Algorithm	**25% ($)**	**50% ($)**	**75% ($)**	**100% ($)**
P1				
AB	121,928	122,573	122,714	2,819,235
PCR	11,944	15,261	21,611	2,780,135
MLR	4,698	13,258	29,210	7,915,281
GLM	20,569	42,941	90,767	4,806,567
RT	2,880	13,270	21,631	1,841,204
M5	3	612	9,103	3,356,926
RF (n = 50)	1,918	10,591	26,615	2,748,779
P2				
AB	50,151	50,254	50,480	3,235,437
PCR	6,578	8,601	13,963	1,738,914
MLR	2,876	8,284	19,863	18,922,901
GLM	19,055	39,334	83,603	3,348,412
RTree	1,432	7,232	13,169	2,925,585
M5	0	72	1,158	4,848,265
RFR (n = 50)	642	5,373	15,229	1,619,046
P3				
AB	19,961	20,156	20,190	7,631,378
PCR	3,176	4,255	8,372	5,919,687
MLR	2,131	5,995	15,362	6,908,047
GLM	19,910	40,047	79,874	7,651,535
RTree	1,696	2,330	6,537	7,213,306
M5	0	7	125	7,621,932
RFR (n= 50)	190	2,090	7,382	7,267,663
P4				
AB	6,833	8,360	8,859	2,048,000
PCR	332	1,438	6,788	2,013,000
MLR	1,525	3,990	9,238	2,020,699
GLM	178	938	4301	2,056,000
RTree	2059	3912	8423	2,029,000
M5	290	915	3,717	2,038,000
RFR (n = 50)	1,341	2,920	8,146	2,018,000

Table 1: Absolute Error distribution summary for models. The values in column 25%, 50%, 75% show quartiles of the absolute error distribution (in dollars). For example, for **P2** scenario, and for model tress, 50% of the test set had a predicted value that was within 72 dollars of the true value. Here, AB = Average Baseline, PCR = Previous Cost Regression (Baseline), MLR = Multiple Linear Regression (Baseline), GLM = Generalized Linear Models (Baseline), RTree = Regression Tree, M5 = M5 Model Tree, and RFR = Random Forest Regression. For RFR, n is the number of trees, and for GLM, we assume a Poisson distribution and use the log link function.