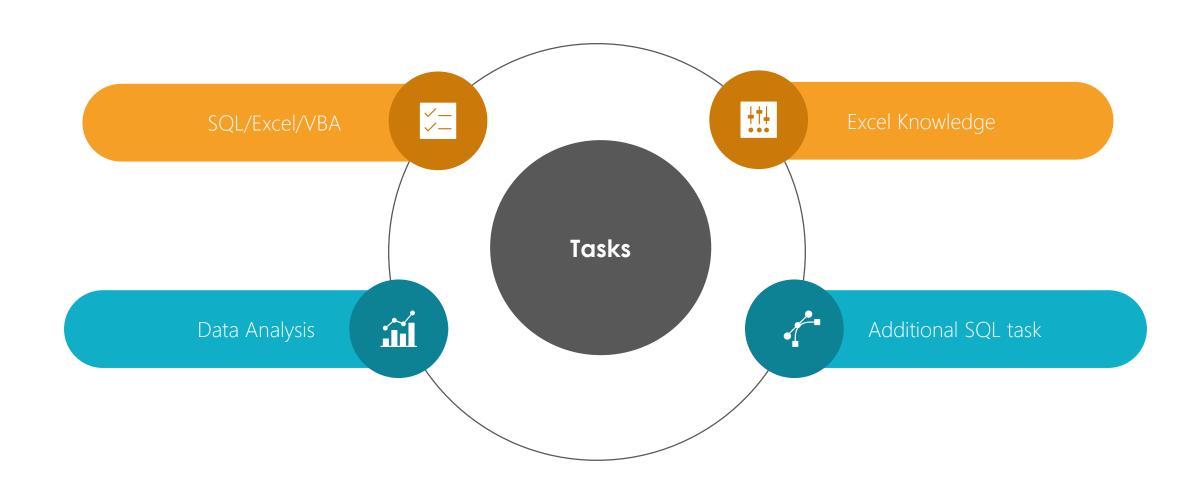


Business Data Analysis Presentation

Prakhar Srivastava

Tasks



Transactions Data Analysis

```
8345;Debit;4342;SE

3138;Debit;5852;SE

5233;Debit;3517;SE

8634;Debit;5325;DK

8076;Credit;4815;SE Text file

5860;Debit;623;SE

5143;Credit;5752;FI

7592;Debit;290;SE

553;Debit;2389;DK

5488;Credit;5655;SE
```

Txt file is imported into excel and delimiter is set to semicolon;

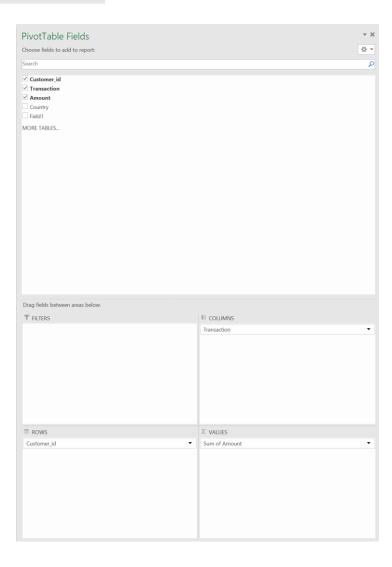
	Α	В	C	D
1	Customer	Transactio	Amount	Country
2	8345	Debit	4342	SE
3	3138	Debit	5852	SE
4	5233	Debit	3517	SE
5	8634	Debit	5325	DK
6	8076	Credit	4815	SE
7	5860	Debit	623	SE
8	5143	Credit	5752	FI
9	7592	Debit	290	SE
10	553	Debit	2389	DK
11	5488	Credit	5655	SE

