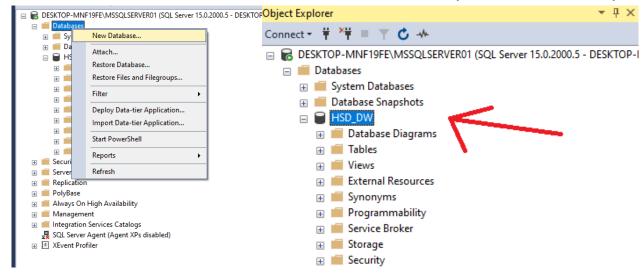
# DTI5126[EG] Fundamentals/Applied Data Science Assignment 1

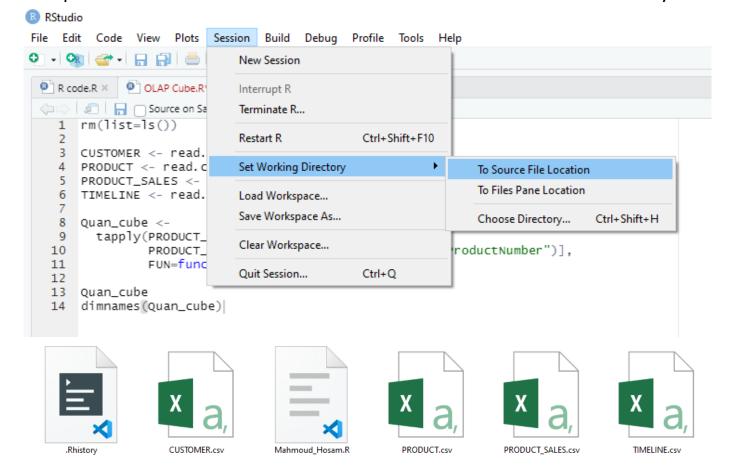


- 1- Hosam Mahmoud Ibrahim 300327269
  - 2- Abdallah Ahmed Ragab 300327288

Please make sure that the database is created in .sql files to run as expected.



And please make sure that the .csv files is with .R file in the same directory.



- Part A: RDBMS & SQL
  - a) Select from TRANS where DateSoldID is not null to get rows with null then delete from TRANS where DateSoldID is not null to delete these rows

```
/* question Part A number a */
|select * from TRANS where DateSold is null
| delete from TRANS where DateSold is null
```

**b)** Use the CONCAT function to get FirstName and LastName and use where to get all rows that have Yellow Blue or White in the artwork title

```
/* question Part A number b */
select WorkID , Title, WORK.Medium , WORK.ArtistID ,CONCAT(TRIM (FirstName) , ' ' ,TRIM (LastName) ) as FullName from WORK
inner join ARTIST on
ARTIST.ArtistID = WORK.ArtistID

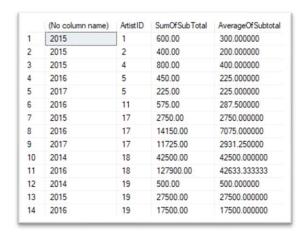
where title LIKE '%Yellow%'
  or title LIKE '%Blue%'
  or title LIKE '%Blue%'
```

	WorkID	Title	Medium	ArtistID	FullName
1	523	On White II	High Quality Limited Print	2	Wassily Kandinsky
2	571	Yellow Covers Blue	Oil and collage	18	Paul Horiuchi
3	590	Blue Interior	Tempera on card	17	Mark Tobey

C) Uses group by YEAR(DateSold), ARTIST.ArtistID to get the sum and average of the sales price For each artist in each year

```
/* question Part A number C */
select YEAR(DateSold) , ARTIST.ArtistID , sum(SalesPrice) as SumOfSubTotal , avg(SalesPrice) as AverageOfSubtotal from TRANS
inner join WORK
on WORK.WorkID = TRANS.WorkID

inner join ARTIST
on ARTIST.ArtistID =WORK.ArtistID
group by YEAR(DateSold), ARTIST.ArtistID
```



**d)** Calculate the average sales prices and compare the average and the sales prices for the artists and get the artist that has a sales price greater than the average

```
/* question Part A number d */
select ARTIST.ArtistID , FirstName , LastName , WORK.WorkID , Title from TRANS
inner join WORK
on WORK.WorkID = TRANS.WorkID
inner join ARTIST
on ARTIST.ArtistID =WORK.ArtistID
where TRANS.SalesPrice> (select AVG(SalesPrice) from TRANS)
```

