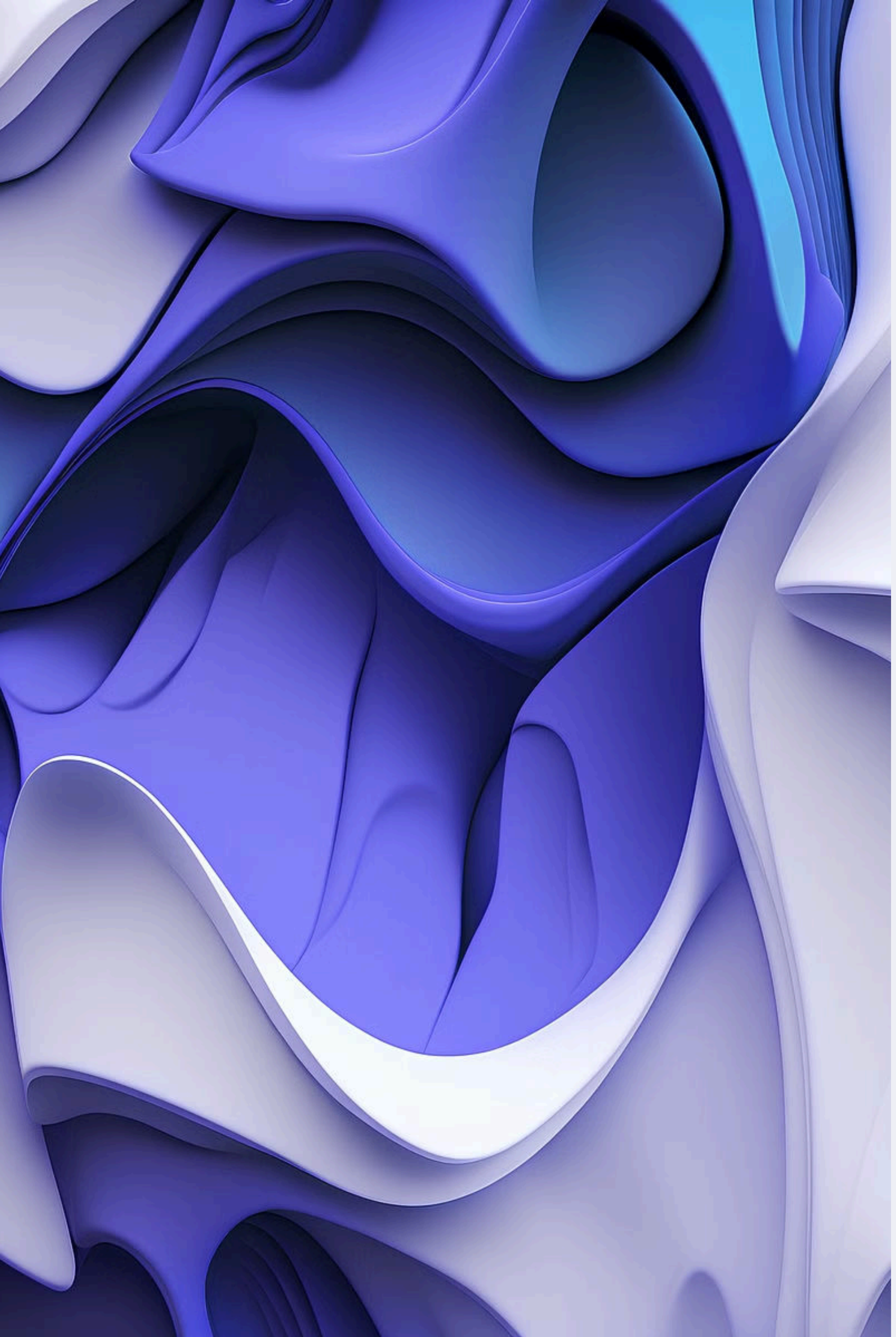
Bank Marketing Analysis

Final Result Interpretation

by Xiyuan Wu

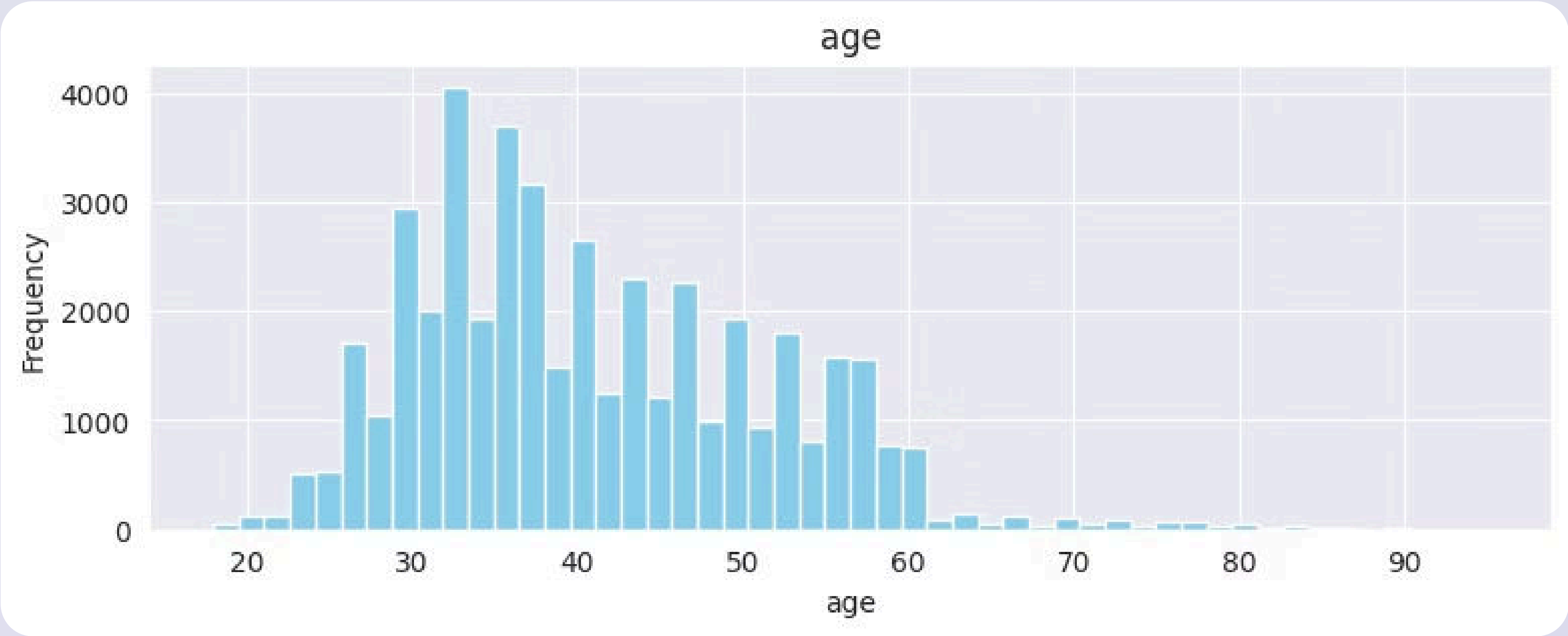
Table of contents



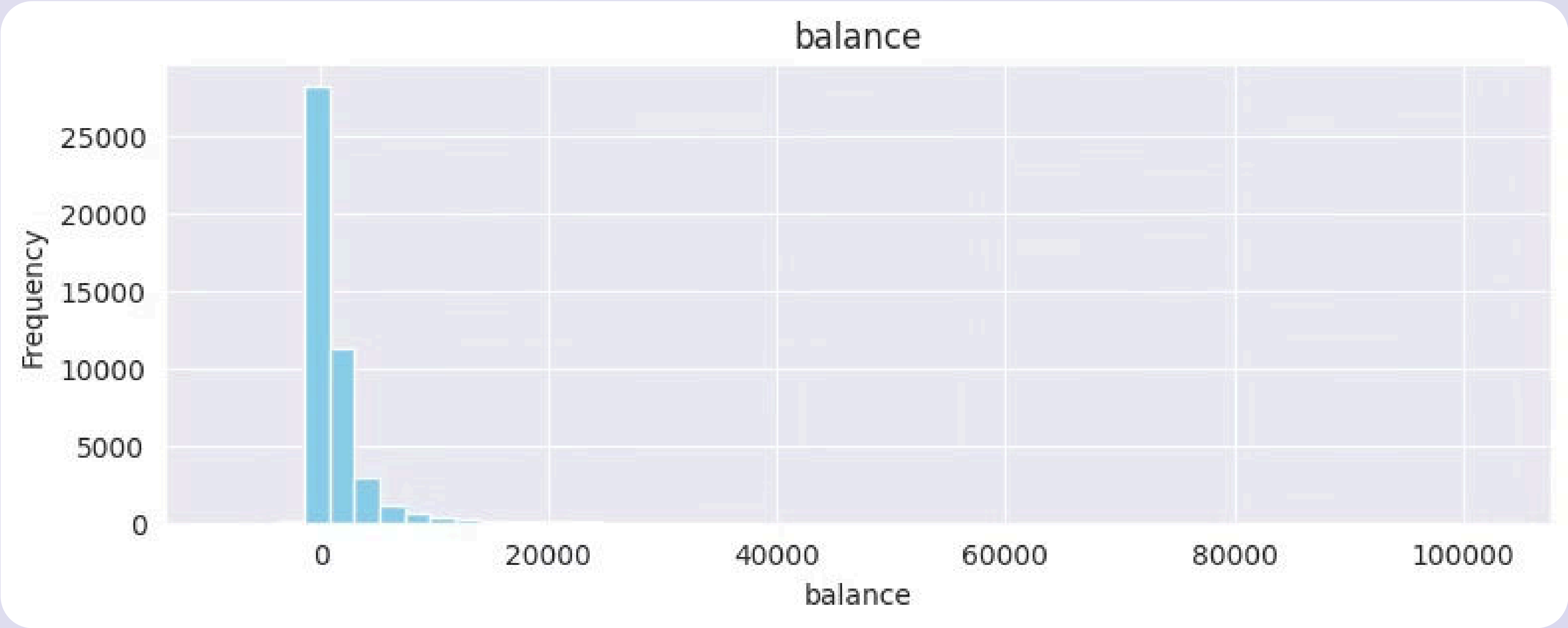
Background Information

The data is related with direct marketing campaigns (phone calls) of a Portuguese banking institution. The classification goal is to predict if the client will subscribe a term deposit (variable y).

