Case: CoffeePlus

Hindupur Ramakrishna
University of Redlands

Avijit Sarkar
University of Redlands

Bindiganavale Vijayaraman
The University of Akron

ABSTRACT

CoffeePlus is a decision-based role-playing case focusing on data-driven analysis in the context of an upscale coffee franchise making retail site-location decisions for growth (expansion) and to address competitive disadvantages. To date, managers at CoffeePlus had primarily used their intuition for site-location decisions rather than data analysis. Their intuitive process had resulted in failure to meet profitability goals.

The students' role is as analysts/consultants responding to a request from the CoffeePlus Business Development Executive seeking to improve the site-location decision-making process through data-driven analysis. The case is in the form of a memo from the Business Development Executive. The case memo lays out the background, nature of the problem, and general guidance in the form of a three-part (4-Step) process beginning with problem definition. Data for students to select and analyze is provided with the case. Students are specifically introduced to location analytics in this case as they are required to evaluate two different solutions: solution with geo-influenced data and solution without geo-influenced data.

The case was originally developed for use in an MBA course titled "Data-Driven Decision-Making" which focuses on descriptive and predictive analytics. CoffeePlus is an integrated case linking various parts of these types of courses. As data-driven decision-making processes are central for students to address in analyzing the case, a model of the data-driven decision-making process was developed for the MBA course by Ramakrishna, Sarkar, and Vijayaraman, (2022). It is presented as a part of the teaching notes. This is fictitious case with mostly real data and some fabricated data.

Keywords: Business problems; problem-solving; decision-making; predictive analytics; location analytics; evaluating solutions.

This is a fictitious case. All information contained herein was fabricated by the author(s). Any similarity contained herein to actual persons, businesses, events, etc., is purely coincidental and is the responsibility of the author(s). Please contact the case author(s) directly with any concerns.

COFFEEPLUS – MEMO

TO: Chai Coffey, Analyst
FROM: John Expander, Director of Business Development

As you know, ten years ago, CoffeePlus, an upscale retailer focusing on coffee and related items, was started to compete with retailers like Starbucks and Peet's coffee. Over the years our expansion has had many ups and downs partly due to our non-scientific approach to selecting locations for CoffeePlus retail stores. The location decision has been made by a few top level managers using their intuition. As a result, about 400 of our 2000 stores are not as profitable as promised. This poses a problem for the future viability of this franchise. I was hired to fix the issue with site selection. Given my background in data driven decision-making I would like to approach the site selection problem with data and analysis.

The plan for development calls for adding approximately 200 stores per year for the next decade. I would like you to develop an approach using relevant data and analysis for selecting locations. Though our executive committee members are not very familiar with analytic techniques, I recommend you use the right analytics techniques and tools to develop and present an approach for site selection. However, your final report should be in simple English so that anyone (including our executive committee members) can understand it.

To tackle the problem in a manageable manner, here is the three-part (four-step) process I recommend:

PART A

Develop a clear definition of the right problem and state it clearly. Make sure to state it in such a way it can be used as a guide to understand data needs. Keep in mind that "a problem well defined is half solved." Please ensure the right problem is stated and it is stated right (i.e., correctly and precisely). All subsequent analytic work depends on this step and, hence, make sure you spend an adequate amount of time in this step.

Identify some variables that could be used for analyses to provide a good solution to the problem posed/stated in step 1. In this step, list each of the variables and their characteristics – i.e., whether the variable is qualitative or quantitative, dependent, or independent, and insure you specify the unit of measurement (like pounds, dollars, miles, etc.). Also list some sample values (you can make these up) for each variable listed. Please use the structure of the following table for listing all the variables:

#	Variable	Unit of measurement	Dependent? (Yes or No)	Quantitative? (Yes or No)	Sample values
1					
2					
3					

Note: Getting data sometimes can be costly. Please keep this in mind as you come up with the list of variables in this step.

What analyses (note, this is plural) would you recommend with the data you have specified, to address the problem?

For each analysis you are recommending, please specify the analysis, list all the variables you plan to use in that analysis, and then defend the choice of the analysis (i.e., justify why you need the analysis and describe how the results of the analysis will help address the stated problem in step 1).

PART B

CoffeePlus has the following (most recent) data for the last year from existing stores:

- Average weekly sales for the year (in dollars)
- Average traffic volume in front of the store (in thousands per week)
- Number of competitors within one mile of the store?
- Number of people who live within one mile of the store
- Average family income for people who live within one mile of the store (in dollars)
- Median family income for people who live within one mile of the store (in dollars)
- Ease of entry/exit into and out of the store (a maximum score of 100 – 0 indicates very difficult)

Data from 111 stores, randomly selected from 2000 stores, is provided in Table 1 (included at the end of this case) in the format listed below:

Sales	Traffic	Competition (1Mile)	Population (1Mile)	AvgIncome (1Mile)	MedianIncome (1Mile)	Ease

Note: Your instructor can provide this data in an Excel worksheet. Alternatively, you can also cut the data from the Word file and paste the data into Excel.

Before you run any analysis, identify any assumptions (about the data, stores, etc.) you must make, list them, and justify why each of the assumptions is necessary.