	ArtistID	FirstName	LastName	WorkID	Title
1	18	Paul	Horiuchi	500	Memories IV
2	19	Momis	Graves	548	Night Bird
3	19	Morris	Graves	561	Sunflower
4	17	Mark	Tobey	570	Untitled Number 1
5	18	Paul	Horiuchi	571	Yellow Covers Blue
6	18	Paul	Horiuchi	500	Memories IV

e) Use Update statement to update EmailAddress and EncryptedPassword for specific Customer Lynda

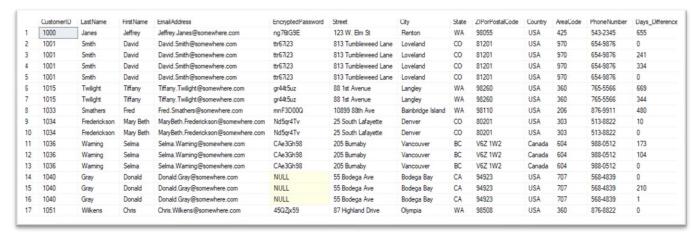
```
/* question Part A number E */
UPDATE CUSTOMER set
EmailAddress = 'Johnson.lynda@somewhere.com' ,
EncryptedPassword = 'aax1xbB'
where FirstName='Lynda' and LastName = 'Johnson'
```

```
(1 row affected)

Completion time: 2022-06-09T08:16:18.4318512+02:00
```

**f)** Use the <u>lead</u> function to get the next purchase then use <u>DATEDIFF</u> to calculate days between two purchases and drop the rows that have null in the Days Difference column

```
/* question Part A number f */
select * from (
    select Customer.* ,
    (DATEDIFF( day, LEAD(DateSold,1) over (PARTITION BY TRANS.CustomerID ORDER BY DateSold DESC ) , DateSold )) as Days_Difference
    from TRANS
    inner join CUSTOMER on
    CUSTOMER.CustomerID =TRANS.CustomerID
) as My_Table
where Mv Table.Davs Difference is not null
```



g) Create View and select the data from the view

Commands completed successfully.

Completion time: 2022-06-09T08:57:47.8657605+02:00



DateAcquired DateSoid rullivame 2015-09-21 2015-11-28 12500.00 Momis Graves Night Bird 2014-11-04 2 Paul Horiuchi Memories IV 2014-12-14 12500.00 3 Paul Horiuchi Yellow Covers Blue 2016-08-23 2016-09-29 20000.00 Paul Horiuchi Memories IV 2016-09-29 2016-12-18 32500.00

**h)** Use with statement to get min and max of the purchase date for each customer and combine the result with the rest of the requested rows then use into #TableName to create and insert into the temporary table

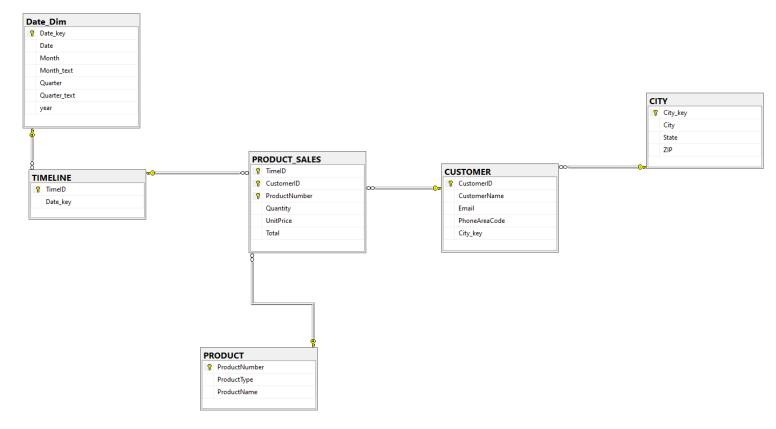
```
/* question Part A number h */
WITH CTEPurchase(CustomerID, MinAcquisitionDate, MaxAcquisitionDate)
AS
    (SELECT
    CustomerID,
    MIN(TRANS.DateAcquired) AS MinAcquisitionDate,
    MAX(TRANS.DateAcquired) AS MaxAcquisitionDate
    FROM TRANS
    GROUP BY TRANS.CustomerID)
select
TransactionID,
DateAcquired,
CTEPurchase.CustomerID,
Customer.LastName,
Customer.FirstName,
MinAcquisitionDate,
MaxAcquisitionDate,
WHEN Medium='High Quality Limited Print' THEN 1
WHEN Medium ='Color Aquatint' THEN 2
WHEN Medium ='Water Color and Ink' THEN 3
WHEN Medium ='Oil and Collage' THEN 4
ELSE 5 END AS Medium
into #Purchase
FROM CTEPurchase
INNER JOIN CUSTOMER ON
CTEPurchase.CustomerID = CUSTOMER.CustomerID
INNER JOIN TRANS ON
TRANS.CustomerID = CTEPurchase.CustomerID
INNER JOIN WORK ON
TRANS.WorkID = WORK.WorkID
where Year(DateAcquired)>=2015 AND Year(DateAcquired)<=2017</pre>
```

