

# GASKIN GREEN AUTOS

DATA ANALYSIS REPORT

GROUP A

SARA JAMIL

MANHA ABDELLAH

NEHA ALAM

M. HAMZA ALI

SHEHLA ASGHAR

## Capstone Project Report 2022

### Table of Contents

INTRODUCTION .....	2
BACKGROUND .....	2
OBJECTIVE .....	2
OUR PLAN.....	2
Data Set.....	3
Data Dictionary.....	3
Stakeholders.....	6
Our Business Problems .....	6
Entity Relationship Diagram (ERD).....	6
BUSINESS PROBLEM 1 .....	7
BUSINESS PROBLEM 2 .....	8
BUSINESS PROBLEM 3 .....	9
POSSIBLE COURSE OF ACTIONS .....	10
SQL QUERIES .....	11
Main Business Problems .....	11
Potential Business Problems .....	14

## INTRODUCTION

The report has been written to present the analysis performed on the dataset of Gaskin Green autos. The Gaskin Green autos is a company that believes in environment friendly technology and works to sell the same. A team of data analysts has been dedicated to gather, analyze and then provide solution to identified problem areas.

## BACKGROUND

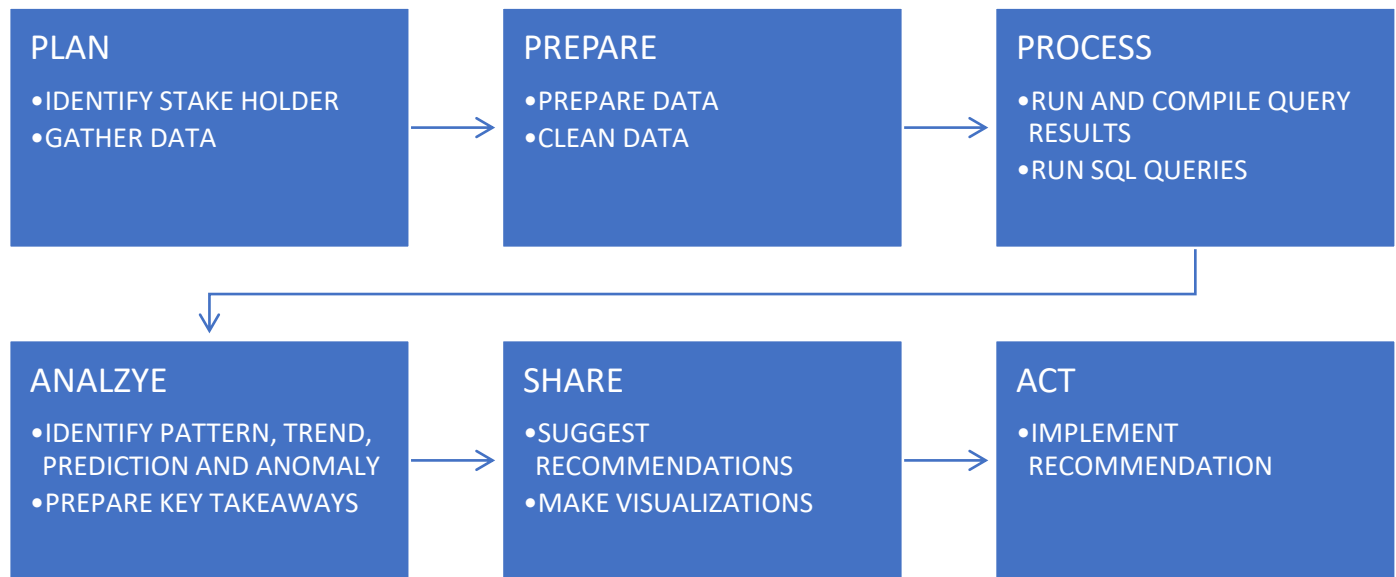
Gaskin Green Autos is a car manufacturing company that uses environmentally friendly technology. The company has hired us to gather, analyze data and provide actionable recommendations.

## OBJECTIVE

To explore and analyze the data set at hand to identify potential opportunities and improvement areas.

## OUR PLAN

The objective is to provide a substantial and accurate analysis of the data we have at our hands. The plan initiates from the planning stage where the data is gathered and the stakeholders are identified. The stakeholders are involved and are directly impacted by the decisions this analysis proposes to implement. We gathered the data from the relevant source and moved to the second stage of our plan which is to prepare the data. We cleaned the data, and removed any unwanted or empty fields. The data is now to be used in the third step of our which is to process by running the data through some SQL queries. The results of the SQL queries are then run through the analysis phase where we predict or identify some trends or patterns. Gaps or missing data help identify any anomaly or abnormal value as well. After analysis, we begin to put our focus on how to share the same data with the stakeholders. The data would be visualized and converted into charts and graphs easier to interpret. The key takeaways and recommendations are then shared with stakeholders to take action.



*Figure 1 Plan to perform data analysis*

#### Data Set

The following is the link to our data set. The details of the same can be seen in figure 3.

[Click here](#) to view Data set.

#### Data Dictionary

A data dictionary serves the purpose of identifying the fields in a data set and is set for all those using it accordingly. Setting a data dictionary standardizes the fields and their data types.

TABLE NAME	COLUMN	DATA TYPE	DESCRIPTION	EXAMPLE
	CREDITID	Integer	The ID of the credit type	1,2,3 till 7.
	CREDITDESCRIPTION	string	The description of type	Very poor, poor, satisfactory etc.
	MINFICO FICO credit scores are a method of quantifying and evaluating an individual's creditworthiness.	Float	Minimum range for a particular credit	300.00, 499.00 etc Max:

CREDITRATING	FICO scores are used in 90% of mortgage application decisions in the United States. <a href="https://www.investopedia.com/terms/f/ficoscore.asp">https://www.investopedia.com/terms/f/ficoscore.asp</a>		description.	700.00 which is very good
	MAXFICO	Float	Maximum range for a particular credit description	499.00, 699.00 etc 850.00 being highest which is very good
SALESPERSON	SALESID	Integer	The ID allotted to each customers	1,2,3,...129
	SFIRSTNAME	string	The first name of the sales person	Harry
	SLASTNAME	String	The last name of the sales person	Potter
	SHIREDATE	Date/time	THE date on which the sales person was hired	2/13/1996
	SALARY	Currency	The salary they have	\$140000
	ENCID	Integer	The ID of the encounter	1,2,3
	SALESID	Integer	The ID of the salesperson	1,2,33...

