**Project Coversheet**

|  |  |
| --- | --- |
| Full Name | Savil Rosario |
| Email | [rosariosavil1905@gmail.com](mailto:rosariosavil1905@gmail.com) |
| Contact Number | 0892217344 |
| Date of Submission | 05/07/2025 |
| 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**](mailto: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**](mailto: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.

**Week 1 – Customer Sign-Up Behaviour   
Savil Rosario**

**Task 1: Load and Clean the Data**

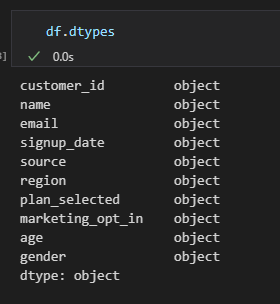
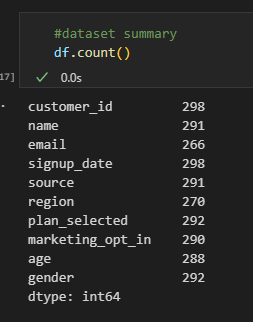
**Missing Values & Inconsistencies**

* **customer\_id**: Two missing values were identified and filled using sequential logic, maintaining consistency with existing IDs.
* **name**: Missing entries were populated by extracting names from the email address where possible. Remaining nulls were labeled as "Not Specified".
* **email**: Null and blank values were replaced with "Not Specified".
* **signup\_date**: Non-date strings, blanks, and missing values were coerced to NaT and filled using the column's median date.
* **region**: Missing values were replaced with "Unknown".
* **age**: The string "thirty" was converted to numeric (30), the outlier 206 was removed, and both null and "unknown" values were filled using the column median.
* **gender**: Inconsistent and invalid entries (e.g., blank, 123) were replaced with "Other" to maintain clean categorical data.
* **marketing\_opt\_in**: Inconsistent entries like null, "Nil", and "none" were standardized to False, while "yes" was mapped to True.
* **plan\_selected**: Text values were standardized to a consistent set: "Basic", "Pro", "Premium" and "Unknown Plan"
* **source**: Blank entries and those containing "??" were replaced with "Unknown".

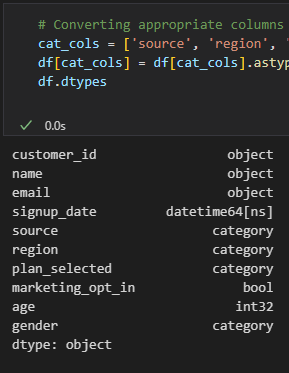
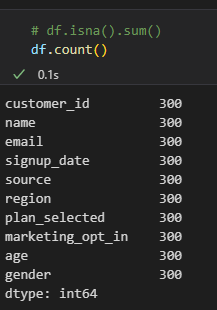
**Duplicates & Data Types**

* Checked for and removed duplicate rows based on customer\_id, treated as a primary key.
* Reviewed and updated data types for all columns to ensure they match their expected formats (e.g., datetime for signup\_date, category for plan\_selected).

**Fig:1.1-Dataset summary Before:**



**Fig:1.2-Dataset summary After**

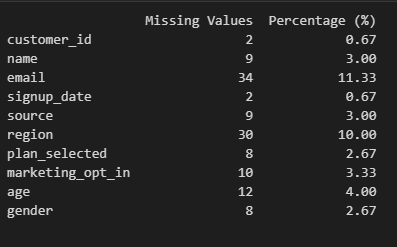


**Task 2. Data Quality Summary**

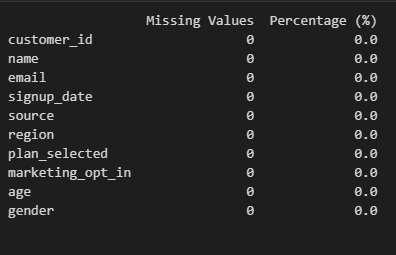
* 1. **Count of Missing values per column**

The analysis confirms that there are no missing values in any column. This is the result of a prior data cleaning process, during which all missing entries were identified and filled with appropriate and context-relevant values. The dataset is therefore complete and ready for further analysis.