### And when selecting data from the temporary table it is show as the follow

	TransactionID	DateAcquired	CustomerID	LastName	FirstName	MinAcquisition Date	MaxAcquisition Date	Mediu
1	115	2015-03-03	1033	Smathers	Fred	2015-03-03	2016-06-28	5
2	121	2015-09-21	1015	Twilight	Tiffany	2014-11-07	2017-08-29	5
3	125	2015-11-21	1001	Smith	David	2014-11-17	2016-05-18	1
4	127	2015-11-21	1034	Frederickson	Mary Beth	2014-11-17	2015-11-21	1
5	128	2015-11-21	1036	Warning	Selma	2015-11-21	2016-09-29	1
6	129	2015-11-21	1036	Warning	Selma	2015-11-21	2016-09-29	1
7	151	2016-05-07	1036	Warning	Selma	2015-11-21	2016-09-29	5
8	152	2016-05-18	1001	Smith	David	2014-11-17	2016-05-18	1
9	153	2016-05-18	1001	Smith	David	2014-11-17	2016-05-18	1
10	154	2016-05-18	1040	Gray	Donald	2016-05-18	2017-02-28	1
11	156	2016-05-18	1040	Gray	Donald	2016-05-18	2017-02-28	1
12	161	2016-06-28	1033	Smathers	Fred	2015-03-03	2016-06-28	5
13	171	2016-08-23	1000	Janes	Jeffrey	2014-11-04	2016-08-23	4
14	175	2016-09-29	1036	Warning	Selma	2015-11-21	2016-09-29	5
15	201	2017-02-28	1040	Gray	Donald	2016-05-18	2017-02-28	2
16	202	2017-02-28	1040	Gray	Donald	2016-05-18	2017-02-28	2
17	225	2017-06-08	1051	Wilkens	Chris	2017-06-08	2017-06-08	1
18	227	2017-06-08	1051	Wilkens	Chris	2017-06-08	2017-06-08	1
19	241	2017-08-29	1015	Twilight	Tiffany	2014-11-07	2017-08-29	5

# • Part B: Data Warehousing & OLAP

# 1- snowflake schema for the data warehouse



```
----- SNOWFLAKE Dimintions -----
                                                                                   /*CREATE TABLE CITY(
/*CREATE TABLE Date_Dim(
                                                                                   City_key int NOT NULL
Date_key Int Not Null
,Date Date Not NULL
,Month int NOT NULL
                                                                                   ,ZIP int NOT NULL
,Month_text Char(50) not NULL
,Quarter int not null
,Quarter_text char(50) not null
,year int not null
,CONSTRAINT Date_key_PK PRIMARY KEY (Date_key)
CREATE TABLE TIMELINE(
TimeID Int NOT NULL
,Date key Int Not Null
,CONSTRAINT TimeID_PK PRIMARY KEY (TimeID)
,CONSTRAINT Date_key_fk FOREIGN KEY(Date_key) REFERENCES Date_Dim(Date_key)
```

```
/*CREATE TABLE CITY(
City_key int NOT NULL
,City VarChar(255) NOT NULL
,State Char(50) NOT NULL
,ZIP int NOT NULL
CONSTRAINT City_key_PK PRIMARY KEY(City_key)
);

CREATE TABLE CUSTOMER(
CustomerID Int NOT NULL
,CustomerName VarChar(255) NOT NULL
,Email VarChar(255) NOT NULL
,PhoneAreaCode int NOT NULL
,City_key int NOT NULL
,CONSTRAINT CustomerID_PK PRIMARY KEY(CustomerID)
,CONSTRAINT CustomerID_PK PRIMARY KEY(CustomerID)
,CONSTRAINT City_key_FK FOREIGN KEY(City_key) REFERENCES CITY(City_key)
);*/
```

