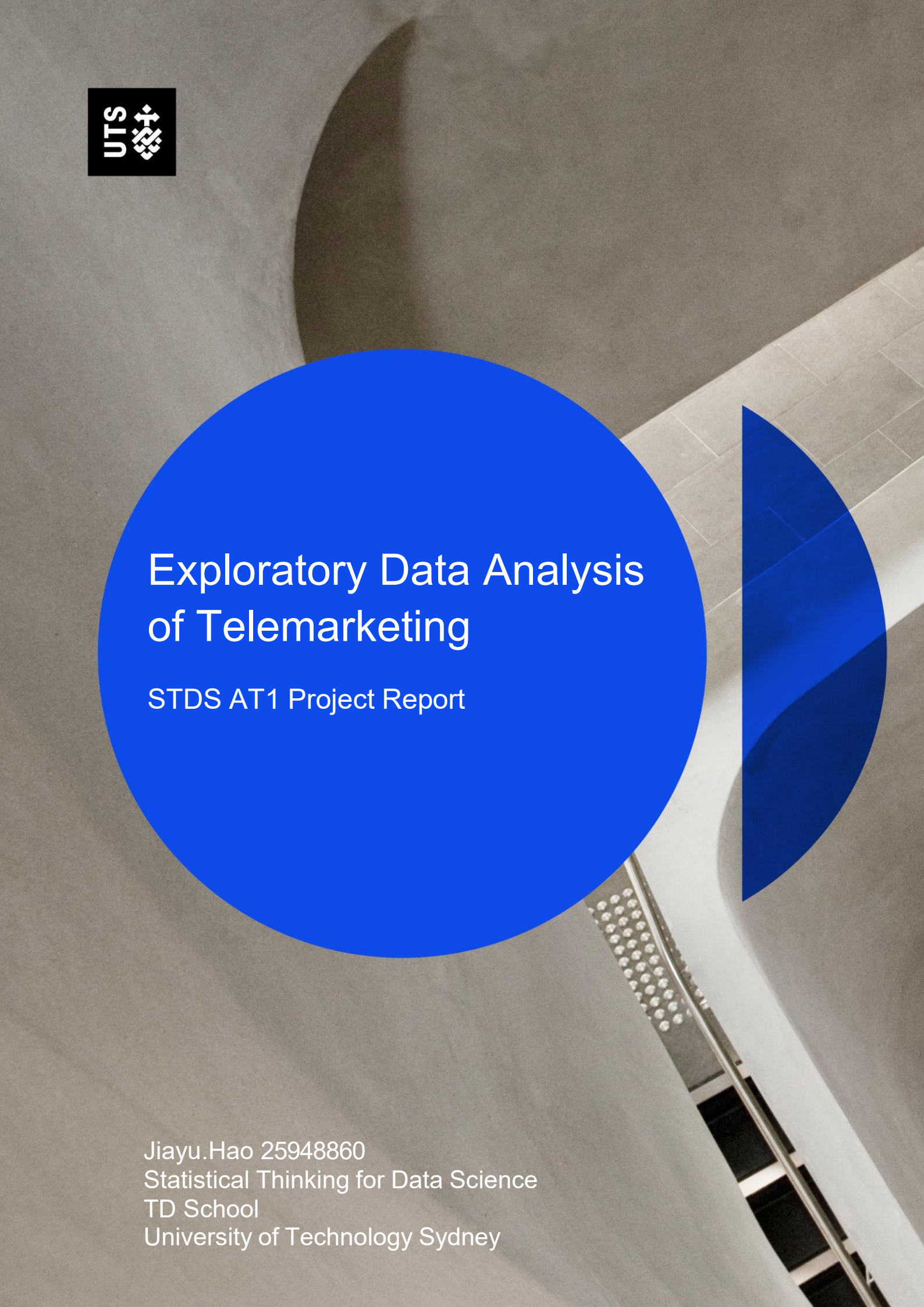

Exploratory Data Analysis of Telemarketing

STDS AT1 Project Report

Jiayu.Hao 25948860
Statistical Thinking for Data Science
TD School
University of Technology Sydney

UTS CRICOS 00099F

Executive summary

Telemarketing is an effective way to increase subscriptions, but its success is influenced by many factors. This report analysis company telemarketing data to identify the main drivers of subscription outcomes.

After data cleaning, this project applied univariate analysis, bivariate analysis, and correlation analysis to examine how customer traits and external factors influence telemarketing subscription outcomes.

The results show that 65.06% of previous subscribers subscribed again during telemarketing. Students had a notably higher rate of 31.43% than other people. Choosing the right channel and timing is also an important factor influencing subscription outcomes.

The conclusion is that telemarketing should target high-response customer groups, use repeated calls via mobile phones, and adjust campaign scale according to economic indicators.

65.06% of previous subscribers subscribed again during

Choosing the right channel and timing influence subscription outcomes.

Table of Contents

EXECUTIVE SUMMARY **2**

TABLE OF CONTENTS **3**

INTRODUCTION **5**

PROBLEM STATEMENT 5
RATIONALE 5
PROJECT AIMS AND OBJECTIVES 5

METHODOLOGY **8**

METHODS OVERVIEW 8
METHODS DETAILS 8

RESULTS **11**

KEY FINDINGS 11
IN-DEPTH RESULTS 11
Do customer characteristics(such as age, education, job type, and financial status) significantly affect the success rate of telemarketing? *11*
Do economic indicators and marketing channels influence telemarketing subscription outcomes? *12*
Based on the key factors identified, can the likelihood of a customer's subscription be predicted? *12*

CONCLUSION **14**

TAKE HOME MESSAGE 14
DISCUSSION 14
PROJECT LIMITATIONS AND CAVEATS 14
STAKEHOLDER ANALYSIS AND PROJECT OUTCOMES 14

ACKNOWLEDGEMENTS **16**

REFERENCES **17**

APPENDICES **18**

Introduction

Introduction

Problem statement

Telemarketing is a widely used strategy to drive paid subscriptions. However, its success depends on multiple customer and market factors. This report analyses a telecom marketing dataset to uncover the key factors influencing subscription uptake and to highlight customer segments most likely to respond.

