CHAPTER –1

NORMALIZATION

- **It is the process of creating an efficient database system from a denormalized dataset to reduce redundancy and perform analysis using SQL queries.**

1.1. Original Columns:

- First Name
- Middle Name
- Last Name
- Candidate Source category name
- Job Posting title
- Total Experience
- City

1.2. 3NF Schema:

Table 1: Candidates

- CandidateID (Primary Key)
- FirstName
- MiddleName
- LastName
- ExperienceID (Foreign Key referencing Experience table)

Table 2: Experiences

- ExperienceID (Primary Key)
- Position
- Years
- SourceOfApplication
- City

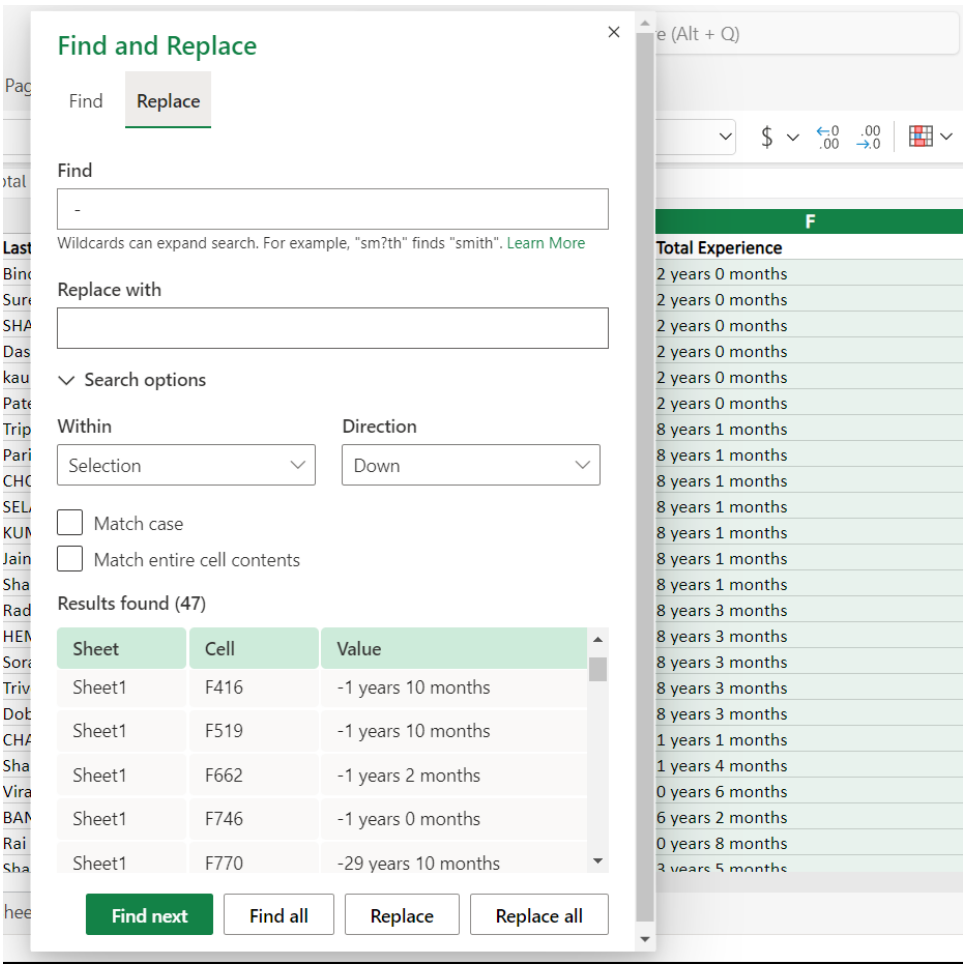
CHAPTER-2

2.1. PREPROCESSING DATASET

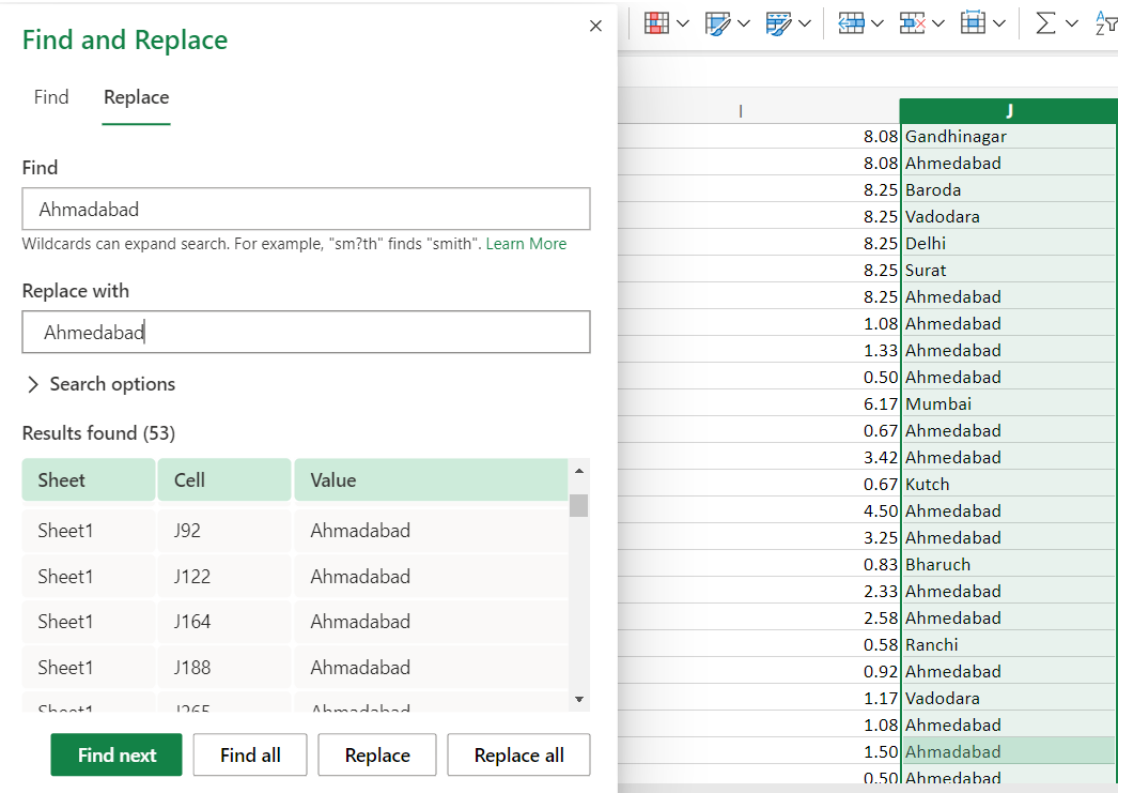
2.1.1. Replacing the unnecessary symbol '-' from column [Experience].

* A simple find and replace option within the section used to do so.

"Minus sign is replaced using a space".



2.1.2. Replacing all the entries of column [City] as ‘Ahmadabad’ to “Ahmedabad”



2.1.3. Converting years and months from column [Experience] to numeric ‘year’ value.

--> This is done to perform a better analysis over the experience column as it will be a numeric variable.

Benefits:

statistical analysis, visualization, modeling compatibility, correlation exploration, feature engineering possibilities, time series analysis, support for machine learning models, and streamlined data preprocessing.

