

Desafio_v1.0

July 26, 2019

1 Data: Bank Marketing

1.1 O seguinte desafio pretende resolver uma série de perguntas associadas à base de dados que pode ser encontrada nos sites:

1.2 Fonte oficial do dataset : <https://archive.ics.uci.edu/ml/datasets/bank+marketing>

1.3 Dados: <https://archive.ics.uci.edu/ml/machine-learning-databases/00222/bank.zip>

Arquivos incluídos no link acima:

1. bank. csv: uma versão reduzida do conjunto de dados;
2. bank-full.csv: o conjunto completo;
3. bank-names.txt: com a descrição

As bases de dados encontram-se associadas aos artigos :

1. S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Telemarketing. Decision Support Systems, Elsevier, 62:22-31, June 2014
2. S. Moro, R. Laureano and P. Cortez. Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. In P. Novais et al. (Eds.), Proceedings of the European Simulation and Modelling Conference - ESM'2011, pp. 117-121, Guimaraes, Portugal, October, 2011. EUROSIS.

Os métodos de análise de sensibilidade e random forrest usados em > 1. podem ser vistos em:

3. Paulo Cortez, Mark J. Embrechts, Using sensitivity analysis and visualization techniques to open black box data mining models, Information Sciences 225 (2013) 1–17.

1.4 Inicialmente é realizada uma análise preliminar dos dados, empregando histogramas.

1.5 As graficas são elaboradas usando as bibliotecas e dados a seguir:

```
In [110]: import pandas as pd
          import matplotlib.pyplot as plt
          import numpy as np
          ### arquivo csv
```

```
col_df = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/bank-full.csv', sep=';')
print("Conjunto de dados a serem usados na análise:")
print(col_df)
plt.close("all")
```

Conjunto de dados a serem usados na análise:

	age	job	marital	education	default	balance	housing	loan	\
0	58	management	married	tertiary	no	2143	yes	no	
1	44	technician	single	secondary	no	29	yes	no	
2	33	entrepreneur	married	secondary	no	2	yes	yes	
3	47	blue-collar	married	unknown	no	1506	yes	no	
4	33	unknown	single	unknown	no	1	no	no	
5	35	management	married	tertiary	no	231	yes	no	
6	28	management	single	tertiary	no	447	yes	yes	
7	42	entrepreneur	divorced	tertiary	yes	2	yes	no	
8	58	retired	married	primary	no	121	yes	no	
9	43	technician	single	secondary	no	593	yes	no	
10	41	admin.	divorced	secondary	no	270	yes	no	
11	29	admin.	single	secondary	no	390	yes	no	
12	53	technician	married	secondary	no	6	yes	no	
13	58	technician	married	unknown	no	71	yes	no	
14	57	services	married	secondary	no	162	yes	no	
15	51	retired	married	primary	no	229	yes	no	
16	45	admin.	single	unknown	no	13	yes	no	
17	57	blue-collar	married	primary	no	52	yes	no	
18	60	retired	married	primary	no	60	yes	no	
19	33	services	married	secondary	no	0	yes	no	
20	28	blue-collar	married	secondary	no	723	yes	yes	
21	56	management	married	tertiary	no	779	yes	no	
22	32	blue-collar	single	primary	no	23	yes	yes	
23	25	services	married	secondary	no	50	yes	no	
24	40	retired	married	primary	no	0	yes	yes	
25	44	admin.	married	secondary	no	-372	yes	no	
26	39	management	single	tertiary	no	255	yes	no	
27	52	entrepreneur	married	secondary	no	113	yes	yes	
28	46	management	single	secondary	no	-246	yes	no	
29	36	technician	single	secondary	no	265	yes	yes	
...	
45181	46	blue-collar	married	secondary	no	6879	no	no	
45182	34	technician	married	secondary	no	133	no	no	
45183	70	retired	married	primary	no	324	no	no	
45184	63	retired	married	secondary	no	1495	no	no	
45185	60	services	married	tertiary	no	4256	yes	no	
45186	59	unknown	married	unknown	no	1500	no	no	
45187	32	services	single	secondary	no	1168	yes	no	
45188	29	management	single	secondary	no	703	yes	no	
45189	25	services	single	secondary	no	199	no	no	
45190	32	blue-collar	married	secondary	no	136	no	no	

45191	75	retired	divorced	tertiary	no	3810	yes	no
45192	29	management	single	tertiary	no	765	no	no
45193	28	self-employed	single	tertiary	no	159	no	no
45194	59	management	married	tertiary	no	138	yes	yes
45195	68	retired	married	secondary	no	1146	no	no
45196	25	student	single	secondary	no	358	no	no
45197	36	management	single	secondary	no	1511	yes	no
45198	37	management	married	tertiary	no	1428	no	no
45199	34	blue-collar	single	secondary	no	1475	yes	no
45200	38	technician	married	secondary	no	557	yes	no
45201	53	management	married	tertiary	no	583	no	no
45202	34	admin.	single	secondary	no	557	no	no
45203	23	student	single	tertiary	no	113	no	no
45204	73	retired	married	secondary	no	2850	no	no
45205	25	technician	single	secondary	no	505	no	yes
45206	51	technician	married	tertiary	no	825	no	no
45207	71	retired	divorced	primary	no	1729	no	no
45208	72	retired	married	secondary	no	5715	no	no
45209	57	blue-collar	married	secondary	no	668	no	no
45210	37	entrepreneur	married	secondary	no	2971	no	no

	contact	day	month	duration	campaign	pdays	previous	poutcome	y
0	unknown	5	may	261	1	-1	0	unknown	no
1	unknown	5	may	151	1	-1	0	unknown	no
2	unknown	5	may	76	1	-1	0	unknown	no
3	unknown	5	may	92	1	-1	0	unknown	no
4	unknown	5	may	198	1	-1	0	unknown	no
5	unknown	5	may	139	1	-1	0	unknown	no
6	unknown	5	may	217	1	-1	0	unknown	no
7	unknown	5	may	380	1	-1	0	unknown	no
8	unknown	5	may	50	1	-1	0	unknown	no
9	unknown	5	may	55	1	-1	0	unknown	no
10	unknown	5	may	222	1	-1	0	unknown	no
11	unknown	5	may	137	1	-1	0	unknown	no
12	unknown	5	may	517	1	-1	0	unknown	no
13	unknown	5	may	71	1	-1	0	unknown	no
14	unknown	5	may	174	1	-1	0	unknown	no
15	unknown	5	may	353	1	-1	0	unknown	no
16	unknown	5	may	98	1	-1	0	unknown	no
17	unknown	5	may	38	1	-1	0	unknown	no
18	unknown	5	may	219	1	-1	0	unknown	no
19	unknown	5	may	54	1	-1	0	unknown	no
20	unknown	5	may	262	1	-1	0	unknown	no
21	unknown	5	may	164	1	-1	0	unknown	no
22	unknown	5	may	160	1	-1	0	unknown	no
23	unknown	5	may	342	1	-1	0	unknown	no
24	unknown	5	may	181	1	-1	0	unknown	no
25	unknown	5	may	172	1	-1	0	unknown	no

26	unknown	5	may	296	1	-1	0	unknown	no
27	unknown	5	may	127	1	-1	0	unknown	no
28	unknown	5	may	255	2	-1	0	unknown	no
29	unknown	5	may	348	1	-1	0	unknown	no
...
45181	cellular	15	nov	74	2	118	3	failure	no
45182	cellular	15	nov	401	2	187	5	success	yes
45183	cellular	15	nov	78	1	96	7	success	no
45184	cellular	16	nov	138	1	22	5	success	no
45185	cellular	16	nov	200	1	92	4	success	yes
45186	cellular	16	nov	280	1	104	2	failure	no
45187	cellular	16	nov	411	1	-1	0	unknown	yes
45188	cellular	16	nov	236	1	550	2	success	yes
45189	cellular	16	nov	173	1	92	5	failure	no
45190	cellular	16	nov	206	1	188	3	success	yes
45191	cellular	16	nov	262	1	183	1	failure	yes
45192	cellular	16	nov	238	1	-1	0	unknown	yes
45193	cellular	16	nov	449	2	33	4	success	yes
45194	cellular	16	nov	162	2	187	5	failure	no
45195	cellular	16	nov	212	1	187	6	success	yes
45196	cellular	16	nov	330	1	-1	0	unknown	yes
45197	cellular	16	nov	270	1	-1	0	unknown	yes
45198	cellular	16	nov	333	2	-1	0	unknown	no
45199	cellular	16	nov	1166	3	530	12	other	no
45200	cellular	16	nov	1556	4	-1	0	unknown	yes
45201	cellular	17	nov	226	1	184	4	success	yes
45202	cellular	17	nov	224	1	-1	0	unknown	yes
45203	cellular	17	nov	266	1	-1	0	unknown	yes
45204	cellular	17	nov	300	1	40	8	failure	yes
45205	cellular	17	nov	386	2	-1	0	unknown	yes
45206	cellular	17	nov	977	3	-1	0	unknown	yes
45207	cellular	17	nov	456	2	-1	0	unknown	yes
45208	cellular	17	nov	1127	5	184	3	success	yes
45209	telephone	17	nov	508	4	-1	0	unknown	no
45210	cellular	17	nov	361	2	188	11	other	no

[45211 rows x 17 columns]

2 construção das classes para os histogramas

In [111]: `class Dados_iniciais:`

```

def __init__(self, col_df):
    self.col_df = col_df
    self.col_df_grouped = col_df.groupby("y")
    self.nome_classe_nao = "no"

```

```

self.nome_classe_sim = "yes"
self.col_df_grouped_nao = self.col_df_grouped.get_group(self.nome_classe_nao)
self.col_df_grouped_sim = self.col_df_grouped.get_group(self.nome_classe_sim)

def plot_histograma_continuo(self, nome_carateristica, bin_tamanho):
    plt.figure()
    plt.hist(self.col_df_grouped_nao[nome_carateristica], bins=bin_tamanho, label='nao')
    plt.hist(self.col_df_grouped_sim[nome_carateristica], bins=bin_tamanho, label='sim')
    plt.legend()
    plt.title("Histograma de Carateristicas - "+nome_carateristica)
    plt.xlabel("Valores Carateristicos")
    plt.ylabel("grouped")

def plot_histogram_categorical(self, nome_carateristica):
    carateristica_df = pd.DataFrame()
    carateristica_df["no"] = self.col_df_grouped_nao[nome_carateristica].value_counts()
    carateristica_df["yes"] = self.col_df_grouped_sim[nome_carateristica].value_counts()

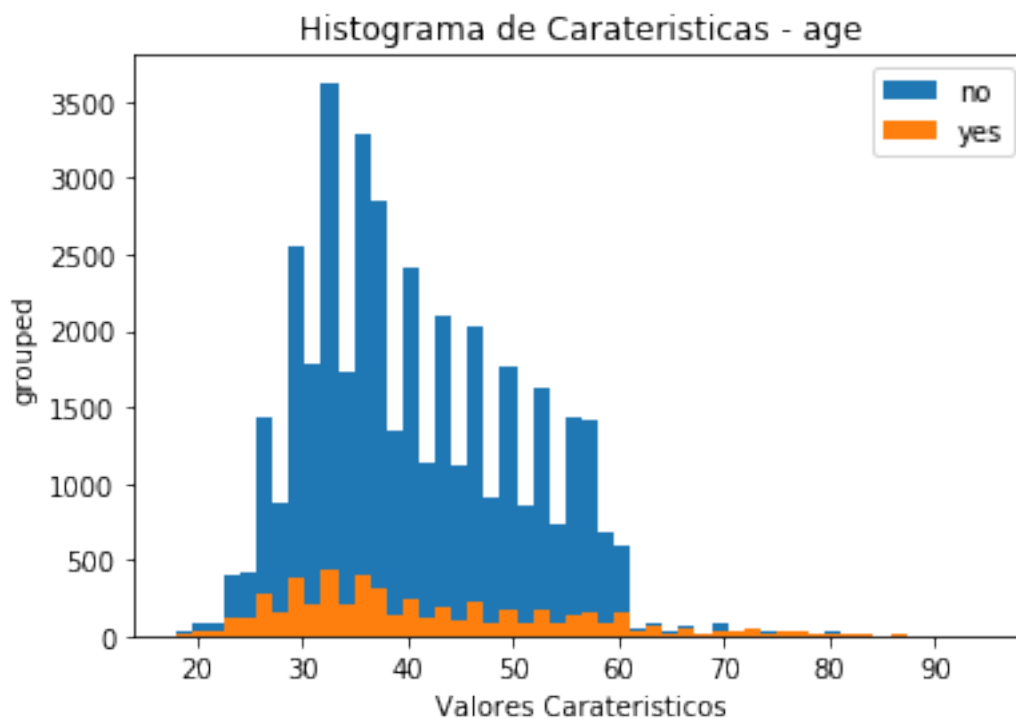
    carateristica_df.plot(kind='bar')
    plt.title("Histograma de Carateristicas - "+nome_carateristica)
    plt.ylabel("grouped")
    plt.xlabel("Valores unicos carateristicos")
    plt.tight_layout()

# uso da classe dados iniciais
Dados_iniciais_obj = Dados_iniciais(col_df)

```

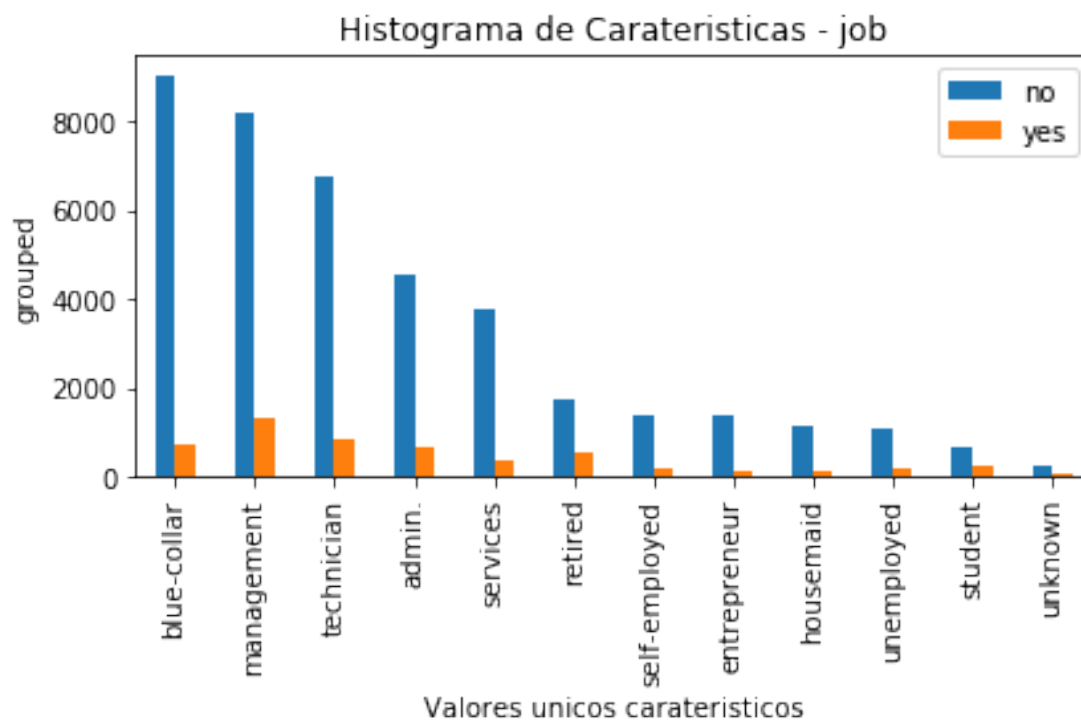
In [112]: *# 1 age*

```
Dados_iniciais_obj.plot_histograma_continuo("age", 50)
```



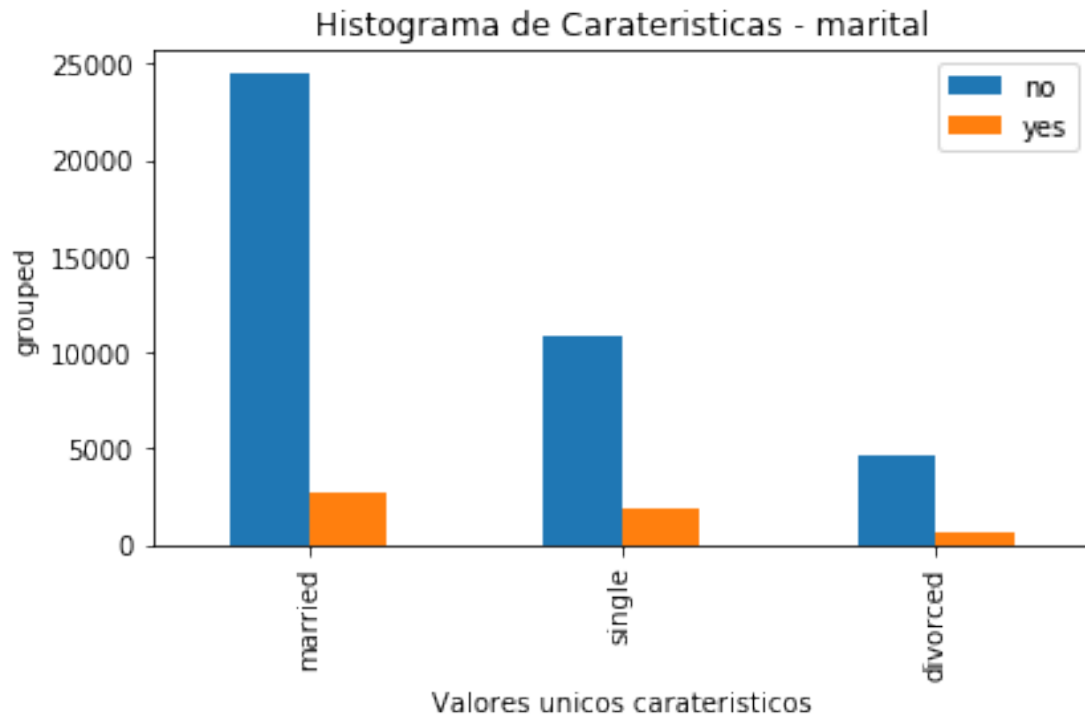
In [113]: # 2 job

```
Dados_iniciais_obj.plot_histogram_categorical("job")
```