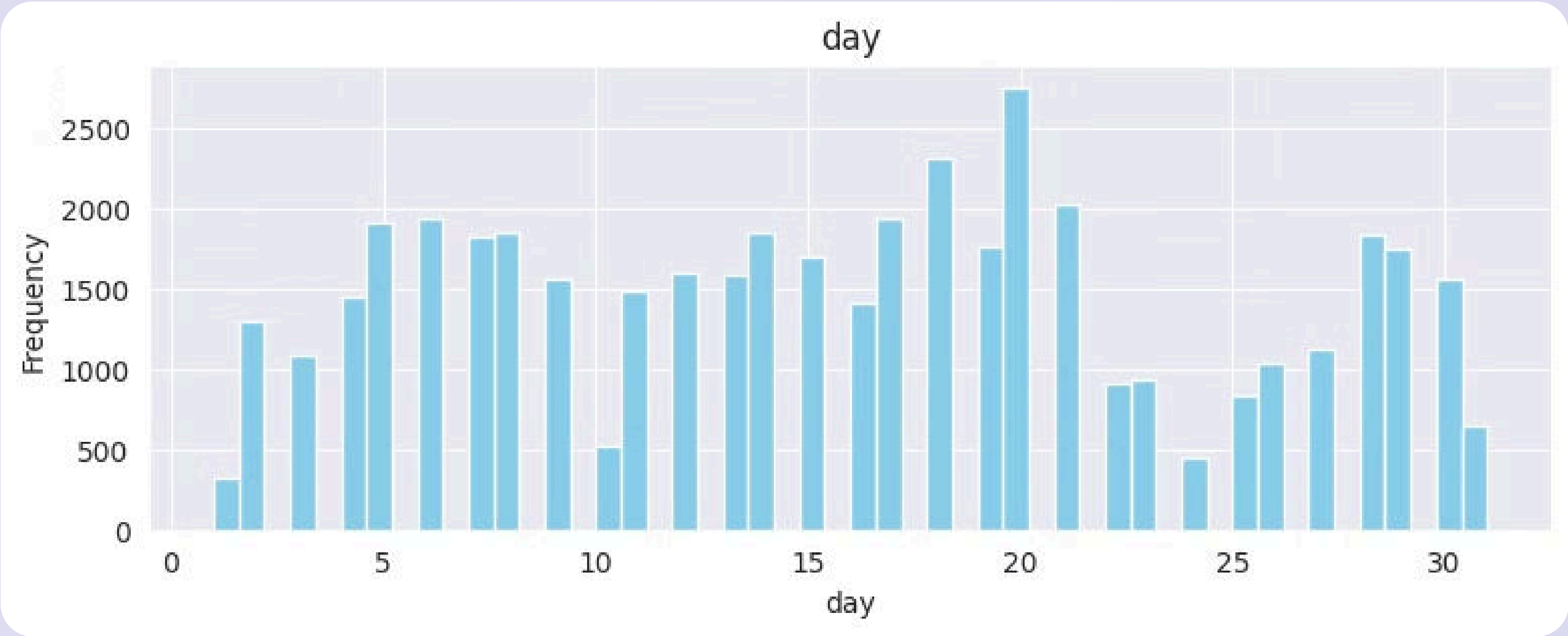
Data Distribution



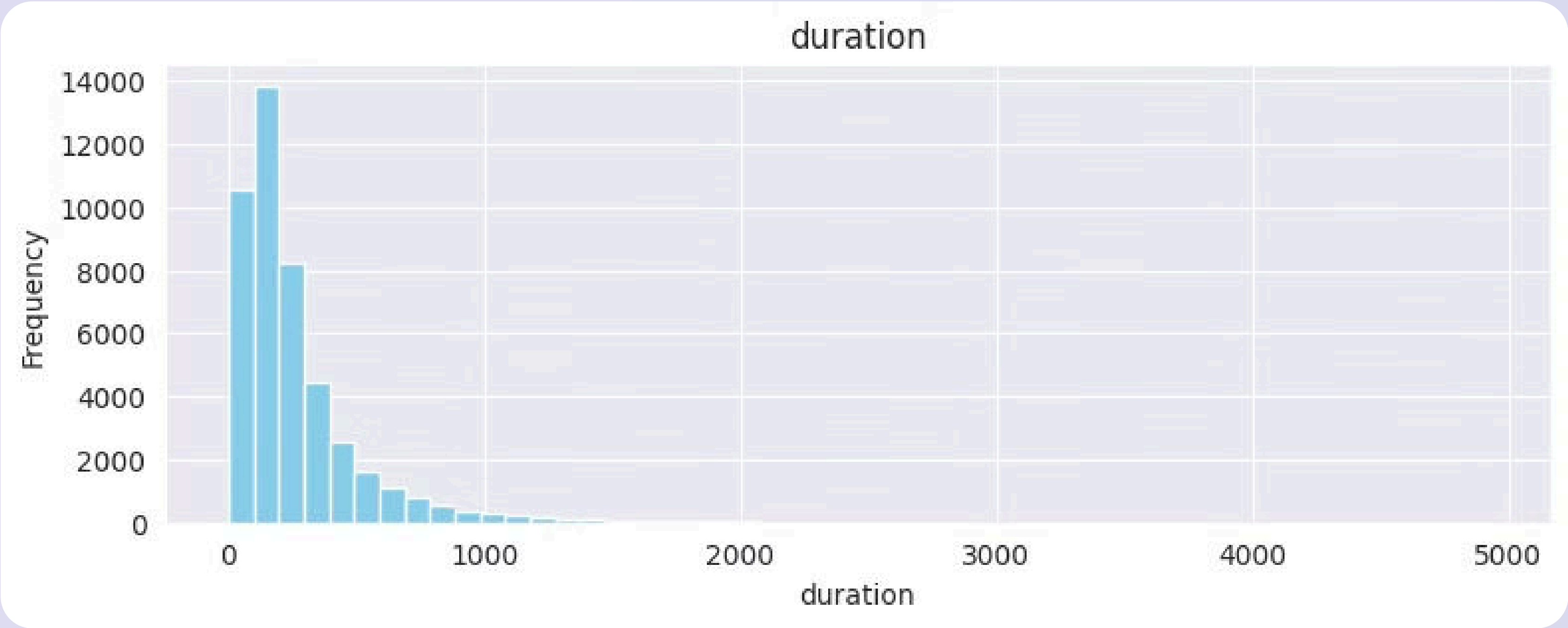
age



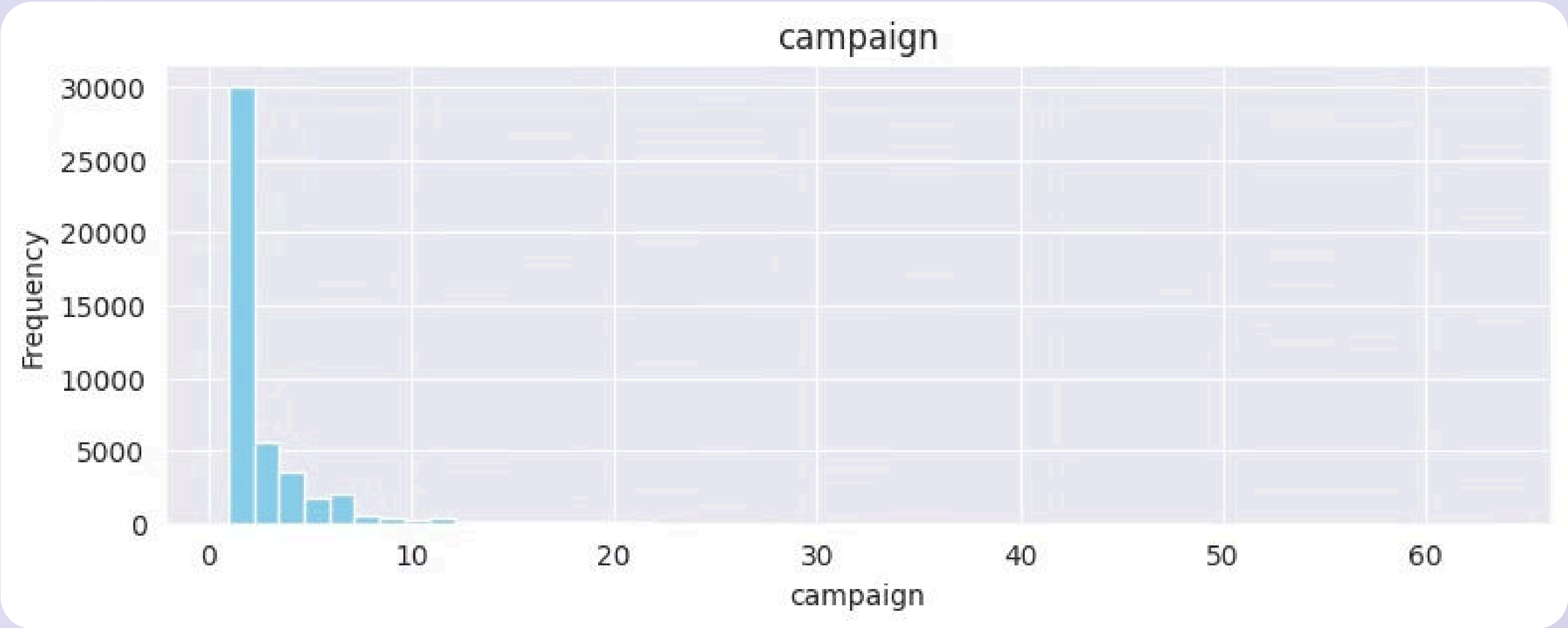
average yearly balance, in euros