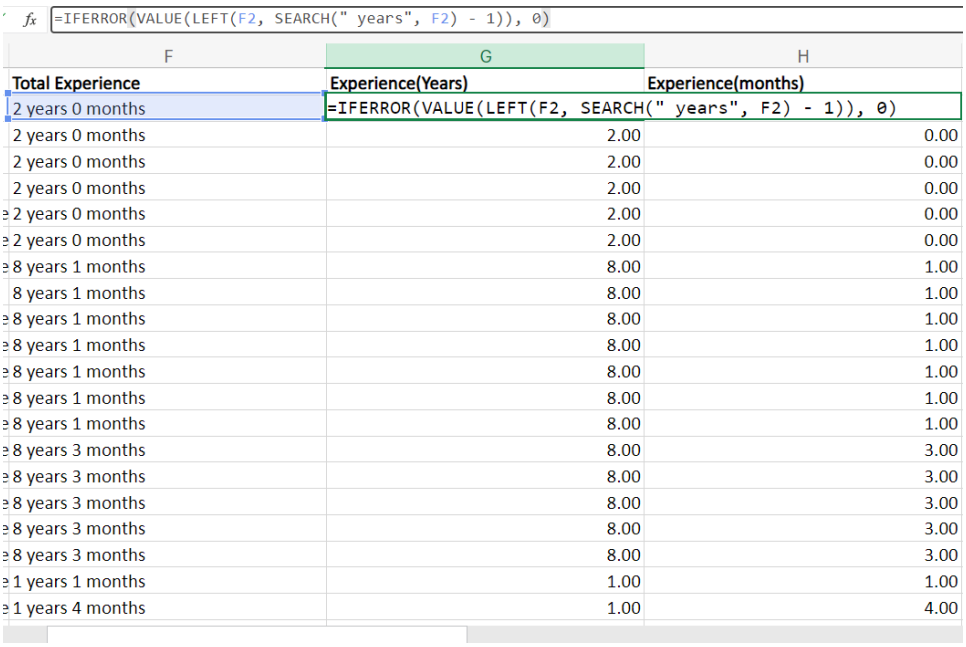
Formulas used:

→ **For extracting years: =** IFERROR (VALUE (LEFT (F2, SEARCH (" years", F2) - 1)), 0)

→ **For extracting months: =** IFERROR (VALUE (MID (F2, SEARCH (" years", F2) + 6, SEARCH (" months", F2) - SEARCH (" years", F2) - 7)), 0)

→ **For total years:** =G2 + H2 / 12

Fig:1



=IFERROR(VALUE(LEFT(F2, SEARCH(" years", F2) - 1)), 0)

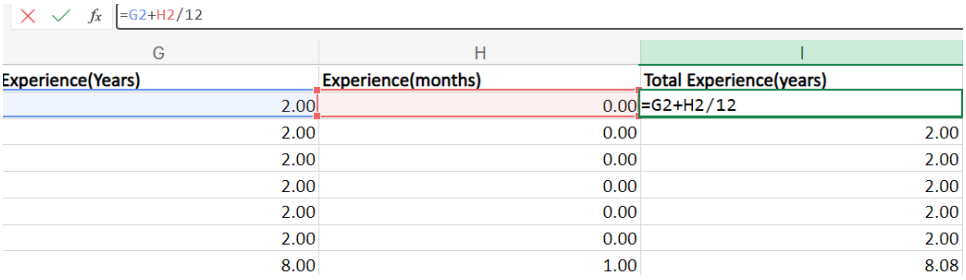
F	G	H
Total Experience	Experience(Years)	Experience(months)
2 years 0 months	=IFERROR(VALUE(LEFT(F2, SEARCH(" years", F2) - 1)), 0)	
2 years 0 months	2.00	0.00
2 years 0 months	2.00	0.00
2 years 0 months	2.00	0.00
2 years 0 months	2.00	0.00
2 years 0 months	2.00	0.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 1 months	8.00	1.00
8 years 3 months	8.00	3.00
8 years 3 months	8.00	3.00
8 years 3 months	8.00	3.00
8 years 3 months	8.00	3.00
8 years 3 months	8.00	3.00
1 years 1 months	1.00	1.00
1 years 4 months	1.00	4.00

Fig:2

", F2) - SEARCH(" years", F2) - 6)), 0)

H	I	J
Experience(months)	Total Experience(years)	City
=IFERROR(VALUE(MID(F2, SEARCH(" years", F2) + 6, SEARCH(" months", F2) - SEARCH(" years", F2) - 6)), 0)		
0.00	2.00	Ahmedabad
0.00	2.00	Indore
0.00	2.00	Ahmedabad
0.00	2.00	Ahmedabad
1.00	8.08	Ahmedabad
1.00	8.08	Ahmedabad
1.00	8.08	Veraval
1.00	8.08	Baroda
1.00	8.08	Bokaro
1.00	8.08	Gandhinagar
1.00	8.08	Ahmedabad
3.00	8.25	Baroda
3.00	8.25	Vadodara
3.00	8.25	Delhi
3.00	8.25	Surat
3.00	8.25	Ahmedabad
1.00	1.08	Ahmedabad
4.00	1.33	Ahmedabad

Fig: 3

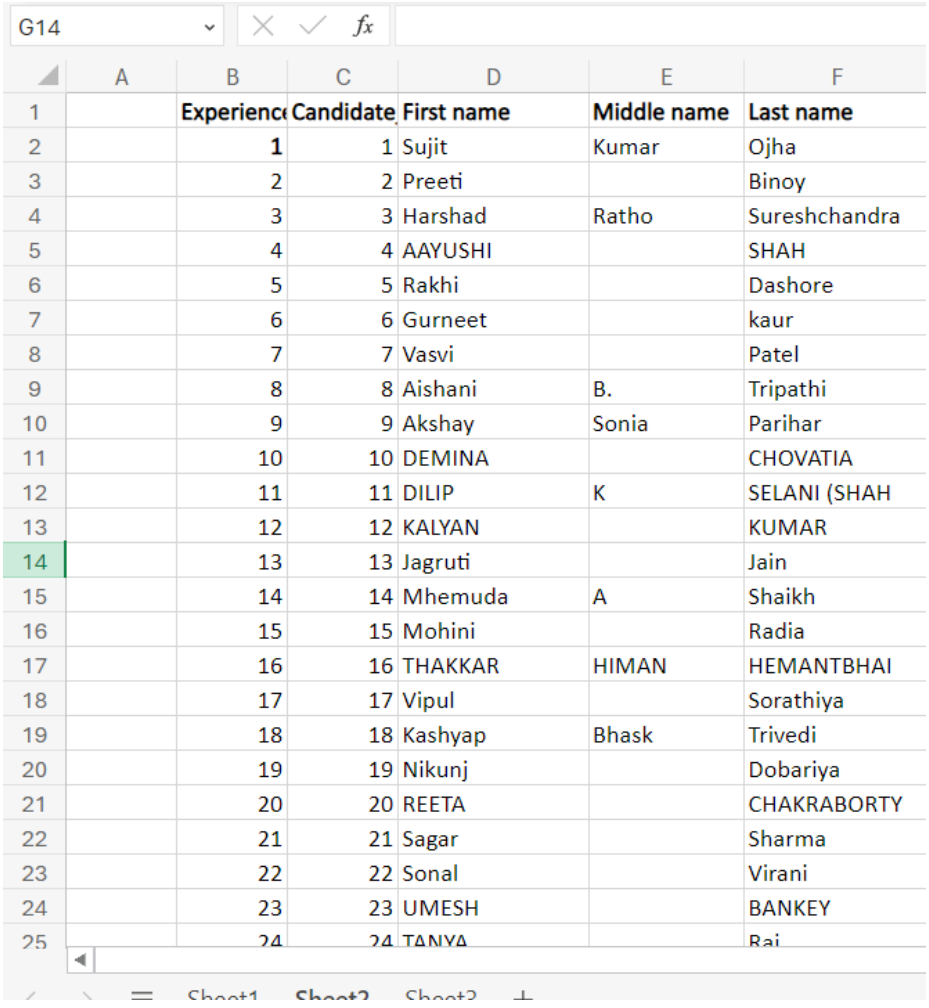

=G2+H2/12
G
H
I
Experience(Years)
Experience(months)
Total Experience(years)
2.00
0.00
=G2+H2/12
2.00
0.00
2.00
2.00
0.00
2.00
2.00
0.00
2.00
2.00
0.00
2.00
2.00
0.00
2.00
8.00
1.00
8.08

