
Fiied Terrm me ʃnglesst ór fixed fied Decquts For Fiixxed for Pedsit

Predicting Term Deposit Subscriptions

This report outlines the analysis, insights, and modeling performance for predicting term deposit subscriptions. Key findings from exploratory data analysis (EDA), model evaluation, and recommendations for future marketing campaigns are presented.

REPORT EXPECTATIONS

1. Exploratory Data Analysis which covers :

Key insights from categorical and socio-economic features.

Analysis of customer behavior and communication effectiveness.

2.Hypothesis and Validation.

3. Modeling Performance.

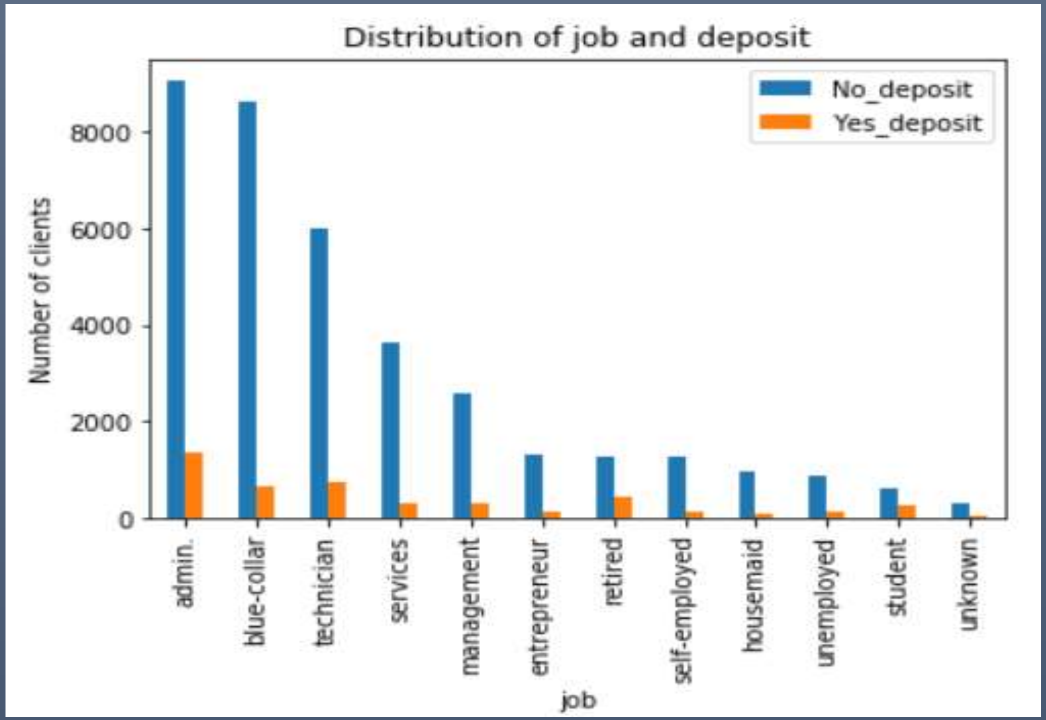
4. Recommendations for improving future marketing campaigns.

