Full Name	Zahra Hayati
Project Title	Week 1: Data Analysis for Business Insights

1. Introduction

This report analyses Rapid Scale's customer sign-up data to support the Monthly Business Review (MBR). The primary goal is to assess data quality and identify key user acquisition trends. Based on 300 customer records and a secondary file of support tickets, the analysis provides insights to help Marketing optimize campaigns and Onboarding improve user engagement by answering five key business questions.

2. Data Cleaning Summary

The dataset contained multiple issues typical of raw sign-up data. Key cleaning steps included:

1. Data Type Standardisation
   - Converted signup_date to datetime
   - Standardised text values for plan_selected, gender, source, and region
   - Lowercased and stripped emails
2. Missing Values
   - Missing region values were set to 'Unknown'
   - Missing ages were imputed with the median age
   - Created email_missing and region_missing flags
3. Duplicate Removal
   - The only duplicate customer_id is an empty string which generated new IDs for customers with missing customer_id.
4. Age Cleaning
   - Converted non-numeric ages (e.g., 'thirty') to integers
   - Flagged unusual ages outside 16–100 as UNUSUAL
5. Outcome
   - Significant reduction in missing values and inconsistencies
   - Dataset now has clean, standardised customer information ready for analysis

Table 1: Missing Data Before vs After Cleaning

| Title	Missing (Raw)	% Missing (Raw)	Missing (Cleaned)	% Missing (Cleaned)	Change (Count)	Change (%)	Trend
email	34	11.33	34	11.33	0	0	No Change ⚪
region	30	10	0	0	30	10	Decreased 🟢
age	12	4	18	6	-6	-2	Increased 🔴
marketing_opt_in	10	3.33	11	3.67	-1	-0.34	Increased 🔴
name	9	3	9	3	0	0	No Change ⚪
source	9	3	9	3	0	0	No Change ⚪
plan_selected	8	2.67	8	2.67	0	0	No Change ⚪
gender	8	2.67	14	4.67	-6	-2	Increased 🔴
customer_id	2	0.67	0	0	2	0.67	Decreased 🟢
signup_date	2	0.67	6	2	-4	-1.33	Increased 🔴

3. Key Findings & Trends

- Weekly Sign-ups: Customer acquisition activity peaks in early February, showing strong seasonal engagement likely tied to marketing campaigns.
- Top Acquisition Sources: the top acquisition channel is YouTube, which dominates the highest signup counts primarily for the Basic plan in the South and East regions.
- Regional Distribution: The North and East regions lead in customer sign-ups. However, a notable portion of users 10% are listed under "Unknown", suggesting incomplete geographic data that may require validation or better collection at sign-up.
- Plan Popularity: The Premium plan dominates user selection, followed by Pro and Basic. This pattern indicates strong interest in higher-value offerings, possibly due to effective upsell marketing or product feature appeal.

Figure 1: Customer sign-ups by acquisition source
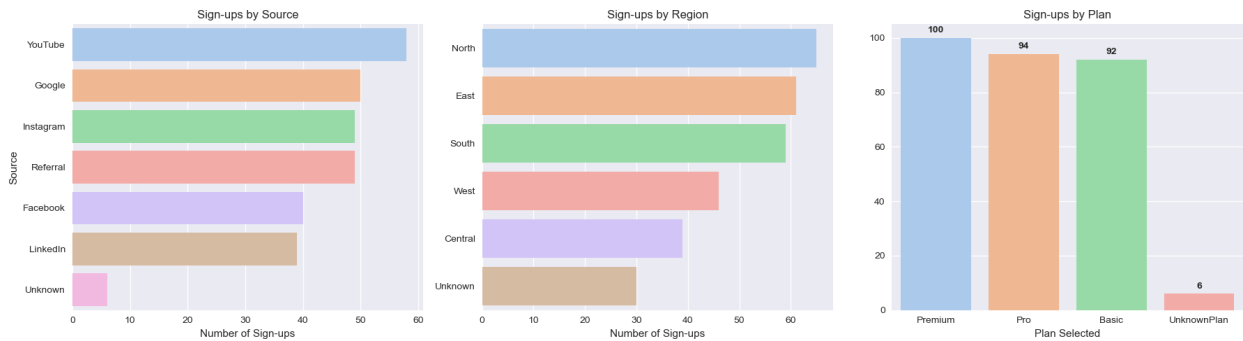
- Age statistics: Min 21, Max 60, Mean 35.4, Median 34, Null count 18.
- Support Tickets: The Pro and Basic plans account for the vast majority of support tickets, totalling nearly 75% of all issues (Pro: 47, Basic: 42). Critically, 54.6% of support tickets occurred within the first 2 weeks of sign-up, highlighting a significant challenge in the early onboarding experience.

