# results

*Exponential Smoothing Model*

Initial investigation and visualization revealed that the data had a noticeable seasonality with a magnitude of variation that appeared to remain *roughly* constant, indicating an additive seasonality (displayed in Figure 1). After evaluating several candidate models on the training data, the Holt/Winters model with additive seasonality was determined to minimize the mean absolute percentage error (MAPE).

*Visualizations*

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**

*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**

*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.

**Figure 1:**