

```
import pandas as pd
import pandas_profiling as pp
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
matplotlib.interactive = True
```

Import and Encode Data

```
In [2]: df = pd.read_csv("bank_data.csv")
df.head()
```

```
Out[2]:
```

	age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	outcome
0	20	student	single	secondary	no	502	no	no	cellular	30	apr	261	1	-1	0	u
1	68	retired	divorced	secondary	no	4189	no	no	telephone	14	jul	897	2	-1	0	u
2	32	management	single	tertiary	no	2536	yes	no	cellular	26	aug	956	6	-1	0	u
3	49	technician	married	tertiary	no	1235	no	no	cellular	13	aug	354	3	-1	0	u
4	78	retired	divorced	primary	no	229	no	no	telephone	22	oct	97	1	-1	0	u

```
In [3]: X = df.iloc[:,1:16]
y = df.iloc[:,16:]
```

```
In [4]: X = pd.get_dummies(X)
```

```
In [5]: from sklearn.preprocessing import LabelEncoder
```

```
In [6]: le = LabelEncoder()
```

```
In [7]: y = LabelEncoder().fit_transform(np.ravel(y))
```

```
In [8]: from sklearn.model_selection import train_test_split
```

```
In [9]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=101)
```

Logistic Regression (Part A)

```
In [10]: import statmodels.api as sm
import statmodels.formula.api as smf
```

Full Model

```
In [11]: full_mod = sm.GLM(y_train, sm.add_constant(X_train), family=sm.families.Binomial()).fit()
```

```
In [12]: full_mod.summary()
```

```
Out[12]:
```

Generalized Linear Model Regression Results									
Dep. Variable:	y	No. Observations:	833						
Model:	GLM	DF Residuals:	790						
Model Family:	Binomial	DF Model:	42						
Link Function:	Logit	Scale:	1.0000						
Method:	IRLS	Log-Likelihood:	-331.47						
Date:	Wed, 07 Dec 2022	Deviance:	662.94						
Time:	20:18:20	Pearson chi2:	7.26e+04						
No. Iterations:	7	Pseudo R-sq. (CS):	0.4459						
Covariance Type:	nonrobust								
	coef	std err	z	P> z	[0.025	0.975]			
const	-0.1655	0.244	0.677	0.498	-0.313	0.644			
age	-0.0103	0.013	-0.783	0.433	-0.036	0.015			
balance	-2.084e-05	3.45e-05	-0.604	0.546	-8.84e-05	6.68e-05			
duration	0.0052	0.014	1.113	0.267	-0.012	0.042			
campaign	-0.0116	0.007	-2.308	0.021	-0.043	-0.020			
pdays	-0.0322	0.002	-1.102	0.270	-0.006	0.002			
previous	-0.1213	0.098	-1.234	0.217	-0.314	0.071			
job_admin.	0.4162	0.339	1.229	0.219	-0.247	1.080			
job_blue-collar	-0.6601	0.292	-2.260	0.024	-1.233	-0.088			
job_entrepreneur	0.1630	0.551	1.153	0.249	-0.444	1.714			
job_housemaid	-0.0656	0.524	-0.318	0.751	-1.194	0.861			
job_management	0.0982	0.290	0.338	0.735	-0.471	0.667			
job_retired	0.7708	0.428	1.800	0.072	-0.068	1.610			
job_self-employed	-0.2479	0.472	-0.525	0.600	-1.173	0.578			
job_services	-0.5707	0.362	-1.390	0.165	-1.279	0.118			
job_student	0.3052	0.575	0.531	0.595	-0.821	1.432			
job_technician	-0.1292	0.271	-0.476	0.634	-0.661	0.403			
job_unemployed	-1.9232	0.721	-2.668	0.008	-3.366	-0.510			
job_unknown	1.5979	0.391	1.716	0.086	-0.227	3.423			
marital_divorced	0.3131	0.245	1.277	0.201	-0.167	0.794			
marital_married	-0.0063	0.168	-0.406	0.684	-0.398	0.261			
marital_single	-0.0794	0.189	-0.420	0.675	-0.450	0.391			
education_primary	0.1468	0.291	0.505	0.614	-0.423	0.717			
education_secondary	0.0224	0.215	0.104	0.917	-0.400	0.445			
education_tertiary	0.2311	0.254	0.910	0.363	-0.266	0.729			
education_unknown	-0.2349	0.481	-0.488	0.625	-1.178	0.708			
default_no	0.3974	0.424	-0.938	0.348	-1.228	0.389			
default_yes	0.5629	0.526	1.071	0.284	-0.467	1.593			
housing_no	0.2367	0.166	1.428	0.153	-0.088	0.562			
housing_yes	-0.0712	0.169	-0.420	0.674	-0.403	0.261			
loan_no	0.5521	0.209	2.546	0.011	0.123	0.942			
loan_yes	-0.3666	0.218	-1.682	0.093	-0.794	0.061			
contact_cellular	0.6658	0.194	3.528	0.000	0.305	1.067			
contact_telephone	0.0332	0.330	0.101	0.920	-0.613	0.680			
contact_unknown	-0.5535	0.265	-2.087	0.037	-1.073	-0.034			
month_apr	-0.0899	0.360	-0.250	0.803	-0.796	0.616			
month_aug	-0.7384	0.312	-2.367	0.018	-1.350	-0.127			
month_dec	1.4435	1.300	1.110	0.267	-1.104	3.991			
month_feb	-0.6254	0.388	-1.611	0.107	-1.387	0.136			
month_jul	-1.8681	0.568	-3.289	0.001	-2.981	-0.755			
month_jun	-1.3114	0.327	-4.015	0.000	-1.952	-0.671			
month_may	-0.1320	0.363	-0.364	0.716	-0.843	0.579			
month_nov	2.7845	1.002	2.778	0.005	0.820	4.749			
month_mar	-0.8936	0.304	-2.936	0.003	-1.490	-0.297			
month_may	-1.4059	0.348	-4.047	0.000	-2.092	-0.727			
month_oct	2.2414	0.759	2.955	0.003	0.755	3.728			
month_sep	0.7646	0.671	1.140	0.254	-0.550	2.080			
poutcome_failure	-0.8082	0.353	-2.287	0.022	-1.501	-0.106			
poutcome_other	-0.2497	0.383	-0.653	0.514	-1.000	0.500			
poutcome_success	2.8718	0.658	4.368	0.000	1.583	4.161			
poutcome_unknown	-1.6484	0.451	-3.657	0.000	-2.532	-0.765			

