# Assignment 1 B. Data Analytic Stage

# **Table of Contents**

3. Create the following reports using OLAP queries.	2
a. Simple reports:	2
REPORT 1 (Top n report)	2
REPORT 2 (Top n% report)	3
b. Reports with proper sub-totals:	4
REPORT 3 (Cube)	4
REPORT 4 (Partial Cube)	5
REPORT 5 (Roll-up)	6
REPORT 6(Partial Roll-up)	7
c. Reports with moving and cumulative aggregates:	8
REPORT 7:	8
REPORT 8:	10
d. Reports with Partitions:	11
REPORT 9:	11
4. Business Intelligence (BI) Reports.	12

## 3. Create the following reports using OLAP queries.

## a. Simple reports:

Produce two reports. Each report contains two attributes from two different dimensions, and one fact measurement.

For the report itself, the first report must be about Top n and the second report is Top n%.

#### REPORT 1 (Top n report)

#### (a) The query questions

Find out the top 10 equipment categories by total sales in 2020.

## (b) Your explanation

Managers will be interested in the sales data of the past years, especially the most recent year, and listing the best-selling categories can help managers make decisions on how to place orders in the coming years.

## (c) The SQL commands

**SELECT** \*

**FROM** 

(SELECT t.SALES\_YEAR,c.CATEGORY\_DESCRIPTION, SUM(s.TOTAL\_SALES\_REVENUE) as SALES\$,

RANK() OVER (ORDER BY SUM(TOTAL\_SALES\_REVENUE) DESC ) AS CATEGORY\_RANK

FROM salesFACT s, Sales\_TimeDIM t, categoryDIM c

WHERE c.CATEGORY\_ID = s.CATEGORY\_ID

AND s.SALES\_TIMEID = t.SALES\_TIMEID

AND t.SALES\_YEAR = '2020'

GROUP BY t.SALES\_YEAR,c.CATEGORY\_DESCRIPTION)

WHERE CATEGORY\_RANK <= 10;

## (d) The screenshots of the query results

```
--REPORT 1
    --Top n RANKING
    --Find out the top 10 equipment categories by total sales in 2020.
   □ SELECT *
   FROM
   (SELECT t.SALES YEAR, c.CATEGORY DESCRIPTION, SUM (S.TOTAL SALES REVENUE) as SALES$,
    RANK () OVER (ORDER BY SUM (TOTAL SALES REVENUE) DESC ) AS CATEGORY RANK
    FROM salesFACT s, Sales TimeDIM t, categoryDIM c
    WHERE C.CATEGORY ID = s.CATEGORY ID
    AND s.SALES TIMEID = t.SALES TIMEID
    AND t.SALES YEAR = '2020'
    GROUP BY t.SALES YEAR, c.CATEGORY DESCRIPTION)
    WHERE CATEGORY RANK <= 10;
D Query Result x | D Query Result 1 x | D Query Result 2 x | D Query Result 3 x | D Query Result 4 x D Query Result 5 x
📌 📇 🙀 📚 SQL | All Rows Fetched: 10 in 0.041 seconds
    $ SALES_YEAR $ CATEGORY_DESCRIPTION $ SALES$
$ CATEGORY_RANK
            Earthmoving 1190000
   1 2020
   2 2020
          Air Compressor 915000
            Landscaping 843400 363600
   з 2020
                                              3
   4 2020
   5 2020
            Generators 338500
            Vehicles 2/0400
265600
   6 2020
   7 2020
   8 2020
             Safety
                              250800
   9 2020
                              225600
             Rail
                                             10
  10 2020
             Plumbing
                             156200
```

## REPORT 2 (Top n% report)

## (a) The query questions

Find out the top 30% company branches by total sales in 2020.

#### (b) Your explanation

As a manager, he or she needs to know the branches with the top sales rankings. The company can reward these branches. At the same time, it can also allow these branches with good sales performance to pass on the experience to other branches with poor performance.