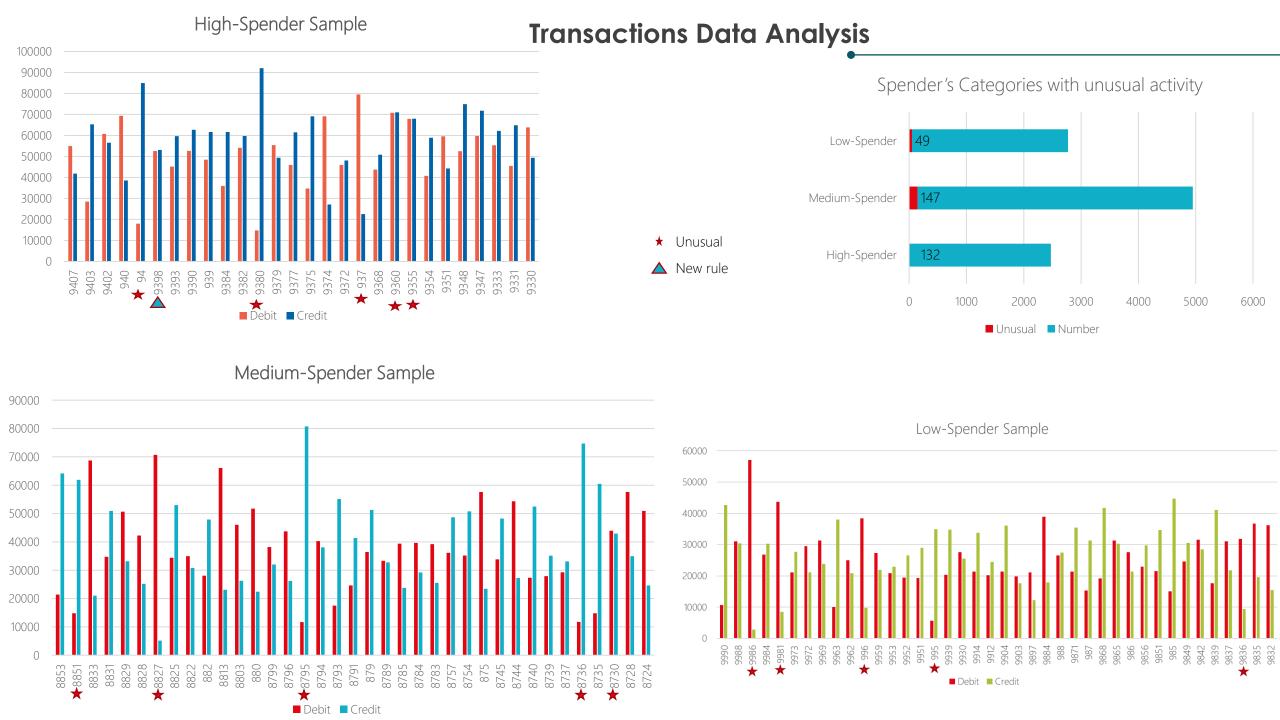
 $CustomerSegment \\ = IF(D5 > (AVERAGE(D\$5:D\$10004) + (0.2*AVERAGE(D\$5:D\$10004))), \\ "High-spender", IF(D5 < (AVERAGE(D\$5:D\$10004) - (0.2*AVERAGE(D\$5:D\$10004))), \\ "Low-spender", \\ "Medium-spender")) \\ = (0.2*AVERAGE(D\$5:D\$10004) + (0.2*AVERAGE(D\$5:D\$10004))), \\ "High-spender", \\ "IF(D5 < (AVERAGE(D\$5:D\$10004) - (0.2*AVERAGE(D\$5:D\$10004))), \\ "High-spender", \\ "$

4	Customer_id <mark>→</mark> 1 Debit		Credit	Grand Total	CustomerSegment *
5	0	34679	42490	77169	Medium-spender
6	1	43973	25248	69221	Medium-spender
7	10	59939	36989	96928	High-spender
8	100	27281	53875	81156	Medium-spender
9	1000	44432	21900	66332	Medium-spender
10	1001	68151	4499	72650	Medium-spender
11	1002	54295	48123	102418	High-spender
12	1003	36330	61566	97896	High-spender
13	1004	14901	39708	54609	Low-spender
14	1005	17314	35867	53181	Low-spender
15	1006	72468	32556	105024	High-spender

D_Q1	=QUARTILE.INC(O5:O2477,1)
D_Q3	=QUARTILE.INC(05:02477,3)
D_IQR	=R5-Q5
DebitActivity	=IF(ABS(O5-Q5)>S5, "Unusual", "")
Activity	=IF(AND(U5="Unusual",Y5="Unusual"),"Suspicious","")
	-II (AND(03- Oliusual ,13- Oliusual), Suspicious ,)

0 11 4 11 11	
CreditActivity	Activity
Jnusual	
Jnusual	
Jnusual	Suspicious
Jnusual	
Jnı	usual usual





My SQL Skills

Table: loans

loan_id	customer_id	loan_amount	loan_date
1	101	5000	2022-01-01
2	102	8000	2022-02-01
3	103	9000	2022-03-01
4	104	6000	2022-04-01
5	105	7000	2022-05-01
6	106	6000	2022-06-01
7	107	8000	2022-07-01
8	108	9000	2022-08-01
9	109	7000	2022-09-01
10	110	8000	2022-10-01

