

Shark Tank India

Data Cleaning Project in MySQL Workbench

```
1
2 • CREATE DATABASE SharkTankIndia;
3
4 • Use SharkTankIndia;
5
6 -- Tables 'deals' and 'presenters' imported with [Table Data Import Wizard]
7
8 -- view 'deals' table
9
10 • SELECT *
11 FROM deals;
```

	episode	pitch_no	company	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal
▶	1	1	BluePine Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A
	1	2	Booz scooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A
	1	3	Heart up my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A
	2	4	Tagz Foods	No Deal	Y	N	N	N	N	N/A	N/A
	2	5	Head and Heart	No Deal	N	N	N	N	N	N/A	N/A
	2	6	Agro tourism	No Deal	N	N	N	N	N	N/A	N/A
	3	7	Qzense Labs	No Deal	N	N	N	N	N	N/A	N/A
	3	8	Peeschute	75 lakhs for 6% equity	N	N	N	N	Y	N/A	N/A

```
12
13 -- view 'presenters' table
14
15 • SELECT *
16 FROM presenters;
17
```

	episode	pitch_no	company	no_of_presenters	male_presenter	female_presenter
▶	1	1	BluePine Industries	3	2	1
	1	2	Booz scooters	1	1	0
	1	3	Heart up my Sleeves	1	0	1
	2	4	Tagz Foods	2	2	0
	2	5	Head and Heart	4	1	3
	2	6	Agro tourism	2	1	1

```
18 -- make a new table 'sti_data' by joining deals and presenters table
```

```
19
```

```
20 • CREATE TABLE sti_data AS
```

```
21 SELECT d.*,
```

```
22         p.no_of_presenters, p.male_presenter, p.female_presenter
```

```
23 FROM deals d
```

```
24 join presenters p
```

```
25 on d.pitch_no = p.pitch_no;
```

```
26
```

```
27 • SELECT *
```

```
28 FROM sti_data;
```

```
29
```

	episode	pitch_no	company	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter
▶	1	1	BluePine Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A	3	2	1
	1	2	Booz scooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A	1	1	0
	1	3	Heart up my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A	1	0	1
	2	4	Tagz Foods	No Deal	Y	N	N	N	N	N/A	N/A	2	2	0
	2	5	Head and Heart	No Deal	N	N	N	N	N	N/A	N/A	4	1	3
	2	6	Agro tourism	No Deal	N	N	N	N	N	N/A	N/A	2	1	1

```
30 -- breaking down deal column into other columns so that data is more easy to understand
```

```
31
```

```
32 • ALTER TABLE sti_data
```

```
33 ADD deal_amount text;
```

```
34
```

```
35 • ALTER TABLE sti_data
```

```
36 ADD deal_equity text;
```

```
37
```

```
38 • ALTER TABLE sti_data
```

```
39 ADD deal_debt text;
```

```
40
```

```
41 • SELECT *
```

```
42 FROM sti_data;
```

```
43
```

	episode	pitch_no	company	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
▶	1	1	BluePine Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A	3	2	1	NULL	NULL	NULL
	1	2	Booz scooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A	1	1	0	NULL	NULL	NULL
	1	3	Heart up my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A	1	0	1	NULL	NULL	NULL
	2	4	Tagz Foods	No Deal	Y	N	N	N	N	N/A	N/A	2	2	0	NULL	NULL	NULL
	2	5	Head and Heart	No Deal	N	N	N	N	N	N/A	N/A	4	1	3	NULL	NULL	NULL
	2	6	Agro tourism	No Deal	N	N	N	N	N	N/A	N/A	2	1	1	NULL	NULL	NULL
	2	6	Agro Labs	No Deal	N	N	N	N	N	N/A	N/A	2	0	2	NULL	NULL	NULL

```

44 -- Extract deal Amount from deal columns to deal_amount column
45
46 • SET SQL_SAFE_UPDATES = 0;
47
48 • UPDATE sti_data
49 SET deal_amount = SUBSTRING_INDEX(deal, 'for', 1)
50 WHERE deal like '%equity%';
51
52 • UPDATE sti_data
53 SET deal_amount = REPLACE(
54     REPLACE(
55         REPLACE(
56             REPLACE(deal_amount,
57                 ' Lakhs', ''),
58                 ' lakhs', ''),
59                 ' lakh', ''),
60                 ' crore', '00'),
61                 ' Crore', '00');
62
63
64 • SELECT *
65 FROM sti_data;
66

