GROUP BY Quality Control Form

| **Team Name** | GROUP BY |
| --- | --- |
| **Team Members** | Matthew Fricker, Phoebe Beggs, Charles Ikegwu, Gavin Oliver |
| **Client Name** | Offuture |
| **Date of file received** | 24/04/2024 |
| **Name of file received** | 1 - Offuture 2011-2014 |
| **File Format received** | .csv |
| **Size of file (KB) received** | 11,807 |
| **Encoding of the file** | UTF-8 |
| **Recorded Number of Columns** | 24 |
| **Recorded Number of Rows** | 51,290 |
| **Name of Schema of Destination Table** | student |
| **Name of Destination Table** | groupby\_offuture\_v2 |

| **Test performed** | **Excel** | | | | **SQL** | | | | **SQL code reference** | **PASS or FAIL** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Count of the rows** | 51,290 | | | | 51,290 | | | | Q02 | PASS |
| **Count of the columns** | 24 | | | | 24 | | | | Q03 | PASS |
| **Count of distinct rows** | 51,290 | | | | 51,290 | | | | Q04 | PASS |
| **Sum of row sums** | 1,882,578,768.62 | | | | 1,882,578,768.62 | | | | Q05 | PASS |
| **Sum of column sums** | 1,882,578,768.62 | | | | 1,882,578,768.62 | | | | Q06 | PASS |
| **Date check** | dd-mm-yyyy | | | | yyyy-mm-dd | | | | Q07 | PASS |
| **Eyeball check** | IDs checked were:  1026, 25389, 26312, 28568, 49085 | | | | IDs checked were:  1026, 25389, 26312, 28568, 49085 | | | | Q10 | PASS |
| **Count of nulls** | 41,296.00 | | | | 41,296.00 | | | | Q09 | PASS |

| **Test performed** | **Excel** | | | | **SQL** | | | | **SQL code reference** | **PASS or FAIL** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Numeric Check** | **Column Name** | **MIN** | **MAX** | **SUM** | **Column Name** | **MIN** | **MAX** | **SUM** | Q08 | PASS |
| **Row ID** | 1.00 | 51,290.00 | 1,315,357,695.00 | **Row ID** | 1.00 | 51,290.00 | 1,315,357,695.00 |
| **Order Date** | 2011-01-01 | 2014-12-31 | N/A | **Order Date** | 2011-01-01 | 2014-12-31 | N/A |
| **Ship Date** | 2011-01-03 | 2015-01-07 | N/A | **Ship Date** | 2011-01-03 | 2015-01-07 | N/A |
| **Postal Code** | 1,040.00 | 99,301.00 | 551,572,652.00 | **Postal Code** | 1,040.00 | 99,301.00 | 551,572,652.00 |
| **Sales** | 0.44 | 22,638.48 | 12,642,501.91 | **Sales** | 0.44 | 22,638.48 | 12,642,501.91 |
| **Quantity** | 1.00 | 14.00 | 178,312.00 | **Quantity** | 1.00 | 14.00 | 178,312.00 |
| **Discount** | 0.00 | 0.85 | 7,329.73 | **Discount** | 0.00 | 0.85 | 7,329.73 |
| **Profit** | -6,599.98 | 8,399.98 | 1,467,457.29 | **Profit** | -6,599.98 | 8,399.98 | 1,467,457.29 |
| **Shipping Cost** | 0.00 | 933.57 | 1,352,820.69 | **Shipping Cost** | 0.00 | 933.57 | 1,352,820.69 |
| **Decimal Truncation** | **Column Name** | **Total of character length in columns** | | | **Column Name** | **Total of character length in columns** | | | Q11 | PASS |
| **Sales** | 290389 | | | **Sales** | 290389 | | |
| **Discount** | 100438 | | | **Discount** | 100438 | | |
| **Profit** | 281851 | | | **Profit** | 281851 | | |
| **Shipping Cost** | 224677 | | | **Shipping Cost** | 224677 | | |

| **Test performed** | **Excel** | | | | **SQL** | | | | **SQL code reference** | **PASS or FAIL** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Character Count** | **Column** | **MIN** | **MAX** | **AVG** | **Column** | **MIN** | **MAX** | **AVG** | Q11 | PASS |
| **Order ID** | 9.00 | 15.00 | 13.57 | **Order ID** | 9.00 | 15.00 | 13.57 |
| **Ship Mode** | 8.00 | 14.00 | 12.84 | **Ship Mode** | 8.00 | 14.00 | 12.84 |
| **Customer ID** | 5.00 | 8.00 | 7.82 | **Customer ID** | 5.00 | 8.00 | 7.82 |
| **Customer Name** | 7.00 | 22.00 | 12.95 | **Customer Name** | 7.00 | 22.00 | 12.95 |
| **Segment** | 8.00 | 11.00 | 8.85 | **Segment** | 8.00 | 11.00 | 8.85 |
| **City** | 2.00 | 35.00 | 8.42 | **City** | 2.00 | 35.00 | 8.42 |
| **State** | 3.00 | 36.00 | 9.42 | **State** | 3.00 | 36.00 | 9.42 |
| **Country** | 4.00 | 32.00 | 8.84 | **Country** | 4.00 | 32.00 | 8.84 |
| **Market** | 2.00 | 6.00 | 3.62 | **Market** | 2.00 | 6.00 | 3.62 |
| **Region** | 4.00 | 14.00 | 6.64 | **Region** | 4.00 | 14.00 | 6.64 |
| **Product ID** | 15.00 | 16.00 | 15.20 | **Product ID** | 15.00 | 16.00 | 15.20 |
| **Category** | 9.00 | 15.00 | 12.86 | **Category** | 9.00 | 15.00 | 12.86 |
| **Sub-category** | 3.00 | 11.00 | 7.23 | **Sub-category** | 3.00 | 11.00 | 7.23 |
| **Product Name** | 5.00 | 127.00 | 30.86 | **Product Name** | 5.00 | 127.00 | 30.86 |
| **Order Priority** | 3.00 | 8.00 | 5.41 | **Order Priority** | 3.00 | 8.00 | 5.41 |

| **Appendix** | * Headings updated by Charles in the csv * Date Format check - initial import misinterpreted date formats, second version imports these dates as varchars and then casts them as dates * Date counted differently in excel and SQL, date therefore excluded from ‘Character Count’ |
| --- | --- |