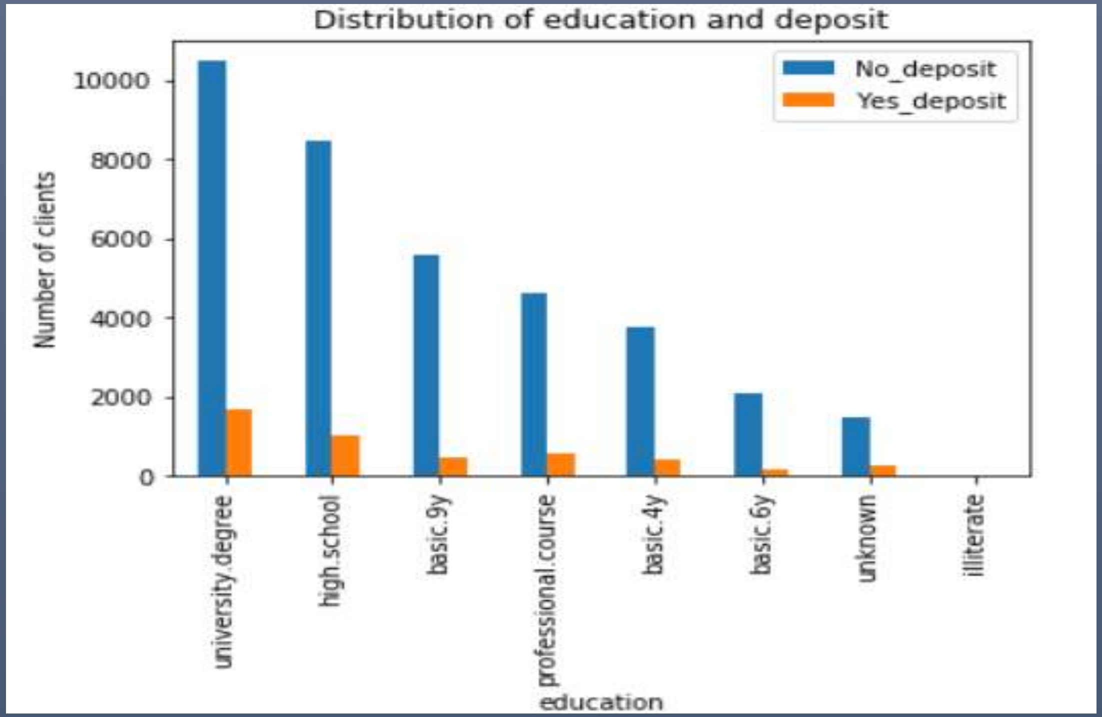
5. Potential applications and real-world impact of the project.

A **Fixed Term Deposit (FTD)** is a financial product offered by banks or financial institutions where an individual deposits a sum of money for a fixed period at a predetermined interest rate. The money cannot be withdrawn until the maturity date without incurring a penalty.

How It's Used:

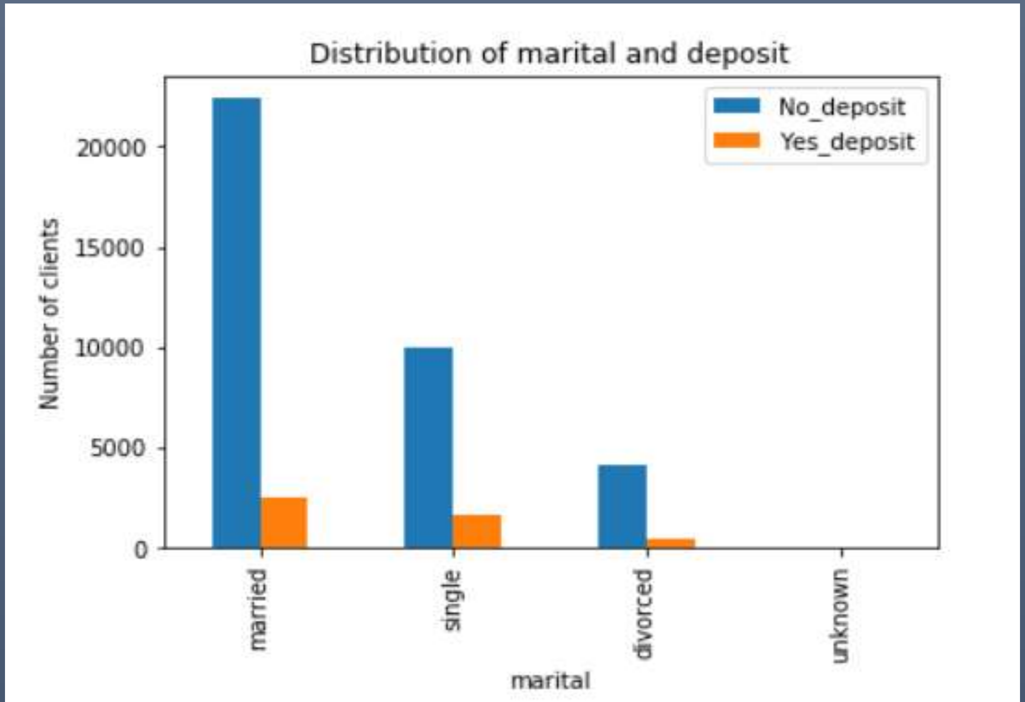
1. **Savings Growth**
2. **Financial Planning**
3. **Corporate Use**
4. **Diversified Investment Strategy**