CHAPTER – 3

CREATING NORMALIZED DATABASE SYSTEM USING MYSQLWORKBENCH

3.1. Create different excel sheets from existing Sheet using 3-NF form as described in chapter 1.

- **Step 1: Add an extra column "candidate_id" which will be used as a Primary key in MySql in Candidate.excel sheet.**

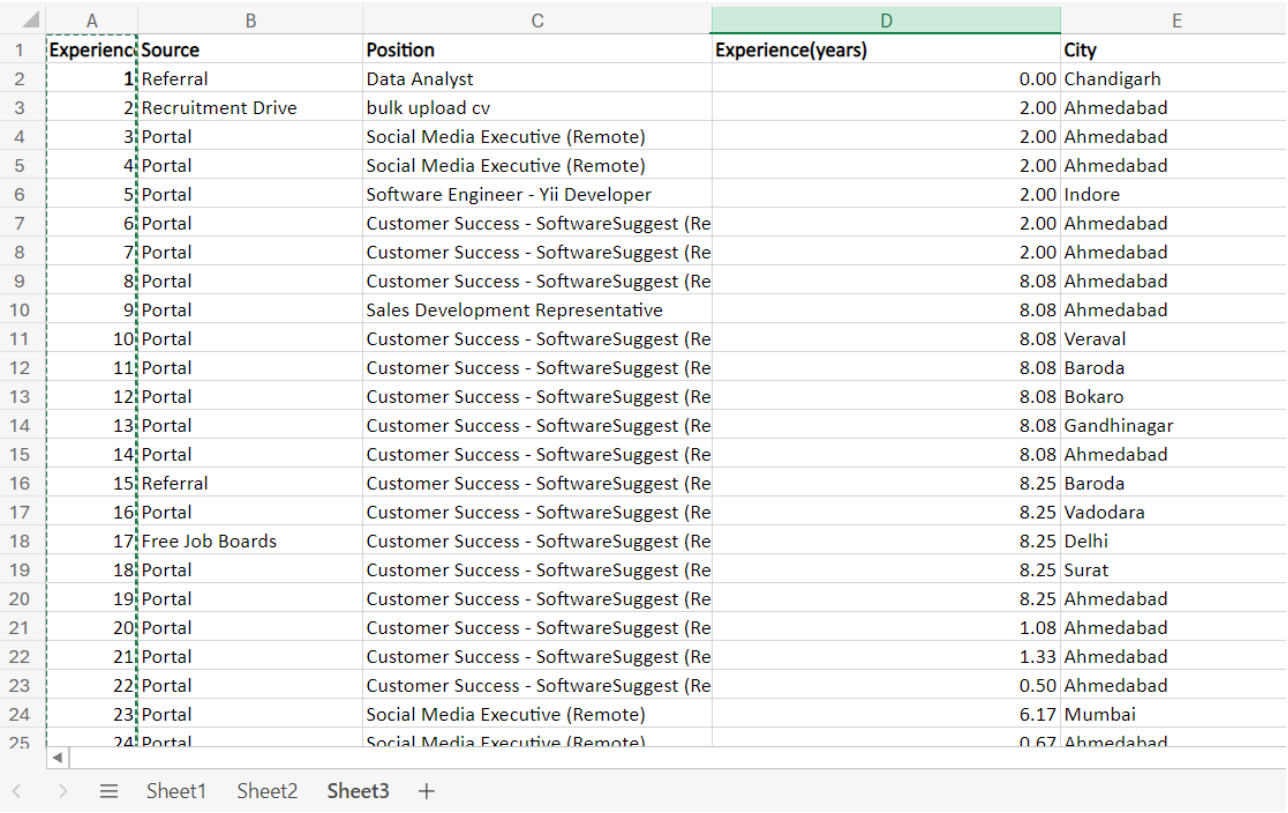


G14

	A	B	C	D	E	F
1		Experienc	Candidate_	First name	Middle name	Last name
2		1	1	Sujit	Kumar	Ojha
3		2	2	Preeti		Binoy
4		3	3	Harshad	Ratho	Sureshchandra
5		4	4	AAYUSHI		SHAH
6		5	5	Rakhi		Dashore
7		6	6	Gurneet		kaur
8		7	7	Vasvi		Patel
9		8	8	Aishani	B.	Tripathi
10		9	9	Akshay	Sonia	Parihar
11		10	10	DEMINA		CHOVATIA
12		11	11	DILIP	K	SELANI (SHAH
13		12	12	KALYAN		KUMAR
14		13	13	Jagruti		Jain
15		14	14	Mhemuda	A	Shaikh
16		15	15	Mohini		Radia
17		16	16	THAKKAR	HIMAN	HEMANTBHAI
18		17	17	Vipul		Sorathiya
19		18	18	Kashyap	Bhask	Trivedi
20		19	19	Nikunj		Dobariya
21		20	20	REETA		CHAKRABORTY
22		21	21	Sagar		Sharma
23		22	22	Sonal		Virani
24		23	23	UMESH		BANKEY
25		24	24	TANYA		Rai

Sheet1 Sheet2 Sheet3

- **Step 2: Add a column “Experience_key” which will be used as primary key for Experience Table in MySQL, in Experience Excel sheet.**



