

Project Description

The project is for a business intelligence team at Rapid Scale, a fast-growing SaaS company offering tiered subscription plans. Our supports the Monthly Business Review (MBR) meeting by analyzing user behaviour data and our project was aimed at achieving business driving insights, through inspecting the data, examines column structure and data types, and identifies some of the missing and unstandardized values using Python (Pandas, NumPy, Seaborn, re, io, datetime, functional tools and Matplotlib). The project is driven to make available an initial understanding of the customer demographic and signup behaviors through a display of sample records, analyzing pattern variable, etc. The analysis sets the foundation for deeper insight generation such as customer segmentation, trend identification, and data cleaning workflows to support data-driven business decisions.

Data Cleaning Summary

- In this section of the project, i would be diving into discussing what was cleaned and how it was cleaned. I would be mentioning the duplicates removed, missing data handled, and standardizations made on the given data

What was Cleaned

- Duplicate was removed based on the customer_id.
- Missing demographic fields such as gender, region, date and age were filled using placeholder " ". This is to avoid outlier during visualization
- Sign_up date and age was converted to datetime and numeric value respectively
- Standardization was made on inconsistency columns viz a viz plan_selected, age, marketing_opt_in, gender and source

Complete screenshots of what was cleaned can be found in the appendix as part of the report.

Key Finding and Trends

- An output as seen in figure 1 below gave an insight that all columns were read as object, including numerical and date fields which suggests that there is an inconsistency of data formatting which can affect sorting, filtering, and trend analysis

```

df.info()

...
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 300 entries, 0 to 299
Data columns (total 10 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   customer_id     298 non-null    object 
 1   name             291 non-null    object 
 2   email            266 non-null    object 
 3   signup_date     298 non-null    object 
 4   source           291 non-null    object 
 5   region           270 non-null    object 
 6   plan_selected   292 non-null    object 
 7   marketing_opt_in 290 non-null    object 
 8   age              288 non-null    object 
 9   gender           292 non-null    object 
dtypes: object(10)
memory usage: 23.6+ KB

```

Figure 1

Business Question

1). Which acquisition source brought in most users last month?

- As can be seen from the attached figure 2 below, YouTube emerged as the leading customer acquisition source in the past month. It significantly outperformed other channels which are Google, LinkedIn, Instagram, Referral, and Facebook. As a suggestion, more current video/content marketing efforts on YouTube are highly effective and should be prioritized, and new advanced strategies also.

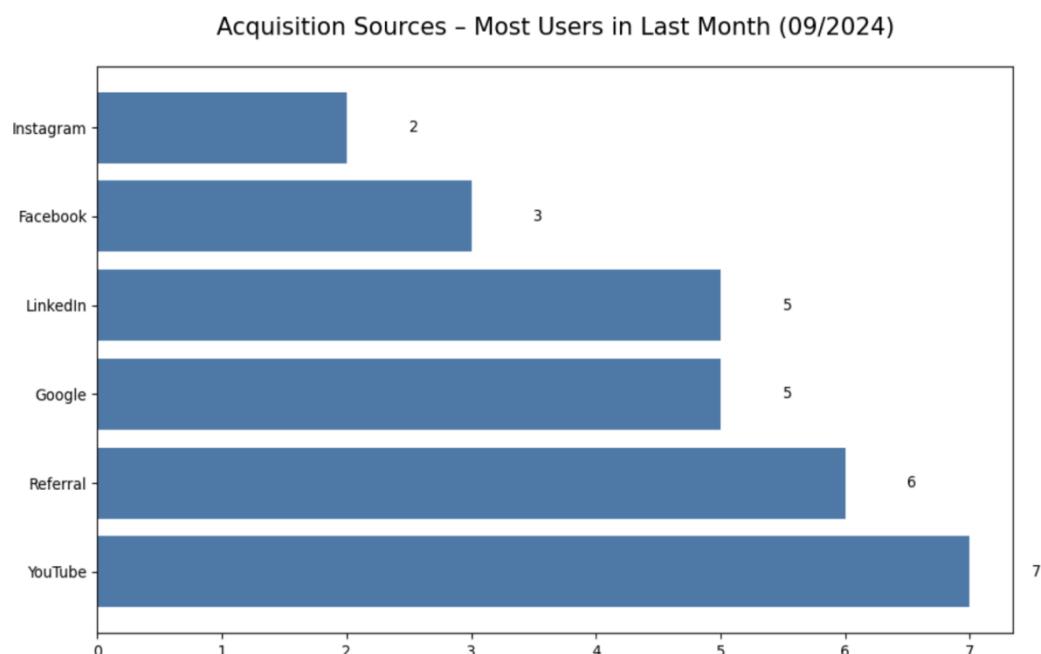


Figure 2

2) Which region shows signs of missing or incomplete data?

