Varun Nair

Overview of the Assignment:

In this assignment we will create a few analytical queries on a data warehouse. We will also explore some aspects of Tableau.

Part 1 – Restore database

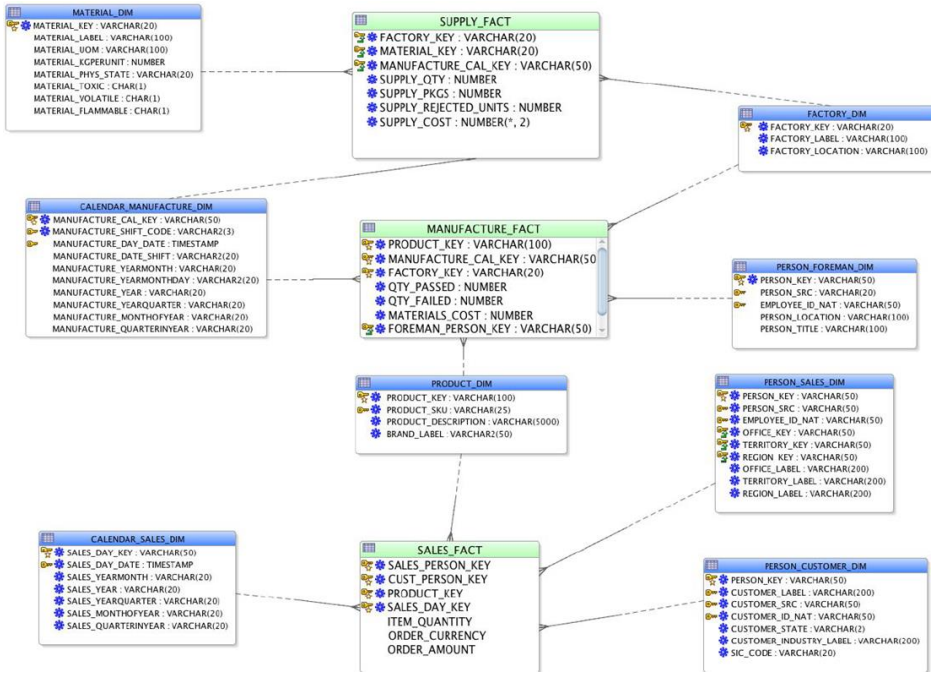
Depending on whether you've chosen SQL Server or PostgreSQL: download the appropriate file and restore the database to your machine.
Write a query showing count of records in the manufacture fact database, as well as today's date and your name (two additional columns), take a screenshot of the result.
Paste the code and a screen shot of the results.
SQL Code: select count(*) as count, getdate() as date, 'Varun Nair' as name from MANUFACTURE_FACT
Screenshot of result:

Results | Messages

	count	date	name
1	304200	2024-11-10 18:14:14.460	Varun Nair

Part 2 – Examine the Schema and Data to familiarize yourself.

Query the dimension tables to see what they look like. Refer to the schema diagram below. Then answer the questions below.



 Copyright 2023 Boston University; All rights reserved

Varun Nair

A. Pick two dimensions and list the hierarchy columns in each of these dimensions which have them. Provide the dimension name and the columns in order from highest to lowest (drill-down).

Dimension 1: CALENDAR_MANUFACTURE_DIM :
MANUFACTURE_DAY_DATE,MANUFACTURE_QUATERINYEAR,MANUFACTURE_YEARQUARTER,MANUFACTURE_MONTHOFYEAR,MANUFACTURE_YEAR,MANUFACTURE_YEARMONTH,MANUFACTURE_YEARMONTHDAY
Dimension 2: PERSON_SALES_DIM :
REGION_LABEL,REGION_KEY,TERRITORY_LABEL,TERRITORY_KEY,OFFICE_LABEL,OFFICE_KEY

B. **Would any of the dimensions benefit from SCD2 or SCD3 implementation? Which ones and what would you recommend as a change. Provide two SCD changes you would suggest implementing. Describe the columns you would add, and how the ETL process would change to maintain these. Hint: It's not just about the above design, review the data as well and explain your reasoning briefly. Dimension 1 to change to SCD and how would you change it:**

PERSON_SALES_DIM can have additional columns like START_DATE and END_DATE along with an IsActive flag highlighting the current assignment. Along with this, there can be additional columns detailing each salespersons previous locations (PREV_OFFICE_LABEL,PREV_OFFICE_KEY)-something to keep in mind that PREV_OFFICE_KEY & PREV_OFFICE_LABEL will be null for new salespeople and each salesperson's first assignment AKA when they joined the company, thus making it a SCD 2 and SCD 3 Dimension
For maintenance, if a salesperson changes his location, then the current OFFICE_LABEL and OFFICE_KEY will be the PREV_OFFICE_LABEL and PREV_OFFICE_KEY in the salespersons new record. To make this record further accurate, we can update the start date as the date of when the salesperson join the new location. The end date for their previous assignment will be whenever they have left that location.
Dimension 2 to change to SCD and how would you change it:
MATERIAL_DIM can have a START_DATE and END_DATE column talking about whether the material is being provided or not. An IsActive flag can highlight all the materials that are currently being provided. In the future, if certain materials are unavailable, having a record of those materials would be helpful. This would make it a SCD Type 2

C. Pick one fact table from the design what are the measures?

Fact table: SALES_FACT
Measures : ITEM_QUANTITY,ORDER_AMOUNT,ORDER_CURRENCY

 Copyright 2023 Boston University; All rights reserved

Varun Nair

Part 3 – Dimensional Queries

1. **Write and execute a query that identifies for each year, three factories which produced (passed) the most units. Your output should have these five columns and each year should show the top three factories:**
   - **Year**
   - **Factory name (label)**
   - **Total Units produced (passed) for each factory for each year**
   - **Total Units failed for each factory for each year**
   - **Factory name rank (based on total units produced)**

   Sort your result by the latest year first, with oldest year last.
   - **Hint: Solve this query in multiple steps, then put it all together. For example, join the data first, compute your aggregates, finally filter the aggregates checking your results in each step. Using an inline view or CTE might be helpful to solve this question by breaking it into two steps.**

   Paste the SQL code and a screen shot of the results.
   SQL Code:

```
select
YearlyProduction.Year,
YearlyProduction.FactoryLabel,
YearlyProduction.QtyPassed  as TotalUnitsPassed,
YearlyProduction.QtyFailed as TotalUnitsFailed,
YearlyProduction.FactoryRank from(
      select
      c.MANUFACTURE_YEAR  as Year,
      f.FACTORY_LABEL  AS FactoryLabel,
      sum(m.QTY_PASSED) as QtyPassed,
      sum(m.QTY_FAILED)  as QtyFailed,
      ROW_NUMBER() over (partition by c.MANUFACTURE_YEAR  order by
sum(m.QTY_PASSED) desc) as FactoryRank
      from
      MANUFACTURE_FACT m
      JOIN
      CALENDAR_MANUFACTURE_DIM  c  ON m.MANUFACTURE_CAL_KEY = c.MANUFACTURE_CAL_KEY
      join
      FACTORY_DIM f ON  m.FACTORY_KEY = f.FACTORY_KEY
      GROUP BY
      C.MANUFACTURE_YEAR,f.FACTORY_LABEL
      ) AS YearlyProduction where YearlyProduction.FactoryRank <= 3 order by
YearlyProduction.Year DESC,YearlyProduction.FactoryRank
```

 Copyright 2023 Boston University; All rights reserved

Screenshot of result:

