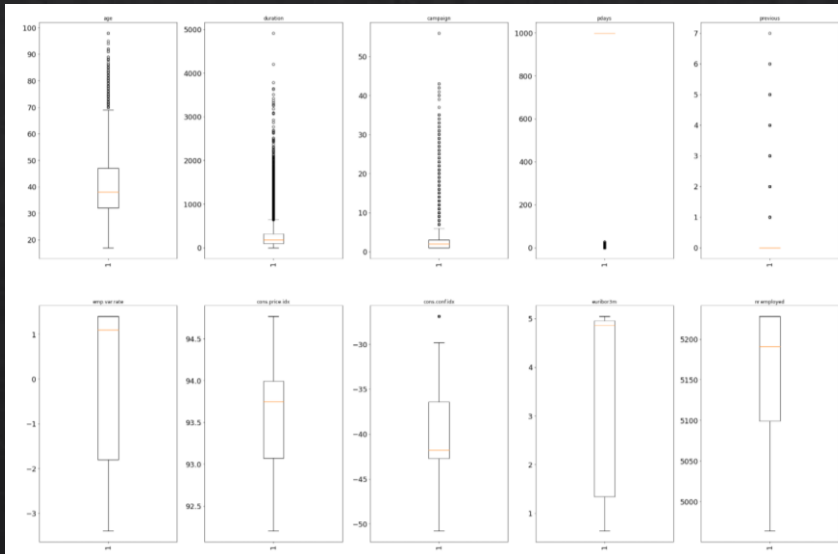


DTSA 5509 Supervised Learning Final Project

Marketing Campaign for a Retail Bank

# Objective and Data

- Build and evaluate “propensity to buy” classification models for a retail bank
- Use “Bank Marketing” dataset from UCI Machine Learning Repository



#	Column	Non-Null Count		Dtype
0	age	41188	non-null	int64
1	job	41188	non-null	object
2	marital	41188	non-null	object
3	education	41188	non-null	object
4	default	41188	non-null	object
5	housing	41188	non-null	object
6	loan	41188	non-null	object
7	contact	41188	non-null	object
8	month	41188	non-null	object
9	day_of_week	41188	non-null	object
10	duration	41188	non-null	int64
11	campaign	41188	non-null	int64
12	pdays	41188	non-null	int64
13	previous	41188	non-null	int64
14	poutcome	41188	non-null	object
15	emp.var.rate	41188	non-null	float64
16	cons.price.idx	41188	non-null	float64
17	cons.conf.idx	41188	non-null	float64
18	euribor3m	41188	non-null	float64
19	nr.employed	41188	non-null	float64
20	y	41188	non-null	object