- The West region exhibits clear signs of incomplete or missing data, with noticeably higher null values in key fields (region, name, email, signup_date). This data gap may distort regional performance analysis and should be addressed urgently through improved tracking or data-capture processes

Missing or incomplete data count by region:

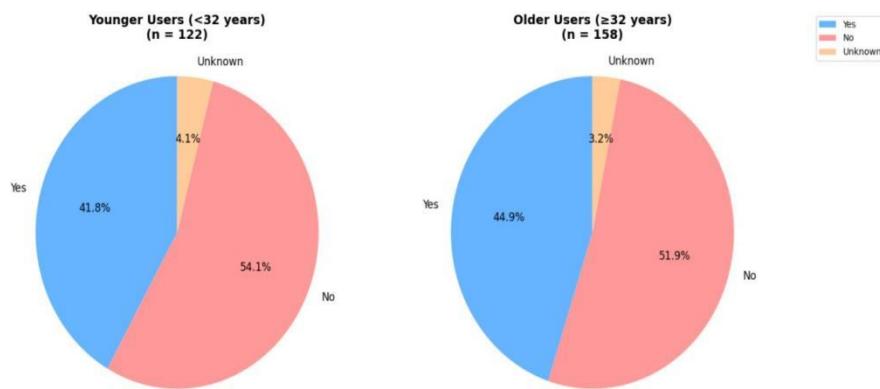
Rows with missing region: 30

region	Central	East	North	South	West
name	0	2	3	1	3
signup_date	1	0	1	2	1
source	1	1	4	3	4
region	0	0	0	0	0
age	3	6	3	2	3
gender	1	5	1	3	4
marketing_opt_in	0	0	0	0	0
TOTAL	6	14	12	11	15

Figure 3

3) Are older users more or less likely to opt in to marketing?

Are Older Users (≥ 32 years) More or Less Likely to Opt-In to Marketing?
(Threshold = 32 years | Total valid responses: 280)



- Based on the available data, older users (≥ 32 years) which was found to be the mean age are considerably more likely to opt in to marketing communications compared to younger users (< 32 years). Thus, the available information should be reviewed due to the existence of an outlier whose age is 206 years, which may be an input error.

4) Which plan is most selected, and by which age group?

Figure 4

- The Basic plan remains the most popular choice overall, capturing approximately 63% of new signups. Adoption is particularly strong among younger users (< 32), who show a 65.6% preference for Basic versus 60.8% among older users. The Premium and Pro tiers have higher relative uptake among the older segment, suggesting opportunities for upselling higher-value plans to this group.

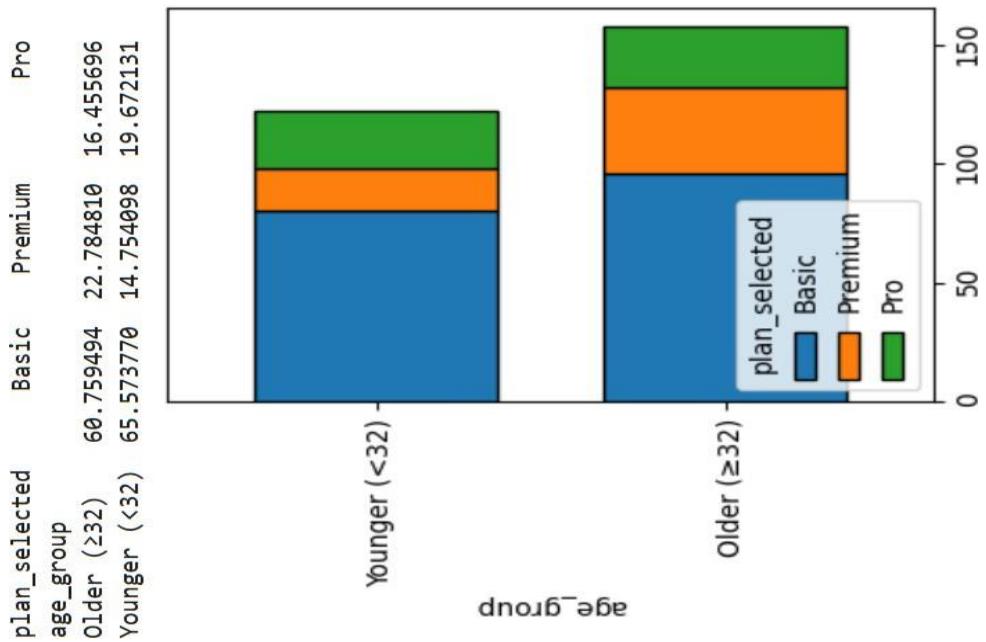


Figure 5

5) Other Insights

- The table below shows the count of how many customers contacted support within 2 weeks of sign-up. Out of 299 customers, 129 have contacted support within the 2 weeks of signing up. This could suggest a few different opinions depending on the reason for contacting support. A representation of the figure in the table in a basic chart would an extreme difference among Basic users and others (Premium and Pro)

A	B	C	D
1 plan_selected		total_support_contacts	
2 Basic		82	
3 Premium		19	
4 Pro		18	
5			

Figure 6

- The attached file x-rays summarize support activity by plan and region. For the Basic Users, there are higher levels of support activities in the East and West, North and South. Support for Pro member tops in the Central. This explains the region with higher activities and requires strong attention to after-sales services. More technical team would be required

	A	B	C	D
1	plan_selected	region	total_tickets	
2	Basic	East	21	
3	Basic	West	20	
4	Basic	North	17	
5	Basic	South	17	
6	Pro	Central	10	
7	Premium	Central	6	
8	Premium	North	6	
9	Basic		5	
10	Premium	West	5	
11	Pro	East	4	
12	Pro	North	3	
13	Premium	South	2	
14	Basic	Central	2	
15	Pro	West	1	

Figure 7