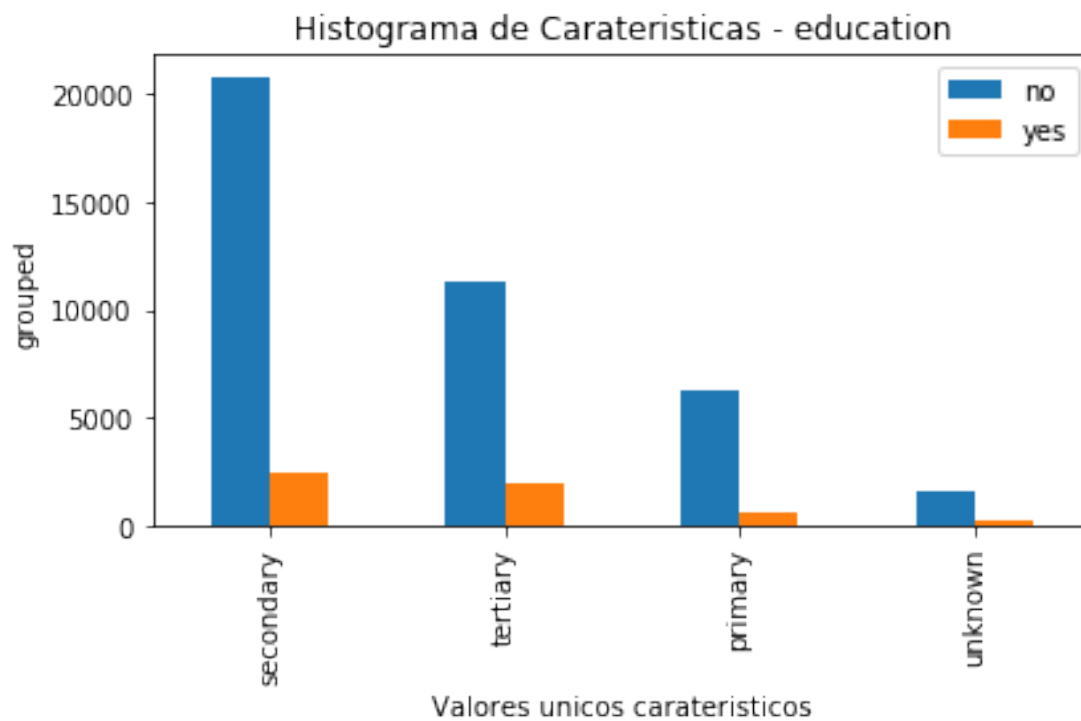
```
In [114]: # 3 marital
```

```
Dados_iniciais_obj.plot_histogram_categorical("marital")
```



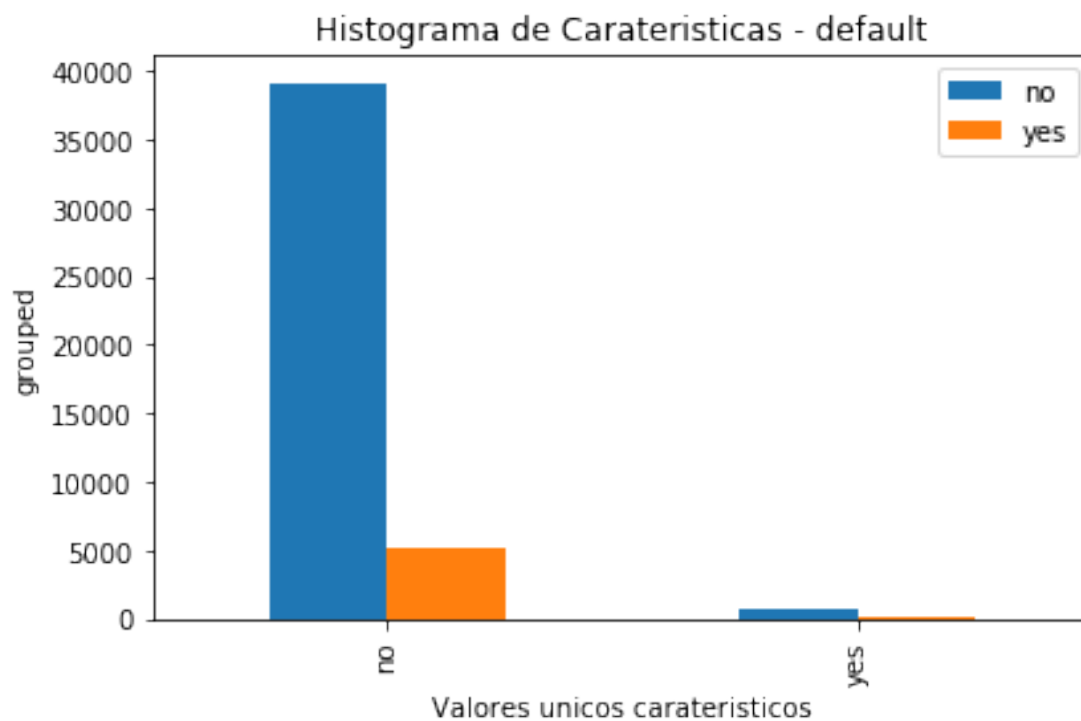
```
In [115]: # 4 education
```

```
Dados_iniciais_obj.plot_histogram_categorical("education")
```



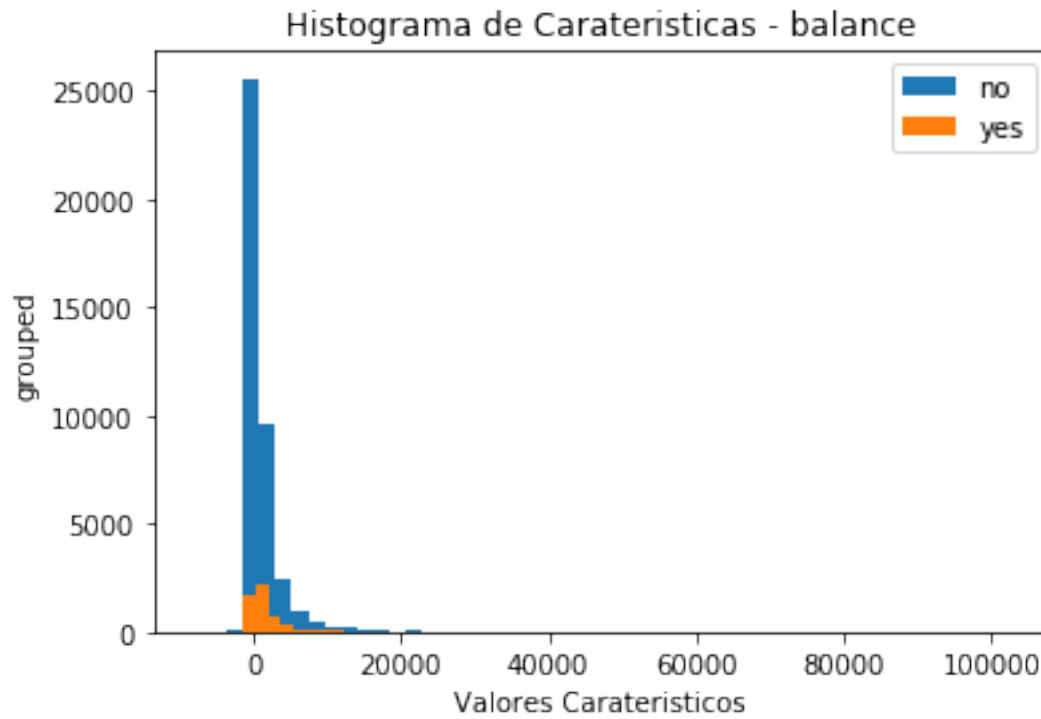
In [116]: # 5 default

Dados_iniciais_obj.plot_histogram_categorical("default")



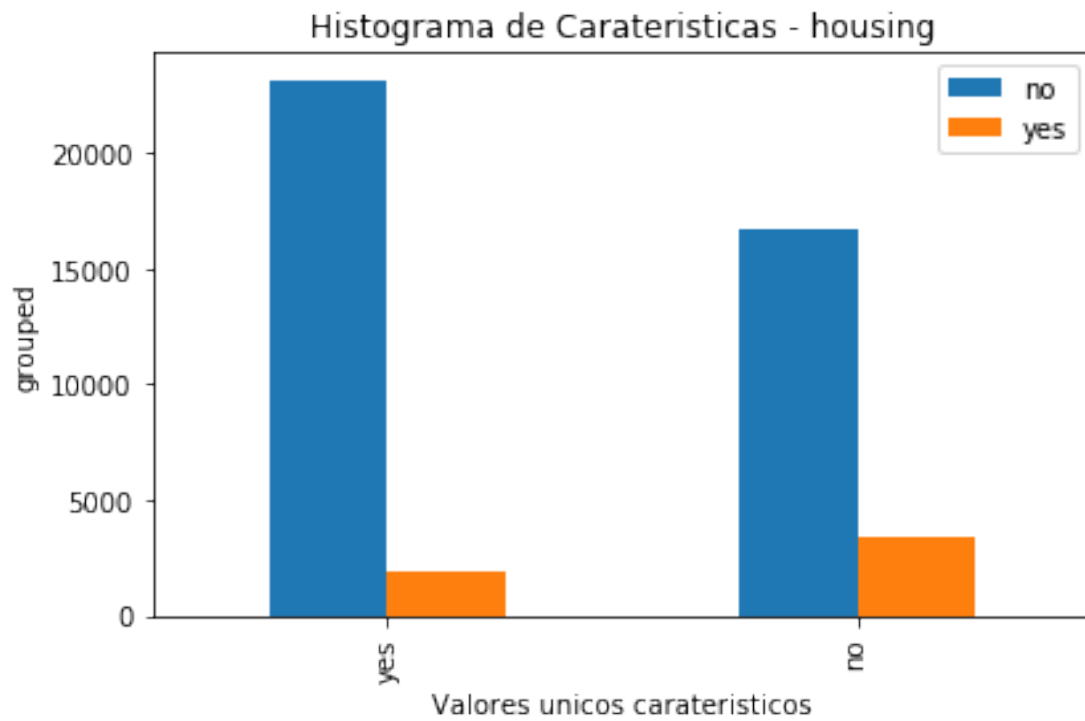

```
In [117]: # 6 balance
```

```
Dados_iniciais_obj.plot_histograma_continuo("balance", 50)
```



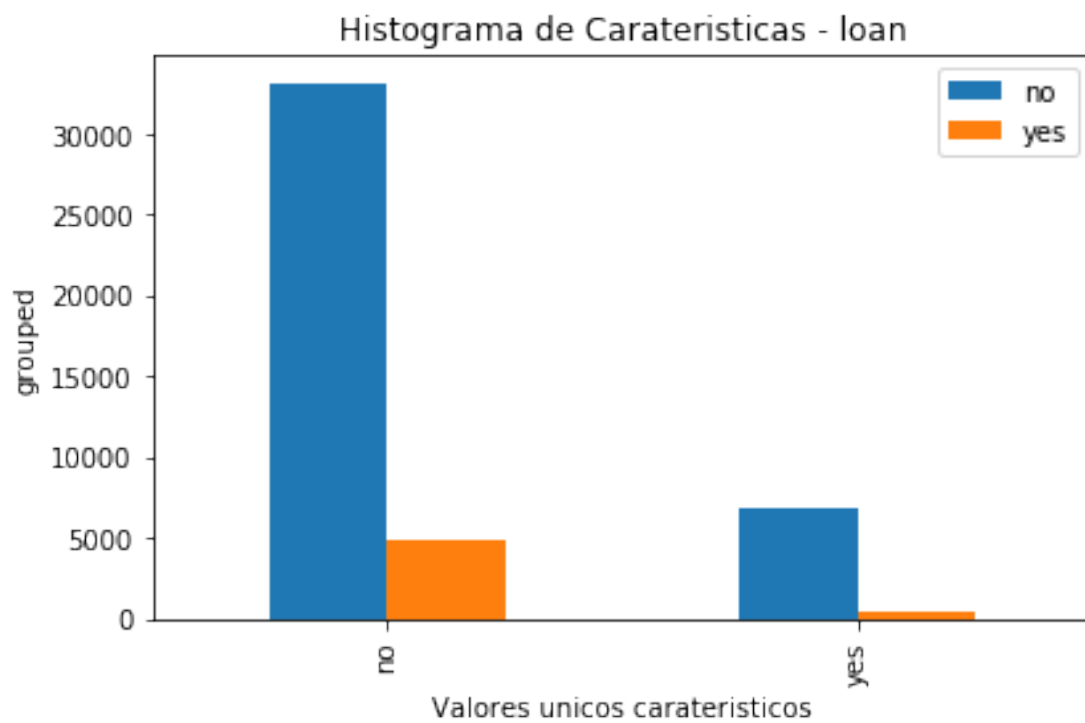
```
In [118]: # 7 housing
```

```
Dados_iniciais_obj.plot_histogram_categorical("housing")
```



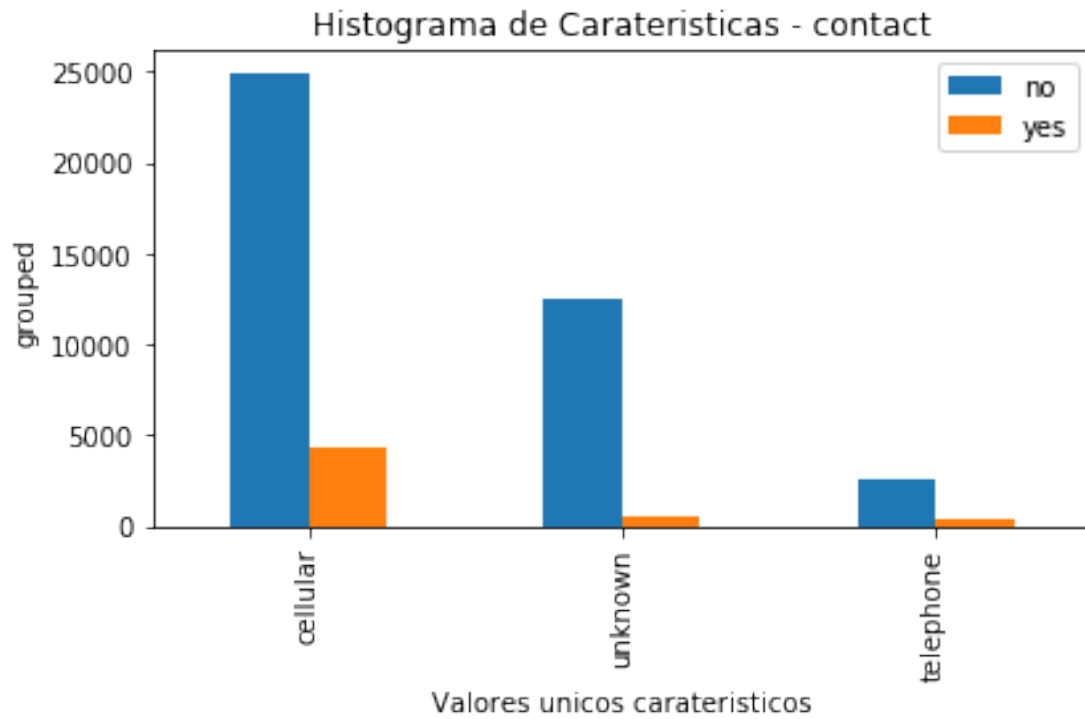
In [119]: # 8 loan

```
Dados_iniciais_obj.plot_histogram_categorical("loan")
```



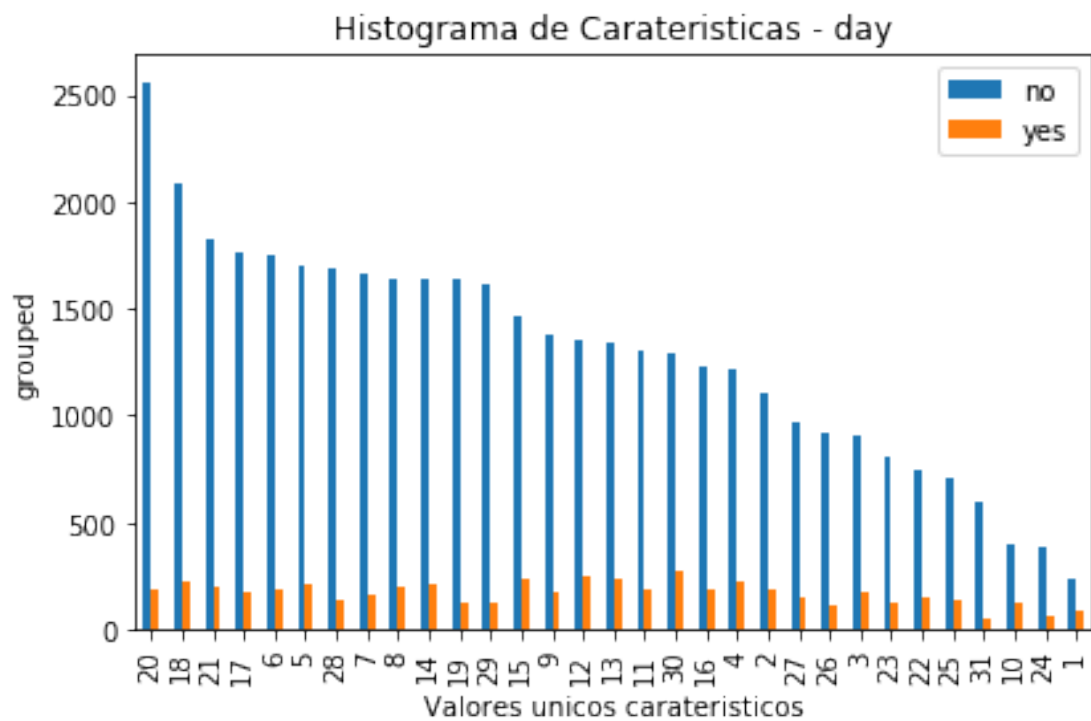
```
In [120]: # 9 contact
```

```
Dados_iniciais_obj.plot_histogram_categorical("contact")
```



