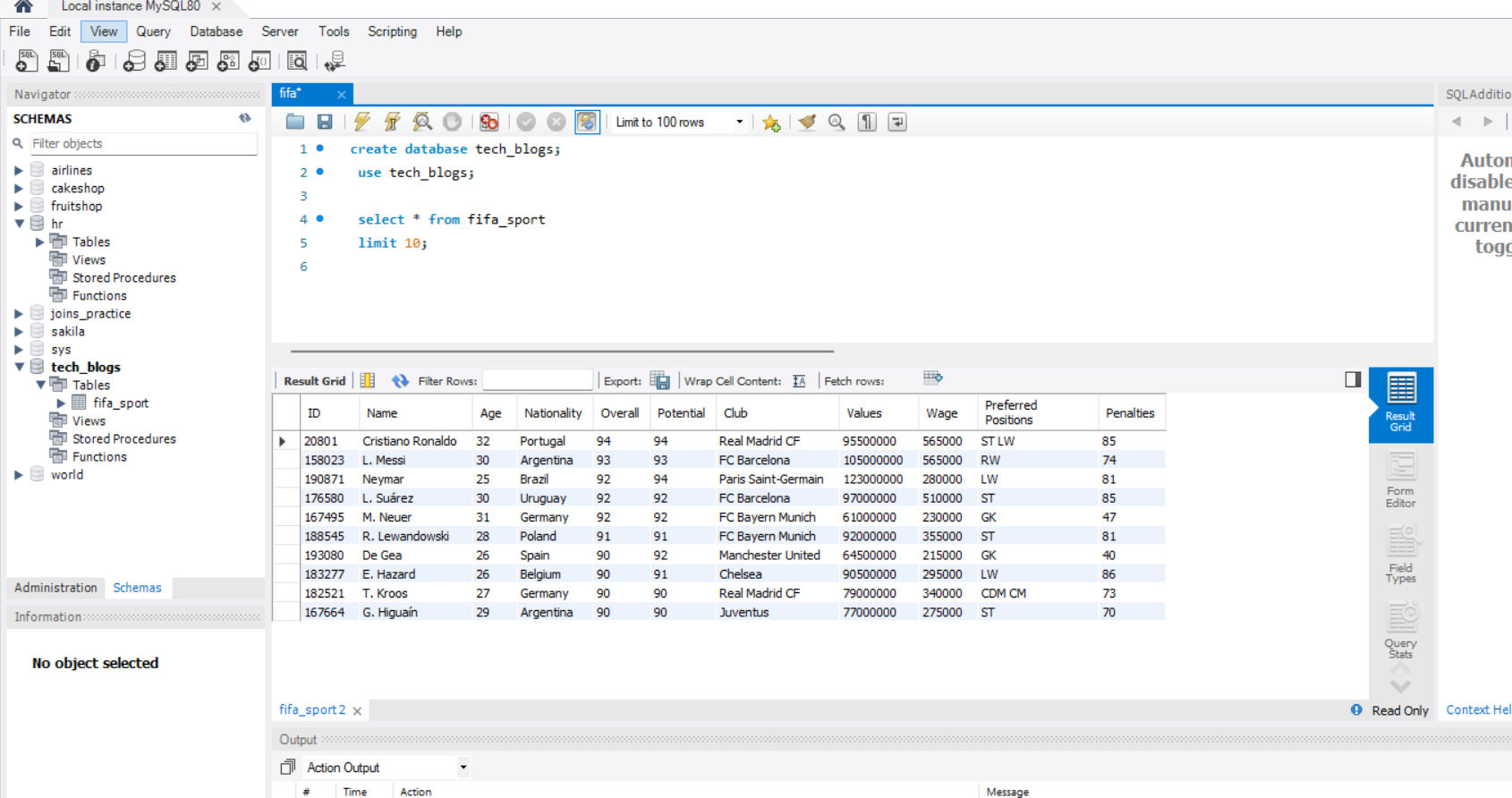
In the article [To Interact with MySQL Database](https://ramya-n.medium.com/to-interact-with-mysql-database-4f7e75e764d5), you can find ways to connect to the MySQL server where I have explained Shell and Workbench tools along with the use-case of essential CRUD operations and about the MySQL Server version in more detail.