Now, run a multiple linear regression analysis with sales as the dependent variable and all the other variables as independent. In your analysis, make sure you identify (through variables selection process in regression analysis such as best subsets analysis, stepwise selection, etc.) the relevant variables and use only those variables in the final regression analysis.

Note: You may want to build the regression model by using the data from 100 stores and use that model to predict sales for the last 11 stores. You can then use the difference between predicted sales and the actual sales as a measure of validity of the model built – the smaller the difference, the better the model.

Finally, make a case for geography to be explicitly considered in the analyses you have completed in this step – i.e., why geography matters and how (for the analyses). Would explicitly considering geography change the analysis for the better (i.e., better selection of retail locations? Why? How?

PART C

You believe that for many locations the local geography (like bodies of water, mountains, etc.) can influence some of the data you have used to build the regression model in step 3. You request that the population and income variables be redefined as follows:

- Number of competitors within ~~one mile~~ *15 minutes driving distance (DD)* of the store?
- Number of people who live within ~~one mile~~ *15 minutes driving distance (DD)* of the store
- Average family income for people who live within ~~one mile~~ *15 minutes driving distance (DD)* of the store
- Median family income for people who live within ~~one mile~~ *15 minutes driving distance (DD)* of the store

Here, we are interested in time and not distance – i.e., driving time (DT) instead of driving distance (DD). The data on the same 111 stores with the data on redefined variables is provided in Table 2 (included at the end of this case) in the format listed below:

Sales	Traffic	Competition (DT)	Population (DT)	AvgIncome (DT)	MedianIncome (DT)	Ease

Note: Your instructor can provide this data in an Excel worksheet. Alternatively, you can also cut the data from the Word file and paste the data into Excel.

Please re-run a multiple linear regression analysis (with sales as dependent variable and the other variables as independent) with the new dataset. In your analysis, make sure you identify (through the variable selection process such as best subsets analysis, stepwise selection, etc.) the relevant variables and use only those variables in the final regression analysis.

Note: You may want to build the regression model by using the data from 100 stores and use that model to predict sales for the last 11 stores. You can then use the difference between predicted sales and the actual sales as a measure of validity of the model built – the smaller the difference, the better the model.

Which predictive model would you recommend (i.e., from step 3 or 4)? Why (i.e., what is the basis for your conclusion)?

Please develop an executive summary of your findings targeted at executives who may not have any background in statistics and hence the summary should be in plain English (and devoid of any statistical terms or symbols). Include all the relevant analysis details in an Appendix – details of data, listing of variables (and their characteristics – quantitative/qualitative, dependent/independent, etc.), type of analysis that is appropriate, any hypotheses, all relevant results of analysis (including screenshots) highlighted to point out important parts, statistical and business conclusions from the results. In your executive summary, make sure you cite relevant sections of the Appendix to support your

discussion/conclusions so that anyone interested in looking at supporting statistical evidence can do so.

Table 1 (Data for Step 3)

ID	Sales	Traffic	Comp.	Pop.	AveInc	MedInc	Ease
1	20538	15	28	85826	126382	85983	55
2	20538	15	71	87604	113026	63597	95
3	26019	19	5	37300	133436	95180	48
4	20538	15	18	79958	162711	108401	95
5	20538	15	71	83688	117471	68611	88
6	20538	15	73	122427	105827	58022	58
7	20538	15	75	110376	109507	61479	82
8	20538	15	16	36020	180188	122582	74
9	20538	15	6	28003	82274	53343	77
10	17096	15	1	21345	87906	57895	24
11	21904	16	79	131634	105082	59141	25
12	17096	15	3	19921	102756	75983	93
13	17096	15	3	19496	100894	74013	44
14	5712	5	2	4	0	0	20
15	12327	9	63	75514	123734	79588	44
16	16423	12	67	89070	116374	70467	94
17	20538	15	48	53075	114620	67071	48
18	20538	15	22	70321	129073	79944	46
19	20538	15	15	57155	138606	89110	79
20	15962	14	8	29845	101937	72406	79
21	20538	15	52	69828	136252	88348	72
22	17096	15	0	31869	135949	89409	68
23	17096	15	13	27803	69224	40189	40
24	17096	15	5	50535	81280	57015	71
25	18231	16	14	33380	62663	36748	84
26	15058	11	13	46539	188275	129283	41
27	20538	15	69	117751	107741	59099	28
28	16423	12	71	124141	106642	58810	59
29	16423	12	64	72517	121768	75382	49
30	20538	15	9	60698	151913	101735	79
31	21904	16	69	93596	115644	68111	65
32	17096	15	14	39301	61800	37390	98
33	27385	20	70	106774	111881	65393	28
34	8212	6	5	53880	173151	118718	64
35	20538	15	17	89425	147123	97716	75
36	19173	14	63	66643	125814	79493	64
37	17096	15	3	25385	87538	65683	59
38	22788	20	6	38190	74489	46562	100
39	22788	20	2	6080	55528	41762	70