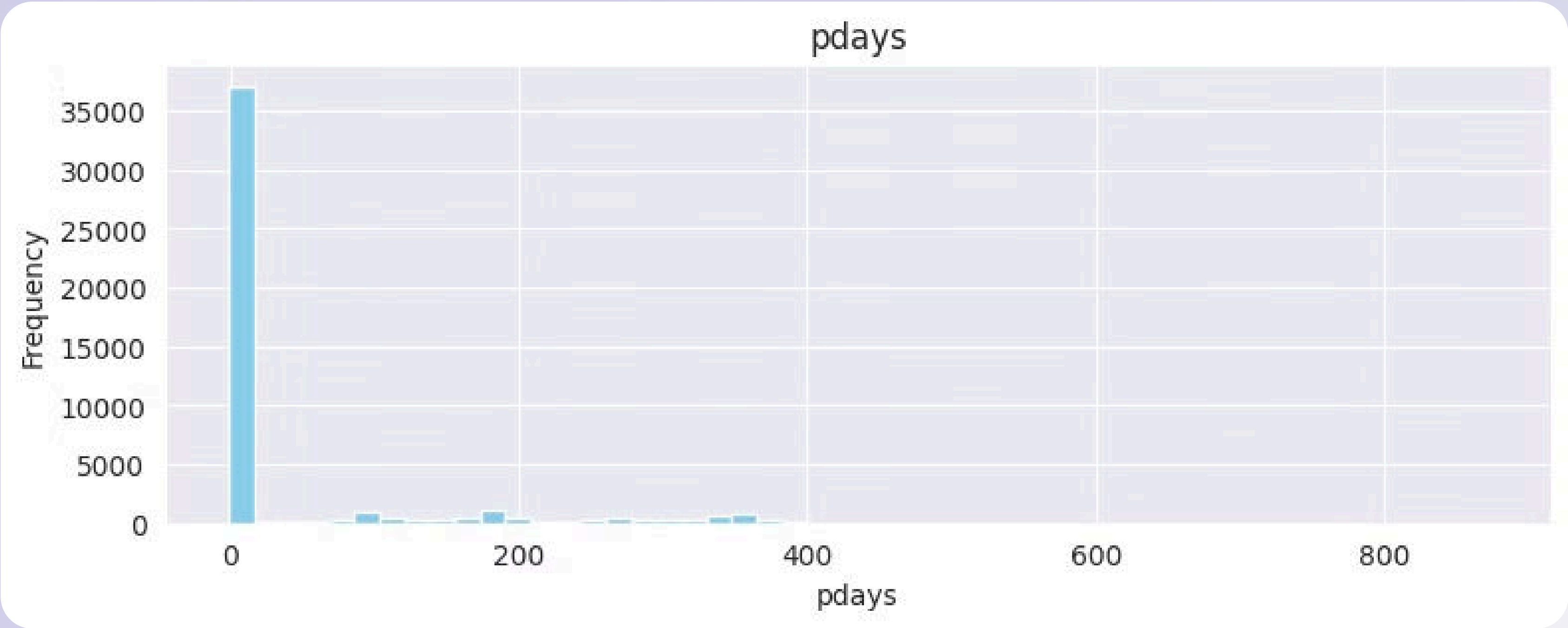
last contact day of the month



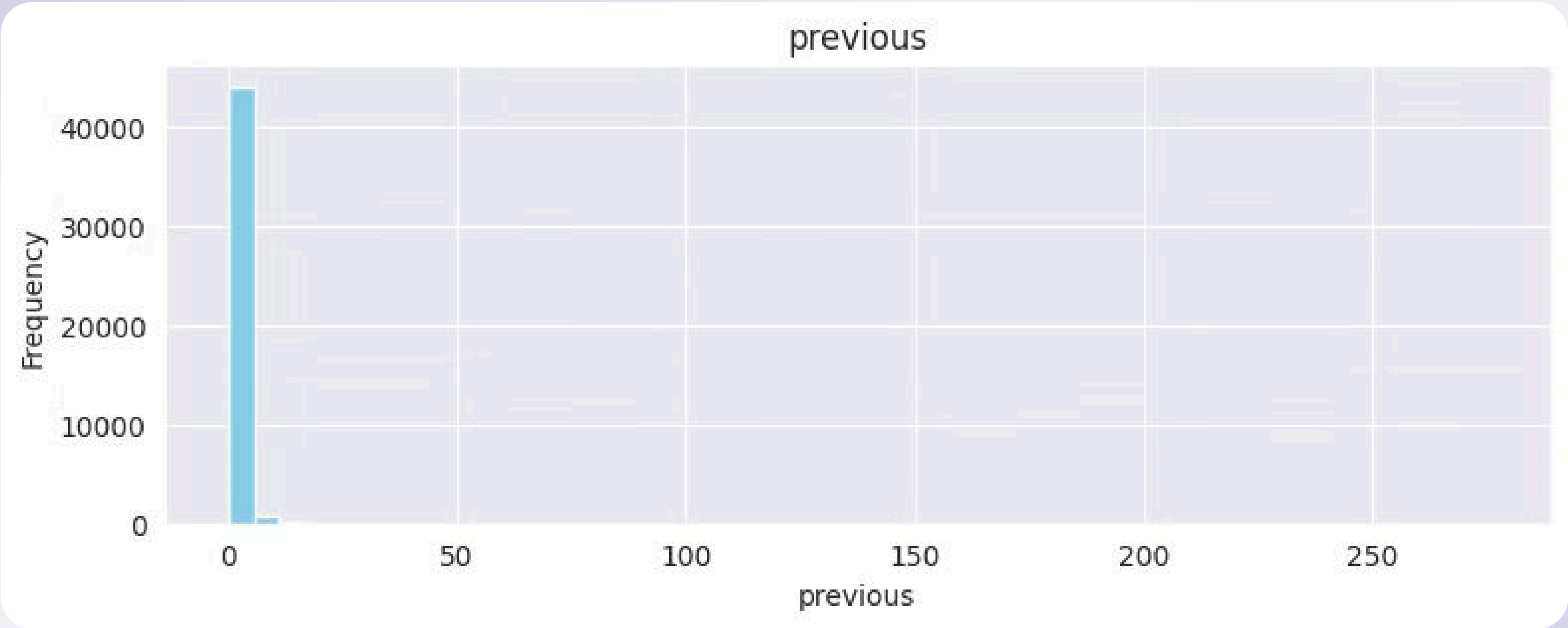
last contact duration, in seconds