#### (c) The SQL commands

SELECT \*

**FROM** 

(SELECT t.SALES YEAR,c.COMPANY BRANCH, SUM(s.TOTAL SALES REVENUE) as SALES\$,

percent\_rank() OVER (ORDER BY SUM(TOTAL\_SALES\_REVENUE) DESC ) AS COMPANY\_BRANCH\_RANK

FROM salesFACT s,Sales\_TimeDIM t,Company\_BranchDIM c

WHERE c.COMPANY\_BRANCH = s.COMPANY\_BRANCH

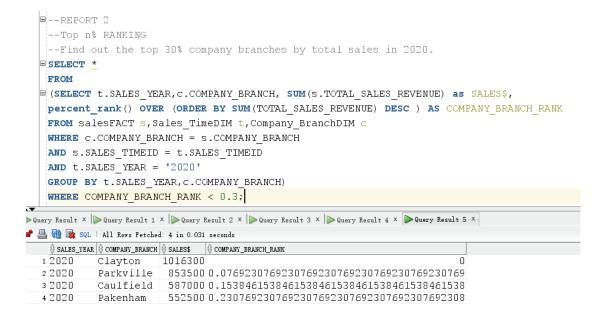
```
AND s.SALES_TIMEID = t.SALES_TIMEID

AND t.SALES_YEAR = '2020'

GROUP BY t.SALES_YEAR,c.COMPANY_BRANCH)

WHERE COMPANY_BRANCH_RANK < 0.3;
```

## (d) The screenshots of the query results



## b. Reports with proper sub-totals:

Produce four reports. These reports must include subtotals, using the Cube or Roll-up or Partial Cube/Roll-up operators

## **REPORT 3 (Cube)**

select cb.COMPANY\_BRANCH,ct.DESCRIPTION as customer\_type,st.SALES\_TIMEID as sales\_time,

SUM(sf.TOTAL\_SALES\_REVENUE) as SALES\$

from Company\_BranchDIM cb,customer\_TypeDIM ct,Sales\_TimeDIM st,salesFACT sf
where sf.COMPANY\_BRANCH = cb.COMPANY\_BRANCH
and ct.CUSTOMER\_TYPE\_ID = sf.CUSTOMER\_TYPE\_ID
and st.SALES\_TIMEID = sf.SALES\_TIMEID
group by cube(cb.COMPANY\_BRANCH,ct.DESCRIPTION,st.SALES\_TIMEID);

```
--b. Reports with proper sub-totals:
   --REPORT 3 (CUBE)
  select cb.COMPANY BRANCH,ct.DESCRIPTION as customer type,st.SALES TIMEID as sales time,
   SUM (sf.TOTAL SALES REVENUE) as SALES$
   from Company BranchDIM cb,customer TypeDIM ct,Sales TimeDIM st,salesFACT sf
   where sf.COMPANY_BRANCH = cb.COMPANY_BRANCH
   and ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID
   and st.SALES_TIMEID = sf.SALES_TIMEID
   group by cube (cb.COMPANY BRANCH, ct.DESCRIPTION, st.SALES TIMEID);
▶Query Result x | DQuery Result 1 x | DQuery Result 2 x | DQuery Result 3 x | Query Result 4 x Query Result 5 x
ి 🖺 🙀 📚 SQL | Fetched 50 rows in 0.032 seconds
    ♦ COMPANY_BRANCH ♦ CUSTOMER_TYPE
♦ SALES_TIME ♦ SALES$
                               (null) 11765050
  1 (null)
               (null)
  2 (null)
               (null)
                               201804
                                           51000
                               201805
                                            83200
  3 (null)
               (null)
  4 (null)
               (null)
                               201806
                                           18000
                                        1270800
                               201807
  5 (null)
              (null)
  6 (null)
               (null)
                               201808
                                           232200
               (null)
                               201809
                                          744700
  7 (null)
  8 (null)
               (null)
                                201810
                                           238700
                                           222000
  9 (null)
               (null)
                               201811
                               201812
  10 (null)
              (null)
                                           325800
```

#### **REPORT 4 (Partial Cube)**

select ct.DESCRIPTION as customer\_type,cb.COMPANY\_BRANCH,st.SALES\_TIMEID as sales time,