ENCOUNTER			who made the encounter	
	CUSTOMERID	Integer	The customer ID for whom the encounter was made	1,2,3..
	ENCDATE	Date/time	The date on which the encounter was made	01/07/2015
	PURCHASE	Boolean	The status of encounter purchase	True, False
CUSTOMER	CUSTOMERID	Integer	The ID of the customer	1,2,3
	CFIRSTNAME	String	The first name of customer	Harry
	CLASTNAME	String	The last name of customer	Potter
	CPHONE	Integer	The contact number of customer	(860)45485461
	ANNUALINCOME	Currency	The annual income for the customer	\$15458744
	CREDITID	String	The type of credit the customer	Good, poor, satisfactory

			has	
--	--	--	-----	--

Table 1 Data dictionary

## Stakeholders

A stakeholder is any individual who might be interested in the analysis we have performed on our data set. We have been thorough in our analysis and identification of the stakeholders of our data analysis and look forward to working with them on this project. Figure 2 provides a summary of who our identified stakeholders are with their relevancy to the project.

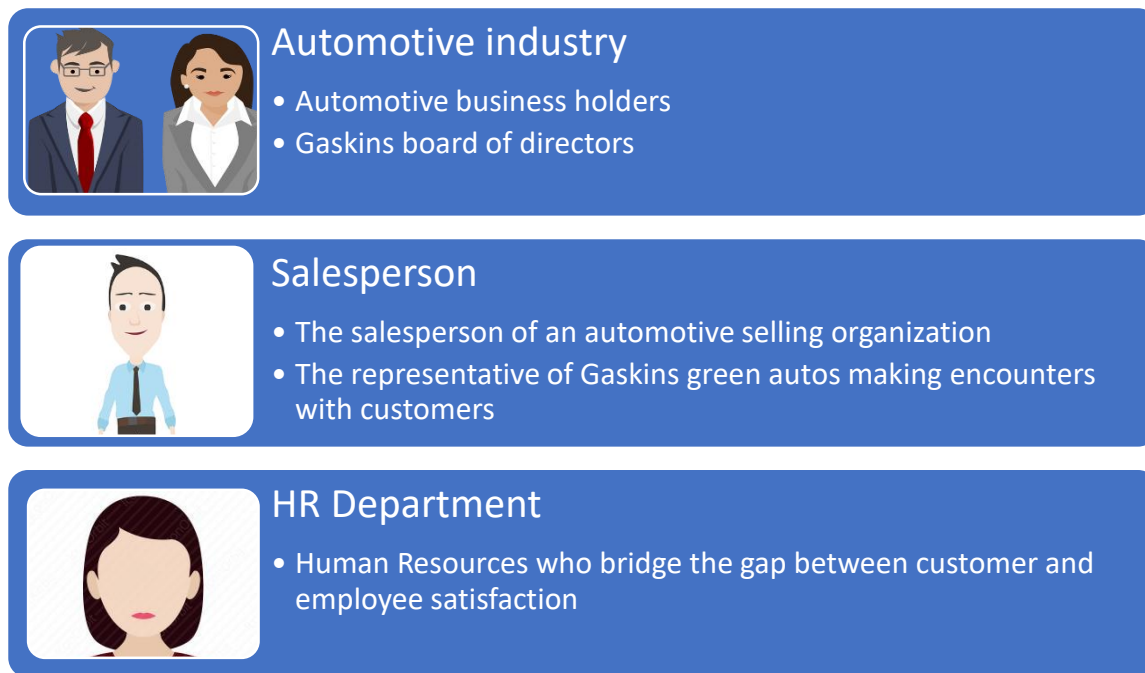


Figure 2 Stakeholders

## Our Business Problems

1. To discover if the annual salaries of customers lead to a better credit ID.
2. To discover any connection between experienced salesperson and their number of successful encounters
3. To find a pattern for the number of successful encounters
4. To discover if there is any correlation between credit rating and number of encounter
5. To identify the loyal customers for Gaskins.

Note: The analysis is our problems is limited to the month of July.

## Entity Relationship Diagram (ERD)

ERD is an entity relationship diagram that provides an insight to the relationships between the fields of our data set.

