

[Open in app](#) ↗[Sign up](#)[Sign in](#)**Medium** Search**The Startup** · [Follow publication](#)

# Data Analysis Project for Retail: Sales Performance Report Using SQL

Analysis of DQLab Store sales by using MySQL

6 min read · Feb 6, 2021

**Adi Sugiarto**[Follow](#)[Listen](#)[Share](#)

Source : [freepik](#)

## About Dataset

This project provided by [DQLab](#), here was given the dataset contains the transactions report from 2009 until 2012 consist 5500 rows, which is the order status field has the value 'Order Finished', 'Order Returned' and 'Order Cancelled'. All data for this project are saved in the table called **dqlab\_store\_sales**. You can see the appearance of the dataset below :

```
#table information
DESC dqlab_sales_store;

#first 5 records
SELECT *
FROM dqlab_sales_store
LIMIT 5;
```

Field	Type	Null	Key	Default	Extra
order_id	int(11)	YES		NULL	
order_status	varchar(100)	YES		NULL	
customer	varchar(100)	YES		NULL	
order_date	date	YES		NULL	
order_quantity	double	YES		NULL	
sales	double	YES		NULL	
discount	double	YES		NULL	
discount_value	double	YES		NULL	
product_category	varchar(100)	YES		NULL	
product_sub_category	varchar(100)	YES		NULL	

fig.1 dqlab\_sales\_store table information

discount	discount_value	product_category	product_sub_category
0.04	20923	Office Supplies	Storage & Organization
0.07	1417223	Office Supplies	Appliances
0.08	794522	Technology	Telephones and Communication
0.08	63083	Office Supplies	Appliances
0.03	5612	Office Supplies	Binders and Binder Accessories

order_id	order_status	customer	order_date	order_quantity	sales
3	Order Finished	Muhammed MacIntyre	2010-10-13	6	523080
293	Order Finished	Barry French	2012-10-01	49	20246040
483	Order Finished	Clay Rozendal	2011-07-10	30	9931519
515	Order Finished	Carlos Soltero	2010-08-28	19	788540
613	Order Finished	Carl Jackson	2011-06-17	12	187080

fig.2 Firts 5 rows appearance of dqlab\_sales\_store table

## What the task given?

Through the data has given, the manager of DQLab Store wants to know :

1. Order numbers and total sales from 2009 until 2012 which order status is finished
2. Total sales for each sub-category of product on 2011 and 2012
3. The effectiveness and efficiency of promotions carried out so far, by calculating the burn rate of the overall promotions by year
4. The effectiveness and efficiency of promotions carried out so far, by calculating the burn rate of the overall promotions by sub-category of product on 2012
5. The number of customers transactions for each year
6. The number of new customers for each year

## Let's solving this project !

### 1. Order Numbers and Total Sales by Years

```
SELECT YEAR(order_date) years,
       SUM(sales) sales,
       COUNT(order_status) 'number of order'
FROM dqlab_sales_store
WHERE order_status = 'Order Finished'
GROUP BY 1;
```

*order\_date* field has a *date* format, we can get the year part of *date* format by using YEAR() function. Next, we use the SUM() function to get total sales, then COUNT() function to get the number of order. Don't forget to add WHERE clause to filter *order\_status* just for finished order. Finally, to get the value by years, we should put GROUP BY 1 on the query since year field is located in the first column.

output :

years	sales	number of order
2009	4613872681	1244
2010	4059100607	1248
2011	4112036186	1178
2012	4482983158	1254

fig.3 Number of sales and order

We can see, total sales of DQLab store are changed over the year. The highest total sales were in 2009 and it doesn't get higher after that. But different from the number of order, it goes ride except in 2011. Although the change isn't too significant over the years.

