

Project Coversheet

Full Name	Onyekachukwu Emmanuel Agada
Email	yeksagads@gmail.com
Contact Number	+44 7546779925
Date of Submission	21/07/2025
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.

YOU CAN START YOUR PROJECT FROM HERE

Data Analysis for Business Insights

Introduction:

After recently joining the Business Intelligence team at Rapid Scale, the assigned task is to complete a data quality audit and provide insights into their user acquisition trends. The report from this analysis will assist both the Marketing and Onboarding teams in optimising their campaigns and engagement workflows. The dataset provided contains information on the customers, including their details (name, email), the date they signed up, the source that led them to register, the region they live in, and finally, the plan they selected after signing up.

There is an optional dataset that covers the customer support after signing up.

Data Cleaning Summary:

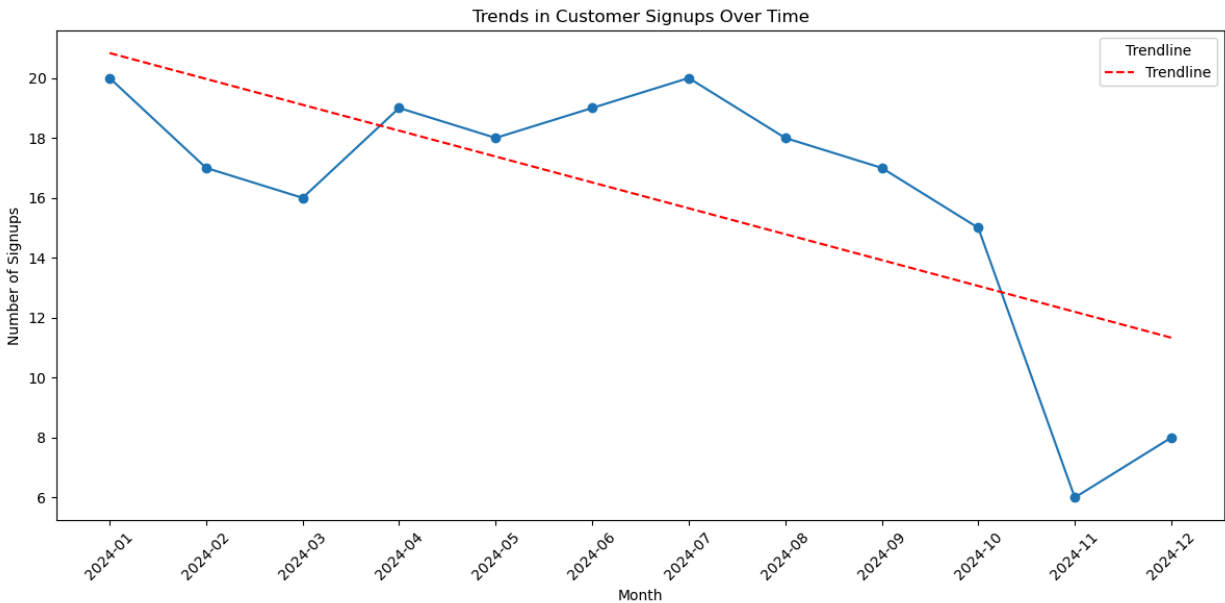
The necessary libraries, such as pandas, numpy, and matplotlib, were imported, after which the dataset, 'customer_signups.csv', was loaded into a dataframe called *custsup* using the pandas read_csv function. On initial inspection of the dataframe, there are 300 entries with 10 columns. All the columns are of the 'object' data type, and there are a total of 123 missing values, which account for 41% of the overall entries. It is also evident that there are various inconsistencies in the data input within certain columns. First, all missing values were removed from the dataframe, reducing the number of entries from 300 to 196. Then, the data types of the following columns were modified to better reflect their contents; the signup_date column was changed from object to datetime, the age column from object to integer, and the name column from object to string.

Second, I checked for duplicate entries in the dataframe and found none before or after removing the rows with missing values. This enabled me to correctly handle the

inconsistent entries in the *gender* and *plan_selected* columns by converting all labels to lowercase for easier readability.

Key Findings & Trends

One notable trend is that as the year progresses, the number of signups to the company decreases after peaking in July. This trend is illustrated in the graph below.