2-(a): Customers that at least made five products with different product numbers.

```
--Q2 a
   SELECT customer.CustomerName, customer.CustomerID FROM customer
     Where customer.CustomerID = SOME (
     SELECT product_sales.CustomerID from product_sales
     GROUP BY product_sales.CustomerID
     HAVING COUNT(DISTINCT product_sales.ProductNumber) >= 5
     );
121 % 🕶 🔻
Results Messages
    CustomerName CustomerID
    Able, Ralph
1
2
    Baker, Susan
3
   Foxtrot, Kathy
   Pearson, Bobbi 9
5
    Tyler, Jenny
               11
6
   Wayne, Joan 12
```

2-(b): customers that made the largest orders.

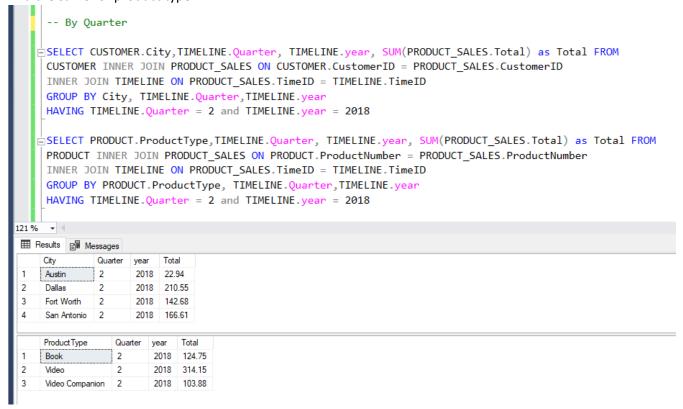
2-(c): Roll-Up operation to summaries the total sales per Year.

# 3-(a): customers that made the largest orders.

Here we aggregated the next for columns together (**City** and **Quarter** and **Year** and **Total**) because we want to know which (city or product type) is responsible for decreasing.

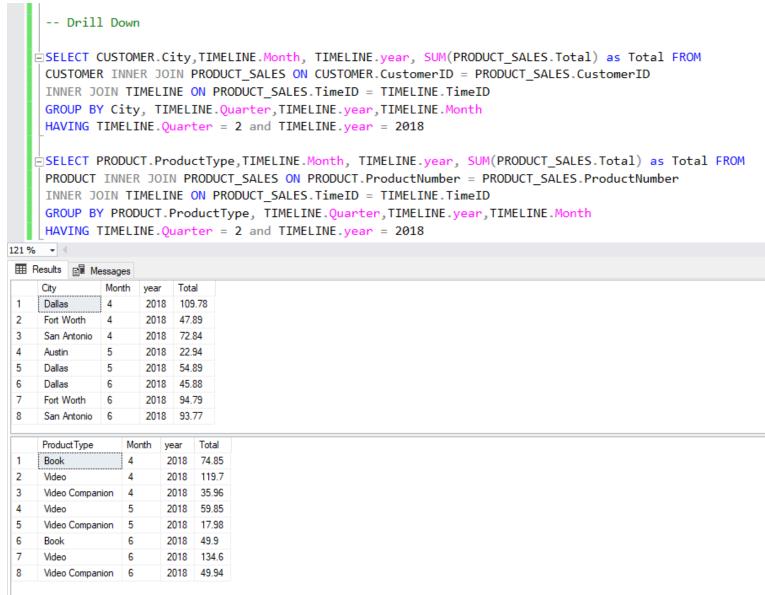
And we observed that the city **Austin** has the minimum value and that would indiacted that this city has a relation with the decrease that happened.

And the same for product type...



And as we will see in **3-(b)** when we apply (Drill Down Operation) we will know when (city or product type) has decreased in each month from **April 2018 to June 2018**.

