

## Summary

This study aims to investigate the factors that drive banks to utilize incentive offers to attract deposits. Incentive offers are characterized by two main variables: the dollar amount of the bonus and the required deposit amount. To achieve this objective, two datasets consisting of banks' incentive text and their quarterly financial variables are provided. To match the bank names of different formats in the two datasets, I utilize the FuzzyWuzzy package to calculate the similarity scores, match those with the highest score if the score is above threshold 0.9 and verify that the mapping is 1-1. A sample matching result is presented in Table 1, where the "bankA" column contains lower case bank names in dataset A and "convert\_above\_threshold" contains corresponding lower case bank names in dataset B. This method matches 499 out of 1023 (48.8%) unique bank names in dataset A and 5549 out of 11182 (49.6%) campaigns. The unmatched names are either not found in dataset B or contain ambiguous information.

To extract possible factors affecting the two response variables, correlation matrix analysis of the dependent variables using the merged data frame is performed, and most pairs have correlation coefficients higher than 0.5, as shown in Fig 1. Subsequently, the Variance Inflation Factor is computed to screen out variables highly collinear with other variables. Using the selected variables ( $VIF < 10$ ) in Table 2, linear regressions are performed for bonus and deposit dependent variables, respectively.

The results in Table 3 indicate that the statistically significant variables ( $p < 0.1$ ) for the bonus regression model include "cash", "domdepservicecharges", "nbranch", and "netinc". When a bank has more cash, lower service charges on domestic deposits, a greater number of branches, and lower net income, it tends to offer a higher bonus. The residual histogram in Fig 2 and the q-q plot in Fig 3 suggest that residuals generally follow a normal distribution, indicating a good fit of the linear model to the bonus data. Moreover, the statistically significant variables ( $p < 0.1$ ) for the deposit regression model include "cash", "domdepservicecharges", "netinc", and "domdepservicecharges" as shown in Table 4. When a bank has more cash, lower service charges on domestic deposits, lower uninsured time deposits, and more net income, it tends to require a higher deposit. Similarly, Fig 4 and Fig 5 indicate a good fit of the linear model to the deposit variable. The high positive correlation of 0.63 between bonus and deposit shows that a higher bonus corresponds to a higher deposit required, which is also consistent with the previous result that they share three common factors.

However, a limitation of the utilization of linear regression is that observations are not entirely independent for campaigns of the same bank and in the same period. Future studies could include these factors as dummy explanatory variables or utilize other non-linear models to achieve better fitting results. Furthermore, the current explanatory variables only include bank financial variables, and certain macroeconomic variables may also be significant in stimulating banks' bonus incentives. Specific assumptions about processing data are shown in the comments of the code.

## Tables & Figures

	bankA	convert	score	convert_above_threshold
0	onewest bank	onewest bank fsb	95.0	onewest bank fsb
1	bb&t	bb&t bankcard corp	90.0	bb&t bankcard corp
2	suntrust	suntrust bank south central tn n	90.0	suntrust bank south central tn n
3	santander	santander bank national association	90.0	santander bank national association
4	discover	discover bank	90.0	discover bank
5	fifth district savings bank	fifth dist savings bank	92.0	fifth dist savings bank
6	mufg union bank, n.a.	ion bank	90.0	ion bank
7	usaa	usaa fsb	90.0	usaa fsb
8	tbk bank	tbk bank ssb	90.0	tbk bank ssb
9	new mexico bank & trust	new mexico bank and trust company	95.0	new mexico bank and trust company

Table 1. Sample Bank Names Conversion Data Frame

	VIF	Column
2	9.113261	timedep
1	8.733078	domdepservicecharges
5	6.955179	netinc
4	6.571823	nbranch
3	6.222759	timedepuninsured
0	5.868815	cash
6	2.982250	brokereddep