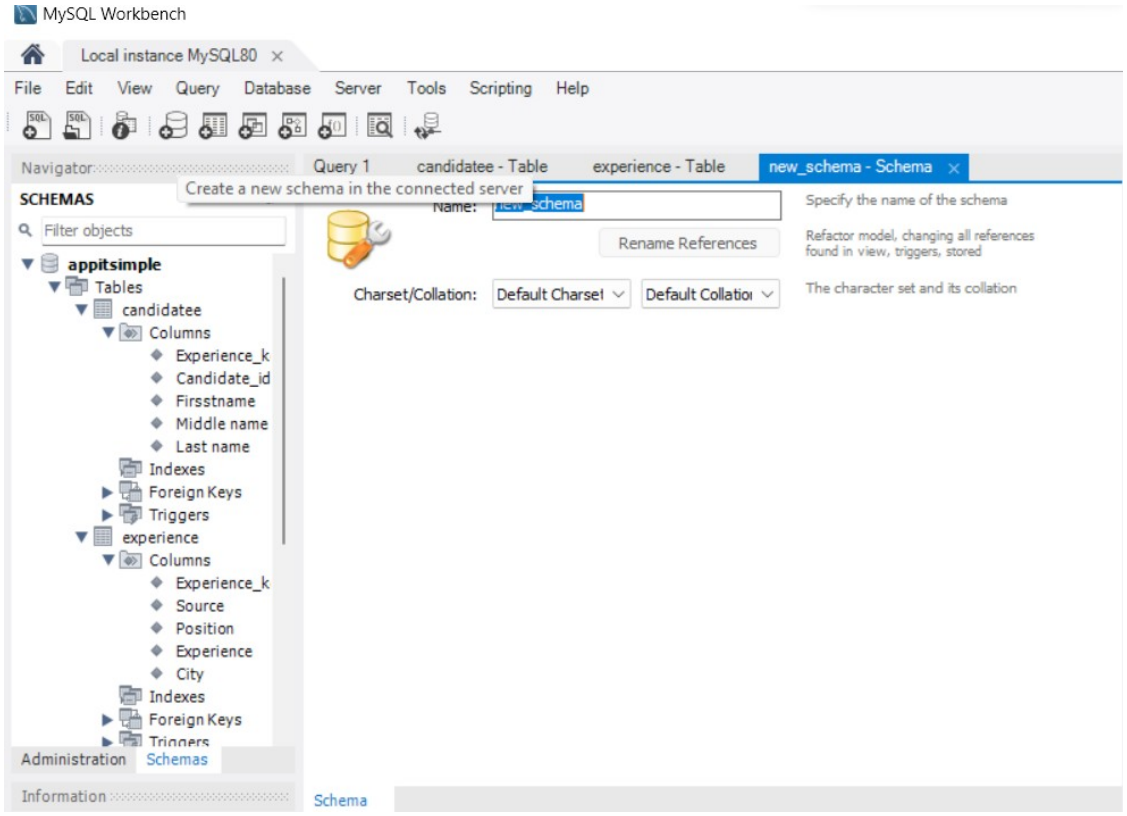
	A	B	C	D	E
1	Experienc	Source	Position	Experience(years)	City
2	1	Referral	Data Analyst	0.00	Chandigarh
3	2	Recruitment Drive	bulk upload cv	2.00	Ahmedabad
4	3	Portal	Social Media Executive (Remote)	2.00	Ahmedabad
5	4	Portal	Social Media Executive (Remote)	2.00	Ahmedabad
6	5	Portal	Software Engineer - Yii Developer	2.00	Indore
7	6	Portal	Customer Success - SoftwareSuggest (Re	2.00	Ahmedabad
8	7	Portal	Customer Success - SoftwareSuggest (Re	2.00	Ahmedabad
9	8	Portal	Customer Success - SoftwareSuggest (Re	8.08	Ahmedabad
10	9	Portal	Sales Development Representative	8.08	Ahmedabad
11	10	Portal	Customer Success - SoftwareSuggest (Re	8.08	Veraval
12	11	Portal	Customer Success - SoftwareSuggest (Re	8.08	Baroda
13	12	Portal	Customer Success - SoftwareSuggest (Re	8.08	Bokaro
14	13	Portal	Customer Success - SoftwareSuggest (Re	8.08	Gandhinagar
15	14	Portal	Customer Success - SoftwareSuggest (Re	8.08	Ahmedabad
16	15	Referral	Customer Success - SoftwareSuggest (Re	8.25	Baroda
17	16	Portal	Customer Success - SoftwareSuggest (Re	8.25	Vadodara
18	17	Free Job Boards	Customer Success - SoftwareSuggest (Re	8.25	Delhi
19	18	Portal	Customer Success - SoftwareSuggest (Re	8.25	Surat
20	19	Portal	Customer Success - SoftwareSuggest (Re	8.25	Ahmedabad
21	20	Portal	Customer Success - SoftwareSuggest (Re	1.08	Ahmedabad
22	21	Portal	Customer Success - SoftwareSuggest (Re	1.33	Ahmedabad
23	22	Portal	Customer Success - SoftwareSuggest (Re	0.50	Ahmedabad
24	23	Portal	Social Media Executive (Remote)	6.17	Mumbai
25	24	Portal	Social Media Executive (Remote)	0.67	Ahmedabad

Sheet1 Sheet2 Sheet3 +

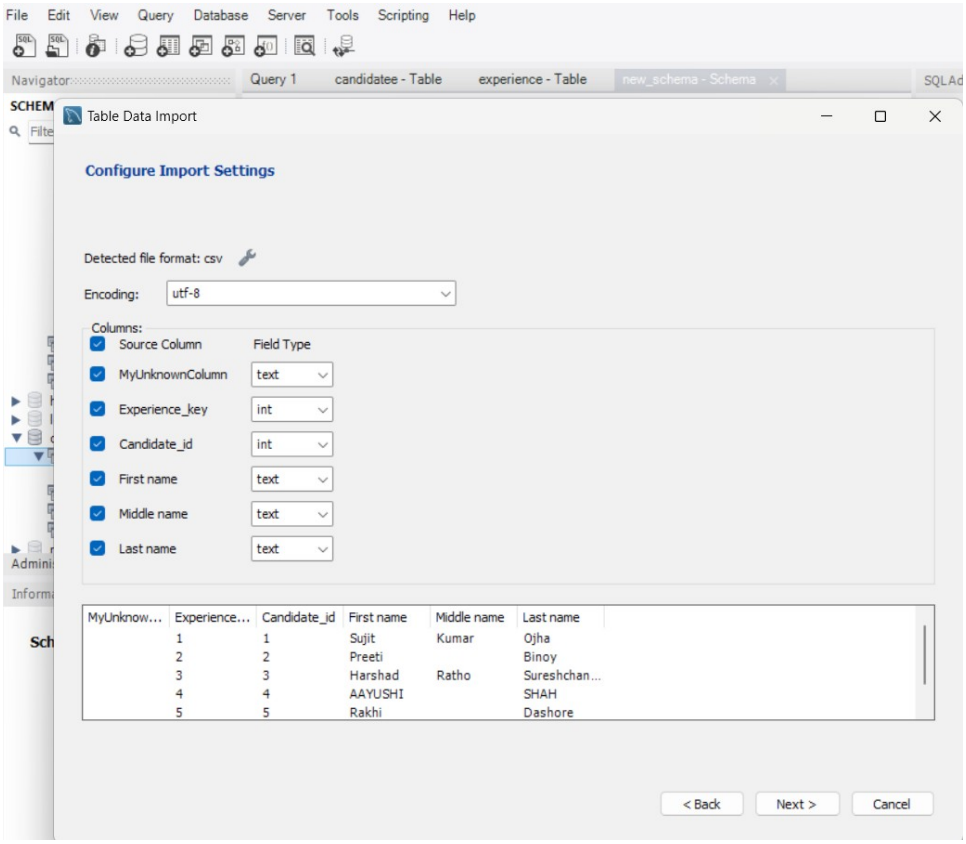
- **Step 3: Copy the column “Experience_key” from Experience.excel sheet to Candidate.excel sheet that we ceated as we will use it as a reference to Experience table.**