Results Messages

	Year	FactoryLabel	TotalUnitsPassed	TotalUnitsFailed	FactoryRank
1	mY2022	Misrata Heating Equipment	16885	2927	1
2	mY2022	Jadu Wholesale	16427	2966	2
3	mY2022	Kufra Bituminous Coal	16156	2982	3
4	mY2021	Misrata Heating Equipment	16935	2890	1
5	mY2021	Jadu Wholesale	16550	2821	2
6	mY2021	Kufra Bituminous Coal	16457	2774	3
7	mY2020	Misrata Heating Equipment	16567	3202	1
8	mY2020	Jadu Wholesale	16432	2758	2
9	mY2020	Kufra Bituminous Coal	16391	2799	3
10	mY2019	Misrata Heating Equipment	16983	2824	1
11	mY2019	Jadu Wholesale	16505	2758	2
12	mY2019	Kufra Bituminous Coal	16315	2804	3
13	mY2018	Misrata Heating Equipment	17010	2856	1
14	mY2018	Jadu Wholesale	16605	2865	2
15	mY2018	Kufra Bituminous Coal	16372	2758	3
16	mY2017	Misrata Heating Equipment	16888	2896	1
17	mY2017	Jadu Wholesale	16639	2714	2
18	mY2017	Kufra Bituminous Coal	16438	2662	3
19	mY2016	Misrata Heating Equipment	17012	2944	1
20	mY2016	Jadu Wholesale	16577	2955	2
21	mY2016	Kufra Bituminous Coal	16363	2902	3
22	mY2015	Misrata Heating Equipment	16872	3003	1

Query executed successfully.

2. Drill down and rollup!

Question 2A Write and execute a query that that identifies total units produced (passed) for each of the factories for each month in 2022 with a subtotal for each factory.

Your result set should have the following four fields. You will want to filter the data for 2022.

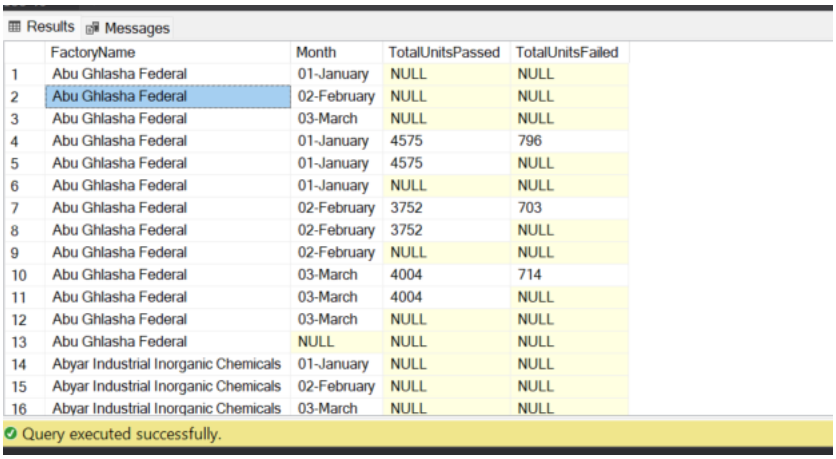
- **Factory name**
- **Month – Notice that the month is not quite in readable format. Transform the month into the following format '01-January', '02-February', etc. Hint – a case statement will be helpful here.**
- **Total Units produced (passed) for each factory for each month**
- **Total Units failed for each factory for each month**

Complete the result by adding a Rollup to show subtotals by factory. With the Rollup function, you will notice a NULL value for the month column for the subtotal on each of the factories.
Paste the SQL code and a screen shot of the results.
SQL Code:

```
WITH MonthlyData AS (
    SELECT
        f.FACTORY_LABEL AS FactoryName,
        CONCAT(
            FORMAT(MONTH(c.MANUFACTURE_DAY_DATE), '00'), '-',
            CASE MONTH(c.MANUFACTURE_DAY_DATE)
                WHEN 1 THEN 'January'
                WHEN 2 THEN 'February'
                WHEN 3 THEN 'March'
                WHEN 4 THEN 'April'
                WHEN 5 THEN 'May'
                WHEN 6 THEN 'June'
```

```
                WHEN 7 THEN 'July'
                WHEN 8 THEN 'August'
                WHEN 9 THEN 'September'
                WHEN 10 THEN 'October'
                WHEN 11 THEN 'November'
                WHEN 12 THEN 'December'
            END
        ) AS Month,
        MONTH(c.MANUFACTURE_DAY_DATE) AS MonthSort,
        SUM(m.QTY_PASSED) AS TotalUnitsPassed,
        SUM(m.QTY_FAILED) AS TotalUnitsFailed
    FROM
        MANUFACTURE_FACT m
    JOIN
        CALENDAR_MANUFACTURE_DIM c ON m.MANUFACTURE_CAL_KEY = c.MANUFACTURE_CAL_KEY
    JOIN
        FACTORY_DIM f ON m.FACTORY_KEY = f.FACTORY_KEY
    WHERE
        YEAR(c.MANUFACTURE_DAY_DATE) = 2022
    GROUP BY
        f.FACTORY_LABEL,
        MONTH(c.MANUFACTURE_DAY_DATE)
)
SELECT
    FactoryName,
    Month,
    TotalUnitsPassed,
    TotalUnitsFailed
FROM MonthlyData
GROUP BY
    FactoryName,
    Month,
    MonthSort,
    TotalUnitsPassed,
    TotalUnitsFailed
WITH ROLLUP
HAVING
    (Month IS NOT NULL OR FactoryName IS NOT NULL)
ORDER BY
    COALESCE(FactoryName, 'ZZZZ'),
    CASE
        WHEN Month IS NULL THEN 'ZZZZ'
        ELSE CAST(MonthSort AS VARCHAR(2))
    END;
```

Screenshot of result:



Results | Messages

	FactoryName	Month	TotalUnitsPassed	TotalUnitsFailed
1	Abu Ghlasha Federal	01-January	NULL	NULL
2	Abu Ghlasha Federal	02-February	NULL	NULL
3	Abu Ghlasha Federal	03-March	NULL	NULL
4	Abu Ghlasha Federal	01-January	4575	796
5	Abu Ghlasha Federal	01-January	4575	NULL
6	Abu Ghlasha Federal	01-January	NULL	NULL
7	Abu Ghlasha Federal	02-February	3752	703
8	Abu Ghlasha Federal	02-February	3752	NULL
9	Abu Ghlasha Federal	02-February	NULL	NULL
10	Abu Ghlasha Federal	03-March	4004	714
11	Abu Ghlasha Federal	03-March	4004	NULL
12	Abu Ghlasha Federal	03-March	NULL	NULL
13	Abu Ghlasha Federal	NULL	NULL	NULL
14	Abyar Industrial Inorganic Chemicals	01-January	NULL	NULL
15	Abyar Industrial Inorganic Chemicals	02-February	NULL	NULL
16	Abyar Industrial Inorganic Chemicals	03-March	NULL	NULL

Query executed successfully.

Question 2B: Outline one suggestion you would implement as part of the design and ETL to make this question easier to solve?
Short Answer:

There can be a separate column for month which only has integer values of the month in the Calendar dimension. This will skip the need of any case statements where we needed to check the for the month number and then assign it to a number.
The month can be extracted from the date via the month() function and can be inserted into the new month_number column

