

FINAL PROJECT

Home Sweet Home, Inc.

Renukalaxmi Dudhe

001239373

Richard A. Chaifetz School of Business,  
Saint Louis University

ITM6400-02 – Applied Business Analytics

Abhimanyu Gupta

The approach to solve the given questions is with the help of Prescriptive Analytics (or Optimization). This includes the below main concepts-

- Strategizing with Data – As the two fields Total value and Remodel for the first 20 houses are empty in the sheet, we must strategize the data to get these values first. After getting these values we can work on the data to get the answers.
- Genetic Algorithm – Apply genetic algorithm to optimize the data to maximize the total selling price of the house. It finds random solutions from the given input file and retains the best effective solutions among them.

### **Step-1 Strategizing with data (with R) in Statsbuddy:**

Open Statsbuddy.net. Go to analytics tab , then go to Prescriptive Analytics (or Optimization) and choose Strategizing with data (with R).

Use the **house.csv** file and upload it. Read the file .

Select the following values:

- ✓ **Dependent variable:** Total Value
- ✓ **Independent variable:** Lot. Sq. ft, Yr. Built, Gross Area, Living Area, Floors, Rooms, Bedrooms, Full bath, Half bath, Kitchen, Fireplace, Remodel.
- ✓ **Optimizing Variable:** Remodel
- ✓ **Type of Optimization:** Maximum
- ✓ **Size of Training data:** 5220-20 (From bottom)
- ✓ **Prediction Technique:** Random Forest (As it gives better result than other techniques)

Create the R code and copy it.

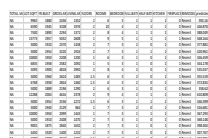
### **Step-2 Using R Studio Cloud:**

- ✓ First of all, upload the **house.csv** file in the R cloud Studio.
- ✓ Verify whether the required package ‘party’ is installed. If not, install it by using the code:  
`install.packages("party")`
- ✓ Run the R code .
- ✓ We will obtain the strategy file containing the 20 houses remodel fields filled and predicted total value data which has an accuracy of **96.59%**. After the code has been executed, export and save the Strategy file in CSV format.
- ✓ The obtained Strategy CSV file is the output of the Strategizing with data.

**RESULT** – house\_Strategy(1).csv is created with predicted total values and Remodel entries.

**Accuracy - 96.59%**

Excel file link-

A small icon representing an Excel spreadsheet, showing a grid of cells with some data.

(Double click the excel logo to open the file)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	TOTAL.VALUE	LOT.SQFT	YR.BUILT	GROSS.AREA	LIVING.AREA	FLOORS	ROOMS	BEDROOMS	FULL.BATH	HALF.BATH	KITCHEN	FIREPLACE	REMODEL	predicted
2	NA	9965	1880	2436	1352	2	6	3	1	1	1	0	Recent	392.161
3	NA	6590	1945	3108	1976	2	10	4	2	1	1	0	Recent	446.8704
4	NA	7500	1890	2294	1371	2	8	4	1	1	1	0	Recent	388.0497
5	NA	13773	1957	5032	2608	1	9	5	1	1	1	1	Recent	568.1644
6	NA	5000	1910	2370	1438	2	7	3	2	0	1	0	Recent	377.8822
7	NA	5000	1954	3220	1916	2	7	3	1	1	1	0	Recent	420.9613
8	NA	10000	1950	2208	1200	1	6	3	1	0	1	0	Recent	346.4595
9	NA	6835	1958	2582	1092	1	5	3	1	0	1	1	Recent	344.1785
10	NA	5093	1900	4818	2992	2	8	4	2	0	1	0	Recent	535.0377
11	NA	5000	1960	2624	1485	1.5	6	3	2	0	1	1	Recent	355.5193
12	NA	6768	1958	2844	1460	1.5	6	3	2	0	1	1	Recent	372.8347
13	NA	5000	1889	2196	1290	2	6	3	1	0	1	0	Recent	358.628
14	NA	12288	2004	4616	2378	2	9	4	2	1	1	1	Recent	610.8098
15	NA	5000	1954	2536	1272	1.5	6	3	1	1	1	1	Recent	346.3999
16	NA	5000	1940	2129	864	1	7	3	2	0	1	0	Recent	316.6811
17	NA	10000	1950	2099	1445	1	7	3	1	1	1	1	Recent	367.2933
18	NA	5000	1910	2408	1470	2	7	3	1	0	1	0	Recent	368.1482
19	NA	9001	1875	2840	1632	2	7	3	1	0	1	0	Recent	398.4047
20	NA	4450	1920	1400	1232	2	7	3	1	0	1	0	Recent	307.7673
21	NA	5000	1889	2560	1302	1.5	6	2	1	0	1	0	Recent	333.8956
22														

### Step-3 Use Genetic Algorithm(With R) in analytics in Statsbuddy:

The file house\_Strategy(1).csv obtained from above step is used as the input file for the genetic algorithm. File is uploaded and read in to the Genetic Algorithm program.