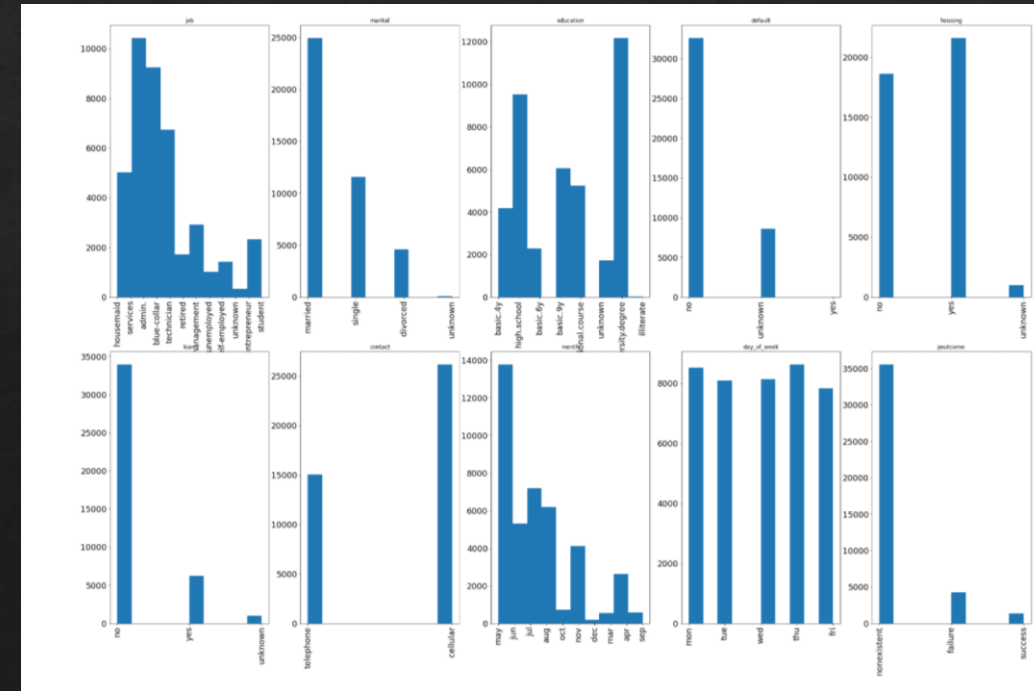
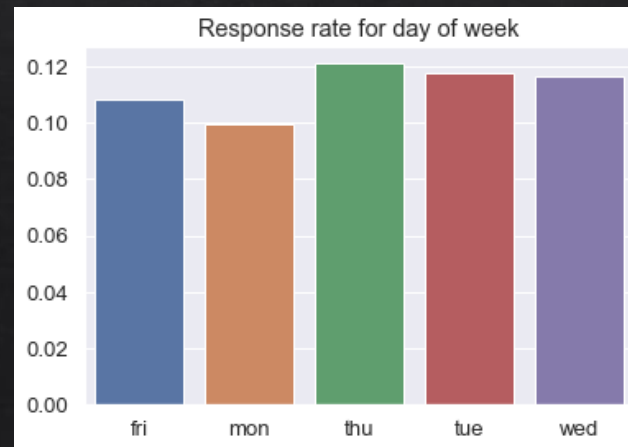
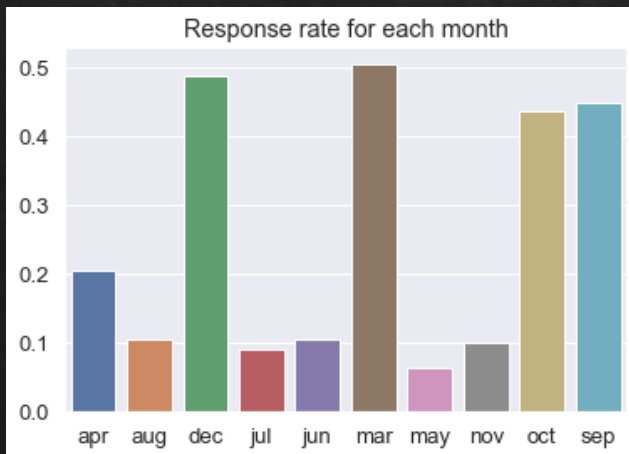
dtypes: float64(5), int64(5), object(11)

