

NAME: SHARATH CHANDRA

BATCH : DATA ENGINEERING

TOPIC :DATA CLEANING

—> Firstly i have created the database and created the table then inserted the values into the tables

```
mysql> create database pet_adoption;
Query OK, 1 row affected (0.00 sec)

mysql> use pet_adoption;
Database changed
mysql> create table adoptions(
  -> animal_id int not null,
  -> name varchar(40),
  -> contact varchar(40)
  -> );
Query OK, 0 rows affected (0.02 sec)

mysql> create table animals(
  -> id int not null,
  -> name varchar(40),
  -> breed varchar(30),
  -> color varchar(30),
  -> gender varchar(10),
  -> status int
  -> );
Query OK, 0 rows affected (0.02 sec)

mysql> insert into animals(id,name,breed,color,gender,status) values(501,'Bellyflop','Beagle','Brown','Male',0);
Query OK, 1 row affected (0.01 sec)

mysql> insert into animals(id,name,breed,color,gender,status) values(502,'Snowy','Husky','White','Female',0);
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status) values(503,'Princes','Pomarnian','Black','Female',0);
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status) values(504,'Cricket','Chihuahua','Brown','Female',0);
```

```
mysql> insert into animals(id,name,breed,color,gender,status) values(504,'Cricket','Chihuahua','Brown','Female',0);
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status) values(505,'Spot','Dalmation','Black and White','Male',0);
Query OK, 1 row affected (0.00 sec)

mysql> ^C
mysql> insert into adoptions(animal_id,name,contact) values(101,'Pinnocchio','pinnocchio@gmail.com');
Query OK, 1 row affected (0.01 sec)

mysql> insert into adoptions(animal_id,name,contact) values(102,'Stella','Stella@gmail.com');
Query OK, 1 row affected (0.01 sec)

mysql> insert into adoptions(animal_id,name,contact) values(103,'Bob','bob@gmail.com');
Query OK, 1 row affected (0.00 sec)

mysql> insert into adoptions(animal_id,name,contact) values(104,'alex','alex@gmail.com');
Query OK, 1 row affected (0.00 sec)

mysql> insert into adoptions(animal_id,name,contact) values(105,'sunny','sunny@gmail.com');
Query OK, 1 row affected (0.01 sec)
```

—>Then I executed the sample queries using where clause .

```
mysql> select id,breed from animals;
+----+-----+
| id | breed |
+----+-----+
| 501 | Beagle |
| 502 | Husky |
| 503 | Pomarnian |
| 504 | Chihuahua |
| 505 | Dalmation |
+----+-----+
5 rows in set (0.00 sec)

mysql> select name from animals where gender='Female';
+-----+
| name |
+-----+
| Snowy |
| Princes |
| Cricket |
+-----+
3 rows in set (0.00 sec)

mysql> select id from animals where status=0;
+----+
| id |
+----+
| 501 |
| 502 |
| 503 |
| 504 |
| 505 |
+----+
5 rows in set (0.00 sec)
```

—> Then i have updated the table and executed the query once again .

```
mysql> update animals set color='Brown and red' where name='Spot';
Query OK, 1 row affected (0.01 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> update animals set gender='Red' where id=501;
Query OK, 1 row affected (0.00 sec)
Rows matched: 1  Changed: 1  Warnings: 0

mysql> update animals set status=1 where gender='Female';
Query OK, 3 rows affected (0.00 sec)
Rows matched: 3  Changed: 3  Warnings: 0

mysql> delete animals where id=501;
ERROR 1064 (42000): You have an error in your SQL syntax; check the manual that corresponds to your MySQL server version for the right syntax to use near 'where id=501' at line 1
mysql> select *from animals;
+----+-----+-----+-----+-----+-----+
| id | name | breed | color | gender | status |
+----+-----+-----+-----+-----+-----+
| 501 | Bellyflop | Beagle | Brown | Red | 0 |
| 502 | Snowy | Husky | White | Female | 1 |
| 503 | Princes | Pomarnian | Black | Female | 1 |
| 504 | Cricket | Chihuahua | Brown | Female | 1 |
| 505 | Spot | Dalmation | Brown and red | Male | 0 |
+----+-----+-----+-----+-----+-----+
5 rows in set (0.00 sec)
```

