



# SQL for Data Science

Author	
Created By	swaraj gogoi
Last Edited	@May 12, 2020 8:00 PM
Property	
Tags	

## What is a database

The database is a collection of inter-related data which is used to retrieve, insert and delete the data efficiently. It is also used to organize the data in the form of a table, schema, views, and reports, etc.

For example: The college Database organizes the data about the admin, staff, students and faculty etc. Using the database, you can easily retrieve, insert, and delete the information.

## Data Base Management System (DBMS) :

- Database management system is a software which is used to manage the database. For example: [MySQL](#), [Oracle](#), etc are a very popular commercial database which is used in different applications.
- DBMS provides an interface to perform various operations like database creation, storing data in it, updating data, creating a table in the database and a lot more.

- It provides protection and security to the database. In the case of multiple users, it also maintains data consistency.

### **Introduction to SQL for Data Science :**

MySQL is an open-source relational database management system.

We are using a database from a company called Crunchbase. Crunchbase is a platform for finding business information about private and public companies. Crunchbase information includes investments and funding information, founding members and individuals in leadership positions, mergers and acquisitions, news, and industry trends.

```
SELECT name, homepage_url, category_code FROM companies
```

	name	homepage_url	category_code
1	8868	http://www.8868.cn	NULL
2	2.10E+07	NULL	NULL
3	.Club Domains	http://dotclub.com	software
4	.Fox Networks	http://www.dotfox.com	advertising
5	[a]list games	http://www.alistgames.com	games_video
6	@Pay	http://atpay.com	mobile
7	&TV Communications	http://enjoyandtv.com	games_video
8	#waywire	http://www.waywire.com	news

To make your data presentable to your employers you can change the column names using SQL query in this way :

```
SELECT name as "Company Name", homepage_url as "Webpage", category_code as "Category"
FROM companies
```

	Company Name	Webpage	Category
1	8868	http://www.8868.cn	NULL
2	2.10E+07	NULL	NULL
3	.Club Domains	http://dotclub.com	software
4	.Fox Networks	http://www.dotfox.com	advertising
5	[a]list games	http://www.alistgames.com	games_video
6	@Pay	http://atpay.com	mobile
7	&TV Communications	http://enjoyandtv.com	games_video
8	#waywire	http://www.waywire.com	news

As in pandas, we use the `head()` method to see only first few rows of our data frame, the keyword "**Limit**" in SQL can be used in a similar way

```
select * from companies limit 5
//which will show only first 5 rows of the companies table
```

	Company Name	Webpage	Category
1	8868	http://www.8868.cn	NULL
2	2.10E+07	NULL	NULL
3	.Club Domains	http://dotclub.com	software
4	.Fox Networks	http://www.dotfox.com	advertising
5	[a]list games	http://www.alistgames.com	games_video

The "**where**" keyword can be used to select table data based on some condition

```
##Find the name of those companies which have a status of "operating"
select name as "Company", status from companies where status= "operating"
```

	Company	status
1	8868	operating
2	2.10E+07	operating
3	.Club Domains	operating
4	[a]list games	operating
5	@Pay	operating
6	&TV Communications	operating
7	0-6.com	operating
8	0xdata	operating

```
##Did any company receive funding greater than 10,000,000
select * from companies where funding_total_usd > 10000000
```

	permalink	name	homepage_url	funding_total_usd
1	/company/1010data	1010data	http://www.1010data.com	35000000
2	/company/1366-technologies	1366 Technologies	http://www.1366tech.com	66500000
3	/company/170-systems	170 Systems	http://www.170systems.com	14000000
4	/company/1c-company	1C Company	http://1c.ru/eng	20000000
5	/company/1life-healthcare	1Life Healthcare	http://www.1life.com	30000000
6	/company/1stdibs	1stdibs	http://www.1stdibs.com	57000000
7	/company/20-20-mobile	20:20 Mobile	http://www.2020mobile.com	29800000
8	/company/2080-media	2080 Media	NULL	41000000

```
## Find where funding_total is NULL  
select * from companies where funding_total_usd is NULL
```

```
## SELECT names which have fox inside  
SELECT * FROM companies WHERE name like "%fox%"
```

```
##Display companies and their total funding in INR  
select name as "Company", funding_total_usd, funding_total_usd*75.30 as "Funding INR"  
from companies
```

	Company	funding_total_usd	Funding INR
1	8868	NULL	NULL
2	2.10E+07	5050000	380265000.0
3	.Club Domains	7000000	527100000.0
4	.Fox Networks	4912394	369903268.2
5	[a]list games	9300000	700290000.0
6	@Pay	3500000	263550000.0
7	&TV Communications	4000000	301200000.0
8	#waywire	1750000	131775000.0

```
## Create a new column called Region ID  
select name as "Company" , country_code ||"\\"|| city as "Region ID" from companies
```