SUM(sf.TOTAL SALES REVENUE) as SALES\$

from Company BranchDIM cb,customer TypeDIM ct,Sales TimeDIM st,salesFACT sf

where sf.COMPANY BRANCH = cb.COMPANY BRANCH

and ct.CUSTOMER TYPE ID = sf.CUSTOMER TYPE ID

and st.SALES TIMEID = sf.SALES TIMEID

group by ct.DESCRIPTION,cb.COMPANY\_BRANCH,cube(st.SALES\_TIMEID);

```
-- REPORT 4 (Partial CUBE)
  select ct.DESCRIPTION as customer type, cb.COMPANY BRANCH, st.SALES TIMEID as sales time,
  SUM (sf.TOTAL SALES REVENUE) as SALES$
   from Company_BranchDIM cb,customer_TypeDIM ct,Sales_TimeDIM st,salesFACT sf
  where sf.COMPANY BRANCH = cb.COMPANY BRANCH
   and ct.CUSTOMER TYPE ID = sf.CUSTOMER TYPE ID
   and st.SALES TIMEID = sf.SALES TIMEID
   group by ct.DESCRIPTION,cb.COMPANY BRANCH,cube(st.SALES TIMEID);
PQuery Result x | ▶ Query Result 1 x | ▶ Query Result 2 x | ▶ Query Result 3 x | ▶ Query Result 4 x ▶ Query Result 5 x
🏲 📇 🙌 攻 SQL | Fetched 150 rows in 0.096 seconds
   1 Business Eltham
                                  22400
                       (null)
  2 Business Eltham
                        201809
                                  11200
  3 Business Eltham
                        202001
                                  11200
                                 499200
  4 Business Toorak
                        (null)
  5 Business Toorak
                        201804
                                  18000
  6 Business Toorak
                        201808
                                  32400
                        201809
  7 Business Toorak
                                 124800
  8 Business Toorak
                        201906
                                  72000
                                 252000
  9 Business Toorak
                        202010
 10 Business Clayton
                       (null) 1697650
```

#### REPORT 5 (Roll-up)

#### (a) The query questions

Find out the total sales of the business and individual customer type in each company branch, and from the year 2018 to 2020.

#### (b) Your explanation

This query question is more comprehensive. First of all, it covers the sales data for the last three years regarding the time dimension. Second, it covers all branches. Third, it covers both business and individual customer types. Last, it also contains all the equipment categories. In this way, the management can see the data information of these dimensions more comprehensively, helping them make better business decisions.

## (c) The SQL commands

```
select st.SALES_YEAR as sales_time,cb.COMPANY_BRANCH,ct.DESCRIPTION as customer_type,ca.CATEGORY_DESCRIPTION as category,

SUM(sf.TOTAL_SALES_REVENUE) as SALES$

from Company_BranchDIM cb,customer_TypeDIM ct,Sales_TimeDIM st,categoryDIM ca,salesFACT sf

where sf.COMPANY_BRANCH = cb.COMPANY_BRANCH

and ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID

and st.SALES_TIMEID = sf.SALES_TIMEID

and ca.CATEGORY_ID = sf.CATEGORY_ID

and ct.DESCRIPTION IN('Individual', 'Business')

and st.SALES_YEAR in ('2018','2019','2020')

group by

rollup(st.SALES_YEAR,cb.COMPANY_BRANCH,ct.DESCRIPTION,ca.CATEGORY_DESCRIPTION);
```

#### (d) The screenshots of the query results

```
--REPORT 5 (Roll-up)
   --Find out the total sales of the business and individual customer type in each company branch, and from the year 2018 to 2020.
select st.SALES_YEAR as sales_time.cb.COMPANY_BRANCH.ct.DESCRIPTION as customer_type.ca.CATEGORY_DESCRIPTION as category,
 SUM(sf.TOTAL_SALES_REVENUE) as SALES$
  from Company BranchDIM cb, customer TypeDIM ct, Sales TimeDIM st, categoryDIM ca, salesFACT sf
  where sf.COMPANY_BRANCH = cb.COMPANY_BRANCH
  and ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID
  and st.SALES_TIMEID = sf.SALES_TIMEID
and ca.CATEGORY_ID = sf.CATEGORY_ID
  and ct.DESCRIPTION IN('Individual','Business')
  and st.SALES_YEAR in ('2018','2019','2020')
  group by rollup (st.SALES YEAR, cb.COMPANY BRANCH, ct.DESCRIPTION, ca.CATEGORY DESCRIPTION);
uery Result × | ▶ Query Result 1 × | ▶ Query Result 2 × | ▶ Query Result 3 × | ▶ Query Result 4 × ▶ Query Result 5 ×
A SQL | Fetched 50 rows in 0.031 seconds
 $ sales_time $ company_branch $ customer_type $ category 1 2018 Eltham Business Plumbin
                         Business Plumbing
Business (null)
Individual Plumbing
                                                                 11200
            Eltham
Eltham
                                                                11200
26400
 4 2018
            Eltham
                         Individual (null)
                                                                26400
5 2018
6 2018
7 2018
            Eltham
Toorak
                         (null)
Business
                                                                37600
                                       Safety
            Toorak
                         Business
            Toorak
Toorak
Toorak
                                      Compaction
Landscaping
(null)
 8 2018
                         Business
                                                                18000
                         Business
Business
                                                               124800
175200
10 2018
```

## REPORT 6(Partial Roll-up)

#### (a) The query questions

Find out the total sales of the individual and business category in each company branch of 2020.