3.2. LOAD DATASET TO MYSQL WORK ENVIRONMENT

- **STEP 1: Create a new Schema in MYSQL server**

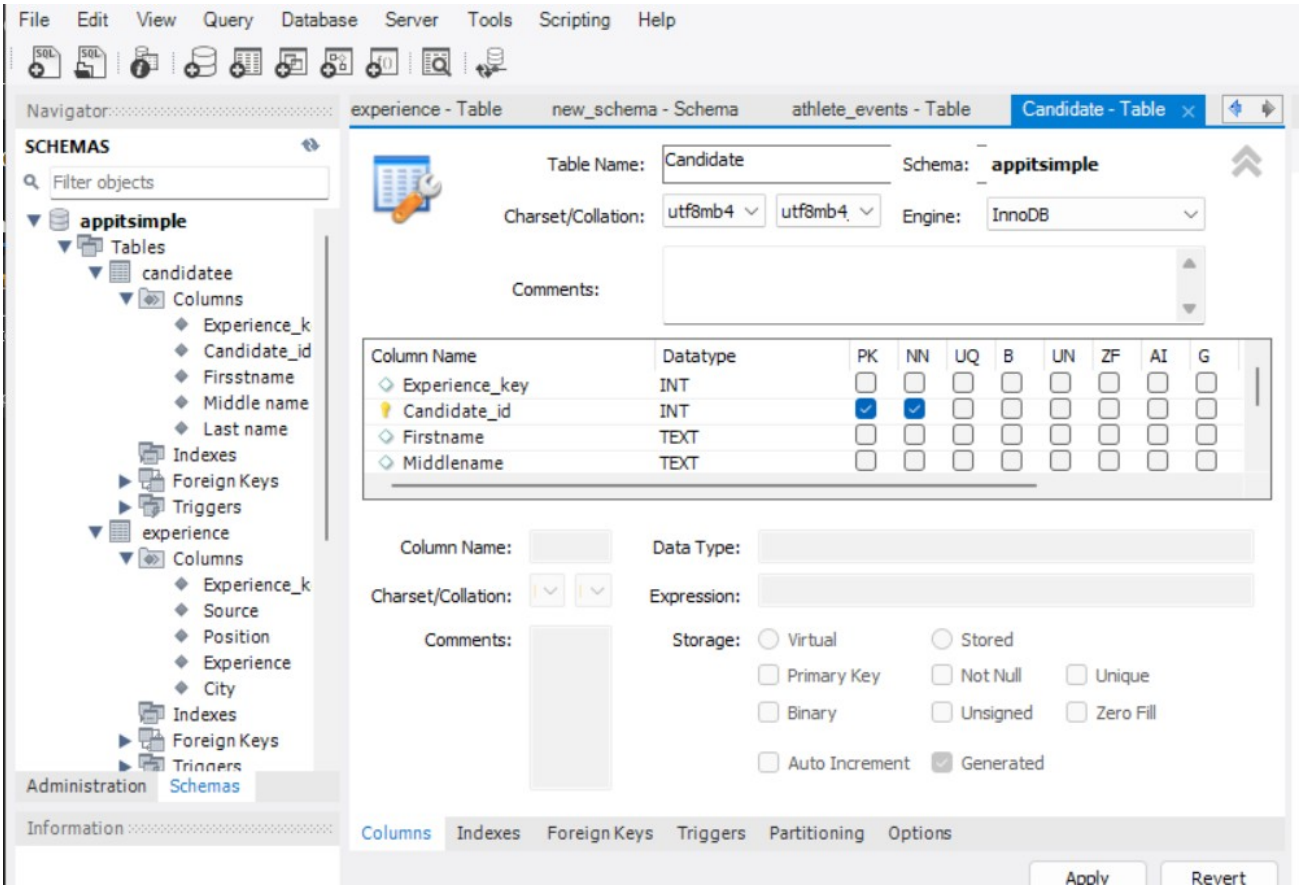


- **STEP 2: Import table data using import table data wizard**



STEP 3: DEFINE THE STRUCTURE OF TABLE

→ **Candidate_id as primary key in Candidate table**
→ **Experience_key as Foreign key in Candidate table**
→ **Experience_key as Primary key in Experience table**



CHAPTER – 4

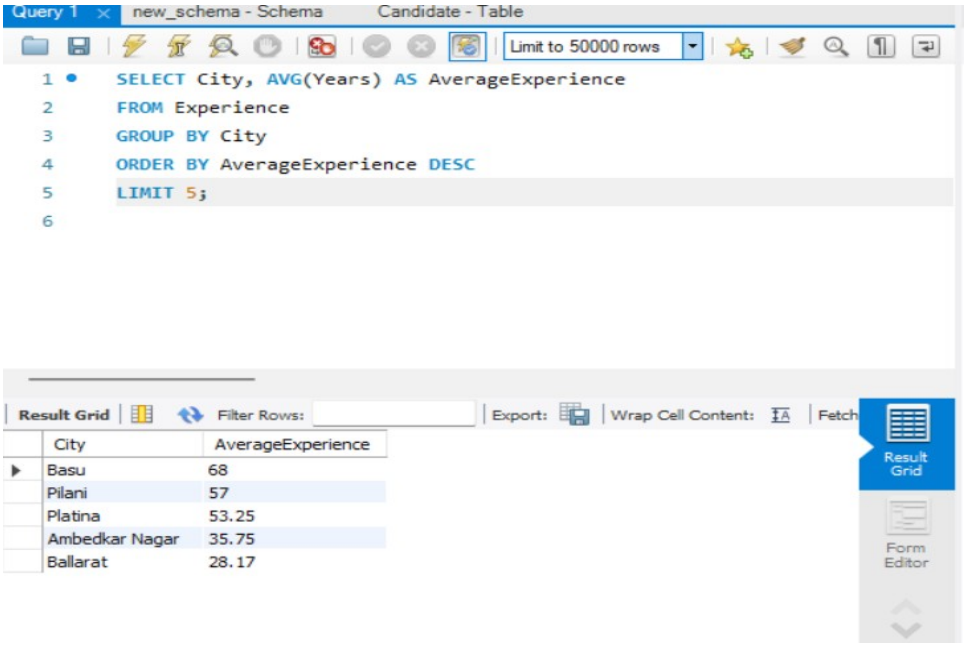
4.1. PERFORMING QUERIES FOR ANALYSIS & GETTING INSIGHTS

1. Find the top 5 cities with the highest average experience.

Query:

```
SELECT City, AVG(Years) AS AverageExperience
FROM Experience
GROUP BY City
ORDER BY AverageExperience DESC
LIMIT 5;
```

Ans:

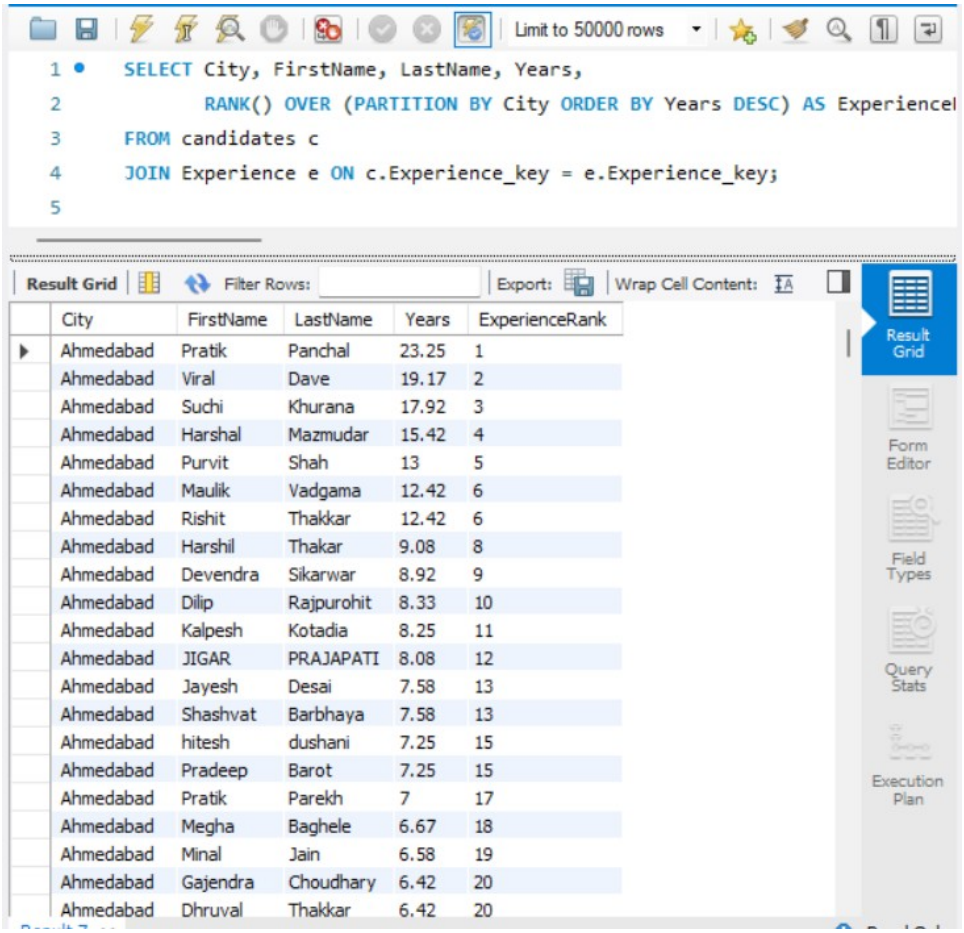


2. Rank Candidates based on their Experiences within each city .

Query:

```
SELECT City, FirstName, LastName, Years,
 RANK() OVER (PARTITION BY City ORDER BY Years DESC) AS ExperienceRank
FROM Candidates c
JOIN Experience e ON c.Experience_key = e.Experience_key;
```

