# 

# **Quality Control Report**

|  |  |
| --- | --- |
| **Team Name** | Team1 SELECT |
| **Team Members** | Jake Wu, Ollie Hanlon, Riya Pathak, Sarah Jennings |
| **Client Name** | Offuture |
| **Date of file received** | 17-01-2024 |
| **Name of file received** | Offuture 2011-2014 |
| **File Format received** | csv |
| **Size of file (KB) received** | 11,807 KB |
| **Encoding of the file** | ‘Windows-1252’ |
| **Recorded Number of Columns** | 24 |
| **Recorded Number of Rows** | 51290 |
| **Name of Schema of Destination Table** | student |
| **Name of Destination Table** | group\_select |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID of SQL** | **Description of Check** | **Result in SOURCE** | **Result in DESTINATION** | **OUTCOME** |
| SQL01 | Total number of columns | 24 | 24 | **PASS** |
| SQL02 | Total number of rows | 51290 | 51290 | **PASS** |
| SQL03 | Check for duplicates and uniqueness of row\_id | 0 Duplicates  Row\_id unique | 0 Duplicates  Row\_id unique | **PASS** |
| SQL04 | Check the occurrence of special characters | All values consistent | All values consistent | **PASS** |
| SQL05 | Check for data truncation | No truncation, max length of respective columns:  Row ID: 5, Order ID, 15, Ship Mode: 14, Customer ID: 8, Customer name: 22, Segment: 11, City: 35, State: 36, Country: 32, Postal Code: 5, Market: 6, Region: 14, Product ID: 16, Category: 15, Sub-Category: 11, Product-Name: 127, Order Priority: 8 | No truncation, max length of respective columns:  Row ID: 5, Order ID, 15, Ship Mode: 14, Customer ID: 8, Customer name: 22, Segment: 11, City: 35, State: 36, Country: 32, Postal Code: 5, Market: 6, Region: 14, Product ID: 16, Category: 15, Sub-Category: 11, Product-Name: 127, Order Priority: 8 | **PASS** |
| SQL06 | Eyeball check for 5 random rows | All values consistent  ROW\_ID:  38956  38867  44477  3232  18557 | All values consistent | **PASS** |
| SQL07 | Check for the Null values | 0 - {Row ID}  0 - {Order ID}  0 - {Order Date}  0 - {Ship Date}  0 - {Ship Mode}  0 - {Customer ID}  0 - {Customer Name}  0 - {Segment}  0 - {City}  0 - {State}  0 - {Country}  41296 - {Postal Code}  0 - {Market}  0 - {Region}  0 - {Product ID}  0 - {Category}  0 - {Sub-Category}  0 - {Product Name}  0 - {Sales}  0 - {Quantity}  0 - {Discount}  0 - {Profit}  0 - {Shipping Cost}  0 - {Order Priority} | 0 - {Row ID}  0 - {Order ID}  0 - {Order Date}  0 - {Ship Date}  0 - {Ship Mode}  0 - {Customer ID}  0 - {Customer Name}  0 - {Segment}  0 - {City}  0 - {State}  0 - {Country}  41296 - {Postal Code}  0 - {Market}  0 - {Region}  0 - {Product ID}  0 - {Category}  0 - {Sub-Category}  0 - {Product Name}  0 - {Sales}  0 - {Quantity}  0 - {Discount}  0 - {Profit}  0 - {Shipping Cost}  0 - {Order Priority} | **PASS** |
| SQL08 | Sum of Column Sums | 1331006116.61916 | 1331006116.61916 | **PASS** |
| SQL09 | Sum of Row sums | 1331006116.61916 | 1331006116.61916 | **PASS** |
| SQL10 | Check the DATE format for date columns (Order Date, Ship Date)  row\_id: 38895, 21583, 4625, 42136, 50430 | row\_id – Order Date, Ship Date  38895 – 17-08-2011, 22-09-2011  21583 – 02-05-2011, 03-05-2011  4625 – 06-10-2014, 08-10-2014  42136 – 24-08-2012, 26-08-2012  50430 – 09-11-2013, 14-11-2013 | row\_id – Order Date, Ship Date  38895 – 17-08-2011, 22-09-2011  21583 – 02-05-2011, 03-05-2011  4625 – 06-10-2014, 08-10-2014  42136 – 24-08-2012, 26-08-2012  50430 – 09-11-2013, 14-11-2013 | **PASS** |
| SQL11 | SUM | MIN | MAX | {row\_id}  1315357695|1|51290  {Sales}  12642501.91|0.444|22638.48  {quantity}  178312|1|14  {discount}  7329.728|0|0.85  {profit}  1467457.29|-6599.98|8399.976  {shipping\_cost}  1352821|0|933.57 | {row\_id}  1315357695|1|51290  {Sales}  12642501.90988|0.444|22638.48  {quantity}  178312|1|14  {discount}  7329.728|0|0.85  {profit}  1467457.29128|-6599.978|8399.976  {shipping\_cost}  1352820.69|0|933.57 | **PASS** |

Notes:

* Some city names contain missing special characters regardless of encoding:
  + Aew?l-li – ROW IDs: 22375, 22376, 22377
  + Cox’s B?z?r ROW IDs: 28323, 28324
* Negative profits noted in Profit column
* Segment includes “Consumer”, “Corporate”, and “Home Office” there is a possibility that “Home Office” should come under “Consumer”.
* Region column is applied inconsistently. For instance, “north” may refer to north of the respective country but the column is also used to specify continent.
* Discount is not applied consistently for some cases. Row\_id 5529 states there should be a discount of size 0.4 where the sale value has been calculated with a 0.6 (i.e. 60%) discount.