—>Then i have executed the query using ORDERBY clause

```
mysql> select *from adoptions;
```

animal_id	name	contact
101	Pinnocchio	pinnocchio@gmail.com
102	Stella	Stella@gmail.com
103	Bob	bob@gmail.com
104	alex	alex@gmail.com
105	sunny	sunny@gmail.com

```
5 rows in set (0.00 sec)
```

```
mysql> select *from animals
-> order by status desc;
```

id	name	breed	color	gender	status
502	Snowy	Husky	White	Female	1
503	Princes	Pomarnian	Black	Female	1
504	Cricket	Chihuahua	Brown	Female	1
501	Bellyflop	Beagle	Brown	Red	0
505	Spot	Dalmation	Brown and red	Male	0

```
5 rows in set (0.00 sec)
```

```
mysql> select *from animals
-> order by color desc;
```

id	name	breed	color	gender	status
502	Snowy	Husky	White	Female	1
505	Spot	Dalmation	Brown and red	Male	0
501	Bellyflop	Beagle	Brown	Red	0
504	Cricket	Chihuahua	Brown	Female	1

```
mysql> update animals set species='Dog';
Query OK, 5 rows affected (0.00 sec)
Rows matched: 5 Changed: 5 Warnings: 0

mysql> insert into animals(id,name,breed,color,gender,status,species) values(401,'Meowcat1','asian','Brown','Male',0,'Cat');
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status,species) values(402,'Meowcat2','africa','white','Female',1,'Cat');
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status,species) values(403,'Meowcat3','africa','Black','FeMale',1,'Cat');
Query OK, 1 row affected (0.00 sec)

mysql> insert into animals(id,name,breed,color,gender,status,species) values(404,'Meowcat4','bengal','Black and white','Male',0,'Cat');
Query OK, 1 row affected (0.00 sec)

mysql> select *from animals
-> order by id ;
+----+-----+-----+-----+-----+-----+-----+
| id | name   | breed | color   | gender | status | species |
+----+-----+-----+-----+-----+-----+-----+
| 401 | Meowcat1 | asian | Brown   | Male   | 0      | Cat     |
| 402 | Meowcat2 | africa | white   | Female | 1      | Cat     |
| 403 | Meowcat3 | africa | Black   | FeMale | 1      | Cat     |
| 404 | Meowcat4 | bengal | Black and white | Male   | 0      | Cat     |
| 501 | BellyFlop | Beagle | Brown   | Red    | 0      | Dog     |
| 502 | Snowy    | Husky | White   | Female | 1      | Dog     |
| 503 | Princes  | Pomarnian | Black | Female | 1      | Dog     |
| 504 | Cricket  | Chihuahua | Brown | Female | 1      | Dog     |
| 505 | Spot     | Dalmation | Brown and red | Male   | 0      | Dog     |
+----+-----+-----+-----+-----+-----+-----+
rows in set (0.00 sec)

mysql>
```

—> again i have created another table and inserted the values and performed the commands

```
mysql> select *from shelter
-> order by id desc;
```

id	name	location
3	Give Life	Blue Hills
2	Adopt animals	Green Town
1	Animals for Homes	Smart city

3 rows in set (0.00 sec)

```
mysql> update animals set shelter=1;
Query OK, 9 rows affected (0.00 sec)
Rows matched: 9 Changed: 9 Warnings: 0
```

```
mysql> insert into shelter(id,name,location) values(2,'Adopt animals','Green Town');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into shelter(id,name,location) values(3,'Give Life','Blue Hills');
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into animals(id,name,breed,color,gender,status,species,shelter) values(301,'Snoopy','asian','Brown','Male',0,'Dog',2);
Query OK, 1 row affected (0.00 sec)
```

```
mysql> insert into animals(id,name,breed,color,gender,status,species,shelter) values(302,'tommy','asian','Brown','Male',1,'Cat',3);
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select *from shelter;
```

id	name	location
1	Animals for Homes	Smart city
2	Adopt animals	Green Town
3	Give Life	Blue Hills

3 rows in set (0.00 sec)