Table: customers

customer_id	customer_name	customer_address	customer_email	customer_phone
101	John Doe	123 Main St	john.doe@example.com	555-555-5555
102	Jane Doe	456 Elm St	jane.doe@example.com	555-555-5556
103	John Smith	789 Oak St	john.smith@example.com	555-555-5557
104	Jane Smith	246 Pine St	jane.smith@example.com	555-555-5558
105	Bob Johnson	135 Maple St	bob.johnson@example.com	555-555-5559
106	Alice Johnson	678 Cedar St	alice.johnson@example.com	555-555-5560
107	Charlie Brown	246 Birch St	charlie.brown@example.com	555-555-5561
108	Sarah Connor	135 Willow St	sarah.connor@example.com	555-555-5562
109	Mike Ross	678 Maple St	mike.ross@example.com	555-555-5563
110	Rachel Zane	246 Oak St	rachel.zane@example.com	555-555-5564

Total loan amount by customer:

```
1 SELECT
2   customers.customer_name,
3   SUM(loans.loan_amount) AS total_loan_amount
4 FROM
5   loans
6   INNER JOIN customers ON loans.customer_id = customers.customer_id
7 GROUP BY
8   customers.customer_name
9
```

customer_name	total_loan_amount
Alice Johnson	6000
Bob Johnson	7000
Charlie Brown	8000
Jane Doe	8000
Jane Smith	6000
John Doe	5000
John Smith	9000
Mike Ross	7000
Rachel Zane	8000
Sarah Connor	9000

<u>Total loan per year:</u>

```
1 SELECT
2 YEAR(loans.loan_date) AS loan_year,
3 SUM(loans.loan_amount) AS total_loan_amount
4 FROM
5 loans
6 GROUP BY
7 YEAR(loans.loan_date)
8
```

loan_y	ear .	total_loan_amount
2022		73000

Top 5 customers with the loan amount:

```
1 SELECT
2   customers.customer_name,
3   SUM(loans.loan_amount) AS total_loan_amount
4   FROM
5   loans
6   INNER JOIN customers ON loans.customer_id = customers.customer_id
7   GROUP BY
8   customers.customer_name
9   ORDER BY
10   total_loan_amount DESC
11 LIMIT 5
```

Number of loans per customer:

```
1 SELECT
2   customers.customer_name,
3   COUNT(loans.loan_id) AS loan_count
4 FROM
5   loans
6   INNER JOIN customers ON loans.customer_id = customers.customer_id
7 GROUP BY
8   customers.customer_name
9
```

Customers who have taken out loan in March:

```
1 SELECT customers.customer_name, customers.customer_email
2 FROM customers
3 INNER JOIN loans
4 ON customers.customer_id = loans.customer_id
5 WHERE MONTH(loans.loan_date) = 3;
```

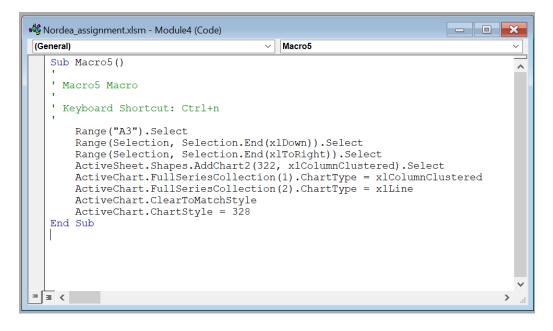
customer_name	total_loan_amount
John Smith	9000
Sarah Connor	9000
Rachel Zane	8000
Charlie Brown	8000
Jane Doe	8000

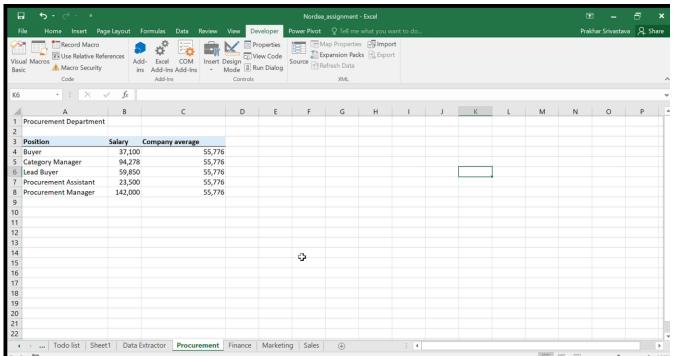
customer_name	loan_count
Alice Johnson	1
Bob Johnson	1
Chartie Brown	1
Jane Doe	1
Jane Smith	1
John Doe	1
John Smith	1
Mike Ross	1
Rachel Zane	1
Sarah Connor	1

customer_name	customer_email
John Smith	john.smith@example.com

Macros

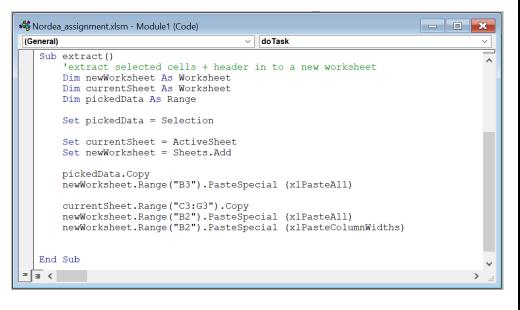
Automating repetative graph building task using Macros recording