Table 2. VIF Table of Selected Variables

# OLS Regression Results of Bonus against Explanatory Variables

<b>Dep. Variable:</b>	y	<b>R-squared:</b>	0.197
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.196
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	193.7
<b>Date:</b>	Tue, 28 Feb 2023	<b>Prob (F-statistic):</b>	9.93e-258
<b>Time:</b>	22:23:44	<b>Log-Likelihood:</b>	-38710.
<b>No. Observations:</b>	5549	<b>AIC:</b>	7.744e+04
<b>Df Residuals:</b>	5541	<b>BIC:</b>	7.749e+04
<b>Df Model:</b>	7		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>const</b>	148.3235	4.176	35.521	0.000	140.138	156.510
<b>cash</b>	2.547e-06	1.26e-07	20.212	0.000	2.3e-06	2.79e-06
<b>domdepservicecharges</b>	-0.0005	5.8e-05	-8.103	0.000	-0.001	-0.000
<b>timedep</b>	-4.05e-08	6.54e-07	-0.062	0.951	-1.32e-06	1.24e-06
<b>timedepuninsured</b>	2.165e-07	9.54e-07	0.227	0.820	-1.65e-06	2.09e-06
<b>nbranch</b>	0.1121	0.012	9.012	0.000	0.088	0.136
<b>netinc</b>	-2.896e-05	7.31e-06	-3.965	0.000	-4.33e-05	-1.46e-05
<b>brokereddep</b>	9.302e-08	2.32e-07	0.401	0.689	-3.62e-07	5.48e-07

<b>Omnibus:</b>	7474.042	<b>Durbin-Watson:</b>	1.352
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	2432171.352
<b>Skew:</b>	7.533	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	104.451	<b>Cond. No.</b>	8.59e+07

Table 3. OLS Regression Results of Bonus against Explanatory Variable

OLS Regression Results of Bonus against Explanatory Variables

<b>Dep. Variable:</b>	y	<b>R-squared:</b>	0.182
<b>Model:</b>	OLS	<b>Adj. R-squared:</b>	0.181
<b>Method:</b>	Least Squares	<b>F-statistic:</b>	176.4
<b>Date:</b>	Tue, 28 Feb 2023	<b>Prob (F-statistic):</b>	1.61e-236
<b>Time:</b>	22:22:39	<b>Log-Likelihood:</b>	-67720.
<b>No. Observations:</b>	5549	<b>AIC:</b>	1.355e+05
<b>Df Residuals:</b>	5541	<b>BIC:</b>	1.355e+05
<b>Df Model:</b>	7		
<b>Covariance Type:</b>	nonrobust		

	coef	std err	t	P> t	[0.025	0.975]
<b>const</b>	3589.0508	778.458	4.610	0.000	2062.967	5115.134
<b>cash</b>	0.0004	2.35e-05	17.716	0.000	0.000	0.000
<b>domdepservicecharges</b>	-0.0577	0.011	-5.328	0.000	-0.079	-0.036
<b>timedep</b>	0.0002	0.000	1.697	0.090	-3.22e-05	0.000
<b>timedepuninsured</b>	-0.0014	0.000	-7.942	0.000	-0.002	-0.001
<b>nbranch</b>	-1.9717	2.319	-0.850	0.395	-6.518	2.574
<b>netinc</b>	0.0051	0.001	3.725	0.000	0.002	0.008
<b>brokereddep</b>	9.111e-05	4.33e-05	2.106	0.035	6.28e-06	0.000

<b>Omnibus:</b>	9862.357	<b>Durbin-Watson:</b>	1.443
<b>Prob(Omnibus):</b>	0.000	<b>Jarque-Bera (JB):</b>	11086378.787
<b>Skew:</b>	12.777	<b>Prob(JB):</b>	0.00
<b>Kurtosis:</b>	220.478	<b>Cond. No.</b>	8.59e+07

