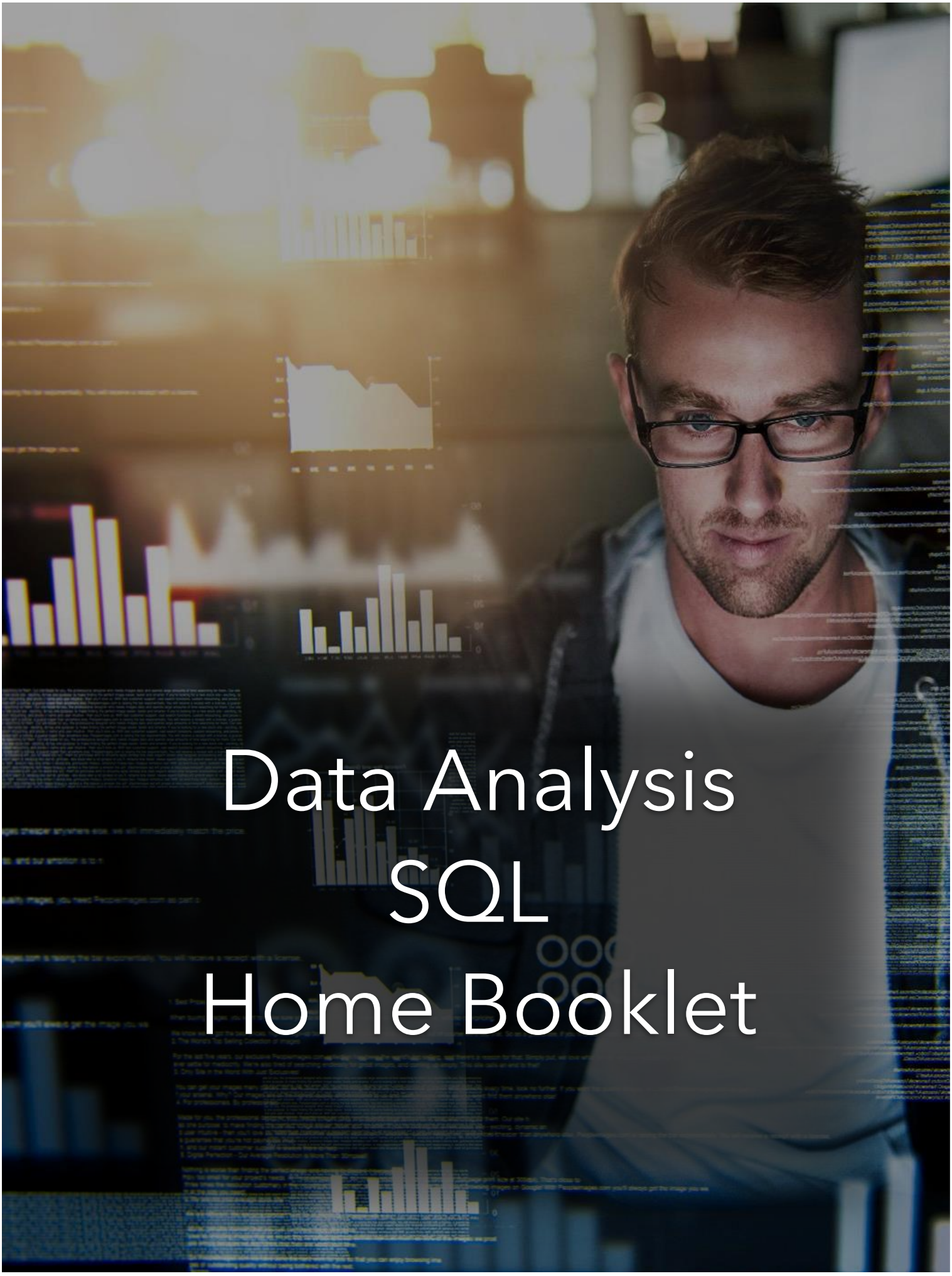
wawiwa

T-M



Data Analysis SQL Home Booklet

Lesson 1 – Introduction to data analysis

Table Column Links, Related Tables, Types of Keys

Examine the attached set of tables and answer the following questions:

1. Study each table and try to determine which entity it represents. Then underline the field or fields that constitute the primary key.
   (E.g., in an Orders table, Primary key = Order number + Line number)

2. Draw a line between the tables that are related. Draw the line from the field in one table to the field in the other table that represents the same value. (E.g., Customer number in Orders → Customer number in Customers.)