—> now we are creating another table to perform DATA CLEANING and TRANSFORMATION.

```
mysql> CREATE TABLE studentdata1 (  
-> id INT PRIMARY KEY AUTO_INCREMENT,  
-> name VARCHAR(255),  
-> age INT,  
-> grade VARCHAR(10)  
-> );  
Query OK, 0 rows affected (0.02 sec)  
  
mysql> insert into studentdata1(id,name,age,grade) values(null,'stella',20,'A+');  
Query OK, 1 row affected (0.04 sec)  
  
mysql> insert into studentdata1(id,name,age,grade) values(1,'appu',20,'A+');  
ERROR 1062 (23000): Duplicate entry '1' for key 'studentdata1.PRIMARY'  
mysql> insert into studentdata1(id,name,age,grade) values(5,'bob',21,'C');  
Query OK, 1 row affected (0.00 sec)  
  
mysql> insert into studentdata1(id,name,age,grade) values(6,'sunny',21,null);  
Query OK, 1 row affected (0.00 sec)  
  
mysql> insert into studentdata1(id,name,age,grade) values(7,null,21,'C');  
Query OK, 1 row affected (0.00 sec)  
  
mysql> SELECT * FROM studentdata1;  
+-----+-----+-----+-----+  
| id | name | age | grade |  
+-----+-----+-----+-----+  
| 1 | stella | 20 | A+ |  
| 5 | bob | 21 | C |  
| 6 | sunny | 21 | NULL |  
| 7 | NULL | 21 | C |  
+-----+-----+-----+-----+
```

—>To perform the data cleaning technique we have 4 steps.They are,

1. Check for Missing Values
2. Check for Duplicates
3. Standardizing and Transforming Data
4. Updating Data / Column Data Types

```
mysql> /* STEP-1 ----> Deleting the duplicate data*/  
mysql> select name,count(name) as Actual_count from studentdata1  
-> group by name  
-> having count(name)>1;  
Empty set (0.00 sec)  
  
mysql>  
mysql> with cte as  
-> (  
-> select name,  
-> ROW_NUMBER() over (partition by name order by name desc) as row_no  
-> from studentdata1)  
->  
-> select *from cte;  
+-----+-----+  
| name | row_no |  
+-----+-----+  
| NULL | 1 |  
| bob | 1 |  
| stella | 1 |  
| sunny | 1 |  
+-----+-----+  
4 rows in set (0.01 sec)
```

```
mysql> /*STEP-2 ----> removing the null values over here*/
```

```
mysql>
```

```
mysql> select *from studentdata1;
```

```
+-----+-----+-----+
| id | name | age | grade |
+-----+-----+-----+
| 1 | stella | 20 | A+ |
| 5 | bob | 21 | C |
| 6 | sunny | 21 | NULL |
| 7 | NULL | 21 | C |
+-----+-----+-----+
```

```
4 rows in set (0.00 sec)
```

```
mysql>
```

```
mysql> /*selecting the data where student name is null*/
```

```
mysql> select * from studentdata1
```

```
-> where name is null;
```

```
+-----+-----+-----+
| id | name | age | grade |
+-----+-----+-----+
| 7 | NULL | 21 | C |
+-----+-----+-----+
```

```
1 row in set (0.00 sec)
```

```
mysql>
```

```
mysql> delete from studentdata1
```

```
-> where name is null;
```

```
Query OK, 1 row affected (0.00 sec)
```

```
mysql> select * from studentdata1;
```

```
+-----+-----+-----+
| id | name | age | grade |
+-----+-----+-----+
| 1 | stella | 20 | A+ |
| 5 | bob | 21 | C |
| 6 | sunny | 21 | NULL |
+-----+-----+-----+
```

```
mysql> /*updating the null values where id and grade column is null*/
```

```
mysql>
```

```
mysql> select *from studentdata1 where id is null;
```

```
Empty set (0.00 sec)
```

```
mysql>
```

```
mysql> update studentdata1 set id=7
```

```
-> where id is null;
```

```
Query OK, 0 rows affected (0.00 sec)
```

```
Rows matched: 0 Changed: 0 Warnings: 0
```

```
mysql>
```

```
mysql> select *from studentdata1;
```

```
+-----+-----+-----+
| id | name | age | grade |
+-----+-----+-----+
| 1 | stella | 20 | A+ |
| 5 | bob | 21 | C |
| 6 | sunny | 21 | NULL |
+-----+-----+-----+
```

```
3 rows in set (0.00 sec)
```

```
mysql>
```

```
mysql> /* Capitalization of names in data*/
```

```
mysql> update studentdata1
```

```
-> set name=UPPER(name);
```

```
Query OK, 3 rows affected (0.01 sec)
```

```
Rows matched: 3 Changed: 3 Warnings: 0
```

```
mysql> select *from studentdata1;
```

```
+-----+-----+-----+
| id | name | age | grade |
+-----+-----+-----+
| 1 | STELLA | 20 | A+ |
| 5 | BOB | 21 | C |
| 6 | SUNNY | 21 | NULL |
+-----+-----+-----+
```

```
3 rows in set (0.00 sec)
```