## 2. Total Sales by Sub-Category on 2011 and 2012

```
SELECT *,
       ROUND((sales2012-sales2011)*100/sales2012, 1) 'growth sales
(%)'
FROM(
  SELECT product_sub_category,
         SUM( IF( YEAR(order_date) = 2011, sales, 0)) sales2011,
         SUM( IF( YEAR(order_date) = 2012, sales, 0)) sales2012
  FROM dqlab_sales_store
  WHERE order_status = 'Order Finished'
  GROUP BY product_sub_category
) sub_category
ORDER BY 4 DESC;
```

Here the pivot table was used to compare the total sales in 2011 with 2012. We can use SUM() function followed by IF() function to do it. SUM() is used to get the total sales and IF() used to filter by year that we want to specify.

Get Adi Sugiarto's stories in your inbox

Join Medium for free to get updates from this writer.

Enter your email

Subscribe

output :

product_sub_category	sales2011	sales2012	growth sales (%)
Scissors, Rulers and Trimmers	12638340	36776400	65.6
Envelopes	36463900	58629280	37.8
Office Machines	545856280	811427140	32.7
Computer Peripherals	232677960	308014340	24.5
Storage & Organization	285991820	356714140	19.8
Rubber Bands	3090120	3837880	19.5
Binders and Binder Accessories	298023200	363879200	18.1
Paper	111080380	126896160	12.5
Office Furnishings	160471500	178927480	10.3
Telephones and Communication	392194658	422287514	7.1
Chairs & Chairmats	622962720	654168740	4.8
Pens & Art Supplies	43093800	43818480	1.7
Appliances	272630020	266131100	-2.4
Bookcases	169304620	159984680	-5.8
Tables	505875008	388993784	-30.0
Copiers and Fax	404074080	292489800	-38.1
Labels	15607780	10007040	-56.0

fig.4 Sales on 2011 and 2012

Most the growth sales are lead the increases, shown by a positive value. But there are some sub-category products that got a decline in sales from 2011 to 2012 which shown by a negative value. *Labels, Copiers & Fax and tables* are the categories that got a decline in sales the most.

### 3. Promotion Effectiveness and Efficiency by Years

```
SELECT YEAR(order_date) years,
       SUM(sales) sales,
       SUM(discount_value) 'promotion value',
       ROUND( SUM(discount_value)*100/SUM(sales), 2) 'burn rate (%)'
FROM dqlab_sales_store
WHERE order_status = 'Order Finished'
GROUP BY 1;
```

In this project, burn rate analysis is used to knowing the effectiveness and efficiency of promotions. It does by comparing the sum of promotion value with total sales. Here DQLab hopes that the burn rate doesn't exceed 4.5%. Burn rate calculating is shown below :

$$\text{burn rate}(\%) = \frac{\text{total discount}}{\text{total sales}} \times 100$$

output :

years	sales	promotion value	burn rate (%)
2009	4613872681	214330327	4.65
2010	4059100607	197506939	4.87
2011	4112036186	214611556	5.22
2012	4482983158	225867642	5.04

fig.5 Burn rate by year

The results inform us that burn rates are above 4.5% for each year as overall. This indicates that the promotions have been carried out haven't been able to reduce the burn rate to a maximum of 4.5%. We can figure out what is the product which made a significant contribution causing the burn rate to be higher than expected by grouping the query by each product.

#### 4. Promotion Effectiveness and Efficiency by Product Sub-Category

```
SELECT product_sub_category,
       product_category,
       SUM(sales) sales,
       SUM(discount_value) promotion_value,
       ROUND(SUM(discount_value)*100/SUM(sales),2) 'burn rate (%)'
FROM dqlab_sales_store
WHERE YEAR(order_date) = 2012
      AND order_status = 'Order Finished'
GROUP BY product_sub_category,
         product_category
ORDER BY 5;
```

output :