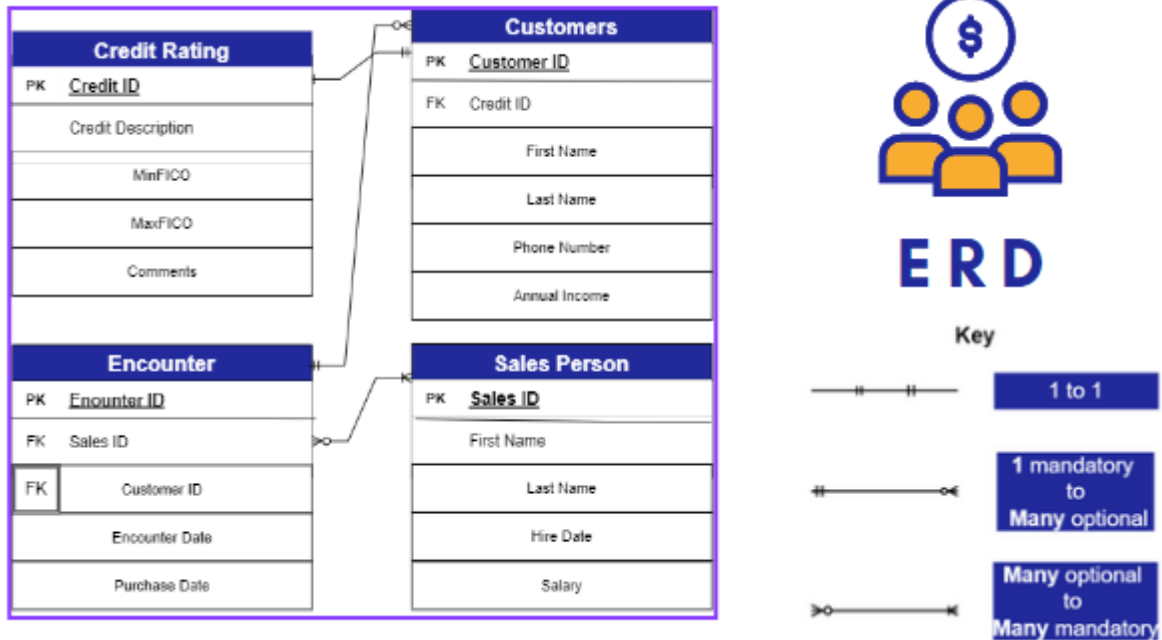
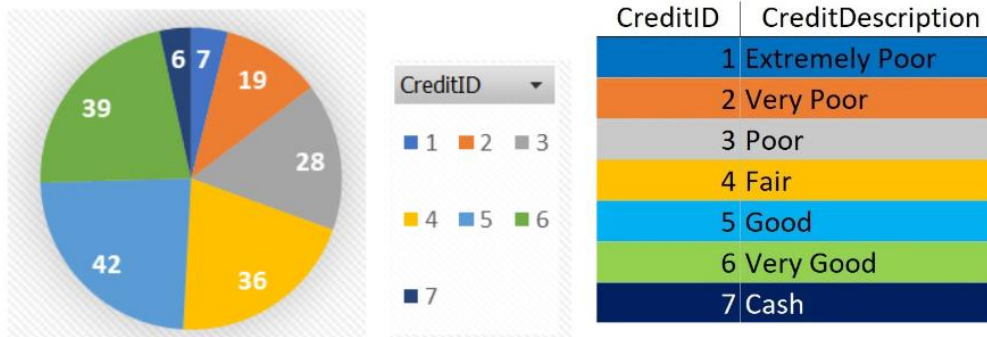


Figure 3 Entity Relationship Diagram

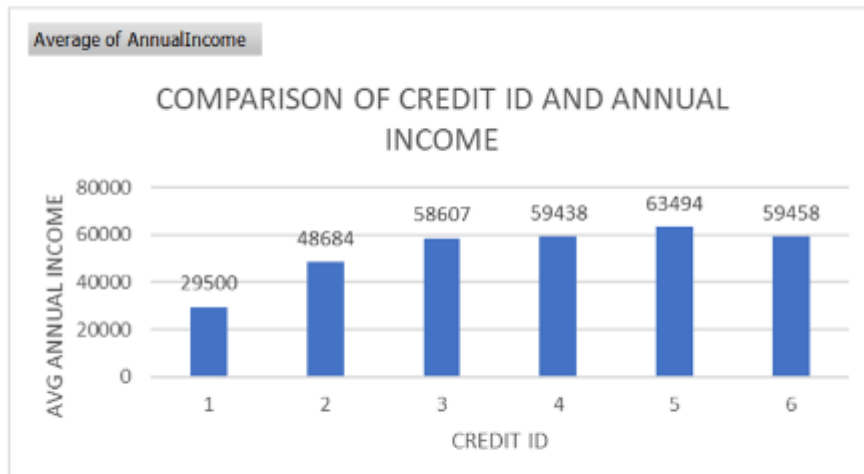
## BUSINESS PROBLEM 1

### Number of Customers and Credit ID





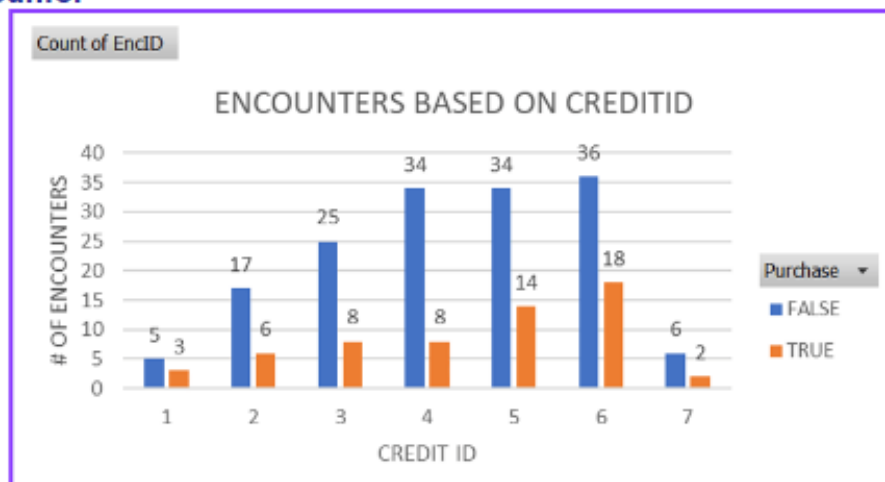
To discover if the annual salaries of customers lead to a better credit ID.



The graph elaborates the relation between credit ID and the corresponding average annual income of customers.