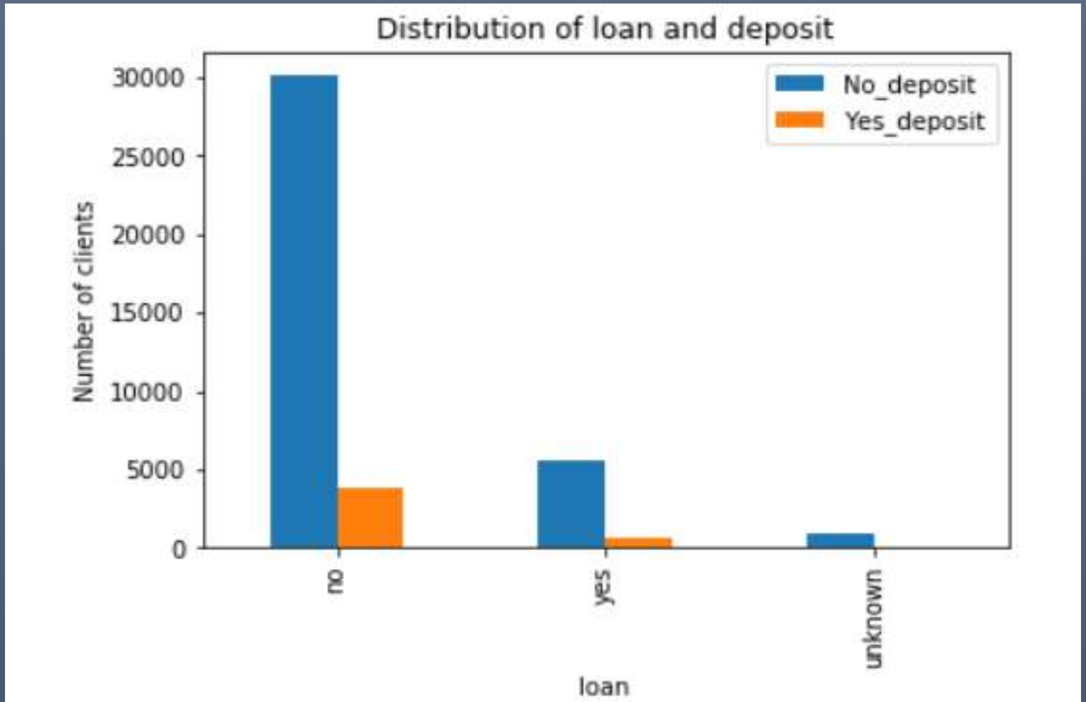




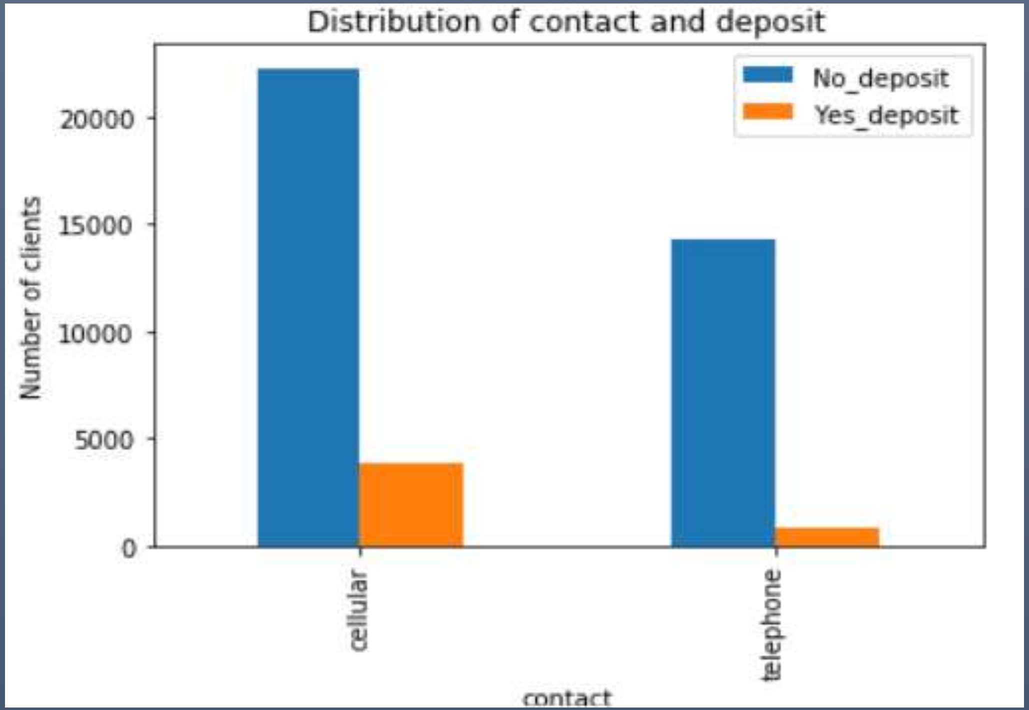
EXPLORATORY DATA ANALYSIS

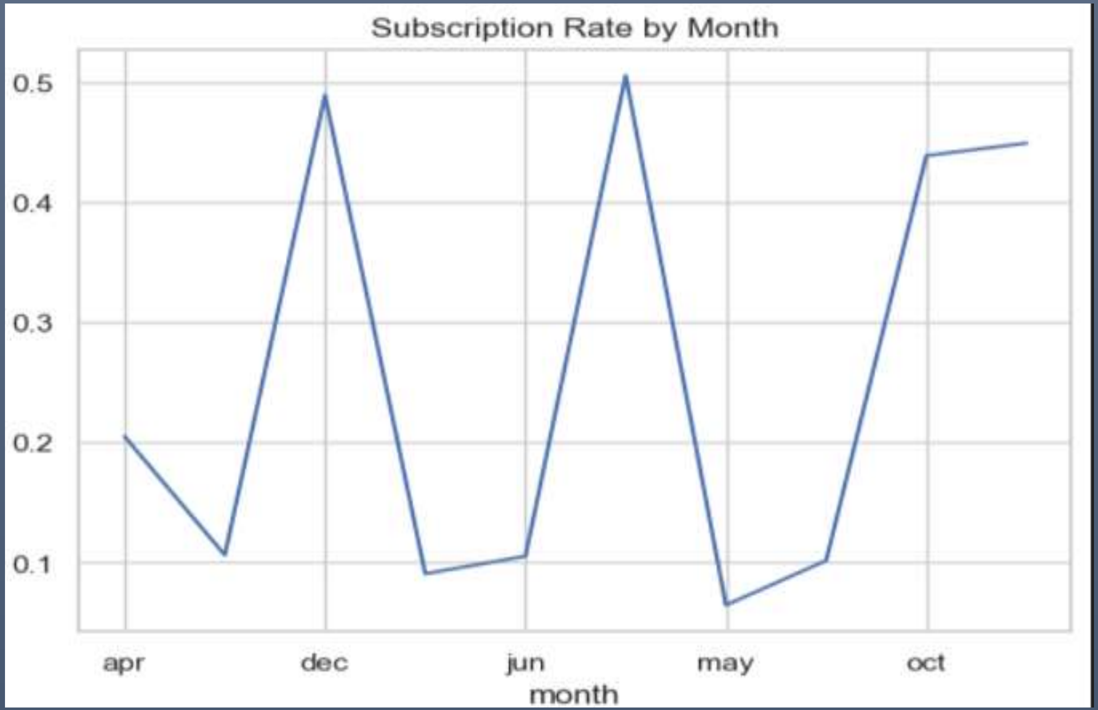
1. Administrative staff and technical specialists were the most likely to open term deposits.
2. Pensioners and students, while fewer in number, had higher relative subscription rates.
3. University educated Clients had the highest Subscription distribution while uneducated clients recorded the lowest.





1. Single customers exhibited better response rates compared to married individuals, despite married individuals being the majority.
2. Customers with existing loans responded differently compared to those without loans.

