六 分组对比

学习目标

- 掌握创建对比分析报表的方法

1 按行比较

- 第一种方法：我们将要对比的两组数据放到不同行中
- 需求：统计运往北美地区和除北美地区外的订单数量

shipping_continent	order_count
North America	180
Other	650

- 使用如下SQL；

```
WITH orders_by_group AS (
  SELECT
    order_id,
    CASE
      WHEN ship_country IN ('USA', 'Canada', 'Mexico')
        THEN 'North America'
      ELSE 'Other'
    END AS shipping_continent
  FROM orders
)
SELECT
  shipping_continent,
  COUNT(order_id) AS order_count
FROM orders_by_group
GROUP BY shipping_continent;
```

- 在内部查询中，使用 `CASE WHEN` 对订单按 `ship_country` 列进行分类(“ North America”或“ Other”)。 分类结果保存在名为 `shipping_continent` 的列中
- 在外部查询中，我们将内部查询的结果按 `shipping_continent` 列进行分组，并使用 `COUNT（order_id）` 对匹配的订单进行计数

练习69

- 需求：统计“ WA ”地区的员工和所有其他员工处理的订单总量
- 结果显示两列： `employee_region` （员工所在区域，“ WA”或“ Not WA”）和 `order_count` 。

```
WITH orders_by_group AS (
  SELECT
    order_id,
    CASE
      WHEN region = 'WA' THEN 'WA'
      ELSE 'Not WA'
    END AS employee_region
  FROM employees e
  JOIN orders o
    ON e.employee_id = o.employee_id
)
SELECT
  employee_region,
  COUNT(order_id) AS order_count
FROM orders_by_group
GROUP BY employee_region;
```

查询结果

employee_region	order_count
WA	605
Not WA	225

练习70

- 需求：统计不同库存水平的产品数量（`product_id`）计数
- 结果显示两列：`availability` 和 `product_count`
  - `'Low'` 库存（`units_in_stock`）<= 10
  - `'Average'` 库存（`units_in_stock`）10~30
  - `'High'` 库存（`units_in_stock`）>30

```
WITH products_by_group AS (
  SELECT
    product_id,
    CASE
      WHEN units_in_stock > 30 THEN 'High'
      WHEN units_in_stock > 10 THEN 'Average'
      ELSE 'Low'
    END AS availability
  FROM products)
SELECT
  availability,
  COUNT(product_id) AS product_count
FROM products_by_group
GROUP BY availability;
```

查询结果

availability	product_count
High	34
Average	29
Low	14

练习71

- 需求：统计法国客户和其他国家/地区的客户消费的总金额（折扣后）
- 结果显示两列：`customer_country`（“ France”或“其他”）和 `discount_revenue`（四舍五入到小数点后两位）

```
WITH orders_by_group AS (
  SELECT
```

```
    o.order_id,
  CASE
    WHEN country = 'France' THEN 'France'
    ELSE 'Other'
  END AS customer_country
FROM orders o
JOIN customers c
  ON o.customer_id = c.customer_id
)
SELECT
  customer_country,
  ROUND(SUM(quantity * unit_price * (1 - discount)), 2) AS
discount_revenue
FROM orders_by_group obg
JOIN order_items oi
  ON obg.order_id = oi.order_id
GROUP BY customer_country;
```

查询结果

customer_country	discount_revenue
Other	1148233.04
France	79742.42

2 按列对比

- 需求：统计已经发货的订单数量和尚未发货的订单数量：

orders_shipped	orders_pending
809	21

看下面的SQL

```
SELECT
  COUNT(CASE
    WHEN shipped_date IS NOT NULL
      THEN order_id
  END) AS orders_shipped,
  COUNT(CASE
    WHEN shipped_date IS NULL
      THEN order_id
  END) AS orders_pending
FROM orders;
```

- 在上面的查询中使用多个 `COUNT（CASE WHEN ...）` 或 `SUM（CASE WHEN ...）` 语句。 每个语句用于在新列中显示不同组的数据
- 在前面介绍的方法中，每组数据用不同行表示，使用单个“ CASE WHEN”语句。 在这里，每组数据用不同列表示，当要对比的分组不多时，可以使用这种方式

练习72

- 需求：统计素食和非素食的商品数量
- 结果包含两列： `non_vegetarian_count` 和 `vegetarian_count`

提示: 非素食产品类别 `category_id` 的值是6或8

```
SELECT
  COUNT(CASE
    WHEN category_id IN (6, 8)
      THEN product_id
  END) AS non_vegetarian_count,
  COUNT(CASE
    WHEN category_id NOT IN (6, 8)
      THEN product_id
  END) AS vegetarian_count
FROM products;
```