40	14808	13	6	30631	73796	50919	73
41	17096	15	1	29531	76181	51328	98
42	20538	15	26	119520	122904	77556	37
43	21904	16	23	110066	139282	90224	34
44	16423	12	62	63260	124661	77863	62
45	6846	6	4	40131	105112	70357	53
46	15058	11	16	87379	143287	95912	26
47	20538	15	16	83300	142710	94084	44
48	20538	15	16	82550	144490	95268	68
49	20538	15	4	55235	122408	83314	26
50	20538	15	77	150097	101846	56746	71
51	20538	15	10	53057	144569	101242	65
52	26019	19	7	54467	128956	95031	81
53	16423	12	57	57846	126615	82854	96
54	21904	16	15	57155	138606	89110	94
55	27385	20	62	136794	94062	52406	97
56	20538	15	75	133259	104705	57284	24
57	26019	19	72	132047	106097	58444	57
58	13692	10	31	54812	162885	110798	21
59	20538	15	32	56147	162816	110755	43
60	17096	15	1	48035	114453	75319	46
61	20538	15	6	37860	195259	134911	60
62	6846	6	0	31447	83356	67606	86
63	17096	15	0	18790	109018	86237	92
64	17096	15	3	16900	85624	63956	53
65	3808	3	3	24574	105617	77386	64
66	27385	20	28	84931	129260	88498	55
67	12327	9	58	55837	128150	82460	24
68	16423	12	68	95002	113066	64014	64
69	17808	13	70	120838	107189	58958	74
70	19173	14	46	135670	91989	52644	93
71	20538	15	67	87074	116004	67201	79
72	20538	15	50	129785	92509	51984	34
73	20538	15	81	143181	104898	58545	41
74	20538	15	28	105114	127506	83499	64
75	21904	16	67	81638	117837	69491	50
76	20538	15	81	134732	103816	57619	84
77	20538	15	62	64075	124712	78214	85
78	19173	14	70	124006	105424	57817	87
79	21904	16	70	124489	107025	59010	82
80	13673	12	0	13276	229702	168988	53
81	20538	15	74	124401	105326	57644	84

82	9596	7	7	60834	123111	96493	30
83	20538	15	5	48684	131259	94377	44
84	20538	15	78	145954	104722	58569	56
85	9596	7	17	37210	176794	121016	28
86	10962	8	63	74278	123110	79216	21
87	22788	20	7	33324	113114	81828	31
88	13692	10	8	55651	159302	108100	66
89	12538	11	1	41424	112636	75052	91
90	20538	15	54	49952	129184	84208	35
91	20538	15	55	52248	130709	85146	34
92	17096	15	0	31689	133098	86965	38
93	22788	20	0	32255	130742	85960	96
94	21904	16	36	135466	119979	72803	49
95	24635	18	3	37114	180074	126900	45
96	23269	17	25	66855	149120	102356	66
97	6846	6	5	27375	112481	81085	65
98	17096	15	5	27160	111894	80473	41
99	17096	15	1	32579	76745	52090	91
100	13692	10	71	105100	112646	65199	44
101	24635	18	55	45169	140668	96057	37
102	21904	16	78	141618	106756	59491	81
103	9115	8	14	54786	71704	47579	67
104	21904	16	28	72078	135289	88719	38
105	23269	17	19	92310	160003	107668	79
106	20538	15	39	159339	103522	58600	84
107	20538	15	45	155204	105351	59084	42
108	17096	15	3	30918	94862	71916	80
109	6846	5	27	127369	118942	74313	77
110	27385	20	3	38062	181249	132212	20
111	20538	15	4	38279	126310	88478	75

Table 2 (Data for Step 4)

ID	Sale	Traffic	Comp.	Pop.	AveInc	MedInc	Ease
1	20538	15	134	648936	126278	84296	55
2	20538	15	138	381062	120858	78470	95
3	26019	19	34	591125	128075	93548	48
4	20538	15	79	609254	128839	89123	95
5	20538	15	146	433303	125139	81631	88
6	20538	15	137	388949	118665	76913	58
7	20538	15	66	516563	124746	85605	82
8	20538	15	126	500385	114669	76711	74
9	20538	15	151	557984	126476	83165	77
10	17096	15	51	376663	98749	61664	24
11	21904	16	155	577509	127852	84161	25
12	17096	15	53	272707	91205	60001	93
13	17096	15	52	268388	90698	59846	44
14	5712	5	9	160770	87610	60570	20
15	12327	9	118	268219	127619	80425	44
16	16423	12	120	266719	127839	80651	94
17	20538	15	112	256177	128307	80702	48
18	20538	15	104	244897	125662	78428	46
19	20538	15	96	213257	125361	77591	79
20	15962	14	56	487056	87691	58409	79
21	20538	15	141	583874	119366	79110	72
22	17096	15	63	436923	97894	61935	68
23	17096	15	59	450838	97446	60385	40
24	17096	15	55	380599	103935	64119	71
25	18231	16	66	459679	98248	62050	84
26	15058	11	98	296029	124551	77989	41
27	20538	15	131	339791	127943	82101	28
28	16423	12	135	355898	130063	83530	59
29	16423	12	125	299292	126785	80890	49
30	20538	15	104	449653	128822	83894	79
31	21904	16	129	331264	128657	82423	65
32	17096	15	65	469658	98415	61928	98
33	27385	20	122	273517	128576	81429	28
34	8212	6	63	646512	130570	91988	64
35	20538	15	155	488696	132179	86086	75
36	19173	14	124	286461	127893	81492	64
37	17096	15	5	387389	86675	62939	59
38	22788	20	63	586506	95557	61818	100
39	22788	20	33	499265	80407	56627	70
40	14808	13	62	510929	92830	60222	73

