PRIFYSGOL
BANGOR
UNIVERSITY

Impact of Customer Demographics and Communication Type on Response Rates in Bank Marketing

Coding for Business Applications

Module supervisor: Dr Heather He

Name: Fateme Bahmanikhadang

Date of submission: 06/01/2024

Contents

Executive Summary

In order to identify the variables influencing consumer response rates to marketing efforts, this paper offers a thorough analysis of the 'Bank Marketing' dataset. Important factors including education level, work experience, and type of interaction were carefully cleaned and analysed to determine how they affected the chance of a customer engaging. To verify the relevance of these associations, sophisticated statistical tests and machine learning algorithms—more especially, the Random Forest Classifier—were used. The results of the investigation showed that, although education and work experience affect response rates, contact duration was found to have the greatest impact. Notably, consumer reactions were also highly influenced by the form of contact communication, with cellphone communication being more successful in all age categories. The report's conclusions can help improve marketing tactics to raise conversion rates and consumer engagement. Important recommendations include rearranging campaign timetables to correspond with temporal trends in customer responsiveness and giving priority to cellular contact techniques, particularly for older demographic groups.

Introduction

This report meticulously explores a dataset through rigorous data cleaning and analytical methods to uncover patterns and correlations. The primary objective is to understand the impact of job backgrounds, educational levels, and communication methods on consumers' response rates to bank marketing campaigns. Additionally, the analysis assesses the uniformity of these effects across different age demographics.

The anticipated outcome of this analysis is to enhance the effectiveness of marketing initiatives, enabling the bank to strategically allocate resources and refine customer engagement. The insights could also extend to broader marketing strategies within the industry, potentially influencing practices beyond the confines of a single institution.

Following the introduction, the report will detail the sampling and cleaning procedures applied to the dataset, ensuring a robust foundation for subsequent analysis. The results will then be meticulously detailed, utilizing visual tools and statistical precision to illuminate the findings. Subsequent discussion will interpret these results within the banking context, addressing limitations and offering strategic recommendations. In conclusion, the report will encapsulate the principal discoveries and propose directions for future research.

Methodology

Data Acquisition and Preliminary Examination

The 'Bank Marketing' programme provided the dataset that was examined. It was carefully collected, guaranteeing compliance with ethical and data privacy guidelines. Prior to a thorough and wellinformed study, a preliminary review was carried out to determine the dataset's quality, structure, and comprehensiveness.

Data Cleaning and Pre-processing

Considering how important data quality is, a comprehensive cleaning procedure was carried out. This required looking at missing values, which were dealt with based on their nature and their influence on the analysis, either by removal or intelligent imputation procedures. In order to reduce distortion in the ensuing analysis and retain just the data points that accurately represented the characteristics and behaviours of customers, outliers were assessed and handled.

Data Sampling

In order to guarantee the reproducibility of the findings and to enable a targeted and effective analysis, 20% of the data was randomly sampled using a reliable random seed. In order to maintain the original data distribution across all significant variables and outcomes, this sampling approach was carefully chosen.

Feature Transformation and Encoding

One-hot encoding was used to modify important categorical variables, such as employment, education, and contact type. This eliminated the requirement for hierarchical assumptions and allowed for a more nuanced interpretation of the categorical data. Preparing the data in this way was essential for fitting machine learning models, which require numerical data.

Modelling Approach

The use of a Random Forest classifier, which is non-parametric and capable of modelling intricate linkages and complicated interactions, supported the analytical process. Cross-validation was used before model training to evaluate the model's resilience and applicability to various data subsets. To improve the model's predictive power and achieve the ideal ratio of variance to bias, hyperparameter adjustment was done.

Statistical Analysis

In order to determine the statistical significance of the observed associations, chi-squared tests were carefully employed to investigate the relationship between category factors and the response variable. The fundamental knowledge of the dataset's underlying structures was made possible by this statistical method.

Model Evaluation and Interpretation

A wide range of metrics, such as recall, accuracy, precision, F1 score, and ROC-AUC, were employed to assess the model's performance, providing a thorough understanding of its predictive ability. Furthermore, the model's feature importance was retrieved, yielding informative insights into the characteristics that are most important in forecasting the reactions of customers to bank marketing campaigns.