**1. With MySQL Workbench:**

Once the database server is turned on, you can open the GUI tool i.e., Workbench as seen in the below screenshot.



Overview of MySQL Workbench tool with the fifa\_sport table in tech\_blogs database:



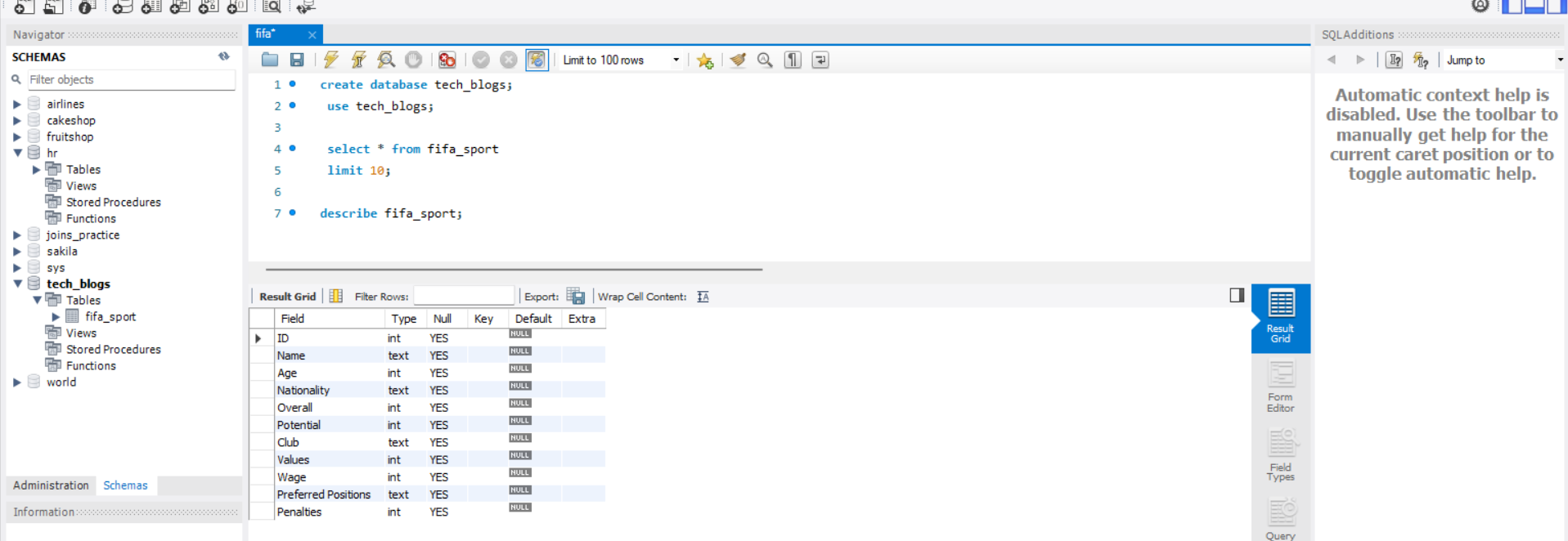
To activate the required database, execute the statements in Workbench as show above on line 2.

Below are the statements can be executed in the Workbench.

Below are the statements to learn about the metrics such as number of records/rows, to fetch records and unique values, structure/schema (columns, data type…) of table and the like.