41	17096	15	55	431268	98913	61741	98
42	20538	15	166	527655	132819	86809	37
43	21904	16	149	442608	130587	84268	34
44	16423	12	144	408760	121328	78632	62
45	6846	6	62	487691	100147	62700	53
46	15058	11	124	586228	133399	88402	26
47	20538	15	56	373272	140577	96802	44
48	20538	15	139	519867	132229	86485	68
49	20538	15	139	519081	131929	86243	26
50	20538	15	154	456458	131621	85464	71
51	20538	15	72	531972	134844	92823	65
52	26019	19	63	497285	136760	94523	81
53	16423	12	116	264285	131243	83734	96
54	21904	16	96	213257	125361	77591	94
55	27385	20	162	618636	130887	86133	97
56	20538	15	144	404206	128320	83210	24
57	26019	19	133	349957	126218	81211	57
58	13692	10	136	665501	118290	79841	21
59	20538	15	137	673783	118330	79862	43
60	17096	15	59	502652	97884	60974	46
61	20538	15	83	347093	131296	84647	60
62	6846	6	4	421126	87486	64601	86
63	17096	15	13	535034	90932	62758	92
64	17096	15	12	429886	82716	59741	53
65	3808	3	56	320672	94613	61162	64
66	27385	20	122	619146	125045	83567	55
67	12327	9	122	280453	126196	80580	24
68	16423	12	139	359578	120149	77505	64
69	17808	13	145	396497	125688	81599	74
70	19173	14	165	649846	133444	88254	93
71	20538	15	129	607339	136138	91051	79
72	20538	15	159	635001	130004	85850	34
73	20538	15	146	401211	127210	82522	41
74	20538	15	129	324705	126702	81084	64
75	21904	16	128	317893	126123	80678	50
76	20538	15	154	536277	127125	83529	84
77	20538	15	136	355616	122166	78969	85
78	19173	14	147	410416	127353	82709	87
79	21904	16	147	411500	127387	82733	82
80	13673	12	50	466795	100345	61506	53
81	20538	15	151	462429	126683	82757	84
82	9596	7	29	374973	130808	93062	30

83	20538	15	37	620990	128770	93355	44
84	20538	15	151	469287	127235	83139	56
85	9596	7	135	668633	121847	81690	28
86	10962	8	118	268298	127624	80423	21
87	22788	20	43	360308	81547	55020	31
88	13692	10	78	507320	135906	93634	66
89	12538	11	57	425612	100206	61657	91
90	20538	15	119	276602	130565	83323	35
91	20538	15	119	280363	128836	82604	34
92	17096	15	55	383441	104036	64208	38
93	22788	20	55	415331	103532	64109	96
94	21904	16	138	362574	129694	82796	49
95	24635	18	73	724203	128067	90304	45
96	23269	17	138	640340	123874	82767	66
97	6846	6	25	179673	89384	59624	65
98	17096	15	25	159217	92046	61516	41
99	17096	15	57	466923	98538	61513	91
100	13692	10	126	313875	129485	82759	44
101	24635	18	131	338545	122707	79710	37
102	21904	16	149	428672	128837	83670	81
103	9115	8	61	428422	99552	62758	67
104	21904	16	151	706020	125754	83742	38
105	23269	17	136	371541	128260	81691	79
106	20538	15	161	493170	134560	87945	84
107	20538	15	157	466021	133920	87329	42
108	17096	15	7	373726	85552	61415	80
109	6846	5	168	545880	133188	87233	77
110	27385	20	55	694141	130268	93213	20
111	20538	15	39	695133	127201	92361	75

Teaching notes – CoffeePlus

CoffeePlus is a decision-based role-playing case focusing on data-driven analysis in the context of an upscale coffee franchise making retail site-location decisions for growth (expansion) and to address competitive disadvantages. To date, managers at CoffeePlus had primarily used their intuition for site-location decisions rather than data-driven analysis. Their intuitive process had resulted in failure to meet profitability goals. The students' role is as an analyst responding to a request from the CoffeePlus Business Development Executive seeking to improve the site-location decision-making process through data-driven analysis. The case is in the form of a memo from the Business Development Executive. The case memo lays out the background, nature of the problem, and general guidance in the form of a three-Part four-Step process beginning with problem definition. Data for students to select and analyze is provided with the case.

The case was developed for use in an MBA course titled "Data Driven Decision-Making" which focuses on descriptive and predictive analytics. CoffeePlus is an integrated case linking various parts of these types of courses. As data-driven-decision-making processes are central for students to address in analyzing the case, a model of the Data-Driven-Decision-Making (DDDM) process was developed for the MBA course by Ramakrishna, Sarkar, and Vijayaraman (2022).

TARGET AUDIENCE

The case is appropriate for MBA students in introductory business analytics courses that address predictive analytics, applied business statistics courses that address multiple regression analysis as well as in specialized master's programs in business analytics. The case can also be used for undergraduate business students in introductory business analytics courses that address predictive analytics as well as applied business statistics courses that covers multiple regression. Depending on the depth of coverage of topics in an undergraduate course, minor modifications to suit the coverage in the course may be helpful in using the case for undergraduate students – for example (1) Steps 3 and 4 can be deleted from the case in courses where multiple regression analysis is not covered, or (2) Steps 3 and 4 can be modified to include analysis of variance (ANOVA), correlation analysis, and/or simple regression analysis in courses where multiple regression analysis is not explicitly covered.

