

# Data Visualization Associate Internship Final Presentation

Transforming Raw Data into Validated Business Intelligence

## TEAM 5

by Team 5

- ⚙️ Erica Hammond (Team Lead)
- 👤 Kshitij Gaikwad (Project Manager)
- ⚙️ Marium Hammad (Project Manager)
- 👤 Swapna Ganne (Project Scribe)



# Objective & Scope

Transform outreach & campaign data into clean, validated, analytics-ready insights



Analyze



Validate



Visualize

From raw data →  
decision-ready insights

## End-to-End Coverage



EDA (*Exploratory Data Analysis*)



Cleaning & Validation



Master Table Dashboards

## Key Focus:

✓ Accuracy • ✓ Consistency • ✓ Usability  
For confident business decisions

## Datasets



Applicant



Outreach (Calls)



Dashboards



Applicant



Outreach (Calls)





# Week 1 – Exploratory Data Analysis (EDA)



## Missing Value Analysis:

- Minimal missing values in Applicant data
- No critical missing fields in Outreach & Campaign data



## Duplicate Analysis:

- Duplicate applicant IDs found
- Valid outreach duplicates (multiple calls per applicant)



## Outlier Detection:

- 37,000+ phone numbers had invalid length or formatting



## Initial Observations:

- High “Not Connected” call outcomes
- Strong dependency on phone number data quality



**Insight:** Early identification of **data quality** risks impacting outreach success







# Key Business Insights

## Application

1. India as the dominant Applicant Source
2. Significant applicants from Ghana and Nigeria
3. Long tail of low frequency counties



## Application

## Outreach

1. Majority of call marked **as not connected**
2. High occurrence of **Wrong Number outcomes**
  - “Will Submit Docs” show moderate engagement



Call Not Connected



## Campaign

1. More “Post Admission” campaigns than “Pre-Admission”
2. Suggest focus on late stage conversion rather than lead generation



## Campaign



## Week 2 – Data Cleaning & Validation (PostgreSQL)



### Data Cleaning Actions:

- Removed duplicate applicant records
- Standardized text fields (country names, outcomes)
- Cleaned phone numbers:
  1. Removed non-numeric characters
  2. Standardized to 10-digit format
  3. Invalid values set to NULL



### Reference ID Validation:

- Removed invalid and non-matching IDs



### Date & Timestamp Handling:

- Converted string fields to proper date formats



**Result:** Creation of a **clean, consistent, and reliable dataset**





# Master Table Creation

Integrate Applicant, Outreach & Campaign data for a unified source of **truth**



## Integrated Datasets

Combine **Applicant** ↔ **Outreach** ↔ **Campaign** data using PostgreSQL joins



## Master Table Includes:

- Applicant details
- Outreach details
- Campaign details



## Key Benefits:

- Centralized data access
- Faster dashboard development
- Consistent KPI calculation

## Outcome:

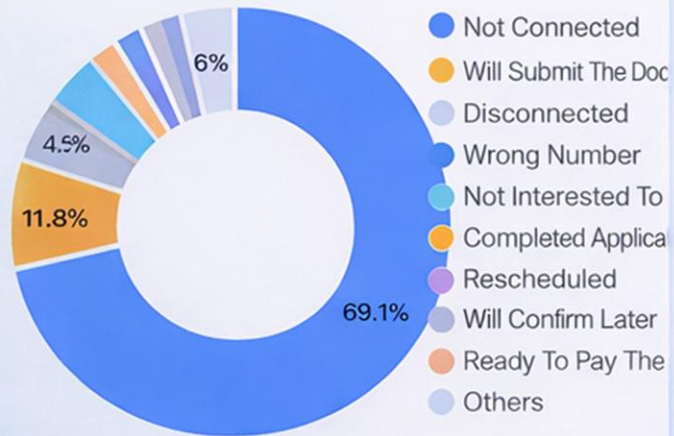
Validated foundation for reporting and analysis





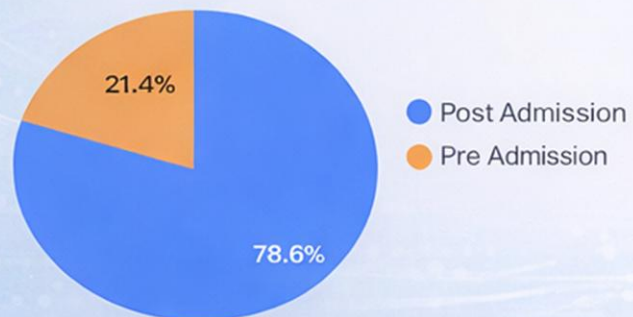
# Week 3: Looker Studio Dashboard

## Call Outcome Distribution



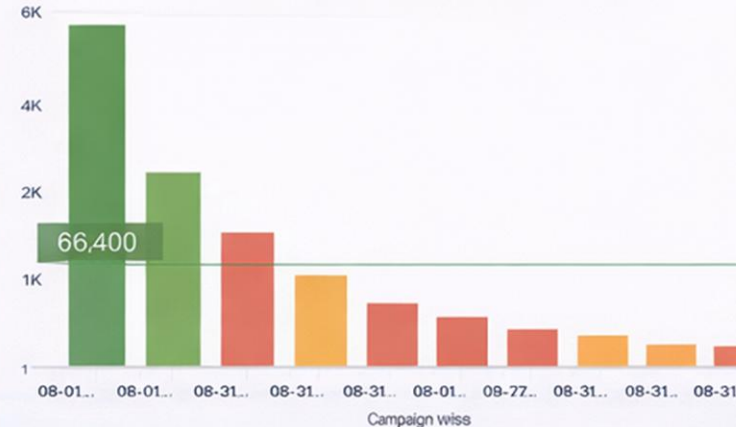
## Call Outcome Distribution

## Category Distribution



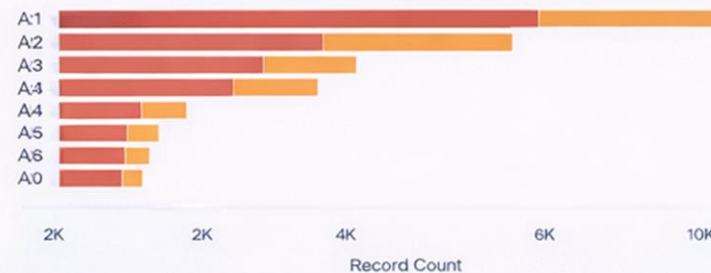
## Category Distribution

## Campaign-wise Call Volume



## Campaign-wise Call Volume

## Agent Performance



## Applicant Distribution



## Scorecards, Donut Charts, Bar Charts



## Geographic Heatmap



## Looker Studio

PostgreSQL

SQL

Google  
Workspace



# Final Insights & Business Impact



## Key Findings



High call volume  
does not guarantee  
better outcomes



## Strategic Insights



Phone data quality  
directly affects  
conversion



## Business Impact



**Closing Statement:**  
**From raw data to validated insights**  
enabling smarter outreach decisions