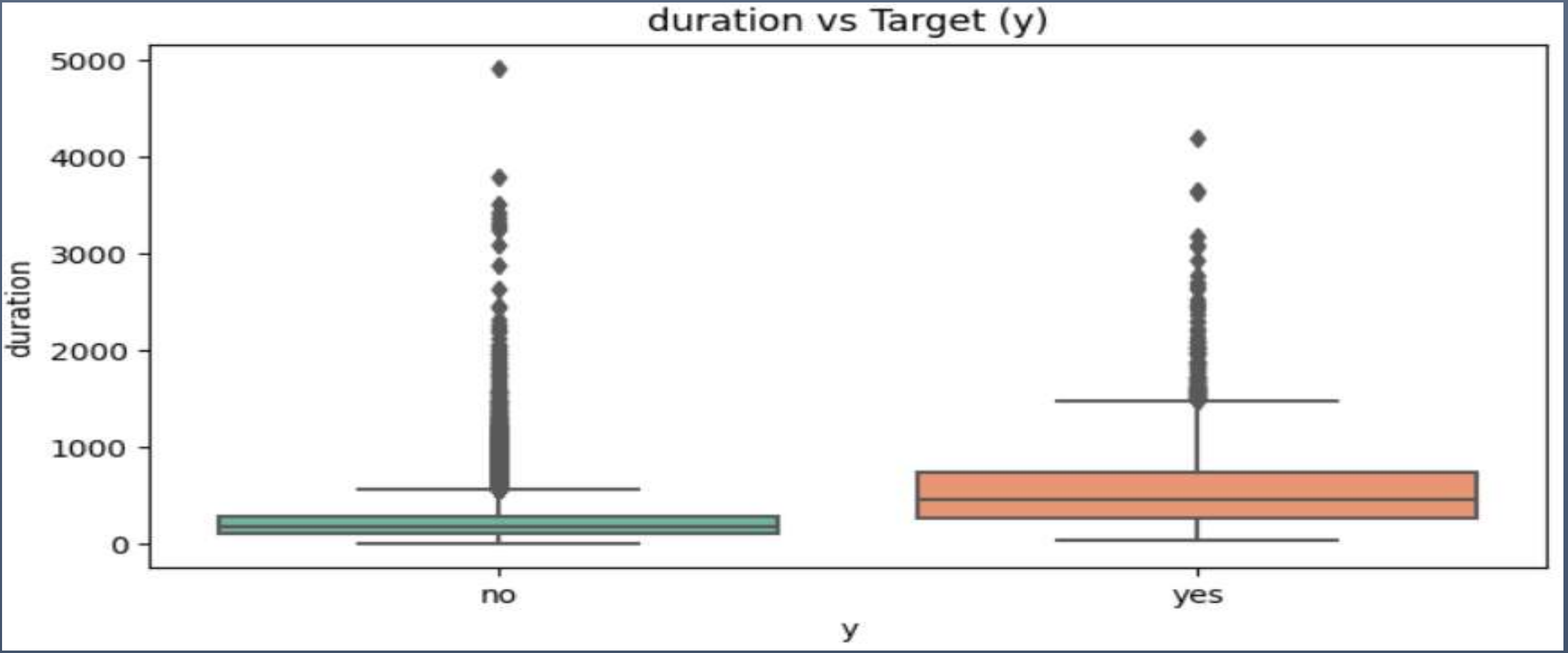




1. The most effective communication channel was cellular with much higher reach than telephone.
2. . The month of May saw a slightly higher subscription rate compared to the other months

	age	duration	campaign	pdays	previous	emp.var.rate	cons.price.idx	cons.conf.idx	euribor3m	nr.employed	y
age	1	-0.000865705	0.00459358	-0.034369	0.0243647	-0.000370685	0.000856715	0.129372	0.0107674	-0.0177251	0.0303988
duration	-0.000865705	1	-0.0716992	-0.047577	0.0206404	-0.0279679	0.00531227	-0.00817287	-0.0328967	-0.0447032	0.405274
campaign	0.00459358	-0.0716992	1	0.0525836	-0.0791415	0.150754	0.127836	-0.0137331	0.135133	0.144095	-0.0663574
pdays	-0.034369	-0.047577	0.0525836	1	-0.587514	0.271004	0.0788891	-0.0913424	0.296899	0.372605	-0.324914
previous	0.0243647	0.0206404	-0.0791415	-0.587514	1	-0.420489	-0.20313	-0.0509364	-0.454494	-0.501333	0.230181
emp.var.rate	-0.000370685	-0.0279679	0.150754	0.271004	-0.420489	1	0.775334	0.196041	0.972245	0.90697	-0.298334
cons.price.idx	0.000856715	0.00531227	0.127836	0.0788891	-0.20313	0.775334	1	0.0589862	0.68823	0.522034	-0.136211
cons.conf.idx	0.129372	-0.00817287	-0.0137331	-0.0913424	-0.0509364	0.196041	0.0589862	1	0.277686	0.100513	0.0548779
euribor3m	0.0107674	-0.0328967	0.135133	0.296899	-0.454494	0.972245	0.68823	0.277686	1	0.945154	-0.307771
nr.employed	-0.0177251	-0.0447032	0.144095	0.372605	-0.501333	0.90697	0.522034	0.100513	0.945154	1	-0.354678
y	0.0303988	0.405274	-0.0663574	-0.324914	0.230181	-0.298334	-0.136211	0.0548779	-0.307771	-0.354678	1

Highly correlated features such as employment rate, consumer confidence index, and consumer price index likely reflect socio-economic factors that influence decision-making. Their variance may support the model’s ability to generalize.