```

hy	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A	3	2	1	75	NA	NA
ooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A	1	1	0	40	NA	NA
p my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A	1	0	1	25	NA	NA
ods	No Deal	Y	N	N	N	N	N/A	N/A	2	2	0	NA	NA	NA
nd Heart	No Deal	N	N	N	N	N	N/A	N/A	4	1	3	NA	NA	NA
urism	No Deal	N	N	N	N	N	N/A	N/A	2	1	1	NA	NA	NA
Labs	No Deal	N	N	N	N	N	N/A	N/A	2	0	2	NA	NA	NA

```

67 -- Extract Deal Equity from deal column to deal_equity column
68
69 • UPDATE sti_data
70 SET deal_equity = SUBSTRING_INDEX(
71     SUBSTRING_INDEX(deal,
72         "%", 1),
73     " ", -1)
74 WHERE deal like '%quity%';
75
76 • SELECT *
77 FROM sti_data;
78

```

hy	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A	3	2	1	75	18	NA
ooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A	1	1	0	40	50	NA
p my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A	1	0	1	25	30	NA
ods	No Deal	Y	N	N	N	N	N/A	N/A	2	2	0	NA	NA	NA
nd Heart	No Deal	N	N	N	N	N	N/A	N/A	4	1	3	NA	NA	NA
urism	No Deal	N	N	N	N	N	N/A	N/A	2	1	1	NA	NA	NA
Labs	No Deal	N	N	N	N	N	N/A	N/A	2	0	2	NA	NA	NA

```
79      -- Extract Deal Debt from deal column to deal_debt column
```

```
80
```

```
81 • UPDATE sti_data
```

```
82 SET deal_debt = SUBSTRING_INDEX(
```

```
83     SUBSTRING_INDEX(
```

```
84         SUBSTRING_INDEX(LOWER(deal),
```

```
85             'equity', -1),
```

```
86             ' ', -3),
```

```
87             'lakh', 1)
```

```
88 WHERE deal like '%debt%';
```

```
89
```

```
90 • SELECT *
```

```
91 FROM sti_data;
```

```
92
```

ty	deal	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
Industries	75 lakhs for 18% equity	Y	N	N	Y	Y	N/A	N/A	3	2	1	75	18	NULL
ooters	40 lakhs for 50% equity	Y	N	N	Y	N	N/A	N/A	1	1	0	40	50	NULL
p my Sleeves	25 lakhs for 30% equity	N	N	Y	Y	N	N/A	N/A	1	0	1	25	30	NULL
ods	No Deal	Y	N	N	N	N	N/A	N/A	2	2	0	NULL	NULL	NULL
nd Heart	No Deal	N	N	N	N	N	N/A	N/A	4	1	3	NULL	NULL	NULL
urism	No Deal	N	N	N	N	N	N/A	N/A	2	1	1	NULL	NULL	NULL
Labs	No Deal	N	N	N	N	N	N/A	N/A	2	0	2	NULL	NULL	NULL
ate	75 lakhs for 6% equity	N	N	N	N	Y	N/A	N/A	1	1	0	75	6	NULL

```
93 -- cleaning sti_data table
```

```
94
```

```
95 • ALTER TABLE sti_data
```

```
96 drop column deal;
```

```
97
```

```
98 • UPDATE sti_data
```

```
99 SET deal_debt = 25
```

```
100 where deal_debt = 'â,â25 ';
```

```
101
```

```
102 -- replace y with 1 (True)
```

```
103
```

```
104 • UPDATE sti_data
```

```
105 SET ashneer = CASE
```

```
106     WHEN ashneer = 'Y' THEN 1
```

```
107     WHEN ashneer = 'N' THEN 0
```

```
108     WHEN ashneer = 'N/A' THEN 0
```

```
109 END,
```

```
110 SET namita = CASE
```

```
111     WHEN namita = 'Y' THEN 1
```

```
112     WHEN namita = 'N' THEN 0
```

```
113     WHEN namita = 'N/A' THEN 0
```

```
114 END,
```

```

115 anupam = CASE
116     WHEN anupam = 'Y' THEN 1
117     WHEN anupam = 'N' THEN 0
118     WHEN anupam = 'N/A' THEN 0
119 END,
120 vineeta = CASE
121     WHEN vineeta = 'Y' THEN 1
122     WHEN vineeta = 'N' THEN 0
123     WHEN vineeta = 'N/A' THEN 0
124 END,
125 aman = CASE
126     WHEN aman = 'Y' THEN 1
127     WHEN aman = 'N' THEN 0
128     WHEN aman = 'N/A' THEN 0
129 END,
130 peyush = CASE
131     WHEN peyush = 'Y' THEN 1
132     WHEN peyush = 'N' THEN 0
133     WHEN peyush = 'N/A' THEN 0
134 END,
135 ghazal = CASE
136     WHEN ghazal = 'Y' THEN 1
137     WHEN ghazal = 'N' THEN 0
138     WHEN ghazal = 'N/A' THEN 0
139 END;
140
141 • SELECT*
142 FROM sti_data;
143