Intermediate Models

```
1
In [13]: X_train_i = X_train.copy()
X_test_i = X_test.copy()
for column in X_train_i.columns:
    if full_mod.pvalues.loc[column] > 0.38:
        X_train_i.drop(column,axis=1, inplace=True)
        X_test_i.drop(column,axis=1, inplace=True)
```

```
In [14]: int_mod = sm.GLM(y_train, sm.add_constant(X_train_i), family=sm.families.Binomial()).fit()
```

```
In [15]: int_mod.summary()
```

```
Out[15]:
```

Generalized Linear Model Regression Results									
Dep. Variable:	y	No. Observations:	833						
Model:	GLM	DF Residuals:	800						
Model Family:	Binomial	DF Model:	32						
Link Function:	Logit	Scale:	1.0000						
Method:	IRLS	Log-Likelihood:	-332.92						
Date:	Wed, 07 Dec 2022	Deviance:	665.83						
Time:	20:18:20	Pearson chi2:	7.54e+04						
No. Iterations:	7	Pseudo R-sq. (CS):	0.4440						
Covariance Type:	nonrobust								
	coef	std err	z	P> z	[0.025	0.975]			
const	-0.3170	0.487	-0.651	0.515	-1.272	0.638			
day	0.0151	0.013	1.157	0.247	-0.010	0.041			
duration	0.0052	0.000	11.869	0.000	0.004	0.006			
campaign	-0.1327	0.056	-2.354	0.019	-0.243	-0.022			
pdays	-0.0021	0.002	-1.093	0.274	-0.006	0.002			
previous	-0.1214	0.099	-1.229	0.219	-0.315	0.072			
job_admin.	0.4899	0.356	1.374	0.169	-0.209	1.188			
job_blue-collar	-0.7039	0.322	-1.784	0.074	-1.204	0.056			
job_entrepreneur	0.7023	0.593	1.185	0.236	-0.460	1.864			
job_retired	0.6140	0.388	1.581	0.114	-0.147	1.375			
job_services	-0.4580	0.411	-1.114	0.265	-1.264	0.348			
job_unemployed	-1.8594	0.776	-2.395	0.017	-3.381	-0.338			
job_unknown	1.3621	0.963	1.415	0.157	-0.525	3.249			
marital_divorced	0.2842	0.304	0.934	0.351	-0.312	0.881			
education_tertiary	0.2861	0.240	1.190	0.234	-0.185	0.757			
default_no	-0.6617	0.419	-1.580	0.114	-1.483	0.159			
default_yes	0.3447	0.617	0.559	0.576	-0.864	1.553			
housing_no	0.3036	0.222	1.368	0.171	-0.131	0.739			
loan_no	0.2807	0.291	0.965	0.334	-0.289	0.851			
loan_yes	-0.5978	0.308	-1.942	0.052	-1.201	0.066			
contact_cellular	0.7217	0.442	1.631	0.103	-0.146	1.589			
contact_unknown	-0.5451	0.513	-1.064	0.288	-1.550	0.459			
month_aug	-0.6501	0.345	-1.885	0.059	-1.326	0.026			
month_dec	1.4502	1.407	1.031	0.303	-1.308	4.208			
month_feb	-0.5302	0.428	-1.238	0.216	-1.370	0.309			
month_jan	-1.7319	0.629	-2.752	0.006	-2.965	-0.498			
month_jul	-1.8672	0.369	-3.160	0.002	-1.891	-0.443			