The mean duration of contact impacts subscription rates as seen in this box plot visual.

HYPOTHESIS TESTING – using well profound testing methods like the chi-square and T-testing this were some of the finds.

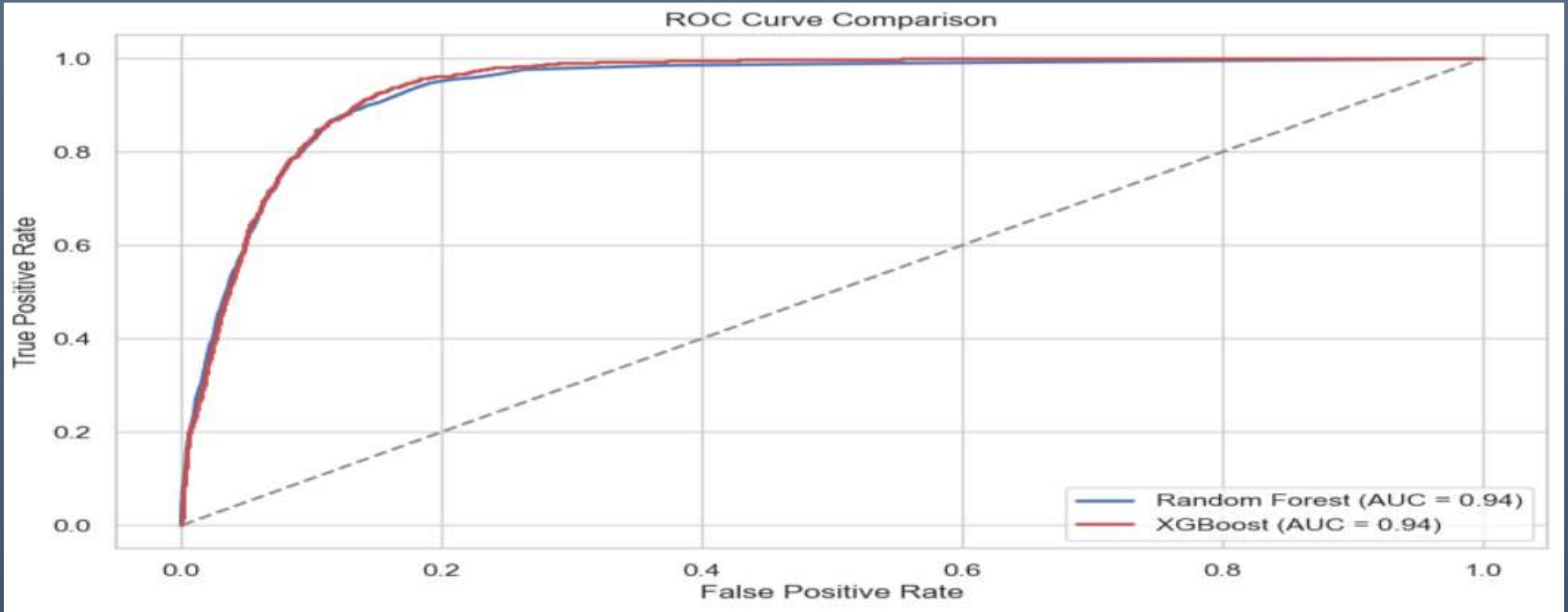
1. With a p-value of 0.5 the hypothesis as confirmed by the earlier visuals showed that employment rate affects subscription.
2. Another insight of interest was that duration also affects subscription which implies that longer conversations increases subscription chances.
3. Education levels also affects subscription as seen by exploratory analysis.

	Model	Accuracy	Precision	Recall	F1-Score
0	RandomForest	0.912722	0.903607	0.912722	0.905996
1	XGBoost	0.911629	0.905668	0.911629	0.907990

	Model	Best Parameters	Best Accuracy
0	Random Forest	{'classifier_max_depth': None, 'classifier_m...	0.548793
1	XGBoost	{'classifier_colsample_bytree': 0.8, 'classif...	0.607007