number of contacts performed during this campaign and for this client



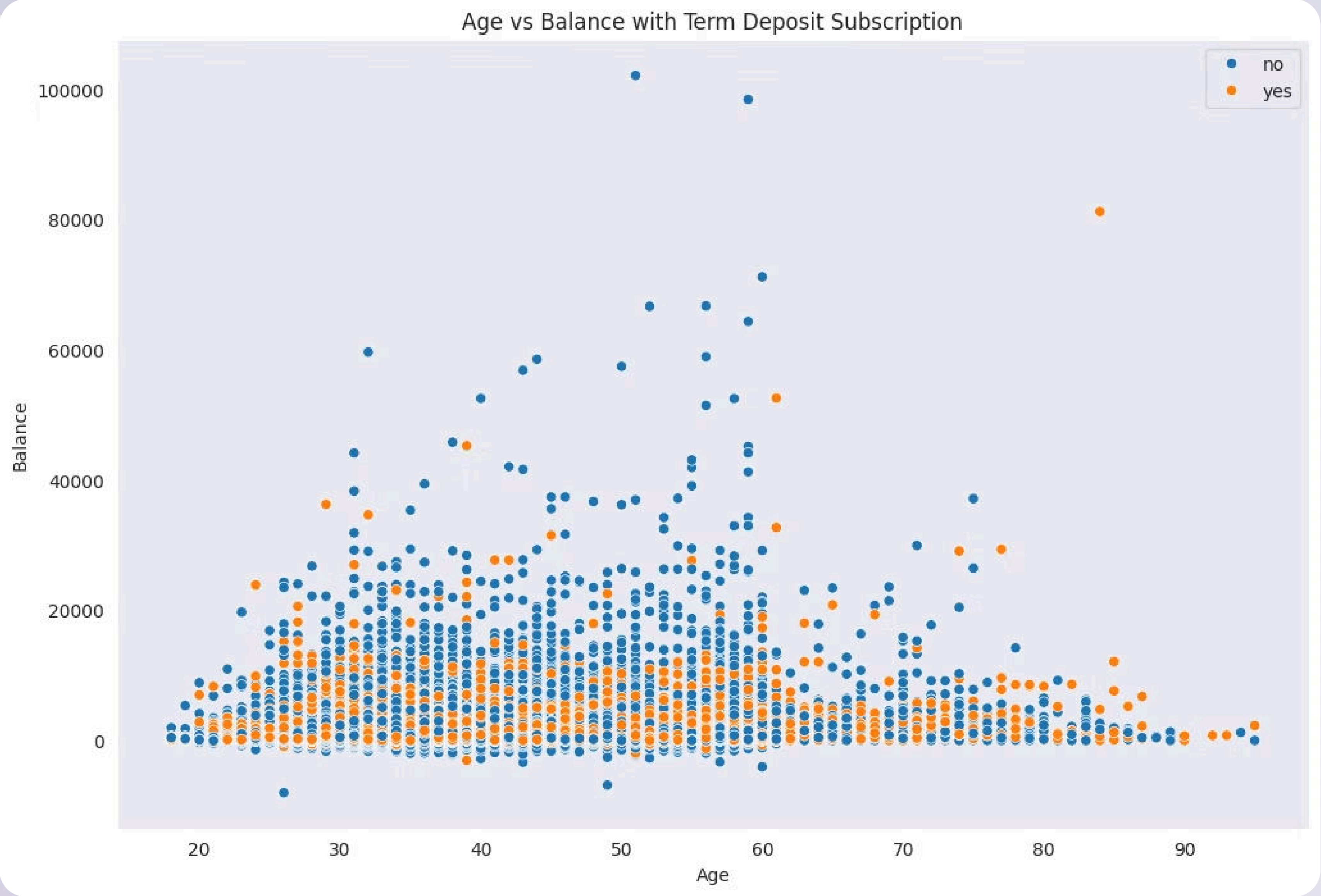
number of days that passed by after the client was last contacted from a previous campaign