month_mar	-2.1659	1.092	-2.643	0.008	-0.746	-3.526			
month_may	-0.7229	0.316	-2.287	0.022	-1.343	-0.103			
month_nov	-1.2980	0.394	-3.298	0.001	-2.069	-0.527			
month_oct	2.3554	0.825	2.856	0.004	0.739	3.972			
month_sep	0.9073	0.722	1.257	0.209	-0.507	2.322			
poutcome_failure	-0.6163	0.488	-1.263	0.207	-1.573	0.340			
poutcome_success	3.0608	0.880	3.479	0.001	1.337	4.785			
poutcome_unknown	-1.3966	0.651	-2.146	0.032	-2.672	-0.121			

```
2
In [16]: X_train_i = X_train_i.copy()
X_test_i = X_test_i.copy()
for column in X_train_i.columns:
    if full_mod.pvalues.loc[column] > 0.2:
        X_train_i.drop(column,axis=1, inplace=True)
        X_test_i.drop(column,axis=1, inplace=True)
```

```
In [17]: int_mod = sm.GLM(y_train, sm.add_constant(X_train_i), family=sm.families.Binomial()).fit()
```

```
In [18]: int_mod.summary()
```

```
Out[18]:
```

Generalized Linear Model Regression Results									
Dep. Variable:	y	No. Observations:	833						
Model:	GLM	DF Residuals:	810						
Model Family:	Binomial	DF Model:	22						
Link Function:	Logit	Scale:	1.0000						
Method:	IRLS	Log-Likelihood:	-339.30						
Date:	Wed, 07 Dec 2022	Deviance:	678.60						
Time:	20:18:20	Pearson chi2:	7.21e+04						
No. Iterations:	7	Pseudo R-sq. (CS):	0.4354						
Covariance Type:	nonrobust								
	coef	std err	z	P> z	[0.025	0.975]			
const	-0.7829	0.428	-1.831	0.067	-1.621	0.055			
duration	0.0051	0.000	11.932	0.000	0.004	0.006			
campaign	-0.1248	0.054	-2.298	0.022	-0.231	-0.018			
job_blue-collar	-0.8813	0.278	-3.175	0.001	-1.425	-0.337			
job_retired	0.5053	0.364	1.387	0.165	-0.209	1.219			
job_services	-0.7556	0.382	-1.956	0.048	-1.505	0.066			
job_unemployed	-1.9651	0.770	-2.572	0.011	-3.474	-0.045			
job_unknown	0.9767	0.966	1.011	0.312	-0.916	2.869			
housing_no	0.2981	0.217	1.372	0.170	-0.128	0.724			
loan_no	0.0069	0.236	0.029	0.977	-0.455	0.469			
loan_yes	-0.7893	0.302	-2.614	0.009	-1.382	-0.198			
contact_cellular	0.8836	0.429	1.594	0.111	-0.157	1.524			
contact_unknown	-0.6256	0.498	-1.256	0.209	-1.602	0.351			
month_aug	-0.7974	0.333	-2.393	0.017	-1.450	-0.141			
month_feb	-0.7335	0.410	-1.789	0.074	-1.537	0.070			
month_jan	-1.6234	0.610	-2.658	0.008	-2.819	-0.426			
month_jul	-1.2735	0.350	-3.638	0.000	-1.960	-0.587			
month_mar	-2.7703	1.080	-2.565	0.010	-0.654	-4.887			
month_may	-0.8171	0.298	-2.746	0.006	-1.400	-0.23			

ALL ROC Plots of Model