Modeling & Tunning

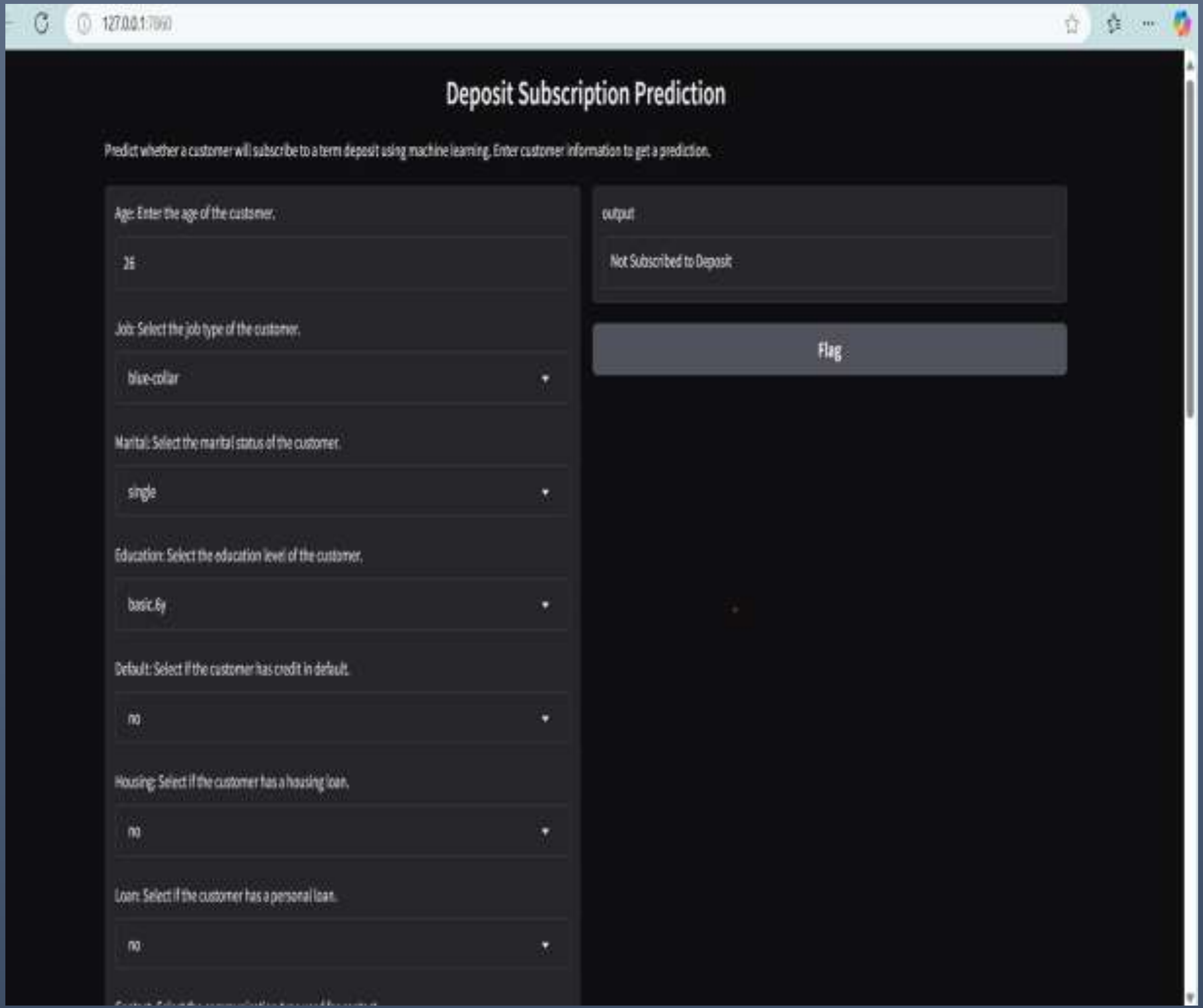
During the modeling stage two models ie Random Forest & Xgboost were trained and evaluated on several metrics like F1 score, Accuracy, Precision & Recall. This was provide a varied bases for choice of best model and also to allow for better model performance perspective. With the models performing at an average of 90% across all the metrics however the tunning of the models to increase it width and performance did not yield much. The performances were debased to an average of 60 proving that the best model is actually data.

