Project Coversheet

Full Name	**Charu Madaan**
Email	**charumadaan88@gmail.com**
Contact Number	**07404932811**
Date of Submission	**24-09-2025**
Project Week	**Week - 1**

Project Guidelines and Rules

1. Submission Format

- **Document Style**:

  - Use a clean, readable font such as *Arial* or *Times New Roman*, size 12.
  - Set line spacing to **1.5** for readability.

- **File Naming**:

  - Use the following naming format:
    Week X – [Project Title] – [Your Full Name Used During Registration]
    Example: Week 1 – Customer Sign-Up Behaviour – Mark Robb

- **File Types**:

  - Submit your report as a **PDF**.
  - If your project includes code or analysis, attach the **.ipynb notebook** as well.

2. Writing Requirements

- Use formal, professional language.
- Structure your content using headings, bullet points, or numbered lists.

3. Content Expectations

- Answer **all** parts of each question or task.

- Reference tools, frameworks, or ideas covered in the programme and case studies.
- Support your points with practical or real-world examples where relevant.
- Go beyond surface-level responses. Analyse problems, evaluate solutions, and demonstrate depth of understanding.

4. Academic Integrity & Referencing

- All submissions must be your own. Plagiarism is strictly prohibited.
- If you refer to any external materials (e.g., articles, studies, books), cite them using a consistent referencing style such as APA or MLA.
- Include a references section at the end where necessary.

5. Evaluation Criteria

Your work will be evaluated on the following:

- Clarity: Are your answers well-organised and easy to understand?
- Completeness: Have you answered all parts of the task?
- Creativity: Have you demonstrated original thinking and thoughtful examples?
- Application: Have you effectively used programme concepts and tools?
- Professionalism: Is your presentation, language, and formatting appropriate?

6. Deadlines and Extensions

- Submit your work by the stated deadline.
- If you are unable to meet a deadline due to genuine circumstances (e.g., illness or emergency), request an extension **before the deadline** by emailing: **support@uptrail.co.uk**
Include your full name, week number, and reason for extension.

7. Technical Support

- If you face technical issues with submission or file access, contact our support team promptly at **support@uptrail.co.uk**.

8. Completion and Certification

- Certificate of Completion will be awarded to participants who submit at least two projects.
- Certificate of Excellence will be awarded to those who:
    - Submit all four weekly projects, and
    - Meet the required standard and quality in each.
- If any project does not meet expectations, you may be asked to revise and resubmit it before receiving your certificate.

Customer Sign-Up Behaviour & Data Quality Audit -
Prepared by Charu Madaan

Rapid Scale SaaS Platform
Business Intelligence Report
Report Period: January - October 2024

---

1. INTRODUCTION :

This report examines customer sign-up behaviour across 300 records from Rapid Scale's SaaS platform over a 10-month period. The objectives were twofold:

- **Data Quality Audit** – Identify and resolve missing, inconsistent, or unrealistic values to restore confidence in insights.

- **Acquisition & Behaviour Analysis** – Understand how customers are signing up, the subscription plans they are opting for, and their engagement with marketing opt-ins.

  The findings will be used to enhance campaign targeting, improve onboarding workflows, and strengthen overall customer acquisition strategies, ensuring Rapid Scale continues to scale effectively in a competitive SaaS market.

---

2. DATA CLEANING SUMMARY :

The original dataset revealed common challenges: missing values, duplicates, inconsistent text, and even unrealistic entries. These issues, if left unresolved, risk misleading decisions.

Key Cleaning Actions:

2.1. Handling Missing Information

- Region (30 records), email (34), and gender/name/source gaps were replaced with "Unknown".

- Age: 19 missing values filled with the median age (34).

2.2. <u>Duplicate Records Removed</u> : 1 duplicate customer record removed → final dataset = 299 unique customers.

2.3. <u>Standardization</u>:

- Subscription plans were aligned to: *Pro, Premium, Basic*.

- Gender entries corrected for format and errors (e.g., "123").

- Source entries like "??" updated to *"Unknown Source"*.

- Signup dates converted into a proper date format for accurate analysis.

