

Bank Marketing Campaign

İlke Candan Bengi

Submitted to: Data Glacier

Agenda

- Business Understanding
- Data Exploration and Preparation
- Model Building
- Model Evaluation
- Result / Outcomes

Business Understanding

A term deposit is a type of investment that a bank or other financial institution offers with a fixed rate (typically better than just opening a deposit account), with the promise that your money will be returned at a certain maturity date.

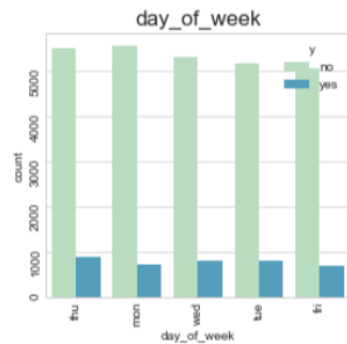
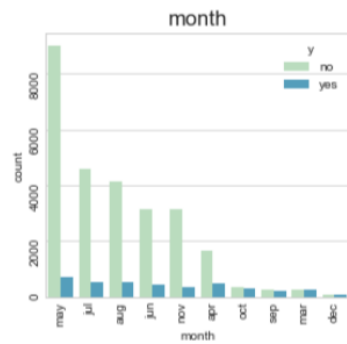
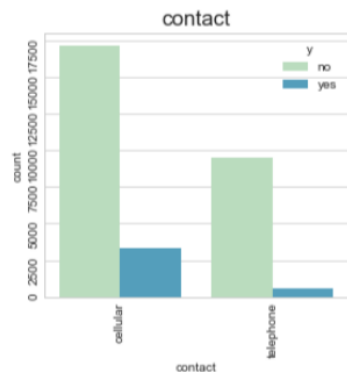
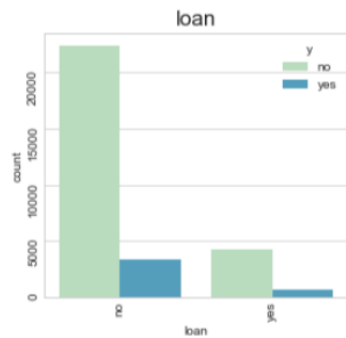
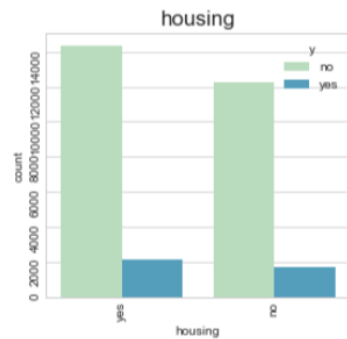
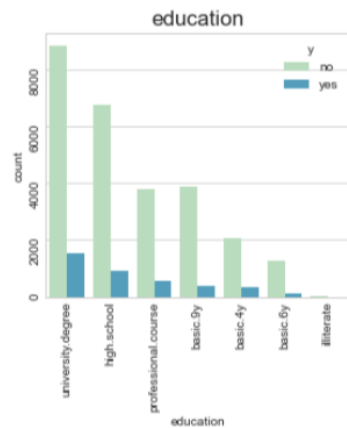
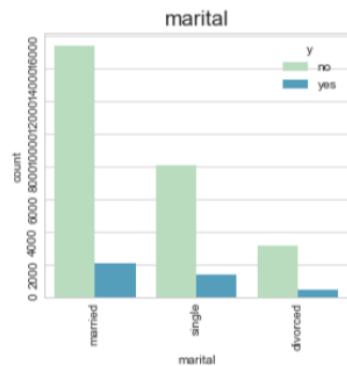
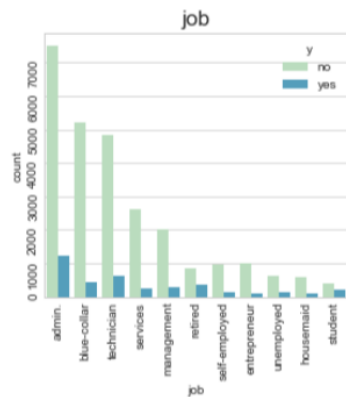
The sample population should be broken down into groups that are most likely to open term deposit accounts.

Data Exploration and Preparation

Data Exploration and Preparation (1/2)

- All programming was done in Python 3 on Jupyter notebook
- Extensive usage of the seaborn and sklearn packages, as well as pandas, numpy, and matplotlib. Dataset contained 20 different features on more than 41,000 clients.

- Target variable was binary (“Yes” or “No”).
- Pandas package was imported and a dataframe was created.
- Categorical variables were looked at first. Visualizations were created using the seaborn package.



Data Exploration and Preparation (2/2)

- Many features had missing values. How do we handle this?
- For categorical features, imputation using other independent variables. For example, cross-tabulation between 'job' and 'education'; 'age' and 'job'; 'home ownership' and 'loan status.'
- Among numerical features, fortunately only column ('pdays') had any missing values. Unfortunately, missing values made up the majority of the column.
- Heatmap using seaborn package was created to show us any particularly strong correlations between the independent variables and the target variable outcome.

Correlation Heatmap :



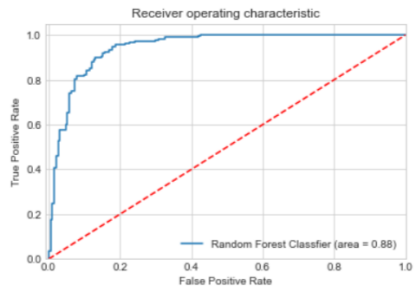
Model Building

Random Forest Report in test set					
		precision	recall	f1-score	support
0		0.96	0.99	0.97	380
1		0.58	0.20	0.30	20
	accuracy			0.95	401
	macro avg	0.77	0.60	0.64	401
	weighted avg	0.94	0.95	0.94	401

Random Forest Classifier perform best

Results / Outcome

```
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
logit_roc_auc = roc_auc_score(y_test, rf.predict(X_test))
fpr, tpr, thresholds = roc_curve(y_test, rf.predict_proba(X_test)[: ,1])
plt.figure()
plt.plot(fpr, tpr, label='Random Forest Classifier (area = %0.2f)' % logit_roc_auc)
plt.plot([0, 1], [0, 1], 'r--')
plt.xlim([-0.01, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc="lower right")
plt.show()
```



Significant Variables

Duration,
Education

Age

Duration, Mode of Contact:
Telephone

Campaign

Thank You