3. Continuing from the previous section, if the relationship you created in the previous section represents a foreign key, indicate this next to the table in which the field is defined as a foreign key.
   (E.g., Customer number in the Orders table is the foreign key of Customer number in the Customers table.)

4. For each relationship that you created in the previous sections, define whether the relationship is one or many.
   (E.g., An item code will only appear once in the item table, but may appear N times in the Orders table.)

Good luck!

SQL Unit – Home practice booklet

Tables Definitions

CallsTable
CallNo
CallDate
EmployeeNoOpenCall
Classification
Status
StreetCode
Description
CloseDate
EmployeeNoHandleCall

EmployeeTable
EmployeeNo
FirstName
LastName
DepartmentCode

DepartmentTable
DepartmentCode
DepartmentName

An Example of Data from the Tables

CallsTable

CallNo	CallDate	EmployeeNo OpenCall	Classification	Status	Description	CloseDate	EmployeeNo HandleCall
100	01/01/2021	451	High	1	Malfunction screen		452
101	01/01/2021	453	Low	1	I can't print anything		450
102	02/01/2021	453	Medium	0	My computer is running too slowly		
103	03/01/2021	451	High	9	Unable to log in	03/01/2021	450

EmployeeTable

EmployeeNo	FirstName	LastName	DepartmentCode
450	James	Smith	1
451	Jennifer	Johnson	4
452	Patricia	Williams	1
453	John	Brown	2

DepartmentTable

DepartmentCode	DepartmentName
1	IT
2	Sales
3	HR
4	Marketing

Lesson 2 – Basic Retrieval and Data Filtering

Subjects:

- Basic SELECT statement
- Basic Filtering – WHERE clause
- Filtering using AND, OR, NOT Operators

1. Write a query that displays the Order number (SalesOrderID), Order date, Customer Number (CustomerID) and Order amount (SubTotal) from the Sales.SalesOrderHeader table, for the orders above $1500 and an Order date from Jan. 1,2013 onwards.

2. Write a query that displays all the data from the Person.Person table, only for people whose BusinessEntityID is above 10,000 and their first name is either Jack or Crystal.

3. Write a query that displays the SalesOrderID, ProductID, and total amount for that order line (LineTotal) only for items with a Line Total between 100 and 1.000, inclusive.

Lesson 3 – Calculated Columns and Sorting

Subjects:

- Calculated Columns
- Aliases
- ORDER BY Keyword

1. Write a query that displays the ProductID, Product Name, Color, Weight and the profit margin (ListPrice – StandardCost) from the Production.Product table.
   Display only the products that have a value for Weight.
   Sort the results by Color, descending, and Weight, ascending.

2. The company wants to check how the prices will change if every order has a $50 delivery charge added.
   Write a query that displays the following columns from the Sales.SalesOrderHeader table:
   Order number (SalesOrderID), Order amount (SubTotal), and Order amount + $50 (named SubTotalPlus50).
   Sort the query results according to Order Amount from the highest to lowest.

Lesson 4 – Grouping and Aggregate

Subjects:

- GROUP BY Keyword
- Aggregate Functions
- YEAR() function
- HAVING Clause

1. Write a query that displays the ProductID and total quantity of that product sold from the Sales.SalesOrderDetail table.
   Display only the products with an Order quantity (OrderQty) between 600 and 850 units.

2. Write a query that displays how many people with each Title there are in the Persons table.
   Sort the results according the number of people with the Title, in descending order.

Lesson 5 – Advanced filtering

Subjects:

- SELECT DISTINCT
- SELECT TOP N
- IN OPERATOR
- BETWEEN OPERATOR
- LIKE OPERATOR
- NULL VALUES

1. Write a query that displays the Business Entity ID, First Name, Middle Name, Last Name, and Modified Date from the Person.Person table.
   Display the data only for the people whose name ends with the letter "O" and for whom the Modified Date is not between the dates March 1, 2008 and Dec. 1, 2008.

2. Write a query that displays the Product number from the Sales.SalesOrderDetail table for the products with a total order quantity over all the years between 600 and 850 units.

Lesson 6 – Diagram and Join Tables

Subjects:

- Database Diagrams
- Join

1. What are the 3 product colors with the highest order amounts?

   Instructions: Write a query that shows the total order amount for each product color. Sort the results from highest to lowest, and display only the first three rows

   Think which tables contain the detailed data for the orders and the products. Use the ERD page for assistance.