product_sub_category	product_category	sales	promotion_value	burn rate (%)
Rubber Bands	Office Supplies	3837880	117324	3.06
Envelopes	Office Supplies	58629280	2334321	3.98
Chairs & Chairmats	Furniture	654168740	26623882	4.07
Tables	Furniture	388993784	16348689	4.20
Telephones and Communication	Technology	422287514	18800188	4.45
Labels	Office Supplies	10007040	452245	4.52
Office Furnishings	Furniture	178927480	8233849	4.60
Paper	Office Supplies	126896160	6224694	4.91
Copiers and Fax	Technology	292489800	14530870	4.97
Computer Peripherals	Technology	308014340	15333293	4.98
Storage & Organization	Office Supplies	356714140	18802166	5.27
Pens & Art Supplies	Office Supplies	43818480	2343501	5.35
Appliances	Office Supplies	266131100	14393300	5.41
Office Machines	Technology	811427140	46616695	5.75
Binders and Binder Accessories	Office Supplies	363879200	22338980	6.14
Bookcases	Furniture	159984680	10024365	6.27
Scissors, Rulers and Trimmers	Office Supplies	36776400	2349280	6.39

fig.6 Burn rate by sub-category of product

There are only five sub-category of product that have the burn rate bellow 4.5 %. It shown on the first five rows, they are starting on *Rubber Bands* to *Telephones and Communication*. Whereas the *Labels* have higher 0.02% from the maximum value of expected burn rate by DQLab Store.

It is very interesting because from these results we know that still many sub-category of product which have the burn rate higher than 4.5%.

## 5. Customers Transactions per Year

```
SELECT YEAR(order_date) years,
       COUNT(DISTINCT customer) 'number of customer'
FROM dqlab_sales_store
WHERE order_status = 'Order Finished'
GROUP BY 1;
```

The calculation of customers number for each year doesn't involve the duplicate value. That is why DISTINCT is used in this query for getting the unique value of customers number.

output :

years	number of customer
2009	585
2010	593
2011	581
2012	594

fig.7 Number of customers by year

The number of customers isn't changing significantly overall. But fortunately, we didn't get a significant decline in customers. We can see that the number of customers tends to be in the values around 580–590.

## 6. New customers over the years

```
SELECT YEAR(first_order) years,
       COUNT(customer) 'new customers'
FROM (
  SELECT customer,
         MIN(order_date) first_order
  FROM dqlab_sales_store
  WHERE order_status = 'Order Finished'
  GROUP BY 1) first
GROUP BY 1;
```

To get the number of new customers for each year, we only need the data that shows the first time transaction from each customer. We can get it by applying MIN() function on the *first\_order* field then calculate the number of customers.

output :

years	new customers
2009	585
2010	141
2011	38
2012	11

fig.8 Number of new customers by year

The growth of new customers for each year is decreasing. It gets extreme in 2012 that only there 11 new customers. But if we back on the result before (fig.7) the number of customers tends to remain overall. This informs us that many previous



customers still back to DQLab Store to do the transaction besides the new customers are decreasing.

## Summary

According to data analysis that we have done, we can conclude that :

1. The total sales and order number of DQLab Store are fluctuated as overall, where the highest total sales happened in 2009. Meanwhile, the number of orders increased except in 2011 though the increase is not too significant.
2. Total sales based on the sub-category of products got increasing between 2011 and 2012 as overall. But there some of them is got sales decreasing, they are Appliances, Bookcases, Tables, Labels, Copiers & Fax.
3. The burn rate over the year still above 4.5%. This indicates that the effectiveness and efficiency of promotions haven't been reaching as DQLab Store hopes.
4. There are many products that have a burn rate above 4.5%. This causes the overall burn rate is above 4.5% for each year. The only products that have a burn rate bellow 4.5% are *Rubber Bands, Envelopes, Chairs & Chairmats, Tables and Telephones & Communication*.
5. The number of customers over the years tends to remain, it is around 580–590.
6. The number of new customers decreasing over the years, with the lowest number of new customers is 11 in 2012.

[Sales Performance](#)[Data Analysis](#)[MySQL](#)[Follow](#)

Published in The Startup

853K followers · Last published 2 days ago

Get smarter at building your thing. Follow to join The Startup's +8 million monthly readers & +772K followers.

[Follow](#)

## Written by Adi Sugiarto

17 followers · 5 following

I'm a data enthusiast and a beginner in webite development

### Responses (1)



Write a response

What are your thoughts?



KRUPALI RAO

Apr 29, 2021



Thank you..I loved it and want to see more. Can you please make Dataset available to do practice?





1 reply

[Reply](#)

### More from Adi Sugiarto and The Startup



 In The Startup by Ellen Frances 

## If You Write A Book, Don't Tell Anyone

Learn from my regret; stay silent.

 Jun 17  3.7K  159



 In The Startup by Divad Sanders 

## 4 Boring Startup Ideas Screaming to Be Built (and How to Build Them)

Build the unsexy thing. Charge for outcomes. Repeat.

 Apr 8  5.8K  180





 In The Startup by Divad Sanders 

## 4 Dumb-Smart Startup Ideas That People Keep Ignoring (and How to Build Them)

Simple. Overlooked. Six-figure obvious.

★ Jun 16 🖱 412 💬 12



 In The Startup by Jano le Roux

## Zuckerberg's \$45 Billion Metaverse Just Died The Most Embarrassing Death

Zuckerberg spent \$45 billion building virtual worlds. Now he's copying Apple's 20-year-old playbook.

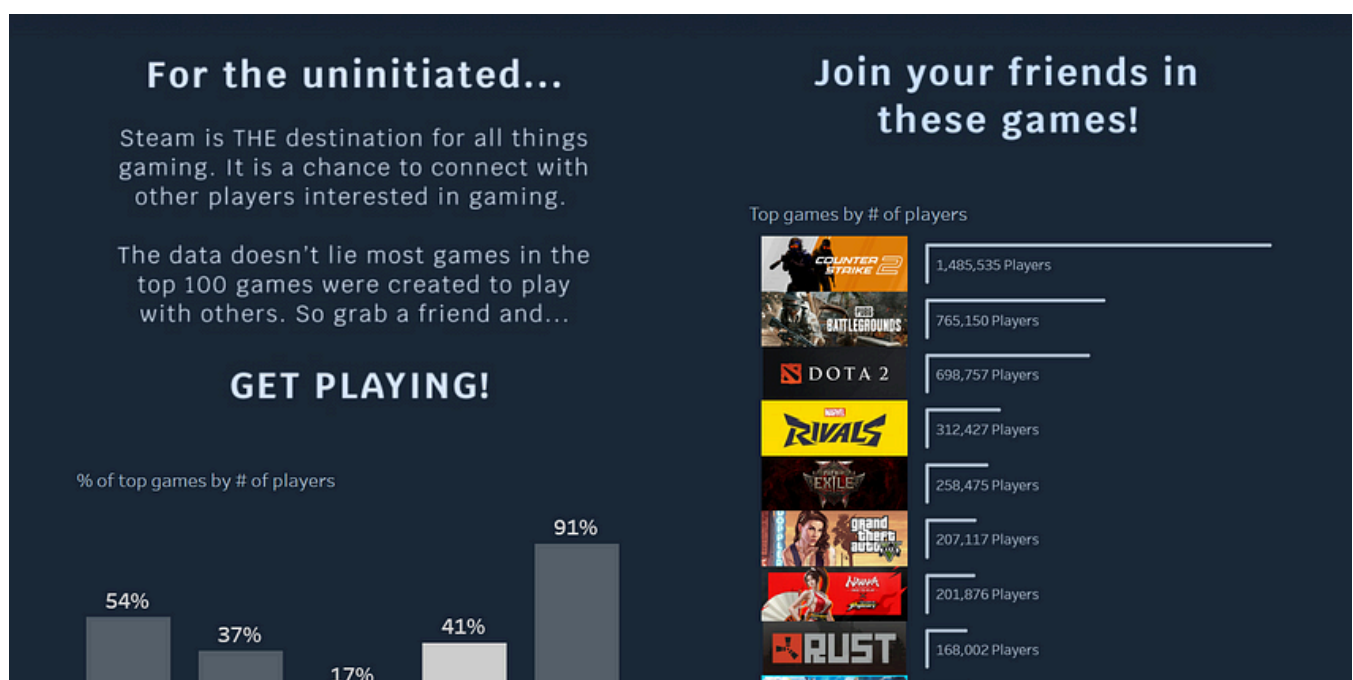
★ Jun 5 🖱 1K 💬 42



See all from Adi Sugiarto

See all from The Startup

## Recommended from Medium



Matt

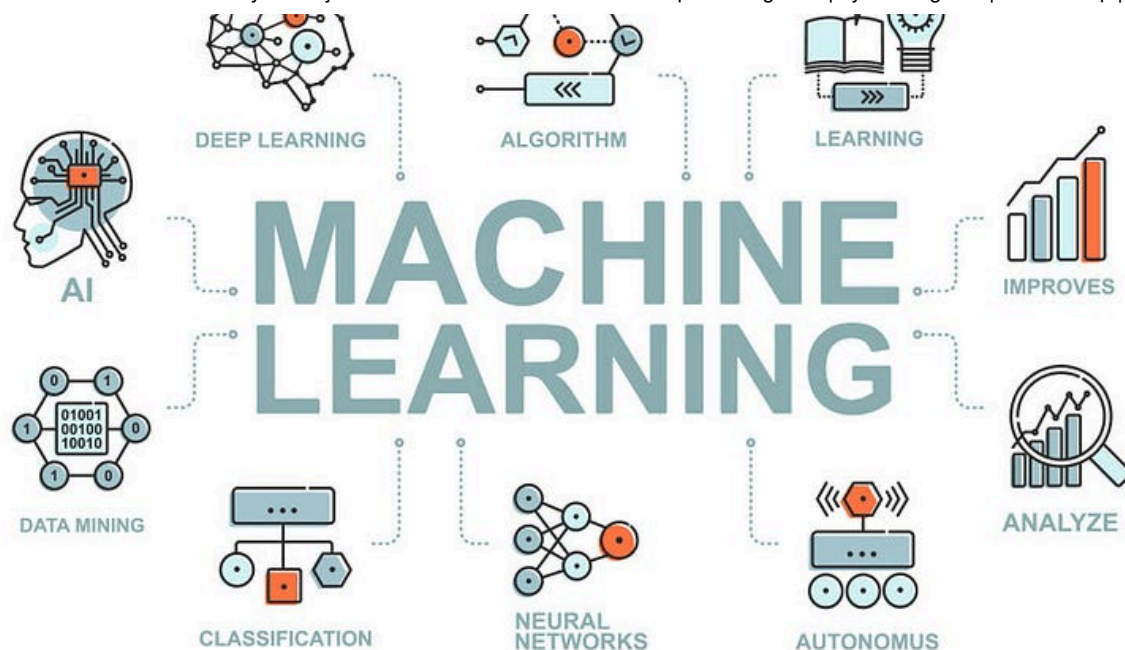
## My Journey with Makeover Monday

As you start your journey with data visualization and Tableau, I recommend the Makeover Monday community project as a great starting point...

Apr 22 🖱 120







 October

## Machine Learning basics in 10 minutes

Machine learning is how machines learn to do stuff. Stuff could be predicting the next apocalypse or your grocery bill based on some data...

★ Jun 3 🖱 52



 Andy Proehl

## Delivering insight not data — Measuring public relations as a case study

Nobody asks for a dashboard. They want to know what's going on and what to do about it. AI will help but we also need better data...

Jun 26  57  2

Vijay Gadhave

## Nested vs. Normalized vs. Denormalized in BigQuery

Note: If you're not a medium member, [CLICK HERE](#)

Jun 27  4

In The BI Corner by Isabelle Bittar

## Smarter Python Visuals in Power BI: 5 UX Tips for Better Insights

From violin plots to field parameters—how to make advanced data visuals both beautiful and understandable.

★ Jun 14 🖱 80 💬 4



 Code With Hannan

## Window Functions in SQL: The Hidden Superpower in Your Toolbox

### Introduction

★ Jun 24 🖱 5



See more recommendations