## BUSINESS PROBLEM 2

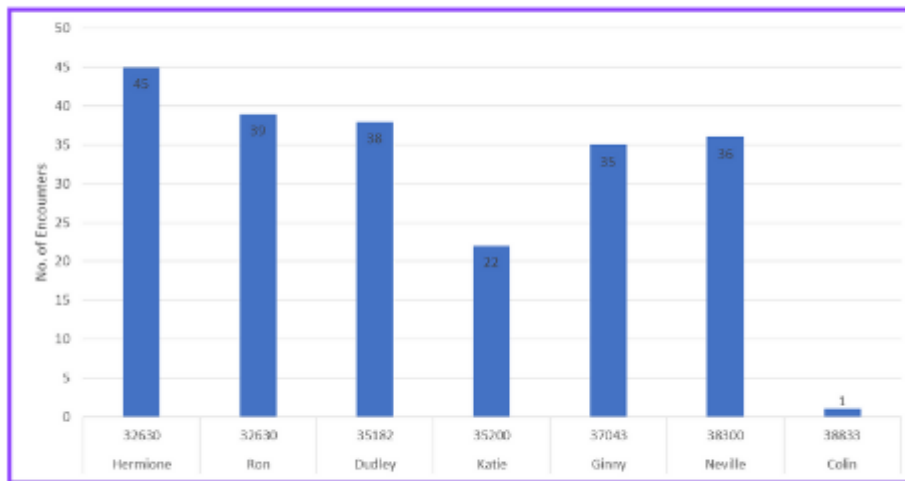
To discover if there is any correlation between credit rating and number of encounter



Customers with higher credit ID have higher purchases

### BUSINESS PROBLEM 3

To discover any connection between experienced salesperson and their number of successful encounters sales.



Experienced salesperson have higher number of encounters.

### BUSINESS PROBLEM 4



The graph elaborates that number of purchases are not affected by how old the employ is.

### BUSINESS PROBLEM 5

## Weekly report with respect to encounters

### NUMBER OF ENCOUNTER/WEEK



The graph explains a pattern of number of encounters increasing as we progress towards the end of the month.

### OUR TAKEAWAYS

Looking at the analyzed data from the eyes of a data analyst we suggest the following;

#### Salesperson experience

- Experience sales persons have greater true encounters. Keeping them engaged through bonuses and incentives that will retain them longer.

#### Each customer is important

- Customer loyalty is crucial to treat every visitor as a potential customer to reap good results.

#### Weekly report

- As the weeks pass each month, the average performance of all employees improves which lead an increase in number of encounters.

### POSSIBLE COURSE OF ACTIONS

Introduce customer loyalty programs. Hold regular training sessions to bridge the gap between veterans and new salespeople. Bonus incentives based on a true number of encounters. Free consultation session annual incommers at the beginning of the credit Follow-ups with old customers to get feedback. Loyalty with customers of higher credit the ID will result into frequent successful encounter.

## SQL QUERIES

### Main Business Problems

#### 1. To discover if the annual salaries of customers lead to a better credit ID.

##### Query:

```
SELECT creditid, round(avg(annualincome))  
FROM customer  
GROUP BY creditid  
order by creditid
```

##### Result:

Result Grid			Filter Rows:	Export:	Wrap Cell Content:
	creditid	round(avg(annualincome))			
▶	1	29500			
	2	48684			
	3	58607			
	4	59438			
	5	63494			
	6	59458			
	7	NULL			

#### 2. To discover if there is any correlation between credit rating and number of encounters

##### Query:

```
SELECT CUSTOMER.creditid, count(purchase) as Total_Number_of_Encounters,  
COUNT(CASE WHEN PURCHASE ='TRUE' THEN 1 END) AS  
Number_of_TRUE_PURCHASE, COUNT(CASE WHEN purchase='FALSE' THEN 1 END)  
AS Number_of_FALSE_PURCHASE  
FROM customer  
JOIN encounter  
ON CUSTOMER.customerid = encounter.customerid  
GROUP BY creditid  
order by creditid
```

##### Result:

Result Grid    Filter Rows: <input type="text"/>   Export:    Wrap Cell Content:				
	creditid	Total_Number_of_Encounters	Number_of_TRUE_PURCHASE	Number_of_FALSE_PURCHASE
▶	1	8	3	5
	2	23	6	17
	3	33	8	25
	4	42	8	34
	5	48	14	34
	6	54	18	36
	7	8	2	6

**3. To discover any connection between experienced salesperson and their number of successful encounters sales.**

**Query:**

```
select sfirstname as salesperson , salesperson.salesid, shiredate, count(encounter.purchase) AS
Number_of_Encounters
FROM salesperson
JOIN encounter
On salesperson.salesid = encounter.salesid
group by sfirstname
order by shiredate
```

**Result:**

Result Grid    Filter Rows: <input type="text"/>   Export:    Wrap Cell Content:				
	salesperson	salesid	shiredate	Number_of_Encounters
▶	Ron	3	1989-05-02	39
	Hermione	2	1989-05-02	45
	Dudley	4	1996-04-27	38
	Katie	5	1996-05-15	22
	Ginny	6	2001-06-01	35
	Neville	7	2004-11-09	36
	Colin	8	2006-04-26	1

**4. Weekly report with respect to encounters.**

**Query:**

```
Select count(*) AS positiveSaleinWeek1 from encounter WHERE encounter.purchase = 'TRUE'
AND encounter.encdate BETWEEN '2015-07-01' AND '2015-07-07'; Select count(*) AS
positiveSaleinWeek2 from encounter WHERE encounter.purchase = 'TRUE'
```

AND encounter.encdate BETWEEN '2015-07-08' AND '2015-07-14'; Select count(\*) AS positiveSaleinWeek3 from encounter WHERE encounter.purchase = 'TRUE' AND encounter.encdate BETWEEN '2015-07-15' AND '2015-07-22'; Select count(\*) AS positiveSaleinWeek4 from encounter WHERE encounter.purchase = 'TRUE' AND encounter.encdate BETWEEN '2015-07-23' AND '2015-07-31';

### Result:

Result Grid		Filter Rows
	positiveSaleinWeek1	
▶	10	

Result Grid		Filter Rows:
	positiveSaleinWeek2	
▶	14	

Result Grid		Filter Rows:
	positiveSaleinWeek3	
▶	16	

Result Grid		Filter Rows:
	positiveSaleinWeek4	
▶	19	

### 5. To identify the loyal customers for Gaskins

#### Query:

```
SELECT cfirstname, count(purchase = 'TRUE')
FROM customer
JOIN encounter
ON customer.CustomerID = encounter.CustomerID
```

```
where purchase = 'True'  
GROUP BY CFirstName  
order by count(purchase = 'TRUE') DESC  
LIMIT 10
```