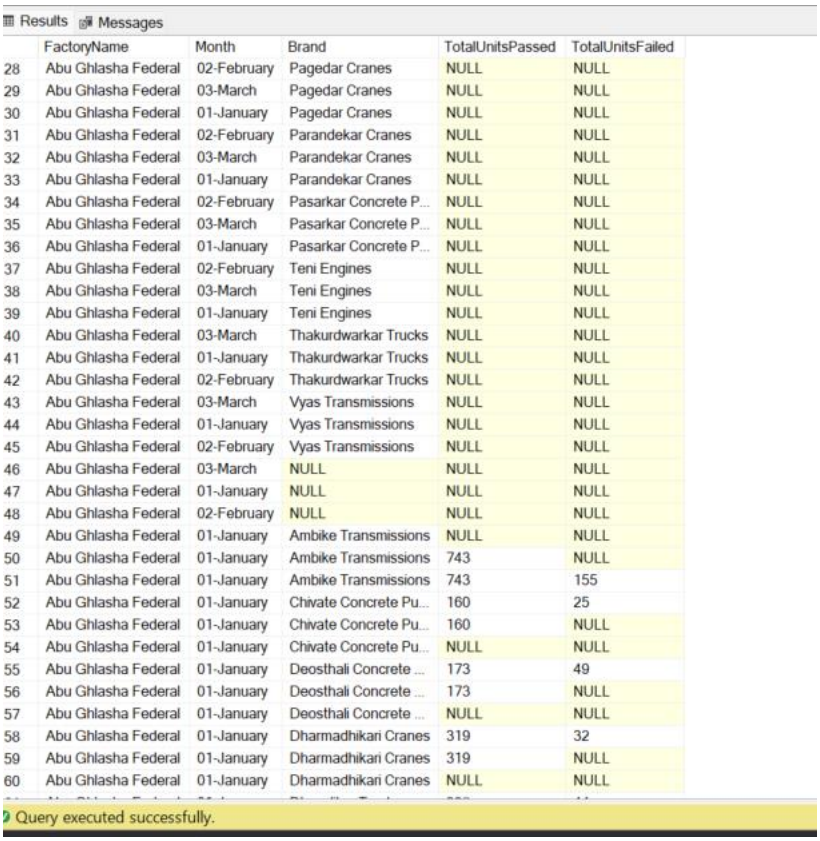
3. **Drill down and rollup! Modify the above query in question 2 (the original query) to now drill down to the brand each of the factories and months within the year you selected. Use ROLLUP to show subtotals by factory, month, and brand. Your output should have these columns:**
   - **Factory name**
   - **Month**
   - **Brand (brand label)**
   - **Total Units produced (passed) for each factory for each month, for each brand**
   - **Total Units failed for each factory for each month, for each brand**

Paste the SQL code and a screen shot of the results.
SQL Code:

```
WITH MonthlyData AS (
    SELECT
        f.FACTORY_LABEL AS FactoryName,
        CONCAT(
            FORMAT(MONTH(c.MANUFACTURE_DAY_DATE), '00'), '-',
            CASE MONTH(c.MANUFACTURE_DAY_DATE)
                WHEN 1 THEN 'January'
                WHEN 2 THEN 'February'
                WHEN 3 THEN 'March'
                WHEN 4 THEN 'April'
                WHEN 5 THEN 'May'
                WHEN 6 THEN 'June'
                WHEN 7 THEN 'July'
                WHEN 8 THEN 'August'
```

```
                WHEN 9 THEN 'September'
                WHEN 10 THEN 'October'
                WHEN 11 THEN 'November'
                WHEN 12 THEN 'December'
            END
        ) AS Month,
            MONTH(c.MANUFACTURE_DAY_DATE) as MonthSort,
            p.BRAND_LABEL  as Brand,
            sum(m.QTY_PASSED) as TotalUnitsPassed,
            sum(QTY_FAILED) as TotalUnitsFailed
            from  MANUFACTURE_FACT m
            JOIN CALENDAR_MANUFACTURE_DIM c ON m.MANUFACTURE_CAL_KEY =
c.MANUFACTURE_CAL_KEY
            JOIN FACTORY_DIM f ON m.FACTORY_KEY = f.FACTORY_KEY
            JOIN  PRODUCT_DIM  p ON  M.PRODUCT_KEY = P.PRODUCT_KEY
            WHERE YEAR(c.MANUFACTURE_DAY_DATE)=2022
            group by f.FACTORY_LABEL,MONTH(c.MANUFACTURE_DAY_DATE),p.BRAND_LABEL
            )
            select FactoryName,Month,Brand,TotalUnitsPassed,TotalUnitsFailed from
MonthlyData
            group by
FactoryName,Month,Brand,MonthSort,TotalUnitsPassed,TotalUnitsFailed
            with rollup  having (Month is not null or FactoryName is  not  null
or  Brand is not null)
            order by coalesce(FactoryName,'ZZZZ'),
            CASE
                  WHEN Month is null then  'ZZZZ'
                  ELSE CAST(MonthSort  as VARCHAR(2))
                  END,
                  COALESCE(Brand,'ZZZZ')
```

Screenshot of result:



Results | Messages

	FactoryName	Month	Brand	TotalUnitsPassed	TotalUnitsFailed
28	Abu Ghlasha Federal	02-February	Pagedar Cranes	NULL	NULL
29	Abu Ghlasha Federal	03-March	Pagedar Cranes	NULL	NULL
30	Abu Ghlasha Federal	01-January	Pagedar Cranes	NULL	NULL
31	Abu Ghlasha Federal	02-February	Parandekar Cranes	NULL	NULL
32	Abu Ghlasha Federal	03-March	Parandekar Cranes	NULL	NULL
33	Abu Ghlasha Federal	01-January	Parandekar Cranes	NULL	NULL
34	Abu Ghlasha Federal	02-February	Pasarkar Concrete P...	NULL	NULL
35	Abu Ghlasha Federal	03-March	Pasarkar Concrete P...	NULL	NULL
36	Abu Ghlasha Federal	01-January	Pasarkar Concrete P...	NULL	NULL
37	Abu Ghlasha Federal	02-February	Teni Engines	NULL	NULL
38	Abu Ghlasha Federal	03-March	Teni Engines	NULL	NULL
39	Abu Ghlasha Federal	01-January	Teni Engines	NULL	NULL
40	Abu Ghlasha Federal	03-March	Thakurdwarkar Trucks	NULL	NULL
41	Abu Ghlasha Federal	01-January	Thakurdwarkar Trucks	NULL	NULL
42	Abu Ghlasha Federal	02-February	Thakurdwarkar Trucks	NULL	NULL
43	Abu Ghlasha Federal	03-March	Vyas Transmissions	NULL	NULL
44	Abu Ghlasha Federal	01-January	Vyas Transmissions	NULL	NULL
45	Abu Ghlasha Federal	02-February	Vyas Transmissions	NULL	NULL
46	Abu Ghlasha Federal	03-March	NULL	NULL	NULL
47	Abu Ghlasha Federal	01-January	NULL	NULL	NULL
48	Abu Ghlasha Federal	02-February	NULL	NULL	NULL
49	Abu Ghlasha Federal	01-January	Ambike Transmissions	NULL	NULL
50	Abu Ghlasha Federal	01-January	Ambike Transmissions	743	NULL
51	Abu Ghlasha Federal	01-January	Ambike Transmissions	743	155
52	Abu Ghlasha Federal	01-January	Chivate Concrete Pu...	160	25
53	Abu Ghlasha Federal	01-January	Chivate Concrete Pu...	160	NULL
54	Abu Ghlasha Federal	01-January	Chivate Concrete Pu...	NULL	NULL
55	Abu Ghlasha Federal	01-January	Deosthali Concrete ...	173	49
56	Abu Ghlasha Federal	01-January	Deosthali Concrete ...	173	NULL
57	Abu Ghlasha Federal	01-January	Deosthali Concrete ...	NULL	NULL
58	Abu Ghlasha Federal	01-January	Dharmadhikari Cranes	319	32
59	Abu Ghlasha Federal	01-January	Dharmadhikari Cranes	319	NULL
60	Abu Ghlasha Federal	01-January	Dharmadhikari Cranes	NULL	NULL

Query executed successfully.

4. Drill down and rollup! Modify the above query in question 3 to use CUBE instead Paste the SQL code and a screen shot of the results.