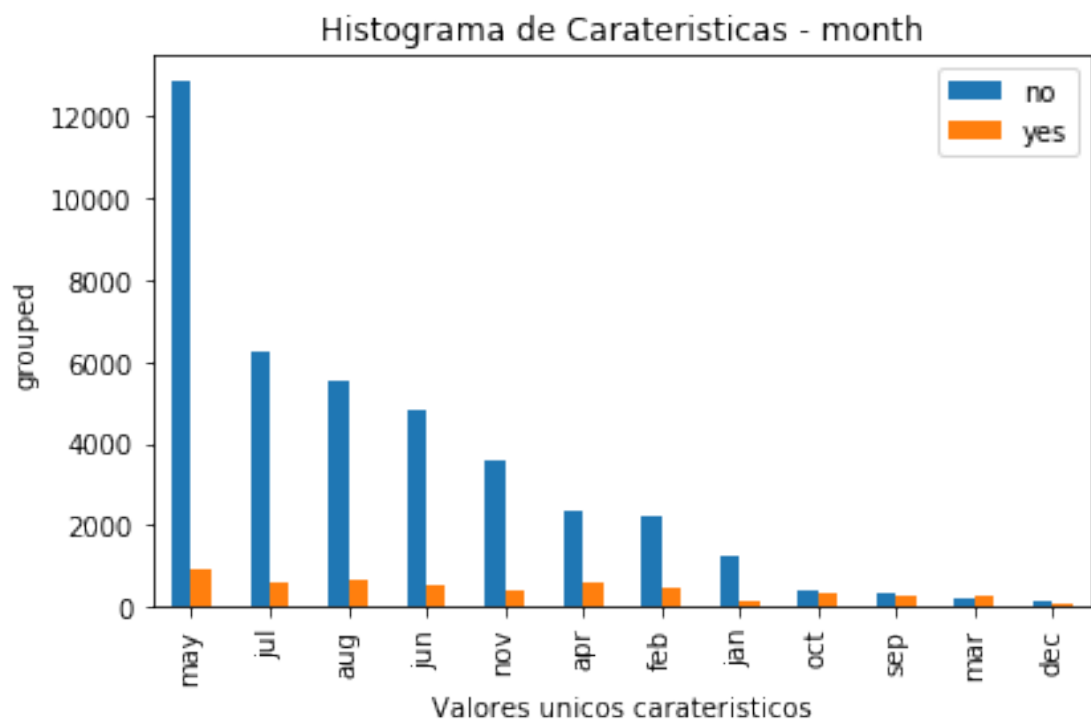
```
In [121]: # 10 day
```

```
Dados_iniciais_obj.plot_histogram_categorical("day")
```



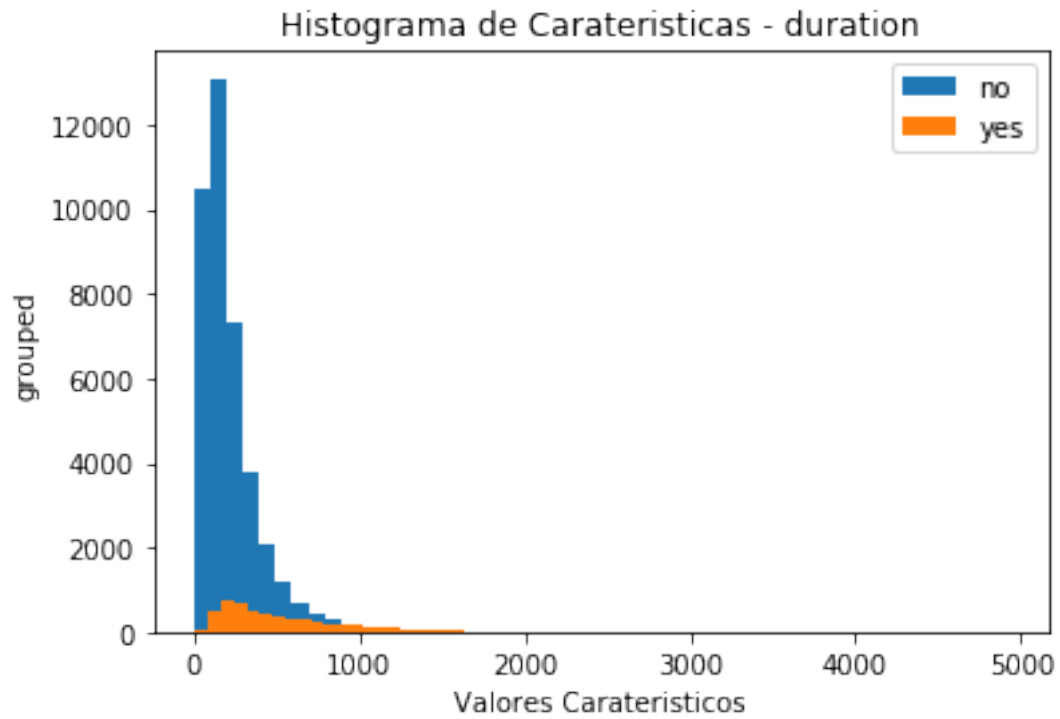
In [122]: # 11 month

```
Dados_iniciais_obj.plot_histogram_categorical("month")
```



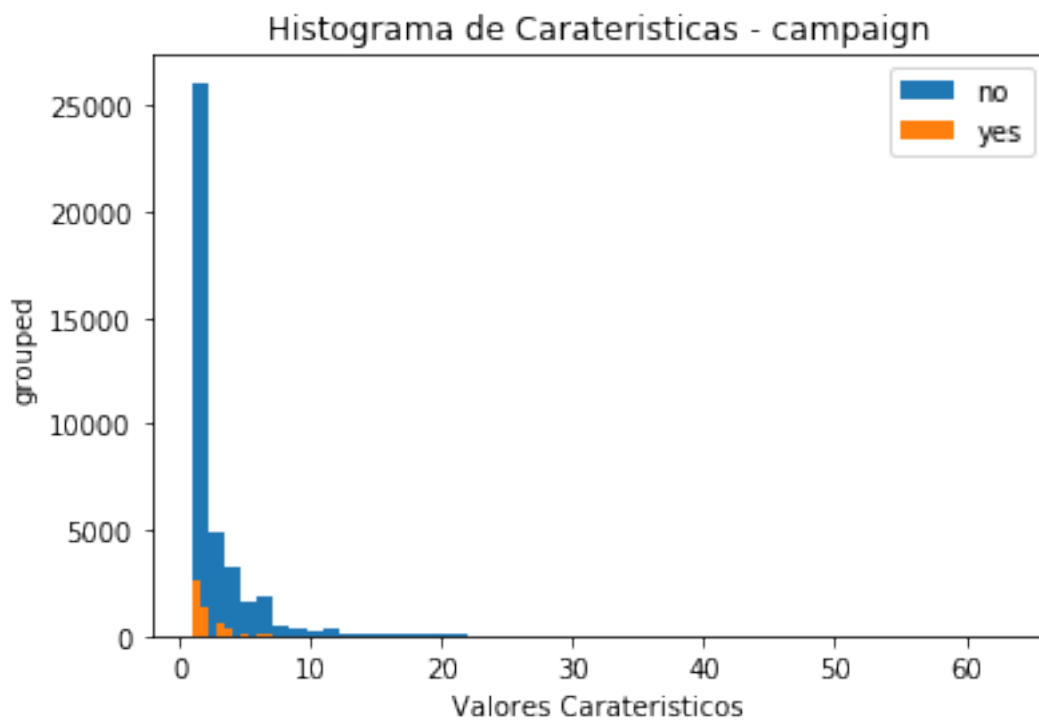
```
In [123]: # 12 duration
```

```
Dados_iniciais_obj.plot_historama_continuo("duration",50)
```

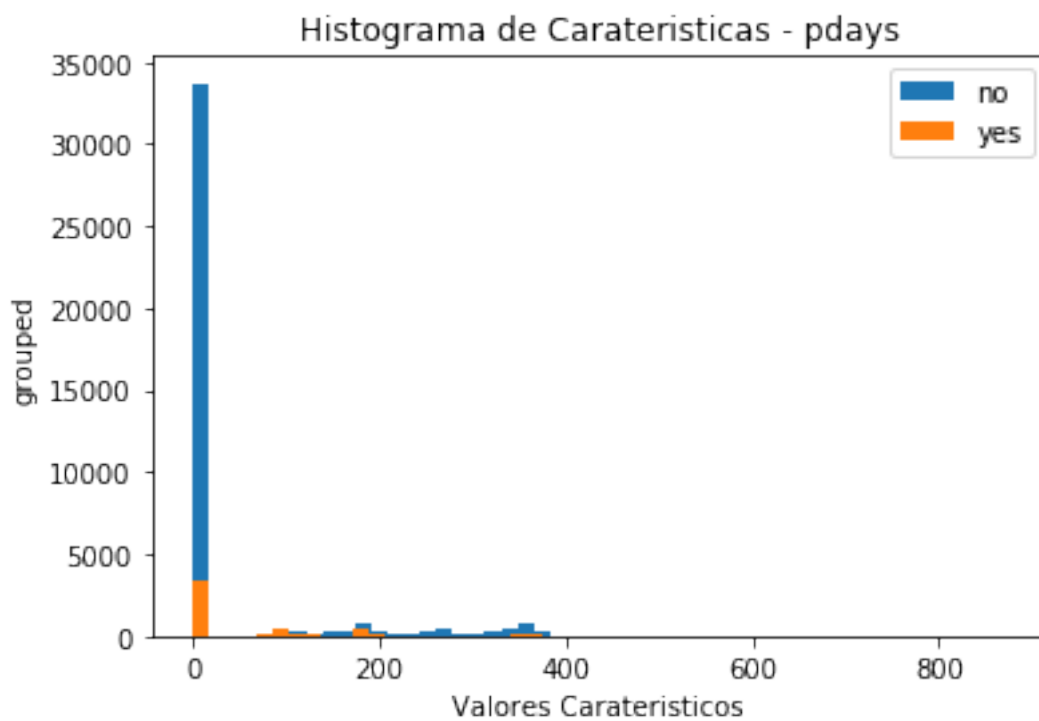


```
In [124]: # 13 campaign
```

```
Dados_iniciais_obj.plot_historama_continuo("campaign", 50)
```

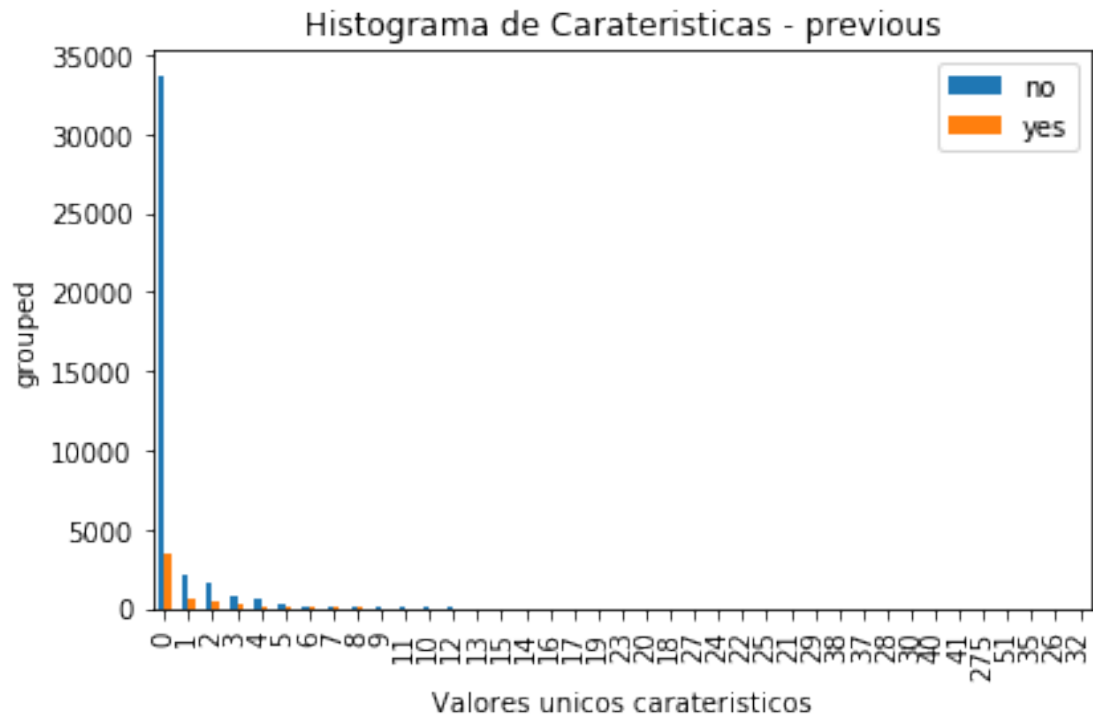


```
In [125]: # 14 pdays
Dados_iniciais_obj.plot_histograma_continuo("pdays", 50)
```



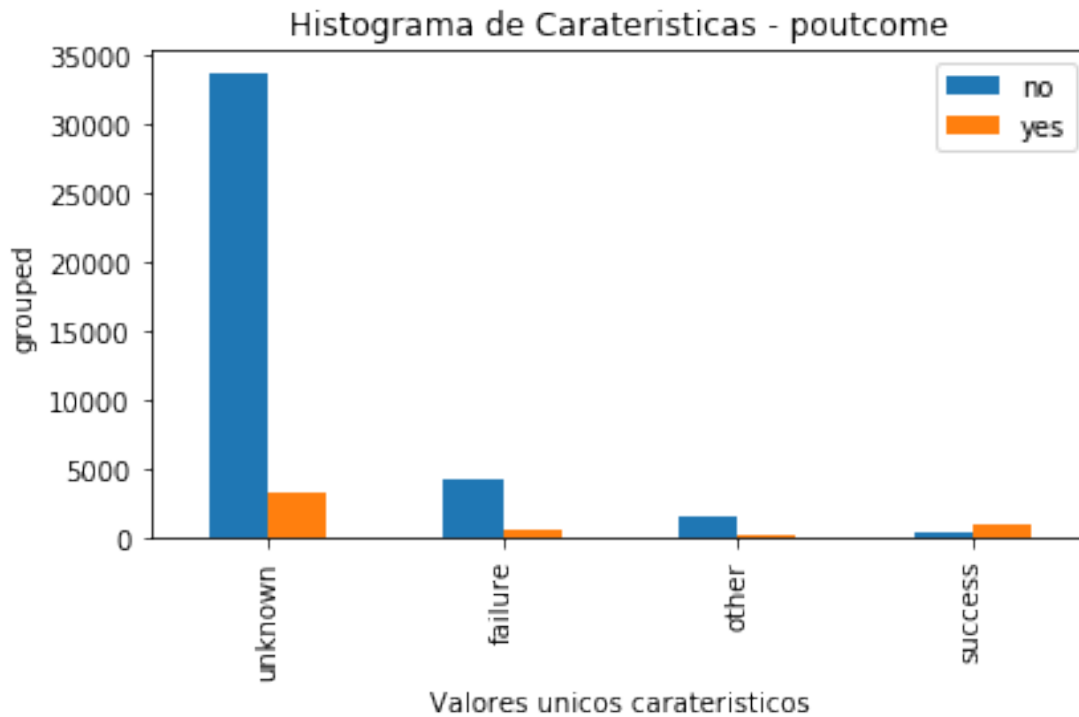
```
In [126]: # 15 previous
```

```
Dados_iniciais_obj.plot_histogram_categorical("previous")
```



```
In [127]: # 16 poutcome
```

```
Dados_iniciais_obj.plot_histogram_categorical("poutcome")
```



```
In [128]: import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder
from sklearn.preprocessing import OneHotEncoder
from sklearn.preprocessing import StandardScaler
from sklearn.preprocessing import MinMaxScaler
from sklearn.svm import SVC

print('Estadisticos:')
descrip_stat=col_df.describe()
print (descrip_stat)
```

Estadisticos:

	age	balance	day	duration	campaign \
count	45211.000000	45211.000000	45211.000000	45211.000000	45211.000000
mean	40.936210	1362.272058	15.806419	258.163080	2.763841
std	10.618762	3044.765829	8.322476	257.527812	3.098021
min	18.000000	-8019.000000	1.000000	0.000000	1.000000
25%	33.000000	72.000000	8.000000	103.000000	1.000000
50%	39.000000	448.000000	16.000000	180.000000	2.000000
75%	48.000000	1428.000000	21.000000	319.000000	3.000000
max	95.000000	102127.000000	31.000000	4918.000000	63.000000

	pdays	previous
--	-------	----------


```

count    45211.000000    45211.000000
mean      40.197828      0.580323
std       100.128746      2.303441
min        -1.000000      0.000000
25%        -1.000000      0.000000
50%        -1.000000      0.000000
75%        -1.000000      0.000000
max       871.000000     275.000000

```

```

In [129]: print('Correlação Canônica entre as variáveis de entrada:')
          descrip_corr=col_df.corr()
          print (descrip_corr)

```

Correlação Canônica entre as variáveis de entrada:

	age	balance	day	duration	campaign	pdays	previous
age	1.000000	0.097783	-0.009120	-0.004648	0.004760	-0.023758	0.001288
balance	0.097783	1.000000	0.004503	0.021560	-0.014578	0.003435	0.016674
day	-0.009120	0.004503	1.000000	-0.030206	0.162490	-0.093044	-0.051710
duration	-0.004648	0.021560	-0.030206	1.000000	-0.084570	-0.001565	0.001203
campaign	0.004760	-0.014578	0.162490	-0.084570	1.000000	-0.088628	-0.032855
pdays	-0.023758	0.003435	-0.093044	-0.001565	-0.088628	1.000000	0.454820
previous	0.001288	0.016674	-0.051710	0.001203	-0.032855	0.454820	1.000000

2.0.1 Como pode ser apreciado na tabela acima as maiores correlações encontram-se nas variáveis *previous* e *pdays*, seguidas de *campaign* e *day*. O qual deve ser levado em conta na análise dos dados.

