# results

*Significant Variables*

Upon completing association testing and preliminary modeling for each variable in the data set, variables that were determined to have a significant relationship (at a 0.0008 significance level) to the event of a customer purchasing an annuity were ranked by decreasing level of significance and grouped by class (binary, ordinal, nominal, continuous). At this significance level, we uncovered a total of 28 variables with a *positive relationship to our response (displayed in Table 1).*

**Table 1: Predictor variables with significant relationship to Annuity Purchase**

|  |  |  |  |
| --- | --- | --- | --- |
| Variable Description | P-Value | Class | Test |
| Certificate of deposit indicator | 0.000000 | Binary | Mantel-Haenszel |
| Checking account indicator | 0.000000 | Binary | Mantel-Haenszel |
| Money market account indicator | 0.000000 | Binary | Mantel-Haenszel |
| Monet market account balance | 0.000000 | Continuous | Logistic Regression |
| Savings account indicator | 0.000000 | Binary | Mantel-Haenszel |
| Number of checking deposits | 0.000000 | Continuous | Logistic Regression |
| Retirement account indicator | 0.000000 | Binary | Mantel-Haenszel |
| Certificate of deposit balance | 0.000000 | Continuous | Logistic Regression |
| Credit card indicator | 0.000000 | Binary | Mantel-Haenszel |
| ATM interaction indicator | 0.000000 | Binary | Mantel-Haenszel |
| Checking account balance | 0.000000 | Continuous | Logistic Regression |
| Investment account indicator | 0.000000 | Binary | Mantel-Haenszel |
| Number of telephone transactions | 0.000000 | Continuous | Logistic Regression |
| Number of money market credits | 0.000000 | Ordinal | Mantel-Haenszel |
| Branch of bank | 0.000000 | Nominal | Chi Squared |
| Value of home | 0.000000 | Continuous | Logistic Regression |
| Number of checks written | 0.000000 | Continuous | Logistic Regression |
| IRA balance | 0.000000 | Continuous | Logistic Regression |
| Direct deposit indicator | 0.000000 | Binary | Mantel-Haenszel |
| Number of insufficient fund issues | 0.000000 | Binary | Mantel-Haenszel |
| Number of credit card purchases | 0.000000 | Ordinal | Mantel-Haenszel |
| Safety deposit box indicator | 0.000000 | Binary | Mantel-Haenszel |
| Total ATM withdrawal amount | 0.000000 | Continuous | Logistic Regression |
| Number of point-of-sale transactions | 0.000001 | Continuous | Logistic Regression |
| Local address indicator | 0.000001 | Binary | Mantel-Haenszel |
| Amount of NSF | 0.000135 | Continuous | Logistic Regression |
| Total amount deposited | 0.000372 | Continuous | Logistic Regression |
| Number of cash back requests | 0.000706 | Ordinal | Mantel-Haenszel |

*Odds Ratios*

After calculating odds ratios, these values were also ranked in decreasing order for each of the 17 binary predictor variables in the data set and are displayed in Table 2. Row one of the table shows that the investment account indicator has the largest odds ratio, and from this we interpreted that customers with an investment account are approximately 3 and a half times more likely to purchase an annuity then customers without an investment account. In general, it appears that customers who participate in some form of saving or investing have higher odds of purchasing an annuity than customers that don’t.

*Linearity Assumption*

Testing of the linearity assumption for continuous variables revealed that of the 25 continuous variables, only nine of them met the assumption of linearity of the logit function. These variables are listed in Table 3.

**Table 3: Continuous predictors failing to meet linearity assumption and their associated p-values**

|  |  |
| --- | --- |
| Variable Description | P-Value |
| Credit score | 0.099214 |
| Income | 0.019015 |
| Credit card balance | 0.018398 |
| Mortgage balance | 0.015575 |
| Line of credit balance | 0.014866 |
| Age | 0.013942 |
| Value of home | 0.012699 |
| Length of residence | 0.011573 |
| Age of oldest account | 0.011394 |

*Data Considerations*

Closer examination of the data set also uncovered a substantial number of missing values among observations, where 13 of the predictors had over one thousand missing observations. Redundant variables were also identified, specifically among indicator variables for account types and their corresponding balances. Other redundancies were (*home ownership indicator, mortgage balance)* and (*credit card indicator*, *line of credit*). We also detected a nearly one-to-one relationship (95% correlation) between the mortgage balance and credit card balance variables.

Figure 1 shows the variables with more than one thousand missing observations, where the variables are sorted by the number of missing values.

Chart

Description automatically generated

**Figure 1:**