TEACHING STRATEGY AND LEARNING OBJECTIVES

Purpose of the case

This is an integrated case that links different parts of the course. Instructors using the case can use the model developed (i.e., the data-driven decision-making process model – (see Ramakrishna, Sarkar, and Vijayaraman, 2022) as the mechanism for integrating different parts of the course or a similar model of their choice given the chosen model facilitates necessary integration.

This case specifically addresses the following:

(1) defining the right business problem,

(2) defining the business problem right (in the form of a statistical problem),

(3) Linking the problem definition with the data needs and explicitly specifying all the

variables and their characteristics,
(4) linking the data specified with the analyses necessary to address the problem defined through the results of the analyses performed (and justifying why the analyses specified are the right ones),
(5) explicitly identifying and stating any assumptions needed regarding the data or the problem context and justifying the need for those assumptions,
(6) performing a multiple regression analysis (including variable selection) and making predictions to validate the model with the holdout sample, and explicitly stating and testing any assumptions necessary for the analyses,
(7) explicitly recognizing that location/geography may have an impact on the analysis performed and perform the analysis (i.e., multiple regression analysis) with data adjusted for geographic impact (geography-influenced data), and, finally
(8) evaluating the results from two different types of analyses and making a recommendation for the predictive model to be used and providing adequate justification for recommendations.

In addition, a key component of the case assignment is the students' executive summary. Guide students to write an executive summary that uses clear and appropriate business language for executives who may not have any analytical/statistical background. With respect to the report, guidance should include directions linking the report with all necessary supporting analytics work completed. A sample guide to "Executive Summary" is included at the end of the teaching notes (see Appendix A).

The following analytics tools are typically used:

- Excel (with data analysis tool pack), Tableau, and Minitab.
- The case analysis is done using Minitab. However, any statistical software will suffice. Potential alternatives include PHStat, MegaStat, SPSS, SAS, JMP and similar tools.
- Predictive analytics that include the impact of geography. Step 4 of the case incorporates the role of geography. As the role of geography is usually missing in most predictive analyses, it may be helpful to reinforce its utility.

Teaching Tactics and Student Learning Objectives

As CoffeePlus is an integrated case, listed below in the recommended sequence are suggestions for introducing different parts of the case in different class sessions.

Part A Step 1 deals with the problem identification and specification – defining the right problem and defining the problem right (in the form of a statistical problem). When regression analysis is discussed in business statistics courses, this part is usually not discussed as much. This part, we believe, is the most important part as "a problem well defined is half solved." It is important to emphasize that in the real world no one gives the decision-maker well-defined problems (as we see in textbooks). Part A Step 2 deals with data and its connection to the problem definition (from Part A step 1) – addressing "do we have the right data?" and then "are we using the data right?"

Part B Step 3 addresses "are we doing the right analysis?" and then "are we doing the analysis right?" The analysis focuses on multiple regression analysis with data that does not explicitly consider geography – i.e., geography-influenced data (for example, we will use data/variable like "average family income for people who live within one mile of the store." This

variable does not consider the impact of geographic features like bodies of water, mountains, etc. in that one mile).

Part C Step 4 is similar to Part B Step 3 except that the data used is geography-influenced (for example, we will use data/variable like "average family income for people who live within one mile 15 minutes driving distance of the store." These types of variables explicitly consider the impact of geographic features like bodies of water, mountains, etc.).

The final part of Part C addresses the evaluation of models developed in Part B and Part C – which model to use and why (and whether geography matters or not, in this specific problem).

PART A

Step 1

Introduce or assign this case step in a class session that includes a discussion about identifying and stating business problems that can be effectively solved using descriptive or predictive analytics. (The case write-up that includes only Step 1 will be handed to students for this class session.)

Case step 1 leads to the learning outcomes below. After analyzing the case step 1, the students will be able to:

1.1 Clearly state the business problem (in plain English) and defend why the problem stated is the real problem (i.e., stating the right problem)

1.2 Restate the problem so that some analytics work can be performed to solve the problem.

This statement will be in terms of data and the type of expected results from the analysis (i.e., stating the problem right)

Key questions to address in the discussion for Step 1 (Ramakrishna, Sarkar, and Vijayaraman, 2022):

1. Are we defining the right problem?
2. Are we defining the problem right?

Deliverables (by students):

- A clear definition of the problem with CoffeePlus and a detailed justification as to why the student believes s/he has defined the right problem.
- A redefinition of the problem in the form of variables and the type of results expected from the analytics on the data available/obtained.

There are no right or wrong problem definitions. Hence it is important to evaluate the justification provided by the students for their definitions. This could be either a group or individual work.

Class discussion (Total time – 50 to 80 minutes)

- Problem solving process and data driven decision-making model (20 - 30 minutes)
- Problem definition and role of problem-solving approaches/processes in defining the right problem (10 - 15 minutes)

- The need for domain knowledge in defining the right problem and the concept of "selective perception" and its impact on problem recognition/definition (5 - 10 minutes)
- Need for explicitly recognizing and stating assumptions (5 - 10 minutes)
- Need for using clearly defined variables in the problem restatement (stating the problem right) (10 - 15 minutes)

Such topics as epistemology (i.e., discussion of Hegelian dialectic or the devil's advocate approach, discussion of Leibnizian or Lockean approach to problem definition, etc.), exploratory data analysis or confirmatory data analysis approaches to finding/defining problems, etc. can be added into classroom discussions as time/interest permits.