Figure 2: Support Ticket Analysis by Plan and Region
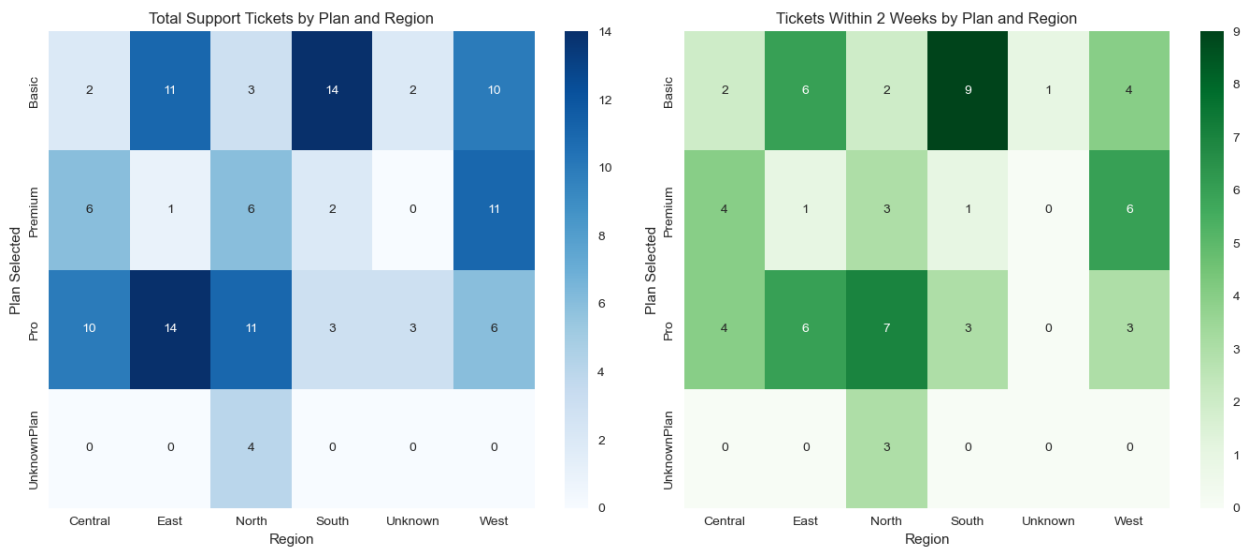
4. Business Question Answers

1. **Which acquisition source brought in the most users last month?**
   Google was the leading acquisition source last month, bringing in the highest number of new users (3), followed by Instagram (2). Other channels such as Referral, Facebook, and LinkedIn contributed smaller but measurable volumes.
   This suggests Google remains the most effective acquisition channel, and future campaigns could prioritize optimization of Google Ads or SEO to sustain growth.

2. **Which region shows signs of missing or incomplete data?**
   The North and West regions show the highest rates of missing or incomplete data, with approximately 43% and 41% of records containing gaps respectively. This suggests potential data entry or tracking issues in these areas. In contrast, the South and Central regions have relatively cleaner data, while the "Unknown" category indicates entries where the region was not recorded at all—highlighting a need for better validation at the data collection stage.

3. **Are older users more or less likely to opt in to marketing?**
   Older users are slightly less likely to opt in to marketing compared to younger users. Marketing opt-in rates peak around the 36–55 age range (≈49%), while younger users aged 16–25 and older users aged 56–65 show lower engagement (~43%).

This pattern suggests that mid-career adults are the most receptive to marketing communications, whereas both the youngest and oldest groups may be less responsive—potentially due to differing digital habits or privacy concerns.

4. **Which plan is most commonly selected, and by which age group?**
The Premium plan is the most commonly selected overall, closely followed by the Pro plan. The 26–35 age group shows the highest engagement across all plans, particularly for Premium (35 users) and Pro (36 users), indicating this demographic is the most active and willing to invest in higher-tier services. In contrast, adoption declines among users above 55, suggesting that older users prefer lower-cost options or are less likely to subscribe altogether.

5. **Which plan's users are most likely to contact support?**
Users on the Pro plan are the most likely to contact support, with 24 support tickets recorded, followed by Basic plan users with 20 tickets. Premium plan users contacted support less frequently (12 tickets), suggesting they may experience fewer issues or receive better onboarding assistance. This trend indicates that mid-tier users (Pro) may require additional support resources or clearer guidance during setup and usage.

5. Recommendations

Based on the analysis, the following strategic actions are recommended:

Acquisition & Marketing
- **Optimize Top Channels:** Focus budget on YouTube (highest overall sign-ups) and Google (top last month). Target YouTube ads promoting the Basic plan in the South/East.
- **Target Marketing Opt-in:** Direct lead-nurturing efforts toward the receptive 36–55 age group (approx. 49% opt-in) and develop tailored content for lower-opting younger and older users.
- **Capitalize on Plan Interest:** Continue to market the benefits of the Premium plan, particularly to the highly engaged 26–35 age group.

Onboarding & Support
- **Address Early Friction:** Immediately enhance the initial user experience to address the 54.6% of tickets occurring within the first two weeks of sign-up.
- **Improve Pro Plan Support:** Allocate additional resources or specific documentation to the Pro plan users, who are the most likely to contact support.

6. Data Issues or Risks

The integrity of future analysis is currently threatened by two primary data quality issues:

1. Missing Geographic Data (Regional Bias):
   - 10% of users are in the "Unknown" region, and data completion is poor in the North (43% missing fields) and West (41% missing fields).
   - Risk: This lack of reliable regional data biases market analysis and prevents effective geographic campaign targeting.
2. Inconsistent Sign-up Data:
   - Missing value counts for age, gender, and signup_date increased after cleaning. This occurred because many initially "filled" values were invalid placeholders (e.g., 'thirty' instead of 30) that had to be correctly converted to missing values.
   - Risk: The source of this poor data quality (e.g., faulty form fields) must be identified and corrected to prevent misleading statistics in future reports.

7. Conclusion

The analysis successfully identified key trends and data quality issues. The 26–35 age group is our most engaged demographic, with YouTube and Google being the top acquisition drivers. The most critical operational finding is that over half of all support tickets occur within the first two weeks (primarily from Pro and Basic users), pointing to a poor early onboarding experience. Future efforts must focus on redesigning onboarding to reduce support load and improving sign-up data validation to eliminate incomplete regional records.