SQL Code:

```
WITH MonthlyData AS (
    SELECT
        f.FACTORY_LABEL AS FactoryName,
        CONCAT(
            FORMAT(MONTH(c.MANUFACTURE_DAY_DATE), '00'), '-',
            CASE MONTH(c.MANUFACTURE_DAY_DATE)
                WHEN 1 THEN 'January'
                WHEN 2 THEN 'February'
```

```
                WHEN 3 THEN 'March'
                WHEN 4 THEN 'April'
                WHEN 5 THEN 'May'
                WHEN 6 THEN 'June'
                WHEN 7 THEN 'July'
                WHEN 8 THEN 'August'
                WHEN 9 THEN 'September'
                WHEN 10 THEN 'October'
                WHEN 11 THEN 'November'
                WHEN 12 THEN 'December'
            END
        ) AS Month,
            MONTH(c.MANUFACTURE_DAY_DATE) as MonthSort,
            p.BRAND_LABEL  as Brand,
            sum(m.QTY_PASSED) as TotalUnitsPassed,
            sum(QTY_FAILED) as TotalUnitsFailed
            from  MANUFACTURE_FACT m
            JOIN CALENDAR_MANUFACTURE_DIM c ON m.MANUFACTURE_CAL_KEY =
c.MANUFACTURE_CAL_KEY
            JOIN FACTORY_DIM f ON m.FACTORY_KEY = f.FACTORY_KEY
            JOIN  PRODUCT_DIM  p ON  M.PRODUCT_KEY = P.PRODUCT_KEY
            WHERE YEAR(c.MANUFACTURE_DAY_DATE)=2022
            group by f.FACTORY_LABEL,MONTH(c.MANUFACTURE_DAY_DATE),p.BRAND_LABEL
            )
            select FactoryName,Month,Brand,TotalUnitsPassed,TotalUnitsFailed from
MonthlyData
            group by
FactoryName,Month,Brand,MonthSort,TotalUnitsPassed,TotalUnitsFailed
            with CUBE  having (Month is not null or FactoryName is  not  null   or
Brand is not null)
            order by coalesce(FactoryName,'ZZZZ'),
            CASE
                  WHEN Month is null then  'ZZZZ'
                  ELSE CAST(MonthSort  as VARCHAR(2))
                  END,COALESCE(Brand,'ZZZZ')
```

Screenshot of result:

Results | Messages

	FactoryName	Month	Brand	TotalUnitsPassed	TotalUnitsFailed
1	Abu Ghlasha Federal	01-January	Ambike Transmissions	NULL	NULL
2	Abu Ghlasha Federal	02-February	Ambike Transmissions	NULL	NULL
3	Abu Ghlasha Federal	02-February	Ambike Transmissions	NULL	20
4	Abu Ghlasha Federal	03-March	Ambike Transmissions	NULL	92
5	Abu Ghlasha Federal	02-February	Ambike Transmissions	147	20
6	Abu Ghlasha Federal	03-March	Ambike Transmissions	463	92
7	Abu Ghlasha Federal	01-January	Ambike Transmissions	743	155
8	Abu Ghlasha Federal	03-March	Ambike Transmissions	463	NULL
9	Abu Ghlasha Federal	01-January	Ambike Transmissions	NULL	155
10	Abu Ghlasha Federal	03-March	Ambike Transmissions	NULL	NULL
11	Abu Ghlasha Federal	02-February	Ambike Transmissions	147	NULL
12	Abu Ghlasha Federal	01-January	Ambike Transmissions	743	NULL
13	Abu Ghlasha Federal	01-January	Chivate Concrete Pumps	NULL	NULL
14	Abu Ghlasha Federal	02-February	Chivate Concrete Pumps	NULL	NULL
15	Abu Ghlasha Federal	03-March	Chivate Concrete Pumps	NULL	11
16	Abu Ghlasha Federal	01-January	Chivate Concrete Pumps	NULL	25
17	Abu Ghlasha Federal	02-February	Chivate Concrete Pumps	NULL	43

Query executed successfully.

5. **Briefly explain the difference you noticed in results between rollup and cube of your results.**

   Your Response:
   The cube had much more detailed rows in the result, with a lot more nulls than the rollup. The rollup had individual subtotals for both produced units and failed units for each month that it was present in(from Manufacture_fact)
   The cube query also listed out all brand labels for each factory first (with NULL values for both units in those rows) and then lists the subtotals for each kind of units, for each month.

6. **Reuse the code from your query in question 1 to create the following data set which we will turn into a PIVOT/Crosstab in question 7. The base query will have the following three columns:**
   - **Year**
   - **Factory name (label)**
   - **Quantity passed**

   Filter this query to the month of February for the most current five years in the data set (the result will only contain data for February for five latest years)
   Paste the SQL code and a screen shot of the results.
   SQL Code:

```
select
YearlyProduction.Year,
YearlyProduction.FactoryLabel,
YearlyProduction.QtyPassed  as TotalUnitsPassed  FROM(
      select
      c.MANUFACTURE_YEAR  as Year,
      f.FACTORY_LABEL  AS FactoryLabel,
      sum(m.QTY_PASSED) as QtyPassed,
      ROW_NUMBER() over (partition by c.MANUFACTURE_YEAR  order by
sum(m.QTY_PASSED) desc) as FactoryRank
      from
      MANUFACTURE_FACT m
      JOIN
      CALENDAR_MANUFACTURE_DIM  c  ON m.MANUFACTURE_CAL_KEY = c.MANUFACTURE_CAL_KEY
      join
      FACTORY_DIM f ON  m.FACTORY_KEY = f.FACTORY_KEY
      where month(c.MANUFACTURE_DAY_DATE) = 2 AND  YEAR(c.MANUFACTURE_DAY_DATE)
in(2022,2021,2020,2019,2018)
      GROUP BY
      C.MANUFACTURE_YEAR,f.FACTORY_LABEL
      ) AS YearlyProduction where YearlyProduction.FactoryRank <= 3 order by
YearlyProduction.Year DESC,YearlyProduction.FactoryRank
```

Screenshot of result:

Results Messages

	Year	FactoryLabel	TotalUnitsPassed
1	mY2022	Teghsat Bottled	5568
2	mY2022	Jadu Wholesale	5539
3	mY2022	Misrata Heating Equipment	5485
4	mY2021	Jadu Wholesale	1413
5	mY2021	Zaltan Broadwoven Fabric Mills	1399
6	mY2021	Misrata Heating Equipment	1375
7	mY2020	Jadu Wholesale	1705
8	mY2020	Qaminis Women'S	1674
9	mY2020	Jalu Water Transportation	1633
10	mY2019	Farzougha Household Audio	1407
11	mY2019	Al Gseibat Petroleum Refining	1406
12	mY2019	Misrata Heating Equipment	1358
13	mY2018	Jadu Wholesale	1611
14	mY2018	Misrata Heating Equipment	1572
15	mY2018	Nalut Retail	1544

Query executed successfully.

7. Crosstab/PIVOT. Use SQL Server PIVOT or PostgreSQL crosstab to create a table based on question 6

- **Year as column headings**
- **Factory Label as rows**
- **Quantity Passed as the data in the body of the table.**

Paste the code and a screen shot of the results.