One potential solution to this step: One definition for the problem may be "how do we reduce the likelihood of opening non-profitable CoffeePlus stores?" Here the discussion could center on clearly defining what does "non-profitable" mean and the timeframe for this non-profitability (i.e., are we defining a store as non-profitable after one year or some other time period?) and on understanding what "likelihood" means. One way this problem can be restated is "Can we predict the average weekly sales of a future store based on local data on characteristics/factors/variables?"

Step 2

Introduce or assign this case step in a class session that includes a discussion of the connection between the problem statement and the data needs, as well as a discussion of the link between data and analyses. (The case write-up that includes only Steps 1 and 2 will be handed to students for this class session. In addition, students will be required to bring their answers to Step 1 of the case for this session).

Case step 2 leads to the learning outcomes below.

After analyzing the case Step 2 assigned/discussed, the students will be able to identify the data needs and explicitly specify all variables and their characteristics as they relate to the problem identified in case steps 1.1 and 1.2.

Connect the data specified with the analyses necessary to address the problem defined through the results of the analyses performed. In other words, justify why the analyses specified are the correct ones. Key questions to address in the discussion of Step 2 are listed in Ramakrishna, Sarkar, and Vijayaraman, (2022).

3. Do we have the right data?
4. Are we using the data right?

Deliverables (by students):

- An exhaustive listing of all the variables that relate to the problem defined in Step 1 and the characteristics (i.e., qualitative/quantitative, dependent/independent, etc.) of each of the variables listed. Here is the suggested structure for this part of the deliverable:

#	Variable	Unit of measurement	Dependent? (Yes or No)	Quantitative? (Yes or No)	Sample values
1					
2					

3					

Sample values for variables could be made-up data at this point in case discussion.

- A listing of all the analysis needed on the variables listed and a justification for each analysis as it relates to "problem to be addressed" as stated in Step 1. At this point, the analysis could be in common language terms instead of statistical terminology. For example, we may say "can we predict the value of variable 1 from variables 2 through 4" instead of stating this as regression analysis.

There are no right, or wrong analyses recommended for addressing the problem stated in Step 1. Hence it is important to evaluate the justification provided by students for the analyses they recommend.

This could be either a group or individual work.

Class discussion (Total time – 100 to 145 minutes)

Good data is critically important for analytics (data driven decision-making or evidence-based decision-making) and hence it is important this discussion is thorough.

- Precision in variable definition and units of measurement (10 – 15 minutes)
- Measurement and variables (15 - 20 minutes)
- Characteristics of variables, qualitative/quantitative, dependent/independent, etc. (15 - 20 minutes)
- Data and normalization (15 – 20 minutes)
- Obtaining data and its cost, cost and benefits of collecting & using data (15 – 20 minutes)
- Variables and possible analyses, i.e., link between data and analysis (15 - 25 minutes)
- Variables, analyses, and types of results (15 - 25 minutes)

One potential solution to this step

We could use the problem definition in Step 1, "Can we predict the average weekly sales of a future store based on local data on characteristics/factors/variables?" as a starting point. We can then identify local factors that may influence/affect "weekly sales." At this point we should be as exhaustive as possible (i.e., divergent in thinking) and on defining each variable precisely. Some variables identified could relate to location characteristics (like how easy/difficult it is to get in and get out, attractiveness, etc.), customer characteristics (who they are, how many, etc.), competition characteristics (what type, where, how many, etc.), etc.

Later, the instructor can bring in a discussion of reducing the number variables for (1) eliminating redundancy, (2) cost of obtaining data compared to its benefit, etc. a bit later (i.e., convergent thinking).

PART B

Step 3

Introduce or assign this case step in a class session that includes a discussion of multiple linear regression. (The case write-up that includes only Steps 1-3 will be handed to students for this class session. In addition, students will be required to bring their answers to Steps 1 and 2 of the case for this session).

Case step 3 leads to the learning outcomes below

After analyzing case step 3 the students will be able to:

3.1 Identify any assumptions about the data or the problem context that need to be made, and state why each assumption is necessary **before** conducting any statistical analysis

3.2 Use a statistical tool to correctly perform a multiple regression with the data provided and use the results to validate the statistical model.

Key questions to address in the discussion of Step 3 (Ramakrishna, Sarkar, and Vijayaraman, 2022):

5. Are we doing the right analysis?
6. Are we doing the analysis right?

Deliverables (by students):

- Performing a correct multiple regression analysis (using a standard statistical tool) with the data provided. This would typically involve the following: (1) Justification for multiple regression analysis (i.e., why is this analysis appropriate for the data provided and for the problem statement from Step 1), (2) Performing the complete regression analysis that includes a variable selection procedure (like Best Subsets, stepwise analysis), a final regression analysis using only the significant variables selected, and, finally, predictions for the 11 "holdout" sample values in the data.
- A report that integrates the results from the analyses performed to interpret the findings. The report should be targeted for management who may not have any background in statistical concepts/analysis.
- Making a case for explicit consideration of geography in refining the data used to obtain better results.