查询结果

non_vegetarian_count	vegetarian_count
18	59

练习73

- 需求：统计运往加拿大的订单中的全价出售的订单数量和打折订单的订单数量
- 显示两列：`full_price_count` 和 `discounted_price_count`

```
SELECT
  SUM(CASE
    WHEN discount = 0
      THEN 1
  END) AS full_price_count,
  SUM(CASE
    WHEN discount != 0
      THEN 1
  END) AS discounted_price_count
FROM orders o
JOIN order_items oi
  ON o.order_id = oi.order_id
WHERE ship_country = 'Canada';
```

查询结果

full_price_count	discounted_price_count
44	31

3 多列对比

- 将对比分类按列显示的一个优点是我们可以在报表中添加其它信息

```
SELECT
  customer_id,
  COUNT(CASE
    WHEN shipped_date IS NOT NULL
      THEN order_id
  END) AS orders_shipped,
  COUNT(CASE
    WHEN shipped_date IS NULL
      THEN order_id
  END) AS orders_pending
FROM orders
GROUP BY customer_id;
```

查询结果(部分)：

customer_id	orders_shipped	orders_pending
ALFKI	6	0
ANATR	4	0
ANTON	7	0
...	...	...

- 我们在 `SELECT` 子句中添加 `customer_id` 列，并在“ GROUP BY”子句中对它进行分组。 这样，我们可以看到每个客户已发货或待处理的订单数量。

练习 74

- 需求：统计每位员工处理的运往德语地区的订单总金额和运往其它国家的订单总金额
  - “ dach_orders”（运往德国，奥地利或瑞士的订单），“other_orders”（运往所有其他国家的订单）。
- 结果显示以下列： `employee_id` ， `first_name` ， `last_name` ， `dach_orders` ， `other_orders`
  - 员工ID，员工名字，员工姓氏，运往德语地区的订单数量，运往其他国家的订单数量

```
SELECT
```

```
  e.employee_id,
  e.first_name,
  e.last_name,
  SUM(CASE
    WHEN ship_country IN ('Switzerland', 'Germany', 'Austria')
      THEN quantity * unit_price
    ELSE 0
  END) AS dach_orders,
  SUM(CASE
    WHEN ship_country NOT IN ('Switzerland', 'Germany',
'Austria')
      THEN quantity * unit_price
    ELSE 0
  END) AS other_orders
FROM orders o
JOIN order_items oi
  ON o.order_id = oi.order_id
JOIN employees e
  ON o.employee_id = e.employee_id
GROUP BY e.employee_id,
  e.first_name,
  e.last_name;
```

查询结果

employee_id	first_name	last_name	dach_orders	other_orders
5	Steven	Buchanan	9349.6	66218.15
6	Michael	Suyama	21136.16	57061.94
4	Margaret	Peacock	66044.78	183694.67
3	John	Smith	78676.17	134375.13
9	Anne	Dodsworth	28447.75	54516.25
1	Nancy	Davolio	49150.91	152992.8
10	John	Smith	0	448
8	Laura	Callahan	39573.85	93727.18
2	Andrew	Fuller	76094.4	101654.86

7 employee_id	Robert first_name	King last_name	48583.14 dach_orders	92712.85 other_orders

4 占比对比

- 在创建对比分析的报表时，我们经常将对比的结果转换成百分比
- 需求：统计每一个员工处理的，收货地为德国和美国的订单中，已发货的订单占已比。
- 我们想要如下结果：

first_name	last_name	germany_perc	usa_perc
Nancy	Davolio	100	95.24
Andrew	Fuller	92.86	100
Janet	Leverling	100	100
...	...	...	...

查询语句如下：

```
WITH germany_orders AS (
  SELECT
    employee_id,
    COUNT(CASE
      WHEN shipped_date IS NOT NULL
        THEN order_id
    END) AS count_shipped,
    COUNT(order_id) AS count_all
  FROM orders o
  WHERE o.ship_country = 'Germany'
  GROUP BY employee_id
),
usa_orders AS (
  SELECT
    employee_id,
    COUNT(CASE
      WHEN shipped_date IS NOT NULL
        THEN order_id
    END) AS count_shipped,
    COUNT(order_id) AS count_all
```

```
  FROM orders o
  WHERE o.ship_country = 'USA'
  GROUP BY employee_id
)
SELECT
  e.first_name,
  e.last_name,
  ROUND(ge_or.count_shipped / ge_or.count_all * 100, 2) AS
germany_perc,
  ROUND(us_or.count_shipped / us_or.count_all * 100, 2) AS
usa_perc
FROM germany_orders ge_or
JOIN usa_orders us_or
  ON ge_or.employee_id = us_or.employee_id
JOIN employees e
  ON ge_or.employee_id = e.employee_id
  OR us_or.employee_id = e.employee_id;
```

