

Thelook_Ecommerce 2

Using the “bigquery-public-data.thelook_ecommerce” data set, perform the following tasks:

1. JOIN the `order_item` table with the `products` table and return the joined table

```
SELECT a.*, b.*
FROM `bigquery-public-data.thelook_ecommerce.order_items` a
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b
on a.product_id = b.id;
```

2. Write a query that shows each product categories sorted by *number of orders* per category

```
SELECT category, count(distinct order_id) as orders_d
FROM `bigquery-public-data.thelook_ecommerce.order_items` a
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b
on a.product_id = b.id
group by category
order by orders_d desc;
```

3. Using the last query as the base, add the average margin per order for each category. Notice how top selling categories may not have the highest relative margins. (*Hint: margin = retail_price - cost*)

```
SELECT category,
count(distinct order_id) as orders_d,
(SUM(b.retail_price) - SUM(b.cost))/count(distinct order_id) as
avg_margin_per_order FROM
`bigquery-public-data.thelook_ecommerce.order_items` a
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b
on a.product_id = b.id
group by category
order by avg_margin_per_order desc;
```

4. Which product(s) is/are the most popular (number sold)?

```
SELECT name,  
count(product_id) as nr_count  
FROM `bigquery-public-data.thelook_ecommerce.order_items` a  
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b  
on a.product_id = b.id  
GROUP BY name  
ORDER BY nr_count desc;
```

5. Which product takes the longest time (in terms of number of minutes) from shipping to delivery?

```
SELECT name,  
avg(date_diff(delivered_at, shipped_at, minute)) as avg_hr_ship_deliver,  
count(product_id) as nr_count  
FROM `bigquery-public-data.thelook_ecommerce.order_items` a  
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b  
on a.product_id = b.id  
GROUP BY name  
ORDER BY avg_hr_ship_deliver desc;
```

6. Which product takes the shortest time?

```
SELECT name,  
avg(date_diff(delivered_at, shipped_at, minute)) as  
avg_hr_ship_deliver, count(product_id) as nr_count  
FROM `bigquery-public-data.thelook_ecommerce.order_items` a  
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b  
on a.product_id = b.id  
GROUP BY name  
ORDER BY avg_hr_ship_deliver asc;
```

7. Back to the query at point 6, add a WHERE condition that excludes *items that have not been delivered yet*:

```
SELECT name,  
avg(date_diff(delivered_at, shipped_at, minute)) as  
avg_hr_ship_deliver, count(product_id) as nr_count  
FROM `bigquery-public-data.thelook_ecommerce.order_items` a  
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b  
on a.product_id = b.id  
where delivered_at is not null  
GROUP BY name  
ORDER BY avg_hr_ship_deliver asc;
```

8. When you answered question five, all those items that took the longest time from shipping to delivery were only ordered once or twice; answer to the same question but only for those products that were sold at least 5 times.

```
SELECT name,  
avg(date_diff(delivered_at, shipped_at, minute)) as  
avg_min_ship_deliver, count(product_id) as nr_count  
FROM `bigquery-public-data.thelook_ecommerce.order_items` a  
LEFT JOIN `bigquery-public-data.thelook_ecommerce.products` b  
on a.product_id = b.id  
where delivered_at is not null  
GROUP BY name  
HAVING count(product_id) > 5  
ORDER BY avg_min_ship_deliver desc;
```