Data Glacier
Your Deep Learning Partner

Business Understanding

Virtual Internship

Felipe I. Crespo

Agenda

Data Understanding

Clients

Last contact

Campaign

Bivarial analysis

Recommended models

Data Glacier
Your Deep Learning Partner

Data Understanding

Features are divided into 4 groups:
1. Data related to clients (age, job, marital, education, default, housing and loan).
2. Data related with the last contact of the current campaign (contact, month, day_of_week and duration).
3. Data related with campaigns (campaign, odays, previous and poutcome).
4. Data related with socio economic context attributes (emp.var.rate, cons.Price.idx, cons.conf.idx, euribor3m, nr.employed).
5. TARGET: "y".

```
Rows = 41188  | Columns =  21

    ['age', 'job', 'marital', 'education', 'default', 'housing', 'loan',
     'contact', 'month', 'day_of_week', 'duration', 'campaign', 'pdays',
     'previous', 'poutcome', 'emp.var.rate', 'cons.price.idx',
     'cons.conf.idx', 'euribor3m', 'nr.employed', 'y'],
```

Client

Customers job (%):
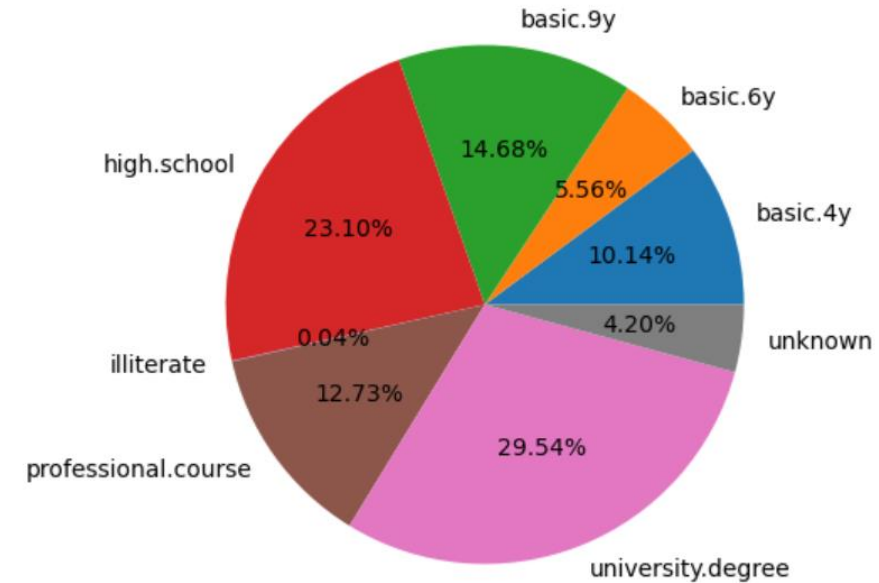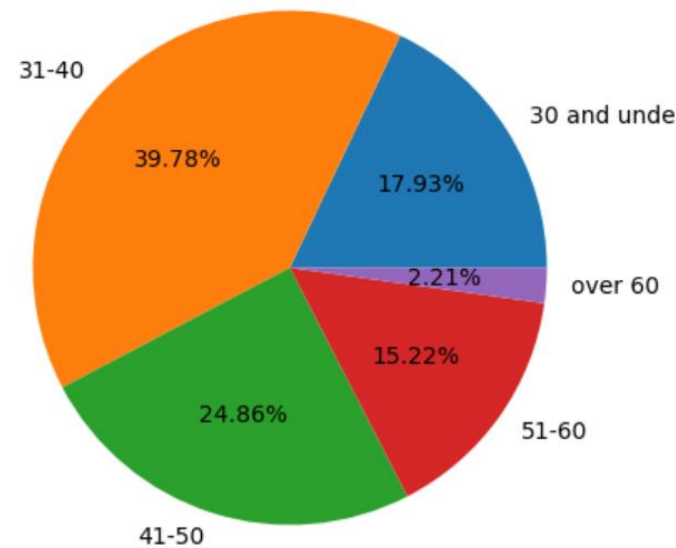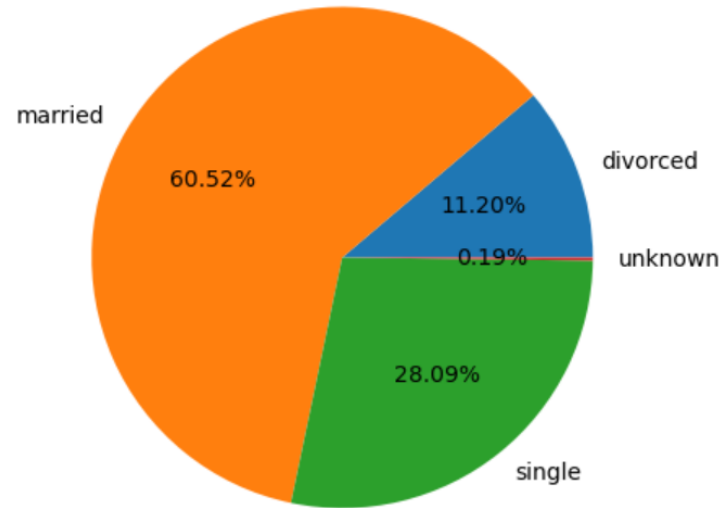
job			
admin.	25.303486	self-employed	3.450034
blue-collar	22.467709	housemaid	2.573565
technician	16.371273	unemployed	2.461882
services	9.636302	student	2.124405
management	7.099155	unknown	0.801204
retired	4.175974		
entrepreneur	3.535010		

Marital state of customers (%):