Select the following values:

- ✓ **Optimizing variable:** predicted
- ✓ **Constraint variable:** GROSS.AREA
- ✓ **Upper limit:** 20000 (as the sum of the gross area of the houses should be less than 20,000 sq feet)
- ✓ **Number of Iterations:** 100
- ✓ **Mutation Chance:** 0.01
- ✓ **Elitism:** True

Create the R code in Statsbuddy and copy it.

### Step- 4 Run the code in R Studio Cloud:

- ✓ Make sure the house\_Strategy(1).csv file is uploaded.
- ✓ Install the below genetic algorithm package  
***install.packages("genalg")***
- ✓ Paste the copied code and run it in the R Studio cloud. The Genetic Algorithm CSV file for the input house\_Strategy(1).csv is created named '**house\_Strategy (1)\_GeneticAlgorithm.csv**' .
- ✓ Export the file. The data in this file tells us which houses we should assign to ourself so that they maximize the total selling price.
- From the available 20 houses, we have got 9 houses with gross area sum = **19730 sq. ft.** with maximum predicted value of **\$3179.084**, and the same nine houses are needed to be remodeled as we have got 'recent' in the remodel column.

	TOTAL IN LAST 57 WEEKS	WEEKS OF MISSING DATA	MISSING	WEEKS OF MISSING DATA BETWEEN	PERCENT MISSING	percent							
NA	980	380	340	132	1	46	33	12	1	0	0	Percent	260.30
NA	700	250	220	171	2	8	4	11	1	0	0	Percent	580.65
NA	500	150	130	110	2	6	3	11	0	0	0	Percent	466.59
NA	500	380	230	120	1	4	6	3	11	0	0	Percent	236.68
NA	500	240	220	184	1	7	3	3	11	0	0	Percent	256.81
NA	1000	250	200	160	1	7	3	11	1	1	0	Percent	367.920
NA	500	150	140	140	2	7	7	3	11	0	0	Percent	580.140
NA	440	150	140	120	1	7	3	11	0	0	0	Percent	367.970
NA	500	260	230	181	1	5	4	2	11	0	0	Percent	580.131

	TOTAL.VALUE	LOT.SQFT	YR.BUILT	GROSS.AREA	LIVING.AREA	FLOORS
1	NA	9965	1880	2436	1352	2.0
3	NA	7500	1890	2294	1371	2.0
7	NA	10000	1950	2208	1200	1.0
12	NA	5000	1889	2196	1290	2.0
15	NA	5000	1940	2129	864	1.0
16	NA	10000	1950	2099	1445	1.0
17	NA	5000	1910	2408	1470	2.0
19	NA	4450	1920	1400	1232	2.0
20	NA	5000	1889	2560	1302	1.5

	ROOMS	BEDROOMS	FULL.BATH	HALF.BATH	KITCHEN	FIREPLACE	REMODEL
1	6	3	1	1	1	0	Recent
3	8	4	1	1	1	0	Recent
7	6	3	1	0	1	0	Recent
12	6	3	1	0	1	0	Recent
15	7	3	2	0	1	0	Recent
16	7	3	1	1	1	1	Recent
17	7	3	1	0	1	0	Recent
19	7	3	1	0	1	0	Recent
20	6	2	1	0	1	0	Recent

	predicted
1	392.1610
3	388.0497
7	346.4595
12	358.6280
15	316.6811
16	367.2933
17	368.1482
19	307.7673
20	333.8956

```
> write.csv(dataset[solution == 1, ], "house_Strategy (1)_GeneticAlgorithm.csv", row.names=F, FALSE)
> -GAmode1$best[iter]
[1] 3179.084
```

1. From the above approach followed with data analytics techniques, I would assign myself the house numbers 1,3,7,12,15,16,17,19,20 so that they maximize the total selling price i.e. **\$3179.084** with the constraint that the sum of the gross area of those houses is less than 20,000 sq. ft. i.e. **19730 sq. ft** as per my findings.
2. To get the maximum value for the houses I should remodel the house numbers **1,3,7,12,15,16,17,19,20** before enlisting them to the organization Home Sweet Home. Inc.,

P.S – Separate file is attached for R Codes.