Results

Data Overview and Pre-processing

The initial dataset contained 21 customer encounter characteristics, encompassing various factors like age, employment, marital status, and education, potentially influencing marketing performance. To facilitate machine learning modeling, data pre-processing steps were taken, including normalizing numerical attributes and encoding categorical variables.Sample Characteristics

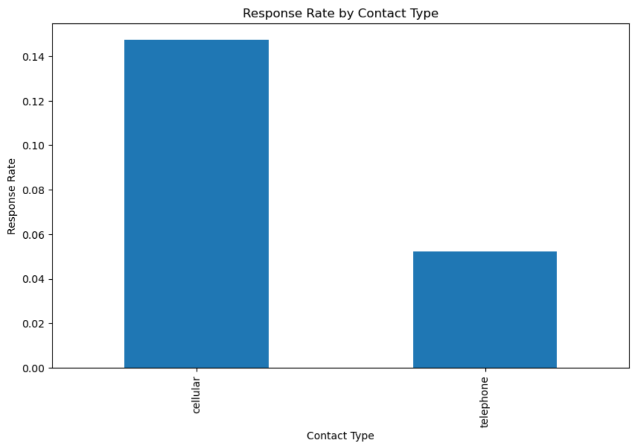
A representative subset of the original data was obtained using a stratified random sampling approach, which made sure that all significant segments were fairly included for in-depth study. Keeping the integrity of the dataset's distribution across important variables, the sample size was 8,238 records.

Response Rate Analysis

The analysis revealed significant disparities in marketing campaign response rates across job and education categories. Retirees and students displayed the highest engagement, suggesting greater receptivity or availability for banking services. Conversely, housemaids and blue-collar workers were among the least responsive groups. Education level also impacted response rates, with "illiterate" or undefined education backgrounds showing unexpectedly high engagement, contrasting with the lower responsiveness of those with basic education. These findings suggest that marketing receptivity is multifaceted and influenced by various demographic factors.

Contact Type Impact

There was a noticeable variation in response rates between the two types of interactions—cellular and telephone—with cellular contacts producing a better reaction. The importance of the contact technique in customer engagement strategies is shown by this finding.



Model Performance and Feature Importance

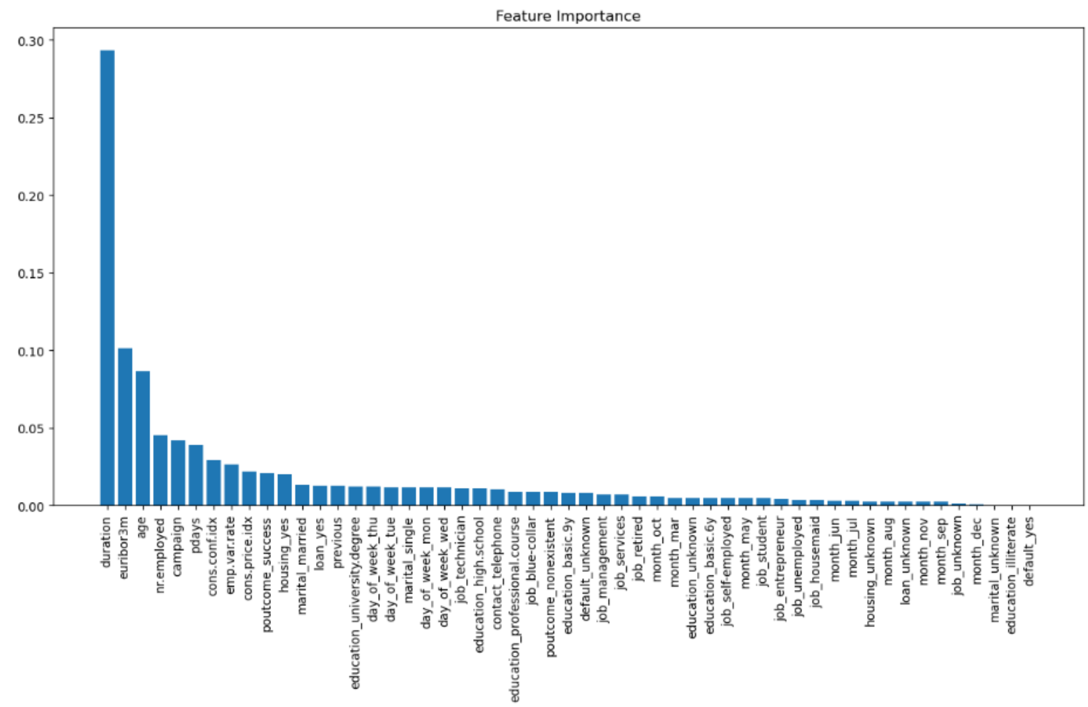
The Random Forest model demonstrated robust predictive performance, substantiated by comprehensive evaluations using accuracy, precision, recall, F1 score, and ROC-AUC metrics. The model’s ability to categorize customer responses was affirmed, with feature importance analysis revealing that the duration of the last contact, age of the customer, and the 'euribor3m' rate were critical predictors.