# 3-(b): here we used Drill Down Operation to stepping down a concept hierarchy for a dimension (change quarter column to months column).



In the above figure we have observed that the value of city **Dallas** has decreased from April (was 109.78) to (54.89 in month 5) to (45.88 in month 6), so it's decreasing constantly and of course it have an effect in some way on the overall decreasing.

And the same happened in the product type table...

### 4- Build an OLAP cube for the Sum of Total Quantity.

```
R • Global Environment •
    CUSTOMER <- read.csv(file = 'CUSTOMER.csv')
PRODUCT <- read.csv(file = 'PRODUCT.csv')
PRODUCT_SALES <- read.csv(file = 'PRODUCT_SALES.csv')
TIMELINE <- read.csv(file = 'TIMELINE.csv')
                                                                                      Data
                                                                                      CUSTOMER
                                                                                                                                       12 obs. of 7 variables
                                                                                      PRODUCT
                                                                                                                                       10 obs. of 3 variables
                                                                                                                                       48 obs. of 6 variables
                                                                                      PRODUCT_SALES
   Ouan_cube
                                                                                                                                       12 obs. of 7 variables
                                                                                      TIMELINE
      values
                                                                                                                                       int [1:12, 1:10, 1:9] NA NA NA NA NA NA NA 1 NA NA ...
11
                                                                                        Ouan cube
12
    Quan_cube
    dimnames(Quan_cube)
```