number of contacts performed before this campaign and for this client

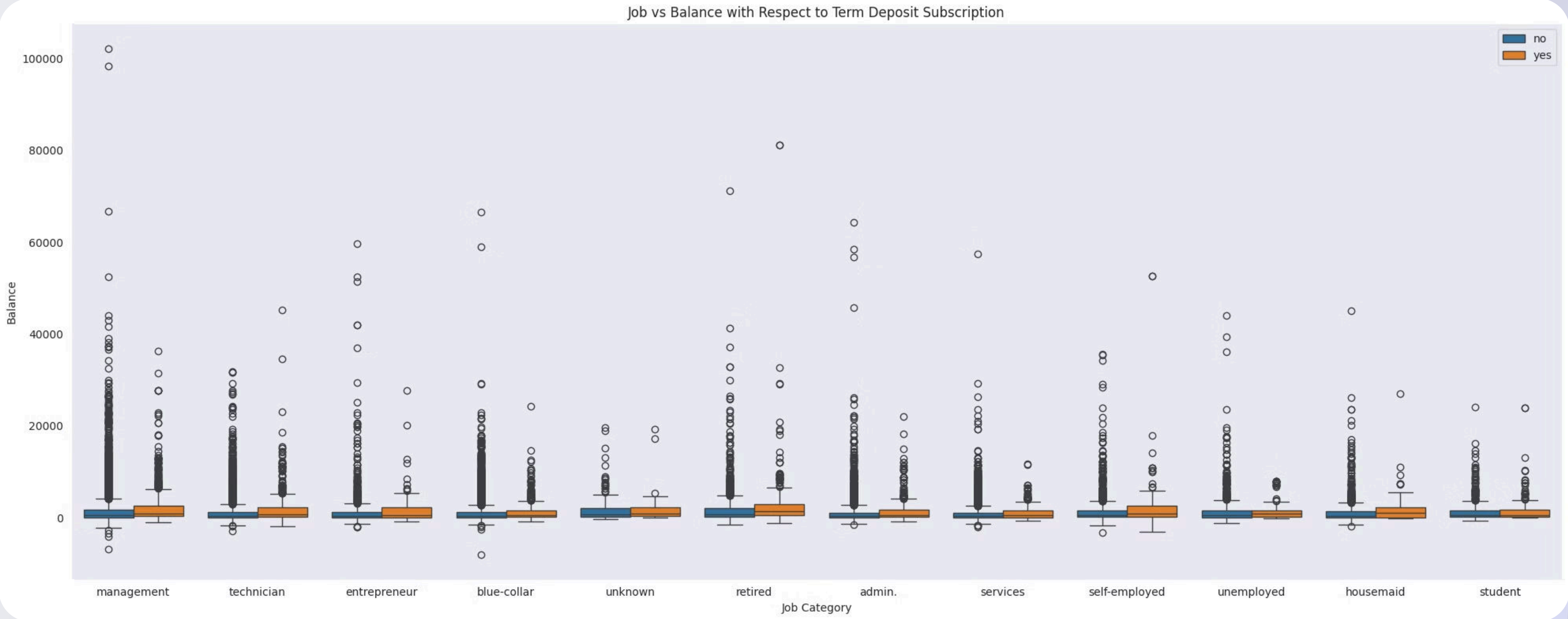
Bank Client Data

Age vs Balance with Term Deposit Subscription



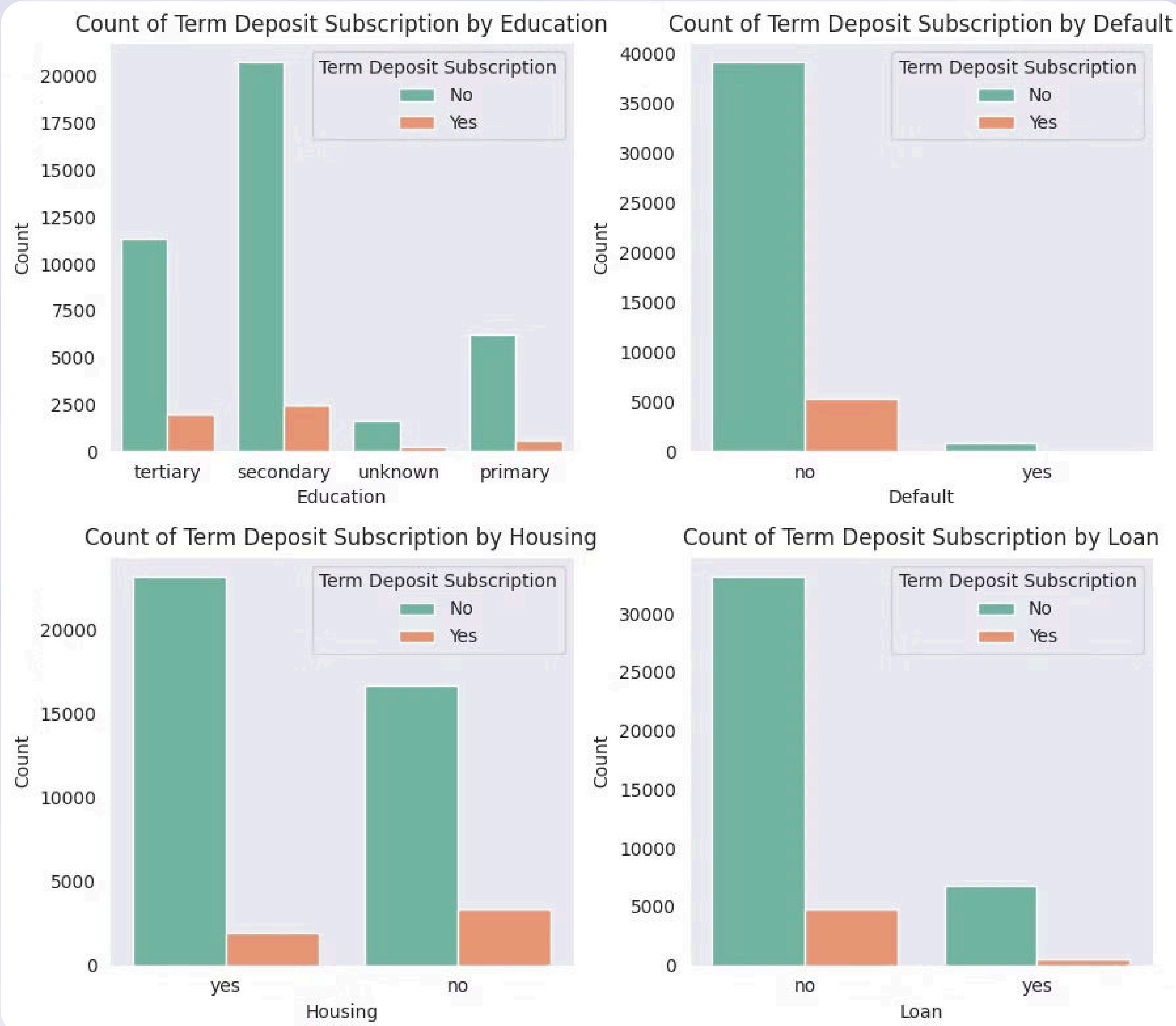
The scatter plot does not reveal an obvious trend between age and balance with regard to term deposit subscriptions. However, it suggests that individuals with a lower balance (less than 20,000 euros) across all age ranges might be inclined to subscribe to a term deposit.