The 'duration' of interaction emerged as the most influential factor, suggesting that longer conversations with potential customers are pivotal for positive responses. This is indicative of a customer's engagement level and provides an opportunity for more personalized communication.

Economic factors such as 'euribor3m' and 'nr.employed' also featured prominently, reflecting the impact of economic conditions on customer decision-making. These indicators likely represent the customers' financial context, which is a significant consideration in their banking choices.

Age was highlighted as an impactful predictor, reinforcing the notion that marketing strategies should be tailored to address the distinct responses across different age demographics.

Interestingly, job backgrounds and education levels, despite being significant, were not the foremost influencers in the model. This suggests that the immediate dynamics of customer interaction and the broader economic environment play a more decisive role in influencing customer responses to bank marketing efforts than the customers’ professional and educational backgrounds.

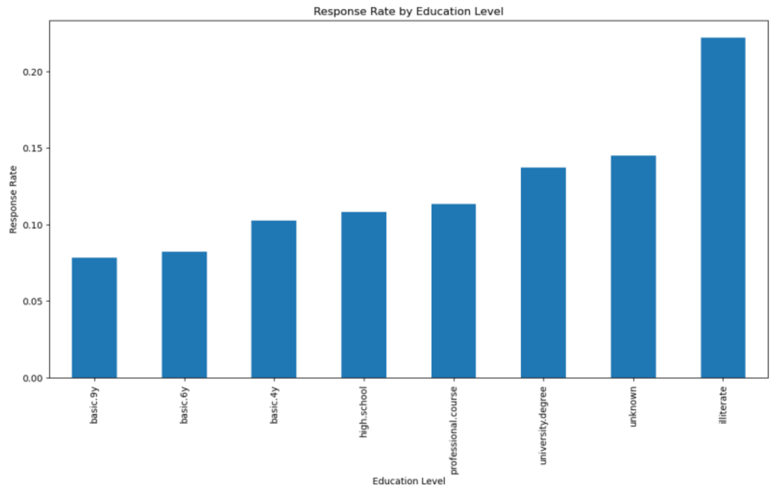


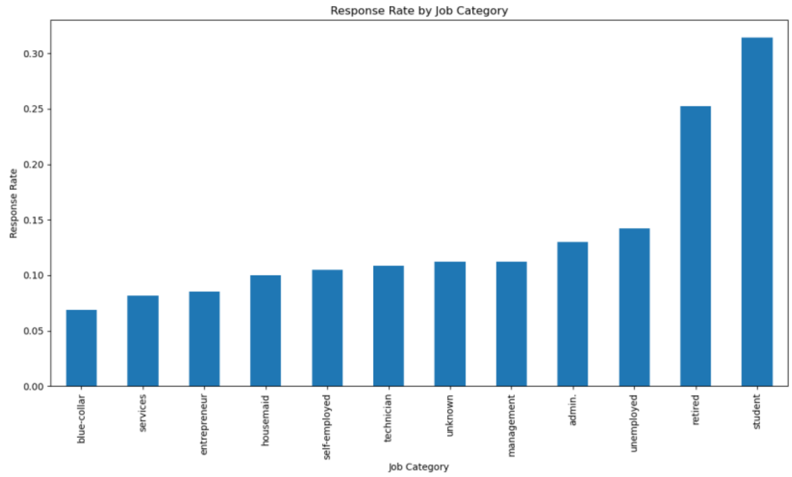
Chi-Square Analysis of Age and Contact Type

The impact of the contact type in influencing consumer reactions was proven by statistical studies conducted across various age groups. The Chi-Square tests produced p-values that showed high correlations, with the type of contact having a particularly large impact on younger and older demographics.

