

SQL for Data Science

Module 6: Data Analysis

- Counting Rows and Items
- Aggregation Functions – SUM, AVG, STDDEV
- Extreme Values Identification – MIN, MAX
- Slicing Data
- Limiting Data
- Sorting Data
- Filtering Patterns
- Groupings and Rolling up data
- Filtering in Groups
- Real life project – Descriptive Analytics of FIFA19 players



- ID: Unique identification of the player
- Name: Name of the player
- Age: Age of the player
- Nationality: Players Nationality
- Overall: Overall Rating
- Potential: Potential Rating
- Club: Player belongs to this club
- Value: Current Market Value in Euros
- Wage: Current Wage in Euros
- Preferred Foot: Preferred Foot for playing
- Jersey Number
- Joined: Joining date
- Height in feet and inches
- Weight in pounds (lbs)
- Penalties: Rating on a scale of 100



- How many players are there in the dataset?
- How many nationalities do these players belong to?
- What is the total wage given to all players? What's the average and standard deviation?
- Which nationality has the highest number of players, what are the top 3 nationalities by # of players?
- Which player has the highest wage? Who has the lowest?
- The player having the – best overall rating? Worst overall rating?
- Club having the highest total of overall rating? Highest Average of overall rating?
- What are the top 5 clubs based on the average ratings of their players and their corresponding averages?



- What is the distribution of players whose preferred foot is left vs right?
- Which jersey number is the luckiest?
- What is the frequency distribution of nationalities among players whose club name starts with M?
- How many players have joined their respective clubs in the date range 20 May 2018 to 10 April 2019 (both inclusive)?
- How many players have joined their respective clubs date wise?
- How many players have joined their respective clubs yearly?



UPDATE Command is used to update data in a tables

UPDATE Command

```
UPDATE `table_name` SET `column_name` = `new_value` [WHERE condition];
```

Things to note

- **Updates have to be done one column at a time**
- **Where clause is used to select the rows on which we want to do the updates**
- **If the where clause is not used it with set all rows of `column_name` to `new_value`**

UPDATE Command is used to update data in a tables

UPDATE Command

```
UPDATE `table_name` SET `column_name` = `new_value` [WHERE condition];
```

DEMO

DELETE Command is used to delete data from a table

DELETE Command

```
DELETE FROM `table_name` [WHERE condition];
```

Things to note

- **All the matching rows corresponding to the where clause will be deleted from the table**
- **If the where clause is not used all rows will be deleted**

Deleting data from a table

DELETE Command is used to delete data from a table

DELETE Command

DEMO

DESCRIBE and SHOW CREATE TABLE commands help us to know the structure of a table

DESCRIBE Command

```
DESCRIBE `table_name`
```

SHOW CREATE TABLE Command

```
SHOW CREATE TABLE `table_name`
```

Difference between the two

- **Describe command gives the schema/ structure of the table**
- **Show create table gives the query which can be used to re-create the table schema**

DESCRIBE and SHOW CREATE TABLE commands help us to know the structure of a table

DESCRIBE Command

DEMO



ALTER command is used to alter a table

ALTER Command

```
ALTER TABLE `table_name` (ADD/DROP/CHANGE/MODIFY)  
COLUMN `column_name` `data_type` AFTER COLUMN2 `column_name2`;
```

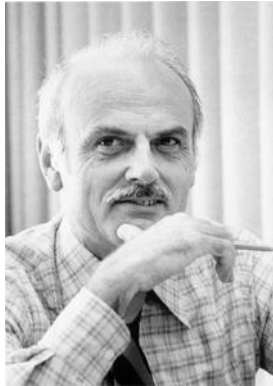
Uses of Alter command

- **Add a column**
- **Drop a column**
- **Change a column**
- **Modify a column**
- **AFTER** is used to put the column at a particular place in the table (after some other column)



SQL started with IBM Researcher Edgar Codd's Research on Relational Databases

1972



Edgar Codd

- **Researcher at IBM Research Center**
- **Mathematician trained from Oxford**
- **Researching on Relational Databases**
- **Chamberlin and Boyce come up with SEQUEL (Structured English Query Language to interact with IBM System R database)**

1979



- **Trademark Issue with a Firm**
- **SEQUEL was changed to SQL**

Connecting to MySQL Server

Connecting to MySQL Server is pretty straightforward

Goto Terminal/ Command Prompt and type

```
[anands-MacBook-Pro:~ analytics$ mysql -uroot -p  
[Enter password:
```

Data Definition Language

Commands used to

- Define the schema of database or its objects (like tables and indexes)
- Create and Modify the structure of database objects
- Examples:
 - CREATE
 - DROP
 - ALTER

Data Manipulation Language

Commands used to

- Manipulate and Select data in the database
- Examples:
 - SELECT
 - INSERT
 - UPDATE
 - DELETE

Data Control Language

Commands dealing with

- Rights, permissions and other controls of the database system
- Examples:
 - GRANT
 - REVOKE

Here we explore some simple commands. Note that all commands end with ; or \G in MySQL

Show all databases

```
mysql> show databases;
```

Work with a particular database

```
mysql> use <database_name>;
```

Get help about commands

```
mysql> help;
```

Get topicwise help

```
mysql> help contents;
```

```
mysql> help Data Manipulation;
```

Here we explore some simple commands. Note that all commands end with ; or \G in MySQL

Show all databases

```
mysql> show databases;
```

Work with a particular database

```
mysql> use <database_name>;
```

Get help about commands

```
mysql> help;
```

Get topicwise help

```
mysql> help contents;
```

```
mysql> help Data Manipulation;
```

We have listed the most commonly used datatypes here. There are a lot more, to learn more: Refer to <https://dev.mysql.com/doc/refman/8.0/en/data-types.html>

Most Popular

- int(10)
- varchar(255)
- text
- TIMESTAMP
- ENUM ('Choice1', 'Choice2', ...)

Not so common

- FLOAT
- DECIMAL
- BLOB
- TINYBLOB
- MEDIUMBLOB
- BIGINT
- SMALLINT
- TINYINT
- DATE
- TIME
- SET
- DOUBLE
- CHAR

Some fields we can keep optional – Others are Mandatory

Difference between NULL and NOT NULL Columns/ Fields

- A column which has NOT NULL constraint means it is mandatory to put some value for the column while inserting the row
- A column which has NULL constraint means its ok to give NULL value – a special value which means blank
- This is defined in the structure of the table