SQL Code:

```
SELECT *
FROM (
    SELECT
        YearlyProduction.Year,
        YearlyProduction.FactoryLabel,
        YearlyProduction.QtyPassed AS TotalUnitsPassed
    FROM (
        SELECT
            YEAR(c.MANUFACTURE_DAY_DATE) AS Year,
            f.FACTORY_LABEL AS FactoryLabel,
            SUM(m.QTY_PASSED) AS QtyPassed,
            SUM(m.QTY_FAILED) AS QtyFailed,
            ROW_NUMBER() OVER (PARTITION BY YEAR(c.MANUFACTURE_DAY_DATE) ORDER BY
SUM(m.QTY_PASSED) DESC) AS FactoryRank
        FROM
            MANUFACTURE_FACT m
        JOIN
            CALENDAR_MANUFACTURE_DIM c ON m.MANUFACTURE_CAL_KEY =
c.MANUFACTURE_CAL_KEY
        JOIN
            FACTORY_DIM f ON m.FACTORY_KEY = f.FACTORY_KEY
        WHERE
            YEAR(c.MANUFACTURE_DAY_DATE) IN (2022, 2021, 2020, 2019, 2018)
            AND MONTH(c.MANUFACTURE_DAY_DATE) = 2
        GROUP BY
            YEAR(c.MANUFACTURE_DAY_DATE), f.FACTORY_LABEL
    ) AS YearlyProduction
    WHERE YearlyProduction.FactoryRank <= 3
) AS SourceTable
PIVOT (
    SUM(TotalUnitsPassed)
    FOR Year IN ([2022], [2021], [2020], [2019], [2018])
) AS PivotTable
ORDER BY FactoryLabel
```

Screenshot of result:

Results | Messages

	FactoryLabel	2022	2021	2020	2019	2018
1	Al Gseibat Petroleum Refining	NULL	NULL	NULL	1406	NULL
2	Farzougha Household Audio	NULL	NULL	NULL	1407	NULL
3	Jadu Wholesale	5539	1413	1705	NULL	1611
4	Jalu Water Transportation	NULL	NULL	1633	NULL	NULL
5	Misrata Heating Equipment	5485	1375	NULL	1358	1572
6	Nalut Retail	NULL	NULL	NULL	NULL	1544
7	Qaminis Women'S	NULL	NULL	1674	NULL	NULL
8	Teghsat Bottled	5568	NULL	NULL	NULL	NULL
9	Zaltan Broadwoven Fabric Mills	NULL	1399	NULL	NULL	NULL

 Copyright 2023 Boston University; All rights reserved

Part 4 – Tableau Data Presentation

In this section, you will be working in at least two Tableau workbooks.

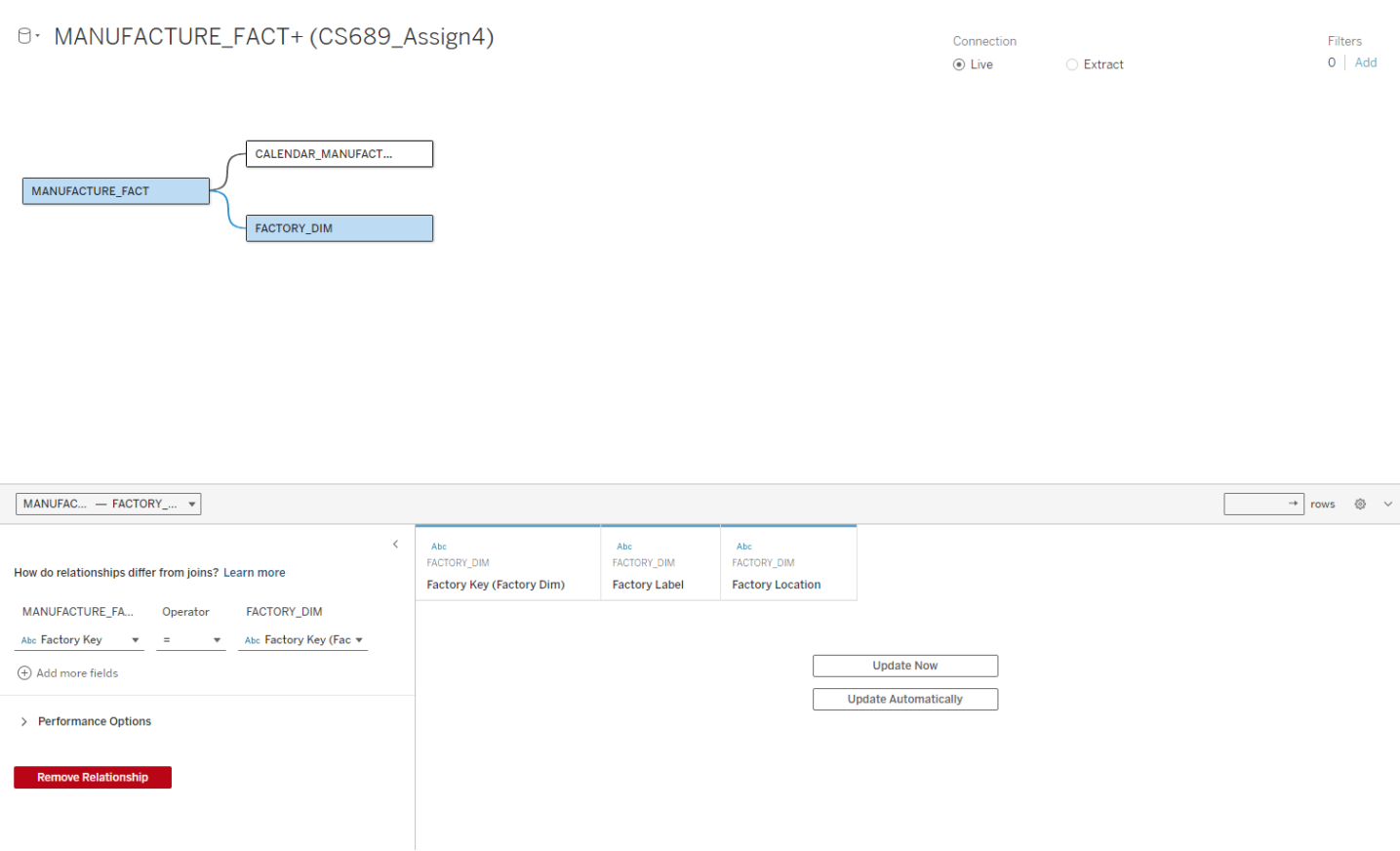
1. Download and install Tableau. You can get a student version here: https://www.tableau.com/academic/students. Note: it make a take few days to get a license. Alternatively, you can download a 14-day free trial version while you wait for your student license: https://www.tableau.com/products/trial

2. Start Tableau and connect it to your new data warehouse database.
   - Under Connect to a Server, choose "more" and select your DBMS (SQL Server or PostgreSQL), you will need to use the same connection options as when you sign into your DBMS.
     - If you are using PostgreSQL on a Mac and having issues connecting, please see Appendix section of this assignment.
   - Select the database you have been using for this assignment.

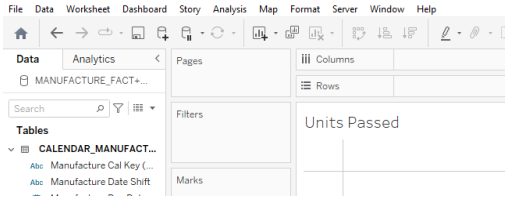
3. You will now see a list of tables you are now familiar with. Drag the MANUFACTURE_FACT, CALENDAR_MANUFACTURE_DIM, AND FACTORY_DIM from the Tables list within the Data Source tab. (The "Drag Tables Here" is where you drag it to). You will see lines connecting the tree tables, these indicate the joins. Click on each line to review that Tableau has joined the tables correctly.

   Take Screenshot of the Data Source table

 Copyright 2023 Boston University; All rights reserved



4. Sheet 1 will show in the bottom right corner, with “Go to Worksheet link”.
   - Click on Sheet 1 to open the worksheet.
   - Rename Sheet 1 worksheet to Units Passed (right click on it to show menu option to rename it)
   - Notice the three tables on the right, columns and rows area at the top.