2. Write a query that shows the 10 orders in 2013 (from the Order Header table) with the highest SubTotals, where the customer's last name contains the string 'lan' and the customer's first name does not contain the letter 'r'.

3. Check whether there are products in the Products table that were never sold. If so, display the products' codes and names.

   Instructions:
   Think what "never sold" means and how it is reflected in the data.
   Think about what type of JOIN is appropriate, and between which tables.

Lesson 7 – Union and Conditions

Subjects:

- Union
- Case when

1. In order to better arrange the store and provide a larger display space for items with a higher average order amount, analyze the average quantity of each product ordered during May 2012.

   To do this, write a query that answers the following:

   a. The query will display the following columns:

      i. Product ID

      ii. Total Order Quantity for that product

      iii. The number of order rows for the product (Count the number of rows in which the product was ordered.)

      iv. A verbal description of the average order amount, as follows:

         1. The average order amount < 3 → "Low quantity"

         2. The average order amount ≤ 6 → "Reasonable quantity"

         3. Above 6 → "High quantity"

   b. Sort the query results according to the average order amount in descending order.

   c. The order data is ProductID and OrderQty from the Sales.SalesOrderDetail table.

   d. Only the sales during May 2012 should be included. The date is calculated from the OrderDate field in the Sales.SalesOrderHeader table.

e. Following is a preview of the query results:

Results | Messages

	ProductID	TotalQty	NoOfSales	AvgQtyDescribe
1	863	357	20	High quantity
2	862	187	20	High quantity
3	852	202	32	High quantity
4	712	215	33	High quantity
5	715	214	32	High quantity
6	801	76	15	Reasonable quantity
7	789	70	14	Reasonable quantity
8	849	142	27	Reasonable quantity
9	843	107	21	Reasonable quantity

Lesson 8 – Built-in and Window Functions

Subjects:

- Built-in Aggregate Functions,
- Built-in Functions on Number Fields
- Built-in Functions on Strings Fields
- Built-in Functions on dates Fields
- Window Functions

1. Write a query based on the data in the Orders table that ranks the years according to the profits from all the orders during each year.

   Instructions:

   a. Decide which tables are involved in solving the query. Use the ERD for assistance. Hint: 3 tables.

   b. How is profit calculated?

   c. Is the profit calculated per item, or for all the items in the order record? (Pay attention to the OrderQty column.)

   d. Following is a preview of the query results:

Results | Messages

	OrderDate	TotalProfit	YearRank
1	2012	948201.191134	1
2	2011	1626845.842854	2
3	2013	3354362.929135	3
4	2014	3442493.663265	4

2. Continuing from the previous question: Are the ranking and comparison done annually analytically correct?
   (To answer this question, examine the order dates and data.)

Lesson 9 – Unrelated Nested Queries

Subjects:

- Unrelated nested queriesthat return a single value
- Unrelated nested queriesthat return a list (one column)
- Unrelated nested queriesthat returna table (several columns)

1. As an analyst working for AdventureWorks, you are required to check the order data for regular customers only, during 2013.

   The company defines a regular customer as a customer who already made a purchase from the store in previous years (i.e., bought in 2012 or 2011).

   For each order in 2013, display the order number, the SubTotal and the difference between the SubTotal of that order and the average for all the orders in 2013, based on the Sales.SalesOrderHeader table,.

   Note: Display the data for regular customers only.

   A preview of the results:

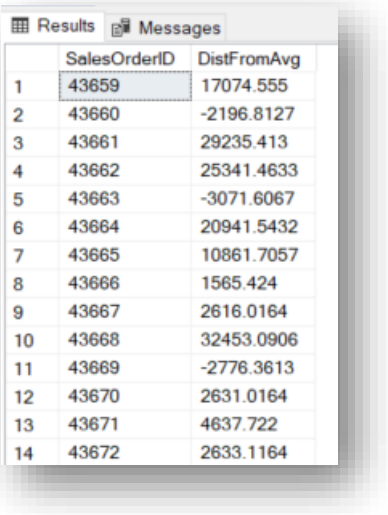
Results | Messages

	SalesOrderID	SubTotal	DistFromAvg
1	49443	67.539	-3008.3656
2	49444	6509.2516	3433.347
3	49445	2818.764	-257.1406
4	49446	13624.026	10548.1214
5	49447	107.988	-2967.9166
6	49448	1275.0469	-1800.8577
7	49449	45.588	-3030.3166
8	49450	32730.2215	29654.3169
9	49451	2595.6818	-480.2228
10	49452	1308.9375	-1766.9671
11	49453	1242.8518	-1833.0528
12	49454	6192.8825	3116.9779

2. Now calculate the same data for all the years and all the customers, similarly to the previous question.

   For each order in the Sales.SalesOrderHeader table, display the order number, the SubTotal and the difference between the SubTotal of that order and the average for all the orders.