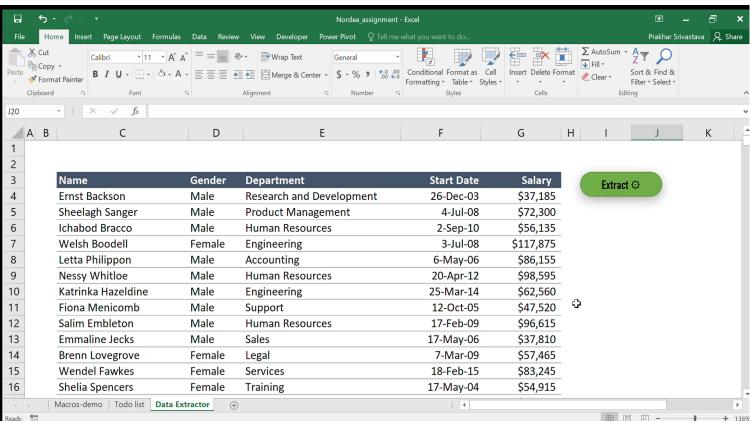




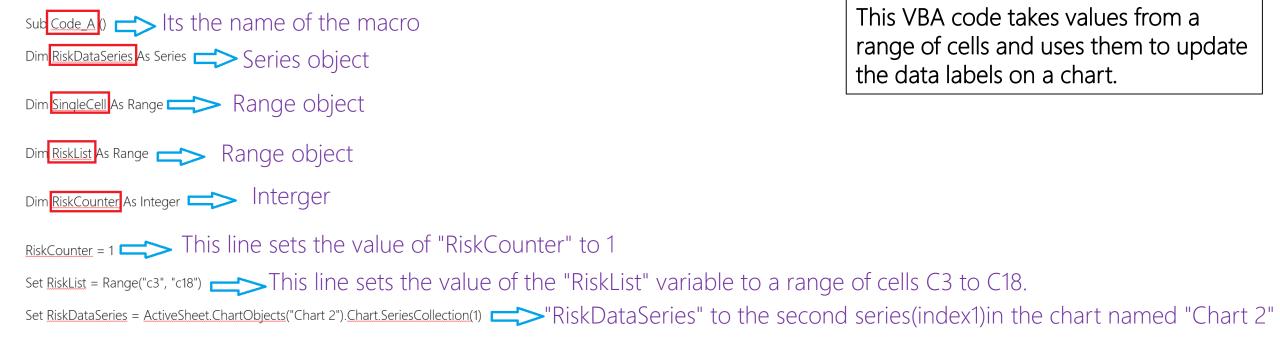
Excel/VBA

Extracting data and header into another sheet with a single click using VBA





VBA code explained



RiskDataSeries.Points(RiskCounter). DataLabel.Text = SingleCell.Value _____ updates the text of the data label to the value of the current cell in the loop

RiskCounter = RiskCounter + 1 increments the value

Next SingleCell end of the loop

RiskDataSeries HasDataLabels = True enables data labels to be displayed on the chart

For Each SingleCell In RiskList a loop that will iterate over each cell in the "RiskList" range

End Sub

VBA code explained

sub Code B() | Its the name of the macro

Set SheetA = Sheets("ProductA") Reference to the worksheet

Set <u>SheetSum</u> = Sheets("Summary") Reference to the worksheet

This VBA code copies rows of data from two different worksheets, "ProductA" and "ProductB", and pastes them into a third worksheet, "Summary"

This line sets the LR variable to the last non-empty row in column 2 (B) of the SheetA worksheet. It uses the Rows. Count property to LR = SheetA.Cells(Rows.Count, 2).End(xlUp).Rowget the total number of rows in the worksheet, and the xIUp constant to find the last non-empty cell in column 2 (B).

Set RNG = SheetA.Range("A4:A" & LR) reference the range of cells in column A of SheetA from row 4 to the last non-empty row

This line copies the entire row of data for each cell in the RNG range and pastes it into the next RNG.EntireRow.Copy SheetSum.Cells(Rows.Count, 1).End(xlUp)(2) empty row at the bottom of column 1 (A) in the SheetSum worksheet

Dim SheetB As Worksheet These two lines declare a new variable called SheetB of type Worksheet Set SheetB = Sheets("ProductB") and set it to reference the worksheet named "ProductB".