2.4. <u>Outliers Adjustment</u> :
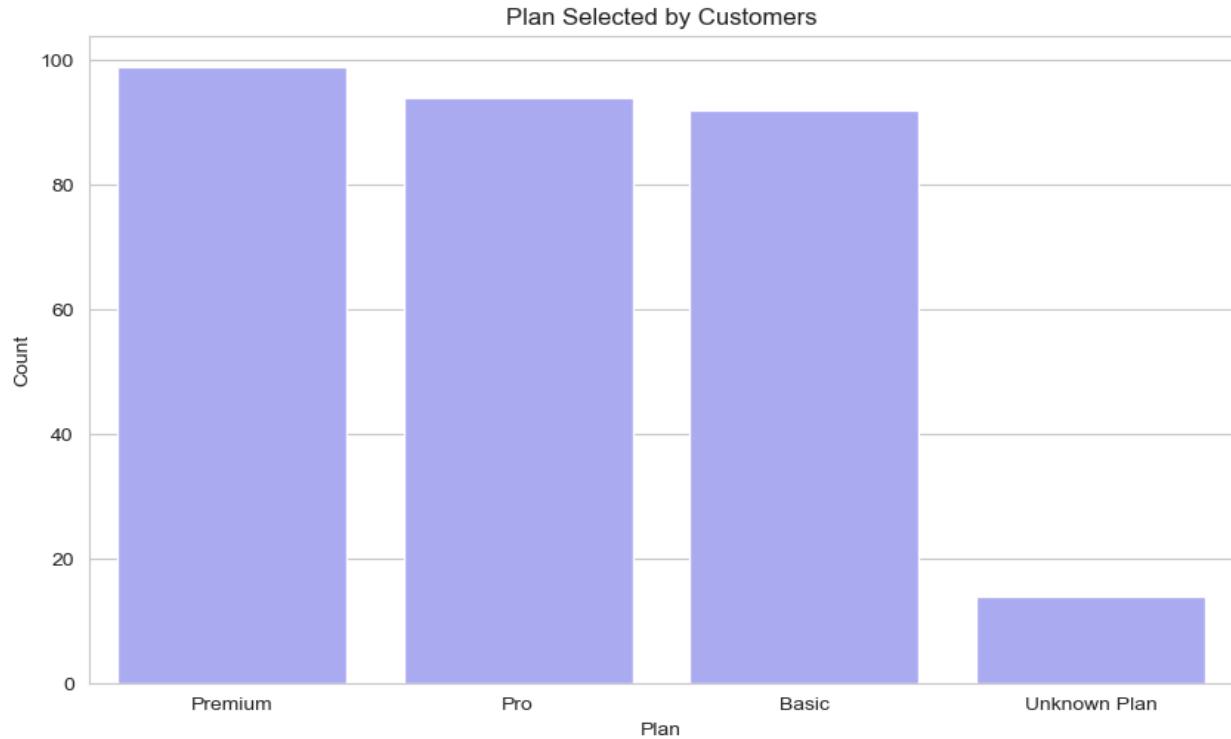
- An unrealistic age entry of **206 years** was identified.
- Any ages outside **15–100 years** were adjusted to the median age.
- The final dataset reflects a realistic customer age range of **21–60 years**.

Result: A clean, standardized dataset enabling trustworthy analysis.

---

3. KEY FINDINGS AND TRENDS :

A. **<u>Premium Plan Drives Revenue Growth</u> :**

Insight: Premium plan is our clear winner with **99 customers (33.1%)**, closely followed by Pro (94) and Basic (92). This nearly equal distribution across our three main tiers, with Premium leading, indicates strong value perception and customers' willingness to invest in advanced features. The fact that our most expensive plan has the highest adoption suggests effective upselling and strong product-market fit.