## (b) Your explanation

This query will narrow the scope and focus the data analysis on the most recent year 2020. At the same time, only the customer type and all branches are selected, so that the management can understand each branch's performance in the past year.

#### (c) The SQL commands

```
select st.SALES_YEAR as sales_time,cb.COMPANY_BRANCH,ct.DESCRIPTION as customer_type,

SUM(sf.TOTAL_SALES_REVENUE) as SALES$

from Company_BranchDIM cb,customer_TypeDIM ct,Sales_TimeDIM st,salesFACT sf where sf.COMPANY_BRANCH = cb.COMPANY_BRANCH

and ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID

and st.SALES_TIMEID = sf.SALES_TIMEID

and ct.DESCRIPTION IN('Individual','Business')

and st.SALES_YEAR in ('2020')

group by st.SALES_YEAR,rollup(cb.COMPANY_BRANCH,ct.DESCRIPTION);
```

#### (d) The screenshots of the query results

```
--REPORT 6(Partial Roll-up)
  --Find out the total sales of the individual and business category in each company branch of 2020.
 select st.SALES_YEAR as sales_time,cb.COMPANY_BRANCH,ct.DESCRIPTION as customer_type,
  SUM (sf.TOTAL_SALES REVENUE) as SALES$
  from Company BranchDIM cb, customer TypeDIM ct, Sales TimeDIM st, salesFACT sf
  where sf.COMPANY BRANCH = cb.COMPANY BRANCH
  and ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID
  and st.SALES_TIMEID = sf.SALES_TIMEID
  and ct.DESCRIPTION IN('Individual','Business')
  and st.SALES_YEAR in ('2020')
  group by st.SALES YEAR, rollup (cb.COMPANY BRANCH, ct.DESCRIPTION);
Query Result X | ▶ Query Result 1 X | ▶ Query Result 2 X | ▶ Query Result 3 X | ▶ Query Result 4 X ▶ Query Result 5 X
📇 🚵 🔯 SQL | All Rows Fetched: 43 in 0.032 seconds
 1 2020
          Eltham
                     Business
                                   11200
 2 2020
                     Individual
                                  52000
          Eltham
 з 2020
          Eltham
                     (null)
                                   63200
                                 252000
 4 2020
          Toorak
                     Business
                  Individual
 5 2020
                                   60000
          Toorak
                                 312000
 6 2020
          Toorak
                     (null)
          Clayton
 7 2020
                                  441800
                     Business
          Clayton
                     Individual
 8 2020
                                 574500
                               1016300
103100
 9 2020
          Clayton
                     (null)
                     Business
10 2020
          Fitzroy
```

## c. Reports with moving and cumulative aggregates:

#### **REPORT 7:**

What are the total revenue for hiring equipment and cumulative total hiring revenue for Site Equipment in each year?

## Total revenue for hiring equipment

```
SELECT t.HIRE_YEAR,

TO_CHAR (SUM(hf.TOTAL_HIRING_REVENUE), '9,999,999,999') AS YEAR_HIRING_REVENUE,

TO_CHAR (SUM(SUM(hf.TOTAL_HIRING_REVENUE)) OVER

(ORDER BY t.HIRE_YEAR

ROWS UNBOUNDED PRECEDING),

'9,999,999,999') AS CUM__HIRING_REVENUES

FROM hireFACT hf,Hire_TimeDIM t

WHERE t.HIRE_TIMEID = hf.HIRE_TIMEID

GROUP BY t.HIRE_YEAR;
```

```
--c. Reports with moving and cumulative aggregates:
   --REPORT 7
  SELECT t.HIRE YEAR,
   TO CHAR (SUM(hf.TOTAL HIRING REVENUE), '9,999,999,999') AS YEAR HIRING REVENUE,
   TO CHAR (SUM (SUM (hf. TOTAL HIRING REVENUE)) OVER
   (ORDER BY t.HIRE YEAR
   ROWS UNBOUNDED PRECEDING),
   '9,999,999,999') AS CUM HIRING REVENUES
   FROM hireFACT hf, Hire TimeDIM t
   WHERE t.HIRE TIMEID = hf.HIRE TIMEID
   GROUP BY t.HIRE YEAR;
▶Query Result x | ▶Query Result 1 x | ▶Query Result 2 x | ▶Query Result 3 x | ▶Query Result 4 x ▶Query Result 5 x
🖈 📇 🙌 🗽 SQL | All Rows Fetched: 3 in 0.031 seconds
   45,610
  1 2018
                                  45,610
  2 2019
                   48,095
                                  93,705
                                 140,980
  з 2020
                   47,275
```

