Customer Sign-Up Behavior and Data Quality Audit

1. Introduction

This analysis examines customer sign-up behaviour and conducts a data quality audit for Rapid Scale, a fast-growing SaaS company. The dataset contains 300 customer records with information about sign-up sources, demographics, plan selections, and marketing preferences. This report provides actionable insights to optimize marketing campaigns and user engagement workflows for the Monthly Business Review.

2. Data Cleaning Summary

Data Quality Issues Identified:

- **Missing values**: Significant gaps across all columns, with email (11.3%) and region (10%) showing the highest percentages
- **Duplicate records**: 1 duplicate customer record removed based on customer_id
- **Inconsistent formatting**: Multiple variations in plan_selected (basic/Basic/PREMIUM/PRO) and gender (male/Male/FEMALE/123)
- **Data type issues**: signup_date stored as text, age as strings instead of numeric

Cleaning Actions Taken:

- Removed 1 duplicate record, reducing dataset from 300 to 299 records
- Standardized plan_selected values (basic→Basic, PRO→Pro, PREMIUM→Premium)
- Standardized gender values (male→Male, FEMALE→Female, removed invalid entries like '123')
- Converted signup_date to datetime format and age to numeric
- Replaced inconsistent missing value indicators ('??', 'Nil', '123') with 'Unknown' or appropriate defaults
- Imputed missing marketing_opt_in values as 'No' (conservative approach)
- Replaced one outlier age value (206) with median age (34)

3. Key Findings & Trends

Customer Acquisition Patterns:

- YouTube emerges as the top acquisition source (58 signups), followed by Google (50) and Referral (49)
- Premium plan shows highest adoption (99 customers), indicating strong value proposition for higher-tier offerings
- Young adults (20-35 years) represent the largest customer segment (186 customers, 62% of total)

Marketing Engagement:

- Overall marketing opt-in rate is 44% (132 of 299 customers)
- Gender distribution shows balanced representation across Male (92), Female (92), and Other (59) categories
- Marketing opt-in rates vary by age group, with middle-aged customers showing highest engagement

Geographic Distribution:

- North region leads in signups (65), followed by East (61) and South (59)
- 30 customers have unknown/missing region data, indicating potential data collection issues

Support Ticket Resolution Performance:

- Strong overall resolution rate with 95 tickets resolved out of 123 total tickets (77% resolution rate)
- Premium plan users show excellent resolution performance (26 tickets with 23 resolved - 88% resolution rate)
- West region demonstrates exceptional support performance with 27 tickets and 23 resolved (85% resolution rate)

4. Business Question Answers

1. Which acquisition source brought in the most users last month? Google brought in the most users in the last month (October 2024) with 7 signups, making it the top-performing channel for recent acquisitions.

2. Which region shows signs of missing or incomplete data? North region shows the most concerning data quality issues, with 10 missing email addresses and 4 missing source attributions. This suggests potential problems with data collection processes in this region.

3. Are older users more or less likely to opt in to marketing? Older users (51+) show a lower marketing opt-in rate at 44% (18 of 41 customers) compared to middle-aged users at 49% (35 of 72 customers). Older users are less likely to opt in to marketing.

4. Which plan is most commonly selected, and by which age group? Premium plan is most commonly selected overall (99 customers). Young adults (20-35 years) represent the largest adopters of Premium plans (57 customers), followed by middle-aged customers (32 customers).

5. Which plan's users are most likely to contact support? Pro plan users generate the most support tickets (47 tickets), followed by Basic plan users (42 tickets). This suggests Pro plan users may need additional onboarding support, or the plan complexity requires more assistance.

5. Recommendations

Marketing Optimization:

- Focus acquisition efforts on YouTube and Google channels, which show strongest performance
- Develop targeted campaigns for middle-aged demographics who show higher marketing opt-in rates
- Investigate why Premium plan appeals most to young adults and leverage these insights for other segments

Data Quality Improvements:

- Implement mandatory email validation for North region signups to reduce missing data
- Establish consistent data collection standards across all regions
- Add data validation rules to prevent inconsistent categorical entries

Customer Success:

- Develop enhanced onboarding materials for Pro plan users to reduce support ticket volume
- Create region-specific engagement strategies, particularly for the high-performing North region
- Investigate and replicate the high resolution rate practices in West region across other regions

6. Data Issues or Risks

Critical Data Quality Risk: The most significant risk is the inconsistent regional data collection, with 10% of customers having unknown regions. This impacts the ability to:

- Allocate regional marketing budgets effectively
- Understand geographic performance patterns
- Comply with regional data protection requirements

Support Resolution Concerns: While overall resolution rates are strong at 77%, some areas need attention:

- East region Premium users show lower resolution rates (1 ticket, 0 resolved)
- Pro plan users in Central region have moderate resolution performance (10 tickets, 7 resolved - 70%)

Recommended Fix: Implement mandatory region selection during signup process and conduct a data enrichment project to backfill missing region data using IP geolocation or other available customer information.

Additional Considerations:

- 40 customers contacted support within 2 weeks of signup, indicating potential onboarding friction
- The support ticket analysis reveals regional variations in support needs, with Basic plan users in South region generating the most tickets (14), suggesting need for targeted support resources