Plan Selected by Customers
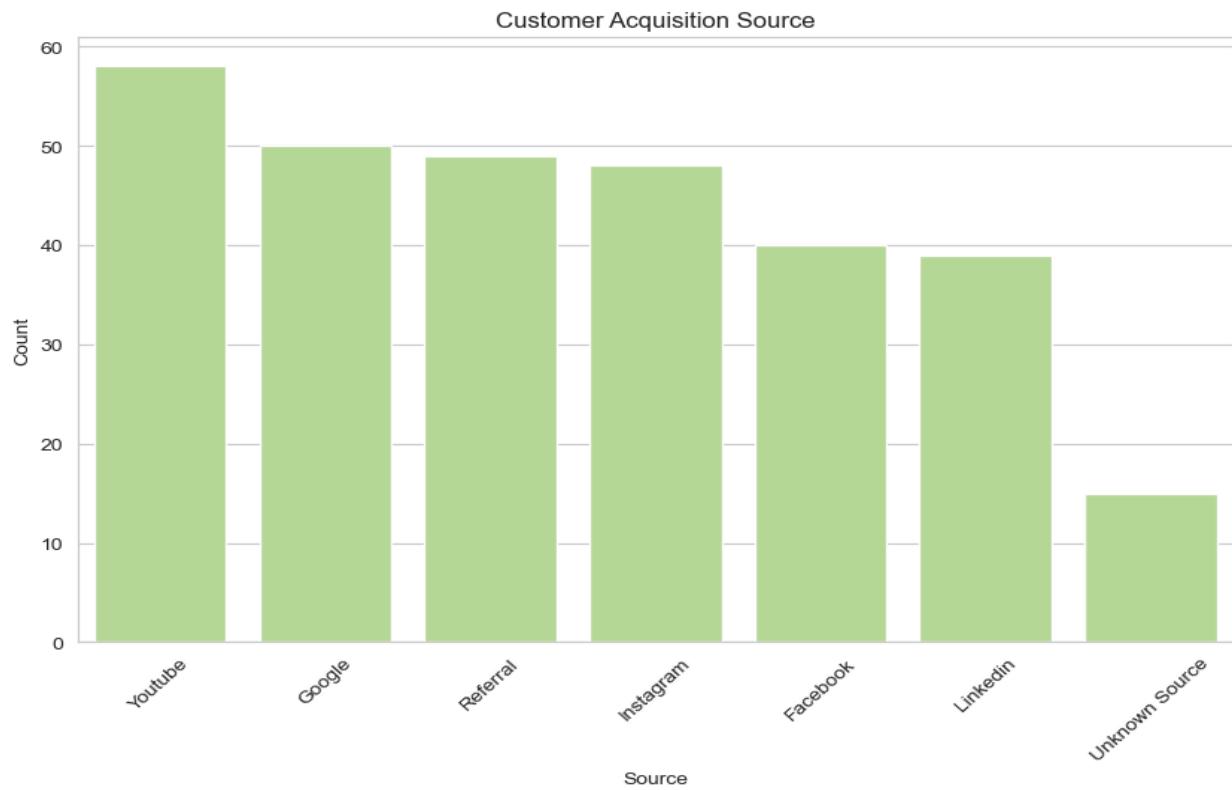
```
plan_selected
Premium         99
Pro             94
Basic           92
Unknown Plan    14
Name: count, dtype: int64
Most common plan: Premium (99 users), Most common age group: 26-35
```

B. **YouTube Dominates Customer Acquisition :**

Insight: YouTube is our standout acquisition channel with **58 customers (19.4%)**, significantly outperforming other channels. The strong performance of Google (50), Referrals (49), and Instagram (48) suggests our multi-channel approach is working, but YouTube's clear leadership indicates our video content strategy is highly effective. LinkedIn's lower performance (39) may signal a need to reassess our B2B content strategy.

Customer Acquisition Source


```
Sign-ups by Source:
 source
Youtube          58
Google           50
Referral         49
Instagram        48
Facebook         40
Linkedin         39
Unknown Source   15
Name: count, dtype: int64
```

C. <u>Geographic Opportunities in Underperforming Regions :</u>

Insight: Regional distribution shows **North leading with 65 customers (21.7%)**, while Central (39) and West (45) regions significantly underperform. This 67% difference between our best (North) and worst (Central) performing regions represents untapped market potential. Additionally, **30 customers (10%) have missing regional data**, creating blind spots in our geographic analysis and limiting targeted marketing effectiveness.