- 上面的SQL中，我们通过两个CTE来分别处理目的地为德国和美国的订单
  - 在每个CTE中都按员工ID分组，分别统计运输到德国和美国的已发货订单和总订单数量
  - 在外部查询中，进行百分比计算

练习75

- 需求：统计ID为1和2的员工处理的订单中，不同收货国家的订单金额，占发往该国的总订单金额百分比（折扣前）保留两位有效数字
- 查询结果显示3列：
  - 订单要运送到的国家/地区（`ship_country`）
  - `percentage_employee_1`-ID为1的员工所产生的收入占比
  - `percentage_employee_2`-ID为2的员工所产生的收入占比

```
WITH revenue_employee_1 AS (
  SELECT
    ship_country,
    SUM(CASE
      WHEN employee_id = 1
```

```
        THEN unit_price * quantity
      ELSE 0
    END) AS employee_revenue,
    SUM(unit_price * quantity) AS total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY ship_country),
revenue_employee_2 AS (
  SELECT
    ship_country,
    SUM(CASE
      WHEN employee_id = 2
        THEN unit_price * quantity
      ELSE 0
    END) AS employee_revenue,
    SUM(unit_price * quantity) AS total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY ship_country)
SELECT
  re_1.ship_country,
  ROUND(re_1.employee_revenue / re_1.total_revenue * 100, 2)
AS percentage_employee_1,
  ROUND(re_2.employee_revenue / re_2.total_revenue * 100, 2)
AS percentage_employee_2
FROM revenue_employee_1 re_1
JOIN revenue_employee_2 re_2
  ON re_1.ship_country = re_2.ship_country
```

查询结果

ship_country	percentage_employee_1	percentage_employee_2
Venezuela	16.42	4.92
Sweden	13.11	14.26
Ireland	5.42	22.91
Portugal	14.21	0.00
Finland	9.25	31.40
France	14.94	11.45
UK	18.68	5.70
Spain	6.39	5.47
Belgium	2.14	8.16
Italy	7.46	34.14
USA	17.60	8.64
Germany	10.92	23.01
Canada	16.86	16.52
Argentina	8.46	5.88
Denmark	19.77	6.89
Switzerland	12.56	0.00
Norway	30.14	10.85
Brazil	26.10	8.99
Austria	13.12	14.20
Mexico	21.98	9.10

练习76

- 需求：对每位员工，统计运送到美国和德国的订单总金额（折扣前）相对于该员工所有订单总金额的百分比（保留两位有效数字）
- 显示以下列：

- `employee_id`.
- `first_name`.
- `last_name`.
- `rev_percentage_usa`.美国订单收入占比
- `rev_percentage_germany` 德国订单收入占比

```
WITH usa_revenue AS (
  SELECT
    employee_id,
    SUM(CASE
      WHEN ship_country = 'USA'
        THEN unit_price * quantity
      ELSE 0
    END) AS country_revenue,
    SUM(unit_price * quantity) AS total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY employee_id),
germany_revenue AS (
  SELECT
    employee_id,
    SUM(CASE
      WHEN ship_country = 'Germany'
        THEN unit_price * quantity
      ELSE 0
    END) AS country_revenue,
    SUM(unit_price * quantity) AS total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY employee_id)
SELECT
  e.employee_id,
  e.first_name,
  e.last_name,
  ROUND(us_rev.country_revenue / us_rev.total_revenue * 100,
2) AS rev_percentage_usa,
  ROUND(ger_rev.country_revenue / ger_rev.total_revenue * 100,
2) AS rev_percentage_germany
FROM usa_revenue us_rev
```

```
JOIN germany_revenue ger_rev
  ON us_rev.employee_id = ger_rev.employee_id
JOIN employees e
  ON e.employee_id = us_rev.employee_id;
```

查询结果

employee_id	first_name	last_name	rev_percentage_usa	rev_percentage_germany
1	Nancy	Davolio	22.95	13.21
2	Andrew	Fuller	12.82	31.67
3	John	Smith	16.34	22.99
4	Margaret	Peacock	20.28	16.53
5	Steven	Buchanan	21.37	11.55
6	Michael	Suyama	23.93	13.27
7	Robert	King	20.06	6.21
8	Laura	Callahan	20.72	20.69
9	Anne	Dodsworth	21.82	19.18
10	John	Smith	0	0

小结

- 本小节中，我们讨论了在单个查询中比较不同业务组的三种方法：

1. 方法1，在CTE中使用 `CASE WHEN` 来标记业务对象。 在外部查询中，我们通过这些标签对行进行分组， **每组对应一行数据。**

```
WITH name_of_cte AS (
  SELECT
    ...,
    CASE
      WHEN ...
      ELSE ...
    END AS group_label
  FROM table_name
)
SELECT
  group_label,
```

```
  COUNT(...)
FROM name_of_cte
GROUP BY group_label;
```