Ans:



```
1  SELECT City, FirstName, LastName, Years,
2         RANK() OVER (PARTITION BY City ORDER BY Years DESC) AS Experience
3  FROM candidates c
4  JOIN Experience e ON c.Experience_key = e.Experience_key;
5
```

Result Grid | Filter Rows: | Export: | Wrap Cell Content:

City	FirstName	LastName	Years	ExperienceRank
Ahmedabad	Pratik	Panchal	23.25	1
Ahmedabad	Viral	Dave	19.17	2
Ahmedabad	Suchi	Khurana	17.92	3
Ahmedabad	Harshal	Mazmudar	15.42	4
Ahmedabad	Purvit	Shah	13	5
Ahmedabad	Maulik	Vadgama	12.42	6
Ahmedabad	Rishit	Thakkar	12.42	6
Ahmedabad	Harshil	Thakar	9.08	8
Ahmedabad	Devendra	Sikarwar	8.92	9
Ahmedabad	Dilip	Rajpurohit	8.33	10
Ahmedabad	Kalpesh	Kotadia	8.25	11
Ahmedabad	JIGAR	PRAJAPATI	8.08	12
Ahmedabad	Jayesh	Desai	7.58	13
Ahmedabad	Shashvat	Barbhaya	7.58	13
Ahmedabad	hitesh	dushani	7.25	15
Ahmedabad	Pradeep	Barot	7.25	15
Ahmedabad	Pratik	Parekh	7	17
Ahmedabad	Megha	Baghele	6.67	18
Ahmedabad	Minal	Jain	6.58	19
Ahmedabad	Gajendra	Choudhary	6.42	20
Ahmedabad	Dhruval	Thakkar	6.42	20

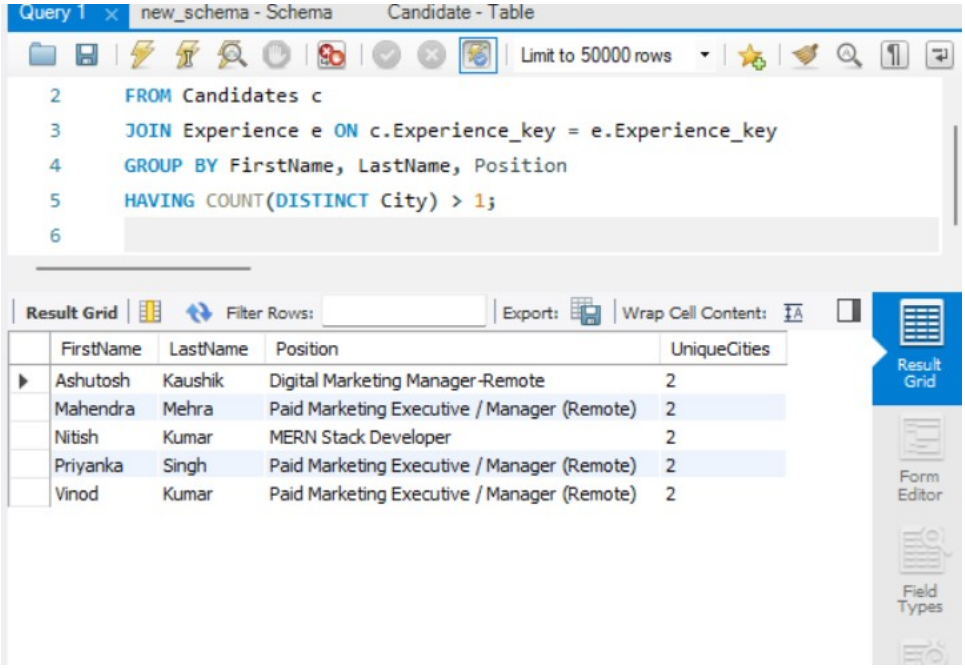
Result Grid | Form Editor | Field Types | Query Stats | Execution Plan

3. Find the name of Candidates who applied for same position from different city.

Query:

```
SELECT FirstName, LastName, Position, COUNT(DISTINCT City) AS UniqueCities
FROM Candidates c
JOIN Experience e ON c.Experience_key = e.Experience_key
GROUP BY FirstName, LastName, Position
HAVING COUNT(DISTINCT City) > 1;
```

Ans:

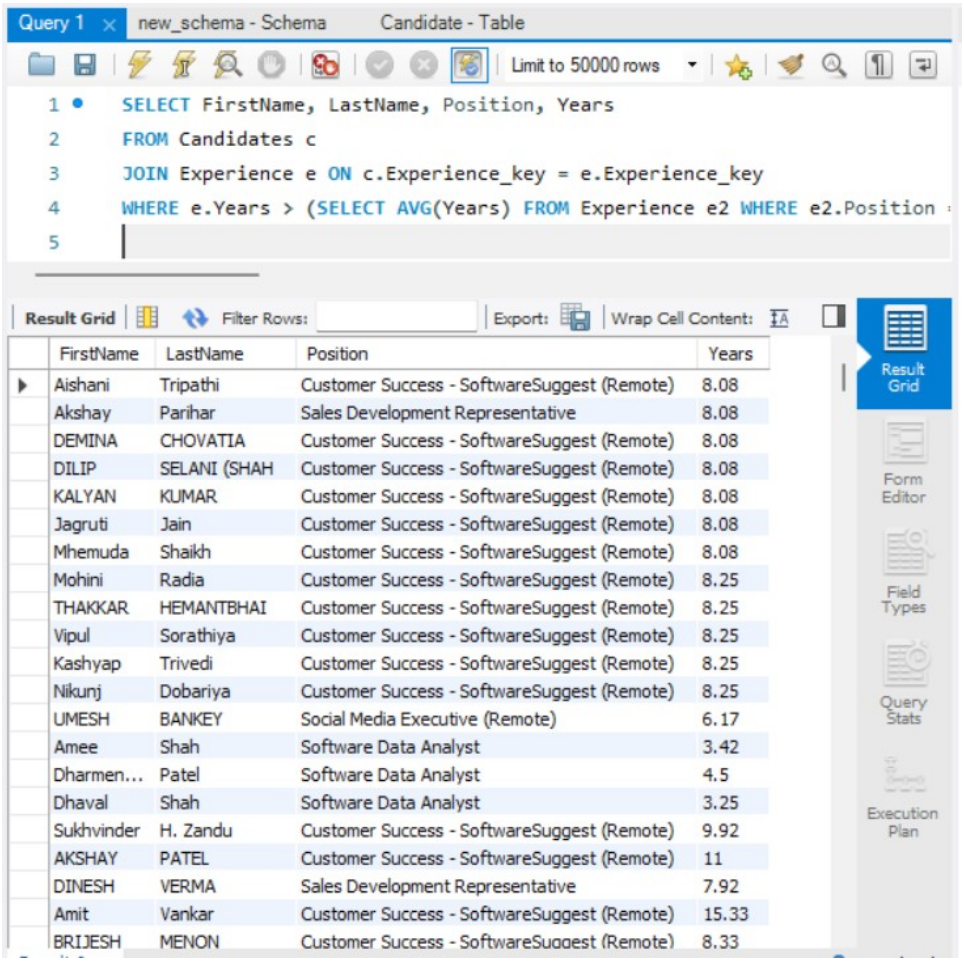


4. Find Candidates who have experience greater than the average experience of candidates applied for same position.

Query:

```
SELECT FirstName, LastName, Position, Years
FROM Candidates c
JOIN Experience e ON c.Experience_key = e.Experience_key
WHERE e.Years > (SELECT AVG(Years) FROM Experience e2 WHERE e2.Position = e.Position);
```