**Fig:2.1a-Missing Values Before**

****

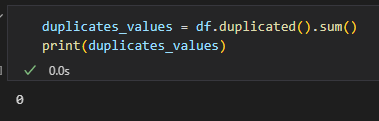
**Fig:2.1b-Missing Values After**

****

* 1. **Number of Duplicates Removed**

The analysis found no duplicate entries, indicating that the dataset contains only unique records.

**Fig 2.2a**



* 1. **Inconsistent Category Values Corrected**

To ensure consistency in categorical data, several corrections were applied across relevant columns:

* In the source column, ambiguous entries labeled as "??" were standardized to Unknown.
* In the plan\_selected column, variations in plan names such as "prem", "PREMIUM", "PRO", "Basic", and "UnkownPlan" were normalized to lowercase values: premium, pro, basic, and unkownplan, respectively.
* The gender column values were also converted to lowercase to maintain uniform formatting.

**Task 3. Summary Outputs**

**3.1 Sign-ups per week (grouped by signup\_date)**

Week 22 recorded the highest number of sign-ups with 14, while all other weeks had either 6 or 7 sign-ups. Overall, the unique weekly sign-up counts were 14, 7, and 6.

**Fig 3.1a**

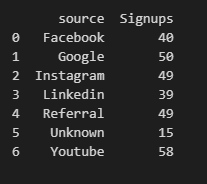
****

**3.2** **Sign-ups by source, region, and plan\_selected**

**3.2.a Sign-ups by Source**

Google leads with 50 sign-ups, followed closely by Youtube (58) and Instagram (49). Facebook and Referral sources also show strong engagement, while Unknown accounts for the lowest sign-ups at 15.

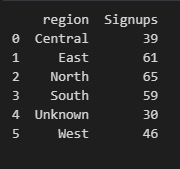
**Fig: 3.2.a**

****

**3.2.b Sign-ups by Region**

The North region has the highest sign-ups at 65, with East (61) and South (59) also performing well. Unknown regions have 30 sign-ups, indicating some missing location data.

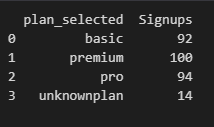
**Fig: 3.2.b**

****

**3.2.c Sign-ups by Plan Selected**

The Premium plan is the most popular with 100 sign-ups, followed by Pro (94) and Basic (92). The UnknownPlan category has the fewest sign-ups at 14.

**Fig: 3.2.c**

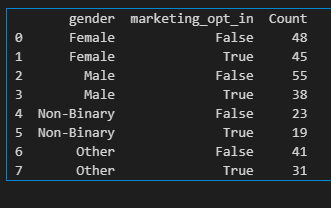


**3.3 Marketing opt-in counts by gender**

Across all gender categories, more customers chose not to opt in to marketing communications.

* Males had the highest number of opt-outs (55), followed by Females (48).
* Non-Binary and Other genders also showed higher opt-out counts than opt-ins.
* Overall, marketing opt-in rates were lower than opt-out rates across all groups**.**

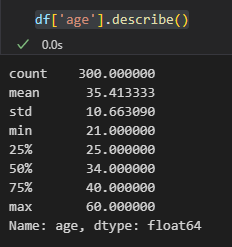
**Fig 3.3.a Group by Marketing and Gender**

****

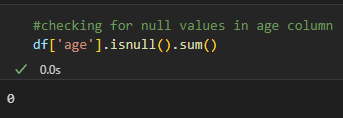
**3.4 Age summary: min, max, mean, median, null count**

* Customer ages range from a minimum of 21 to a maximum of 60 years.
* The average (mean) age is 35.4, and the median age is 34, indicating a fairly even distribution.
* There are no missing values in the age column.

**Fig 3.4.a Column Age Summary**

****

**Fig 3.4.b Column Age Null Values**

****