# Marketing Analytics Project

## Data Quality Report

**Overview**: This report documents the steps taken to clean and validate the raw marketing dataset before performing exploratory analysis, SQL queries, dashboarding, and machine learning.

### Steps Taken:

**1. Missing Values**

* Checked all columns for NULL, blank, or invalid values.
* **No missing values** were found in critical columns such as CustomerID, TotalSpent, and NumberOfOrders.
* Ensured no null values exist in categorical columns like Country, AcquisitionChannel, Gender.

**2. Outlier Detection**

Used the **IQR method** to identify extreme values in numeric fields:

* **TotalSpent**: Flagged customers with unusually high or low spending.
* **NumberOfOrders**: Identified customers with very high order counts.

Created two outlier flag columns:

* TotalSpent\_Outlier\_Flag
* OrderCount\_Outlier\_Flag

**3. Logical Validations (Click Rates)**

Ensured click data followed business logic:

* EmailClicks ≤ EmailOpens
* AdClicks ≤ AdImpressions  
  ❌ Invalid records were flagged for review or correction.

**4. Date Formatting**

* Original RegistrationDate values were in inconsistent or broken formats (e.g., 15:20.8).
* All dates were standardized to format: **YYYY-MM-DD HH:MM:SS**  
  ✔️ This allowed for accurate sorting, time-based analysis, and trend identification.

**5. Numeric Precision**

To maintain consistency and readability:

* Rounded TotalSpent and AvgOrderValue to **2 decimal places**
* Rounded EmailClickRate and AdClickRate to **4 decimal places**  
  📈 This ensured cleaner pivot tables and visualizations.

**6. Boolean Flags**

* Verified Churned column had only binary values:
  + 1 → Churned
  + 0 → Not Churned
* Re-coded any inconsistent or text-based churn values.

**Summary of Fixes**

|  |  |
| --- | --- |
| Issue | Action Taken |
| Missing values | None found |
| Outliers | Detected and flagged |
| Invalid click logic | Checked and corrected |
| Date format issues | Cleaned and converted |
| Inconsistent decimals | Rounded to 2 or 4 places as needed |
| Boolean column errors | Standardized churned to 0 and 1 |