Table 4. OLS Regression Results of Deposit against Explanatory Variables

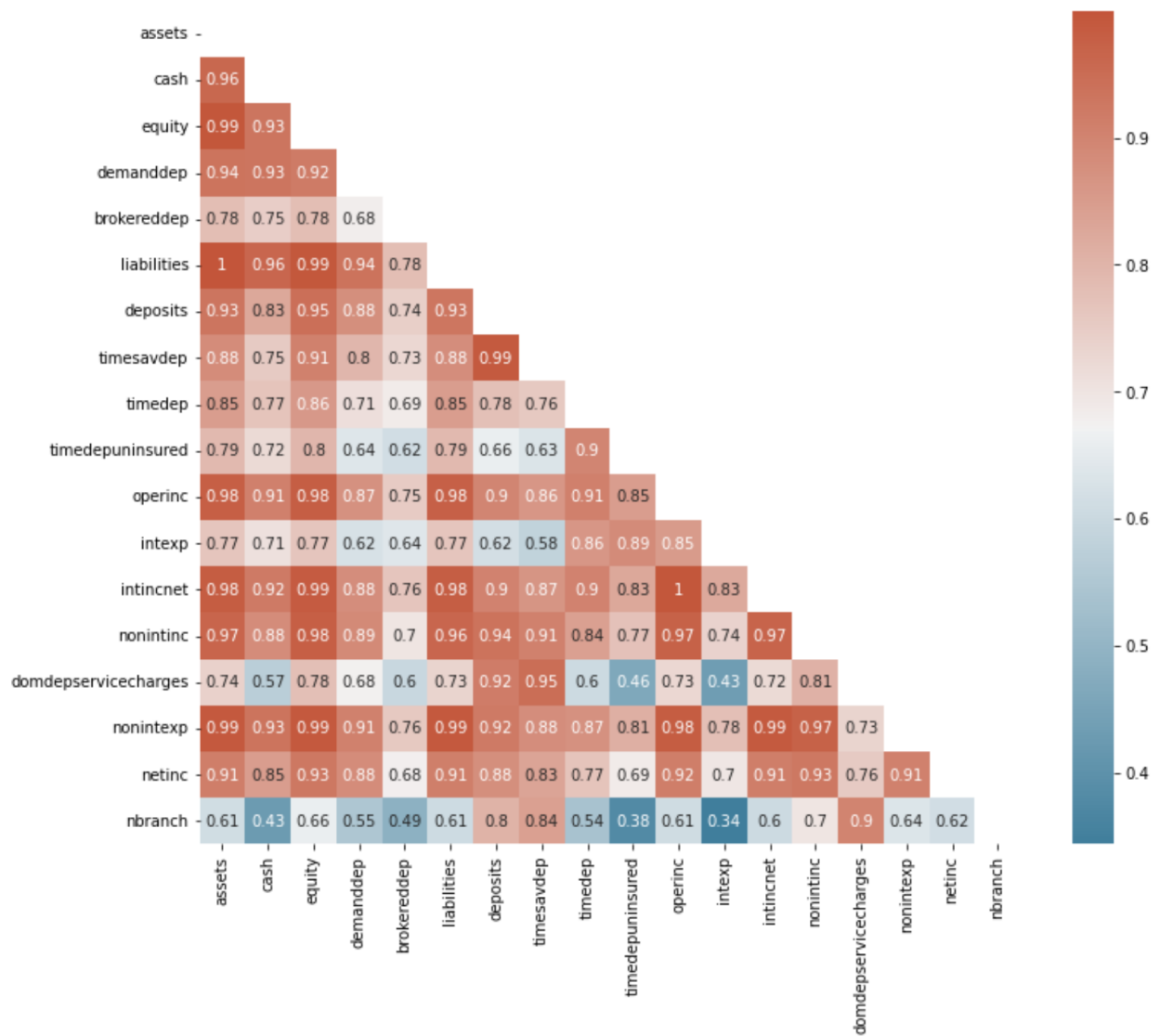


Fig 1. Correlation Heatmap of Explanatory Variables

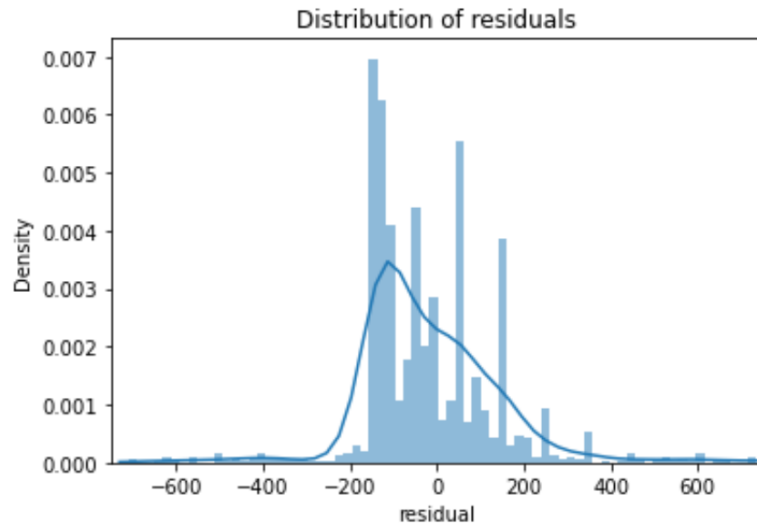


Fig 2. Histogram of Residuals in Bonus Regression Model