Ans:



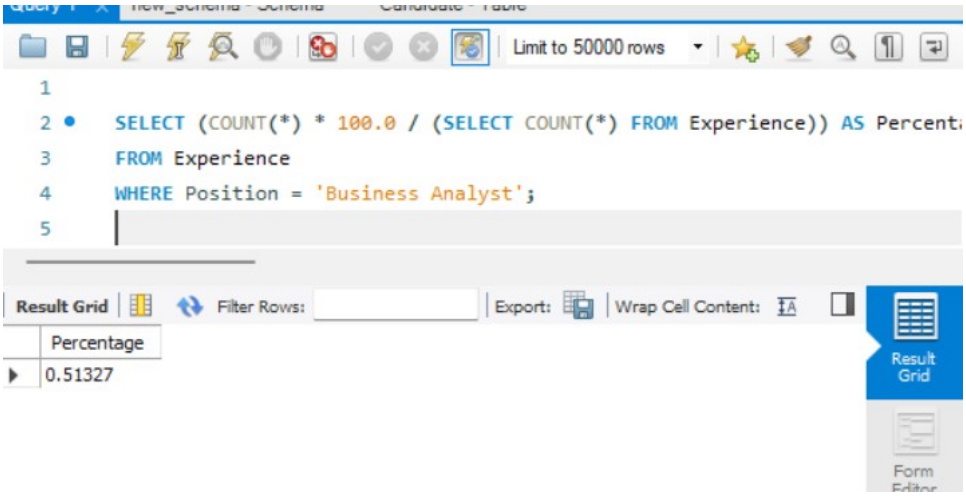
FirstName	LastName	Position	Years
Aishani	Tripathi	Customer Success - SoftwareSuggest (Remote)	8.08
Akshay	Parihar	Sales Development Representative	8.08
DEMINA	CHOVATIA	Customer Success - SoftwareSuggest (Remote)	8.08
DILIP	SELANI (SHAH	Customer Success - SoftwareSuggest (Remote)	8.08
KALYAN	KUMAR	Customer Success - SoftwareSuggest (Remote)	8.08
Jagruti	Jain	Customer Success - SoftwareSuggest (Remote)	8.08
Mhemuda	Shaikh	Customer Success - SoftwareSuggest (Remote)	8.08
Mohini	Radia	Customer Success - SoftwareSuggest (Remote)	8.25
THAKKAR	HEMANTBHAI	Customer Success - SoftwareSuggest (Remote)	8.25
Vipul	Sorathiya	Customer Success - SoftwareSuggest (Remote)	8.25
Kashyap	Trivedi	Customer Success - SoftwareSuggest (Remote)	8.25
Nikunj	Dobariya	Customer Success - SoftwareSuggest (Remote)	8.25
UMESH	BANKEY	Social Media Executive (Remote)	6.17
Amee	Shah	Software Data Analyst	3.42
Dharmen...	Patel	Software Data Analyst	4.5
Dhaval	Shah	Software Data Analyst	3.25
Sukhvinder	H. Zandu	Customer Success - SoftwareSuggest (Remote)	9.92
AKSHAY	PATEL	Customer Success - SoftwareSuggest (Remote)	11
DINESH	VERMA	Sales Development Representative	7.92
Amit	Vankar	Customer Success - SoftwareSuggest (Remote)	15.33
BRIJESH	MENON	Customer Success - SoftwareSuggest (Remote)	8.33

5. Calculate the overall percentage of the Candidates applied for “Business Analyst”.

Query:

```
SELECT (COUNT(*) * 100.0 / (SELECT COUNT(*) FROM Experience)) AS Percentage
FROM Experience
WHERE Position = 'Business Analyst
```

Ans:

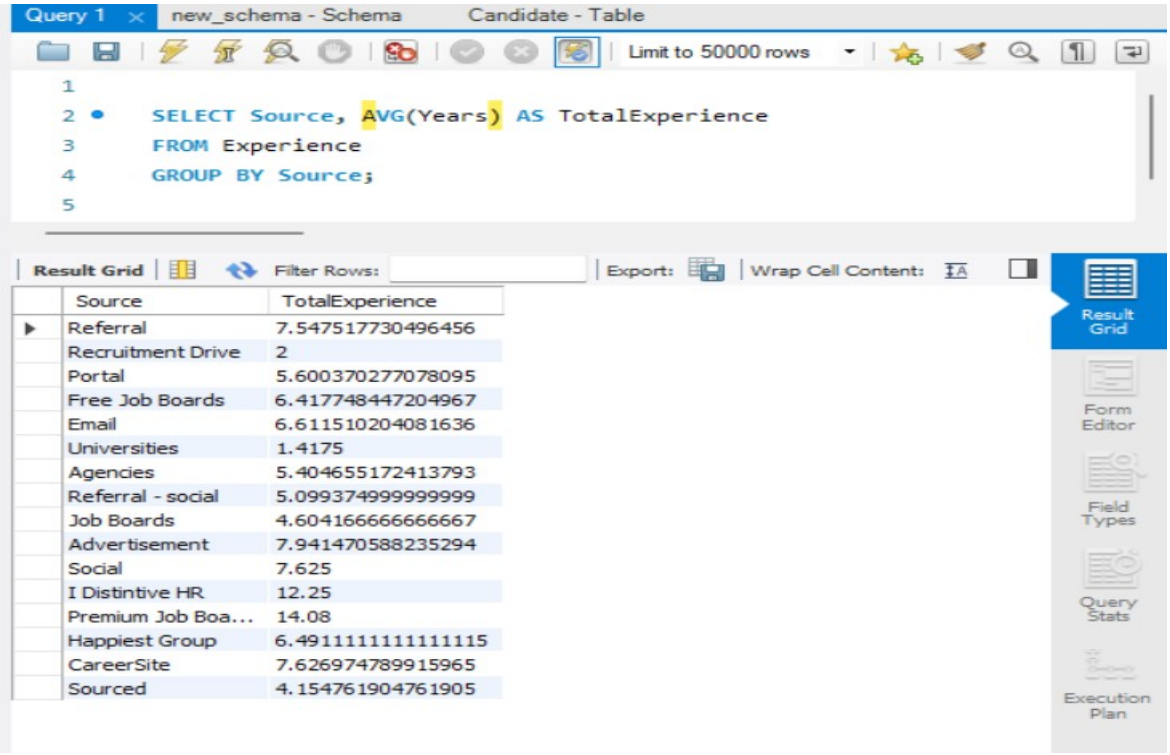


6. Calculate average experience of candidates who applied from each source.

Query:

```
SELECT Source, AVG(Years) AS TotalExperience
FROM Experience
GROUP BY Source;
```

Ans:



