



Week 1 - Project: Data Analysis for Business Insights

Project Title: Customer Sign-Up Behaviour & Data Quality Audit

Please write the answers in the '**Project Coversheet**' and refer to the dataset provided for completing the tasks.

Deliverables:

- Jupyter Notebook (.ipynb)
- PDF Report (1500-word max, with written answers and optional screenshots of outputs. No code screenshots required)

Project Scenario

You've just joined the Business Intelligence team at Rapid Scale, a fast-growing SaaS company offering tiered subscription plans. Every month, your team supports the Monthly Business Review (MBR) meeting by analysing user behaviour data.

Your manager has shared a recent customer sign-up dataset and asked you to complete a data quality audit and provide insights into user acquisition trends. This report will help the Marketing and Onboarding teams optimise their campaigns and engagement workflows.

They're particularly interested in:

- Identifying where the data may be inaccurate or incomplete
- Understanding how users are signing up and which plans they're choosing
- Assessing marketing opt-in behaviour and demographics

Dataset

1. Primary File: **customer_signups.csv**
2. Optional Stretch File: **support_tickets.csv**

Columns include:

- customer_id
- name
- email
- signup_date



- source (e.g., Google, Instagram, Referral)
- region
- plan_selected (Basic, Pro, Premium - expect inconsistencies)
- marketing_opt_in (Yes/No)
- age
- gender (may include inconsistent casing or spelling)

Tasks

1. Load & Clean the Data

- Identify missing values, data types, and column structure
- Convert signup_date to datetime
- Standardise inconsistent text values (plan_selected, gender, etc.)
- Remove duplicate rows based on customer_id
- Handle missing values (e.g., region, email, age)

2. Data Quality Summary

- Count of missing values per column
- % of missing values
- Number of duplicates removed
- Mention inconsistent category values corrected (e.g., PRO → Pro)

3. Summary Outputs (Using Pandas Aggregations)

Use .groupby() or .value_counts() to summarise:

- Sign-ups per week (grouped by signup_date)
- Sign-ups by source, region, and plan_selected
- Marketing opt-in counts by gender
- Age summary: min, max, mean, median, null count

4. Answer These Business Questions

Answer the following using your analysis. Write clear, concise answers in your PDF report:

1. Which acquisition source brought in the most users last month?
2. Which region shows signs of missing or incomplete data?
3. Are older users more or less likely to opt in to marketing?
4. Which plan is most commonly selected, and by which age group?
5. (Optional) Which plan's users are most likely to contact support?

5. Optional Stretch Task

- Load the support_tickets.csv dataset



- Join it to customer_signups.csv on customer_id
- Count how many customers contacted support within 2 weeks of sign-up
- Summarise support activity by plan and region (Group by plan and region)

Report Structure (Submit as a PDF file)

Note: Please write your answers in the 'Project Coversheet' provided.

Your report should be clear, concise, and well-structured, following the format below (maximum 1500 words):

1. Introduction
 - Briefly describe the task and dataset
2. Data Cleaning Summary
 - Explain what you cleaned and how
 - Mention duplicates removed, missing data handled, and standardisations made (optional: include screenshot of output)
3. Key Findings & Trends
 - Write 2-3 short insights based on the outputs
 - Optional: Include screenshots output
4. Business Question Answers
 - Clearly answer each question with short explanations
5. Recommendations
 - Suggest 2-3 ideas based on your findings (e.g., focus campaigns on the most engaged age group, improve data collection for missing regions)
6. Data Issues or Risks
 - Highlight one data quality problem
 - Explain how it could be fixed at the source or in future reporting

Note: Include screenshots from your jupyter notebook where required.

Submission Checklist

Before you submit, ensure you have:

- Your completed Jupyter Notebook (.ipynb)
- A PDF report following the structure above
- Answers written in the 'Project Coversheet' as instructed

Final Tip: Remember, your audience is non-technical. Avoid jargon. Your job is to tell a story with the data that helps the business make better decisions, clear, actionable, and relevant.