**Result:**

Result Grid			Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
	cfirstname	count(purchase = 'TRUE')				
▶	Katrina	2				
	Owen	1				
	Danny	1				
	George	1				
	Stephen	1				
	Mitchell	1				
	Dan	1				
	Bob	1				
	Vicki	1				
	Paul	1				

Potential Business Problems

**1. How many customers paid cash without looking into financing options?**

**Query:**

```
SELECT count(customer.customerid)  
Customers_paid_cash_without_looking_into_financing_options  
FROM customer  
INNER JOIN creditrating  
ON customer.creditid = creditrating.creditid  
WHERE customer.creditid = '7'
```

AS

**Result:**

```

1 • SELECT count(customer.customerid) AS Customers_paid_cash_without_looking_into_financing_options
2 FROM customer
3 INNER JOIN creditrating
4 ON customer.creditid = creditrating.creditid
5 WHERE customer.creditid = '7'
6

```

Customers_paid_cash_without_looking_into_financing_options
6

## 2. Get the total number of customers on each date that purchased.

### Query:

```

SELECT encdate AS Date, count(customerid) AS Total_number_of_Customers
FROM encounter
WHERE purchase = 'TRUE'
GROUP BY (encdate)

```

### Result:

```

1 • SELECT encdate AS Date, count(customerid) AS Total_number_of_Customers FROM encounter
2 WHERE purchase = 'TRUE' GROUP BY (encdate)

```

Date	Total_number_of_Customers
2015-07-02	2
2015-07-03	1
2015-07-04	2
2015-07-05	1
2015-07-06	1
2015-07-07	1
2015-07-08	1
2015-07-09	1
2015-07-10	1
2015-07-11	4
2015-07-12	2
2015-07-13	2
2015-07-14	3
2015-07-15	1
2015-07-17	4
2015-07-18	2
2015-07-19	2
2015-07-20	2
2015-07-21	4
2015-07-22	1
2015-07-23	3
2015-07-24	2
2015-07-26	1
2015-07-27	3
2015-07-28	2
2015-07-29	3

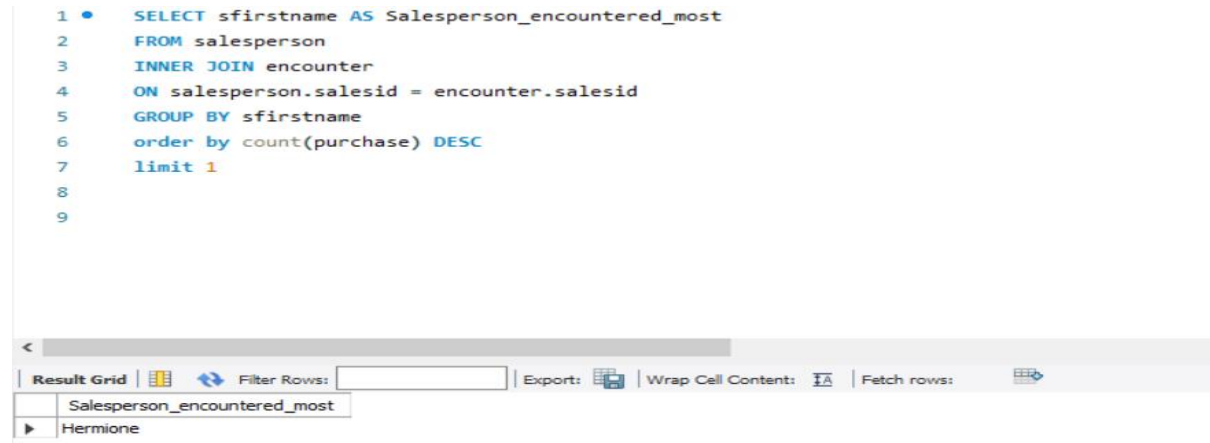
## 3. Which Salesperson have encountered most?

### Query:



```
SELECT sfirstname AS Salesperson_encountered_most
FROM salesperson
INNER JOIN encounter
ON salesperson.salesid = encounter.salesid
GROUP BY sfirstname
order by count(purchase) DESC
limit 1
```

**Result:**



The screenshot shows a SQL query editor with a query window and a result grid. The query is as follows:

```
1 • SELECT sfirstname AS Salesperson_encountered_most
2 FROM salesperson
3 INNER JOIN encounter
4 ON salesperson.salesid = encounter.salesid
5 GROUP BY sfirstname
6 order by count(purchase) DESC
7 limit 1
8
9
```

The result grid below the query shows the following data:

Salesperson_encountered_most
Hermione

**4. Identify the salesperson who sold the most.**

**Query:**

```
SELECT sfirstname AS Salesperson_sold_most
FROM salesperson
INNER JOIN encounter
ON salesperson.salesid = encounter.salesid
where purchase = 'True'
GROUP BY sfirstname
order by count(purchase = 'TRUE') DESC
LIMIT 1
```

**Result:**

```

1 • SELECT sfirstname AS Salesperson_sold_most
2 FROM salesperson
3 INNER JOIN encounter
4 ON salesperson.salesid = encounter.salesid
5 where purchase = 'True'
6 GROUP BY sfirstname
7 order by count(purchase = 'TRUE') DESC
8 LIMIT 1

```

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

Salesperson_solda_most
Hermione

## 5. Get details of customers that purchased most.

### Query:

```

SELECT *
FROM customer
INNER JOIN encounter
ON customer.CustomerID = encounter.CustomerID
where purchase = 'True'
GROUP BY CFirstName
order by count(purchase = 'TRUE') DESC
LIMIT 1

```

### Result:

```

1 • SELECT *
2 FROM customer
3 INNER JOIN encounter
4 ON customer.CustomerID = encounter.CustomerID
5 where purchase = 'True'
6 GROUP BY CFirstName
7 order by count(purchase = 'TRUE') DESC
8 LIMIT 1

```

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

customerid	cfirstname	clastname	cphone	annualincome	credited	endid	salesid	customerid	enddate	purchase
65	Katrina	Newsom	(801) 763-7396	54500	5	91	6	65	2015-07-14	true