```

	episode	pitch_no	company	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
▶	1	1	BluePine Industries	1	0	0	1	1	0	0	3	2	1	75	18	NULL
	1	2	Booz scooters	1	0	0	1	0	0	0	1	1	0	40	50	NULL
	1	3	Heart up my Sleeves	0	0	1	1	0	0	0	1	0	1	25	30	NULL
	2	4	Tagz Foods	1	0	0	0	0	0	0	2	2	0	NULL	NULL	NULL
	2	5	Head and Heart	0	0	0	0	0	0	0	4	1	3	NULL	NULL	NULL
	2	6	Agro tourism	0	0	0	0	0	0	0	2	1	1	NULL	NULL	NULL
	3	7	Qzense Labs	0	0	0	0	0	0	0	2	0	2	NULL	NULL	NULL
	3	8	Peeschute	0	0	0	0	1	0	0	1	1	0	75	6	NULL
	3	9	NOCO	0	0	0	1	0	0	0	2	2	0	20	15	30

```

144 -- Fill null values with 0
145
146 • UPDATE sti_data
147 Set deal_amount = 0
148 WHERE deal_amount is null;
149
150 • UPDATE sti_data
151 Set deal_equity = 0
152 WHERE deal_equity is null;
153
154 • UPDATE sti_data
155 Set deal_debt = 0
156 WHERE deal_debt is null;
157
158 • SELECT *
159 FROM sti_data;
160

```

	episode	pitch_no	company	ashneer	namita	anupam	vineeta	aman	peyush	ghazal	no_of_presenters	male_presenter	female_presenter	deal_amount	deal_equity	deal_debt
▶	1	1	BluePine Industries	1	0	0	1	1	0	0	3	2	1	75	18	0
	1	2	Booz scooters	1	0	0	1	0	0	0	1	1	0	40	50	0
	1	3	Heart up my Sleeves	0	0	1	1	0	0	0	1	0	1	25	30	0
	2	4	Tagz Foods	1	0	0	0	0	0	0	2	2	0	0	0	0
	2	5	Head and Heart	0	0	0	0	0	0	0	4	1	3	0	0	0
	2	6	Agro tourism	0	0	0	0	0	0	0	2	1	1	0	0	0
	3	7	Qzense Labs	0	0	0	0	0	0	0	2	0	2	0	0	0
	3	8	Peeschute	0	0	0	0	1	0	0	1	1	0	75	6	0
	3	9	NOCOD	0	0	0	1	0	0	0	2	2	0	20	15	30
	4	10	Cosiq	0	0	1	1	0	0	0	2	1	1	0	0	0
	4	11	JhaJi Achaar	0	0	0	0	0	0	0	2	0	2	0	0	0

```
161 -- Trim columns we extracted from deal column
```

```
162
```

```
163 • UPDATE sti_data
```

```
164 SET deal_amount = trim(deal_amount);
```

```
165
```

```
166 • UPDATE sti_data
```

```
167 SET deal_equity = trim(deal_equity);
```

```
168
```

```
169 • UPDATE sti_data
```

```
170 SET deal_debt = trim(deal_debt);
```

```
171
```

```
172 -- Change data types
```

```
173
```

```
174 • ALTER TABLE sti_data
```

```
175 MODIFY COLUMN ashneer INT;
```

```
176
```

```
177 • ALTER TABLE sti_data
```

```
178 MODIFY COLUMN namita INT;
```

```
179
```

```
180 • ALTER TABLE sti_data
```

```
181 MODIFY COLUMN anupam INT;
```

```
182
```

```

183 • ALTER TABLE sti_data
184     MODIFY COLUMN vineeta INT;
185
186 • ALTER TABLE sti_data
187     MODIFY COLUMN aman INT;
188
189 • ALTER TABLE sti_data
190     MODIFY COLUMN peyush INT;
191
192 • ALTER TABLE sti_data
193     MODIFY COLUMN ghazal INT;
194
195 • ALTER TABLE sti_data
196     MODIFY COLUMN deal_amount FLOAT;
197
198 • ALTER TABLE sti_data
199     MODIFY COLUMN deal_equity FLOAT;
200
201 • ALTER TABLE sti_data
202     MODIFY COLUMN deal_debt FLOAT;
203
204 • SHOW COLUMNS from sti_data;
205

```

	Field	Type	Null	Key	Default	Extra
►	episode	int	YES		NULL	
	pitch_no	int	YES		NULL	
	company	text	YES		NULL	
	ashneer	int	YES		NULL	
	namita	int	YES		NULL	
	anupam	int	YES		NULL	
	vineeta	int	YES		NULL	
	aman	int	YES		NULL	
	peyush	int	YES		NULL	
	ghazal	int	YES		NULL	
	no_of_p...	int	YES		NULL	
	male_pr...	int	YES		NULL	
	female_...	int	YES		NULL	
	deal_am...	float	YES		NULL	
	deal_eq...	float	YES		NULL	
	deal_debt	float	YES		NULL	