

Bank Marketing Insights — Data Cleaning & Correlation Analysis

Finds what tracks with campaign success, and stops where the data stops supporting a model.

BUILT WITH

Python

Pandas

Feature Engineering

Correlation Analysis

EDA

Jupyter Notebook

~20%

of credit_default values missing

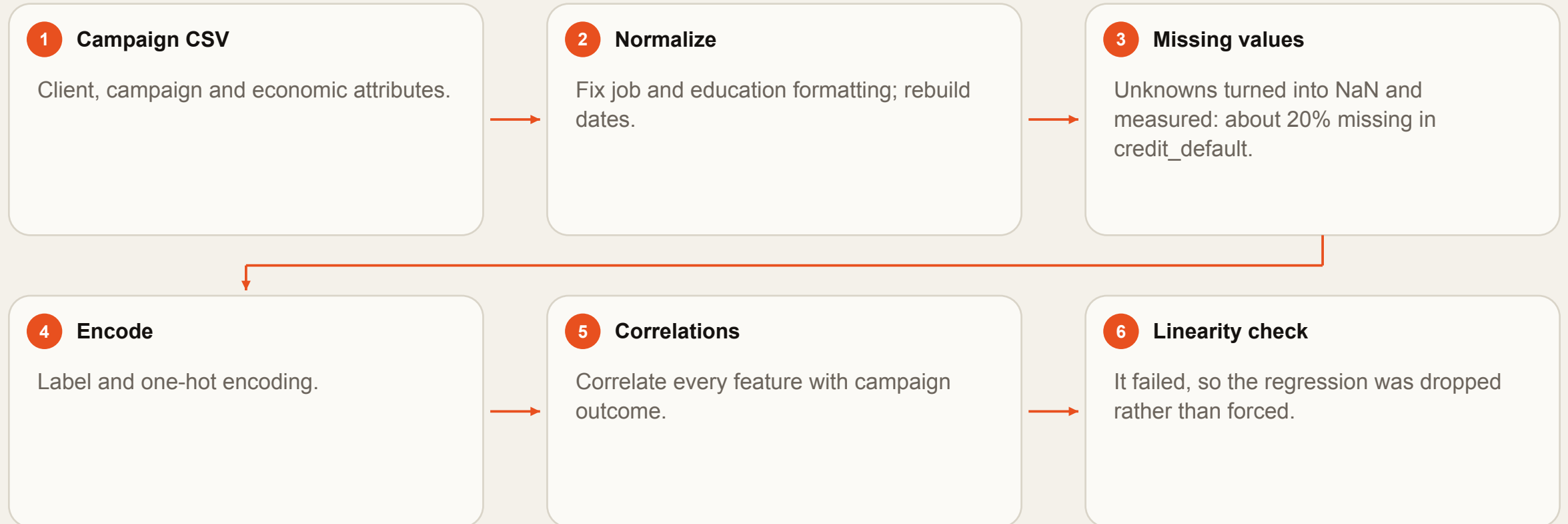
1

regression assumption tested: linearity failed

3

analysis subsets: client, campaign, economic

How it works



Key decisions

01

Stop instead of forcing a model

The linearity check failed, so the regression was dropped rather than shipped on bad assumptions.

02

Measure missing data first

Unknown values became NaN and were counted before deciding what to do, instead of being filled in silently.

03

Split by question

Client, campaign and economic columns were analysed separately so each question stays focused.

Stack, status and what is next

BUILT WITH

Python

Pandas

Feature Engineering

Correlation Analysis

EDA

Jupyter Notebook

KNOWN LIMITS

- No predictive model was built; the findings are correlations.

NEXT STEPS

- Try tree-based models (Random Forest, Gradient Boosting), which do not need linear relationships.

SOURCE

<https://github.com/wailethewave/Bank-Marketing-Insights>