marital	
divorced	11.197436
married	60.522482
single	28.085850
unknown	0.194231

Client

Customers and Default, Housing Loan and Personal Loan:
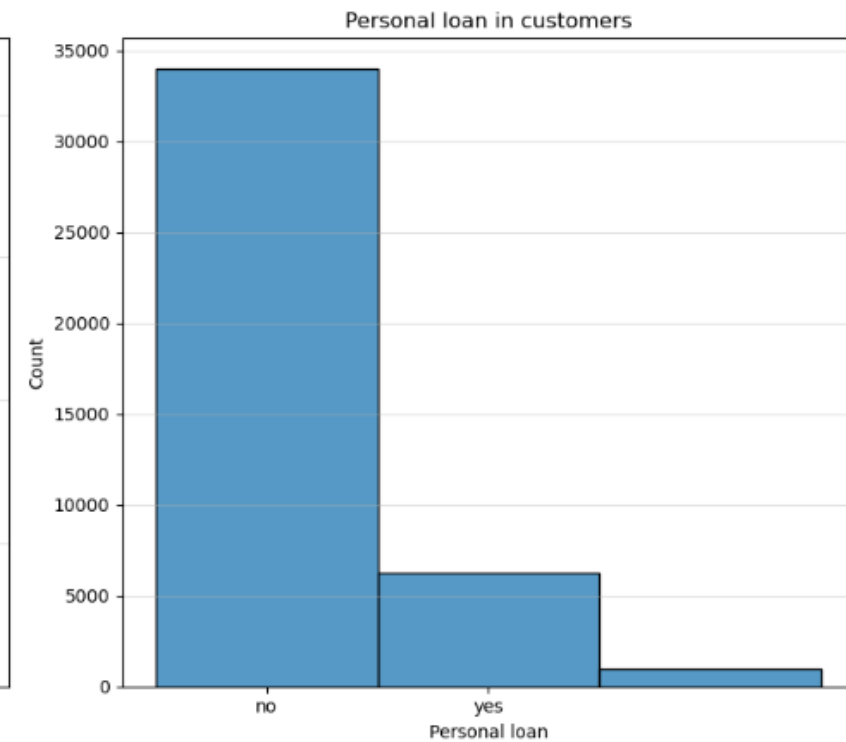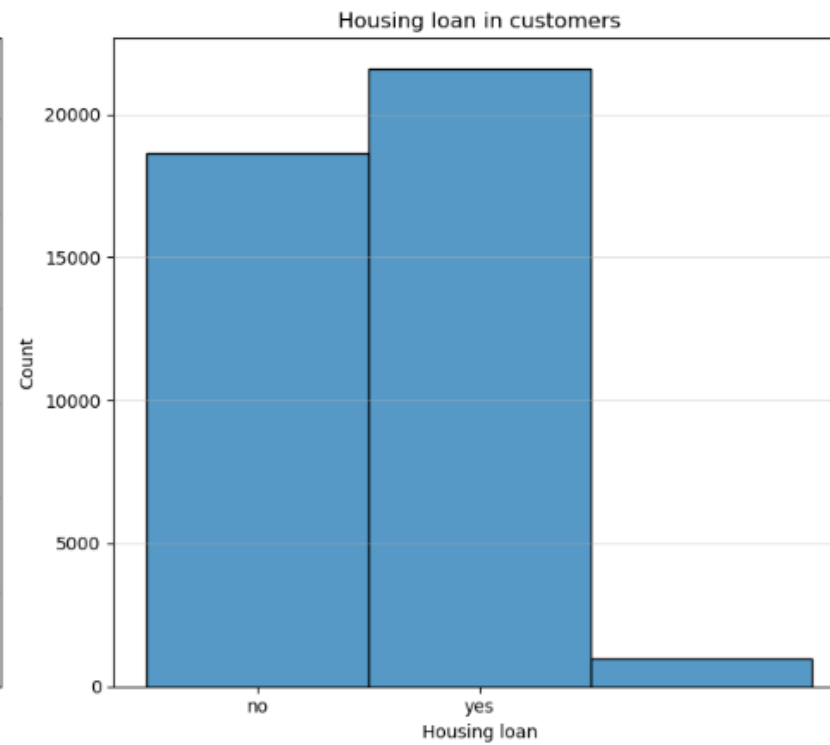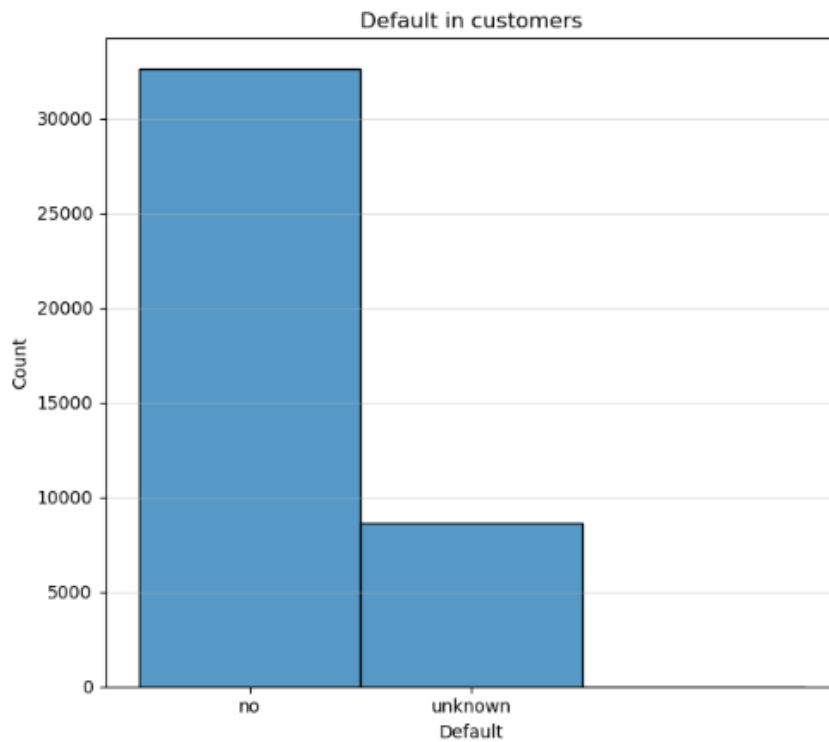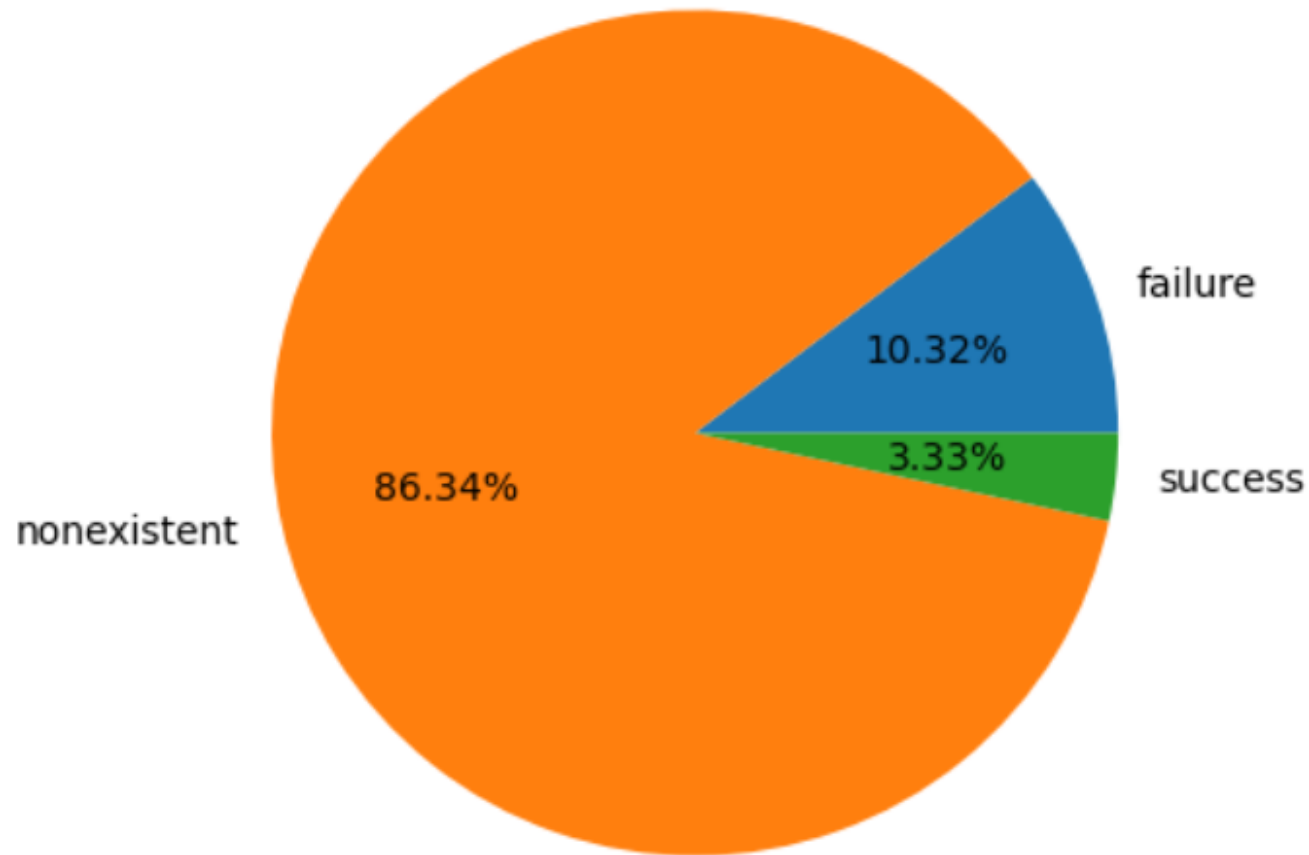
Last Contact