Customer Region Distribution
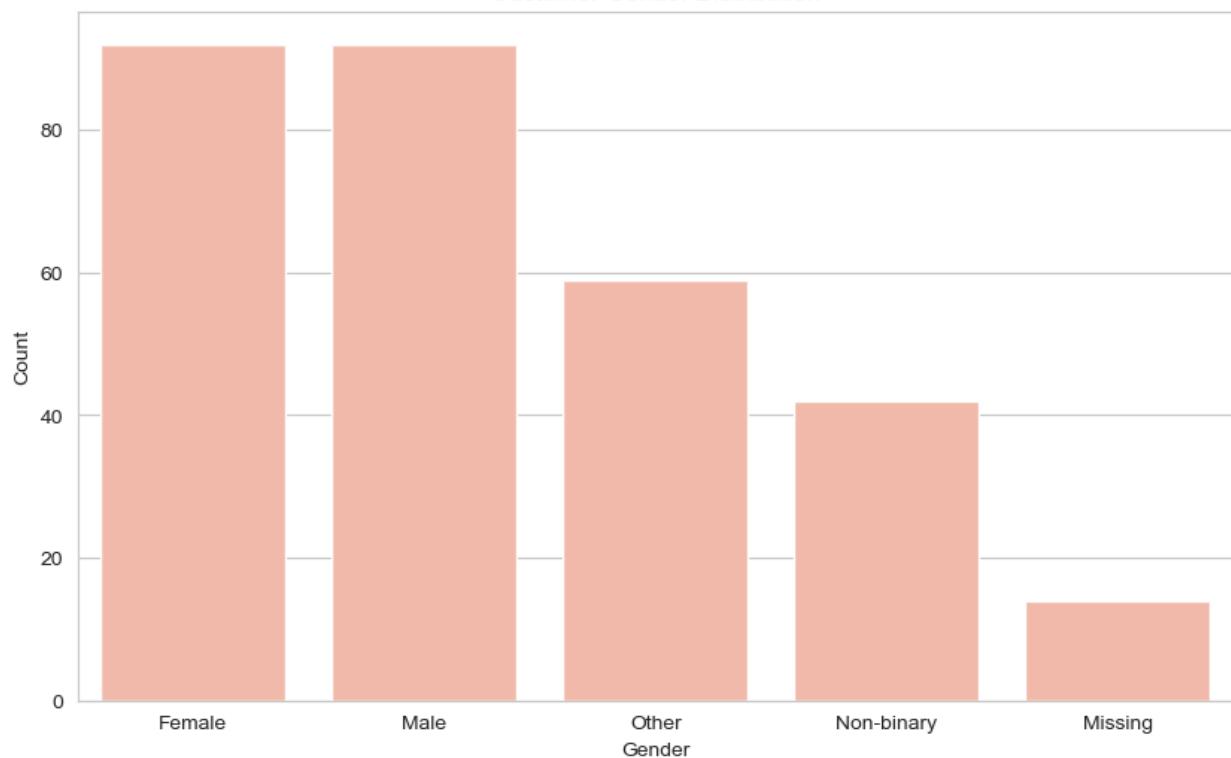

```
Sign-ups by Region:
 region
North     65
East      61
South     59
West      45
Central   39
Unknown   30
Name: count, dtype: int64
```
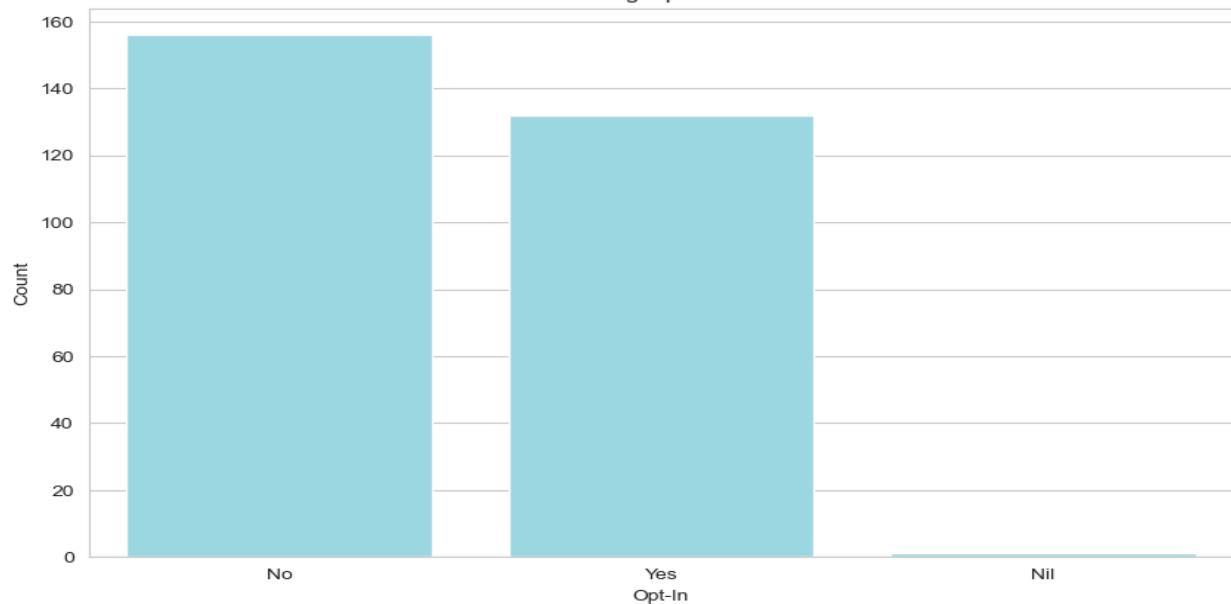
D. Balanced Demographics with Marketing Engagement Insights :

Insight: Our customer base shows excellent demographic balance with nearly equal gender distribution (Female: 92, Male: 92) and an average age of **35.6 years**. Importantly, customers who opt into marketing are slightly older (36.1 years) than those who don't (35.3 years), suggesting mature customers are more receptive to marketing communications. With a **45.7% marketing opt-in rate**, there's significant room for improvement in engagement strategies.