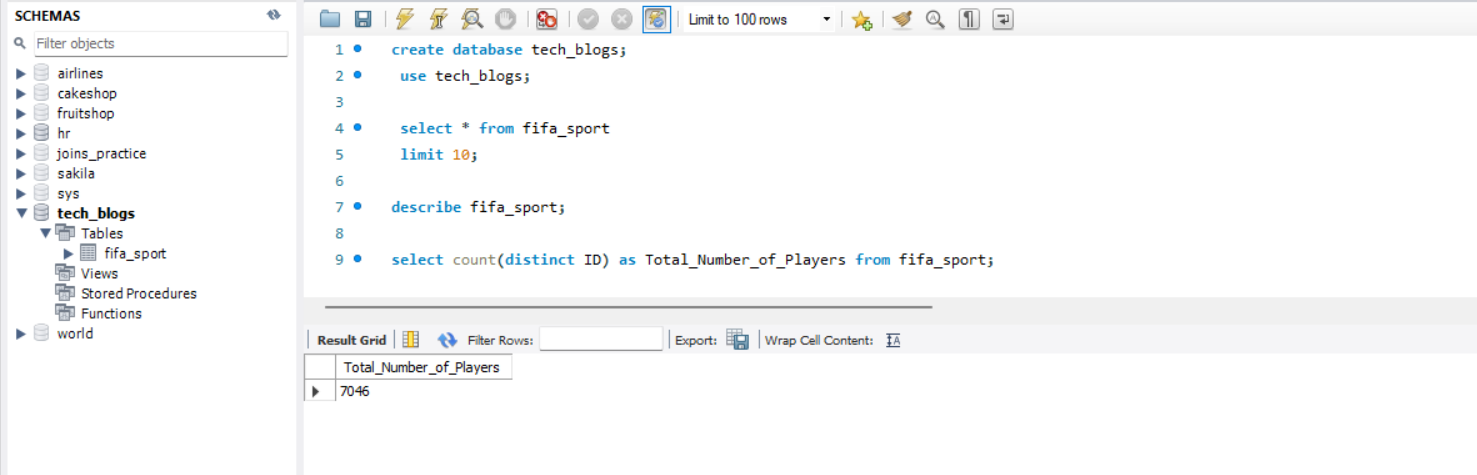
So, without further delay let’s dive in for next steps.

The EDA helps you to know which features (fields) required for the analysis, if there are any outliers, any missing values, unique values, descriptive statistics (average, max and min values, standard deviation) which ofcourse depends on the scenarios - (i) if you are doing EDA to build predictive model (Machine Learning application) or (ii) to extract information for data-driven decisions making, depending on organization’s requirement, here I am explaining EDA for the scenario(ii) i.e., for data-driven decision making!

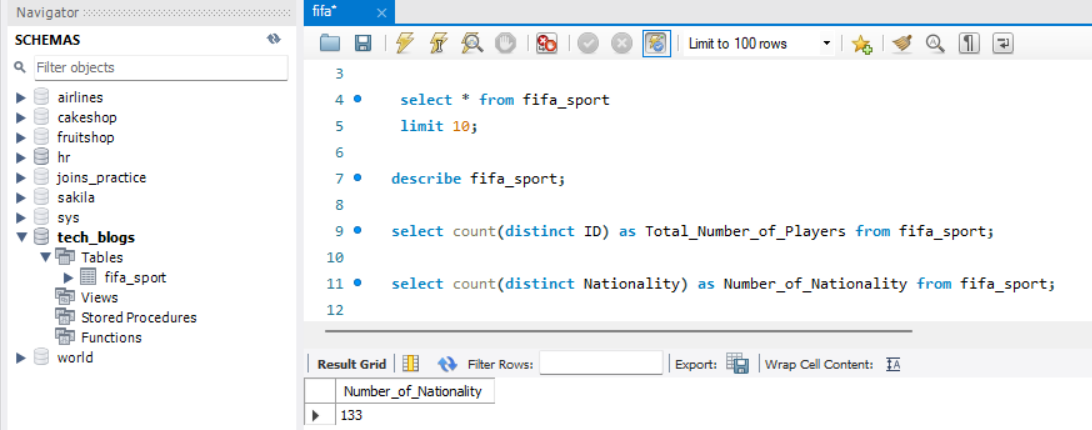


Below are some of the questions, that can be answered from these SQL statements. By now you know that, we are working with MySQL (RDBMS) and SQL (language) here.