Target

y  
no 36548  
yes 4640

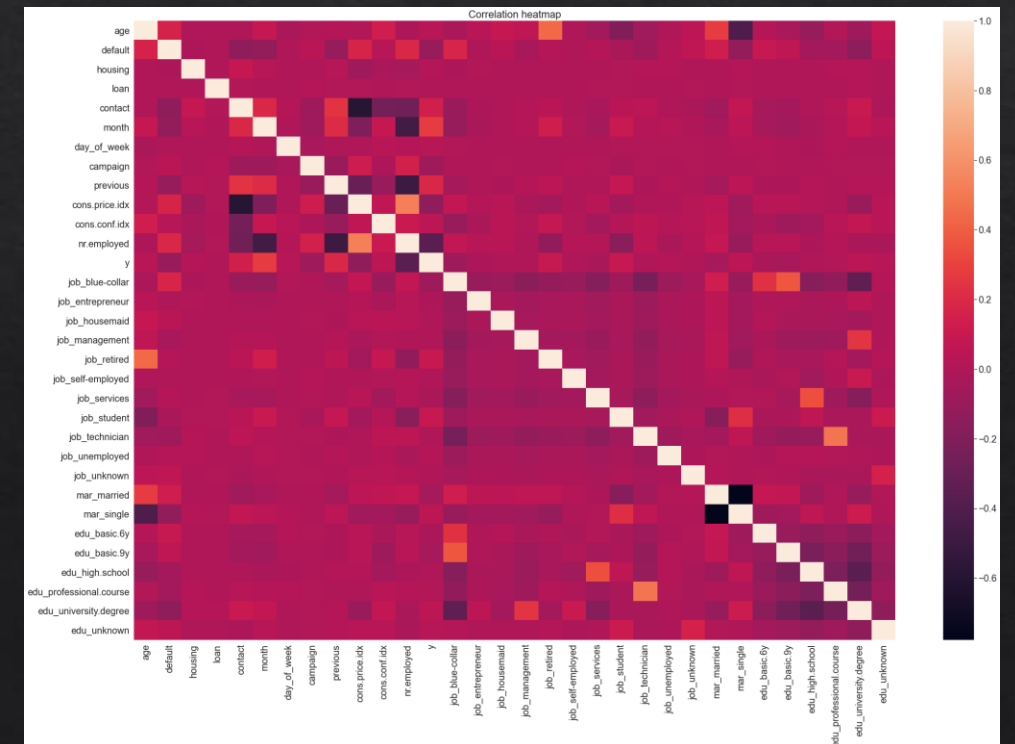
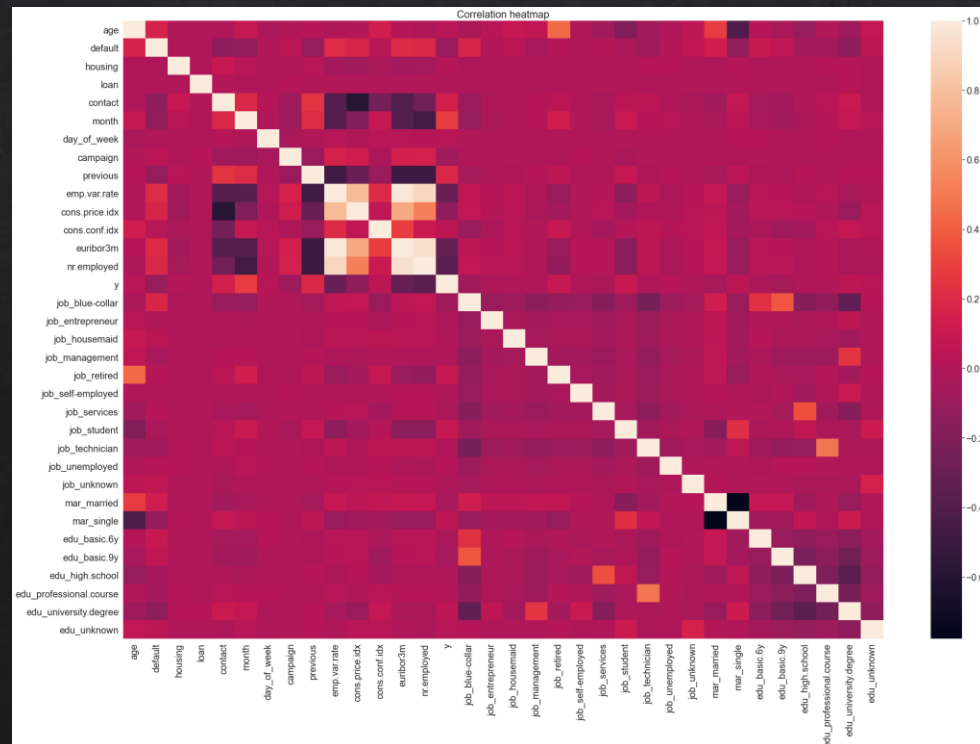
# Preprocessing

- NULL values and Unknown categories replacement
- Delete outlier records and adjusted 3SD+ values
- Binary values substituted with 0 and 1
- Removed irrelevant fields
- Categorical fields converted to indicator fields
- Numerical values scaled
- Date fields converted to response rate



# Exploratory Analysis

- Unpredictive values and fields weeded out
- Correlations examined before and after collinearity analysis



# Models

- Good results from Logistic Regression and Random Forest
- Logistic Regression with CV and SVC with hyperparameter tuning attempted
- Logistic regression preferred for its explainability

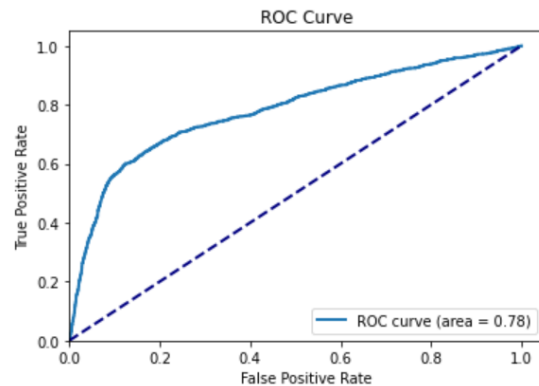
-----logistic regressions-----  
Classification Report

	precision	recall	f1-score	support
0	0.91	0.97	0.94	10953
1	0.53	0.23	0.32	1398
accuracy			0.89	12351
macro avg	0.72	0.60	0.63	12351
weighted avg	0.87	0.89	0.87	12351

