

## SQL

### Data cleaning

- Duplicate the table to clean it up.

```
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean` AS  
SELECT *  
FROM `enduring-wharf-464607-c7.online_retail.sales_raw`
```

- Remove null values

```
5CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step1`  
AS  
SELECT *  
FROM `enduring-wharf-464607-c7.online_retail.sales_clean`  
WHERE CustomerID IS NOT NULL  
AND InvoiceNo IS NOT NULL  
AND Description IS NOT NULL;
```

- Remove negative or incorrect values. Table 2 is created to add a transaction type column based on whether the negative quantity is a return or a sale, and to remove values where the quantity is 0 and the price unit is negative.

```
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step2` AS  
SELECT *  
,  
CASE  
WHEN Quantity < 0 THEN 'Return'  
ELSE 'Sale'  
END AS Transaction_type  
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step1`  
WHERE  
Quantity != 0  
AND UnitPrice > 0;
```

- Remove duplicates and convert columns to strings to use the function.

```
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step3`  
AS  
SELECT *  
EXCEPT (row_num)  
FROM (  
SELECT *  
,  
ROW_NUMBER() OVER (PARTITION BY  
InvoiceNo,  
StockCode,  
Description,  
CAST(ROUND(Quantity, 0) AS INT64),  
InvoiceDate,  
CAST(CustomerID AS STRING),  
Country  
) AS row_num  
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step2`)  
WHERE row_num = 1;
```

- Clean up text. Capital letters and spaces.

```
6CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step4_clean_text` AS  
SELECT  
InvoiceNo,  
StockCode,  
TRIM(UPPER(Description)) AS Description,  
Quantity,  
InvoiceDate,  
UnitPrice,  
CustomerID,  
TRIM(UPPER(Country)) AS Country,  
Transaction_type  
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step3`
```

- Convert Customerid because it was in another format, with CAST.

```
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step5` AS  
SELECT *  
, CAST(CustomerID AS INT64) AS customerID_  
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step4_clean_text`  
- Create new columns for analysis  
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_step7`
```

```
AS
SELECT *
,
Quantity * UnitPrice AS Total,
EXTRACT(YEAR FROM InvoiceDate) AS Year,
EXTRACT(MONTH FROM InvoiceDate) AS Month,
EXTRACT(DAYOFWEEK FROM InvoiceDate) AS Day_of_week,
EXTRACT(HOUR FROM InvoiceDate) AS Hour
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step6`
- Rename table: sales_clean_final
CREATE TABLE `enduring-wharf-464607-c7.online_retail.sales_clean_final`
AS
SELECT *
FROM `enduring-wharf-464607-c7.online_retail.sales_clean_step7`
```