Rationale

Maximizing return on investment is central to telemarketing. By focusing on customers most likely to respond, costs can be reduced without compromising campaign success. For instance, in a charity campaign, cutting call volume by 35% maintained ROI while improving efficiency (Dataro, 2025). Moreover, telemarketing ROI depends on conversion rates, investment costs, and customer value (Intelemark, 2025).

Using EDA on telecom subscription data, this study aims to identify high-response customer segments and design more cost-effective marketing strategies.

Project aims and objectives

The purpose of this research is to analyse a telecom marketing campaign dataset to identify key factors influencing customer subscription decisions. It aims to use customer information (e.g., job, age, loan status) to find groups more likely to subscribe; to examine how objective factors (e.g., economic indicators, marketing channels) affect success rates in order to choose suitable times and channels and reduce costs; and to combine customer and external factors to predict individual subscription likelihood, enabling the company to target the right customers at the right time.

To support our project aim, we asked the following questions:

1. Do customer characteristics(such as age, education, job type, and financial status) significantly affect the success rate of telemarketing?
2. Do economic indicators and marketing channels influence telemarketing subscription outcomes?
3. Based on the key factors identified, can the likelihood of a customer's subscription be predicted?

To address the above research questions, the selected dataset should undergo exploratory data analysis (EDA), data cleaning, and essential statistical evaluation.

To answer these questions we had the following objectives:

1. Data Cleaning: Dataset overview and data quality check
2. Data Cleaning: Identification of missing values and data types
3. Univariate Analysis: Distribution of categorical features (e.g., job, education, marital status)
4. Bivariate Analysis: Relationship between customer attributes and subscription outcomes
5. Bivariate Analysis:Impact of economic indicators and marketing channels on success rates
6. Feature correlation analysis: Detection of key factors influencing subscription decisions

Our results show that certain customer characteristics are linked to telemarketing success rate. Among external factors, the marketing channel affects telemarketing subscription outcomes, and weaker economic conditions are associated with lower success rates of telemarketing.

Methodology

Methodology

Methods overview

The dataset was cleaned and reformatted to ensure it could be applied to statistical analysis. The research uses univariate analysis to examine individual features, bivariate analysis to explore relationships between pairs of features, and correlation analysis to assess feature interactions. These methods support the research objectives and help identify key factors influencing subscription outcomes.

Methods details

The following paragraphs outline the research methodology in detail, covering all steps of the exploratory data analysis.

- **Data Cleaning:**
  - The dataset was loaded into Python and structured into a *41,180 × 21* DataFrame.
  - Delimiters such as quotation marks (‘"’) and semicolons (‘;’) were removed
  - Numerical features were converted into numeric values to allow statistical analysis.
- **Univariate Analysis:**
  - Numerical and categorical features were examined separately.
  - Histograms were used to show the distribution of numerical features(Figure 1).
  - categorical features(Figure 3) were summarized by frequency counts.
  - Binary categorical features were transformed into numeric form for further analysis.
  - The target variable(Figure 2), “subscription to the plan (‘y’)” is a binary categorical feature with an imbalanced distribution.
  - Ordered categorical features were transformed into numeric form for further analysis(Customer education level is ordinal, ranging from no formal education to professional training and university education).
- **Feature Correlation Analysis:**
  - To identify the key factors influencing subscription, numerical features were analysed for correlations and visualized in a heatmap(Figure 4).
  - To identify numerical features with stronger relationships in the

heatmap, the correlation coefficients with the target variable were reordered(Figure 5).

Some notable results show that customer features such as “number of contacts before this campaign(*‘previous’*)”, “previous campaign success(*‘poutcome’*)”, and “last contact duration(*‘duration’*)” are positively correlated with subscription success, while “days since last contact(*‘pdays*’)” is negatively correlated. Among economic indicators, “quarterly employment variation rate(*‘emp.var.rate’*)”, “Euribor 3-month rate(*‘euribor3m’*)”, and “number of employees(*‘nr.employed’*)” are negatively correlated with subscription rates. The results provide a basis for deeper analysis.

Note: According to the dataset description, the correlation of the feature ‘duration’ should not be interpreted as a meaningful result.

- **Bivariate Analysis:**
  - Selected features with strong correlations were analysed in more detail.
  - To address the relationship between customer characteristics and subscription outcomes.

  Compare customer attributes with subscription rates.

  The grouped analysis results show that students and retirees have higher telemarketing success rates than other groups and the overall average, at about 31.43% and 25.20%, respectively. Age grouped analysis supports this conclusion. Grouped statistics show that education, marital status, and credit default or disclosure status significantly affect telemarketing success rates. The same technique shows that education, marital status, and credit default or disclosure status significantly affect telemarketing success rates.

  The correlation matrix(Figure 4) shows a significant positive relationship between prior(feature “poutcome”) and current(feature “y”) telemarketing success.

  - To evaluate the impact of economic indicators and marketing channels on subscription decisions.

  Grouped analysis are used to analyse telemarketing channels against subscription success rates, the results show that cellular is more likely to succeed.

  In correlation matrix, three economic indicators are negatively correlated with telemarketing success rates. At the same time, these three features are highly positively correlated with each other (correlation above 0.90, $p < 0.01$). In future work, they may be combined into a single “economic indicator” feature(Figure 6).

Results

Results

Key findings

Certain customer characteristics are related to subscription outcomes in telemarketing. The marketing channel influences success rates of telemarketing, and indicators of weaker economic conditions show negative correlations with telemarketing success rate.