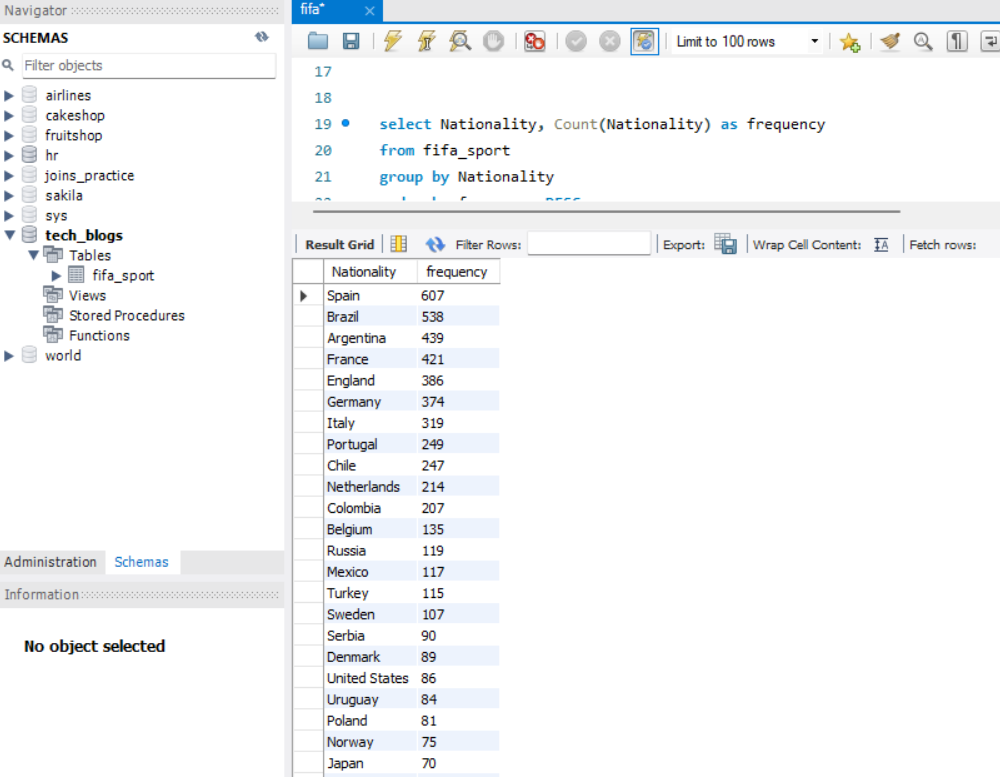
**Q1: How many football players are there in the FIFA organization?**



**Q2: How many nationalities are there?**

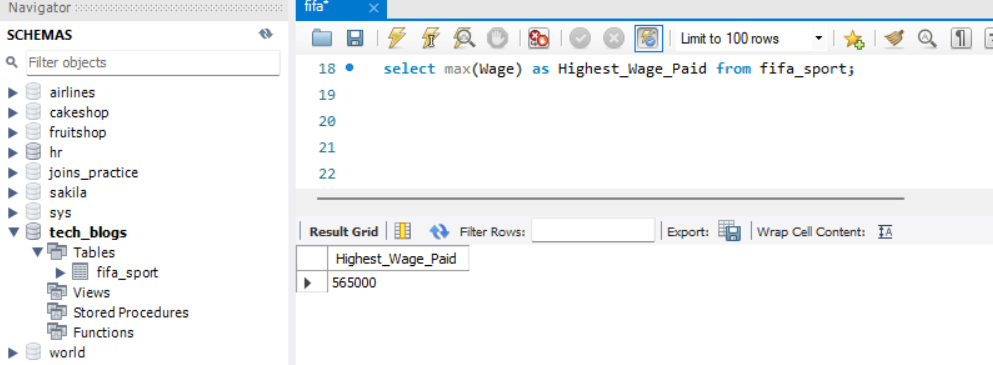


**Q3: How many players from each country - i.e., frequency of nationality?**

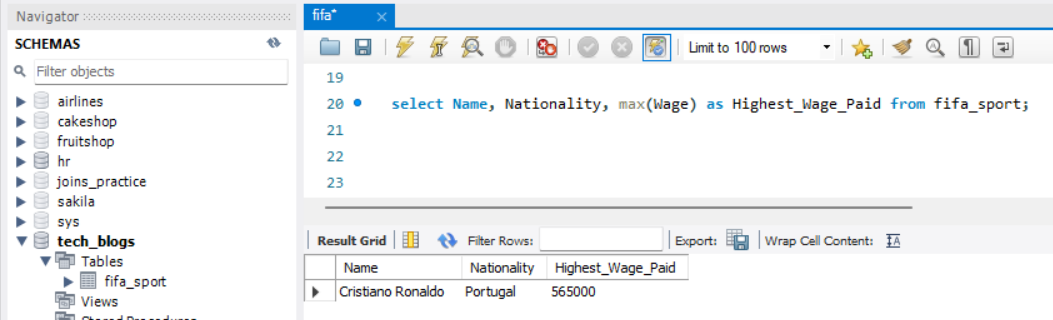
****

Above statement also provides the information on which country has got more players i.e., highest number of players are from Spain.