## Cumulative total hiring revenue for Site Equipment in each year

SELECT t.HIRE\_YEAR, C. CATEGORY\_DESCRIPTION,

TO\_CHAR (SUM(hf.TOTAL\_HIRING\_REVENUE), '9,999,999') AS YEAR\_HIRING\_REVENUE,

TO\_CHAR (SUM(SUM(hf.TOTAL\_HIRING\_REVENUE)) OVER

(ORDER BY t.HIRE\_YEAR

ROWS UNBOUNDED PRECEDING),

'9,999,999,999') AS CUM\_\_HIRING\_REVENUES

FROM hireFACT hf, Hire\_TimeDIM t, categoryDIM C

WHERE t.HIRE\_TIMEID = hf.HIRE\_TIMEID

AND c.CATEGORY\_ID = hf.CATEGORY\_ID

AND CATEGORY\_DESCRIPTION = 'Site Equipment'

GROUP BY t.HIRE\_YEAR,C.CATEGORY\_DESCRIPTION;

```
SELECT t.HIRE YEAR, C.CATEGORY DESCRIPTION,
   TO CHAR (SUM(hf.TOTAL HIRING REVENUE), '9,999,999,999') AS YEAR HIRING REVENUE,
   TO CHAR (SUM (SUM (hf. TOTAL HIRING REVENUE)) OVER
   (ORDER BY t.HIRE YEAR
   ROWS UNBOUNDED PRECEDING),
   '9,999,999,999') AS CUM HIRING REVENUES
   FROM hireFACT hf, Hire TimeDIM t, categoryDIM C
   WHERE t.HIRE TIMEID = hf.HIRE TIMEID
   AND c.CATEGORY ID = hf.CATEGORY ID
   AND CATEGORY DESCRIPTION = 'Site Equipment'
   GROUP BY t.HIRE YEAR, C.CATEGORY DESCRIPTION;
▶ Query Result x | D Query Result 1 x | D Query Result 2 x | D Query Result 3 x | D Query Result 4 x D Query Result 5 x
🏲 📇 🝓 퀋 SQL | All Rows Fetched: 3 in 0.031 seconds
  ♦ HIRE_YEAR ♦ CATEGORY_DESCRIPTION ♦ YEAR_HIRING_REVENUE
  1 2018 Site Equipment 3,665
                                                    3,665
  2 2019
           Site Equipment
                                    2,230
                                                     5,895
         Site Equipment
  3 2020
                                    4,090
                                                     9,985
```

#### **REPORT 8:**

Produce one other moving/cumulative aggregate report that is useful for management.

The outputs of this task are:

## (a) The query questions

For individual and business customer types, find out the three-month moving average of total sales revenue.

#### (b) Your explanation

Managers will be interested in the sales revenue of different types of customers. So I choose to analyse the three-month moving average of the two customer types. Therefore, Managers can easily see the growth of sales of different kind of customers.

## (c) The SQL commands

```
SELECT ct.DESCRIPTION as customer_type,s.SALES_TIMEID,

TO_CHAR (SUM(sf.TOTAL_SALES_REVENUE), '9,999,999') AS Month_SALES_REVENUE,

TO_CHAR (AVG(SUM(sf.TOTAL_SALES_REVENUE)) OVER

(ORDER BY s.SALES_TIMEID

ROWS 2 PRECEDING),

'9,999,999,999') AS MOVING_3_MONTH_AVG

FROM salesFACT sf,Sales_TimeDIM s,customer_TypeDIM ct
```

