

Customer Sign-Up Behaviour & Data Quality Audit

1. Introduction

This project aims to audit the quality of the customer sign-up data and derive insights into user acquisition trends for Rapid Scale, a SaaS company offering tiered subscription plans. The dataset contains sign-up information such as customer ID, source of acquisition, region, plan selected, marketing preferences, and demographics.

The Business Intelligence team uses these insights to support monthly strategy meetings, and the goal is to help marketing and onboarding teams enhance user engagement campaigns.

2. Data Cleaning Summary

A thorough cleaning was conducted to prepare the dataset for analysis:

- Duplicates Removed: 1 duplicate based on customer_id was removed.
- Standardisation: Text fields like plan_selected, gender, and source were standardized using title casing and trimming spaces.
- Missing Values: Placeholder values ('??', 'NaN') were replaced, and numeric conversions handled 'unknown' or blank values.
- Age was filled with the median, and categorical blanks in region and gender were labeled as 'Unknown' or 'Not Specified'.
- Signup date was converted to datetime format.

3. Key Findings & Trends

- Weekly sign-ups show stable growth with peaks during certain weeks.
- The 'Pro' plan is most selected, indicating strong interest in mid-tier features.
- Male users are more likely to opt in for marketing. Users aged 25-50 favor premium or pro plans.

4. Business Question Answers

1. Google brought in the most users last month.
2. The 'Unknown' region suggests incomplete regional data (~10%).
3. Older users are less likely to opt into marketing compared to younger users.
4. The 'Pro' plan is most popular, especially among 25-35 and 36-50 age groups.

5. Recommendations

- Ensure 'region' is a required field to reduce missing data.
- Focus campaigns on users under 35 who are more engaged.
- Design better onboarding flows for users over 50 to increase their opt-in and plan engagement.

Customer Sign-Up Behaviour & Data Quality Audit

6. Data Issues or Risks

Missing values in age, region, and gender fields may distort analysis. Input validation should be enforced at data entry forms to improve data reliability.