In this step, there are clearly the right and wrong analysis that can be performed. This is the first thing the student deliverable should be evaluated on. Only after this, the report will be evaluated on proper interpretation of the results and its integration with the report. This could be either a group or individual work.

Class discussion (Total time – 100 to 150 minutes)

- Mapping data and problem statement with the right analysis to be performed (with special focus on justifying the analysis proposed), multiple regression analysis (20 - 30 minutes)

- Performing the analysis right – following all the steps necessary and in the right sequence. This includes using the chosen statistical tool right (including data input, and choosing the right options) (30 - 45 minutes)
- Properly interpreting the results of the analysis performed, for the context of the stated problem (Step 1 of the case) and reporting it to the right audience the right way (20 - 30 minutes)
- Need for explicitly recognizing and stating assumptions, and, if necessary, addressing any issues related to violation of some assumptions through re-analysis (15 - 20 minutes)
- Finally, making a case for explicit (additional) consideration of geography as it impacts the data used in the analysis (15 – 25 minutes)

Discussion of non-linear regression may be appropriate here (This is based on the course and judgment of the instructor).

One potential solution to this step

The analysis results for 100 observations in the data set is provided below (Minitab 19 was used for this analysis). The hold-out sample consists of the last 11 observations and the predictions were done for observations 101 through 111. The important parts/points of the results are highlighted, and the instructor should discuss these in the class. Interpretation of the predictions as they pertain to the stated problem (in Step 1 of the case) is critical for this step – addressing the "Prediction interval (PI)" in terms of its precision and accuracy and in terms of its "usefulness" for solving the problem at hand. A sample "executive report" (developed as per the instruction in Appendix A), developed by the instructor, should be discussed in class.

Analysis results: Non-geography-influenced data (CoffeePlus-Step 3)
Best Subsets Regression: Sales versus Traffic, Competitors, Population, AverageInc, MedianInc, Ease
Response is Sales

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Traffic	Competitors	Population	AverageInc	MedianInc	Ease
1	89.6	89.5	89.2	162.0	1517.4	X					
1	15.2	14.3	11.9	2017.7	4343.3			X			
2	94.3	94.2	94.0	47.1	1128.0	X		X			
2	93.4	93.3	93.0	69.9	1215.8	X	X				
3	96.0	95.8	95.3	8.8	958.48	X		X	X		
3	95.8	95.7	95.2	12.2	974.47	X		X		X	
4	96.2	96.1	95.5	3.6	928.43	X	X	X		X	
4	96.2	96.1	95.6	3.8	929.42	X	X	X	X		

5	96.3	96.1	95.5	5.2	931.11	X	X	X		X	X
5	96.3	96.1	95.5	5.3	931.61	X	X	X	X		X
6	96.3	96.0	95.3	7.0	935.22	X	X	X	X	X	X

Regression Analysis: Sales versus Traffic, Competitors, Population, MedianInc

Regression Equation

Sales = -3300 + 1253.8 Traffic + 17.94 Competitors + 0.01941 Population + 0.02737 MedianInc

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-3300	552	-5.98	0.000	
Traffic	1253.8	27.8	45.07	0.000	1.06
Competitors	17.94	5.47	3.28	0.001	2.92
Population	0.01941	0.00407	4.77	0.000	2.85
MedianInc	0.02737	0.00408	6.70	0.000	1.12

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
928.430	96.24%	96.09%	95.54%

Analysis of Variance

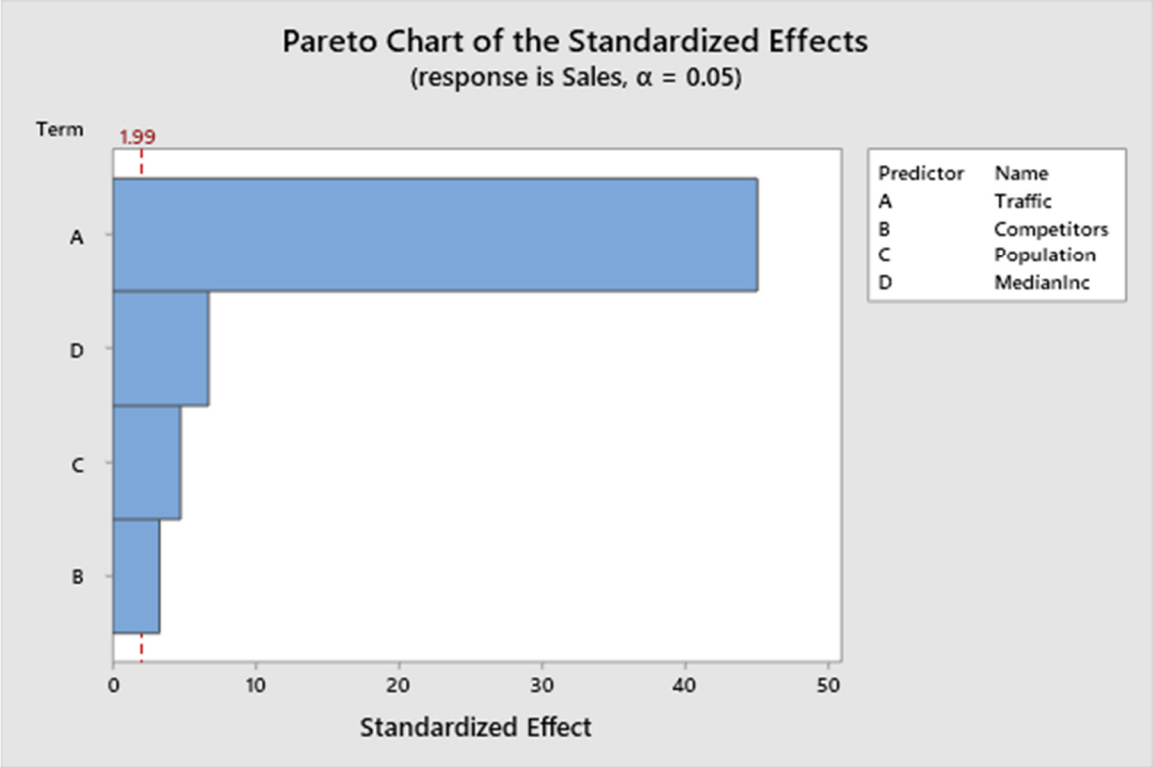
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	4	2097858147	524464537	608.44	0.000
Traffic	1	1750873984	1750873984	2031.22	0.000
Competitors	1	9271564	9271564	10.76	0.001
Population	1	19599908	19599908	22.74	0.000
MedianInc	1	38742227	38742227	44.95	0.000
Error	95	81888391	861983		
Total	99	2179746538			