2.1 os dados apresentados nos histogramas anteriores, podem ser tratados a fim de reduzir o erro devido viés e distribuição. Para isto a seguir apresenta-se a estandarização dos dados

2.2 a estandarização e o código de avaliação do modelo apresentada a seguir basea-se no relatório apresentado em GitHub de ABINAYA MANIMARAN.

```

In [130]: # a estandarizacao

```

```

def caracteristicas_preprocessadas(df_train, df_test, process_continuous):

```

```

    to_delete_features = []

```

```

    continuous_features = ['age', 'balance', 'duration', 'campaign', 'pdays', 'previous']

```

```

    categorical_ordered_features = ['marital', 'default', 'education', 'housing', 'loan_status']

```

```

    categorical_unordered_features = ['job']

```

```

    unknown_present_features = ['contact', 'outcome', 'job', 'education']

```

```

    all_present_features = ['age', 'balance', 'duration', 'campaign', 'pdays', 'previous']

```

```

    ### Delete Features

```

```

for feat in to_delete_features:
    print ("\n----- deleting feature ----- ",feat)
    del df_train[feat]
    del df_test[feat]

### Normalization or Standardization of Continuous Features
if process_continuous == "Standardize":
    print ("\n----- Standardizing Continuous Features (Mean=0, Standard Deviation=1) -----")
    standardization = StandardScaler()
    standardization.fit(df_train[continuous_features])
    print ("Mean: ",standardization.mean_)
    print ("Variance: ",standardization.var_)
    df_train[continuous_features] = standardization.transform(df_train[continuous_features])
    df_test[continuous_features] = standardization.transform(df_test[continuous_features])

elif process_continuous == "Normalize":
    print ("\n----- Normalizing Continuous Features (Min=0, Max=1) -----")
    min_max_scaling = MinMaxScaler()
    min_max_scaling.fit(df_train[continuous_features])
    print (min_max_scaling.data_min_)
    print (min_max_scaling.data_max_)
    df_train[continuous_features] = min_max_scaling.transform(df_train[continuous_features])
    df_test[continuous_features] = min_max_scaling.transform(df_test[continuous_features])

### Label Categorical Ordered Features -- Features used for Imputation (All Present)
label_dict = {'education':{'primary':0, 'secondary':1, 'tertiary':2},
              'housing':{'no':0, 'yes':1},
              'loan':{'no':0, 'yes':1},
              'contact':{'telephone':0, 'cellular':1},
              'month':{'jan':1, 'feb':2, 'mar':3, 'apr':4, 'may':5, 'jun':6, 'jul':7, 'aug':8},
              'default':{'no':0, 'yes':1},
              'marital':{'single':0, 'married':1, 'divorced':2},
              'poutcome':{'other':0, 'failure':1, 'success':2}}

for feat in categorical_ordered_features:
    if feat not in unknown_present_features:
        print ("\n----- Labelling feature Before Imputation ----- ",feat)
        df_train = df_train.replace({feat:label_dict[feat]})
        df_test = df_test.replace({feat:label_dict[feat]})
        print ("Labelled as: ",label_dict[feat])

### Imputation using SVM
df_train_impute = df_train.loc[:,df_train.columns.isin(all_present_features)]
df_test_impute = df_test.loc[:,df_test.columns.isin(all_present_features)]

for feat in unknown_present_features:
    print ("\nFilling Unknowns for Feature: ",feat)

```

```

train_impute = df_train[feat]
test_impute = df_test[feat]

train_impute_no_unknowns = train_impute[train_impute != 'unknown']
train_impute_unknowns = train_impute[train_impute == 'unknown']
test_impute_unknowns = test_impute[test_impute == 'unknown']

df_train_impute_train_features = df_train_impute.loc[train_impute_no_unknowns.index]
df_train_impute_test_features = df_train_impute.loc[train_impute_unknowns.index]
df_test_impute_test_features = df_test_impute.loc[test_impute_unknowns.index]

svm_model = SVC()
svm_model.fit(df_train_impute_train_features, train_impute_no_unknowns)
df_train.loc[df_train_impute_test_features.index, feat] = svm_model.predict(df_train_impute_test_features)
print ("Train Filled with: ",df_train.loc[df_train_impute_test_features.index, feat])
df_test.loc[df_test_impute_test_features.index, feat] = svm_model.predict(df_test_impute_test_features)
print ("Test Filled with: ", df_test.loc[df_test_impute_test_features.index, feat])

### Label Categorical Ordered Features -- Features Imputed (Unknowns were Present)
for feat in categorical_ordered_features:
    if feat in unknown_present_features:
        print ("\n----- Labelling feature After Imputation ----- ",feat)
        df_train = df_train.replace({feat:label_dict[feat]})
        df_test = df_test.replace({feat:label_dict[feat]})
        print ("Labelled as: ",label_dict[feat])

### One hot encoding Categorical Un-ordered Features
for feat in categorical_unordered_features:
    print ("\n----- One Hot Encoding feature ----- ",feat)
    label_encoder = LabelEncoder()
    label_encoder.fit(df_train[feat])
    df_train[feat] = label_encoder.transform(df_train[feat])
    df_test[feat] = label_encoder.transform(df_test[feat])

one_hot_encoder = OneHotEncoder(sparse=False)
one_hot_encoder.fit(df_train[categorical_unordered_features])
one_hot_encoded_array_train = one_hot_encoder.transform(df_train[categorical_unordered_features])
one_hot_encoded_df_train = pd.DataFrame(one_hot_encoded_array_train, index=df_train.index)
one_hot_encoded_array_test = one_hot_encoder.transform(df_test[categorical_unordered_features])
one_hot_encoded_df_test = pd.DataFrame(one_hot_encoded_array_test, index=df_test.index)

df_train = pd.concat([df_train,one_hot_encoded_df_train], axis=1) #concatenate old df with new df
df_test = pd.concat([df_test,one_hot_encoded_df_test], axis=1) #concatenate old df with new df

df_train = df_train.drop(categorical_unordered_features, axis=1) #Delete columns
df_test = df_test.drop(categorical_unordered_features, axis=1) #Delete columns

### Return pre-processed df

```

```

    return df_train, df_test

def classe_preprocessamento(df_train, df_test):
    "\n----- Labelling Class Information ----- "
    class_col = 'y'
    label_dict = {class_col: {'no': 0, 'yes': 1}}
    df_train = df_train.replace({class_col: label_dict[class_col]})
    df_test = df_test.replace({class_col: label_dict[class_col]})
    return df_train, df_test

```

In [131]: # rotina usada para a avaliação da qualidade do modelo usando criterios ROC e AUC

```

import matplotlib.pyplot as plt
from sklearn.metrics import accuracy_score, confusion_matrix
from sklearn.metrics import precision_score, recall_score, f1_score
from sklearn.metrics import roc_curve, auc
from sklearn.metrics import classification_report

def avaliacao_PerformanceC(df_train_class, predicted_train, predicted_prob_train, df_test_class, predicted_test, predicted_prob_test):
    ### Confusion Matrix
    confusion_matrix_train = confusion_matrix(df_train_class, predicted_train)
    confusion_matrix_test = confusion_matrix(df_test_class, predicted_test)
    print ("\nTraining Confusion Matrix:\n ", confusion_matrix_train)
    print ("\nTesting Confusion Matrix:\n ", confusion_matrix_test)

    ### Accuracy score
    score_train = accuracy_score(df_train_class, predicted_train)
    score_test = accuracy_score(df_test_class, predicted_test)
    print ("\nTraining Accuracy Score: ", score_train)
    print ("\nTesting Accuracy Score: ", score_test)

    ### Precision, Recall
    precision_train = precision_score(df_train_class, predicted_train)
    precision_test = precision_score(df_test_class, predicted_test)
    print ("\nTraining Precision: ", precision_train)
    print ("\nTesting Precision: ", precision_test)

    recall_train = recall_score(df_train_class, predicted_train)
    recall_test = recall_score(df_test_class, predicted_test)
    print ("\nTraining Recall: ", recall_train)
    print ("\nTesting Recall: ", recall_test)

    ### Classification Report
    print ("\nTrain Classification Report: \n", classification_report(df_train_class, predicted_train, target_names=['no', 'yes']))
    print ("\nTest Classification Report: \n", classification_report(df_test_class, predicted_test, target_names=['no', 'yes']))

    ### F1 Score
    f1score_train = f1_score(df_train_class, predicted_train) #, average='weighted')
    f1score_test = f1_score(df_test_class, predicted_test) #, average='weighted')

```

```

f1score_test = f1_score(df_test_class, predicted_test)#, average='weighted')
print ("\nTraining F1score: ", f1score_train)
print ("\nTesting F1score: ", f1score_test)

f1score_train = f1_score(df_train_class, predicted_train, average='weighted')
f1score_test = f1_score(df_test_class, predicted_test, average='weighted')
print ("\nTraining Weighted F1score: ", f1score_train)
print ("\nTesting Weighted F1score: ", f1score_test)

### ROC-AUC
if roc_y_n == 'y':
    fpr, tpr, threshold = roc_curve(df_train_class, predicted_prob_train[:,1])
    roc_auc_train = auc(fpr, tpr)
    print ("\nTraining AUC for ROC: ",roc_auc_train)
    plt.figure()
    plt.plot(fpr, tpr, 'b', label = 'AUC = %0.2f' % roc_auc_train)
    plt.plot([0, 1], [0, 1], 'r--')
    plt.xlim([0, 1])
    plt.ylim([0, 1])
    plt.ylabel('True Positive Rate')
    plt.xlabel('False Positive Rate')
    plt.legend(loc = 'lower right')
    plt.title('Training - Receiver Operating Characteristic')

    fpr, tpr, threshold = roc_curve(df_test_class, predicted_prob_test[:,1])
    roc_auc_test = auc(fpr, tpr)
    print ("\nTesting AUC for ROC: ",roc_auc_test)
    plt.figure()
    plt.plot(fpr, tpr, 'b', label = 'AUC = %0.2f' % roc_auc_test)
    plt.plot([0, 1], [0, 1], 'r--')
    plt.xlim([0, 1])
    plt.ylim([0, 1])
    plt.ylabel('True Positive Rate')
    plt.xlabel('False Positive Rate')
    plt.legend(loc = 'lower right')
    plt.title('Testing - Receiver Operating Characteristic')

```

2.3 A seguir usamos a estandarização acima dividimos dados para train e test e posteriormente salvamos os dados.

O procedimento a seguir permite:

1. substituir os atributos yes ou no contidos nas variáveis por valores numericos
2. substituir os labels unknow a través do uso de SVM para ajustar a mostra a valores conhecidos
3. permite a criação de subconjuntos de dados de treino e teste
4. separa as variáveis emtre numericas, categoricas ordenadas e categoricas não ordenadas

5. as variáveis categoricas não ordenadas como jobs são sub classificadas em labels

```
In [132]: from sklearn.model_selection import train_test_split

def train_teste_dados(df):
    df_class = pd.DataFrame(df['y'])
    df_features = df.loc[:, df.columns != 'y']

    df_features_train, df_features_test, df_class_train, df_class_test = train_test_split(
        df_features, df_class, test_size=0.2, random_state=42)

    df_train = pd.concat([df_features_train, df_class_train], axis=1)
    df_test = pd.concat([df_features_test, df_class_test], axis=1)

    return df_train, df_test

df = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/bank-full.csv', sep=';')
df_train, df_test = train_teste_dados(df)
df_train, df_test = caracteristicas_preprocessadas(df_train, df_test, "Standardize")
df_train, df_test = classe_preprocessamento(df_train, df_test)

df_train.to_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_standard.csv')
df_test.to_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_standard_test.csv')

----- Standardizing Continuous Features (Mean=0, Standard Deviation=1) -----
Mean: [4.09314321e+01 1.36675192e+03 2.58952283e+02 2.75681255e+00
       4.01849416e+01 5.85171641e-01 1.57916126e+01]
Variance: [1.12572478e+02 9.30091391e+06 6.67362994e+04 9.60465602e+00
          1.00098333e+04 5.81783132e+00 6.89238391e+01]

----- Labelling feature Before Imputation ----- marital
Labelled as: {'single': 0, 'married': 1, 'divorced': 2}

----- Labelling feature Before Imputation ----- default
Labelled as: {'no': 0, 'yes': 1}

----- Labelling feature Before Imputation ----- housing
Labelled as: {'no': 0, 'yes': 1}

----- Labelling feature Before Imputation ----- loan
Labelled as: {'no': 0, 'yes': 1}

----- Labelling feature Before Imputation ----- month

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/data.py:645: DataConversionWarning:
  return self.partial_fit(X, y)
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:26: DataConversionWarning:
```

/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:27: DataConversionWarning

Labelled as: {'jan': 1, 'feb': 2, 'mar': 3, 'apr': 4, 'may': 5, 'jun': 6, 'jul': 7, 'aug': 8,

Filling Unknowns for Feature: contact

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:196: FutureWarning: The
"avoid this warning.", FutureWarning)

Train Filled with: cellular 9802
Name: contact, dtype: int64
Test Filled with: cellular 3218
Name: contact, dtype: int64

Filling Unknowns for Feature: poutcome

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:196: FutureWarning: The
"avoid this warning.", FutureWarning)

Train Filled with: failure 19426
success 6312
other 1953
Name: poutcome, dtype: int64
Test Filled with: failure 6483
success 2109
other 676
Name: poutcome, dtype: int64

Filling Unknowns for Feature: job

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:196: FutureWarning: The
"avoid this warning.", FutureWarning)

Train Filled with: management 104
blue-collar 79
retired 37
technician 3
student 1
Name: job, dtype: int64
Test Filled with: management 27
blue-collar 22
retired 14

```
technician      1
Name: job, dtype: int64
```

Filling Unkowns for Feature: education

```
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/svm/base.py:196: FutureWarning: The
"avoid this warning.", FutureWarning)
```

```
Train Filled with: secondary    1296
tertiary           84
primary           18
Name: education, dtype: int64
Test Filled with: secondary    428
tertiary           22
primary            9
Name: education, dtype: int64
```

```
----- Labelling feature After Imputation ----- education
Labelled as: {'primary': 0, 'secondary': 1, 'tertiary': 2}
```

```
----- Labelling feature After Imputation ----- contact
Labelled as: {'telephone': 0, 'cellular': 1}
```

```
----- Labelling feature After Imputation ----- poutcome
Labelled as: {'other': 0, 'failure': 1, 'success': 2}
```

```
----- One Hot Encoding feature ----- job
```

```
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/preprocessing/_encoders.py:371: Futu
If you want the future behaviour and silence this warning, you can specify "categories='auto'"
In case you used a LabelEncoder before this OneHotEncoder to convert the categories to integers:
warnings.warn(msg, FutureWarning)
```

2.4 Modelos obtidos e avaliação dos mesmos

2.5 Os modelos a seguir são obtidos a partir dos métodos

1. de regressão logística usando PCA
2. de Perceptron
3. de Random forrest

2.6 LR com PCA

```
In [133]: import numpy as np
          import pandas as pd
```



```

import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
from sklearn.decomposition import PCA
from sklearn.model_selection import StratifiedKFold

df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_
df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_

del df_train['Unnamed: 0']
del df_test['Unnamed: 0']

df_train_class = pd.DataFrame(df_train['y'])
df_train_features = df_train.loc[:, df_train.columns != 'y']

df_test_class = pd.DataFrame(df_test['y'])
df_test_features = df_test.loc[:, df_test.columns != 'y']

# Reducao usando Analise de Componentes Principais
modelo_PCA = PCA(n_components=5)
modelo_PCA.fit(df_train_features)
df_train_features = pd.DataFrame(modelo_PCA.transform(df_train_features))
df_test_features = pd.DataFrame(modelo_PCA.transform(df_test_features))

#Modelo de Regressao Logistica

C_list = np.linspace(0.1, 1, 10)
lista_punicao = ['l1', 'l2']

skf_model = StratifiedKFold(n_splits=5, shuffle=True)

# minimo de 3 iteracoes
numero_iteracoes = 3
for t in range(0, numero_iteracoes):
    print ("---Iteration: ", t)
    AVG_ACC = np.zeros(shape=[len(C_list), len(lista_punicao)])
    STD_ACC = np.zeros(shape=[len(C_list), len(lista_punicao)])

    x_count = 0
    for c_value in C_list:

        y_count = 0
        for punicao in lista_punicao:

            temp_accuracy_list = []
            for train_subset_index, cv_index in skf_model.split(df_train_features, df_
                df_train_features_subset = df_train_features.loc[train_subset_index]
                df_train_class_subset = df_train_class.loc[train_subset_index]