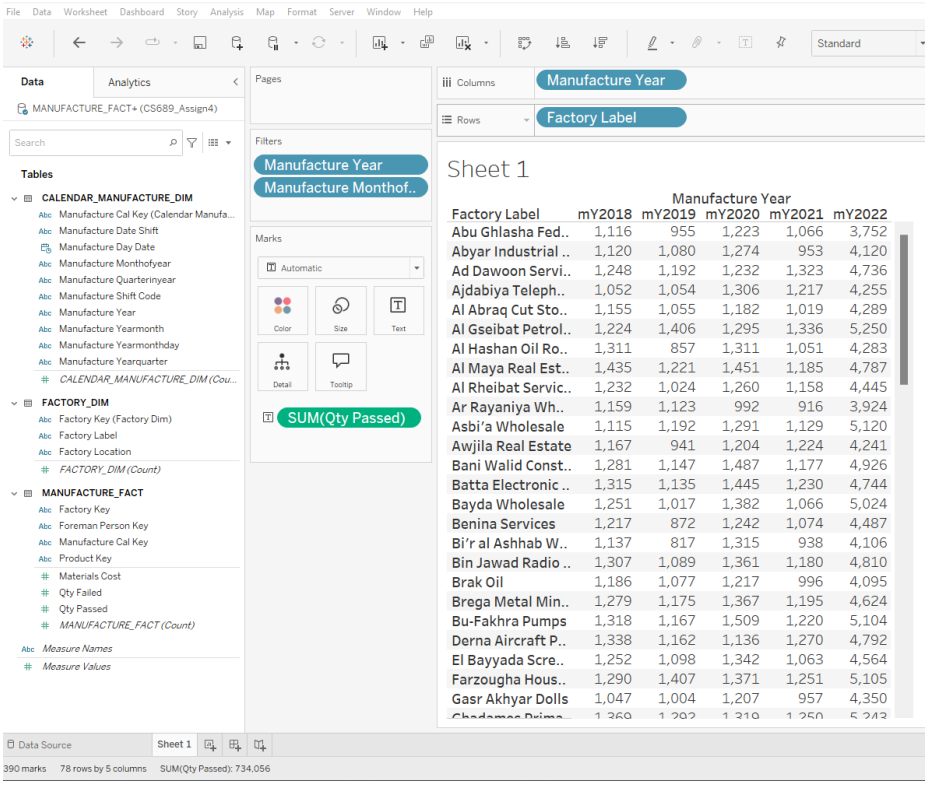


   - Place Manufacture Year (from the CALENDAR_MANUFACTURE_DIM table) as columns and Factory Label (from the FACTORY_DIM table) as Rows. You will now see years as columns and factories as rows.
   - Under the Manufacture fact, drag the Qty Passed into the center of the pivot table (where you see the “abc” for each record). Under the Marks, Sum (Qty Pass will show)
   - Drag the Manufacture Year (from the CALENDAR_MANUFACTURE_DIM table) into the Filters box, in the General tab, select the latest 5 years

- Drag the Manufacture Monthofyear (from the CALENDAR_MANUFACTURE_DIM table) into the Filters box, in the General tab, select February.

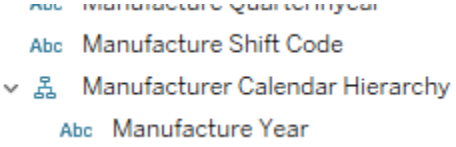
The pivot table we just created is the exact same one you solved in question 7 from part 3! You should see the same data.

Take Screenshot of the worksheet showing the pivot table



5. Create a Manufacture Calendar Hierarchy.

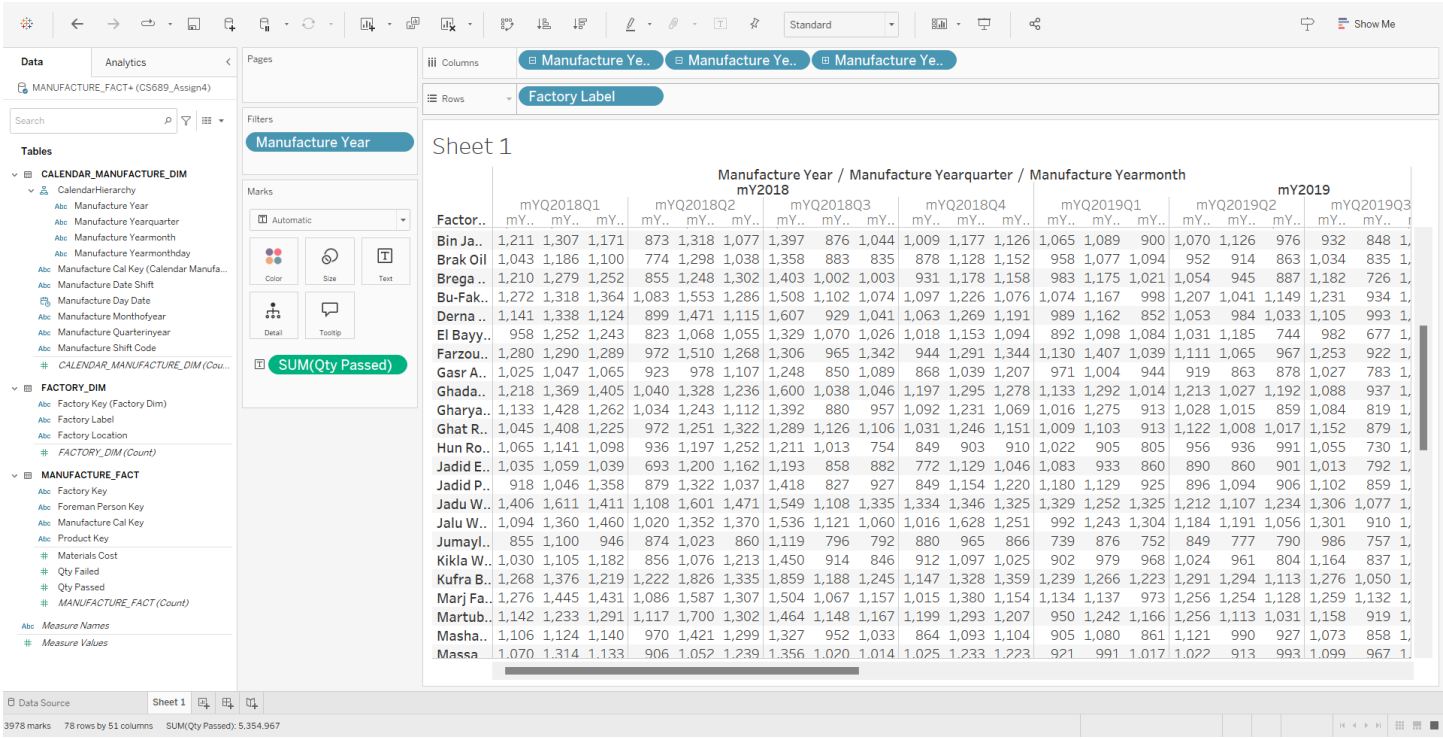
- **Under the CALENDAR_MANUFACTURE_DIM select the four attributes which represent the calendar hierarchy, right click on them and from the Hierarchy option select create Hierarchy. It should use four of the table's fields with the following order – Year, Yearquarter, Yearmonth, Yearmonthdat. The hierarchy should increase in detail as you move down. Here is a hint to what it should look like (the first field is shown):**



- Remove the Manufacture Year from the Columns at the top and replace it with the Manufacture Calendar Hierarchy you just created.
- You will see a plus next to the Manufacture Year in your column area, click on it to expand date hierarchy column to Months which will appear in your pivot table.
- Remove the February month filter, you should see an expanded workbook of the year and all the months under it.

Varun Nair

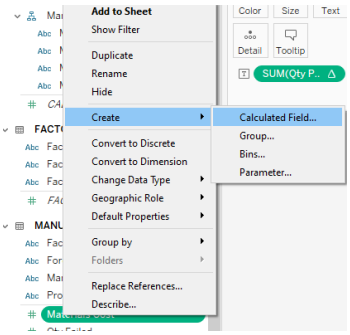
Take Screenshot of the worksheet showing the pivot table including the Manufacture Calendar Hierarchy in the tables area on the left



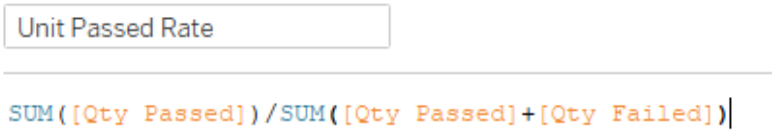
6. Let's create a calculated measure called Unit Passed Rate.