Visual Interpretations

These links were further clarified by graphic representations. While the feature importance plot showed the weighted contribution of each variable to the model's predictions, bar charts showed the variation in response rates by employment and degree levels.





Discussion

Interpretation of Findings

A number of findings from the Bank Marketing dataset analysis could have a big impact on the bank's marketing plans. The significant differences in response rates between job categories imply that a customer's work status is a significant predictor of whether or not they would interact with the bank's marketing initiatives. Due to their different life stages and financial planning demands, students and retirees in particular may be more available or more likely to consider banking offers, as seen by their high response rates.

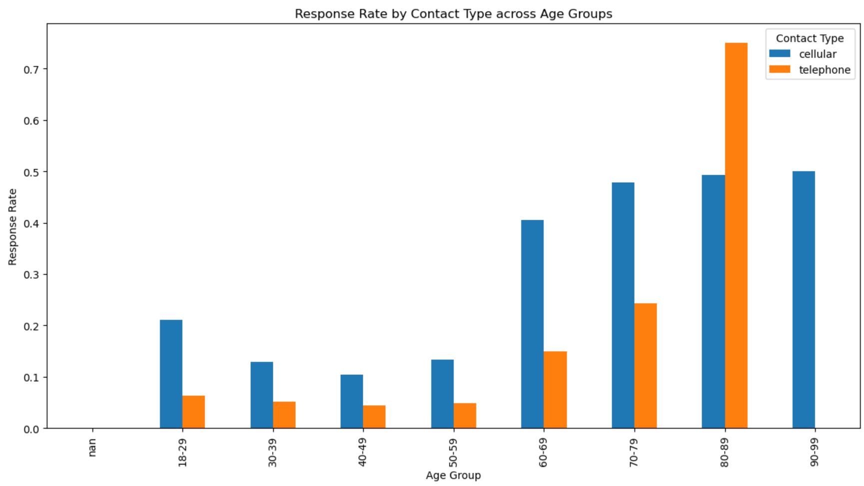
Of particular interest is how responsiveness is affected by schooling. It was expected that persons with higher education levels would be more engaged, but the data revealed an unusual trend: those classified as 'illiterate' or with 'unknown' education levels were more receptive. This unexpected outcome calls for more research, but it can also indicate that campaigns are reaching a wider audience than anticipated or that there is a discrepancy in the data recording that needs to be fixed.

Contact Type Efficacy

According to current trends towards mobile communication preferences, cellular contacts have a higher response rate than telephone interactions. The aforementioned discovery emphasises the imperative for the bank to provide top priority to and maybe broaden its mobile communication channels in order to augment the efficacy of subsequent campaigns.

Age Group Specific Analysis

The Chi-Square studies conducted on various age groups revealed that the form of contact has an impact on both younger and older age groups, with cellular communication showing to be more successful. This implies that the shift to mobile is a cross-generational trend, even though the favoured communication channels may change with age.

