

Project 0: Offuture Project Report - Group WHERE

Introduction:

- The Brief was to upload and quality check sales data from Offuture in postgresql, to obtain a set of data that could be used for further analytics.
- The data was loaded, quality checked and prepared, with the results highlighted in the report below.

Pre-checking coding:

- The data provider was requested to provide the encoding format of the data.
- The data provider returned with the answer of Windows 1251, which was changed from the original format of UTF-8, in the settings while importing the data.
- Please note, when running Gitbash command, "file -i 2_0_2 Offuture 2011-2014.csv", the encoding returns ISO-8859-1.

Check	SQL Database outcome	Original data value	Pass?
1	51290 (total rows)	51290 (total rows)	Pass
2	51290 (distinct rows)	51290 (distinct rows)	Pass
3	24 (columns)	24 (columns)	Pass
4	*Eyeball - MATCH (column names)	*Eyeball - MATCH (column names)	Pass
5	12642501.91 (sum of sales)	12642501.91 (sum of sales)	Pass
6	7329.728 (sum of discount)	7329.728 (sum of discount)	Pass
7	1352820.69 (sum of shipping cost)	1352820.69 (sum of shipping cost)	Pass
8	1467457.291 (sum of profit)	1467457.291 (sum of profit)	Pass
9	178312 (sum of quantity)	178312 (sum of quantity)	Pass
10	1315357695 (sum of row id)	1315357695 (sum of row id)	Pass
11	551572652 (sum of postal code)	551572652 (sum of postal code)	Pass
12	51290 (max row id)	51290 (max row id)	Pass
13	**Untruncated	**Untruncated	Pass
14	51290 (rows with correct data types)	51290 (rows with correct data types)	Pass

15	24 (columns with correct data type)	24 (columns with correct data type)	Pass
16	***return = 20	***return = 0	Fail
17	****return = 6	****return = 0	Fail
18	Translation - MATCH	Translation - MATCH	Pass
19	Excel - MATCH	Excel - MATCH	Pass
20	Excel - MATCH	Excel - MATCH	Pass
21	Excel - MATCH	Excel - MATCH	Pass
22	*****Eyeball - MATCH	*****Eyeball - MATCH	Pass
23	row_id16343 - MATCH	row_id16343 - MATCH	Pass
24	row_id11844 - MATCH	row_id11844 - MATCH	Pass
25	Excel download compare - Eyeball - MATCH	Excel download compare - Eyeball - MATCH	Pass
26	*****postal_code = null (full data null check)	*****Postal_code = null (full data null check)	Pass
27	No nulls (excluding postal code)	No nulls (excluding postal code)	Pass
28	41296 (nulls in postal code column)	41296 (nulls in postal code column)	Pass

*Column names in table visually matched with original data set in MS Excel

**Max length of each column was checked against each “varchar” column settings in SQL, and was found to be below the max column setting. Max lengths also correlated with the data set in MS Excel

***Number of errors were found in the translation, along with names with apostrophes and dashes, however the apostrophes and dashes matched the data set MS Excel

****SQL isolated six names with mistranslated characters which did not match the data set in MS Excel

*****Null values of postal codes returned, correlated to non-US addresses

*****Nulls found but were expected, and matched with original data set in MS Excel

Additional work: Single Customer View (SCV)

A Single Customer View (SCV) table was created across a single touchpoint, to build a consolidated view of customer purchase data. This could be emulated with other customers who have data stored across multiple touchpoints.

customer_id	customer_name	different_locations	product_categories	quantity	total_sales	profit
46996	Emily Burns	5	Office Supplies, Furniture, Technology	8	184.87	-166.82