In-depth results

Do customer characteristics(such as age, education, job type, and financial status) significantly affect the success rate of telemarketing?

Grouped statistics show that job type, education, marital status, and credit default or disclosure status significantly affect telemarketing success rates.

By job type(Figure 7), students have the highest success rate at 31.43%, followed by retirees at 25.20%, both above the average rate of 11.26%, with age group analysis supporting this result.

By education, customers with no formal education show a higher success rate of 22.22%. By marital status, married and divorced customers have lower success rates at 10.15% and 10.32%, respectively. For credit default(Figure 8), customers explicitly reporting no default achieve a 12.87% success rate, while those with default show 0%, and those unwilling to disclose show 5.15%. All results are statistically significant according to the Chi-square test of independence ($p < 0.01$).

The result shows that customers who previously subscribed through telemarketing are much more likely to subscribe again. The correlation matrix(Figure 4) shows a significant positive relationship between prior and current telemarketing success (correlation = 0.49, $p < 0.01$). Among customers with a previous subscription, 65.06% subscribed again, compared to 14.24% of those previously contacted without subscribing, and only 8.83% of those never contacted before. These findings are supported by a positive correlation between the number of previous contacts and subscription success (correlation = 0.23) and a negative correlation between days since last contact and success (correlation = -0.32).

Do economic indicators and marketing channels influence telemarketing subscription outcomes?

Economic indicators have a clear impact on telemarketing success rates. In the correlation matrix(Figure 4), “ quarterly employment variation rate (emp.var.rate),” “Euribor 3-month rate (euribor3m),” and “number of employees (nr.employed)” all show negative correlations with subscription success.

Another finding is that telemarketing via cellular shows a higher success rate of

14.73% compared to 5.23% for telephone, as revealed by grouped statistics on the “contact” feature.

Based on the key factors identified, can the likelihood of a customer’s subscription be predicted?

The likelihood of a customer’s subscription can be predicted. Logistic regression is suitable because the target variable is binary (yes/no). The dataset includes numerical and categorical features that can be encoded, and exploratory analysis shows that factors such as job, education, loan status, and economic indicators are significantly related to subscription. With a large sample size and strong correlations, logistic regression provides a reliable model that both predicts subscription probability and explains the influence of key factors, supporting better decision-making.

Conclusion

Conclusion

Take home message

To minimize costs and maximize returns, telemarketing should target high-response customer groups, use repeated calls via mobile phones, and launch large-scale campaigns during favorable economic periods.

Discussion

The conclusions of this project are supported by literature from multiple fields, indicating that the following findings have practical value and merit further validation.

Customers with certain personal characteristics are more likely to subscribe during telemarketing campaigns. Students and retirees have higher subscription rates than other groups, likely due to lower price sensitivity(Norton, A. 2021; Wuppermann, A, 2014).

Psychology suggests that once consumers form habits, they continue subscribing without much reconsideration. This status quo bias helps explain why customers who subscribed before are more likely to subscribe again (Cybertek Marketing, 2024).

Increasing sales call frequency improves sales, service perception, value perception, and satisfaction, meaning that even failed attempts can raise the chance of future subscriptions compared to no prior contact (Román & Martín, 2008). Subscription success also differs by channel: mobile calls are more effective than landline calls, matching evidence that people are more prone to impulse buying on mobile devices (Alcántara, 2024).

And with rising living costs, consumers are more sensitive to subscription expenses. This may explain why poor economic indicators, marital status, and credit default are negatively linked to subscription decisions (FT Strategies, 2025).

These findings show that personal traits, marketing channels, and economic conditions all play important roles in telemarketing success and can guide further efficient strategies.

Project limitations and caveats

One limitation is that the dataset comes from a single company, so the results may not apply to other companies or industries. Including data from other sectors could make the findings more general, as such data would be broader and more representative. However, this would reduce accuracy when predicting outcomes for the specific company, which is not desirable. Therefore, the

results should be considered valid only for the company that provided the data or for companies with similar customer groups in the same industry.

Stakeholder analysis and project outcomes

The outcomes of this project are a set of conclusions that offer clear benefits to the company providing the telemarketing dataset. By focusing on customer groups with higher response rates, such as students, retirees, and past subscribers, companies can improve efficiency and reduce costs. Mobile phones should be preferred over landlines, as they deliver higher success rates. Marketing efforts should also consider the economic cycle. Campaigns are more effective when economic indicators are stable, while weak economic conditions may lower response. Following these principles allows companies to target the right customers, use the right channels, and choose the right timing, leading to higher subscription rates and better returns.

References

Dataro. (2025, February 7). *How to cut telemarketing costs using propensity modelling*. Dataro Blog. Retrieved from https://www.dataro.io/blog/how-to-cut-telemarketing-costs-using-propensity-modelling

Intelemark. (2025, April 17). *10 key metrics for measuring telemarketing success and ROI*. Intelemark Blog. Retrieved from https://www.intelemark.com/blog/10-key-metrics-for-measuring-telemarketing-success-and-roi/

Alcántara, A.-M. (2024, June 3). Retailers hate that you buy big things on your laptop. The Wall Street Journal. https://www.wsj.com/tech/personal-tech/retailers-hate-that-you-buy-big-things-on-your-laptop-57cf403c

Cybertek Marketing. (2024). The psychology of subscription-based models: Encouraging long-term customer loyalty and retention. https://cybertekmarketing.com/digital-marketing/the-psychology-of-subscription-based-models-encouraging-long-term-customer-loyalty-and-retention/

FT Strategies. (2025). Subscription economy: A transformed world. https://www.ftstrategies.com/en-gb/insights/the-latest-report-on-the-subscription-economy-a-transformed-world

Norton, A., Cherastidtham, I., & Mackey, W. (2021). Increasing the price of a university degree does not significantly affect enrolment if income-contingent loans are available. arXiv preprint. https://arxiv.org/abs/2102.03956

Román, S., & Martín, P. J. (2008). Changes in sales call frequency: A longitudinal examination of the consequences in the supplier–customer relationship. Industrial Marketing Management, 37(5), 554–564. https://doi.org/10.1016/j.indmarman.2006.12.004

Wuppermann, A., Heiss, F., McFadden, D., & Winter, J. (2014). The price sensitivity of health plan choice among retirees: Evidence from the German social health insurance. Annual Conference of the Verein für Socialpolitik 2014. https://www.econstor.eu/bitstream/10419/100352/1/VfS_2014_pid_758.pdf

Appendices

Appendix A Visualization Data

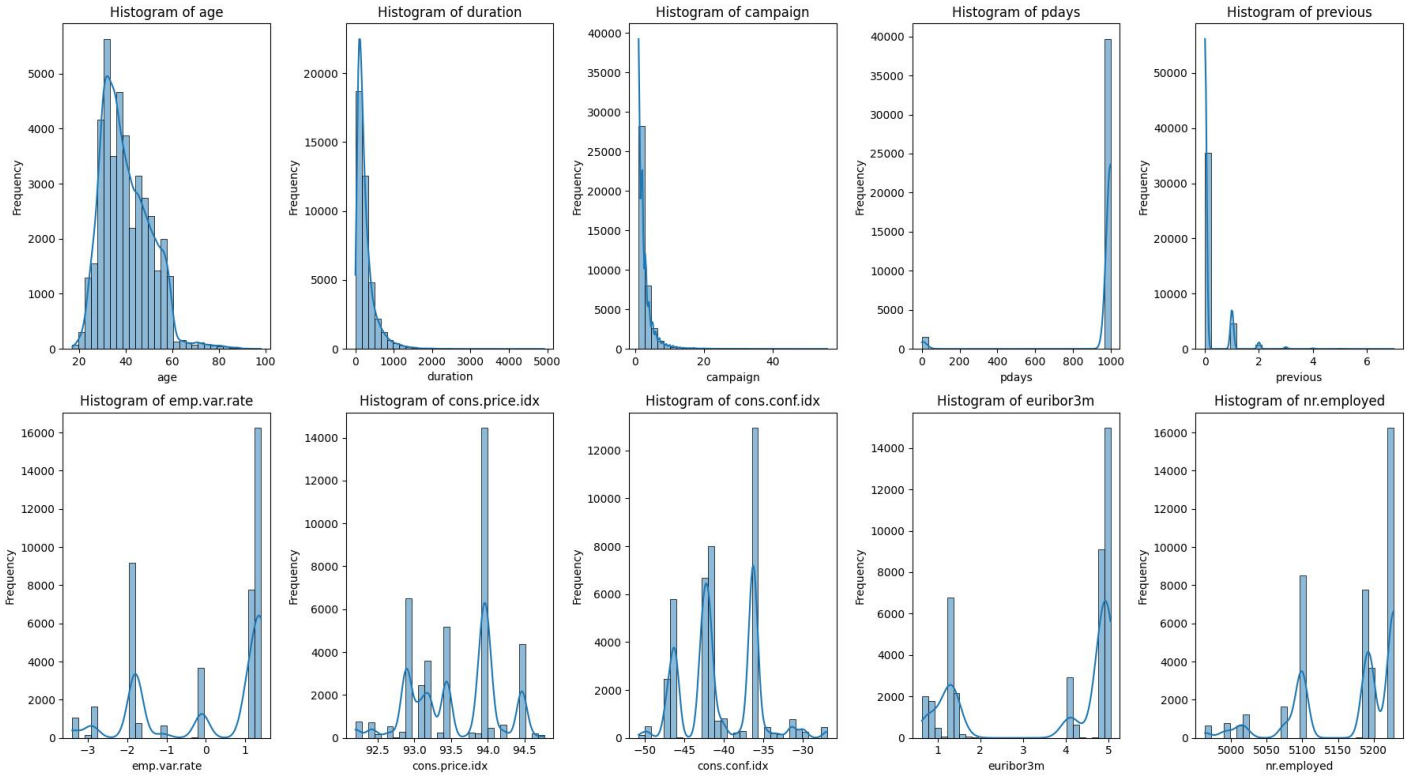

Figure 1

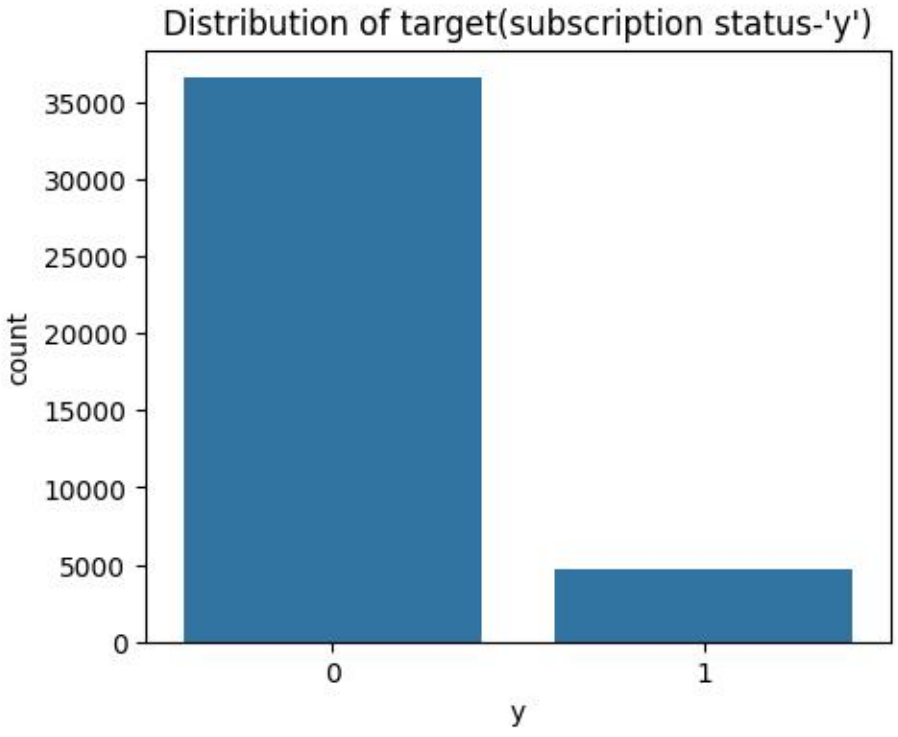

Figure 2

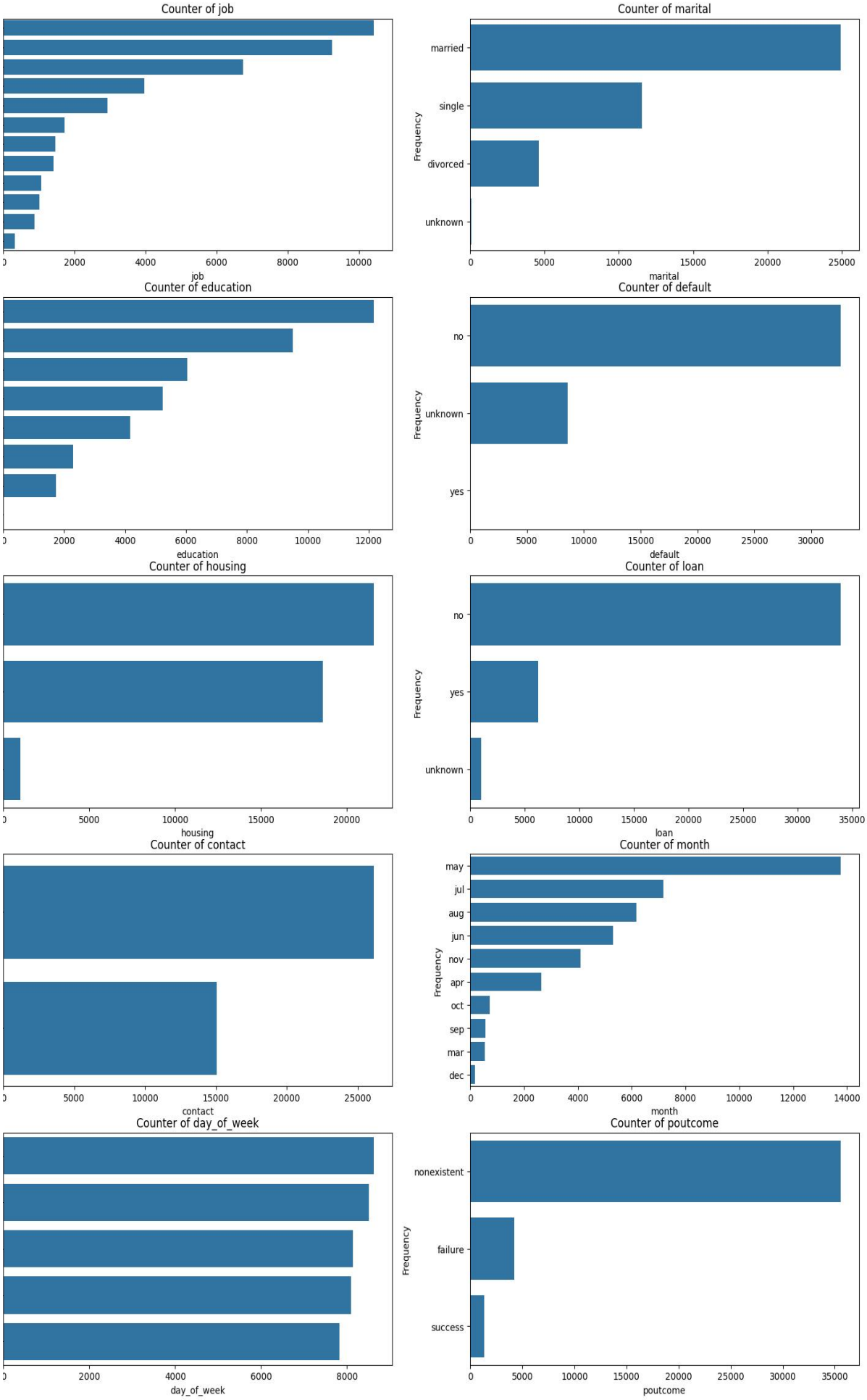
Counter of job
job
Counter of marital
married
single
divorced
unknown
Frequency
marital
Counter of education
education
Counter of default
no
unknown
yes
Frequency
default
Counter of housing
housing
Counter of loan
no
yes
unknown
Frequency
loan
Counter of contact
contact
Counter of month
may
jul
aug
jun
nov
apr
oct
sep
mar
dec
Frequency
month
Counter of day_of_week
day_of_week
Counter of poutcome
nonexistent
failure
success
Frequency
poutcome

Figure 3

Appendix B Correlation

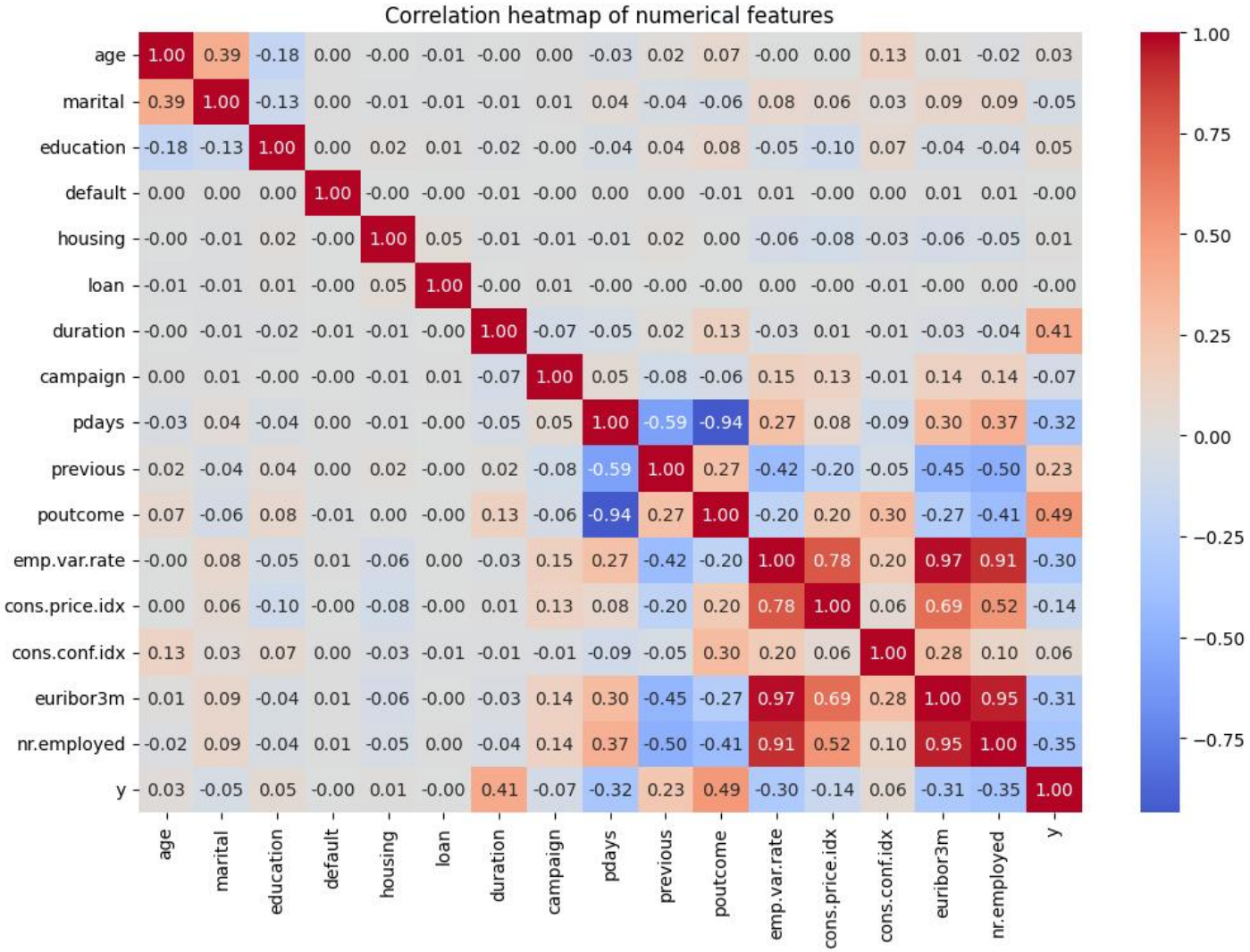

Figure 4

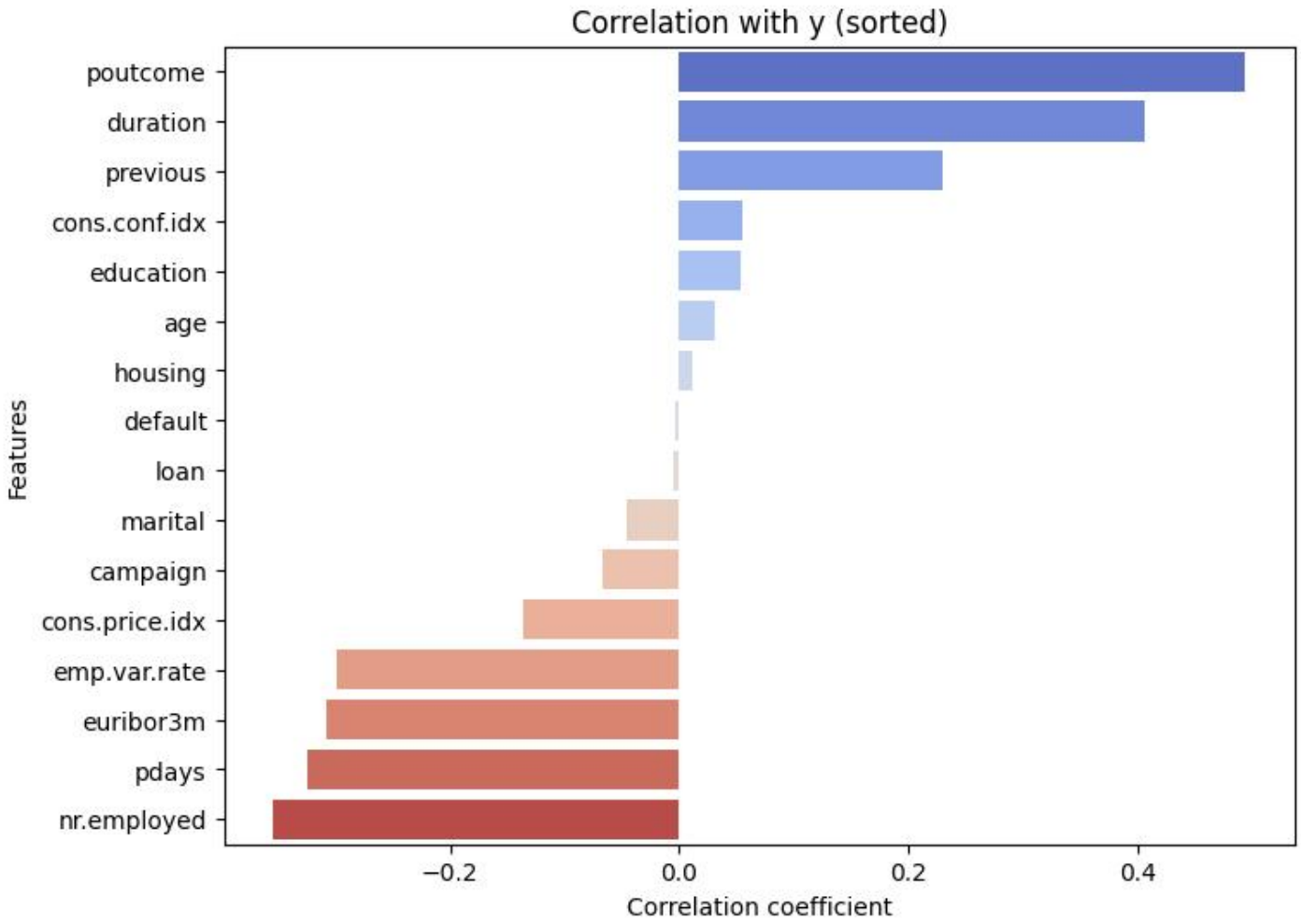