Job vs Balance with Respect to Term Deposit Subscription



From this visual, it does seem that there are no stark differences between the median balances of individuals who subscribed to a term deposit and those who did not, across different job types, and also there is a wide range of balances within each job category.

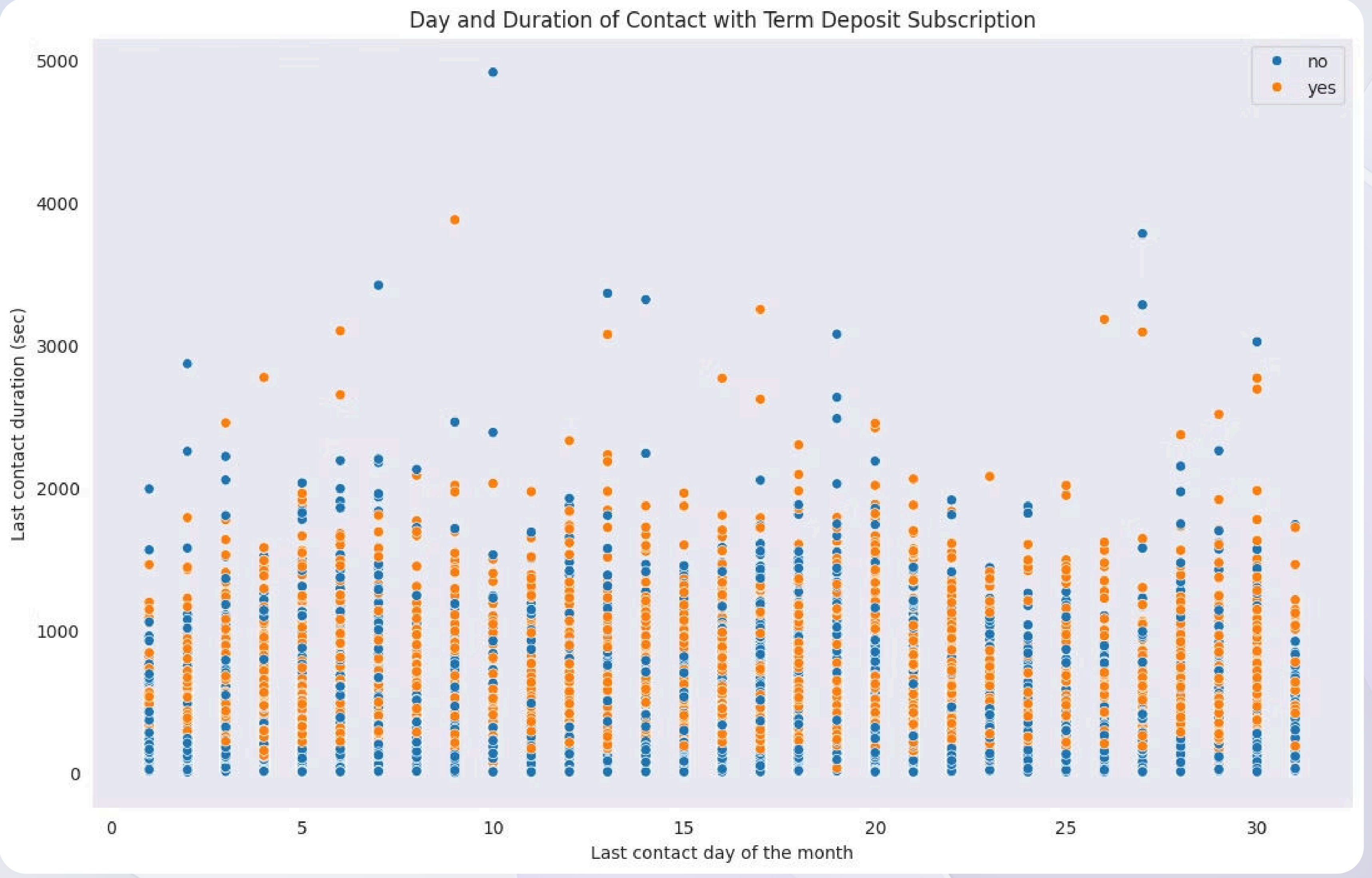
Education, Default, Housing, Loan



The bar chart shows that while there are differences in subscription rates across education levels, credit default status, housing and loan status, the relationship between these and term deposit subscriptions is not very obvious.