```

```

df_train_features_cv = df_train_features.loc[cv_index]
df_train_class_cv = df_train_class.loc[cv_index]

lr_model = LogisticRegression(penalty=punicao, C=c_value, class_weight=
lr_model.fit(df_train_features_subset, df_train_class_subset)
score_value = lr_model.score(df_train_features_cv, df_train_class_cv)
temp_accuracy_list.append(score_value)

AVG_ACC[x_count,y_count] = np.mean(temp_accuracy_list)
STD_ACC[x_count,y_count] = np.std(temp_accuracy_list)
y_count += 1

x_count += 1

if t==0:
    final_AVG_ACC = AVG_ACC
    final_STD_ACC = STD_ACC
else:
    final_AVG_ACC = np.dstack([final_AVG_ACC, AVG_ACC])
    final_STD_ACC = np.dstack([final_STD_ACC, STD_ACC])

final_accuracy_mean_list = np.mean(final_AVG_ACC, axis=2)
max_ind = np.unravel_index(np.argmax(final_accuracy_mean_list, axis=None), final_acc

Escolha_C = C_list[max_ind[0]]
Escolha_punicao = lista_punicao[max_ind[1]]
print ("Cross Validation - C pela Regressao Logistica: ",Escolha_C)
print ("Cross Validation - Punicao pela Regressao Logistica: ",Escolha_punicao)

RL_modelo_F = LogisticRegression(penalty=Escolha_punicao, C=Escolha_C, class_weight=
RL_modelo_F.fit(df_train_features, df_train_class)

Predicao_train = RL_modelo_F.predict(df_train_features)
Predicao_test = RL_modelo_F.predict(df_test_features)

Predicao_prob_train = RL_modelo_F.predict_proba(df_train_features)
Predicao_prob_test = RL_modelo_F.predict_proba(df_test_features)

avaliacao_PerformanceC(df_train_class, Predicao_train, Predicao_prob_train, df_test_

---Iteration: 0

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)

```

---Iteration: 2

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning
    y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning
    FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning

```



```

y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: Future
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConver
y = column_or_1d(y, warn=True)

```

Cross Validation - C pela Regressao Logistica: 0.5
Cross Validation - Punicao pela Regressao Logistica: 12

Training Confusion Matrix:
[[23495 6457]
[1200 2756]]

Testing Confusion Matrix:
[[7909 2061]
[376 957]]

Training Accuracy Score: 0.7741830836380795

Testing Accuracy Score: 0.7843935238432275

Training Precision: 0.29914251600998587

Testing Precision: 0.31709741550695825

Training Recall: 0.6966632962588474

Testing Recall: 0.7179294823705926

Train Classification Report:

	precision	recall	f1-score	support
0	0.95	0.78	0.86	29952
1	0.30	0.70	0.42	3956
micro avg	0.77	0.77	0.77	33908
macro avg	0.63	0.74	0.64	33908
weighted avg	0.88	0.77	0.81	33908

Test Classification Report:

	precision	recall	f1-score	support
0	0.95	0.79	0.87	9970
1	0.32	0.72	0.44	1333
micro avg	0.78	0.78	0.78	11303

macro avg	0.64	0.76	0.65	11303
weighted avg	0.88	0.78	0.82	11303

Training F1score: 0.4185587364264561

Testing F1score: 0.4398988738221099

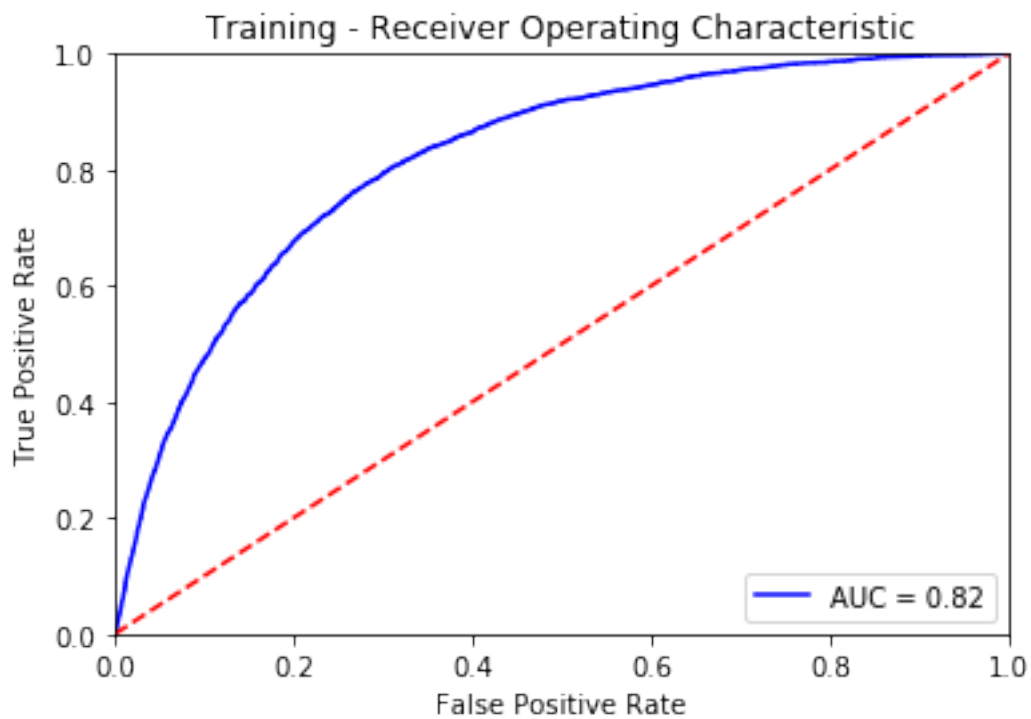
```
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/logistic.py:433: FutureWarning:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
```

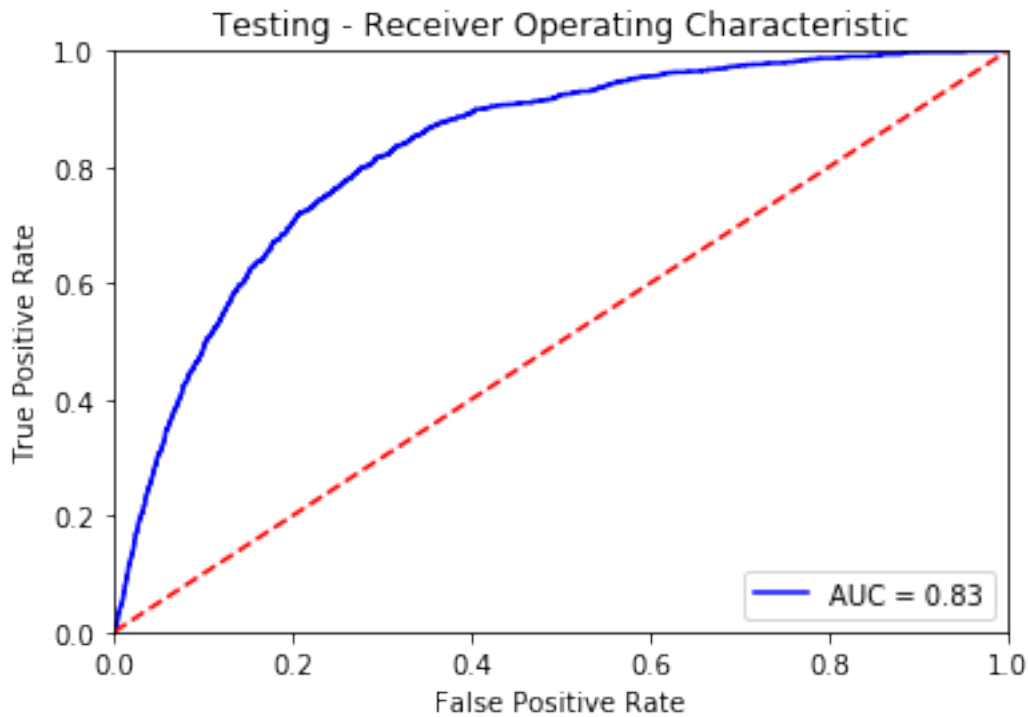
Training Weighted F1score: 0.8083938764225234

Testing Weighted F1score: 0.8161915783644864

Training AUC for ROC: 0.8191824816572034

Testing AUC for ROC: 0.8296755231937374





2.7 Perceptron

```
In [134]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.linear_model import Perceptron
from sklearn.calibration import CalibratedClassifierCV
from sklearn.model_selection import StratifiedKFold

df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_
df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_

del df_train['Unnamed: 0']
del df_test['Unnamed: 0']

df_train_class = pd.DataFrame(df_train['y'])
df_train_features = df_train.loc[:, df_train.columns != 'y']

df_test_class = pd.DataFrame(df_test['y'])
df_test_features = df_test.loc[:, df_test.columns != 'y']
```

```

# Perceptrao
lista_alpha = np.linspace(0.00001, 1, 15)
lista_punicao = ['l1', 'l2', 'elasticnet']

skf_model = StratifiedKFold(n_splits=5, shuffle=True)

numero_iteracoes = 2
for t in range(0, numero_iteracoes):
    print ("---Iteration: ", t)
    AVG_ACC = np.zeros(shape=[len(lista_alpha), len(lista_punicao)])
    STD_ACC = np.zeros(shape=[len(lista_alpha), len(lista_punicao)])

    x_count = 0
    for valor_alpha in lista_alpha:

        y_count = 0
        for punicao in lista_punicao:

            lista_temp_acuidade = []
            for indice_suCbtrain, cv_indice in skf_model.split(df_train_features, df_train_class):
                df_train_features_subset = df_train_features.loc[indice_suCbtrain]
                df_train_class_subset = df_train_class.loc[indice_suCbtrain]
                df_train_features_cv = df_train_features.loc[cv_indice]
                df_train_class_cv = df_train_class.loc[cv_indice]

                modelo_perceptrao = Perceptron(penalty=punicao, alpha=valor_alpha, c=1)
                modelo_perceptrao.fit(df_train_features_subset, df_train_class_subset)
                pontagem = modelo_perceptrao.score(df_train_features_cv, df_train_class_cv)
                lista_temp_acuidade.append(pontagem)

            AVG_ACC[x_count, y_count] = np.mean(lista_temp_acuidade)
            STD_ACC[x_count, y_count] = np.std(lista_temp_acuidade)
            y_count += 1

        x_count += 1

    if t==0:
        final_AVG_ACC = AVG_ACC
        final_STD_ACC = STD_ACC
    else:
        final_AVG_ACC = np.dstack([final_AVG_ACC, AVG_ACC])
        final_STD_ACC = np.dstack([final_STD_ACC, STD_ACC])

final_accuracy_mean_list = np.mean(final_AVG_ACC, axis=2)
max_ind = np.unravel_index(np.argmax(final_accuracy_mean_list, axis=None), final_accuracy_mean_list.shape)

Escolha_alpha = lista_alpha[max_ind[0]]
Escolha_Punicao = lista_punicao[max_ind[1]]

```

```

print ("Cross Validation - alpha pelo Perceptron: ",Escolha_alpha)
print ("Cross Validation - Punicao pelo Perceptron: ",Escolha_Punicao)

modelo_perceptrao_final = Perceptron(penalty=Escolha_Punicao, alpha=Escolha_alpha, c=1)
modelo_perceptrao_final = CalibratedClassifierCV(base_estimator=modelo_perceptrao_final)
modelo_perceptrao_final.fit(df_train_features, df_train_class)

Predicao_train = modelo_perceptrao_final.predict(df_train_features)
Predicao_test = modelo_perceptrao_final.predict(df_test_features)

Predicao_prob_train = modelo_perceptrao_final.predict_proba(df_train_features)
Predicao_prob_test = modelo_perceptrao_final.predict_proba(df_test_features)

avaliacao_PerformanceC(df_train_class, Predicao_train, Predicao_prob_train, df_test_class)

---Iteration: 0

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```

```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```



```

/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)

```


Cross Validation - alpha pelo Perceptron: 1e-05
Cross Validation - Punicao pelo Perceptron: elasticnet

```
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/utils/validation.py:761: DataConversionWarning:
  y = column_or_1d(y, warn=True)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
/home/andres/anaconda3/lib/python3.7/site-packages/sklearn/linear_model/stochastic_gradient.py:
  FutureWarning)
```

Training Confusion Matrix:

```
[[29673  279]
 [ 3505  451]]
```

Testing Confusion Matrix:

```
[[9870  100]
 [1182  151]]
```

Training Accuracy Score: 0.8884039164798867

Testing Accuracy Score: 0.8865787843935239

Training Precision: 0.6178082191780822

Testing Precision: 0.601593625498008

Training Recall: 0.11400404448938321

Testing Recall: 0.11327831957989497

Train Classification Report:

	precision	recall	f1-score	support
0	0.89	0.99	0.94	29952
1	0.62	0.11	0.19	3956
micro avg	0.89	0.89	0.89	33908
macro avg	0.76	0.55	0.57	33908
weighted avg	0.86	0.89	0.85	33908

Test Classification Report:

	precision	recall	f1-score	support
0	0.89	0.99	0.94	9970
1	0.60	0.11	0.19	1333
micro avg	0.89	0.89	0.89	11303
macro avg	0.75	0.55	0.56	11303
weighted avg	0.86	0.89	0.85	11303

Training F1score: 0.19248826291079812

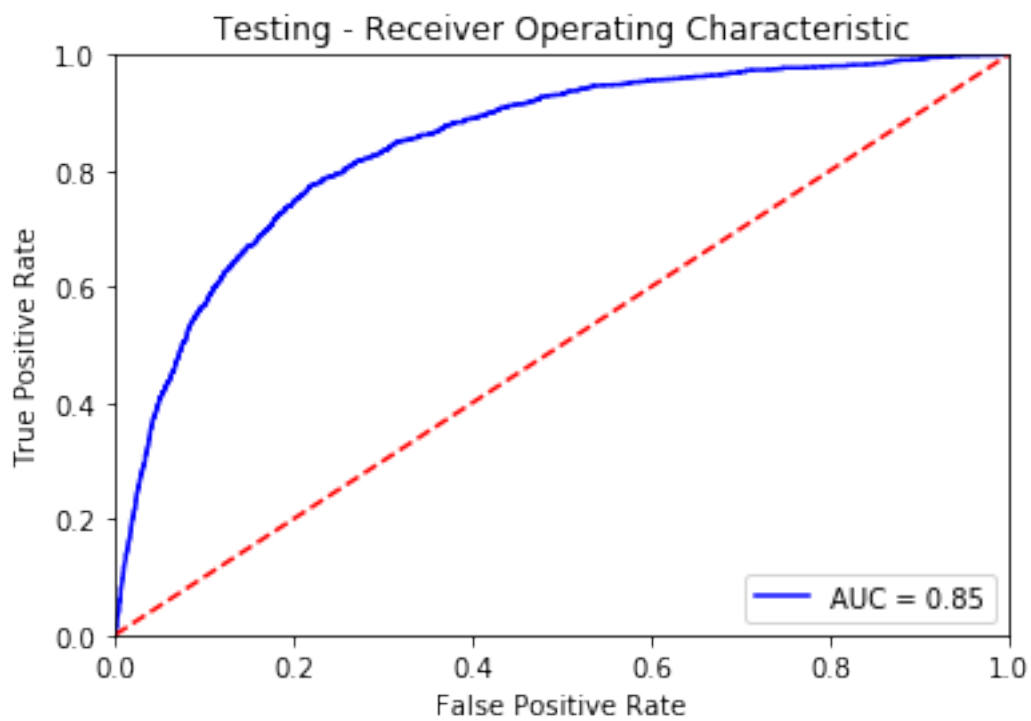
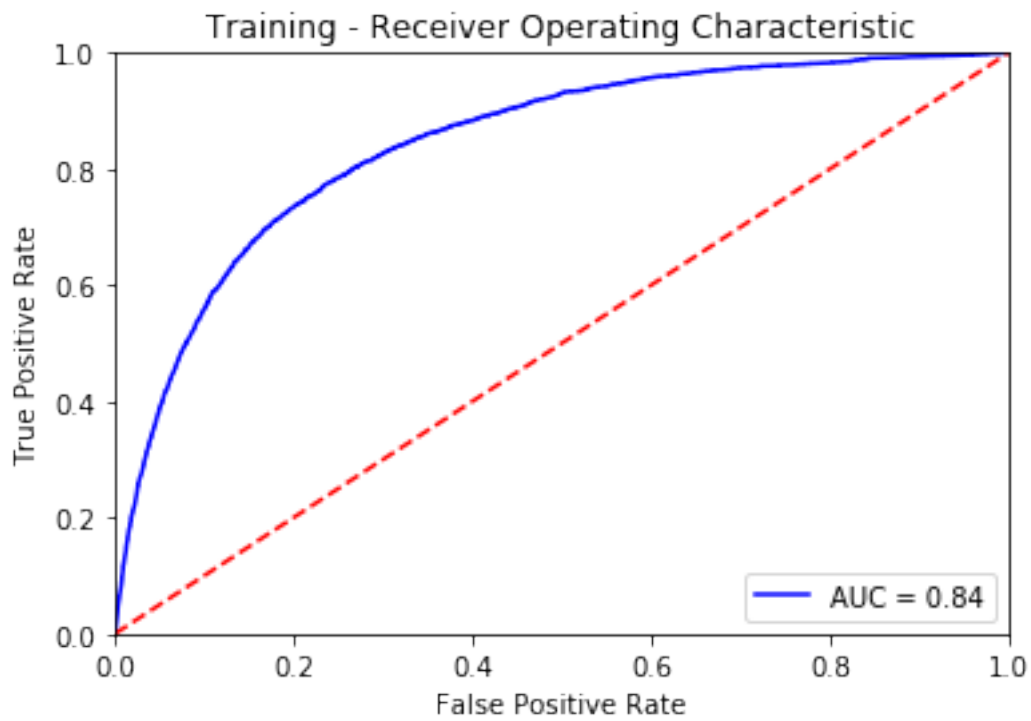
Testing F1score: 0.19065656565656566

Training Weighted F1score: 0.8528419982411115

Testing Weighted F1score: 0.8507597452595979

Training AUC for ROC: 0.8435902102953535

Testing AUC for ROC: 0.8465573013112856



2.8 Floresta Aleatória

```
In [135]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import StratifiedKFold

df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro...')
df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro...')

del df_train['Unnamed: 0']
del df_test['Unnamed: 0']

df_train_class = pd.DataFrame(df_train['y'])
df_train_features = df_train.loc[:, df_train.columns != 'y']

df_test_class = pd.DataFrame(df_test['y'])
df_test_features = df_test.loc[:, df_test.columns != 'y']