Figure 5

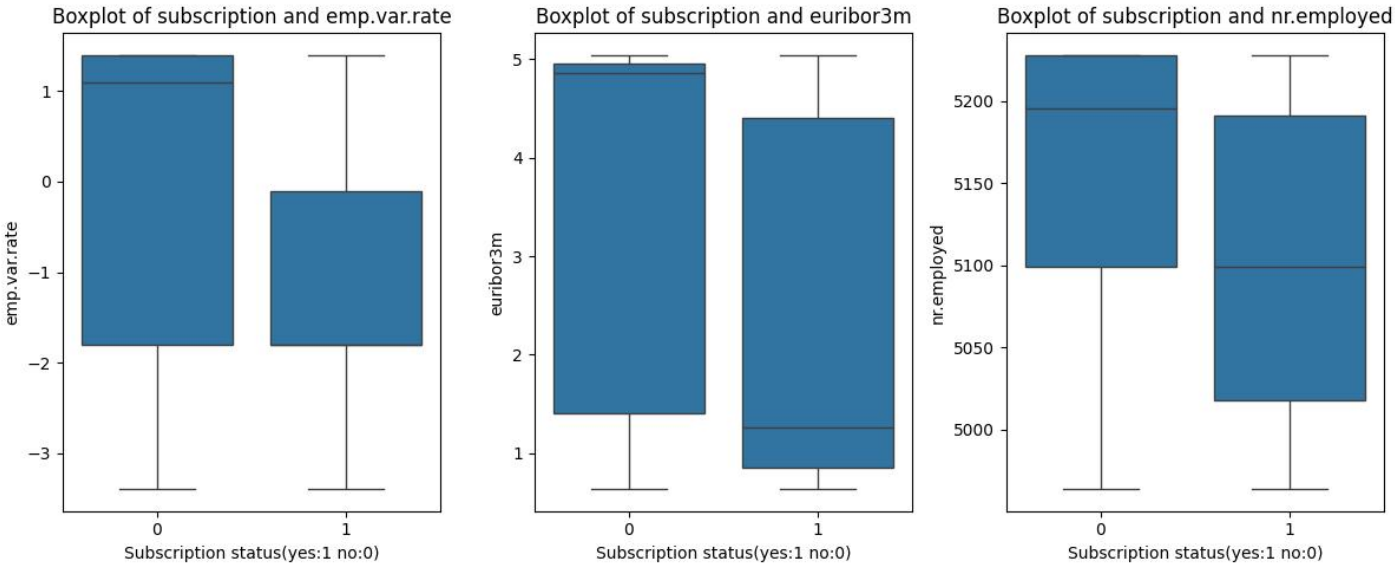

Figure 6 Boxplot of Subscription and Economic indicators