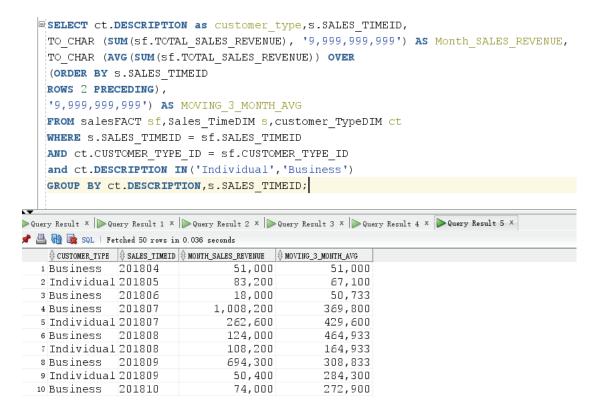
```
WHERE s.SALES_TIMEID = sf.SALES_TIMEID

AND ct.CUSTOMER_TYPE_ID = sf.CUSTOMER_TYPE_ID

and ct.DESCRIPTION IN('Individual','Business')

GROUP BY ct.DESCRIPTION,s.SALES_TIMEID;
```

## (d) The screenshots of the query results



## d. Reports with Partitions:

#### **REPORT 9:**

Show ranking of each equipment category based on the monthly total number of equipment hired and the ranking of each branch based on the monthly total number of equipment hired.

#### (a) The SQL commands that contain partitions

SELECT cb.COMPANY\_BRANCH, ht.HIRE\_TIMEID AS month,

TO\_CHAR(SUM(hf.NUM\_EQUIPMENT\_HIRED)) AS number\_of\_equipment\_hired,

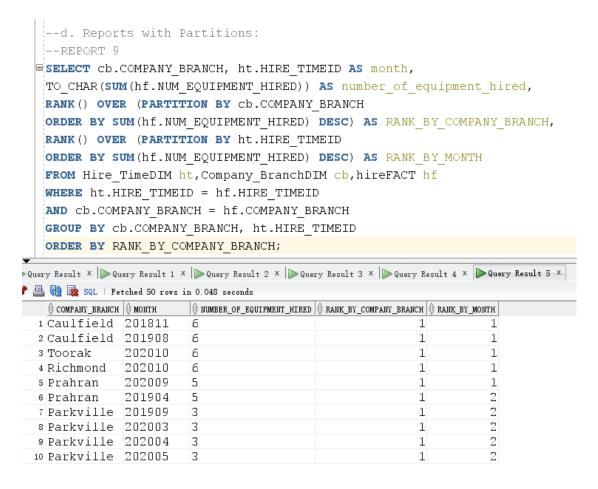
RANK() OVER (PARTITION BY cb.COMPANY\_BRANCH

ORDER BY SUM(hf.NUM\_EQUIPMENT\_HIRED) DESC) AS RANK\_BY\_COMPANY\_BRANCH,

RANK() OVER (PARTITION BY ht.HIRE\_TIMEID

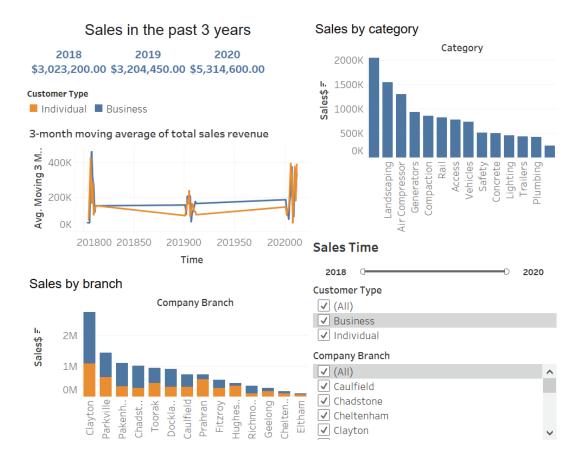
```
ORDER BY SUM(hf.NUM_EQUIPMENT_HIRED) DESC) AS RANK_BY_MONTH
FROM Hire_TimeDIM ht,Company_BranchDIM cb,hireFACT hf
WHERE ht.HIRE_TIMEID = hf.HIRE_TIMEID
AND cb.COMPANY_BRANCH = hf.COMPANY_BRANCH
GROUP BY cb.COMPANY_BRANCH, ht.HIRE_TIMEID
ORDER BY RANK_BY_COMPANY_BRANCH;
```

## (b) The screenshots of the query results



# 4. Business Intelligence (BI) Reports.

About the BI report, I choose to use the report5, 6 and 8 to do the presentation., and use the Tableau put above reports into the dashboard. Furthermore, this presentation can show the sales fact to the management. Managers can directly see the 3 moths moving average sales data of the past three years, the sales revenue of various branches, and the revenue of different equipment category on the dashboard (see the figure below).



This dashboard provides users with three selection buttons: time, company branch, and customer type. Users can easily choose the data they want to know and use these data for comparison (see the figure below).