(Figure 1: Trend of Signups Over the Year)

A key finding in the data is that the average age of users who signed up for the subscription plan was approximately 36 years. This follows into the fact that most users who have a subscription plan fall into the ages between 25-40.

Another interesting discovery is the source linking users to the subscription plan. YouTube brought in the most users, with a total of 39, followed closely by both Google and Instagram. From the figure below, we can see that YouTube was responsible for the majority of signups during July.

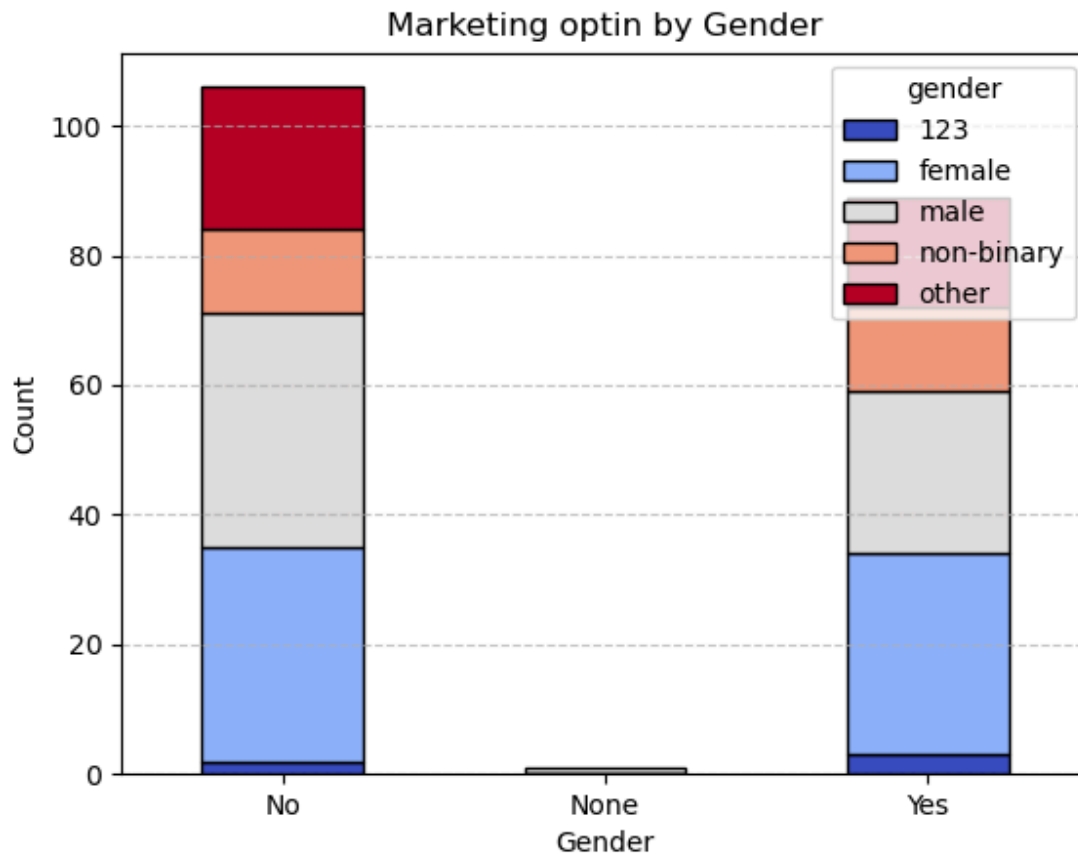
source	??	Facebook	Google	Instagram	LinkedIn	Referral	Youtube
signup_date							
2024-01	0	1	4	4	3	4	4
2024-02	1	2	2	1	4	5	2
2024-03	1	3	3	3	2	0	4
2024-04	0	4	1	5	3	4	2
2024-05	0	1	4	4	4	4	1
2024-06	0	4	4	0	1	2	8
2024-07	0	1	3	5	1	3	7
2024-08	0	4	2	4	1	2	5
2024-09	2	3	1	1	5	3	2
2024-10	0	2	7	2	1	1	2
2024-11	0	1	2	1	1	1	0
2024-12	0	1	0	3	1	1	2

(Figure 2: Source Acquisitions per Month)

YouTube	39
Google	33
Instagram	33
Referral	31
LinkedIn	28
Facebook	28
??	4

(Figure 3: Source Acquisitions Total)

Finally, it seems that males are more likely to opt out of marketing ads than females and vice versa.