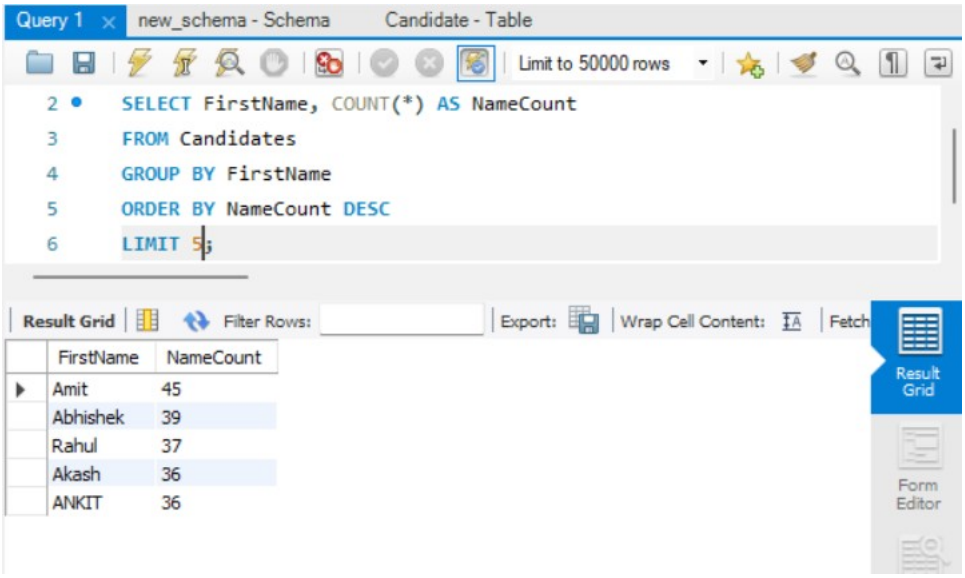
Source	TotalExperience
Referral	7.547517730496456
Recruitment Drive	2
Portal	5.600370277078095
Free Job Boards	6.417748447204967
Email	6.611510204081636
Universities	1.4175
Agencies	5.404655172413793
Referral - social	5.099374999999999
Job Boards	4.604166666666667
Advertisement	7.941470588235294
Social	7.625
I Distintive HR	12.25
Premium Job Boa...	14.08
Happiest Group	6.4911111111111115
CareerSite	7.626974789915965
Sourced	4.154761904761905

7. Candidates with most Common first name

Query:

```
SELECT FirstName, COUNT(*) AS NameCount
FROM Candidates
GROUP BY FirstName
ORDER BY NameCount DESC
LIMIT 5;
```

Ans:

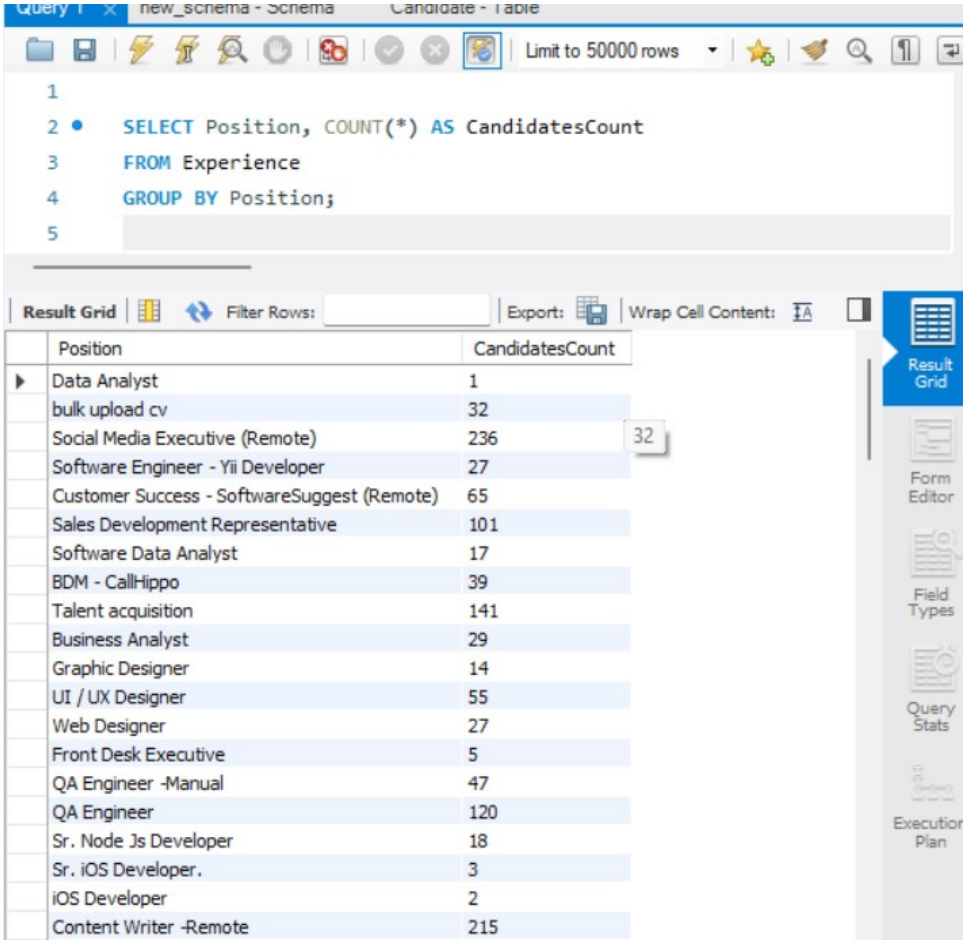


8. Get count of candidates for each job position

Query:

```
SELECT Position, COUNT(*) AS CandidatesCount
FROM Experience
GROUP BY Position;
```

Ans:



THERE ARE SEVERAL MORE QUESTIONS CAN BE ANSWERED USING THIS EFFICIENT DATABASE SYSTEM CREATED.

- **BENIFITS OF THIS NORMALIZED DATABASE:**

1. **FUTURE RECORD CAN BE EASILY ADDED INTO THIS WITHOUT HAVING INSERTION ANAMOLIES**
2. **DELETION OF A CERTAIN RECORD WON'T AFFECT ANY OTHER RECORDS**
3. **SCALABLE: IT CAN GROW EASILY AS THE COMPANY GROWS.**

4.2. LIST OF FEW QUESTIONS THAT CAN BE ANSWERED:

1. Retrieve all columns for all candidates in the dataset.

2. Determine the total number of candidates in the dataset.

3. Retrieve a list of distinct cities where candidates are located.

4. Count the number of candidates for each source of application.

5. Calculate the average experience (in years) for candidates in each city.

6. Calculate the total experience (sum of experience) for each job position.

7. If applicable, join the dataset with another table containing additional information (e.g., outcomes, feedback).

8. Identify candidates with experience higher than the average experience in their city.

9. Count the number of candidates for each combination of source and job position.

4.3. SUGGESTIONS:

1. Hiring Based on ranking:

- **Recruitment board should consider the rankings of the candidates as given in Question no.2 & Question no.4 above, to shortlist Candidates to cut off hiring expenses**

2. Optimizing Recruitment Sources:

- **Allocate more resources to the Source of Application: 'Free Job Boards', 'Distinctive HR' and 'Advertisements' as they have the highest qualified candidates.**

3. Geographical Focus:

- **Target Cities: Bangalore, Mumbai & New Delhi as they constitute the candidates with the highest average Experience.**

4. Name Analysis:

- **Amit is the most common name still these days**
- **Singh as a surname is most common**

5. Data Quality Improvement:

- **Add the Outcome data of hiring status as Hired or Rejected, as it limits the ability to perform certain analyses related to recruitment success, effectiveness of the hiring process, and factors influencing hiring decisions.**

4.3. LIMITATION OF DATASET:

1. Limited Variable:

The Dataset doesn't include additional details such as education, skills, specific project experience, or outcomes (e.g., whether a candidate was hired or not), which could provide a more comprehensive view.

2. Outcome Information (Hired or Not Hired):

The absence of information on whether candidates were hired or not limits the ability to analyze the effectiveness of the recruitment process or factors influencing hiring decisions.

3. Lack of Time component:

It makes the data challenging to analyze trends over time or changes in candidate behavior or characteristics.