2. 方法2，使用多个 `COUNT / SUM（CASE WHEN ...）` 语句来创建分组。 **每组对应一列数据。**

```
SELECT
  group_column,
  COUNT(CASE
    WHEN ... THEN ...
  END) AS group_one,
  COUNT(CASE
    WHEN ... THEN ...
  END) AS group_two
FROM table_name
GROUP BY group_column;
```

3. 在方法3中，每个组均以**单独的CTE** 进行计算。 然后在外部查询进一步处理得到最终结果

```
WITH group_one AS (...),
group_two AS (...)
SELECT
  group_one.column_name,
  group_two.column_name
FROM group_one
FULL OUTER JOIN group_two
  ON ...;
```

练习77

- 需求：将商品根据价格进行分组，分成高价，中等价格，低价三档
- 报表包含两列数据：
  - `price_category` 单价<=30的产品为低价（`'low'`），单价>90的产品为高价（`'high'`），其余为中等（`'average'`）
  - `product_count` 给定组中的产品数。

```
WITH products_by_group AS (
  SELECT
```

```
    product_id,
    CASE
      WHEN unit_price > 90
        THEN 'high'
      WHEN unit_price > 30
        THEN 'average'
      ELSE 'low'
    END AS price_category
  FROM products
)
SELECT
  price_category,
  COUNT(product_id) AS product_count
FROM products_by_group
GROUP BY price_category;
```

查询结果

price_category	product_count
average	21
high	3
low	53

练习78

- 需求：统计使用不同语言的用户数量，根据国家/地区来判断使用的语言
- 结果显示三列：
  1. `english_count` -来自英国，加拿大，美国和爱尔兰的客户数量
  2. `german_count` -来自德国，瑞士和奥地利的客户数量
  3. `other_count` -来自其他国家的客户数量

```
SELECT
  COUNT(CASE
    WHEN country IN ('UK', 'Canada', 'USA', 'Ireland')
      THEN 'English'
  END) AS english_count,
  COUNT(CASE
```

```
    WHEN country IN ('Germany', 'Switzerland', 'Austria')
      THEN 'German'
  END) AS german_count,
  COUNT(CASE
    WHEN country NOT IN ('UK', 'Canada', 'USA', 'Ireland',
'Germany', 'Switzerland', 'Austria')
      THEN 'Other'
  END) AS other_count
FROM customers;
```

查询结果

english_count	german_count	other_count
24	15	52

练习79

- 需求：创建报表，统计运送到的每个国家中运费值低（小于或等于90.0）和高（大于90.0）的订单所产生的折扣后总收入百分比。
- 报告中显示如下信息：
  - `ship_country`，`percentage_low_freight`，`percentage_high_freight`
  - 目的地国家，高运费收入占比，低运费收入占比（保留两位有效数字）

```
WITH revenue_low_freight AS (
  SELECT
    ship_country,
    SUM(CASE
      WHEN freight <= 90.0
        THEN unit_price * quantity * (1 - discount)
      ELSE 0
    END) AS freight_revenue,
    SUM(unit_price * quantity * (1 - discount)) AS
total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY ship_country),
```

```
revenue_high_freight AS (
  SELECT
    ship_country,
    SUM(CASE
      WHEN freight > 90.0
        THEN unit_price * quantity * (1 - discount)
      ELSE 0
    END) AS freight_revenue,
    SUM(unit_price * quantity * (1 - discount)) AS
total_revenue
  FROM orders o
  JOIN order_items oi
    ON o.order_id = oi.order_id
  GROUP BY ship_country)

SELECT
  lf.ship_country,
  ROUND(lf.freight_revenue / lf.total_revenue * 100, 2) AS
percentage_low_freight,
  ROUND(hf.freight_revenue / hf.total_revenue * 100, 2) AS
percentage_high_freight
FROM revenue_low_freight lf
JOIN revenue_high_freight hf
  ON lf.ship_country = hf.ship_country;
```

查询结果

ship_country	percentage_low_freight	percentage_high_freight
France	49.15	50.85
Germany	37.82	62.18
Brazil	57.11	42.89
Belgium	60.15	39.85
Switzerland	47.19	52.81
Venezuela	62.72	37.28
Austria	15.22	84.78
Mexico	82.4	17.6
USA	34.86	65.14
Sweden	31.9	68.1
Finland	70.2	29.8
Italy	89.91	10.09
Spain	65.51	34.49
UK	66.74	33.26
Ireland	35.29	64.71
Portugal	77.04	22.96
Canada	54.49	45.51
Denmark	67.68	32.32
Poland	100	0
Norway	81.55	18.45
Argentina	61.18	38.82