**Title** : Bank data prediction using Logistic regression

**Project Duration:** 1 Days  
**Date Completed:** April 15, 2025  
**Author:** *sagar khese*

**1 .project overview :**

Bank data uses for logistic regression to predict whether customer will subscribe or not to term deposit which based on demographic and contact-related from the bank data

**2. Solution Architecture :**

**Data pipeline flow :**

**Raw data > data processing > Feature encoding > Model Training (LR) > Model Evaluation > Model saving**

**And EDA visualization done in other notebook**

**Tool**

* **Python (sklearn ,seaborn ,Matplotlib)**
* **Excel**
* **Word docs**
* **Ppt**

**3. Methodology :**

| **Step** | **Reason** |
| --- | --- |
| **Data Collection** | Loaded CSV from local directory |
| **Data Cleaning** | Checked for null values, selected relevant columns |
| **Exploratory Data Analysis (EDA)** | Visualized categorical feature distributions and correlation heatmaps and bar graph |
| **Feature Engineering** | Label encoding for categorical features |
| **Modeling** | Logistic Regression |
| **Evaluation** | Accuracy score for training data |
| **Deployment Prep** | Saved the model using Joblib for future use |

**4. Model Performance :**

* **Accuracy**: **89.5%**

**5. Time Taken :**

| **Phase** | **Duration** |
| --- | --- |
| Data Cleaning & EDA | 1 hours |
| Feature Engineering | 0.5 hour |
| Model Training & Evaluation | 2 hour |
| Documentation & Visualization | 1.5 hour |
| **Total** | **~4 hours** |
| **6. Challenges Faced :**   * Some categorical variables had many unique values (e.g., job), affecting model interpretability. * The duration feature was highly predictive but needed removal to avoid data leakage. * Balancing class imbalance could be considered for future improvement.   **7. Project hardness :**   * **Complexity Level**: *Low to Medium*   + Logistic regression is interpretable and quick to implement |  |