QUALITY CHECK FORM

OFFUTURE

THE FUTURE OF OFFICE SUPPLIES

Team Having

OFFUTURE

The future of office supplies

OFFUTURE

PROJECT SUMMARY

Team Name	HAVING
Team Members	Daniel, Sado, Sum-Ye, Yassim
Client Name	Offuture
Date of file received	22/04/2024
Name of file received	1 - Offuture 2011-14.csv
File Format received	csv
Size of file (KB) received	11,806.558
Encoding of the file	UTF-8
Recorded Number of Columns	24
Recorded Number of Rows	51290
Name of Schema of Destination Table	student
Name of Destination Table	team_having_offuture

STATUS SUMMARY

This report explores the data sales records between the years 2011 and 2014. The objective of this assessment is to verify that the data has been successfully transferred to the database, data types are consistent across the database and identify potential changes and points of improvement for data storage and tabulation.

The quality checks performed include: count of columns, rows and distinct rows, sum of columns and rows, min/max of columns, null value checks, string length, decimal number count, distinct categories and sub-categories, presence of special characters and 'eyeball check' of five random entries from the table.

AUTOMATED CHECKS - CROSS-CHECKING

TEST CASE	DESCRIPTION OF CHECK	RESULT IN SOURCE	RESULT IN DESTINATION	OUTCOME	NOTES
TA01.1	Order Date – date format	DD/MM/YYYY	DD/MM/YYYY	PASS	Annex a)
TA01.2	Ship Date – date format	DD/MM/YYYY	DD/MM/YYYY	PASS	Annex a)
QC01.1	Count of Rows	51,290	51,290	PASS	
QC01.2	Count of distinct Rows	51,290	51,290	PASS	
QC01.3	Count of Columns	24	24	PASS	
QC01.4	Sum of Columns	1,331,006,116.61916	1,331,006,116.61916	PASS	Annex b)
QC01.5	Sum of Rows	1,331,006,116.61916	1,331,006,116.61916	PASS	Annex b)
QC01.6	Count of NULLS	41,296	41,296	PASS	
QC02.1	Row ID NULLS	0	0	PASS	
QC02.2	Order ID NULLS	0	0	PASS	
QC02.3	Order Date NULLS	0	0	PASS	
QC02.4	Ship Date NULLS	0	0	PASS	

TEST CASE	DESCRIPTION OF CHECK	RESULT IN SOURCE	RESULT IN DESTINATION	OUTCOME	NOTES
QC02.5	Ship Mode NULLS	0	0	PASS	
QC02.6	Customer ID NULLS	0	0	PASS	
QC02.7	Customer Name NULLS	0	0	PASS	
QC02.8	Segment NULLS	0	0	PASS	
QC02.9	City NULLS	0	0	PASS	
QC02.10	State NULLS	0	0	PASS	
QC02.11	Country NULLS	0	0	PASS	
QC02.12	Postal Code NULLS	41,296	41,296	PASS	
QC02.13	Market NULLS	0	0	PASS	
QC02.14	Region NULLS	0	0	PASS	
QC02.15	Product ID NULLS	0	0	PASS	
QC02.16	Category NULLS	0	0	PASS	
QC02.17	Sub-Category NULLS	0	0	PASS	
QC02.18	Product Name NULLS	0	0	PASS	
QC02.19	Sales NULLS	0	0	PASS	
QC02.20	Quantity NULLS	0	0	PASS	
QC02.21	Discount NULLS	0	0	PASS	
QC02.22	Profit NULLS	0	0	PASS	
QC02.23	Shipping Cost NULLS	0	0	PASS	
QC02.24	Order Priority NULLS	0	0	PASS	
QC03.1	MIN Row ID	1	1	PASS	**Annex c)**
QC03.1	MAX Row ID	51,290	51,290	PASS	**Annex c)**
QC03.1	Sum Row ID	1,315,357,695	1,315,357,695	PASS	**Annex c)**
QC03.2	MIN Sales	0.444	0.444	PASS	**Annex c)**
QC03.2	MAX Sales	22,638.48	22,638.48	PASS	**Annex c)**
QC03.2	Sum Sales	12,642,501.90988	12,642,501.90988	PASS	**Annex c)**
QC03.3	MIN Quantity	1	1	PASS	**Annex c)**
QC03.3	MAX Quantity	14	14	PASS	**Annex c)**
QC03.3	Sum Quantity	178,312	178,312	PASS	**Annex c)**
QC03.4	MIN Discount	0	0	PASS	**Annex c)**
QC03.4	MAX Discount	0.85	0.85	PASS	**Annex c)**
QC03.4	Sum Discount	7,329.728	7,329.728	PASS	**Annex c)**
QC03.5	MIN Profit	-6.599.978	-6.599.978	PASS	**Annex c)**
QC03.5	MAX Profit	8,399.976	8,399.976	PASS	**Annex c)**
QC03.5	Sum Profit	1,467,457.29128	1,467,457.29128	PASS	**Annex c)**
QC03.6	MIN Shipping Cost	0	0	PASS	**Annex c)**
QC03.6	MAX Shipping Cost	933.57	933.57	PASS	**Annex c)**
QC03.6	Sum Shipping Cost	1,352,820.69	1,352,820.69	PASS	**Annex c)**