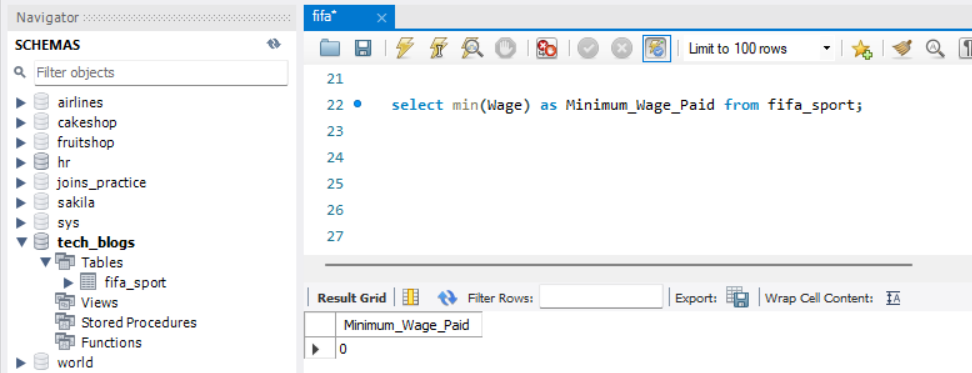
**Q4: What is the highest amount of wage paid to the player?**

****

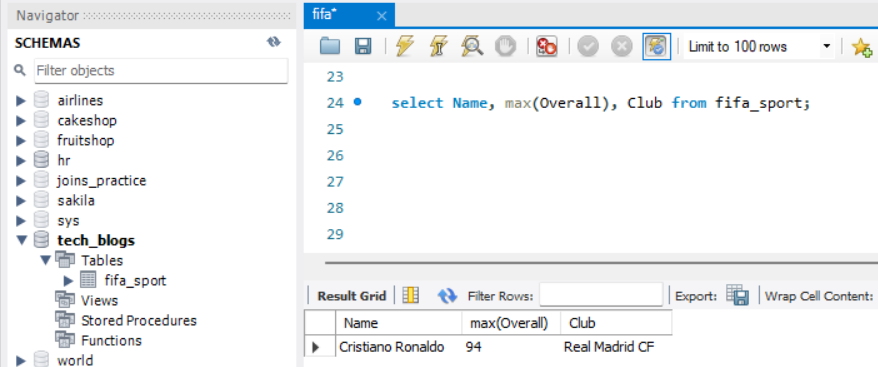
**Q5: Which player is getting paid highest and from which country?**