## 6. What is the percentage of poor and good credit ratings in customers?

### Query:

```

SELECT COUNT(customerid) FROM project.'customer (1)';
SELECT COUNT(customerid) FROM project.'customer (1)'
WHERE credited>=4

```

WHERE credited<4

**Result:**

```
1 • SELECT COUNT(customerid) FROM project.`customer (1)`  
2 ;
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content
	COUNT(customerid)			
▶	177			

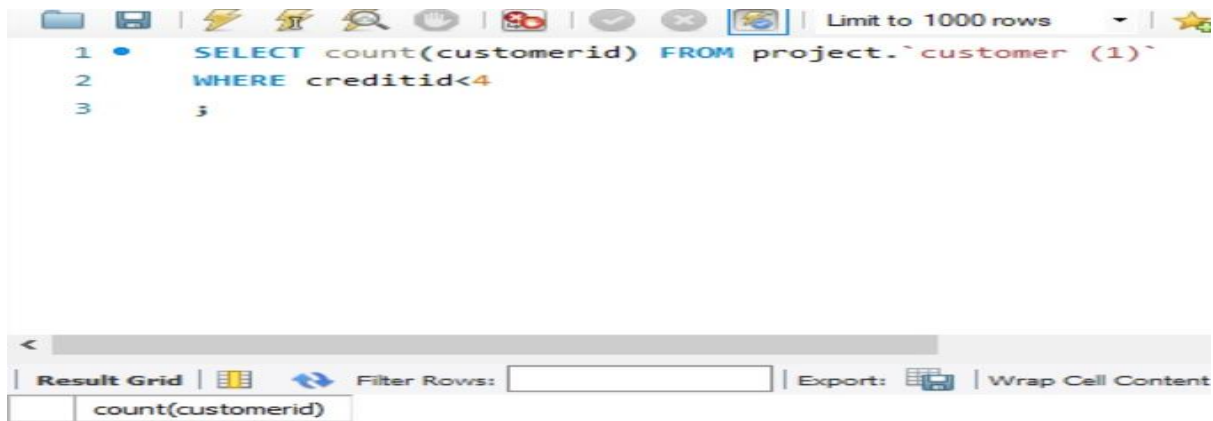
**GOOD:**

```
1 • SELECT count(customerid) FROM project.`customer (1)`  
2 WHERE creditid>=4  
3 ;
```

Result Grid		Filter Rows:	Export:	Wrap Cell Content:
	count(customerid)			
▶	123			

$(123/177)*100=69.5\%$

**BAD:**

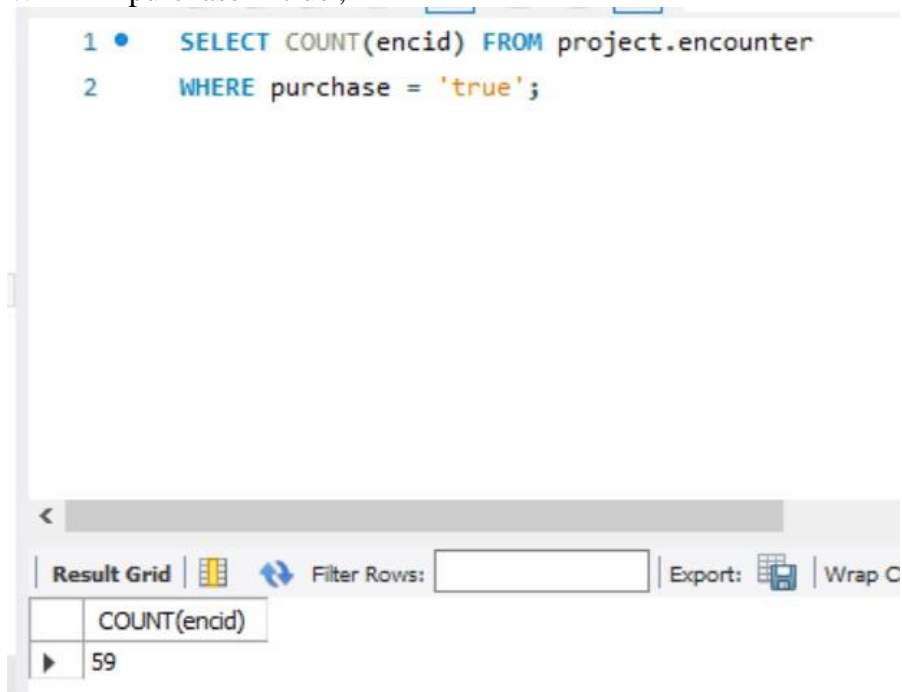


$(54/177) * 100 = 30.5\%$

## 7. How many sales were there?

Query:

SELECT COUNT(encid) FROM project.encounter  
 WHERE purchase = 'true';



## 8. Average Credit of Customers

Query:

SELECT avg(creditrating.creditid) FROM creditrating;

### Result:

14 • `SELECT avg(creditrating.creditid) FROM creditrating;`

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
avg(creditrating.creditid)			
4.0000			

### 9. What number of customers paid with cash

#### Query:

Highest Annual Income of customers

Select MAX(annualincome) as HighestIncome from customer;

#### Result:

16 • `Select MAX(annualincome) as HighestIncome from customer;`

Result Grid	Filter Rows:	Export:	Wrap Cell Content:
HighestIncome			
124375			

### 10. Lowest Annual Income of customers

#### Query:

Select Min(annualincome) as LowestIncome from customer;

#### Result:

16 • `Select Min(annualincome) as LowestIncome from customer;`

LowestIncome
20250

## 11. Employee having highest salary and their sales

### Query:

```
SELECT salesperson.sfirstname, encounter.purchase, salesperson.slastname, salesperson.salary,
salesperson.salesid from salesperson
inner join project.encounter ON salesperson.salesid = encounter.salesid WHERE
encounter.purchase = 'true'
group by salesid;
```

### Result:

Query 1