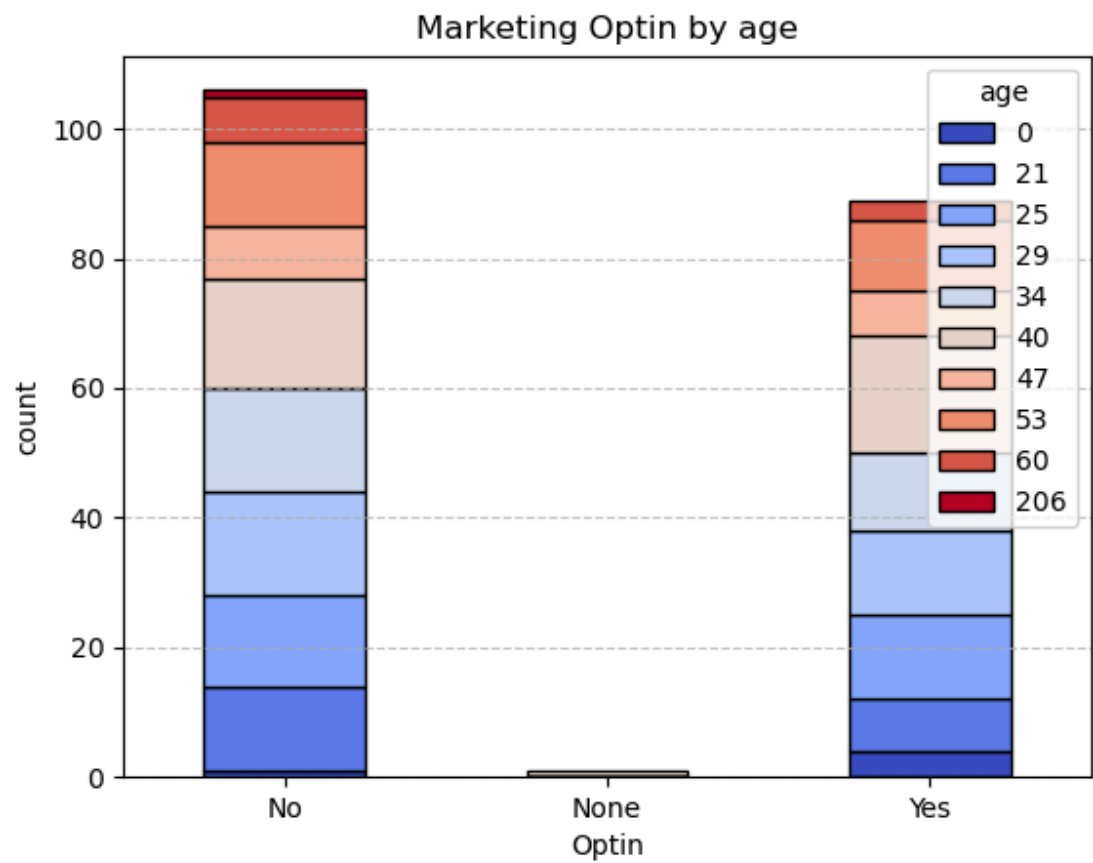


(Figure 3: Marketing Opt in by Gender)

Business Question Answers:

1. Instagram brought in the most users in December, with only 3 signing up.
2. The West and Central regions have incomplete data, as there are 30 missing values from the column, and they both have only 30 entries each.
3. From the data, older users are less likely to opt in to marketing. Assuming the range of older users starts at 40, the graph and screenshots below show that of the 196

users who signed up, 22% chose not to opt in to marketing, while only 20% chose to opt in.