Confusion Matrix

```
[[10662  291]
 [ 1075  323]]
```

accuracy score 0.8894016678811433



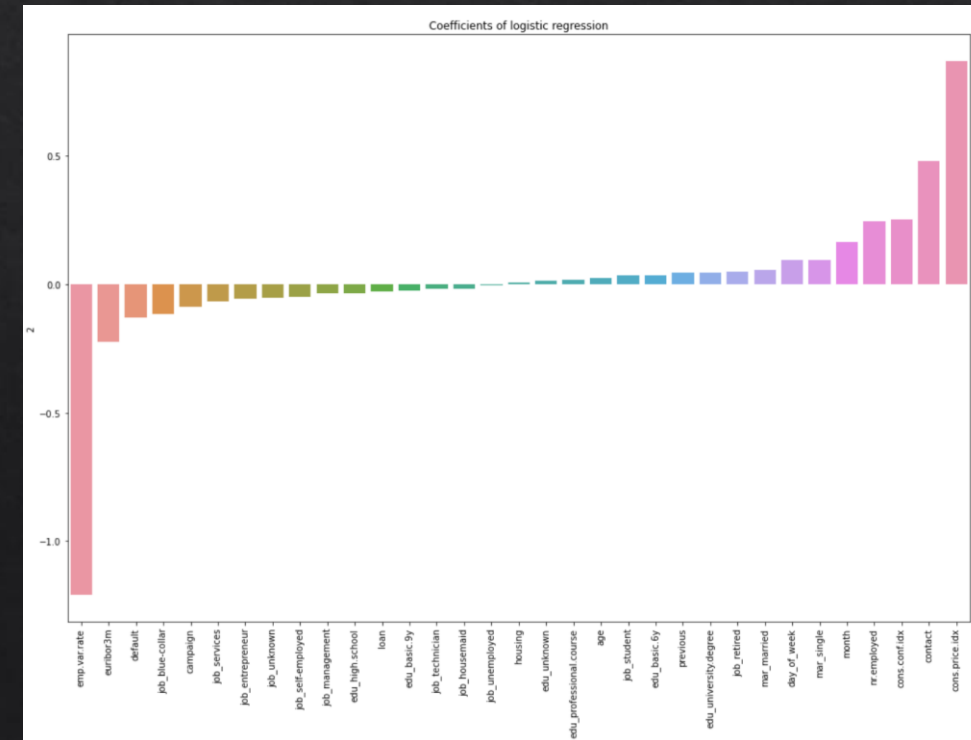
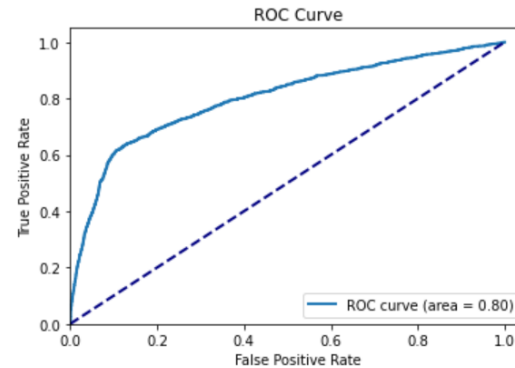
-----random forest-----  
Classification Report

	precision	recall	f1-score	support
0	0.91	0.98	0.94	10953
1	0.58	0.24	0.34	1398
accuracy			0.89	12351
macro avg	0.75	0.61	0.64	12351
weighted avg	0.87	0.89	0.87	12351

Confusion Matrix

```
[[10717  236]
 [ 1069  329]]
```

accuracy score 0.8943405392275929





# Evaluation and Conclusion

- Achieved 90% accuracy rate across algorithms achieved (OK)
- Estimation with simple strategy shows 104% gain in the response rate
- Machine learning steps were experienced, various lessons were learned
- Domain knowledge and exploratory analysis are critical

```
-----logistic regressions-----
Classification Report

              precision    recall  f1-score   support

     0       0.91       0.97       0.94     10953
     1       0.53       0.23       0.32      1398

 accuracy          0.89     12351
 macro avg       0.72     12351
 weighted avg    0.87     12351

Confusion Matrix

[[10662   291]
 [ 1075   323]]

accuracy score 0.8894016678811433
```

Population response rate:

$$1398 / (1398 + 10953) = 11.3\%$$

All predicted response rate:

$$323 / (323 + 1075) = 23.1\%$$

$$23.1/11.3 = 204\%$$