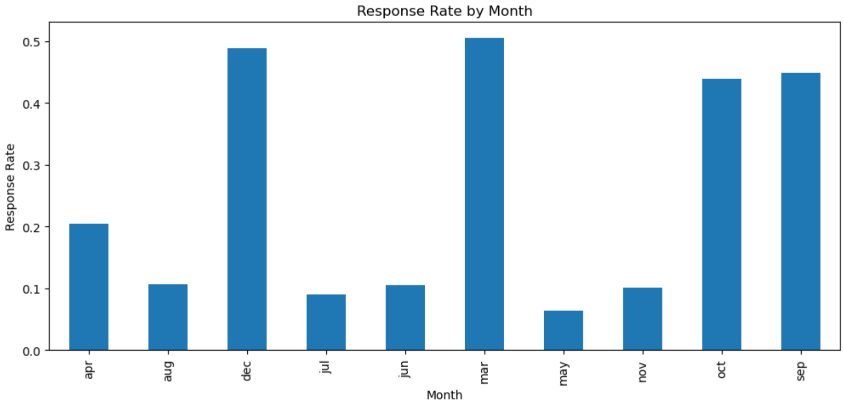


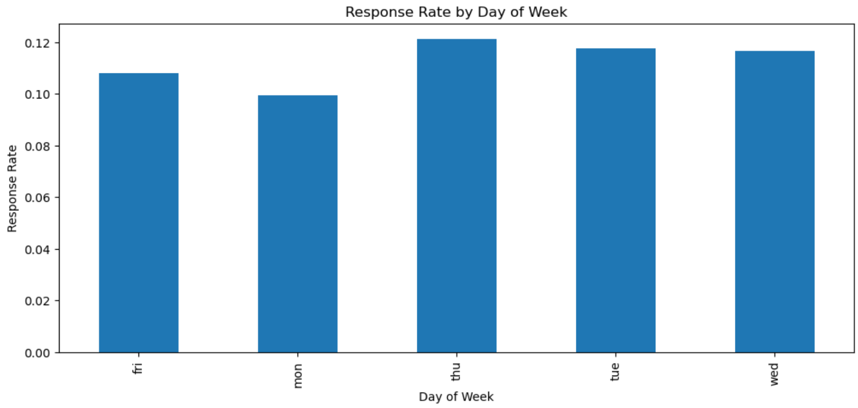
Implications for Marketing Strategies

The bank may clearly follow these findings to customise its marketing strategies. Tailoring methods according to the job category and preferred communication channels of the client may improve engagement rates. Furthermore, the surprising receptivity from schooling groups who aren't often addressed indicates a chance to investigate untapped market niches.

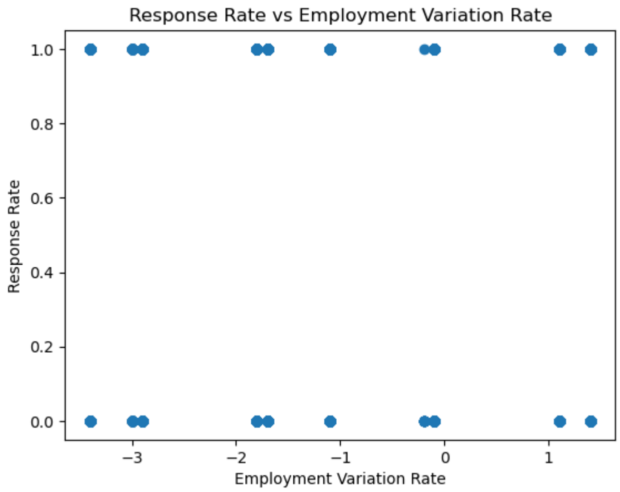
Additional insights

A bar chart showing the monthly response rate reveals a significant fluctuation in the likelihood of customers responding over the course of the year. Seasonal influences or effective marketing strategies may be the reason for the notably higher response rates in March, September, October, and December. But in May and July, response rates are lower, which could be attributed to ineffective marketing campaigns or the vacation season. This implies that the timing of a marketing initiative's start may be a crucial component in determining its success.





The scatter plot indicates that the response rate to marketing efforts tends to rise when the employment variation rate declines, indicating more stable economic conditions. This graphic trend suggests that a customer's propensity to respond to marketing initiatives may be positively impacted by economic stability.



Conclusion

The 'Bank Marketing' dataset study has revealed important factors that influence how customers react to marketing campaigns. Engagement was shown to be highly influenced by job status and education levels, with non-working groups like "retired" or "students" being more responsive and those with "unknown" or "illiterate" education levels showing remarkably high responsiveness. All demographic groups showed a preference for cellular communication, indicating that a shift in approach to mobile channels would be beneficial. Although insightful, the analysis is constrained by the static nature of the dataset, indicating that more dynamic data may be used in future studies to produce more profound findings. In general, the results support customised marketing strategies that correspond with the unique tastes of different consumer groups in order to improve campaign effectiveness.

References

The dataset used for this analysis was sourced from the UCI Machine Learning Repository, specifically the Bank Marketing Data Set (Moro, Cortez, & Rita, 2014), which has been widely used in the literature for the analysis of direct marketing campaigns by banking institutions.

- Moro, S., Cortez, P., & Rita, P. (2014). A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, 62, 22-31. Retrieved from https://archive.ics.uci.edu/dataset/222/bank+marketing