	Company	Region ID
1	8868	NULL
2	2.10E+07	USA\San Francisco
3	.Club Domains	USA\Oakland Park
4	.Fox Networks	ARG\Buenos Aires
5	[a]list games	NULL
6	@Pay	USA\Albuquerque
7	&TV Communications	USA\Los Angeles
8	#waywire	USA\New York

```
## Find company names that have second character as 'a'  
SELECT * FROM companies where name like "_a%"
```

	permalink	name	homepage_url	category_code	funding_total_usd
1	/company/a-list-games	[a]list games	http://www.alistgames.com	games_video	9300000
2	/company/7ac-technologies	7AC Technologies	http://7actech.com	cleantech	9860000
3	/company/8awEEK	8awEEK	http://8awEEK.com	web	NULL
4	/company/aaipharma-services	AAI Pharma Services	http://www.aaipharma.com	biotech	20000000
5	/company/aarden-pharmaceuticals	Aarden Pharmaceuticals	http://aardenpharma.com	biotech	500000
6	/company/aastrom-biosciences	Aastrom Biosciences	http://www.aastrom.com	biotech	40000000
7	/company/aava-mobile	Aava Mobile	http://www.aavamobile.com	mobile	4260000

```
##Display only those companies whose names are 8868, 0-6.com and @Pay
SELECT * FROM companies where name in ('8868','0-6.com','@Pay')
```

	permalink	name	homepage_url	category_code	funding_total_usd	status	country_code
1	/company/8868	8868	http://www.8868.cn	NULL	NULL	operating	NULL
2	/company/pay-mobile-checkout	@Pay	http://atpay.com	mobile	3500000	operating	USA
3	/company/0-6-com	0-6.com	http://www.0-6.com	web	2000000	operating	NULL

```
## ORDER COMPANIES BY TOTAL FUNDING AMOUNT.
select * from companies where funding_total_usd is Not null order by
funding_total_usd DESC
```

	permalink	name	homepage_url	category_code	funding_total_usd
1	/company/abound-solar	Abound Solar	http://www.abound.com	cleantech	5100000
2	/company/a123systems	A123 Systems	http://a123systems.com	nanotech	5030000
3	/company/admi-holdings	ADMI Holdings	NULL	NULL	3680000
4	/company/airbnb	Airbnb	http://airbnb.com	travel	3260000
5	/company/aiming	Aiming	http://aiming-inc.com/en	games_video	3000000
6	/company/aircell	AirCell	http://www.aircell.com	messaging	2760000
7	/company/aersale-holdings	AerSale Holdings	http://www.aersale.com	manufacturing	2500000
8	/company/adknowledge	Adknowledge	http://adknowledge.com	advertising	2480000

```
## Multiple Column Ordering
SELECT * FROM companies WHERE category_code IS NOT NULL
ORDER BY category_code, funding_total_usd DESC
```

	permalink	name	homepage_url	category_code	funding_total_usd
1	/company/adknowledge	Adknowledge	http://adknowledge.com	advertising	2480000
2	/company/58-com	58.com	http://www.58.com	advertising	1410000
3	/company/adchemy	Adchemy	http://www.adchemy.com	advertising	1190000
4	/company/adconion-media-group	Adconion Media Group	http://www.adconion.com	advertising	1140000
5	/company/affinity-com	Affinity.com	http://www.affinity.com	advertising	850000
6	/company/aggregateknowledge	Aggregate Knowledge	http://www.aggregateknowled...	advertising	743000
7	/company/addthis	AddThis	http://www.addthis.com	advertising	730000
8	/company/4info	4INFO	http://www.4info.com	advertising	643000

```
##Find number of companies that have funding greater than 100000
SELECT COUNT(*) FROM companies WHERE funding_total_usd > 100000
```

	COUNT(name)
1	811

```
SELECT COUNT(homepage_url) FROM companies WHERE funding_total_usd > 100000
## Count has lower value, since some are NULL
```

	COUNT(homepage_url)
1	762

```
SELECT DISTINCT category_code AS "Dist Cat"
From companies WHERE category_code IS NOT NULL
```

	Dist Cat
1	software
2	advertising
3	games_video
4	mobile
5	news
6	web
7	analytics
8	health
9	ecommerce

```
SELECT COUNT(DISTINCT category_code) AS "Dist Cat"
From companies WHERE category_code IS NOT NULL
```

	Dist Cat
1	40

```
##Find Total Funding received in SF Bay
SELECT SUM(funding_total_usd) AS "Funding Total"
From companies WHERE region = "SF Bay"
```

Funding Total
1 3328169119