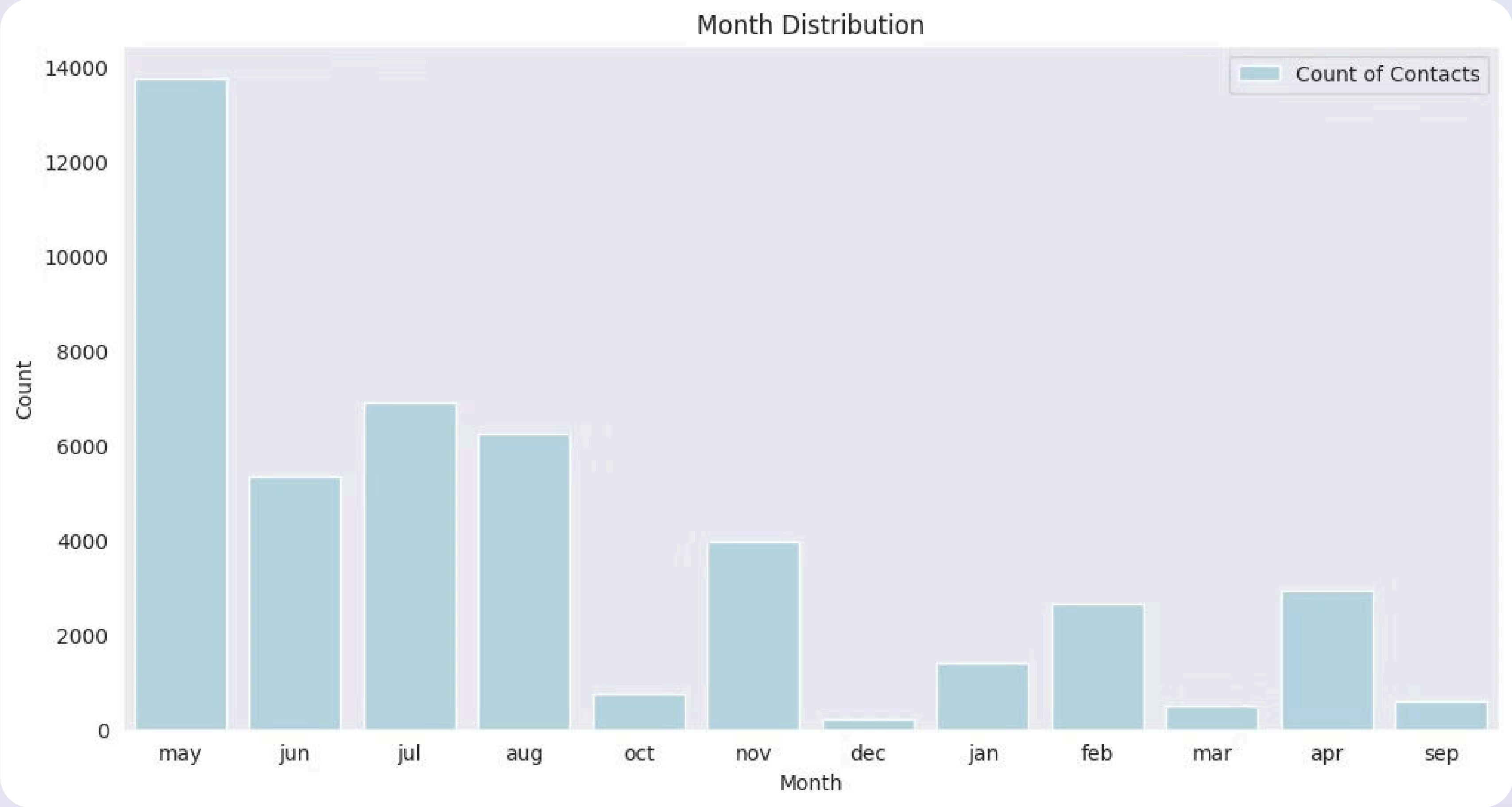
Last Contact of the Current Campaign

Last contact day of the month vs Duration



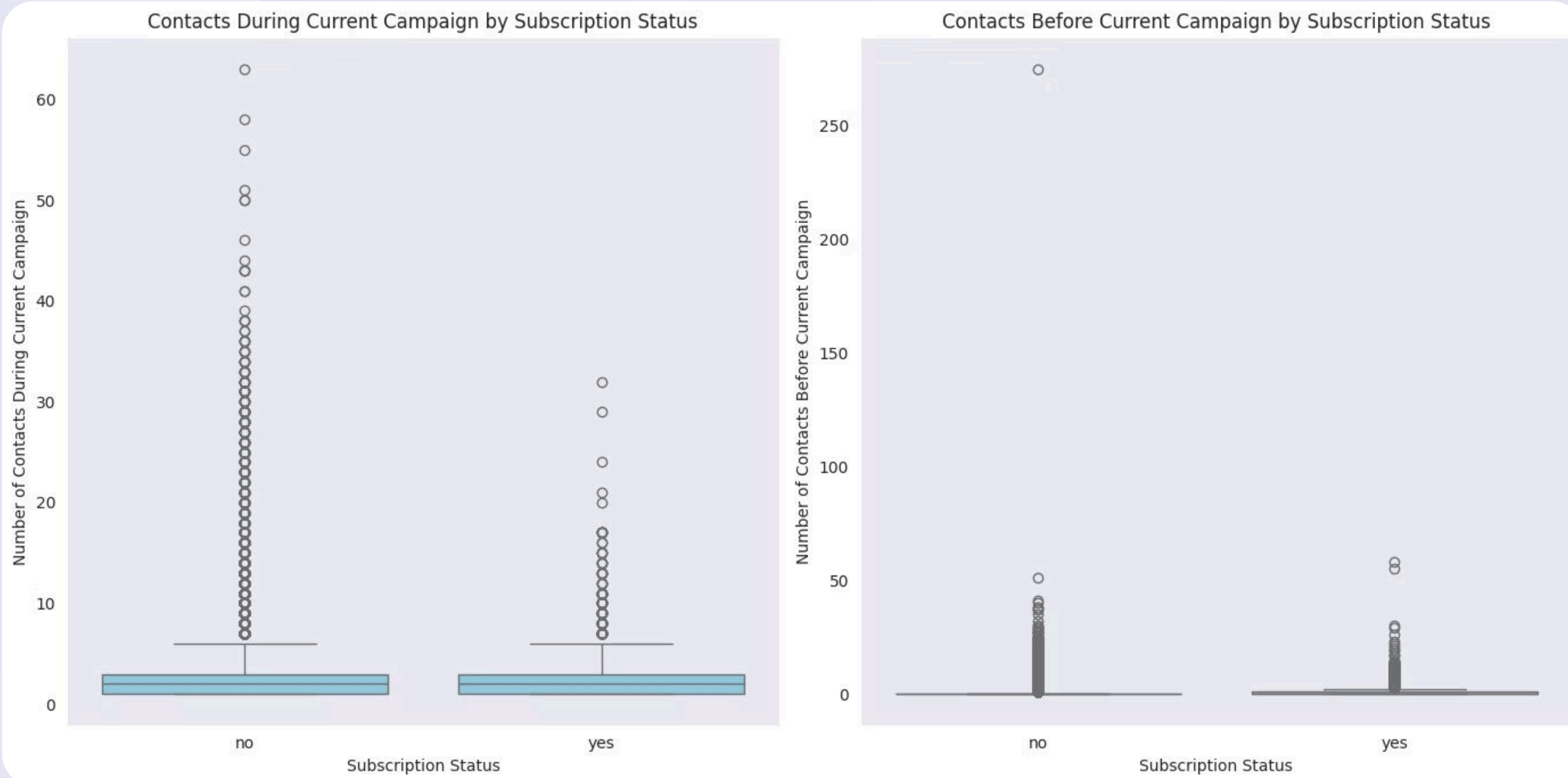
Subscription 'yes' and 'no' are interspersed throughout the month without a visible pattern indicating that contacts made on certain days are more likely to result in a subscription.

Month Distribution



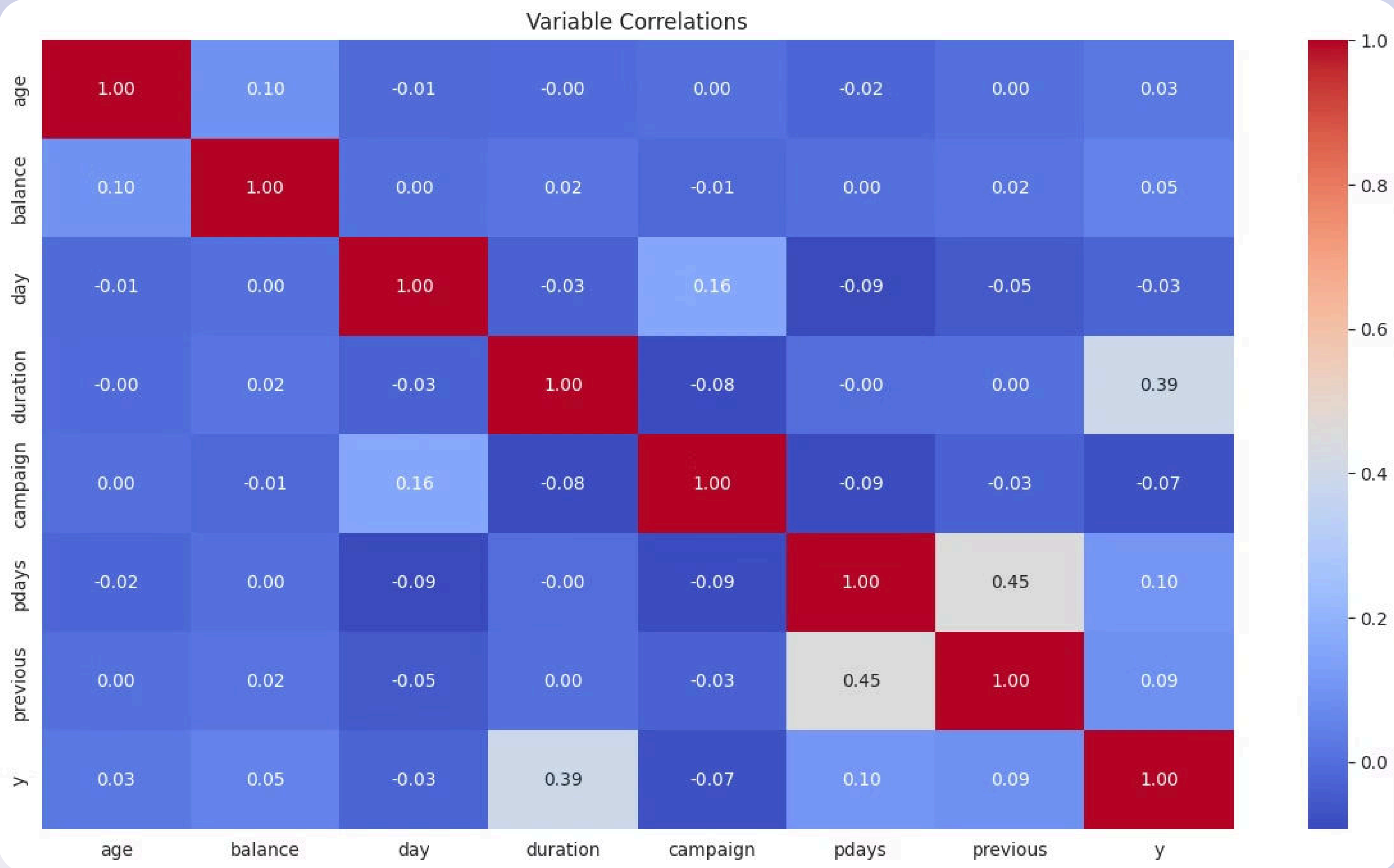
May stands out as the month with the highest number of contacts, approximately double the amount observed in the months with the next highest figures, which are July and August. This significant difference suggests that May was a particularly active period for contacting clients regarding term deposits.