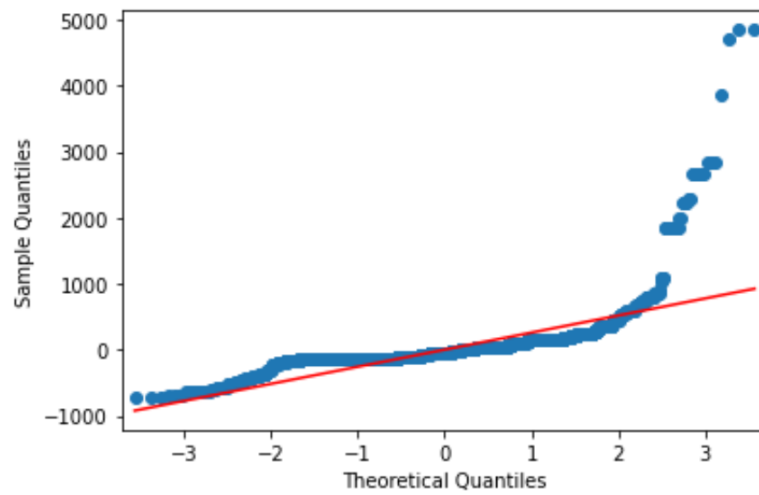


Fig 3. Q-Q Plot of Residuals in Bonus Regression Model

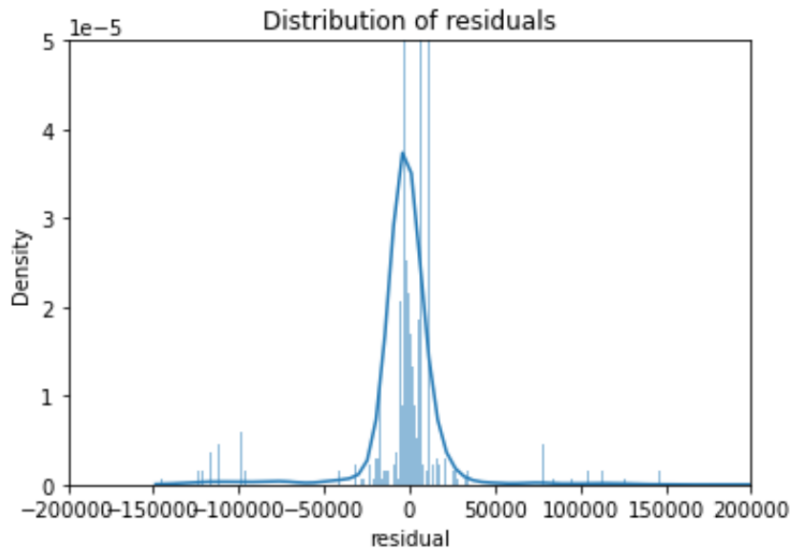


Fig 4. Histogram of Residuals in Deposit Regression Model

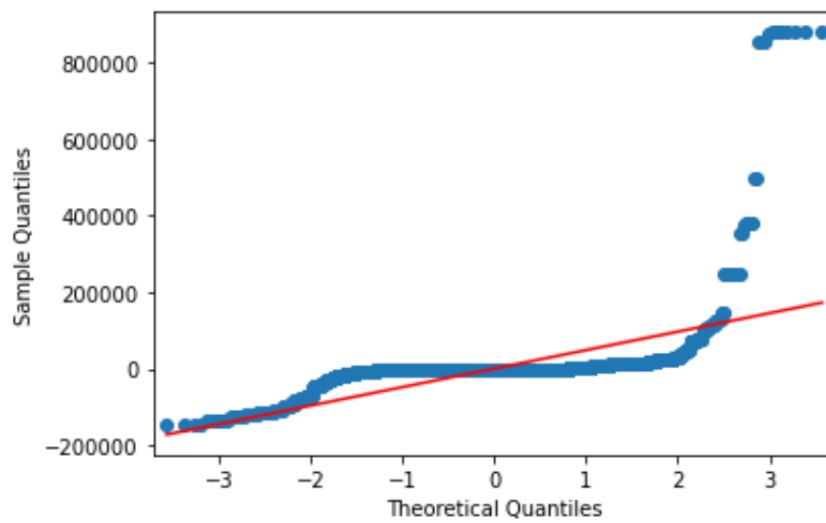


Fig 5. Q-Q Plot of Residuals in Deposit Regression Model