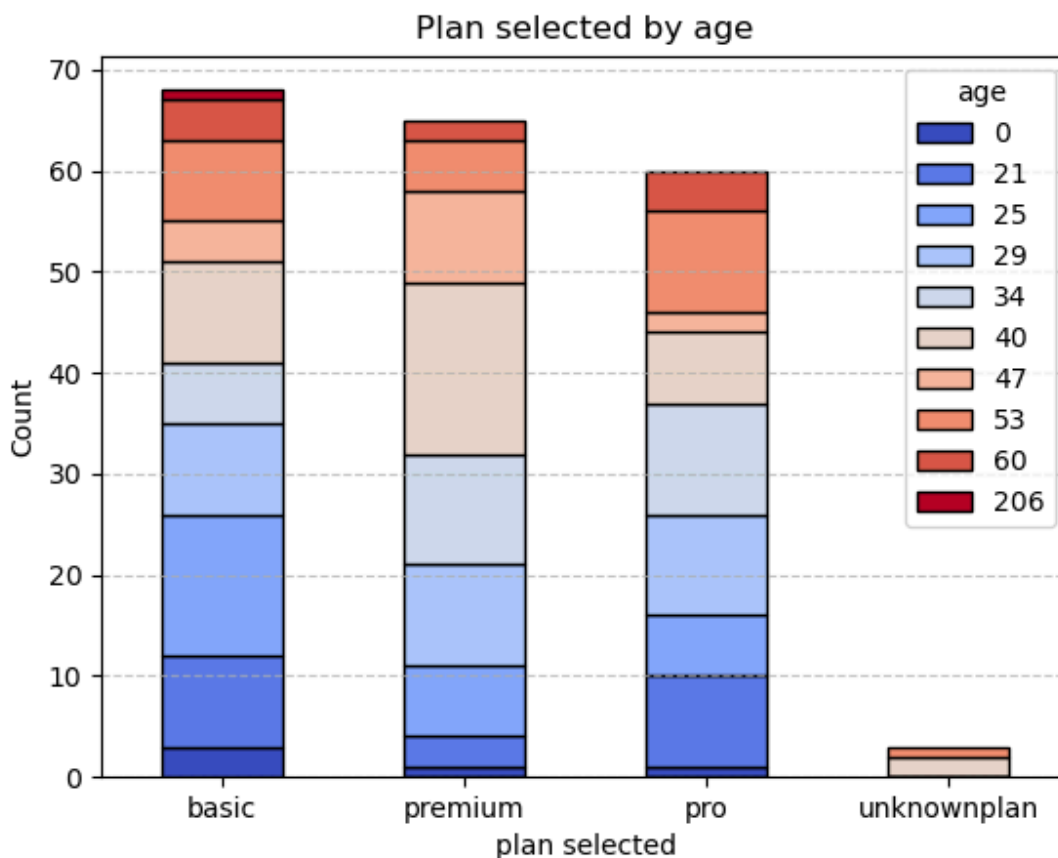


(Figure 4: Marketing Opt-in by Age)

age	0	21	25	29	34	40	47	53	60	206
marketing_opt_in										
No	1.0	13.0	14.0	16.0	16.0	17.0	8.0	13.0	7.0	1.0
None	NaN	NaN	NaN	NaN	NaN	1.0	NaN	NaN	NaN	NaN
Yes	4.0	8.0	13.0	13.0	12.0	18.0	7.0	11.0	3.0	NaN
marketing_opt_in										
No	106.0									
None	1.0									
Yes	89.0									
dtype:	float64									

(Figure 5: Marketing Opt in Output)

4. The most popular subscription is the basic plan, chosen by 68 users. This is because it is usually free or low-cost, appealing to those who may not have a high income or stable finances to easily afford the other plans. The age group most likely to select this subscription ranges from 25 to 40, accounting for over 50% of the users who signed up for the basic plan.



(Figure 6: Subscription Plan selected by Age Group)

5. Those who opted for the pro subscription plan were the most likely to contact support, with over 50% of its users contacting the support team and over half of that number having their issues resolved.

Recommendations:

To conclude, the Marketing team's focus should be on advertising the subscription plan on social media platforms that support video, as seen by the number of acquisitions both YouTube and Instagram brought in over the year. Platforms like Twitter or TikTok could be viable options as the age group with the most subscriptions regularly uses those platforms.

Another suggestion is to improve the data collection for certain columns, which will be further elaborated on in the Data Issues section.

Data Issues/Risks:

An issue found in the data is the large number of missing values. This makes it hard to perform proper Exploratory Data Analysis, as those missing values can greatly affect the output returned. A solution for this can be to make it compulsory for the users to input certain data when signing up for the subscription.

Another issue found in the data is the presence of extreme outliers. In the case of ages, some users inputted ages that make it difficult to observe the results. A solution for this may be to remove the option of entering custom ages or refining the age range selection.