Contact to Costumers
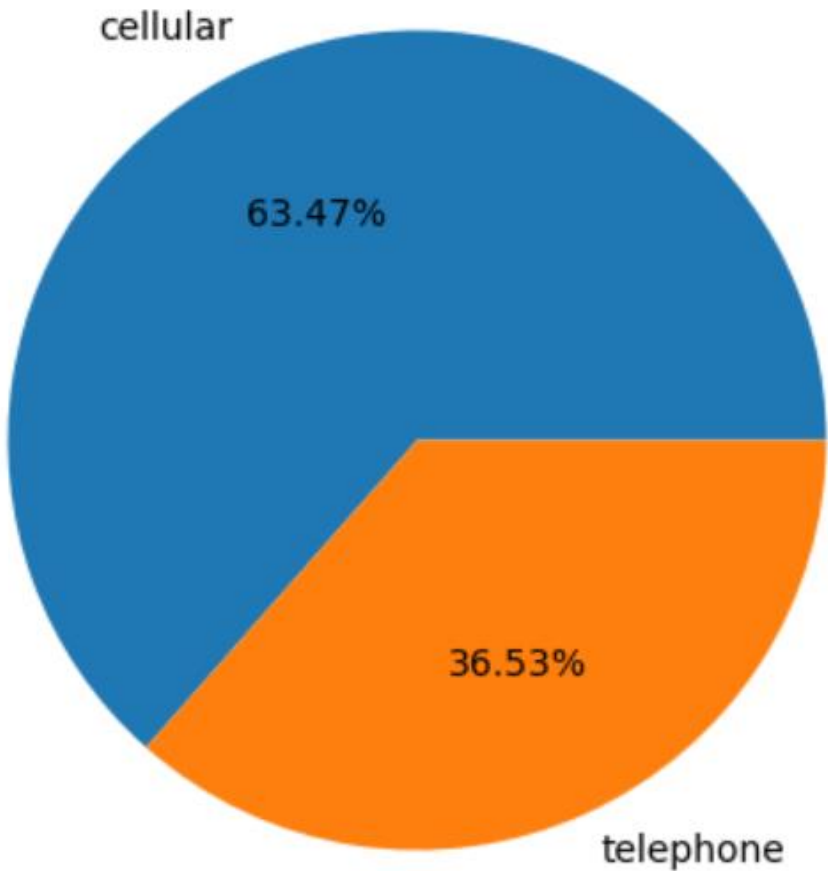
contact		
month		
may	13769	
jul	7174	
aug	6178	
jun	5318	
nov	4101	

```
duration_range
0-200         22760
201-400       11264
401-600        3700
601-800        1707
801-1000        801
over 1000       956


contact
cellular      63.474798
telephone     36.525202
```

Campaing

Previous Relation


```
campaign_range
0-3               81.463047
4-6               12.695445
7-9                3.185394
10-12              1.279499
More than 12       1.376615
```
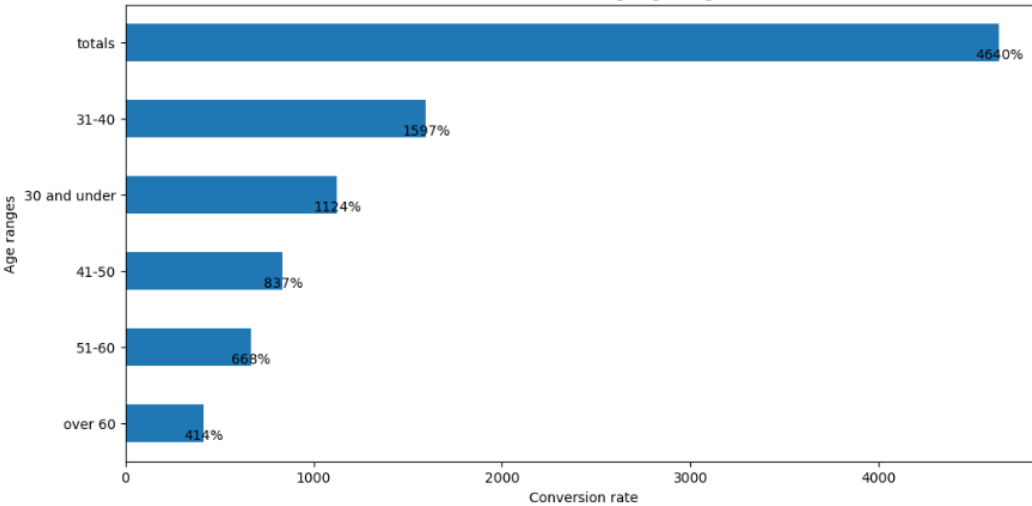
- The majority of customers (86%) have no previous relation with the bank
- Only 14% of customers have had previous contacts with the bank
- Among those who had previous contacts with the bank, the success rate is 3.3%.
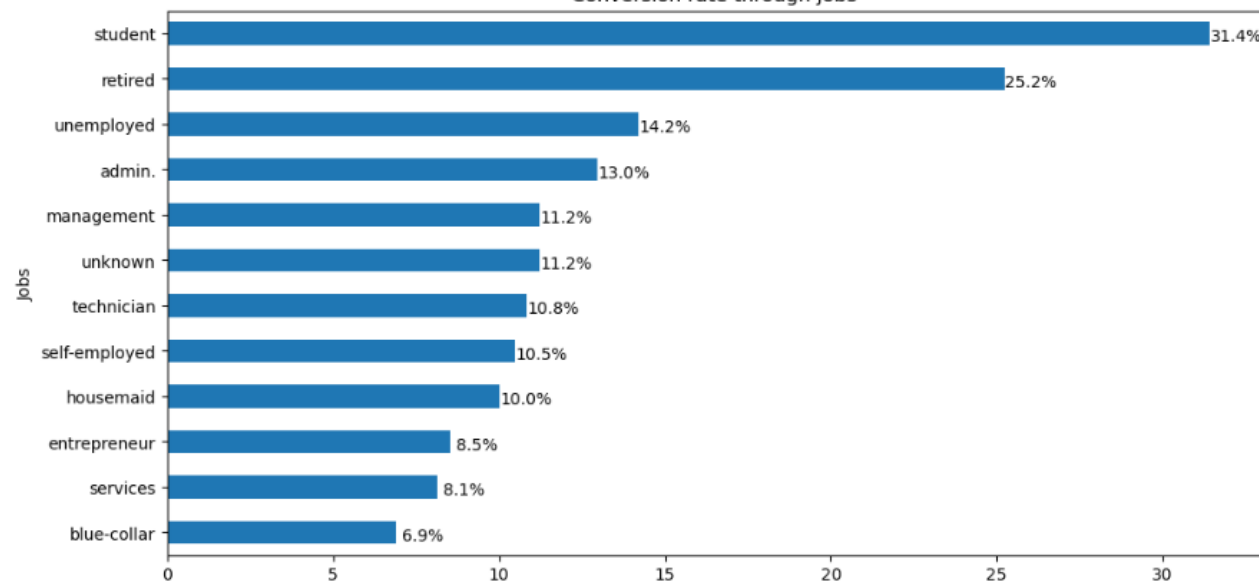
Biavarial analysis


Conversion rate through age ranges

Conversion rate through jobs

y	no	yes	totals	% no	% yes
job					
admin.	9070	1352	10422	87.027442	12.972558
blue-collar	8616	638	9254	93.105684	6.894316
entrepreneur	1332	124	1456	91.483516	8.516484
housemaid	954	106	1060	90.000000	10.000000
management	2596	328	2924	88.782490	11.217510
retired	1286	434	1720	74.767442	25.232558
self-employed	1272	149	1421	89.514426	10.485574
services	3646	323	3969	91.861930	8.138070
student	600	275	875	68.571429	31.428571
technician	6013	730	6743	89.173958	10.826042
unemployed	870	144	1014	85.798817	14.201183
unknown	293	37	330	88.787879	11.212121
totals	36548	4640	41188	88.734583	11.265417

- Customers under 30 and over 60 years old have higher term deposit acceptance rates (15% and 45%, respectively)
- Student and retired are the top jobs with the highest acceptance rates.

Biavarial analysis

Conversion rate along contact mean

Conversion rate along months
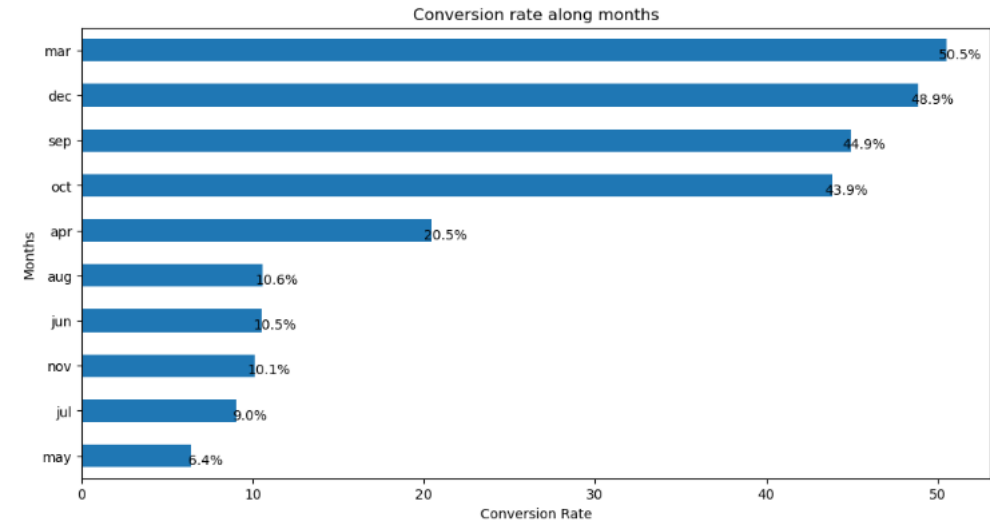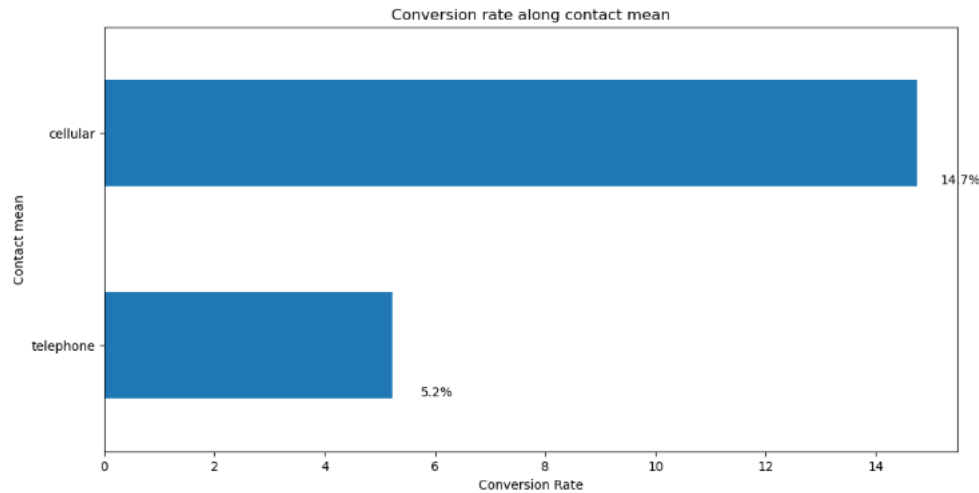
| y	no	yes	totals	% no	% yes
duration_range					
0-300	28466	1518	29984	94.937300	5.062700
301-600	6302	1438	7740	81.421189	18.578811
601-900	1259	924	2183	57.672927	42.327073
901-1200	329	439	768	42.838542	57.161458
1201-1500	115	192	307	37.459283	62.540717
1501-1800	36	69	105	34.285714	65.714286
1801-2100	22	34	56	39.285714	60.714286
2101-2400	6	8	14	42.857143	57.142857
More than 2400	13	18	31	41.935484	58.064516
totals	36548	4640	41188	88.734583	11.265417

- Contacting customers by cell phone resulted in a higher term deposit acceptance rate (15%) compared to telephone contact (5%)
- Low customer numbers in December, March, October, and September had a high conversion rate (over 40%)
- Longer call duration was associated with higher conversion rates, but only a small proportion of

Biavarial analysis

Conversion rate through housing

Conversion rate through loan
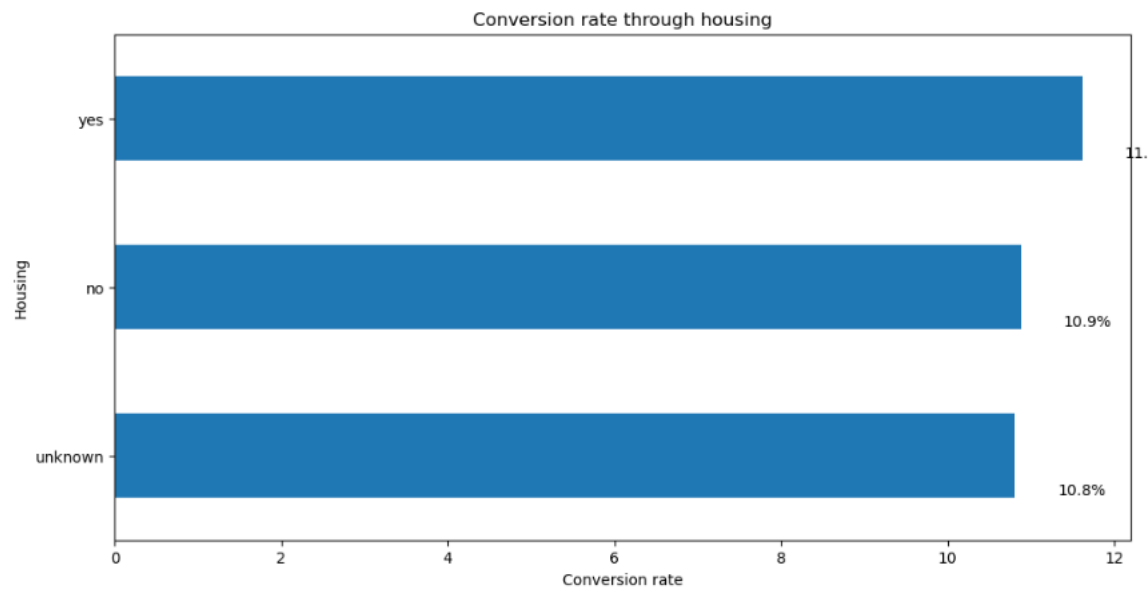
```
Default and target:
y           no        yes
default
no        87.121026  12.878974
unknown   94.847040   5.152960
yes      100.000000   0.000000
```
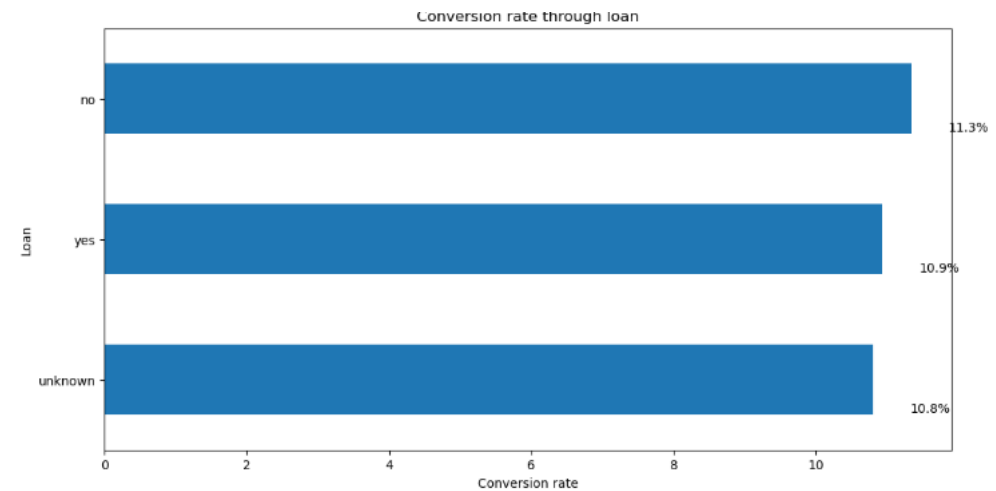
- Conversion rates for Loan and Housing products are similar across all categories of people (known, unknown, and without relevant information)
- Default product conversion rate varies based on default
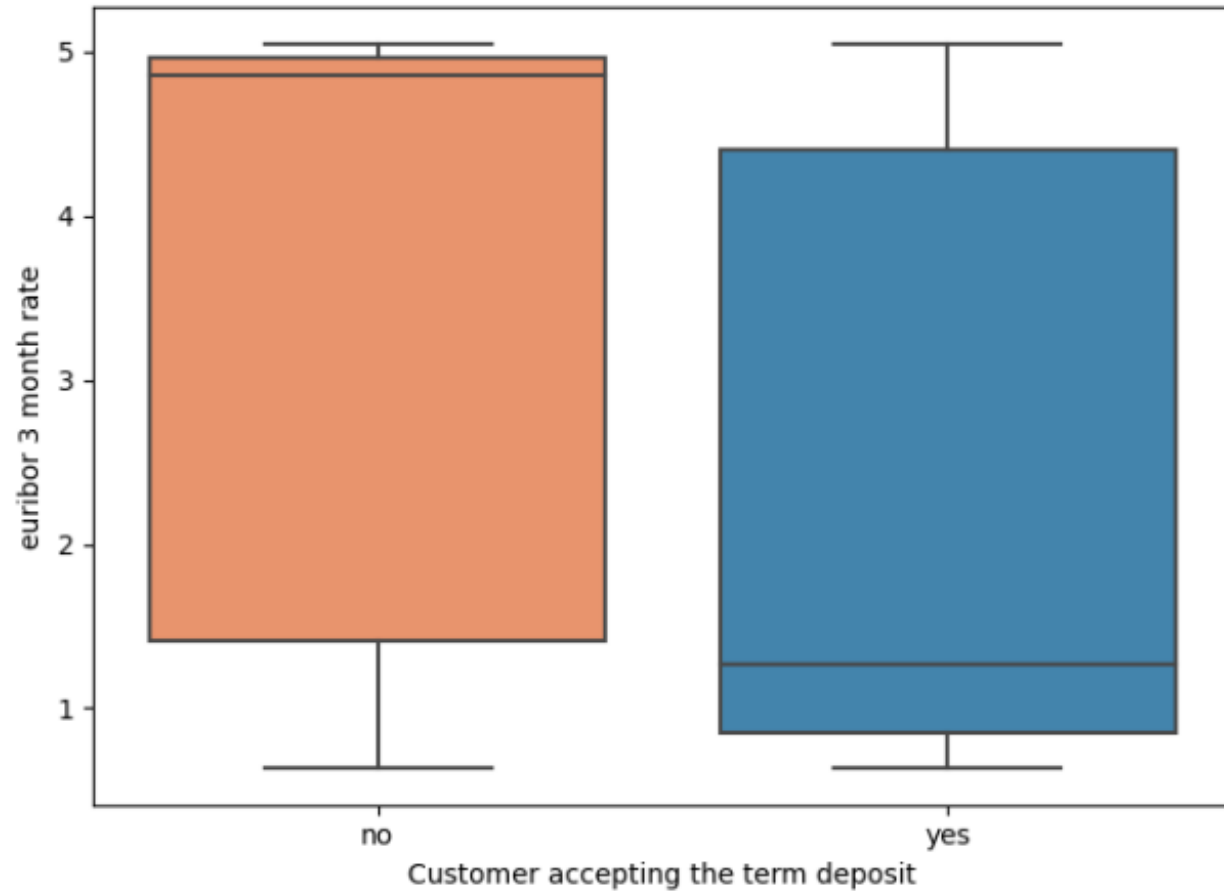
Biavarial analysis
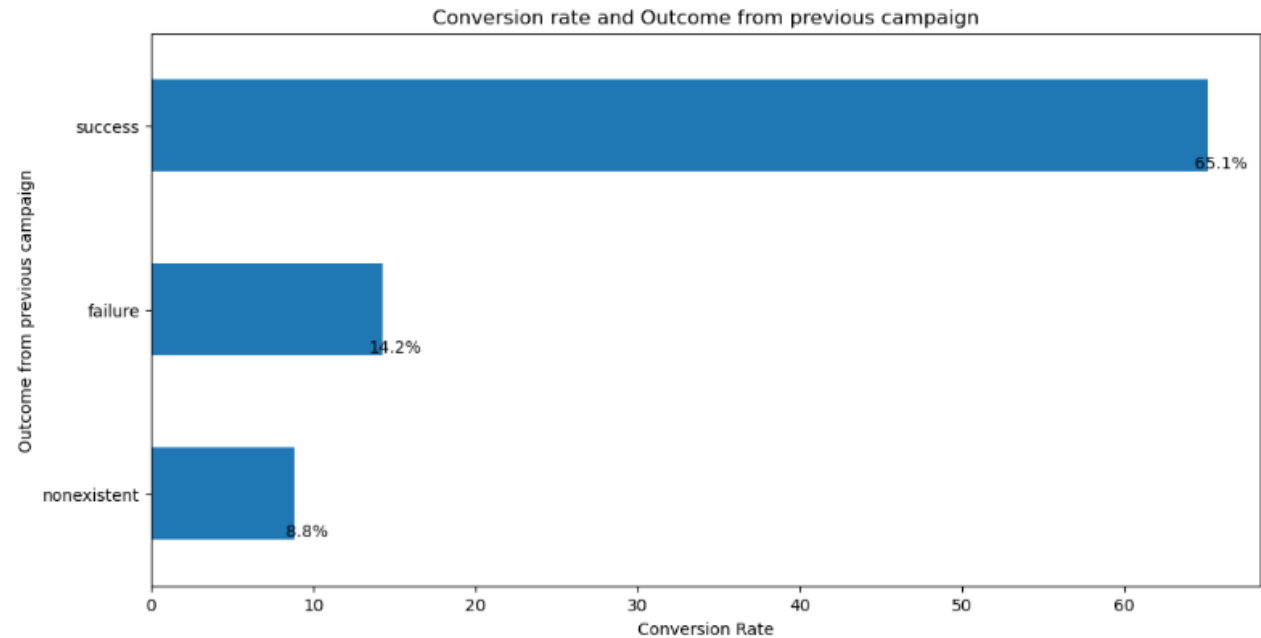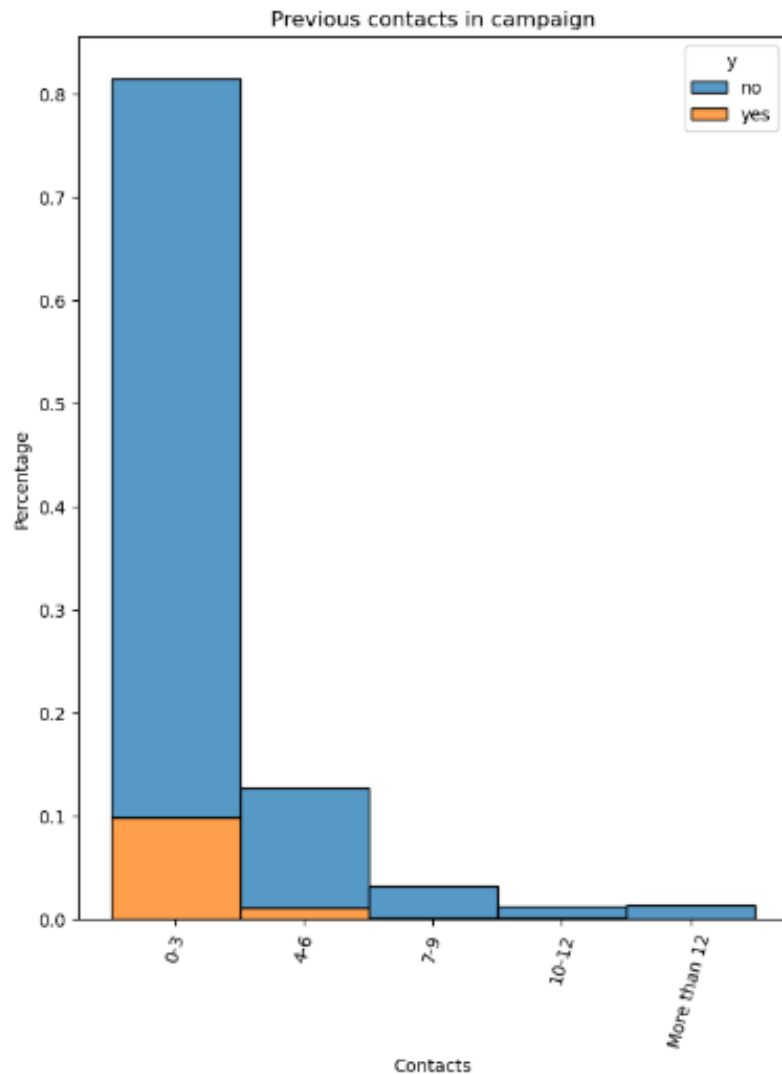
- A lower number of contacts during the campaign is associated with higher conversion rates: 12% for 0-3 contacts and 9% for 4-6 contacts
- Previous campaign success is a strong predictor of term deposit acceptance, with 65% of customers who had previously accepted now also accepting it.
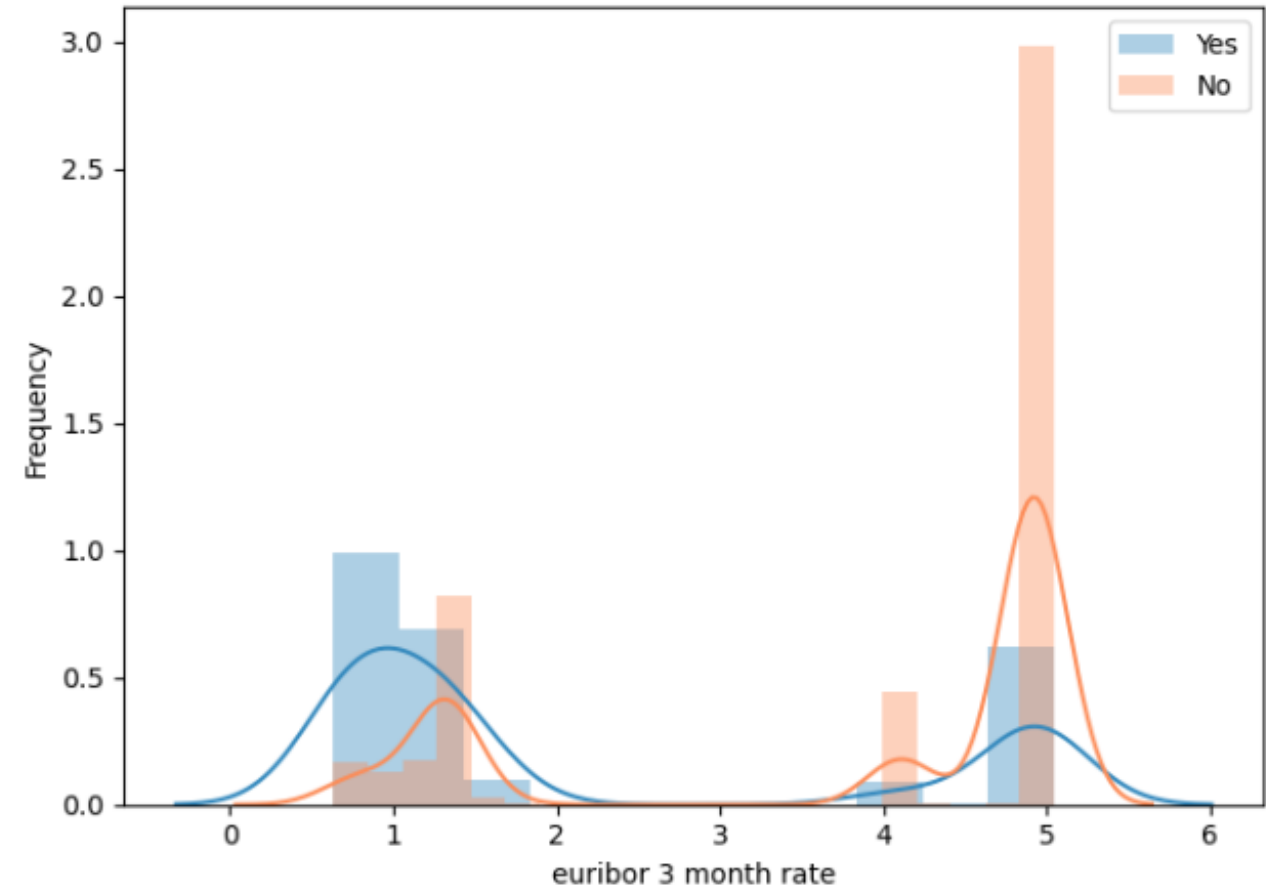
Biavarial analysis
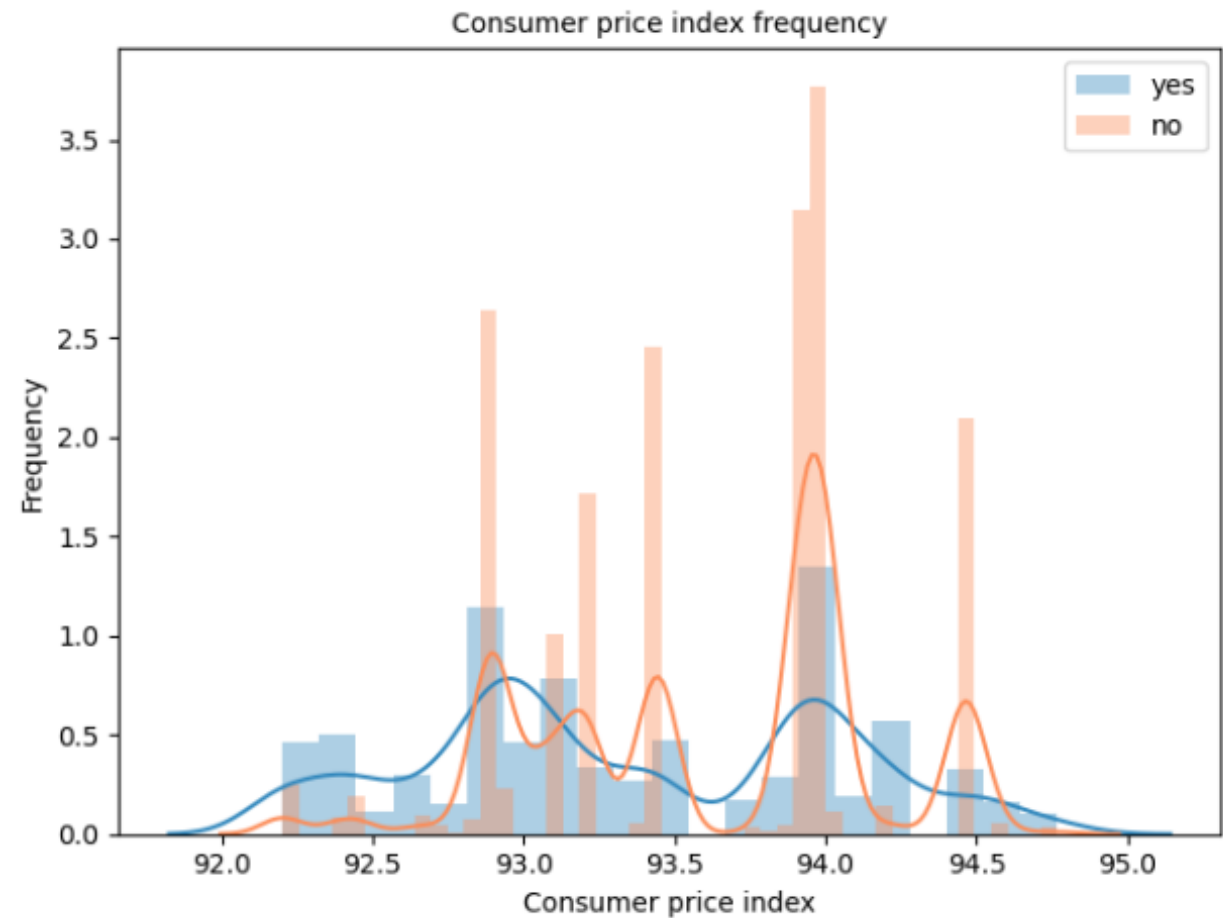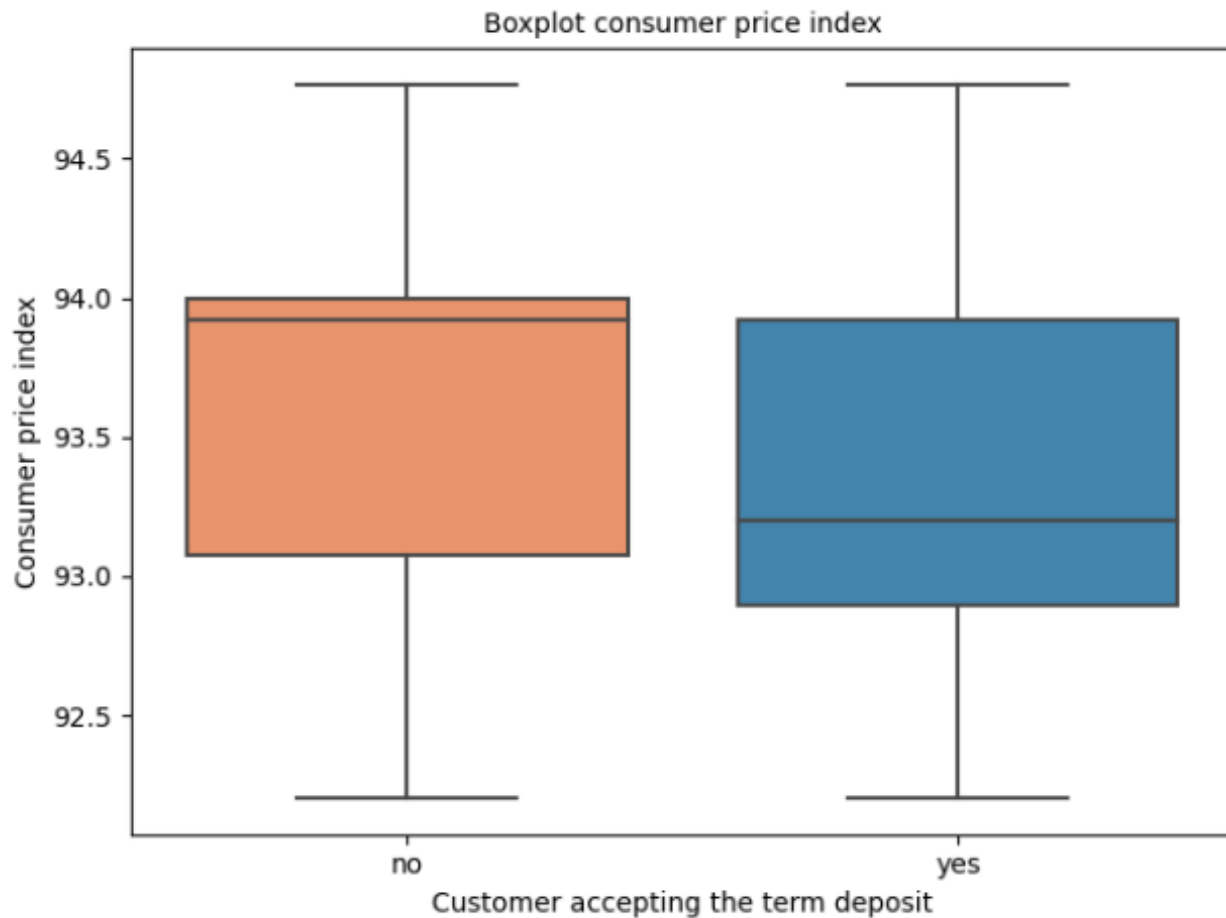
Biavarial analysis

Recommended Model

Used Pycaret library to select best model for business problem based on AUC
Preprocessed three datasets using different methods and found treating "Unknown" as separate category produced best results

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
lightgbm	Light Gradient Boosting Machine	0.8970	0.8007	0.3267	0.5767	0.4167	0.3649	0.3829	0.2900
catboost	CatBoost Classifier	0.8973	0.7962	0.3045	0.5855	0.4001	0.3501	0.3728	7.1710
lr	Logistic Regression	0.8231	0.7903	0.6352	0.3458	0.4476	0.3532	0.3764	0.3960
lda	Linear Discriminant Analysis	0.8210	0.7895	0.6352	0.3421	0.4446	0.3492	0.3730	0.1390

Selected LightGBM and CatBoost as top-performing models for hyperparameter tuning
Plan to implement Grid Search and optimize hyperparameters next week to further

Plan to implement Grid Search and optimize hyperparameters next week to further improve model performance.

Thank You