   A preview of the results:



Results | Messages

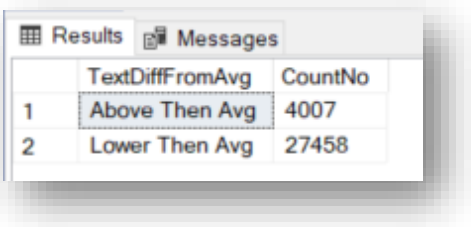
	SalesOrderID	DistFromAvg
1	43659	17074.555
2	43660	-2196.8127
3	43661	29235.413
4	43662	25341.4633
5	43663	-3071.6067
6	43664	20941.5432
7	43665	10861.7057
8	43666	1565.424
9	43667	2616.0164
10	43668	32453.0906
11	43669	-2776.3613
12	43670	2631.0164
13	43671	4637.722
14	43672	2633.1164

3. Continuing on from the previous question and based on the query you wrote, display the number of orders that are equal to and above the average, and the number of orders below the average.

   Hint: You will need to use the tools from previous lessons

   Instruction: Before you begin writing the query, think how you would solve it manually (i.e., if you got an order table with 10 rows). What calculations would you perform? How would you approach the solution? Once you have found a way to solve this manually, you can convert it into an SQL query.

   A preview of the results:



Results | Messages

	TextDifffFromAvg	CountNo
1	Above Then Avg	4007
2	Lower Then Avg	27458

Lesson 10 – Related Nested Queries

Subjects:

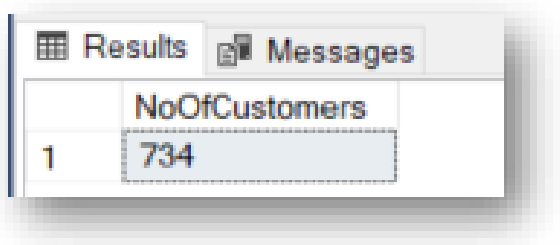
- Unrelated Nested Queries
- Exists

1. Write a query that displays the CustomerID, first name and last name of all the customers who ordered at least one product with the word "women" in its model name (Production.ProductModel table, Name column) in 2013. The word can appear at the beginning, middle or end.

   Instructions:
   Make sure you have included all the relevant tables in the query.
   Solve it using "exists".

2. Continuing from the previous question, use the query and the tools learned in previous lessons to answer how many customers appeared in question 1.

   A preview of the results:



Lesson 11 – CTE (Common Table Expressions)

Subjects:

- CTE (Common Table Expressions)

1. Write a query that shows the salespeople's (SalesPersonID) average annual order amount (from the OrderDate column).
   First process the data from the Order header table to get the total amount of orders generated by each salesperson in each year. Then calculate the average of the orders per salesperson for each year.

   Instructions (note all sections):

   a. Define a CTE that includes the SalesPersonID, year and total orders generated for each year (OrderDate) and salesperson. Calculate only for order records that have a SalesPersonID.

      A preview of the CTE result:

Results | Messages

	SalesPersonID	SaleYear	TotalYearOrders
1	281	2014	777941.6519
2	279	2011	1521289.1881
3	285	2013	151257.1152
4	289	2012	3014278.0472
5	275	2012	3375456.8947
6	286	2014	585755.8006

   b. Use the CTE you already defined to display the average amount of sellers' orders for each year.
      A preview of the query result:

Results | Messages

	SaleYear	Yearly Average Orders Per SalesPerson
1	2011	877855.1999
2	2012	1938121.527
3	2013	1934726.5719
4	2014	687358.7694

Lesson 12 – Review and Summary Practice on Data Retrieval

Subjects:

- All subjects learned so far

Please repeat the material studied, and make sure that you master all the topics studied so far.

If there is a question, please contact the unit lecturer.