- First copy the worksheet you have been using
- Right-click in the Measures section at lower left (on any of the existing measures) and select Create to add a calculated field called Unit Passed Rate.

Varun Nair



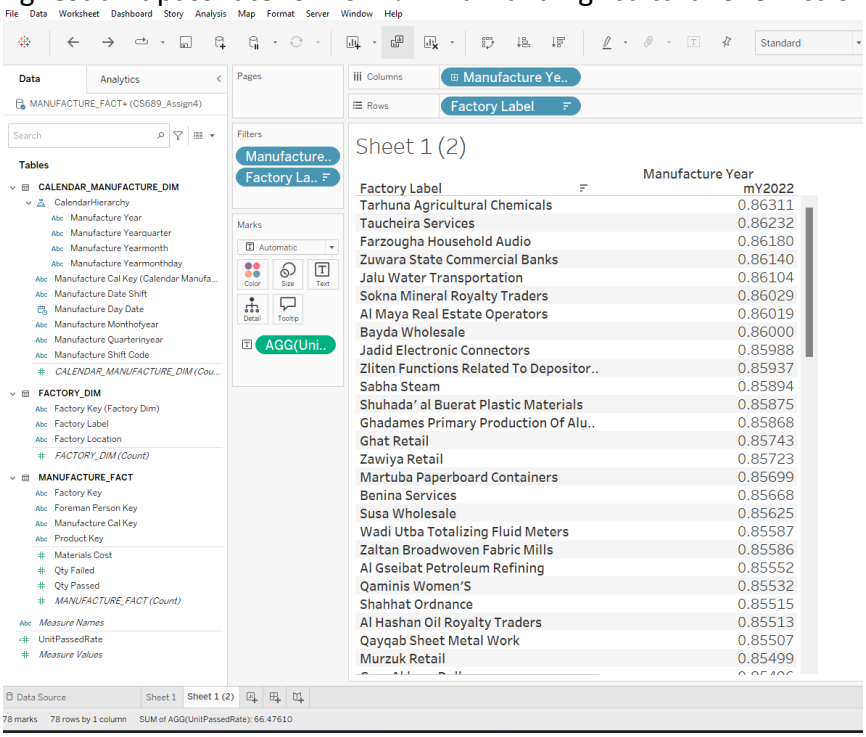
- The formula should divide the sum of quantity passed by the total quantities (passed and failed) summed.



- Replace the Qty Passed sum measure with Unit Passed Rate calculated measure in the pivot, and in addition add Factory location as the first hierarchy in the rows.

What factory in Quebec has the highest unit pass rate for 2022? - Tarhuna Agricultural Chemicals

Take a screenshot showing Quebec and its factories along with the unit pass rate



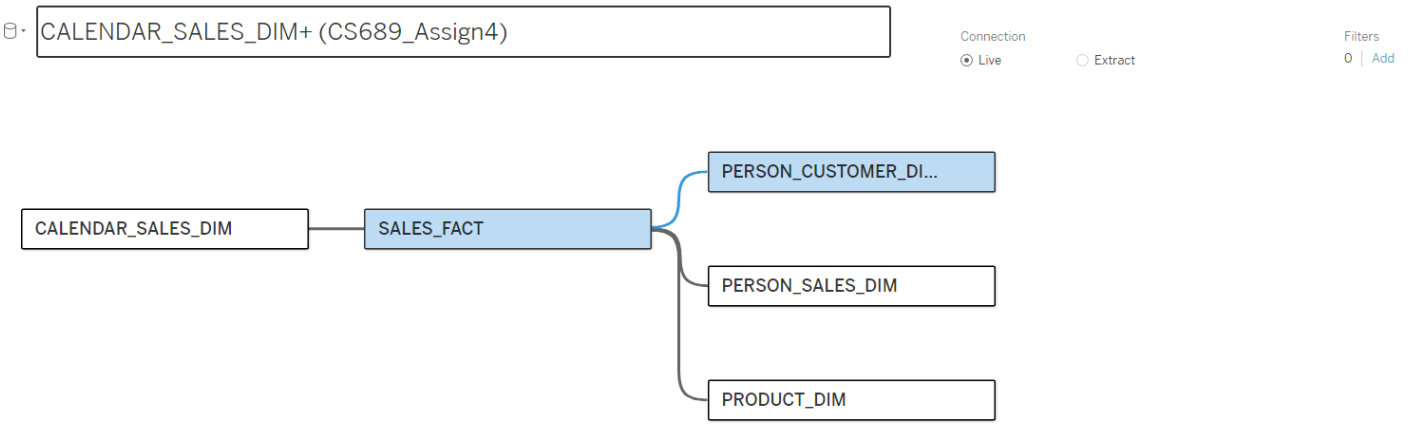
 Copyright 2023 Boston University; All rights reserved

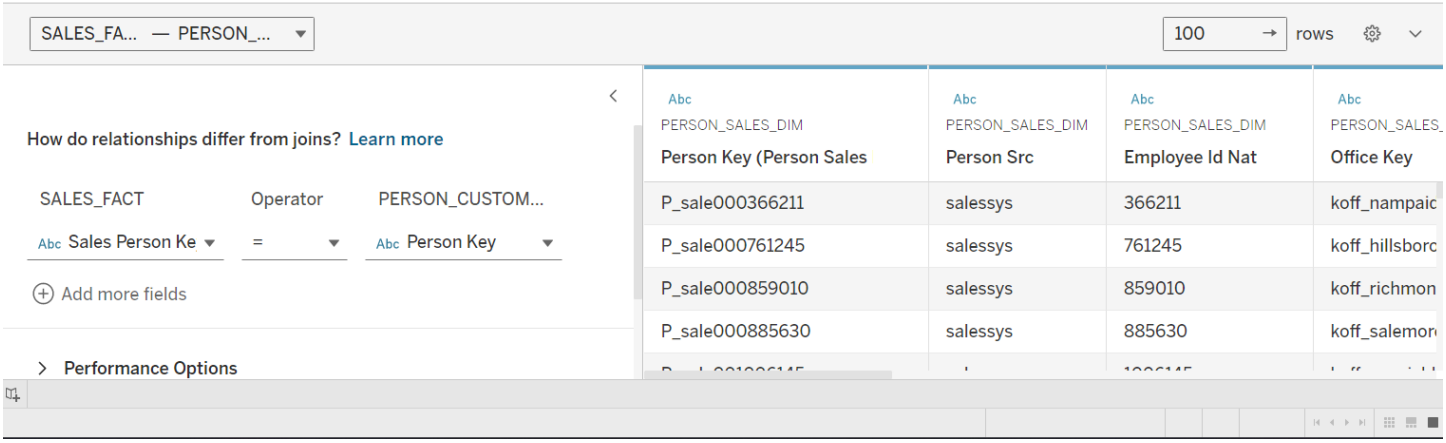
Varun Nair

7. Create a New Tableau Workbook for Sales and set up a Data Source for Sales Facts.

   - You will need CALENDAR_SALES_DIM, SALES_FACT, PERSON_CUSTOMER_DIM, PERSON_SALES_DIM, AND PRODUCT_DIM. When adding the two person DIM tables, review the key relationships as the names are not consistent, match the salesperson key and customer person key with the appropriate person key in both of the DIM tables.

