**Codveda Data Analytics Internship - Task 1 Report**

**Name:** Monil Sunilkumar Panchal

**Task Title:** Data Cleaning and Preprocessing

**Dataset Used:** churn-bigml-80.csv

**Objective:**  
To clean and preprocess a raw dataset by identifying and resolving issues like missing values, duplicates, and inconsistent data formats.

**Steps Performed:**

1. Loaded the dataset using pandas.
2. Verified the structure using .info()
3. Checked for missing values using df.isnull().sum() and confirmed that none were present.
4. Detected and removed duplicate rows using df.drop\_duplicates().
5. Cleaned categorical columns like 'International plan', 'Voice mail plan', and 'Churn':
   * Replaced 'yes'/'no' with 1/0.
6. Standardized column names (e.g., spaces replaced with underscores).
7. Saved the cleaned dataset as cleaned\_churn\_data.csv.

**Conclusion:**  
The dataset was successfully cleaned. No missing values were found, but duplicates were removed and categorical columns were standardized.