```
1 • use project;
2 • SELECT salesperson.sfirstname, encounter.purchase, salesperson.slastname, salesperson.salary,
3 salesperson.salesid from salesperson
4 inner join project.encounter ON salesperson.salesid = encounter.salesid WHERE encounter.purchase = 'true'
5 group by salesid;
```

sfirstname	purchase	slastname	salary	salesid
Ron	true	Weasley	93000	3
Hermione	true	Granger	82000	2
Dudley	true	Dursley	56000	4
Ginny	true	Weasley	75000	6
Neville	true	Longbottom	71000	7
Katie	true	Bell	67000	5

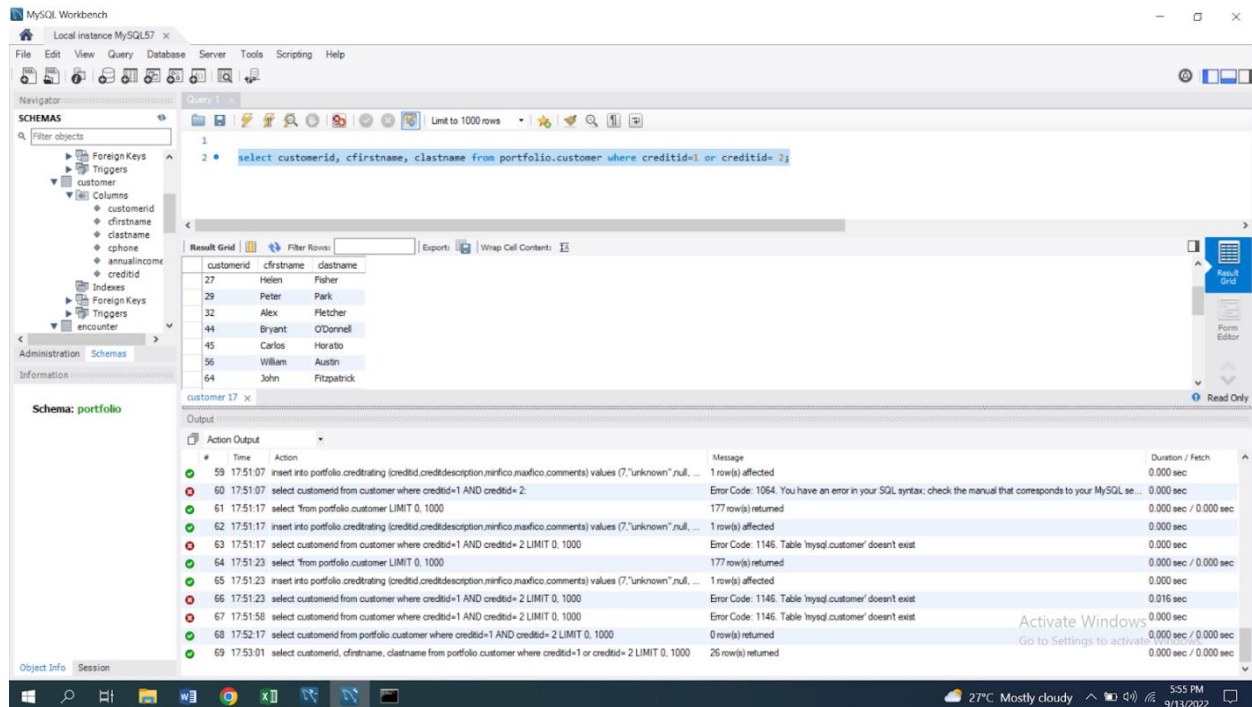
## 12. IDENTIFY THE RECORDS WHERE THE CUSTOMERS HAVE A BAD CREDIT (id 1,2)

### Query:

```
select customerid, cfirstname, clastname from portfolio.customer where creditid=1 or creditid=2;
```

### Result:

26 ROWS FOUND

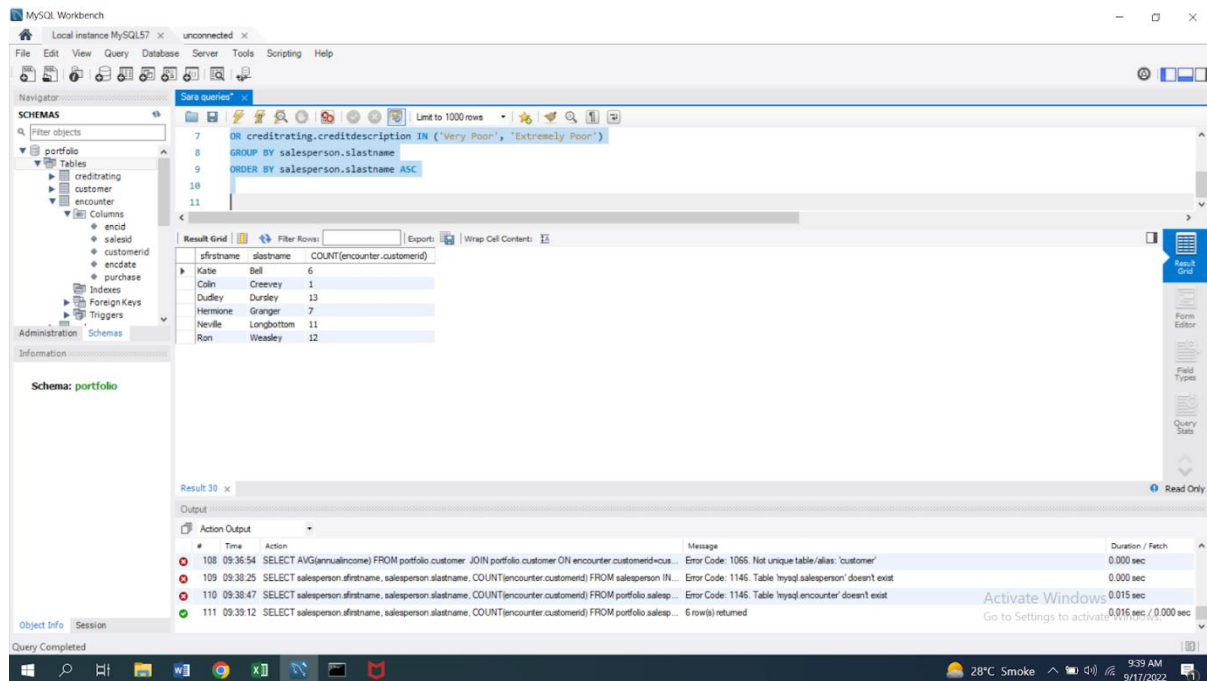


**13. Create a list that shows each salesperson (first and last name) and the number of encounters each has had with customers that have an annual income of \$25,000 or less or have a credit description that is “Very Poor” or “Extremely Poor.” Order the results alphabetically by the salesperson’s last name**

**Query:**