   Take a screenshot of the Sales Data Source

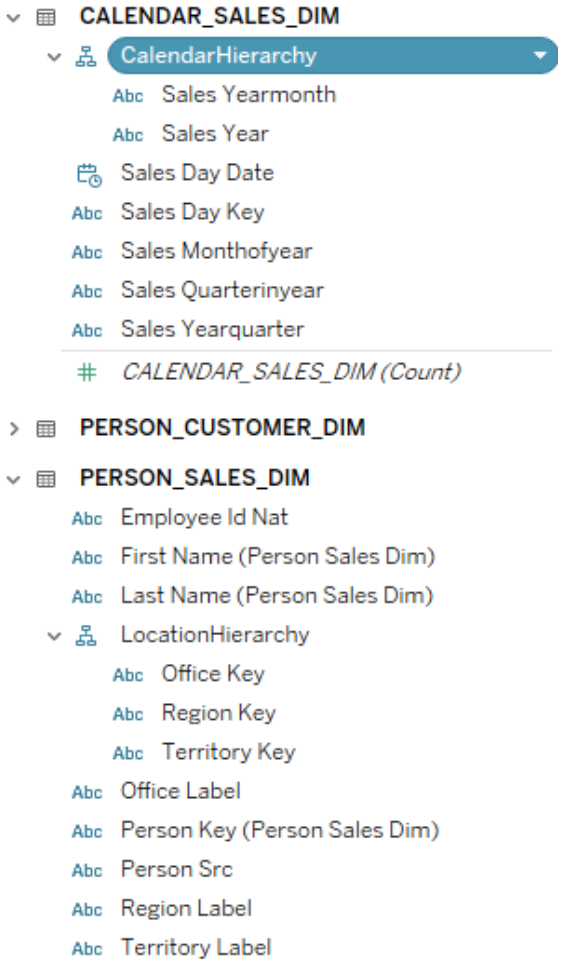




 Copyright 2023 Boston University; All rights reserved

8. Set up hierarchies.
   - Click into "Sheet 1" at the bottom to start a worksheet. You will see a completely empty sheet, with dimensions and measures.
   - Review all the dimension tables and set up hierarchies which make sense to you. Two hierarchies in total should be fine here.

   Provide screenshots of the two hierarchies

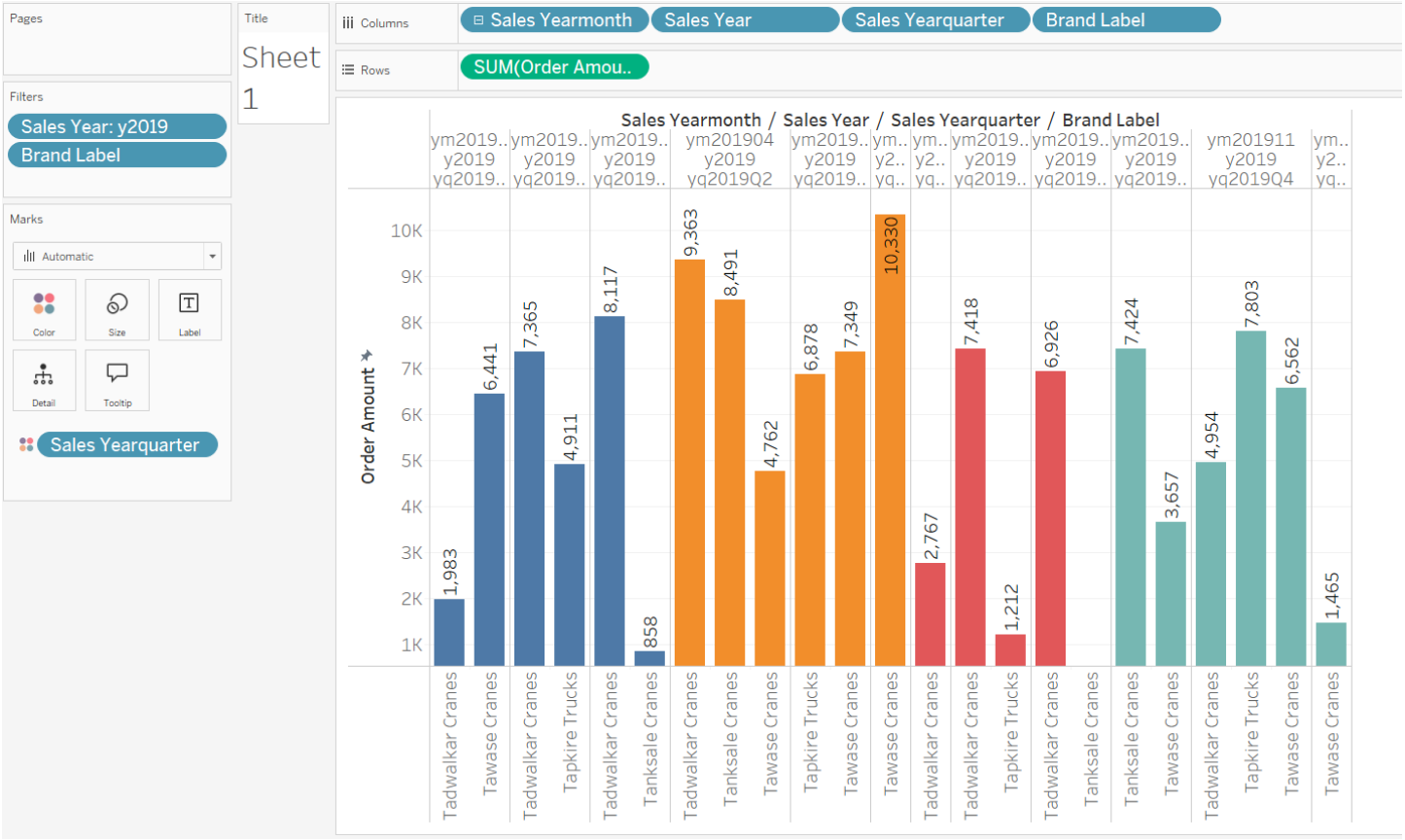


9. Create a Bar Chart for Sales Revenue

   We want to compare performance quarter-by-quarter for 2019 between the brands whose names begin with "Ta". As you complete each of these bullet points, the bar chart will take shape.
   - Add the date hierarchy as a column and expand to quarters.
   - Rows don't just have to be dimensions, we can add a measure for the sum of order amounts from the sales fact here.
   - Add a filter from the calendar dimension for 2019.
   - Add Brand Label from the Products dimension to the columns

- Add a filter for brand label, selecting only the brands beginning with "Ta", there should be 4 product labels that being with “Ta”
- Your bar chart is complete, however let’s add color! Select use the “Side-by-Side” bars. If you recall we are comparing products by a single variable, and side-by-side chart works great here!
- In the Marks click on the Label, and check the box to Show mark labels so that the amounts are shown.
- In the columns area, you can drag the Sales Yearquarter and the Brand label to have them switch spaces, note how the chart will re-organize.

Provide screenshot of your chart, including all the filters, columns, rows, etc.



Which brand had the worst quarter in 2019 and what was the order amount?
TankSale Cranes in Q3
Which brand had the best quarter in 2019 and what was the order amount?
Tawase Cranes in Q2

 Copyright 2023 Boston University; All rights reserved

Varun Nair

Criterion	A	B	C	D	F	Letter Grade
Technical mastery (50%)	Evidence of excellent mastery throughout	Evidence of good mastery throughout	Evidence of basic mastery throughout or good mastery intermittently	Minimal mastery evidenced	Virtually no mastery evidenced	
Depth and thoroughness of coverage (25%)	Excellent depth and coverage of significant topics and issues	Good depth and coverage of significant topics and issues	Basic depth and coverage of significant topics and issues	Minimal depth and coverage of significant topics and issues	Virtually no depth and coverage of significant topics and issues	
Clarity in presentation (25%)	Ideas and designs are exceptionally clear and organized throughout	Ideas and designs are clear and organized throughout	Ideas and designs are somewhat clear and organized throughout	Ideas and designs are mostly obscure and disorganized	Ideas and designs are entirely obscure and disorganized	
					Assignment Grade:	

 Copyright 2023 Boston University; All rights reserved