TEST CASE	DESCRIPTION OF CHECK	RESULT IN SOURCE	RESULT IN DESTINATION	OUTCOME	NOTES
QC03.7	MIN Order Date	2011-01-01	2011-01-01	PASS	Annex d)
QC03.7	MAX Order Date	2014-12-31	2014-12-31	PASS	Annex d)
QC03.8	MIN Ship Date	2011-01-03	2011-01-03	PASS	Annex d)
QC03.8	MAX Ship Date	2015-01-07	2015-01-07	PASS	Annex d)
QC04.1	MIN length Region	4	4	PASS	
QC04.1	MAX length Region	14	14	PASS	
QC04.2	MIN length Product Name	5	5	PASS	
QC04.2	MAX length Product Name	127	127	PASS	
QC04.3	Decimal count Sales	(0, 5)	(0, 5)	PASS	Annex e)
QC04.4	Decimal count Profit	(0, 5)	(0, 5)	PASS	Annex e)
QC05.1	Distinct Sub-Category	17	17	PASS	Annex f)
QC05.2	Distinct Order Priority	4	4	PASS	Annex f)
QC06.1	Special Character Customer Name	6	6	PASS	Annex g)
QC06.2	Special Character City	202	202	PASS	Annex g)
QC06.3	Special Character State	93	93	PASS	Annex g)
QC06.4	Special Character Country	0	0	PASS	Annex g)
MC	Manual Check: 1 / 1,346 6,222 / 7,812 / 40,000	Full match	Full match	PASS	

ANNEX

List of different notes and annotations of the different checks and finds during the quality check process:

Annex a)
To maintain the correct format date on the database, 'Order Date' and 'Ship Date' has to be imported as varchar. Once import was completed, the table was altered to ensure these columns had date type "Date" in the format DD/MM/YYYY.

Annex b)

Only columns with integer values were used for these checks. Columns/rows used for these checks include:
- Row ID
- Sales
- Quantity
- Discount
- Profit
- Shipping Cost

Annex c)

Only columns with integer values were used for checking minimum, maximum and sum. Columns used for these checks include:
- Row ID
- Sales
- Quantity
- Discount
- Profit
- Shipping Cost

Annex d)

For data type "Date" columns, such as 'Order Date' and 'Ship Date', only minimum and maximum values were checked.

Annex e)

For decimal value count, the objective was to ensure that no decimal numbers were lost during the import process. To do so, decimal places were calculated for 'Sales' and 'Profit' columns, listing decimal place count and sales and profit entries count as shown in table below.

'Sales' Entry Count	Decimal Place Count	'Profit' Entry Count	Decimal Place Count
461	5	461	5
2,184	4	8,661	4
14,217	3	12,951	3
27,329	2	22,824	2
6,156	1	5,060	1
943	0	1,333	0

Annex f)

To ensure no data was lost, a count and listing of all the distinct labels for 'Sub-Category' and 'Order Priority' columns was performed. List of labels for column 'Sub-Category':

- Accessories
- Appliances
- Art
- Blinders
- Bookcases
- Chairs
- Copiers
- Envelopes
- Fasteners
- Furnishings
- Labels
- Machines
- Paper
- Phones
- Storage
- Supplies
- Tables

List of labels for column 'Order Priority':

- Critical
- High
- Medium
- Low

Annex g)

Using the Regular Expression *'[^a-zA-Z0-9\s"\-\()]'*, we filtered through 'Customer Name', 'City', 'State' and 'Country columns, to ensure that data imported with special characters was correctly added to the data base. This also ensures that the encoding matches the SOURCE encoding.

Tables with listing and counts of the column labels with special characters can be found in the Offture_Quality_Check.sql file (list for 'City' and 'State' are too large to include.

Annex h)

We observe data inconsistency with country of Mongolia as it is listed as APAC, EMEA for the market and North/Asia in region and only the United States have postal codes.

As for the 'Region' labels (North, South, East and West), there is no distinction for these regions for EU or US market. This makes difficult to filter the data by region, as EU and US market data would be clustered together.

Annex i)

Distinct 'Customer Name' appear to have two 'Customer IDs'. Customer ID will always have Customer Name initials followed by '-'and some numbers i.e. AB-15.

CONCLUSION

With this form, we can confirm that all Data has been correctly transferred to the DBeaver data base. No data has been lost, and outside of changing the Data Type of 'Order Date' and 'Ship Date' columns to 'Date', no modifications were needed to amend the SOURCE Data.