Appendix C Group Statistics Results

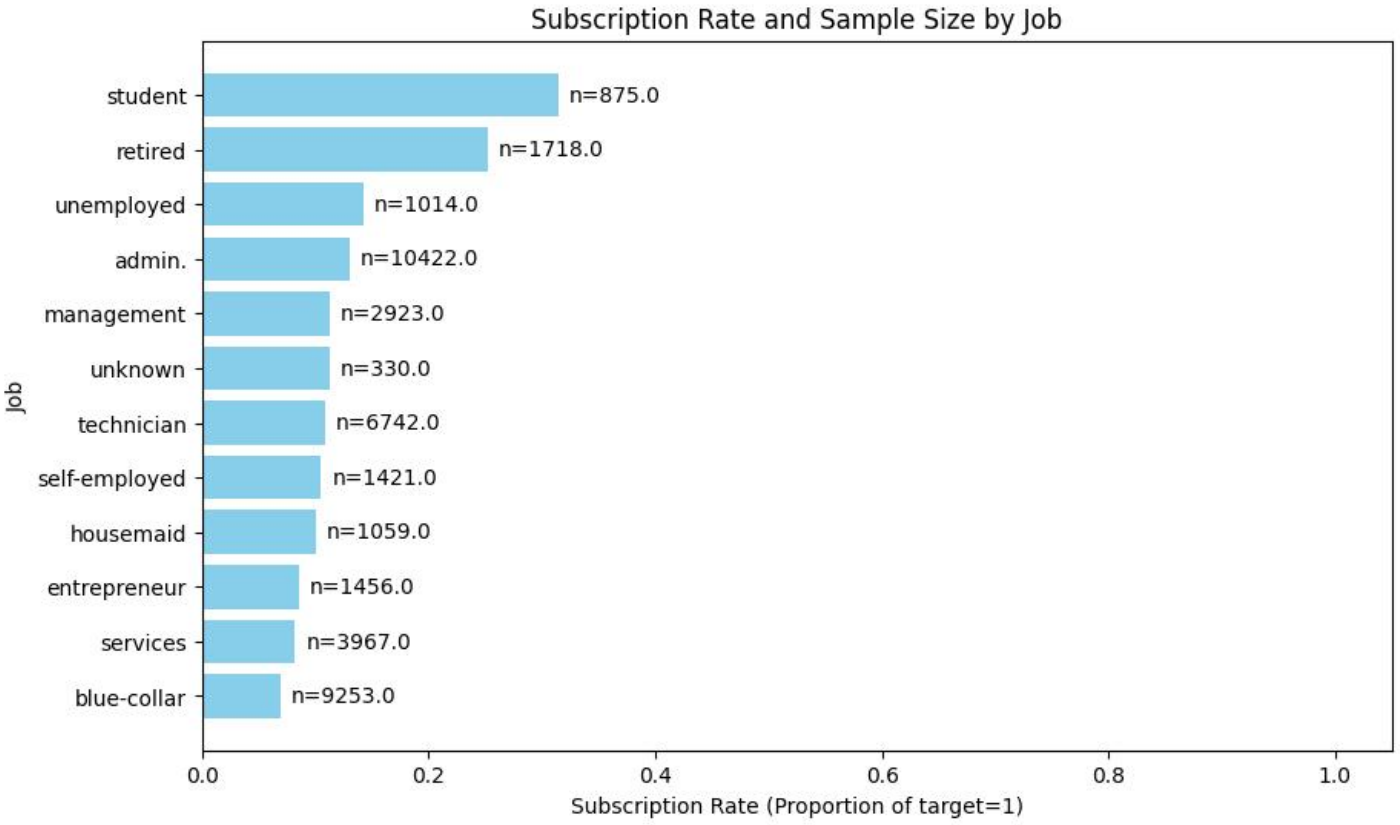

Figure 7

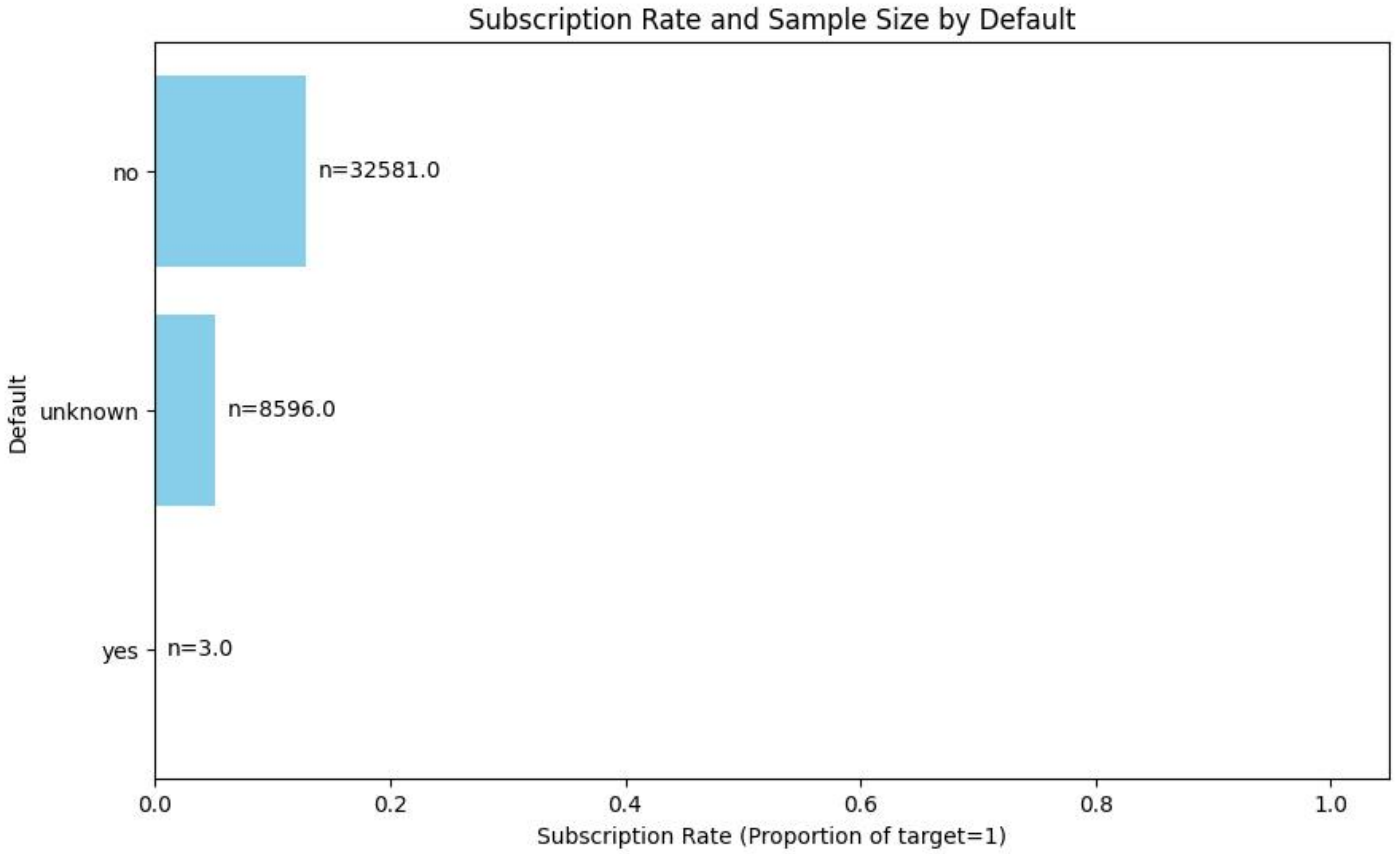

Figure 8