#### **Cube layers:**

```
, , ProductNumber = VB003
, , ProductNumber = BK001
                                                                                                     . . ProductNumber = VB001
                                                        CustomerID
      CustomerID
                                                                                                           CustomerID
                                                 TimeID
                                                            3 4 5 6 7 8 9 11 12
TimeID
               5 6 7 8 9 11 12
                                                                                                      1 3 4
                                                   43023 NA NA NA NA NA NA NA NA NA NA
 43033 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43033 NA NA 1 NA NA NA NA NA NA NA
                                                   43089 NA NA NA NA NA NA NA NA NA
 43089 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43089 NA NA NA NA NA NA NA NA NA NA
                                                   43184 NA NA NA NA NA NA NA NA NA NA
 43184 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43184 NA NA NA NA NA NA NA NA NA NA
 43186 NA NA NA NA NA 1 NA NA NA NA
                                                   43186 NA NA NA NA 1 NA NA NA NA NA
                                                                                                       43186 NA NA NA NA NA NA NA NA NA NA
                                                   43190 NA NA NA NA NA NA NA NA NA
 43190 NA NA NA NA NA NA
                           1 NA NA
                                                                                                       43190 NA NA NA NA NA NA NA
 43193 NA NA NA NA NA NA NA NA NA NA
                                                   43193 NA NA NA NA NA NA NA NA
                                                                                                       43193 NA NA NA NA NA NA NA NA NA NA
 43198
       1 NA NA 1 NA NA NA NA NA NA
                                                   43198 ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ ΝΔ
                                                                                                       43198
                                                                                                             1 NA NA 1 NA NA NA NA NA NA
          1 NA NA NA NA NA NA NA NA
                                                  43213 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43213 NA NA NA NA NA NA NA NA NA NA
 43227 NA NA NA NA NA NA NA NA NA NA
                                                   43227 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43227 NA NA NA NA NA NA NA NA NA NA
 43241 NA NA NA NA NA NA NA NA NA NA
                                                   43241 NA NA NA NA NA NA 1 NA NA NA
                                                                                                       43241 NA NA NA NA NA NA NA NA NA NA
 43256 NA NA NA NA NA NA NA NA NA NA
                                                   43256 NA 1
                                                                                                       43256 NA 1 NA NA NA NA NA NA NA NA
, , ProductNumber = BK002
                                                 , , ProductNumber = VK001
                                                                                                     . . ProductNumber = VB002
CustomerID
TimeID 1 3 4
 imeID 1 3 4 5 6 7 8 9 11 12
43023 NA NA NA NA NA NA NA NA NA NA
                                                       CustomerID
                                                                                                           CustomerID
                                                  TimeID
                                                                                                     TimeID
                                                                                                             1 3 4 5 6 7 8 9 11 12
                                                                                                       43023 NA NA NA NA NA NA NA NA NA NA
 43033 NA NA NA NA NA NA NA NA NA NA
  43089 NA NA NA NA NA NA NA NA NA
                                                   43033 NA NA 1 NA NA NA NA NA NA NA
                                                                                                       43033 NA NA NA NA NA NA NA NA NA NA
                                                   43089 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43089 NA NA NA NA NA NA NA NA NA NA
 43184 NA NA 1 NA NA NA NA NA NA NA
 43186 NA NA NA NA
                      1 NA NA NA NA
                                                                                                       43184 NA NA NA NA NA NA NA NA NA NA
                                                   43184 NA NA NA NA NA NA NA NA NA NA
 43190 NA NA NA NA NA NA NA NA NA NA
                                                   43186 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43186 NA NA NA NA NA NA NA NA NA NA
 43193 NA NA NA NA NA NA NA NA
                                                                                                       43190 NA NA NA NA NA NA NA NA NA
                                                   43190 NA NA NA NA NA NA NA 1 NA NA
 43198 NA NA NA NA NA NA NA NA NA NA
                                                   43193 NA NA NA NA NA NA NA NA NA NA
                                                                                                       43193 NA NA NA NA NA NA NA NA NA NA
                                                   43198
                                                         1 NA NA 1 NA NA NA NA NA NA
                                                                                                       43198 NA NA NA NA NA NA NA NA NA NA
 43227 ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ ΝΑ
                                                                                                       43213 NA NA NA NA NA NA NA NA NA NA
                                                   43213 NA NA NA NA NA NA NA NA NA NA
 43241 NA NA NA NA NA NA NA NA NA
                                                                                                       43227 NA NA NA NA NA NA NA 1 NA NA
                                                   43227 NA NA NA NA NA NA NA NA NA NA
 43256 NA 1 NA NA NA NA NA NA NA 1
                                                                                                       43241 NA NA NA NA NA NA NA NA NA NA
                                                   43241 NA NA NA NA NA NA NA NA NA
                                                                                                       43256 NA 2 NA NA NA NA NA NA 2 NA
                                                   43256 NA 1 NA NA NA NA NA NA NA NA
```

```
, , ProductNumber = VK002
         CustomerID
  imeID 1 3 4 5 6 7 8 9 11 12
43023 NA NA NA NA NA NA NA NA NA NA
TimeID
   43033 NA NA NA NA NA NA NA NA NA
   43089 NA NA NA NA NA NA NA NA NA NA
  43184 NA NA 1 NA NA NA NA NA NA NA A43186 NA NA NA NA NA 1 NA NA NA NA NA NA NA
   43190 NA NA NA NA NA NA NA NA NA NA
   43198 NA NA NA NA NA NA NA NA NA NA
   43241 NA A
43256 NA 2 NA NA NA NA NA NA A
2 1
 , , ProductNumber = VK003
        TimeID
  43023 NA NA
43033 NA NA
   43089 NA NA NA NA NA NA NA NA NA NA
   43186 NA NA NA NA
                         1
                             1 NA NA NA NA
   43190 NA NA NA NA NA NA NA NA NA
   43193 NA NA NA NA NA NA NA NA
   43198 NA NA NA NA NA NA NA NA NA NA
   43227 NA NA
   43241 NA NA NA NA NA NA
   43256 NA NA NA NA NA NA NA NA NA
 , , ProductNumber = VK004
         CustomerID
TimeID
                     567
                                 8 9 11 12
           1 3
  IMBEID I 3 4 5 6 7 8 9 11 12
43023 NA NA NA NA NA NA NA NA NA
43033 NA NA NA NA NA NA NA NA NA
43089 NA NA NA NA NA I NA NA NA NA
43184 NA NA I NA NA NA NA NA NA NA
```

#### **Dimentions**

```
> dimnames (Quan_cube)
$TimeID
[1] "43023" "43033" "43089" "43184" "43186" "43190" "43193" "43198" "43213" "43227" "43241" "43256"

$CustomerID
[1] "1" "3" "4" "5" "6" "7" "8" "9" "11" "12"

$ProductNumber
[1] "8K001 " "8K002 " "V8002 " "V8002 " "V8001 " "VK001 " "VK001 " "VK001 " "VK001 " "VK004 " "VK04 "
```