



OFFUTURE
THE FUTURE OF OFFICE SUPPLIES

Quality Control Form

Team Name	HAVING
Team Members	Ben Sarsfield, Elif Varli, Fatima Gonzalez, Katie Kitchen, Rowan Jarvis
Client Name	OFFUTURE
Date of file received	17/01/2024
Name of file received	1 - Offuture 2011-2014
Size of file received	12.1 MB
File Format received	csv
Encoding of the file	WINDOWS-1251
Delimiter character	,
Escape character	\
Recorded Number of Columns	24
Recorded Number of Rows	51290
Name of Schema of Destination Table	student
Name of Destination Table	h_offuture
Name of Attached SQL quality control code file	team_having_quality_control.sql
Name of Attached SQL table creation file	team_having_table_creation.sql

ID of SQL or row ID of checked rows	Description of Check	Result in DESTINATION (same as SOURCE if not stated otherwise)	OUTCOME
H_01	Count of rows	51,290	PASS
H_02	Count of columns	24	PASS
H_03	Count of duplicates rows	0	PASS
H_04	Sum of rows sums	1,882,578,769	PASS
H_05	Sum of column sums	1,882,578,769	PASS
H_06 49 756	Date format Check	OK OK	

50384 25003 39201		OK OK OK	
H_07 43550 1088 9375 50 24623	Eyeball Check of 5 random rows	OK OK OK OK OK	PASS
H_08	Count of Nulls	NULL only present in column postal_code “Postal Code” 41296 NULL VALUES PRESENT	PASS
H_09	Spaces before or after strings (Varchar Columns)	Not present	PASS
7081 2694 1616 2920 51220	Special Characters (Varchar Columns)	Special Characters Match	PASS
H_10	Sum Min Max Avg row_id “Row ID” (int)	Sum -> 1,315,357,695 Min -> 1 Max -> 51,290 Avg -> 25,645.50	PASS
H_12	Min Max order_date “Order Date” (date)	Min -> 2011-01-01 Max -> 2014-12-31	PASS
H_12	Min Max ship_date “Ship Date” (date)	Min -> 2011-01-03 Max -> 2015-01-07	PASS
H_10	Sum Min Max Avg postal_code “Postal Code” (int)	Sum -> 551,572,652 Min -> 1,040 Max -> 99,301 Avg -> 55,190.38	PASS
H_10	Sum Min Max Avg sales “Sales” (float)	Sum -> 12,642,502 Min -> 0.444 Max -> 22,638 Avg -> 246.49	PASS
H_10	Sum Min Max Avg quantity	Sum -> 178,312 Min -> 1	PASS

	"Quantity" (int)	Max -> 14 Avg -> 3.48	
H_10	Sum Min Max Avg discount "Discount" (float)	Sum -> 7330 Min -> 0 Max -> 0.85 Avg -> 0.14	PASS
H_10	Sum Min Max Avg profit "Profit " (float)	Sum -> 1,467,457 Min -> -6,600 Max -> 8,400 Avg -> 28.61	PASS
H_10	Sum Min Max Avg shipping_cost "Shipping Cost" (float)	Sum -> 1,352,821 Min -> 0 Max -> 934 Avg -> 26.38	PASS
H_11	Min Length Max Length order_id "Order ID" (varchar)	Min Length -> 9 Max Length -> 15	PASS
H_11	Min Length Max Length ship_mode "Ship Mode" (varchar)	Min Length -> 8 Max Length -> 14	PASS
H_11	Min Length Max Length customer_id "Customer ID" (varchar)	Min Length -> 5 Max Length -> 8	PASS
H_11	Min Length Max Length customer_name "Customer Name" (varchar)	Min Length -> 7 Max Length -> 22	PASS
H_11	Min Length Max Length segment "Segment" (varchar)	Min Length -> 8 Max Length -> 11	PASS
H_11	Min Length Max Length city "City" (varchar)	Min Length -> 2 Max Length -> 35	PASS

H_11	Min Length Max Length state "State" (varchar)	Min Length -> 3 Max Length -> 36	PASS
H_11	Min Length Max Length country "Country" (varchar)	Min Length -> 4 Max Length -> 32	PASS
H_11	Min Length Max Length market "Market" (varchar)	Min Length -> 2 Max Length -> 6	PASS
H_11	Min Length Max Length region "Region" (varchar)	Min Length -> 4 Max Length -> 14	PASS
H_11	Min Length Max Length product_id "Product ID" (varchar)	Min Length -> 15 Max Length -> 16	PASS
H_11	Min Length Max Length category "Category" (varchar)	Min Length -> 9 Max Length -> 15	PASS
H_11	Min Length Max Length sub_category "Sub-Category" (varchar)	Min Length -> 3 Max Length -> 11	PASS
H_11	Min Length Max Length product_name "Product Name" (varchar)	Min Length -> 5 Max Length -> 127	PASS
H_11	Min Length Max Length order_priority "Order Priority" (varchar)	Min Length -> 3 Max Length -> 8	PASS

N.B. All sums, minimums and maximums greater than the value 1 are rounded to the nearest integer. Averages are rounded to two decimal places.

NOTES

- Some cases of special characters found in source data, assumed to be errors. For example:

City = 'Aew?l-li' where Row ID = 22377,
City = 'Cox's B?z?r' where Row ID = 28323,
State = 'T?nh C?n Th?' where Row ID = 28352,
State = 'Th? Đô Hà N?i' where Row ID = 28915

These have been replicated in our SQL table.

- Data type enum ('Low', 'Medium', 'High', 'Critical') created for column "Order Priority" (order_priority), to facilitate correct input of new data in future. As this data is historical, enum types have not been created for other relevant columns. Code to create this type can be replicated for future use by clients, using our **team_having_table_creation.sql** file.
- Eyeball check conducted using code to randomly select rows, to ensure fair testing. This can be seen in code H_07 in **team_having_quality_control.sql**.