Rapid Scale Customer Sign-Up Data Quality Audit & Insights

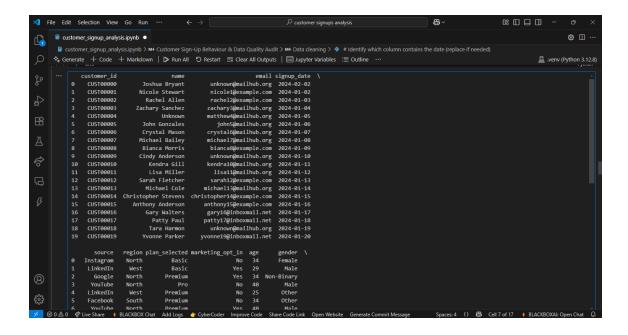
1. Introduction

This report presents a data quality audit and user acquisition analysis for Rapid Scale, a fast-growing SaaS company. The dataset reviewed contains customer sign-up records, including customer demographics, subscription plans, marketing opt-in status, and acquisition sources. The objective is to assess data accuracy, identify trends, and provide actionable recommendations to support the Monthly Business Review (MBR) and guide marketing and onboarding strategies.

2. Data Cleaning Summary

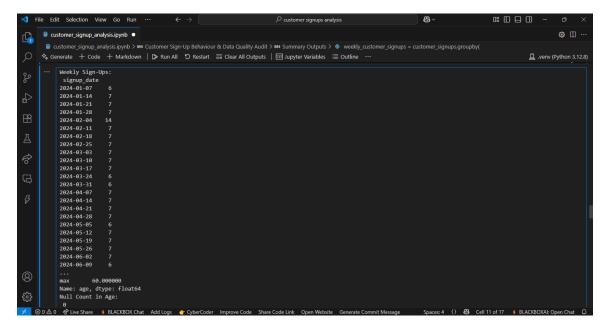
The dataset required extensive cleaning to ensure reliability. Key actions included:

- a. Changed the data types of all respective columns
- b. Removed duplicate customer IDs.
- c. Handled missing values:
 - Customer IDs: Replaced blanks with sequential IDs.
 - Names: Filled missing entries with 'Unknown' and standardized capitalization.
 - Emails: Standardized format, converted to lowercase, replaced missing/invalid with 'unknown@mailhub.org'.
 - Regions: Replaced blanks with 'North' because it was the modest.
 - Source: Trimmed and replaced blanks and ?? with 'YouTube'
 - Signup dates: Standardized multiple date formats using 'regex', filled missing with most common date.
 - Ages: Converted to numeric, replaced outliers (206) and missing with median, converted to integers.
 - Marketing opt-ins: Replaced Nil and None with "No" and blanks with "Yes".
 - Gender: Standardized categories, replaced blanks with 'Male' because it was modest and invalid codes, that is 123 with Female.
 - Plan selected: Corrected inconsistent labels, that is Prem to Premium, basic to Basic, PRO to Pro and assigned Basic and Pro randomly for blanks at 50%/50%



3. Key Findings & Trends

- Weekly sign-up trends indicate fluctuating acquisition activity, with identifiable peaks. This date, "2024-02-04" had the highest sign ups of 14 customers.
- YouTube emerged as the top acquisition source.
- Premium is the most selected marketing plan.
- Marketing opt-in rates are higher among 36-46 age group and opted for Yes.
- The 'North' region shows the highest sign-ups, partly due to missing-region imputation.



4. Business Question Answers

- Data Accuracy: Issues included inconsistent date formats, blank regions, and invalid categorical values.
- User Acquisition: Most users sign up via YouTube, with Premium being the most popular plan.
- Marketing Opt-In Behavior: Higher opt-in rates occur among older age groups (36-60)

5. Recommendations

- Enforce stricter data validation at the point of entry to prevent blanks and invalid codes.
- Focus marketing efforts on strong-performing acquisition channels like YouTube.
- Develop targeted campaigns for younger age groups to improve marketing opt-in rates.

6. Data Issues or Risks

One significant data quality issue is the lack of complete regional data. Replacing blanks with 'North' may distort regional performance metrics. Future data collection processes should require mandatory region input, validated against a predefined list.