

## RESPONSE MODEL

THE RESPONSE.CSV FILE CONTAINS OBSERVATIONS TAKEN FROM A RECENT DIRECT MARKETING CAMPAIGN TO WIN BACK FORMER CUSTOMERS. THE GOAL OF THIS EXERCISE IS TO

1. DEVELOP A PREDICTIVE MODEL TO BE USED TO SELECT RECORDS TO BE INCLUDED IN A FUTURE CAMPAIGN.
2. APPLY THE MODEL TO A NEW DATASET AND SELECT RECORDS TO MEET THE MARKETING TEAMS COST PER SALE GOALS.
3. DESIGN A SIMPLE AB COMPARISON TEST TO UNDERSTAND PERFORMANCE DIFFERENCES ATTRIBUTABLE TO VARIATION IN MARKETING TREATMENT.

PLEASE USE SAS OR R TO PERFORM THIS ANALYSIS.

### PART 1

PLEASE PERFORM THE FOLLOWING.

DIVIDE DATA SET INTO TRAINING AND VALIDATION SETS. SET ASIDE VALIDATION SET TO USE ONLY FOR EVALUATING MODEL OUT OF SAMPLE MODEL PERFORMANCE.

PERFORM DATA EXPLORATION TO REDUCE AND TRANSFORM THE NUMBER OF VARIABLES IN THE DATA SET TO **SET OF 10 PREDICTORS** THAT LOOK MOST PROMISING FOR PREDICTING SALES.

FIT A **LOGISTIC REGRESSION** MODEL TO PREDICT SALES USING THE TRAINING DATA SET.

EVALUATE THE PERFORMANCE OF THE MODEL ON THE TRAINING SET.

SCORE THE VALIDATION SET WITH THE MODEL AND EVALUATE ITS PERFORMANCE ON THE VALIDATION SET.

### PART 2

ASSUME THE VALIDATION SET REPRESENTS A NEW BASE FOR SELECTING RECORDS TO BE INCLUDED IN THE NEXT DIRECT MARKETING CAMPAIGN.

ASSUME A COST PER MARKETING TREATMENT OF \$2.

SELECT RECORDS FROM THE SCORED VALIDATION SET TO MAXIMIZE SALES WHERE THE GOAL OF THE CAMPAIGN IS FOR AVERAGE COST PER SALE TO BE \$6.

### PART 3

THE MARKETING TEAM WOULD LIKE TO TEST 2 VARIATIONS OF MARKETING TREATMENT IN THIS CAMPAIGN.

SPECIFY THE MINIMUM SAMPLE SIZE REQUIRED TO DETECT A 5% DIFFERENCE IN SALES RATE BETWEEN THE TWO TREATMENTS WITH 90% CONFIDENCE AND STATISTICAL POWER OF 80%.