

Exploratory Data Analysis (EDA) Report

DVT TEAM 14

1. Introduction

This analysis was conducted as part of the **Excellerate Internship Program (Week 2 Deliverable)**.

The goal of this EDA is to understand and visualize applicant and outreach campaign data - identifying trends, inconsistencies, and insights that support decision-making for future campaigns.

The dataset combines three original sheets:

- **Applicant Data**
- **Campaign Data**
- **Outreach Data**

After merging and cleaning, the dataset reflects how different campaigns performed, which applicants were reached, and what outcomes were achieved.

2. Dataset Overview



Key Dataset Attributes:

Pre-Cleaning:

- **Total Rows:** 11,841
- **Total Columns:** 12
- **Total Campaigns:** 12
- **Total Valid Applicant IDs:** 6497
- **Missing Applicant IDs:** 5343
- **Data Coverage:** Applicant details, campaign details, outreach outcomes

Post-Cleaning:

- **Total Rows:** 6498
- **Total Columns:** 16
- **Total Campaigns:** 12
- **Total Valid Applicant IDs:** 6497
- **Missing Applicant IDs:** 0
- **Data Coverage:** Applicant details, campaign details, outreach outcomes

Key Variables:

Variable	Description
Applicant_id	Unique identifier for each applicant
Phone_number	Applicant's contact number
Country	Country of applicant (some entries had emails)
University	Illinois Institute of Technology
Received_at	Date the application was received
Caller_name	The outreach specialist who contacted the applicant
Outcome	Result of outreach (Interested / Not Interested / No Response)
Remarks	Notes from outreach call
Campaign_id	Unique campaign identifier
Campaign_name	Campaign title
Campaign_category	Pre-admission or Post-admission
Campaign_start_date	When the campaign began
Application_date	When the applicant submitted their form

Data Types:

Column	Data Type	Notes
Applicant_id	Numeric (some text mixed)	Cleaned to 6-digit integers
Phone_number	Text	Uniform formatting applied
Country	Text	Some entries contained emails
Dates	DateTime	Converted to standard date format
Outcome	Categorical	Contains missing or blank entries

Initial Observations:

- Several Applicant IDs were missing or in an incorrect format (scientific notation, long IDs, non-numeric).
- Emails were misplaced under the “Country” column.
- Date columns had timestamps, which were cleaned.
- Mismatch between Applicant IDs and Outreach Reference IDs (only ~5,658 matched).
- Significant missing outcome values, making conversion analysis harder.

3. Missing Data

Column	Missing Values	Issue Description
Applicant_id	5,343	Missing or invalid IDs
Outcome	Several hundred	Applicant not reached or data not logged
Remarks	Many	Optional field, often blank
Country	Few	Some replaced by emails

Solution:

- Standardized all Applicant IDs → removed non-numeric, extra-long, and exponent-form values.
- Replaced blank outcomes with “No Response.”
- Identified emails in “Country” column using:
`=ISNUMBER(SEARCH("@", A2))` → manually corrected or moved.
- Converted all dates using custom format `yyyy-mm-dd`.
- Applied validation filters to detect and remove duplicates.

4. Key Findings







4.1 Outcome Distribution

Outcome	Count	% Share
Not connected	4,054	62.4%
Will Submit the docx	971	15.0%
Completed application	287	4.4%
Not interested to IIT	219	3.4%
Disconnected	204	3.1%
Will confirm later	138	2.1%
Ready to pay / Paid deposit	66	1.0%
Others	558	8.6%

Interpretation:

- The majority (62%) were *Not connected* — major opportunity area.
- About **4.4%** completed applications → the strict conversion rate.
- Including *soft conversions* (Will Submit, Ready to Pay) gives a potential **conversion funnel of 21–23%**.

4.2 Top Performing Callers

Caller	Calls Made	Relative Share
Rudra	1,395	 21.5%
Prajwal	1,303	 20.1%
Isha	1,073	 16.5%
Palak	947	 14.6%
Poppy	920	 14.2%
Shailja	741	 11.4%

4.3 Campaign Performance Snapshot

Campaign ID	Calls Logged	Category	Remarks
IANF23	1,683	Pre-Admission	Highest activity volume
FA24SIC	1,251	Post-Admission	Consistent call outcomes
FA24IP	833	Pre-Admission	Balanced distribution
FA24AND	586	Pre-Admission	Moderate scale
IND23	223	Post-Admission	Smaller, focused
Others (7 campaigns)	1,921	Mixed	Lower scale

4.4 Outcome Insights Summary

Positive Outcomes (Completed + Ready + Connected): 4,665 → **71.8%**

Negative/No Contact Outcomes: ~1,830 → **28.2%**

Strict conversions (Completed Application): **4.4%**

Soft conversions (Ready to Pay, Will Submit): **~17–20%**

Key Takeaway:

High engagement exists, but actual application completion remains low - pipeline leak between *interest* and *action* stages.

Conclusion & Next Steps

Main Takeaways:

- Dataset is clean, complete, and analysis-ready.
- **62%** Not Connected → biggest opportunity area.
- **4.4%** Strict Conversion Rate (Completed Applications).
- High activity in campaigns [IANF23](#) & [FA24SIC](#).
- Top callers (Rudra & Prajwal) handle most volume — review their scripts.

Next Steps

1. **Improve Connect Rate:**
Use better timing, validation, and reminder texts.
2. **Caller Performance Dashboard:**
Track conversion % per caller.
3. **Outcome Categorization:**
Standardize 16+ outcome labels into 5 clear groups.
4. **Campaign ROI Dashboard:**
Add metrics: cost per call, cost per completion.
5. **Automate QA:**
Flag missing remarks or low-quality entries for review.