# Classificador Aleatorio de floresta
n_estimador_lista = range(10, 50, 5)

skf_model = StratifiedKFold(n_splits=5, shuffle=True)

Numero_iteracoes = 1
for t in range(0, Numero_iteracoes):
    print ("---Iteration: ", t)
    AVG_ACC = np.zeros(shape=[len(n_estimador_lista)])
    STD_ACC = np.zeros(shape=[len(n_estimador_lista)])

    x_count = 0
    for k_val in n_estimador_lista:
        Lista_Acuracia_Temp = []

        for Indice_subCindice, cv_index in skf_model.split(df_train_features, df_train_class):
            df_train_features_subset = df_train_features.loc[Indice_subCindice]
            df_train_class_subset = df_train_class.loc[Indice_subCindice]
            df_train_features_cv = df_train_features.loc[cv_index]
            df_train_class_cv = df_train_class.loc[cv_index]

            BA_modelo = RandomForestClassifier(n_estimators=k_val, class_weight='balanced')
            BA_modelo.fit(df_train_features_subset, df_train_class_subset)
            contagem = BA_modelo.score(df_train_features_cv, df_train_class_cv)
            Lista_Acuracia_Temp.append(contagem)

        AVG_ACC[x_count] = np.mean(Lista_Acuracia_Temp)
```


[illegible]

Cross Validation - Numero de Estimadores pela Floresta A : 10

```
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:64: DataConversionWarning
```

Training Confusion Matrix:

$$\begin{bmatrix} [29941 & 11] \\ [275 & 3681] \end{bmatrix}$$

Testing Confusion Matrix:

```
[[9740 230]
 [ 946 387]]
```

Training Accuracy Score: 0.9915654122920845

Testing Accuracy Score: 0.8959568256215165

Training Precision: 0.997020585048754

Testing Precision: 0.6272285251215559

Training Recall: 0.9304853387259858

Testing Recall: 0.2903225806451613

Train Classification Report:

	precision	recall	f1-score	support
0	0.99	1.00	1.00	29952
1	1.00	0.93	0.96	3956
micro avg	0.99	0.99	0.99	33908
macro avg	0.99	0.97	0.98	33908
weighted avg	0.99	0.99	0.99	33908

Test Classification Report:

	precision	recall	f1-score	support
0	0.91	0.98	0.94	9970
1	0.63	0.29	0.40	1333
micro avg	0.90	0.90	0.90	11303
macro avg	0.77	0.63	0.67	11303
weighted avg	0.88	0.90	0.88	11303

Training F1score: 0.9626046025104602

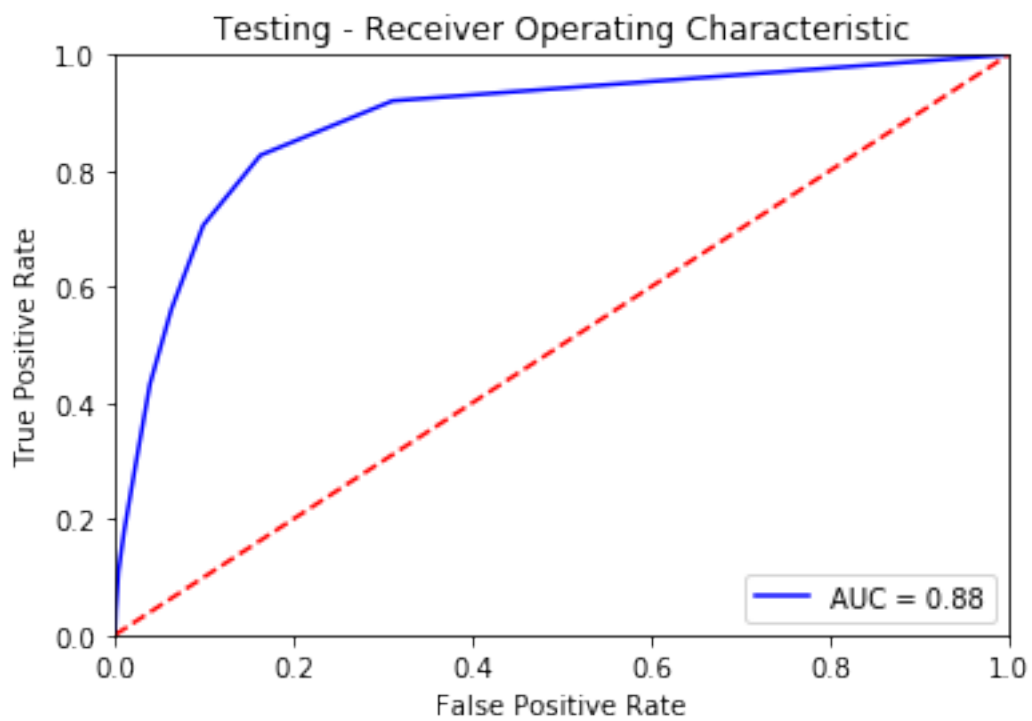
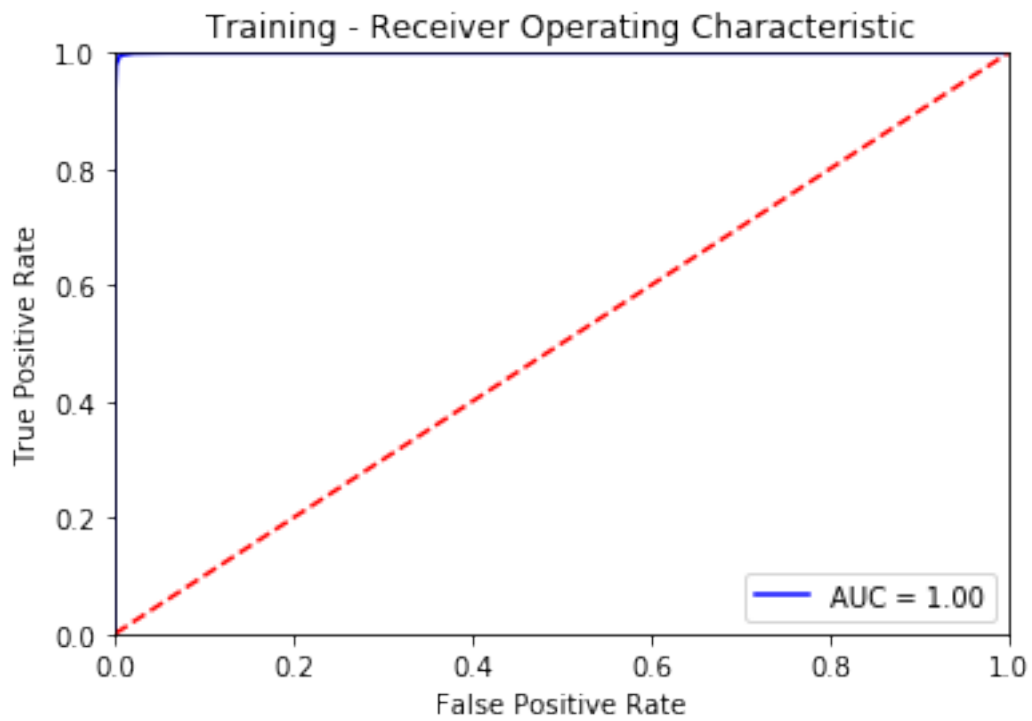
Testing F1score: 0.39692307692307693

Training Weighted F1score: 0.9914383405299793

Testing Weighted F1score: 0.878658792891652

Training AUC for ROC: 0.9997920754771503

Testing AUC for ROC: 0.8844188981046666



3 Nesta seção são apresentados alguns resultados analíticos

3.1 A primeira análise, está associado à tendencia a fazer um empréstimo entre as profissões dos participantes da campanha de adesão via telemarketing.

3.1.1 Os dados associados à cedito são:

1. housing: has housing loan? (binary: "yes","no")
2. loan: has personal loan? (binary: "yes","no")
3. default: has credit in default? (binary: "yes","no")

com os quais pretende-se responder à pergnta:

1. Qual profissão tem mais tendência a fazer um empréstimo? De qualtipo?

Primeiramente revisa-se a tendencia dos dados

In [136]: *## declaração de classe a usar para os histogramas associados a empréstimos*

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
plt.close("all")
col_df = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/bank-full.csv', sep=';')
loan=col_df.groupby("loan")
housing=col_df.groupby("housing")
default=col_df.groupby("default")

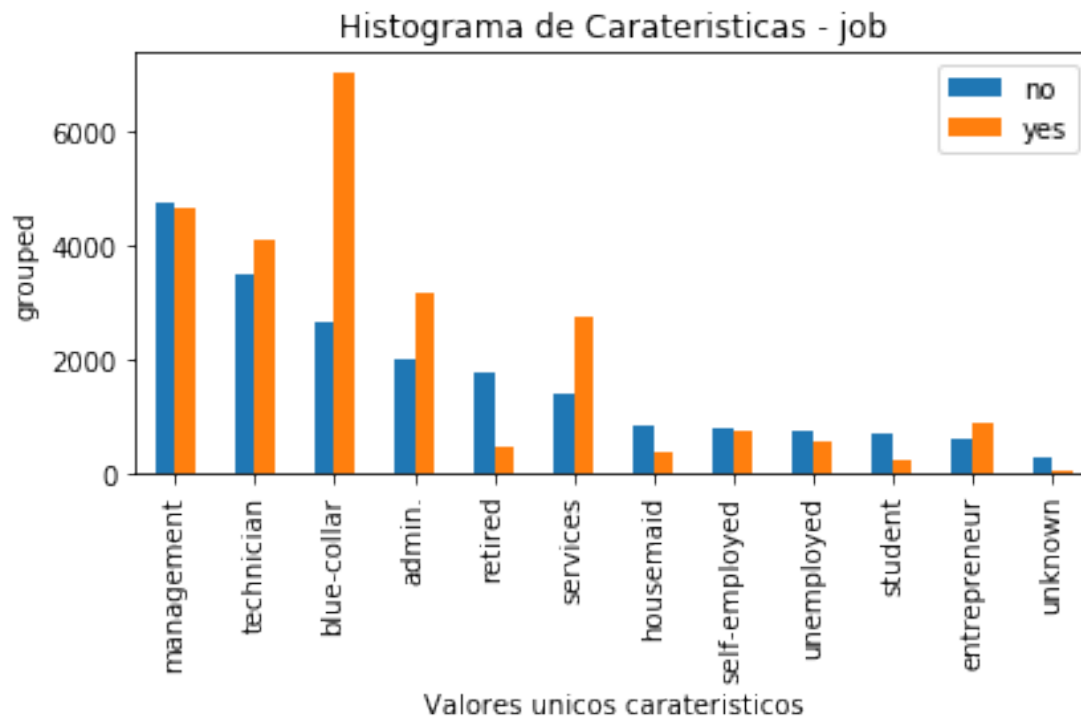
class Dados_iniciais1:

    def __init__(self, col_df, var):
        self.col_df = col_df
        self.col_df_grouped = var
        self.nome_classe_nao = "no"
        self.nome_classe_sim = "yes"
        self.col_df_grouped_nao = self.col_df_grouped.get_group(self.nome_classe_nao)
        self.col_df_grouped_sim = self.col_df_grouped.get_group(self.nome_classe_sim)

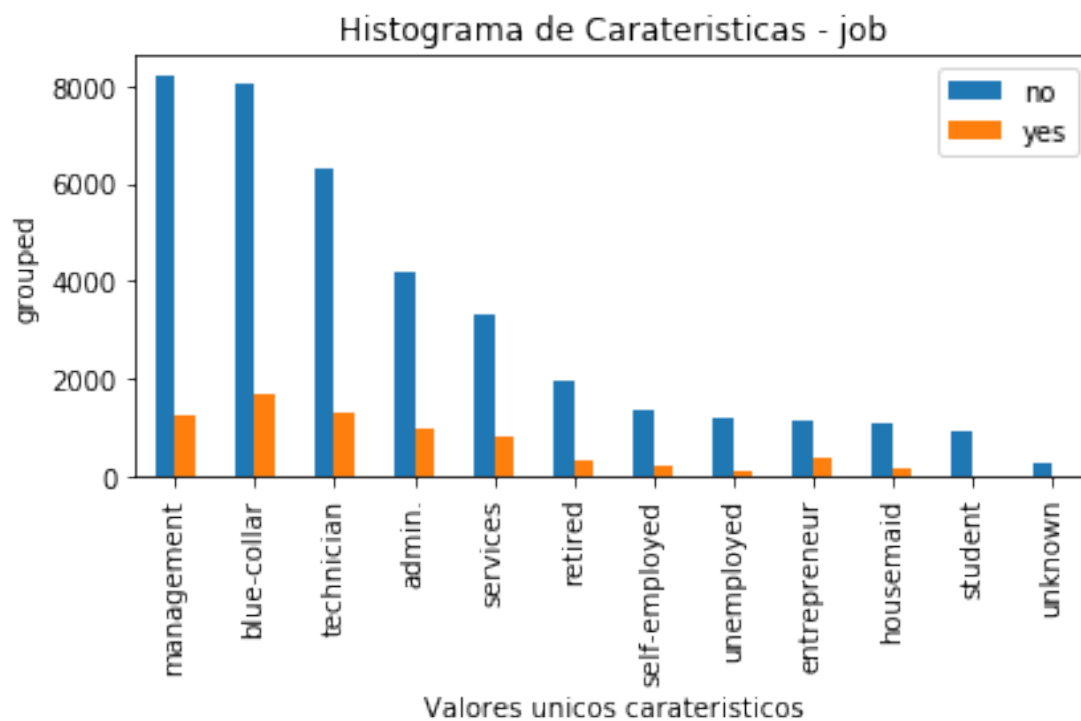
    def plot_histogram_categorical(self, nome_carateristica):
        carateristica_df = pd.DataFrame()
        carateristica_df["no"] = self.col_df_grouped_nao[nome_carateristica].value_counts()
        carateristica_df["yes"] = self.col_df_grouped_sim[nome_carateristica].value_counts()

        carateristica_df.plot(kind='bar')
        plt.title("Histograma de Carateristicas - "+nome_carateristica)
        plt.ylabel("grouped")
        plt.xlabel("Valores unicos carateristicos")
        plt.tight_layout()
```

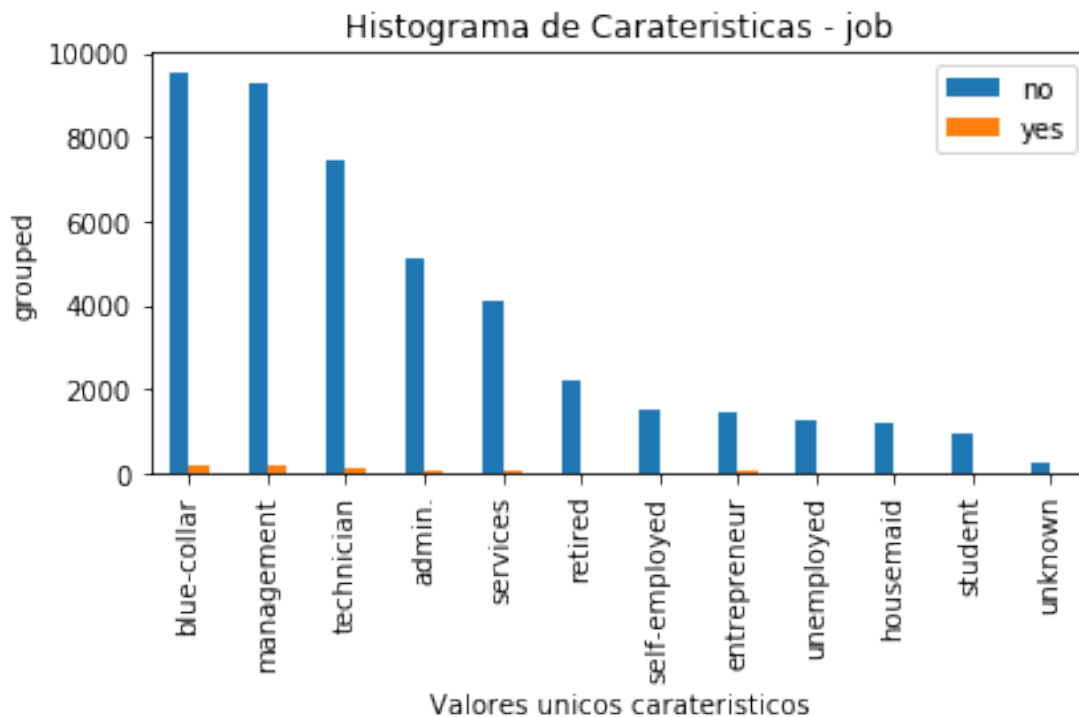
```
In [137]: Dados_iniciais_obj = Dados_iniciais1(col_df,housing)
          Dados_iniciais_obj.plot_histogram_categorical("job")
```



```
In [138]: Dados_iniciais_obj = Dados_iniciais1(col_df,loan)
          Dados_iniciais_obj.plot_histogram_categorical("job")
```



```
In [139]: Dados_iniciais_obj = Dados_iniciais1(col_df,default)
          Dados_iniciais_obj.plot_histogram_categorical("job")
```



3.1.2 A partir dos histogramas acima podemos formular que o blue-collar, management, e technician tem o maior número de empréstimos segundo os dados, e que por sua vez este grupo pode ser considerado como o grupo de maior tendência a realizar empréstimos de vivienda e pessoais.