```
SELECT salesperson.sfirstname, salesperson.slastname, COUNT(encounter.customerid)
FROM portfolio.salesperson
INNER JOIN portfolio.encounter ON salesperson.salesid=encounter.salesid
INNER JOIN portfolio.customer ON encounter.customerid=customer.customerid
INNER JOIN portfolio.creditrating ON customer.creditid=creditrating.creditid
WHERE customer.annualincome <= 25000
OR creditrating.creditdescription IN ('Very Poor', 'Extremely Poor')
GROUP BY salesperson.slastname
ORDER BY salesperson.slastname ASC
```

**Result:**

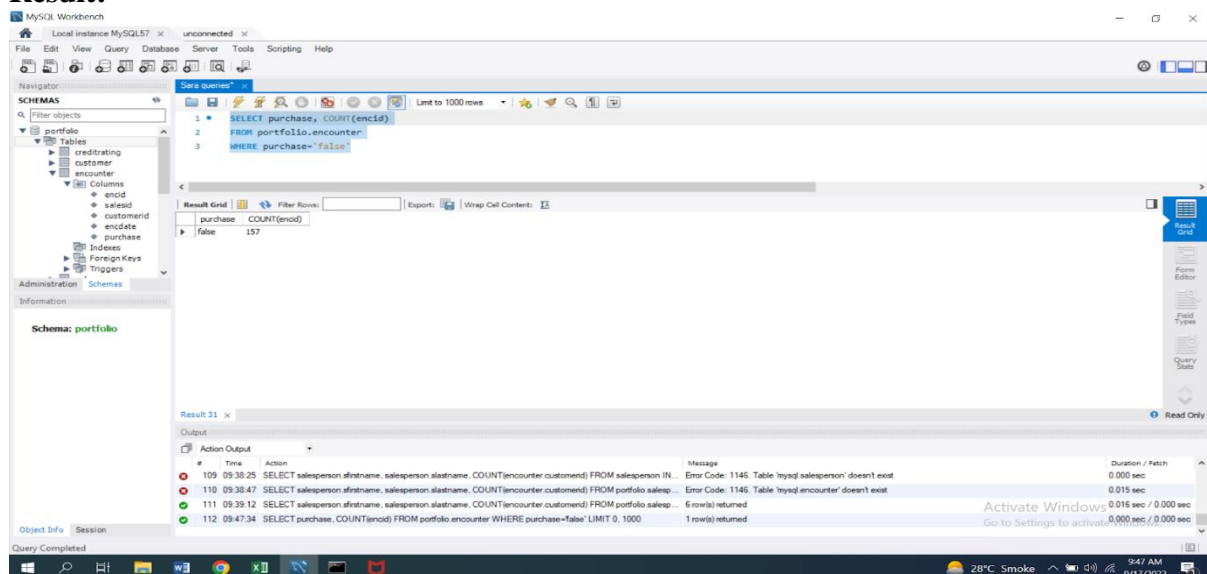


## 14. Count no. of false purchase

### Query:

SELECT purchase, COUNT(encid)  
FROM portfolio.encounter  
WHERE purchase='false'

### Result:



- In which day do they get the maximum order quantity?



## Query:

```
SELECT encdate AS Date, count(customerid) AS Total_number_of_Customers FROM  
encounter WHERE purchase = 'TRUE' GROUP BY (encdate) order by count(purchase =  
'TRUE') DESC
```

## Result:

```
1 • SELECT encdate AS Date, count(customerid) AS Total_number_of_Customers FROM encounter WHERE purchase = 'TRUE'  
2 GROUP BY (encdate) order by count(purchase = 'TRUE') DESC  
3  
4
```

Date	Total_number_of_Customers
2015-07-11	4
2015-07-17	4
2015-07-21	4
2015-07-14	3
2015-07-23	3
2015-07-27	3
2015-07-29	3
2015-07-30	3
2015-07-01	2
2015-07-02	2
2015-07-04	2
2015-07-12	2
2015-07-13	2
2015-07-18	2
2015-07-19	2
2015-07-20	2
2015-07-24	2
2015-07-28	2
2015-07-31	2
2015-07-03	1
2015-07-05	1
2015-07-06	1
2015-07-07	1
2015-07-08	1
2015-07-09	1

## 15. Monthly/yearly salary of an employ

### Query:

```
SELECT sfirstname, salary FROM salesperson  
monthly/yearly salary of an employ SELECT sfirstname, salary FROM salesperson
```

### Result:

```
1 • SELECT sfirstname, salary  
2 FROM salesperson  
3
```

sfirstname	salary
Harry	140000
Hermione	82000
Ron	93000
Dudley	56000
Katie	67000
Ginny	75000
Neville	71000
Colin	62000

## 16. Which customer have more income?

**Query:**

```
SELECT annualincome, cfirstname FROM customer GROUP BY cfirstname order by  
max(annualincome) DESC LIMIT 1
```

**Result:**

```
1 • SELECT annualincome, cfirstname
2 FROM customer
3 GROUP BY cfirstname
4 order by max(annualincome) DESC
5 LIMIT 1
```

Result Grid			Filter Rows:	Export:	Wi
	annualincome	cfirstname			
▶	124375	Sue			

## Conclusion

To conclude this report, it can be seen that Gaskins has problem areas that have been identified and can be mitigated if proper measures are taken on time. The analysis has been shared with the stakeholders to take relevant actions.