Other Attributes



The graph show that the campaign's approach was fairly consistent in terms of the number of contacts per client but also that making more contacts does not necessarily result in a higher subscription rate.

Overall Correlations



The heatmap shows that none of the two variables have a strong positive or negative relationship.

Model Interpretation

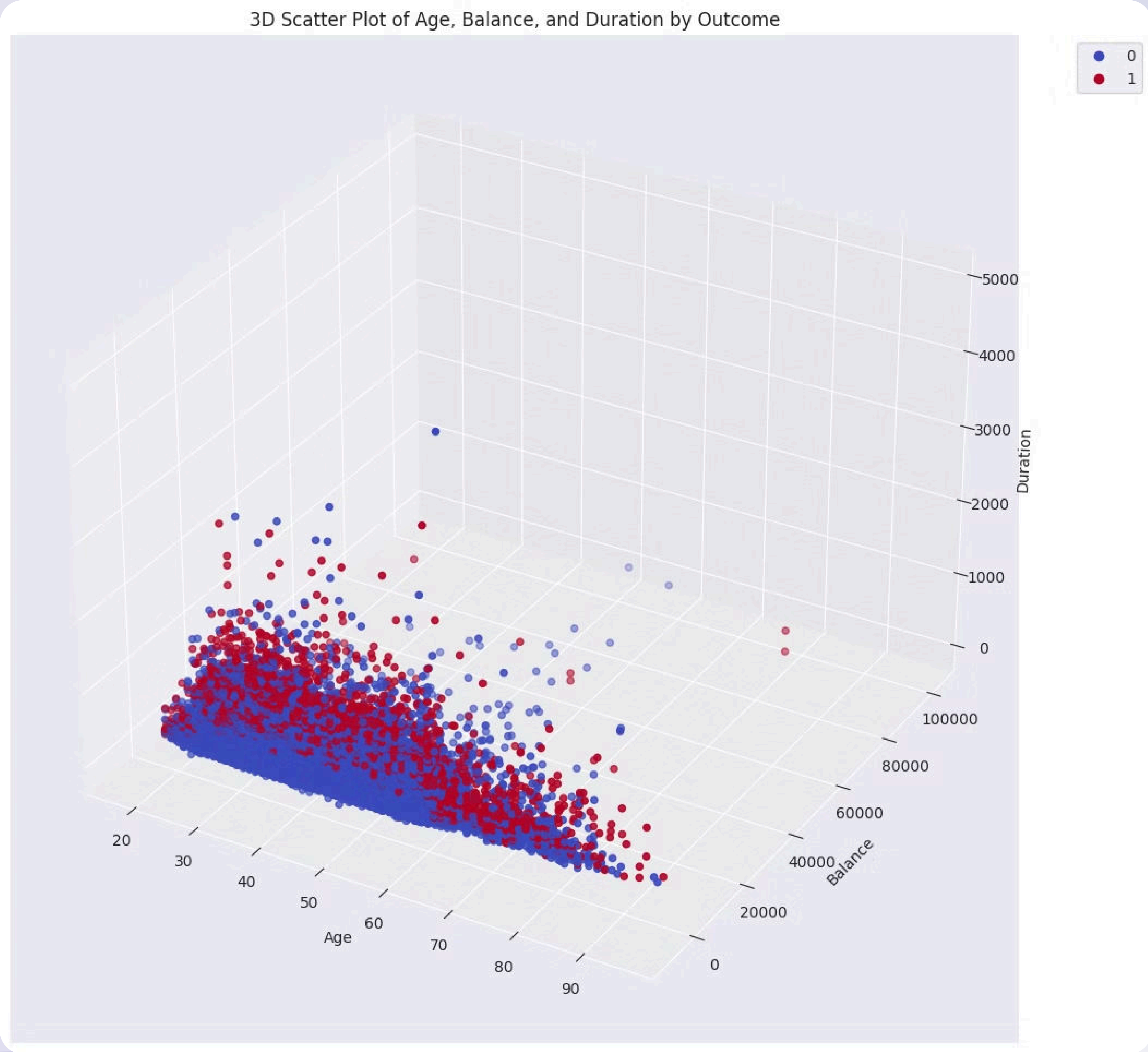
In this analysis, I tried 8 different model, here is their score:

Model	Accuracy	Precision	Recall	F1 Score	AUC-ROC
Logistic Regression	0.8993	0.6257	0.3295	0.4317	0.9081
GBM	0.9048	0.6417	0.4076	0.4985	0.9260
Random Forest	0.9060	0.6520	0.4086	0.5023	0.9274
Decision Tree	0.8826	0.4947	0.5343	0.5137	0.7313
Support Vector Classification	0.9037	0.6630	0.3467	0.4553	-
K-Nearest Neighbors	0.8926	0.5678	0.3152	0.4054	0.8028
XGBoost	0.9081	0.6287	0.5095	0.5629	0.9329
LightGBM	0.9099	0.6449	0.4981	0.5621	0.9361

After we evaluated all models, we found that the **Random Forest** model is the best overall, and that's my recommendation.

Top Feature Affect Term Deposit Subscription

Duration, age, and balance are the top 3 features that affect results. We can try to plot it using a 3D scatterplot with respect to y and see what we can find.



- Most data points, especially those with lower balances and durations, are blue, indicating a high frequency of non-subscriptions within these ranges.
- Red dots are scattered throughout, but appear more frequently over longer periods of time, regardless of age or balance. This suggests that longer durations are associated with larger subscriptions.
- There is no clear distinction between the two categories, and while duration has the greatest impact, it is not the only determinant of subscription status.
- The distributions showed no clear pattern, suggesting that the relationship between these variables and the outcome is not simple or linear.

Conclusion

Focus on call duration

During marketing campaigns, allocate more resources to customers with longer call durations, as duration is the most influential characteristic for predicting subscriptions.

Consider demographics

Pay attention to age and balance as key demographics. Tailored marketing strategies can be developed for individuals of different age groups and levels of balance.

Thank You