4 2. Fazendo uma relação entre o número de contactos e sucesso da campanha quais são os pontos mais relevantes a serem observados?

```
In [140]: from sklearn.ensemble import RandomForestClassifier
          from sklearn.model_selection import StratifiedKFold

          df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_
          df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_

          del df_train['Unnamed: 0']
          del df_test['Unnamed: 0']
```



```

df_train_class = pd.DataFrame(df_train['y'])
df_train_features = df_train.loc[:, df_train.columns != 'y']

# Classificador Aleatorio de floresta
n_estimador_lista = range(10, 50, 5)

skf_model = StratifiedKFold(n_splits=5,shuffle=True)

Numero_iteracoes = 1
for t in range(0,Numero_iteracoes):
    print ("---Iteration: ",t)
    AVG_ACC = np.zeros(shape=[len(n_estimador_lista)])
    STD_ACC = np.zeros(shape=[len(n_estimador_lista)])

    x_count = 0
    for k_val in n_estimador_lista:
        Lista_Acuracia_Temp = []

        for Indice_subCindice, cv_index in skf_model.split(df_train_features,df_train_class):
            df_train_features_subset = df_train_features.loc[Indice_subCindice]
            df_train_class_subset = df_train_class.loc[Indice_subCindice]
            df_train_features_cv = df_train_features.loc[cv_index]
            df_train_class_cv = df_train_class.loc[cv_index]

            BA_modelo = RandomForestClassifier(n_estimators=k_val, class_weight='balanced')
            BA_modelo.fit(df_train_features_subset, df_train_class_subset)
            contagem = BA_modelo.score(df_train_features_cv, df_train_class_cv)
            Lista_Acuracia_Temp.append(contagem)

        AVG_ACC[x_count] = np.mean(Lista_Acuracia_Temp)
        STD_ACC[x_count] = np.std(Lista_Acuracia_Temp)
        x_count += 1

    if t==0:
        final_AVG_ACC = AVG_ACC
        final_STD_ACC = STD_ACC
    else:
        final_AVG_ACC = np.vstack([final_AVG_ACC, AVG_ACC])
        final_STD_ACC = np.vstack([final_STD_ACC, STD_ACC])

Lista_Acc_meia_final = np.mean(final_AVG_ACC, axis=0)
final_k_indice = np.argmax(Lista_Acc_meia_final)

Escolha_k= n_estimador_lista[final_k_indice]
print ("Cross Validation - Numero de Estimadores pela Floresta A : ",Escolha_k)

BA_modelo_final = RandomForestClassifier(n_estimators=Escolha_k, class_weight='balanced')

```



```
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarn
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarn
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarn
```

Cross Validation - Numero de Estimadores pela Floresta A : 10

```
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:58: DataConversionWarn
```

```
In [141]: print("ranking das Caracteristicas:")

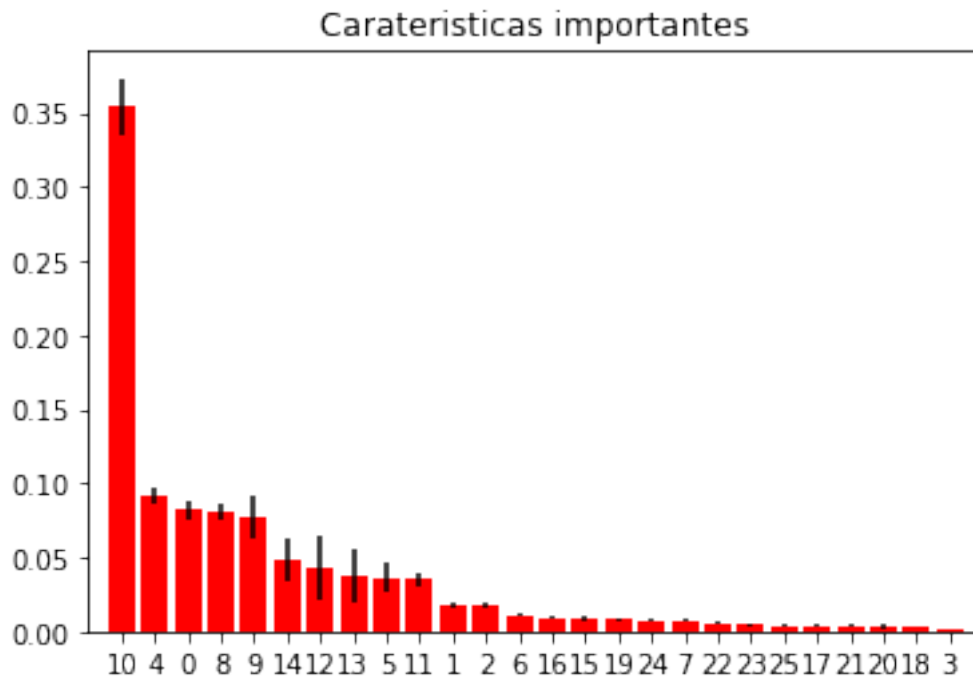
          for f in range(df_train_features_subset.shape[1]):
              print("%d. feature %d (%f)" % (f + 1, indices[f], importance[indices[f]]))

          # Plot the feature importances of the forest
          plt.figure()
          plt.title("Carateristicas importantes")
          plt.bar(range(df_train_features_subset.shape[1]), importance[indices],
                  color="r", yerr=std1[indices], align="center")
          plt.xticks(range(df_train_features_subset.shape[1]), indices)
          plt.xlim([-1, df_train_features_subset.shape[1]])
          plt.show()
```

ranking das Caracteristicas:

```
1. feature 10 (0.353740)
2. feature 4 (0.090681)
3. feature 0 (0.081947)
4. feature 8 (0.081156)
5. feature 9 (0.077302)
6. feature 14 (0.048171)
7. feature 12 (0.042330)
8. feature 13 (0.036917)
9. feature 5 (0.036398)
10. feature 11 (0.035134)
11. feature 1 (0.017955)
12. feature 2 (0.017305)
13. feature 6 (0.011233)
14. feature 16 (0.009257)
15. feature 15 (0.008145)
16. feature 19 (0.007789)
17. feature 24 (0.007619)
18. feature 7 (0.007405)
19. feature 22 (0.005513)
20. feature 23 (0.004223)
21. feature 25 (0.004129)
22. feature 17 (0.003930)
23. feature 21 (0.003690)
```

24. feature 20 (0.003599)
 25. feature 18 (0.003059)
 26. feature 3 (0.001372)



5 A través da análise ao modelo pode se corroborar a hipotese apresentada em [1] onde se se determina como um dos mais importantes elementos da toma de decisão dos clientes está associada ao tempo da ligação; em [1] determinam que um tempo superior a 8 seg, garante o maior numero de adesões à campanha.

5.1 Por outra parte pode se ver que outros elementos importantes para a adesão à campanha estão associados ao balance idade o dia e o mês de contacto. Por outra parte os clientes que se aderem à campanha estão ubicados maioritariamente no setor de serviço, desempregados, auto-empregados ou blue-collar.

[1] S. Moro, P. Cortez and P. Rita. A Data-Driven Approach to Predict the Success of Bank Tele-marketing. Decision Support Systems, Elsevier, 62:22-31, June 2014

6 3. Baseando-se nos resultados de adesão desta campanha qual é o número medio e máximo que você indica para otimizar a adesão?

```
In [142]: train_col_df = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_pro
```

```
col_df = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/bank-full.csv', sep
```

```
class Dados_iniciais2:
```

```
    def __init__(self, col_df, var, val_n, val_p):
        self.col_df = col_df
        self.col_df_grouped = var
        self.nome_classe_nao = val_n
        self.nome_classe_sim = val_p
        self.col_df_grouped_nao = self.col_df_grouped.get_group(self.nome_classe_nao)
        self.col_df_grouped_sim = self.col_df_grouped.get_group(self.nome_classe_sim)
```

```
    def plot_histograma_continuo(self, nome_carateristica, bin_tamanho):
        plt.figure()
        plt.hist(self.col_df_grouped_nao[nome_carateristica], bins=bin_tamanho, label=self.nome_classe_nao)
        plt.hist(self.col_df_grouped_sim[nome_carateristica], bins=bin_tamanho, label=self.nome_classe_sim)
        plt.legend()
        plt.title("Histograma de Carateristicas - "+nome_carateristica)
        plt.xlabel("Valores Carateristicos")
        plt.ylabel("grouped")
```

```
y=col_df.groupby("y")
val_n=col_df.nome_classe_nao = "no"
val_p=col_df.nome_classe_sim = "yes"
```

```
Dados_iniciais_obj = Dados_iniciais2(col_df,y,val_n,val_p)
```

```
Dados_iniciais_obj.plot_histograma_continuo("campaign", 50)
plt.axis([ 0, 10, 0, 25000])
plt.title("Histograma de Carateristicas Campaign")
plt.ylabel("Adesão à Campanha Atual")
plt.xlabel("Numero de Contactos ao Cliente")
```

```
Dados_iniciais_obj.plot_histograma_continuo("campaign", 50)
plt.axis([ 0, 2.5, 0, 4500])
plt.title("Histograma de Carateristicas Campaign")
plt.ylabel("Adesão à Campanha Atual")
plt.xlabel("Numero de Contactos ao Cliente")
```

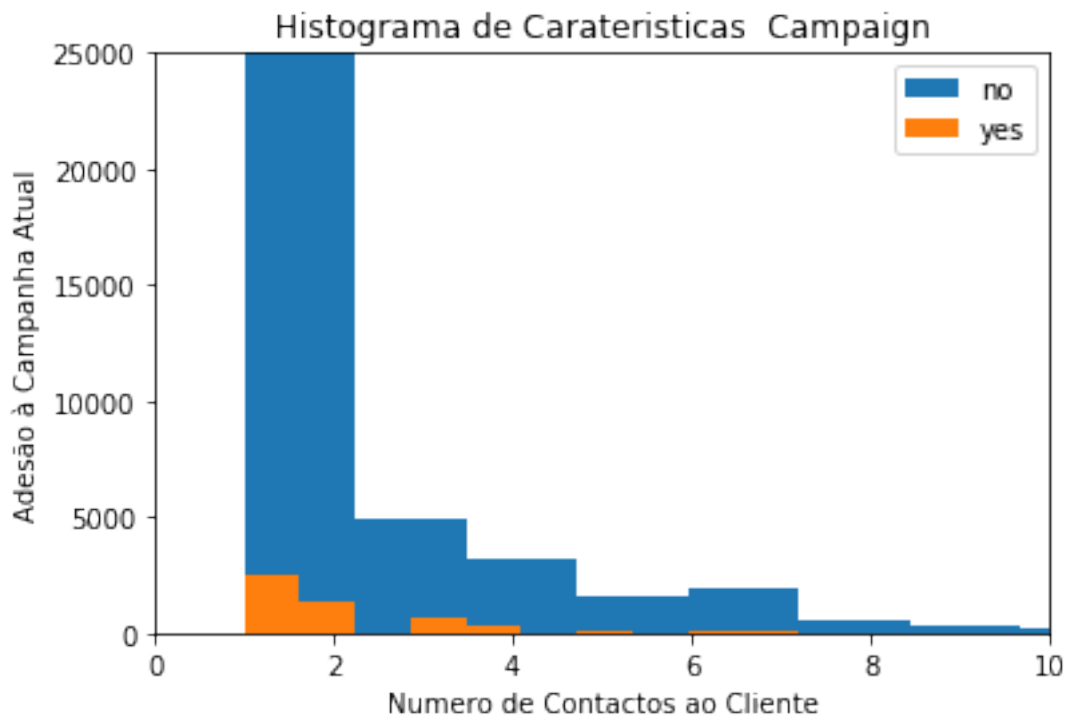
```

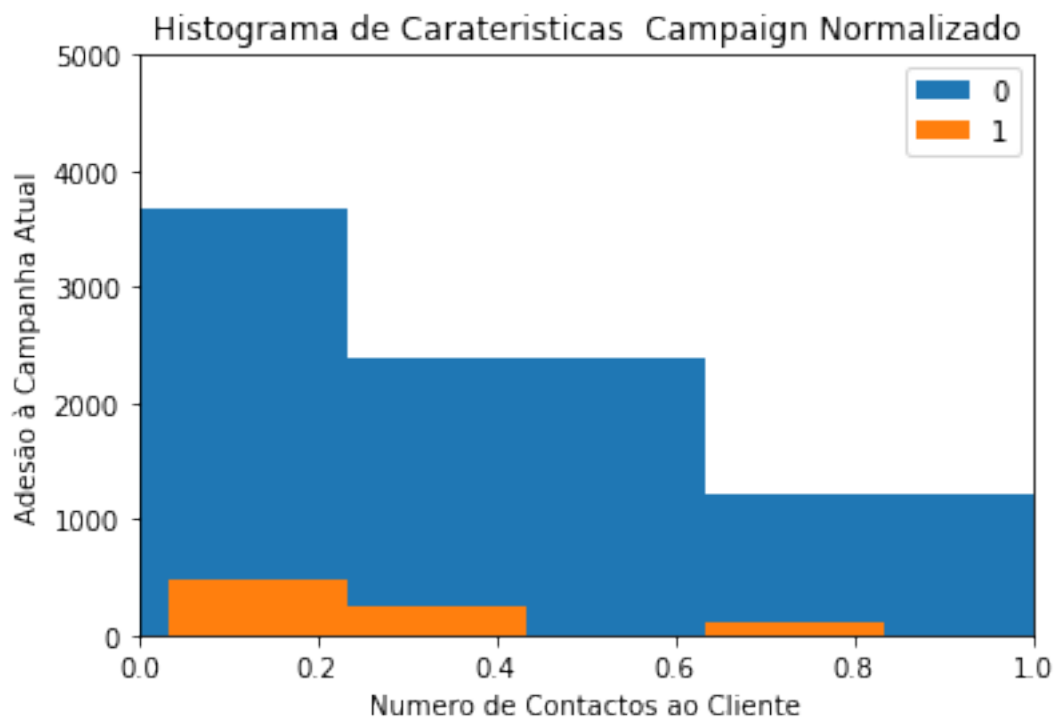
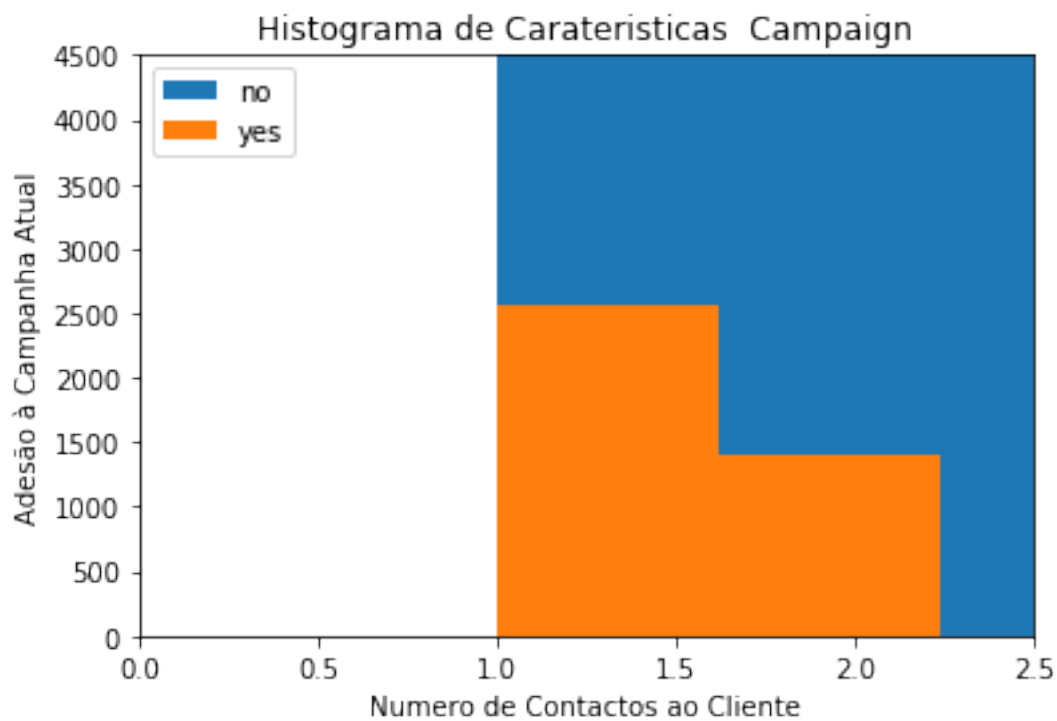
y=train_col_df.groupby("y")
val_n=train_col_df.nome_classe_nao = 0
val_p=train_col_df.nome_classe_sim = 1
Dados_iniciais_obj = Dados_iniciais2(train_col_df,y,val_n,val_p)

Dados_iniciais_obj.plot_historama_continuo("campaign", 50)
plt.axis([ 0, 1., 0, 5000])
plt.title("Histograma de Carateristicas Campaign Normalizado")
plt.ylabel("Adesão à Campanha Atual")
plt.xlabel("Numero de Contactos ao Cliente")

```

Out[142]: Text(0.5, 0, 'Numero de Contactos ao Cliente')





6.1 deacordo com os dados e a frequencia apresentada na tabela abaixo, pode-se concluir que o maior numero de adesões ocorre um numero de ligações inferior a 6.

6.2 Portanto, recomenda-se em media ligar 4 vezes e no máximo 6.

```
In [143]: Camp=col_df.groupby(["campaign","y"])
          Camp_count=Camp['y'].count().unstack()
          Camp_count.head(15)
```

```
Out[143]: y          no      yes
campaign
1          14983.0  2561.0
2          11104.0  1401.0
3           4903.0   618.0
4           3205.0   317.0
5           1625.0   139.0
6           1199.0    92.0
7            688.0    47.0
8            508.0    32.0
9            306.0    21.0
10           252.0    14.0
11           185.0    16.0
12           151.0     4.0
13           127.0     6.0
14            89.0     4.0
15            80.0     4.0
```

7 4. O resultado da campanha anterior tem relevância na campanha atual?

7.1 Para isto usaremos os resultados apresentados em na resposta à pergunta 1.

7.2 Usando os valores 9 a 16 deacordo com o bank-names.csv fornecido com o problema, como valores de entrada no modelo. E y (17) como saída do mesmo.