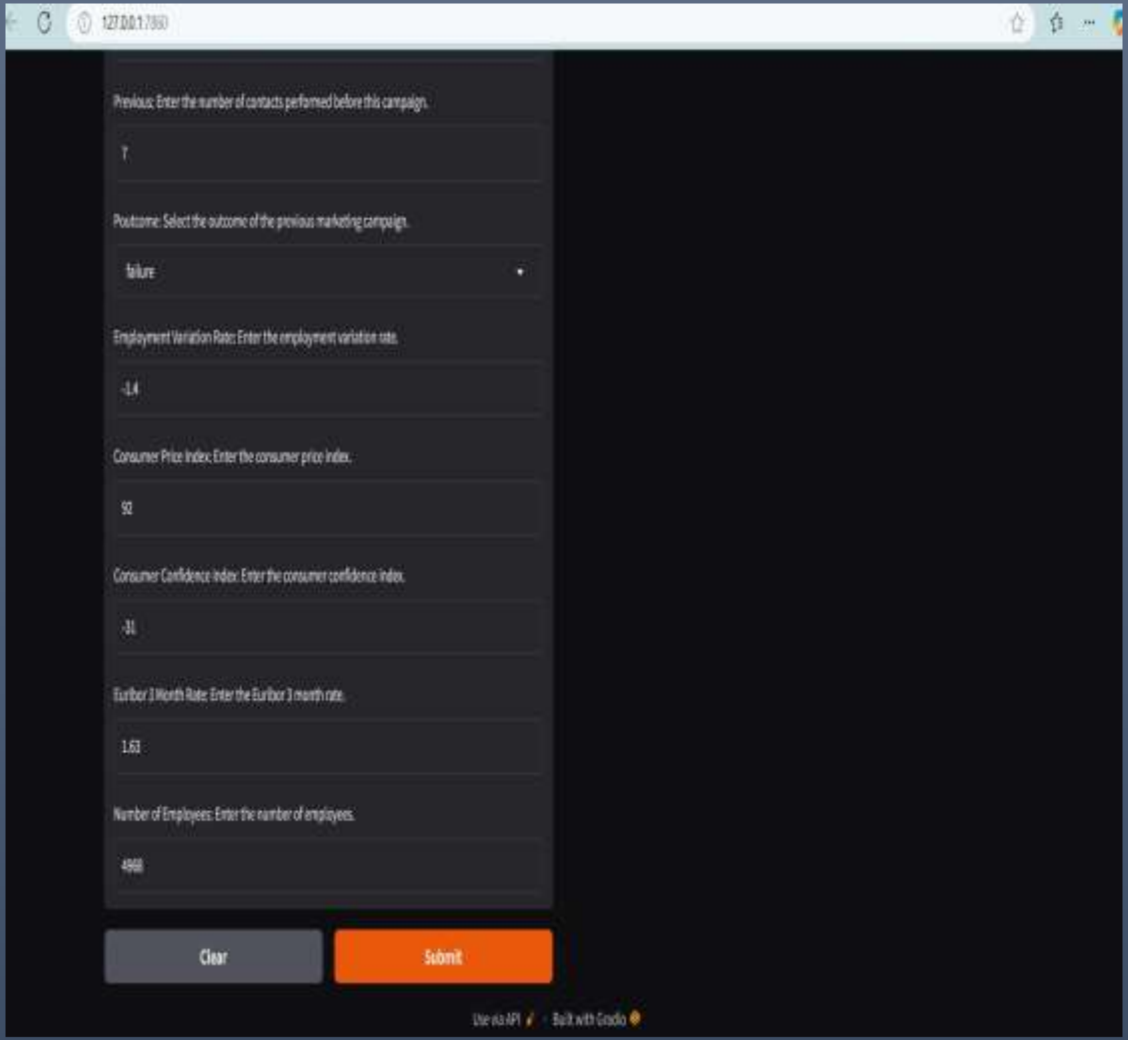


Model validation

1. A high AUC validates the models' ability to predict outcomes accurately across various threshold levels, confirming their strong performance.

2. An AUC of 94% suggests these models are reliable for deployment, as they are effective in identifying true positives while minimizing false positives and negatives.





Deployment

The model (Random Forest) was deployed into a GUI using the Gradio Framework for 3rd party usage and prediction.

Insights:

1. Socio-economic features are critical for segmentation and targeting.

2. Specific demographics, such as students and senior citizens, are more responsive to marketing campaigns.

3. Timing is crucial: May was identified as the most effective month for conversions.

4. Longer contact durations and alternative communication strategies may further improve outcomes.

Recommendation :

1. Segment customers based on socio-economic factors such as age, income level, and profession.
2. Focus marketing efforts on high-response groups (eg Admins).
3. Conduct campaigns during peak months (e.g., May).
4. Explore and refine alternative communication strategies to extend engagement.
5. Regularly re-evaluate model performance using updated data and diverse metrics to ensure consistent generalization.