LR = SheetB.Cells(Rows.Count, 2).End(xlUp).Row These two lines set the LR variable to the last non-empty row in column 2 (B) of the SheetB worksheet, and set the RNG variable to reference the range of cells in column A of SheetB from row 4 to the last non-empty row. Set RNG = SheetB.Range("A4:A" & LR)

This line copies the entire row of data for each cell in the RNG range and RNG.EntireRow.Copy SheetSum.Cells(Rows.Count, 1).End(xlUp)(2) pastes it into the next empty row at the bottom of column 1 (A) in the SheetSum worksheet

End Sub

Additional SQL tasks _____

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	Niharika	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

worker table2

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	<u>Niharika</u>	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to fetch the names of workers who earn the highest salary

```
1 SELECT First_name, Last_name, Salary
```

2 FROM worker

3 WHERE Salary = (SELECT MAX(Salary) FROM worker)

First_name	Last_name	Salary
Amitabh	Singh	500000
Vivek	Bhati	500000

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	Niharika	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to fetch departments along with the total salaries paid for each of them.

1 SELECT Department, SUM(Salary) as Total_Salary

2 FROM worker

3 GROUP BY Department

Department	Total_Salary
Account	775000
Admin	670000
HR	400000

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	Niharika	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to print the name of employees having the highest salary in each department

```
SELECT Department, First_name, Last_name, Salary
FROM worker w1
WHERE Salary =
   (SELECT MAX(Salary) FROM worker w2 WHERE w1.Department = w2.Department)
GROUP BY Department, First_name, Last_name, Salary;
```

Department	First_name	Last_name	Salary
Account	Vivek	Bhati	500000
Admin	Amitabh	Singh	500000
HR	Vishal	Singhal	300000

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	Niharika	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to fetch the departments that have less than five people in it

```
1 SELECT Department
2 FROM
3  (SELECT Department, COUNT(*) as num_workers
4   FROM worker
5   GROUP BY Department) as subquery
6 WHERE num_workers < 5;</pre>
```

Department	
Account	
Admin	
HR	

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	<u>Niharika</u>	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to print details of the Workers whose FIRST_NAME ends with 'h' and contains six alphabets

```
1 SELECT *
2 FROM worker
3 WHERE LENGTH(First_name) = 6 AND Right(First_name, 1) = 'h';
4
```

Worker_id	First_name	Last_name	Salary	Joining_date	Department
7	Satish	Kumar	75000	2014-01-20T09:00:00Z	Account

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	Niharika	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query that fetches the unique values of DEPARTMENT from Worker table and prints its length

1 SELECT department, COUNT(*) as Number_of_Workers

2 FROM worker

3 GROUP BY department;

department	Number_of_Workers
Account	3
Admin	3
HR	2

WORKER_ID	FIRST_NAME	LAST_NAME	SALARY	JOINING_DATE	DEPARTMENT
1	Monika	Arora	100000	2014-02-20 09:00:00	HR
2	<u>Niharika</u>	Verma	80000	2014-06-11 09:00:00	Admin
3	Vishal	Singhal	300000	2014-02-20 09:00:00	HR
4	Amitabh	Singh	500000	2014-02-20 09:00:00	Admin
5	Vivek	Bhati	500000	2014-06-11 09:00:00	Admin
6	Vipul	Diwan	200000	2014-06-11 09:00:00	Account
7	Satish	Kumar	75000	2014-01-20 09:00:00	Account
8	Geetika	Chauhan	90000	2014-04-11 09:00:00	Admin

WORKER_ REF_ID	WORKER_TITLE	AFFECTED_FROM
1	Manager	2016-02-20 00:00:00
2	Executive	2016-06-11 00:00:00
8	Executive	2016-06-11 00:00:00
5	Manager	2016-06-11 00:00:00
4	Asst. Manager	2016-06-11 00:00:00
7	Executive	2016-06-11 00:00:00
6	Lead	2016-06-11 00:00:00
3	Lead	2016-06-11 00:00:00

Query to print details of the Workers who are also Managers

```
1 SELECT w.*
2 FROM worker w
3 JOIN table2 t2 ON w.Worker_id = t2.Worker_ref_id
4 WHERE t2.Worker_title = 'Manager';
```

Worker_id	First_name	Last_name	Salary	Joining_date	Department
1	Monika	Arora	100000	2014-02-20T09:00:00Z	HR
5	Vivek	Bhati	500000	2014-06-11T09:00:00Z	Account

Excel Knowledge