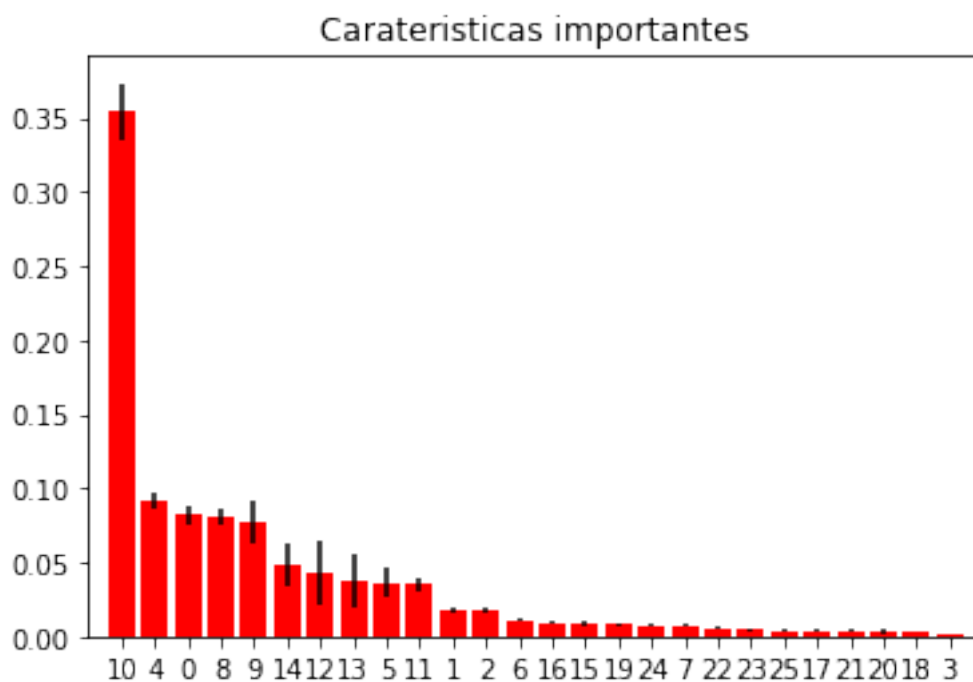
```
In [144]: print("ranking das Caracteristicas:")

          for f in range(df_train_features_subset.shape[1]):
              print("%d. feature %d (%f)" % (f + 1, indices[f], importance[indices[f]]))

          # Plot the feature importances of the forest
          plt.figure()
          plt.title("Carateristicas importantes")
          plt.bar(range(df_train_features_subset.shape[1]), importance[indices],
                  color="r", yerr=std1[indices], align="center")
          plt.xticks(range(df_train_features_subset.shape[1]), indices)
          plt.xlim([-1, df_train_features_subset.shape[1]])
          plt.show()
```


ranking das Caracteristicas:

1. feature 10 (0.353740)
2. feature 4 (0.090681)
3. feature 0 (0.081947)
4. feature 8 (0.081156)
5. feature 9 (0.077302)
6. feature 14 (0.048171)
7. feature 12 (0.042330)
8. feature 13 (0.036917)
9. feature 5 (0.036398)
10. feature 11 (0.035134)
11. feature 1 (0.017955)
12. feature 2 (0.017305)
13. feature 6 (0.011233)
14. feature 16 (0.009257)
15. feature 15 (0.008145)
16. feature 19 (0.007789)
17. feature 24 (0.007619)
18. feature 7 (0.007405)
19. feature 22 (0.005513)
20. feature 23 (0.004223)
21. feature 25 (0.004129)
22. feature 17 (0.003930)
23. feature 21 (0.003690)
24. feature 20 (0.003599)
25. feature 18 (0.003059)
26. feature 3 (0.001372)



7.3 Levando em conta o anteriora tabela acima pode ser reescrita como:

1. duration
2. month
3. pdays
4. poutcome
5. previous
6. campaign

7.4 sendo confirmado que os dados mais relevantes associados à campanha anterior são a duração o mês de contacto e a ligação poutcome. Por outra parte campaign contes a informação do ultimo contacto portanto a informação está correlata com previous.

8 6. Quais são as características mais prominentes de um cliente para que possua um emprestimo imobiliario

8.1 Para este fim creamos um modelo de regressão entre a variável housing que inclui o credito imobiliario e as variáveis características incluídas as de jobs, de acordo com a lassificação do item 3 e dos dados standarizados e salvos como bank_prepro_standardize_train.csv

```
In [145]: from sklearn.ensemble import RandomForestClassifier
          from sklearn.model_selection import StratifiedKFold

df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_
df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro_

del df_train['Unnamed: 0']
del df_test['Unnamed: 0']

df_train_class = pd.DataFrame(df_train['housing'])
df_train_features = df_train.loc[:, df_train.columns != 'housing']

# Classificador Aleatorio de floresta
n_estimador_lista = range(10, 50, 5)

skf_model = StratifiedKFold(n_splits=5,shuffle=True)

Numero_iteracoes = 1
for t in range(0,Numero_iteracoes):
    print ("---Iteration: ",t)
    AVG_ACC = np.zeros(shape=[len(n_estimador_lista)])
    STD_ACC = np.zeros(shape=[len(n_estimador_lista)])
```

```

x_count = 0
for k_val in n_estimador_lista:
    Lista_Acuracia_Temp = []

    for Indice_subCindice, cv_index in skf_model.split(df_train_features, df_train_class):
        df_train_features_subset = df_train_features.loc[Indice_subCindice]
        df_train_class_subset = df_train_class.loc[Indice_subCindice]
        df_train_features_cv = df_train_features.loc[cv_index]
        df_train_class_cv = df_train_class.loc[cv_index]

        BA_modelo = RandomForestClassifier(n_estimators=k_val, class_weight='balanced')
        BA_modelo.fit(df_train_features_subset, df_train_class_subset)
        contagem = BA_modelo.score(df_train_features_cv, df_train_class_cv)
        Lista_Acuracia_Temp.append(contagem)

    AVG_ACC[x_count] = np.mean(Lista_Acuracia_Temp)
    STD_ACC[x_count] = np.std(Lista_Acuracia_Temp)
    x_count += 1

    if t==0:
        final_AVG_ACC = AVG_ACC
        final_STD_ACC = STD_ACC
    else:
        final_AVG_ACC = np.vstack([final_AVG_ACC, AVG_ACC])
        final_STD_ACC = np.vstack([final_STD_ACC, STD_ACC])

Lista_Acc_meia_final = np.mean(final_AVG_ACC, axis=0)
final_k_indice = np.argmax(Lista_Acc_meia_final)

Escolha_k= n_estimador_lista[final_k_indice]
print ("Cross Validation - Numero de Estimadores pela Floresta A : ", Escolha_k)

BA_modelo_final = RandomForestClassifier(n_estimators=Escolha_k, class_weight='balanced')
BA_modelo_final.fit(df_train_features, df_train_class)

importance=BA_modelo_final.feature_importances_

std1 = np.std([tree.feature_importances_ for tree in BA_modelo_final.estimators_],
               axis=0)
indices = np.argsort(importance)[::-1]

```

---Iteration: 0

```

/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarning:
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarning:
/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarning:

```



```

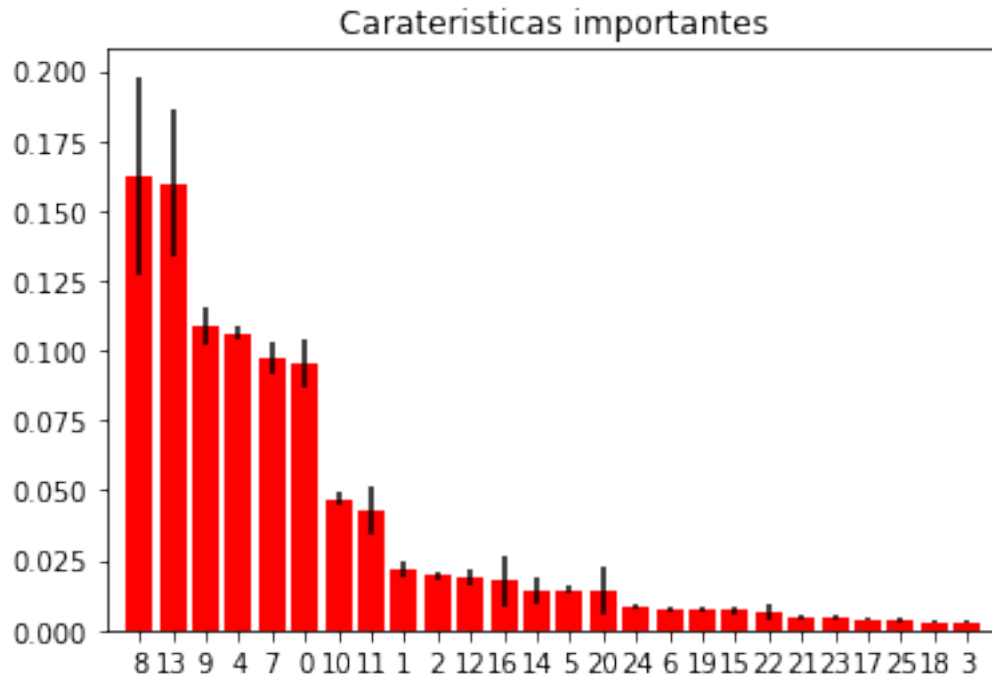
print("%d. feature %d (%f)" % (f + 1, indices[f], importance[indices[f]]))

# Plot the feature importances of the forest
plt.figure()
plt.title("Carateristicas importantes")
plt.bar(range(df_train_features_subset.shape[1]), importance[indices],
        color="r", yerr=std1[indices], align="center")
plt.xticks(range(df_train_features_subset.shape[1]), indices)
plt.xlim([-1, df_train_features_subset.shape[1]])
plt.show()

```

ranking das Caracteristicas:

1. feature 8 (0.162519)
2. feature 13 (0.159573)
3. feature 9 (0.108606)
4. feature 4 (0.106350)
5. feature 7 (0.097721)
6. feature 0 (0.095319)
7. feature 10 (0.047008)
8. feature 11 (0.042828)
9. feature 1 (0.021796)
10. feature 2 (0.019566)
11. feature 12 (0.019360)
12. feature 16 (0.017659)
13. feature 14 (0.014380)
14. feature 5 (0.014318)
15. feature 20 (0.014080)
16. feature 24 (0.008648)
17. feature 6 (0.007427)
18. feature 19 (0.007372)
19. feature 15 (0.007062)
20. feature 22 (0.006427)
21. feature 21 (0.004410)
22. feature 23 (0.004246)
23. feature 17 (0.003681)
24. feature 25 (0.003627)
25. feature 18 (0.003018)
26. feature 3 (0.002999)



9 deacordo ao modelo e os dados analisados, pode se concluir que as 5 variáveis mais importantes associadas ao credito de vivienda são:

1. Month
2. Outcome
3. duration
4. balance
5. age

10 5. Qual o fator determinante para que o banco exija um seguro de credito

```
In [147]: from sklearn.ensemble import RandomForestClassifier
          from sklearn.model_selection import StratifiedKFold

df_train = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro...
df_test = pd.read_csv('/media/andres/dados1/semantix_teste/desafio/Data/bank_prepro...

del df_train['Unnamed: 0']
del df_test['Unnamed: 0']

df_train_class = pd.DataFrame(df_train['default'])
```

```

df_train_features = df_train.loc[:, df_train.columns != 'default']

# Classificador Aleatorio de floresta
n_estimador_lista = range(10, 50, 5)

skf_model = StratifiedKFold(n_splits=5,shuffle=True)

Numero_iteracoes = 1
for t in range(0,Numero_iteracoes):
    print ("---Iteration: ",t)
    AVG_ACC = np.zeros(shape=[len(n_estimador_lista)])
    STD_ACC = np.zeros(shape=[len(n_estimador_lista)])

    x_count = 0
    for k_val in n_estimador_lista:
        Lista_Acuracia_Temp = []

        for Indice_subCindice, cv_index in skf_model.split(df_train_features,df_train_class):
            df_train_features_subset = df_train_features.loc[Indice_subCindice]
            df_train_class_subset = df_train_class.loc[Indice_subCindice]
            df_train_features_cv = df_train_features.loc[cv_index]
            df_train_class_cv = df_train_class.loc[cv_index]

            BA_modelo = RandomForestClassifier(n_estimators=k_val, class_weight='balanced')
            BA_modelo.fit(df_train_features_subset, df_train_class_subset)
            contagem = BA_modelo.score(df_train_features_cv, df_train_class_cv)
            Lista_Acuracia_Temp.append(contagem)

        AVG_ACC[x_count] = np.mean(Lista_Acuracia_Temp)
        STD_ACC[x_count] = np.std(Lista_Acuracia_Temp)
        x_count += 1

    if t==0:
        final_AVG_ACC = AVG_ACC
        final_STD_ACC = STD_ACC
    else:
        final_AVG_ACC = np.vstack([final_AVG_ACC, AVG_ACC])
        final_STD_ACC = np.vstack([final_STD_ACC, STD_ACC])

Lista_Acc_meia_final = np.mean(final_AVG_ACC, axis=0)
final_k_indice = np.argmax(Lista_Acc_meia_final)

Escolha_k= n_estimador_lista[final_k_indice]
print ("Cross Validation - Numero de Estimadores pela Floresta A : ",Escolha_k)

BA_modelo_final = RandomForestClassifier(n_estimators=Escolha_k, class_weight='balanced')
BA_modelo_final.fit(df_train_features, df_train_class)

```

```
importance=BA_modelo_final.feature_importances_  
  
std1 = np.std([tree.feature_importances_ for tree in BA_modelo_final.estimators_],  
              axis=0)  
indices = np.argsort(importance)[::-1]
```

[illegible]

/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:36: DataConversionWarn

Cross Validation - Numero de Estimadores pela Floresta A : 10

/home/andres/anaconda3/lib/python3.7/site-packages/ipykernel_launcher.py:58: DataConversionWarn

```
In [148]: print("ranking das Caracteristicas:")

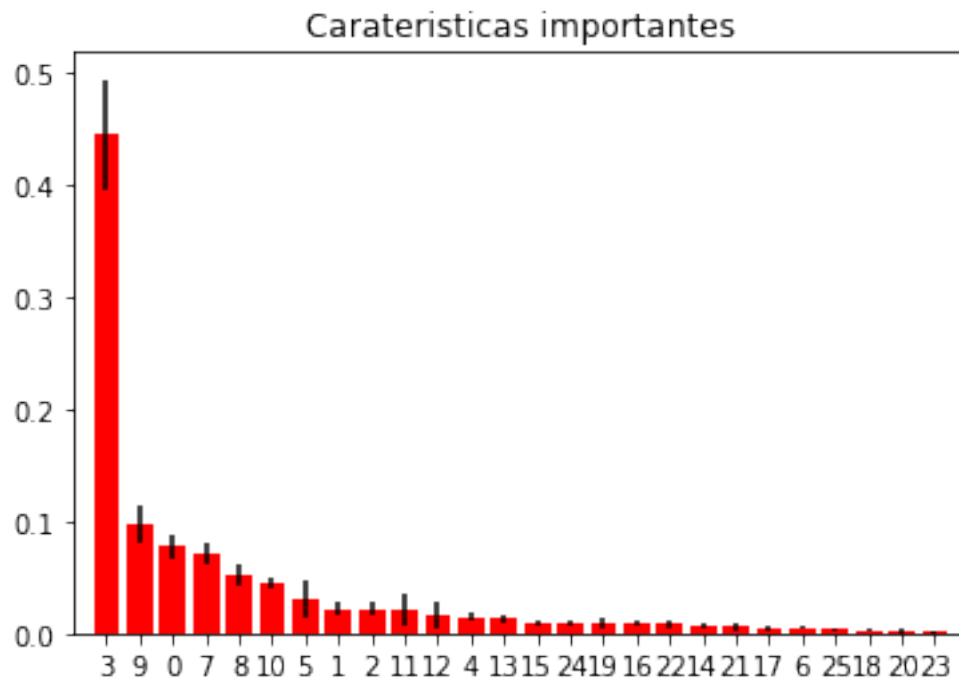
          for f in range(df_train_features_subset.shape[1]):
              print("%d. feature %d (%f)" % (f + 1, indices[f], importance[indices[f]]))

          # Plot the feature importances of the forest
          plt.figure()
          plt.title("Carateristicas importantes")
          plt.bar(range(df_train_features_subset.shape[1]), importance[indices],
                  color="r", yerr=std1[indices], align="center")
          plt.xticks(range(df_train_features_subset.shape[1]), indices)
          plt.xlim([-1, df_train_features_subset.shape[1]])
          plt.show()
```

ranking das Caracteristicas:

1. feature 3 (0.443842)
2. feature 9 (0.097340)
3. feature 0 (0.076962)
4. feature 7 (0.070953)
5. feature 8 (0.051983)
6. feature 10 (0.044437)
7. feature 5 (0.031281)
8. feature 1 (0.021719)
9. feature 2 (0.021610)
10. feature 11 (0.020790)
11. feature 12 (0.016020)
12. feature 4 (0.014135)
13. feature 13 (0.012688)
14. feature 15 (0.009145)
15. feature 24 (0.008780)
16. feature 19 (0.008718)
17. feature 16 (0.008671)
18. feature 22 (0.008200)
19. feature 14 (0.007275)
20. feature 21 (0.005696)
21. feature 17 (0.005054)
22. feature 6 (0.004727)
23. feature 25 (0.003779)
24. feature 18 (0.003027)
25. feature 20 (0.002227)

26. feature 23 (0.000942)



11 Analizando também o modelo de balance vs as outras carateristica também podemos inferir algumas informações relevantes, dado que o cliente com saldo negativo anual pode também representar um risco e precisar de um seguro de emprestimo. No ponto 6. pode ser visto que o balance é um ponto relevante para adquirir um emprestimo. Dado isto e a informação acima, pode se concluir que as 5 variáveis mais importantes associadas ao seguro para credito são:

1. balance (3)
2. age (0)
3. loan (5)
4. marital (1)
5. education (2)

In []: