**SQL Question – Fast & Furious**  
Shilo Babila & Pe’er Nahman

* **What is the volume of shopping and customers in each of the stores?**

select store\_id, count(distinct order\_id) cnt\_order, COUNT(distinct customer\_id) cnt\_customer

from [dbo].[orders]

group by store\_id

-- I use 'group by' the calculate the volume in each store, how much order, and how much customers buy there (one customer can do multi order).  
-- the 'distinct' in 'cnt\_customer' is necessary because we know one value can be show several times.

[Option to add the name of store (now only number)]

* Customers usually buy bike once, rarely they came back again.
* **What are the most profitable categories?**

select pro.category\_id,

sum(cast(item.list\_price as float)

\*item.quantity

\*(1-cast(item.discount as float))

) income\_by\_category

from [dbo].[order\_items] item

join [dbo].[products] pro

on item.product\_id = pro.product\_id

group by pro.category\_id

order by income\_by\_category desc

-- I notice not all info in [dbo].[order\_items] -so I did 'join'

-- then group by category and do sum

-- multiplied by quantity and discount

-- NOTE! - can be calculate only the income, for profit need to know the expense

-- for the most category will first we use 'order by'

* Most of the profit comes from useful bikes, and less trend.
* **Which range of price sell more? Is this different in each store?**

select store\_id,price\_tag, sum(cast(quantity as int)) cnt\_per\_price\_tag

from

(select store\_id,i.order\_id,product\_id,quantity,list\_price,

case

when cast(list\_price as float)<250 then '1. Under250'

when cast(list\_price as float)<500 then '2. 250-500'

when cast(list\_price as float)<750 then '3. 500-750'

when cast(list\_price as float)<1000 then '4. 750-1000'

when cast(list\_price as float)<1500 then '5. 1000-1500'

when cast(list\_price as float)<2000 then '6. 1500-2000'

when cast(list\_price as float)<3000 then '7. 2000-3000'

when cast(list\_price as float)<6000 then '8. 3000-6000'

else '9. Over6000'

end price\_tag

from [dbo].[order\_items] i

join [dbo].[orders] o

on i.order\_id = o.order\_id)q1

group by rollup(price\_tag,store\_id)

-- 1. distinct 'list\_price' and order them to see easily the all prices

-- 2. notice that is comper them like a text so we need to convert them

-- 3. decide to divide by 250, 500, 750, 1000, 1500, 2000, 3000, 6000 or more

-- 4. use case to give each the right title

-- 5. join 'orders' because I dont have info about which store sell the order

-- 6. make it subquary and use 'group by rollup' by price\_tag and store\_id to calculate sum quantity

[Option to add the name of store (now only number)]

* store 2 sell more than double from the other expensive bikes
* **How much is the profit of each category in each store?**

-- I have a question that look like that {What are the most profitable categories}, so I use it in the base.  
-- The stores information is missing so I joined another table to the query.  
-- changed the 'group by' to 'rollup' and add the store\_id.  
-- join category table for showing the name (not a number)

select cat.category\_name ,ord.store\_id,

sum(cast(item.list\_price as float)

\*item.quantity

\*(1-cast(item.discount as float))

) income\_by\_category

from [dbo].[order\_items] item

join [dbo].[products] pro

on item.product\_id = pro.product\_id

join [dbo].[orders] ord

on ord.order\_id = item.order\_id

join [dbo].[categories] cat

on pro.category\_id = cat.category\_id

group by rollup(cat.category\_name, ord.store\_id)

* Shopping trends are the same across all stores

* **Which month sell more?**

select year(order\_date) year\_order, month(order\_date) month\_order, COUNT(distinct order\_id) cnt\_order

from [dbo].[orders]

group by rollup (month(order\_date), year(order\_date))

-- first I isolated the year and the month

-- after that I group them by the year and month and count the orders.

* Until 04.2018 the sells was the same, but in April was very high sell that maybe cause to next months to get really low sells.
* **Which store leads in meeting the delivery times? and which store needs improvement?**

select store\_id,delivery\_in\_time, cast(cnt\_orders\*100.0/total\_orders as decimal(4,2)) persent\_in\_time

from

(select \*, sum(cnt\_orders) over(partition by store\_id) total\_orders

from

(select store\_id, delivery\_in\_time, COUNT(distinct order\_id) cnt\_orders

from

(select order\_id,store\_id, --

case

when required\_date > shipped\_date then 'NotInTime'

else 'InTime'

end 'delivery\_in\_time'

from [dbo].[orders])q1

group by delivery\_in\_time, store\_id

)q2

)q3

-- 1. add new culumn 'delivery\_in\_time' that tell us if order delivired in time.

-- 2. count them per store and 'delivery\_in\_time'

-- 3. for knowing the total order per store, I use 'sum over()'

-- 4. finally for percent we to divide cnt in total.

* Store 1 is the best, after that store 2 although it has a lot of orders, but store 3 need improvement.
* **What is the company's profit in each of the stores? Is there a big difference in profits between stores?**

-- I need to calculate total price of each order including the discount(from order items)  
-- once I have total price per order, i need to connect it to orders table to see which store made this order

with order\_price as (

select oi.order\_id,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'total\_order'

from [dbo].[order\_items] oi

group by oi.order\_id

)

select o.store\_id,

sum(op.total\_order) as 'store\_profit'

from [dbo].[orders] as o

join order\_price as op

on o.order\_id = op.order\_id

group by o.store\_id

-- Firstly, I didn't think about WITH AS. after writing the query that is now inside WITH AS i understood that i just made 'new table'

-- and that i need to JOIN this 'new table' with orders table, and it would be the best to use WITH AS

* The profit difference is big.
* store 3 is least profitable with 764K.
* store 1 in the middle with 1.5M.
* store 2 the most with almost 4.8M.
* **What products would you recommend for the company to continue selling and marketing? Which products would you recommend to change marketing / remove?**

------------- I want to check total profit of each product, maybe add amount sold (order items)

------------- I might wanna check it per category (cross with products table)... (maybe children's bike are not so profitable, but i need to have them too, so maybe only keep a bit)

------------- maybe check stocks, if i have a lot of some item that isn't bought often i need to lower the price

------------- mid-work i decided to really join products, but with right join, in order to see if there are products that weren't sold at all (i would see nulls)

------------- mid-work i thought that i also need to consider model year... the newer the product the less sales it has.

------------- lets say profitable is: above 7 in 2018, above 30 in 2017, above 130 in 2016

with is\_profit\_table as(

select distinct q1.product\_id,

case

when (q1.model\_year = 2018 and q1.amount\_sold>=7) OR

(q1.model\_year = 2017 and q1.amount\_sold>=30) OR

(q1.model\_year = 2016 and q1.amount\_sold>=130) then 'profitable'

else 'not profitable'

end as 'is\_profitable'

from

(select p.product\_id, p.model\_year,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'total\_product\_profit',

sum(cast(oi.quantity as int)) as 'amount\_sold'

from [dbo].[order\_items] as oi

right join [dbo].[products] p

on oi.product\_id = p.product\_id

group by p.product\_id, p.model\_year

) q1

),

------------ I also want to check now what was the last time the product was sold.

------------ i will check the latest order and consider it as current date

------------ i will say that if a produt wasn't sold in the last 12 months it is not profitable

last\_bought\_table as (

select a2.product\_id,

case

when datediff(month, a2.last\_bought, a2.c\_date)<=12 then 'during\_last\_year'

else 'more\_than\_year'

end as 'when\_bought'

from

(select distinct

\*,

max(a1.last\_bought) over() as 'c\_date'

from

(select --top 1

oi.product\_id,

o.order\_date,

LAST\_VALUE(o.order\_date) over (partition by oi.product\_id order by oi.product\_id) as 'last\_bought'

from [dbo].[order\_items] oi

join [dbo].[orders] o

on oi.order\_id = o.order\_id

) a1

where a1.order\_date = a1.last\_bought

) as a2

)

----------- the query above will give us the last order which i consider as current date... and if the product was sold in last year

----------- i figured i want to save both queries from above with WITH AS and try to see, if a product is not profitable + wasn't sold last year, we should remove it

select \*

from

(select ipt.product\_id, lbt.when\_bought, ipt.is\_profitable,

case

when lbt.when\_bought = 'more\_than\_year' and ipt.is\_profitable = 'not profitable' then 'remove'  
 when lbt.when\_bought is null then 'remove'

when lbt.when\_bought = 'during\_last\_year' and ipt.is\_profitable = 'profitable' then 'great product, keep high stock!'

else 'ok product, keep medium stock'

end as 'remove or keep'

from last\_bought\_table as lbt

right join is\_profit\_table as ipt

on lbt.product\_id = ipt.product\_id

) as a3

order by a3.[remove or keep]

--------- FINALYYYY!!!! these are the products we need to REMOVE, keep medium stock, or high stock

--------- i had to try several times and had to change my way of thinking a bit, each time i had a clearer vision of what i want and need to do.

* Every product got a tag
* **Which brands are worthwhile to own?**

----------- i want to check the profit of each brand, in each store

----------- i will use order items, orders, products, and brands table

----------- i plan to join order items with products and orders, sum the price grouped by brand\_id p

----------- adding brands to view names

select q1.brand\_id,

b.brand\_name,

q1.brand\_profit,

concat(round(((q1.brand\_profit/q1.total\_profit)\*100), 2), '%') as 'percent\_of\_total\_profit'

from

(select p.brand\_id,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'brand\_profit',

max(sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount))) over() as 'total\_profit'

from [dbo].[order\_items] oi

join [dbo].[products] p

on oi.product\_id = p.product\_id

left join [dbo].[orders] o

on oi.order\_id = o.order\_id

group by rollup( p.brand\_id)

) q1

join [dbo].[brands] b

on q1.brand\_id = b.brand\_id and q1.brand\_id is not null

* brand 9(trek) is really worth owning. brand 1 and 8 (electra and surly) are kinda good too
* the rest of the brands are not as good as the mentioned above
* brand 6(strider) is making very little profit, consider removing it
* **what is the total sales of each category, per brand?**

----- joining order items and products, to see sum price grouped by brand and cat..

select q1.category\_id,

q1.brand\_id,

round(((q1.brand\_profit/q1.category\_profit)\*100), 2) as 'brand\_%profit\_of\_category'

from

(select p.category\_id,

p.brand\_id,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'brand\_profit',

last\_value(sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount))) over (partition by p.category\_id order by p.category\_id) 'category\_profit'

from [dbo].[order\_items] oi

join [dbo].[products] p

on oi.product\_id = p.product\_id

group by rollup (p.category\_id, p.brand\_id)

) q1

where q1.brand\_id is not null

* this query helps me to see in each category, how profitable is every brand.
* i wrote additonal query that check what are the products where the %profit is less than 15
* but i decided to remove it, because some of them were actually profitable, because they were unique in their brand&category
* so even if they make only 5%... 5% for one product is great value
* you can use this table to track profits of each brand in each cat
* **Percent of profit of each brand in each store**

------ from this query i can find out how much stock from each brand in each store

------ joining order itemsand products and orders.

select q1.store\_id,

q1.brand\_id,

round(((q1.brand\_profit/q1.store\_profit)\*100), 2) as 'brand\_%profit\_of\_store'

from

(select o.store\_id,

p.brand\_id,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'brand\_profit',

last\_value(sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount))) over (partition by o.store\_id order by o.store\_id) 'store\_profit'

from [dbo].[order\_items] oi

join [dbo].[orders] o

on oi.order\_id = o.order\_id

join [dbo].[products] p

on oi.product\_id = p.product\_id

group by rollup (o.store\_id, p.brand\_id)

) q1

where q1.brand\_id is not null

order by q1.store\_id, round(((q1.brand\_profit/q1.store\_profit)\*100), 2)

* I can see that in each store, the order is almost identical...
* **trying to find interesting info based on customers details**

----------- firstly i will find incomes from each state, city.

----------- i will sum the profit from order items, join with orders (customer id) and with customers (state, city)

----------- i will count the customers

----------- lets say i need to advertise the company in cities with 10 customers or less

select c.state,

c.city,

sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) as 'profit',

count(c.customer\_id) 'customer\_amount'

from [dbo].[order\_items] oi

join [dbo].[orders] o

on oi.order\_id = o.order\_id

join [dbo].[customers] c

on o.customer\_id = c.customer\_id

group by rollup(c.state, c.city)

having count(c.customer\_id) <= 10 or sum(cast(oi.list\_price as float)\*oi.quantity\*(1-oi.discount)) <15000

* some cities actually have nice profit - but i want to reach out for more customers
* some cities have nice amount of customers, but seems like they are not buying a lot (due to profit)
* I will advertise more in these cities.