Bank Marketing Data Analysis

Exploratory Data Analysis using Python

Project Overview

This project applies **Exploratory Data Analysis (EDA)** to a real-world bank marketing dataset, uncovering customer behaviour patterns and the key factors that drive term deposit subscription decisions.

Understand

Decode dataset structure and customer demographics

Analyse

Extract meaningful patterns through visualisation

Prepare

Ready the data for future machine learning models

BUSINESS PROBLEM

Why Do Some Customers Subscribe — and Others Don't?

Banks invest heavily in marketing campaigns to promote term deposits, yet response rates remain low. This project analyses customer data to answer three core questions:

- → **Which factors influence subscription decisions?**
- → **What patterns exist in customer behaviour?**
- → **How effective are current campaign strategies?**

Project Objectives

Six clearly defined objectives guided the analysis from raw data to actionable insight.

1

Understand Structure

Inspect data types, shape, and feature definitions

2

Assess Quality

Identify missing values and data inconsistencies

3

Analyse Target

Examine subscription distribution and class balance

4

Explore Relationships

Measure correlations across numerical features

5

Visualise Patterns

Surface insights through charts and heatmaps

6

Enable Modelling

Prepare a clean, structured dataset for ML pipelines

Dataset Description

Each row represents a unique customer record from a bank marketing campaign, spanning demographic, financial, and campaign-related attributes.

Customer Demographics

- Age
- Job & Marital Status
- Education Level

Financial Information

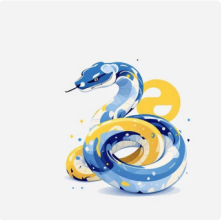
- Account Balance
- Housing Loan Status
- Personal Loan Status

Campaign Information

- Contact Duration
- Number of Contacts
- Previous Campaign Outcomes

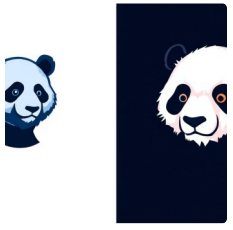
Tools & Technologies

The project was built entirely in Python, leveraging a modern data science stack for cleaning, analysis, and visualisation.



Python

Core programming language

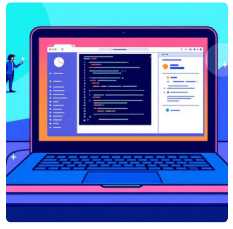


Pandas & NumPy

Data manipulation and numerical computation

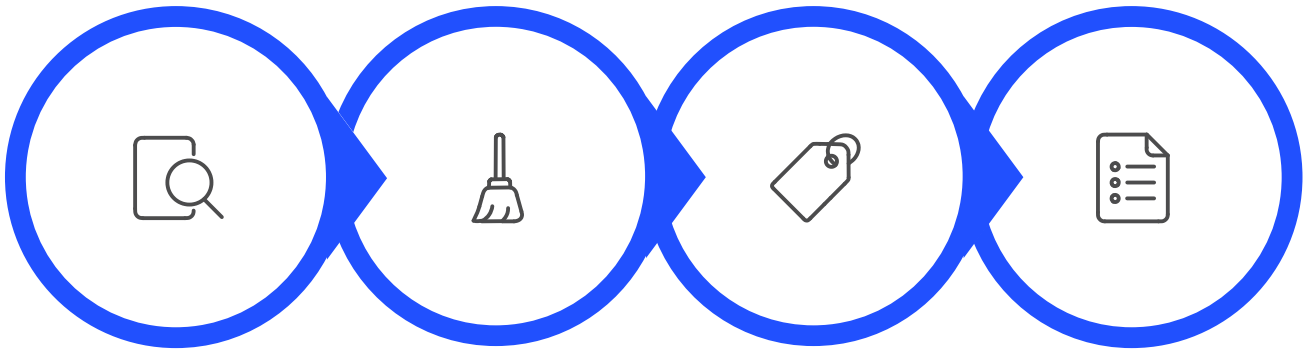
Matplotlib & Seaborn

Statistical charts and visual exploration



Jupyter Notebook

Interactive development environment



Inspect Clean Classify Select

DATA PREPARATION

Laying the Foundation for Reliable Analysis

Before any analysis could begin, the dataset was rigorously prepared to ensure accuracy and consistency.

No Missing Values Found

All features were complete, removing the need for imputation.

Data Types Verified

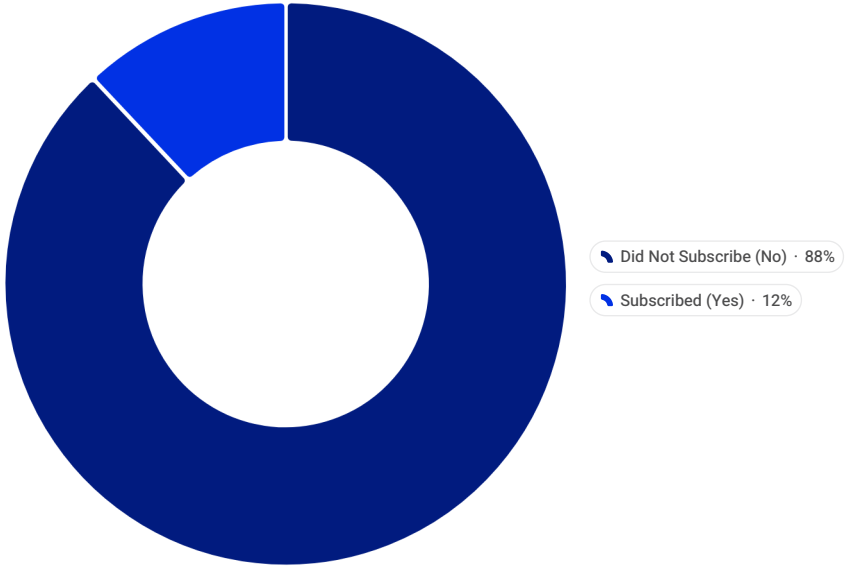
Categorical and numerical columns were correctly classified.

Numerical Subset Selected

Relevant columns isolated for downstream correlation analysis.

Target Variable: Class Imbalance Uncovered

The target variable `y` captures whether a customer subscribed to a term deposit. Analysis revealed a **significant class imbalance** — a critical finding for any future predictive modelling work.



This imbalance means future ML models must account for bias using techniques such as oversampling, undersampling, or weighted classification.

~88%

Customers did **not** subscribe

~12%

Customers **subscribed** successfully

Key Insights from EDA

Contact Duration Matters Most

Longer call durations showed the strongest association with successful subscriptions — quality of engagement outweighs quantity.

Demographics Show Weak Direct Links

Age, job, and marital status alone are poor predictors. Subscription decisions are multi-dimensional, not demographic-driven.

Campaign Variables Carry More Signal

Number of contacts and previous campaign outcomes provided stronger predictive indicators than financial attributes.

Class Imbalance Requires Attention

The 88/12 split demands deliberate handling before training any classification model to avoid skewed predictions.

Skills Demonstrated & Next Steps

Skills Demonstrated

- Data Cleaning
- EDA
- Visualisation
- Statistical Thinking
- Python
- Analytical Thinking

Future Scope

01

Feature Engineering

Create derived variables to improve signal strength

02

Encode Categorical Variables

Apply label and one-hot encoding for ML readiness

03

Build Classification Models

Train and evaluate models to predict subscriptions

🔗 **View the full project on GitHub:**
github.com/Ankurvashistha07