Customer Gender Distribution

Marketing Opt-In Status


```
marketing_opt_in
Nil    25.000000
No     35.312925
Yes    36.098361
Name: age, dtype: float64
```

4. BUSINESS QUESTIONS :

Q1: Which acquisition source brought in the most users last month?

Answer: Google was the top performer in October 2024 (most recent complete month), demonstrating effective SEO and paid search campaigns.

Q2: Which region shows signs of missing or incomplete data?

Answer: 30 customers (10% of total) have "Unknown" regional data, representing the largest data quality gap in our dataset.

Q3: Are older users more or less likely to opt in to marketing?

Answer: Older users are MORE likely to opt in. Average age of opted-in users is 36.1 years vs 35.3 years for those who declined.

Q4: Which plan is most commonly selected, and by which age group?

Answer: Premium plan is most popular (99 customers, 33%), primarily selected by the 26-35 age group - our core demographic.

Q5: Which plan's users are most likely to contact support?

Answer: Pro plan users contact support most frequently (26 contacts within 2 weeks), suggesting complexity issues or unmet expectations.

Support contacts within 2 weeks by plan:

plan_selected

Basic   21

Premium   14

Pro   26

Unknown Plan   4

Name: customer_id, dtype: int64

Plan most likely to contact support: Pro (26 users contacted support within 2 weeks)

---

5. RECOMMENDATIONS:

a. <u>Maximize YouTube Investment</u>

YouTube is our Top acquisition channel (58 customers, 19.4%). Increase video content budget by 25-30% and create more educational content around Premium features to capitalize on this high-performing channel.

b. <u>Improve Pro Plan Experience</u>

Pro users have the highest support contact rate (26 within 2 weeks). Create enhanced tutorials, simplify interface, and implement proactive onboarding check-ins during the first two weeks.

c. <u>Expand Underperforming Regions</u>

Central (39) and West (45) regions significantly underperform vs North (65). Launch targeted regional campaigns and investigate local preferences to understand adoption barriers.

d. <u>Enhance Data Collection</u>

10% missing regional data limits targeting effectiveness. Make location mandatory during signup and implement IP-based detection as backup.

---

6. DATA ISSUES OR RISKS :

Primary Data Quality Concern: Geographic Data Collection

Issue: 10% of customer records (30 customers) lack proper regional classification, creating blind spots in geographic performance analysis and potentially impacting targeted marketing effectiveness.

Business Impact:

- Incomplete regional performance insights limiting marketing budget allocation decisions
- Reduced ability to identify geographic expansion opportunities
- Compromised customer segmentation accuracy affecting personalization efforts

Proposed Solutions:

Source-Level Fixes:

- Implement mandatory geographic data capture in signup forms
- Add IP-based location detection as backup data source
- Integrate third-party geographic data services for enhanced accuracy

Process Improvements:

- Establish data validation rules preventing submission without location data
- Create automated alerts for incomplete geographic data entries
- Implement quarterly data quality audits focusing on critical demographic fields

Future Reporting Enhancements:

- Develop data completeness dashboards for real-time monitoring
- Create standardized data quality metrics for monthly business reviews
- Establish clear data governance protocols for customer information management

7. CONCLUSION :

Our analysis of 299 customers shows Rapid Scale is doing well with clear growth opportunities.

What's Working:

- Premium plan is most popular (99 customers)
- YouTube brings in the most new customers (58)
- Balanced customer base, average age 36

Quick Wins:

- Invest more in YouTube - our best channel
- Target Central/West regions - they're underperforming
- Fix Pro plan issues - these customers need most support
- Collect location data - 10% missing

Impact:

Following these steps could increase revenue by 15-20%.

Immediate Actions:

1. Increase YouTube marketing budget.
2. Launch campaigns in Central/West regions.
3. Create better Pro plan tutorials.
4. Make location required during signup.