****

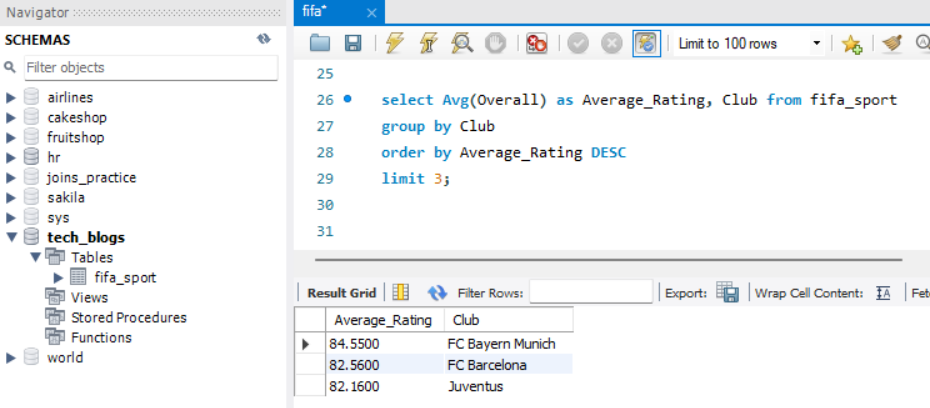
**Q6: What is the minimum wage paid?**

****

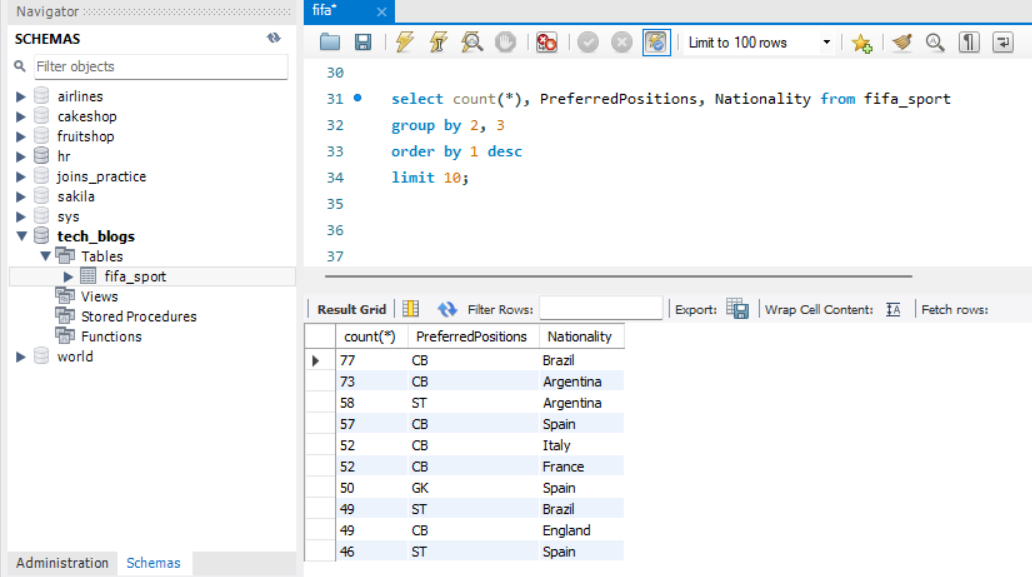
**Q7: Which player has got overall highest rating and from which club?**

****

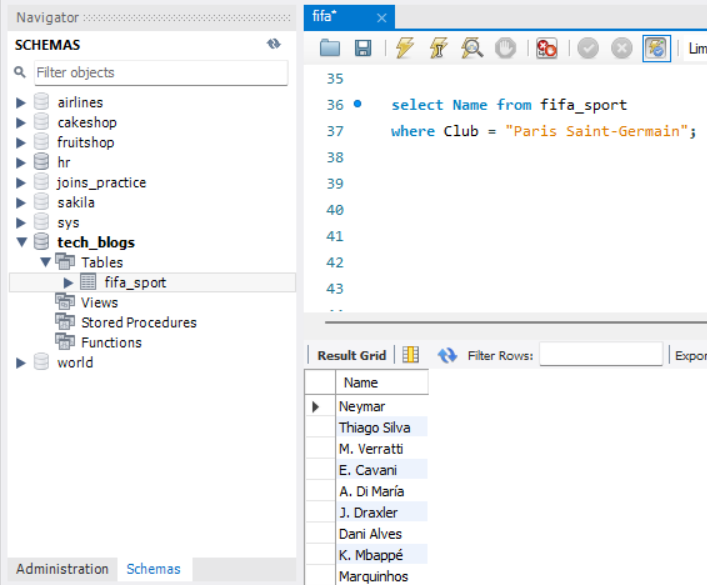
**Q8: Which are top-3 clubs based on the overall rating?**

****

**Q9: How many player preferred position with their nationality and show top-10 records?**

****

**Q 10: Who are the players registered at ‘Paris Saint-Germain’ club?**

****

Summary: So, depending on the questions, we can analyze the given dataset, thus it can be presented to the respective team with the report & requested information. For representing data in graph visualizations, you can use Tableau or Python with Seaborn/Matplotlib libraries or R language to enhance the report and presentation.