Fits and Diagnostics for Unusual Observations

Obs	Sales	Fit	Resid	Std Resid		
3	26019	23942	2077	2.29	R	
9	20538	17619	2920	3.21	R	
14	5712	3005	2706	3.36	R	X
80	13673	16630	-2956	-3.49	R	X
87	22788	24789	-2001	-2.22	R	
93	22788	24756	-1967	-2.18	R	

R Large residual

X Unusual X



Prediction for Sales

Regression Equation

Sales = -3300 + 1253.8 Traffic + 17.94 Competitors + 0.01941 Population + 0.02737 MedianInc

Settings

Variable	Setting
Traffic	18
Competitors	55
Population	45169
MedianInc	96057

Prediction

Fit	SE Fit	95% CI	95% PI
23762.1	291.217	(23184.0, 24340.3)	(21830.4, 25693.8)

Settings

Variable	Setting
Traffic	16
Competitors	78
Population	141618
MedianInc	59491

Prediction

Fit	SE Fit	95% CI	95% PI
22538.8	201.561	(22138.6, 22938.9)	(20652.7, 24424.9)

Settings

Variable	Setting
Traffic	8

Competitors	14
Population	54786
MedianInc	47579

Prediction

Fit	SE Fit	95% CI	95% PI
9347.96	255.181	(8841.36, 9854.56)	(7436.44, 11259.5)

PART C

Step 4

Introduce or assign this case step in a class session where a discussion of the impact of geography on analytics takes place. (The complete case write-up that includes all steps, Steps 1-4, will be handed to students for this class session). This will lead to the following learning outcomes:

After analyzing the case step assigned/discussed, the students will be able to:

4.1 Correctly perform a multiple regression with the new data provided and use the results to validate the statistical model using a statistical tool (possibly the same one used in step 3)

4.2 Select and recommend the predictive model (between the two developed, in Case Steps 3 and 4) and justify their choice.

4.3 Write an executive summary in plain English and link the report with all necessary supporting analytics work done, as listed below (Ramakrishna, Sarkar, and Vijayaraman, 2022).

7. Do we have the right results?
8. Are we interpreting the results right?

Deliverables (by students):

- Performing a correct multiple regression analysis (using a standard statistical tool) with the data provided. This would typically involve the following: (1) Justification for multiple regression analysis (i.e., why is this analysis appropriate for the data provided and for the problem statement from Step 1), (2) Performing the complete regression analysis that includes a variable selection procedure (like Best Subsets, stepwise analysis), a final regression analysis using only the variables selected, and, finally, predictions for the 11 "holdout" sample values in the data.
- A report that integrates the results from the analyses performed in Step 3 and Step 4 to interpret the findings. The report should be targeted for management who may not have any background in statistical concepts/analysis.
- Making an evaluation as to the impact of geography-influenced data used in Step 4 on the final results.

In this step, there are clearly the right and wrong analysis that can be performed. This is the first thing the student deliverable should be evaluated on. Only after this, the report will be evaluated on proper interpretation of the results and its integration with the report. This could be either a group or individual work.

Class discussion (Total time – 100 to 150 minutes)

As the discussion here is, for the most part, almost the same as the discussion in Step 3 the time that needs to be spent is a bit less.

- Mapping data and problem statement with the ***right analysis*** to be performed (with special focus on justifying the analysis proposed), multiple regression analysis (10 - 15 minutes)
- Performing the ***analysis right*** – following all the steps necessary and in the right sequence. This includes using the chosen statistical tool right (including data input, and choosing the right options) (15 - 25 minutes)
- Properly interpreting the results of the analysis performed, for the context of the stated problem (Step 1 of the case) and reporting it to the right audience the right way (15 - 20 minutes)

- Need for explicitly recognizing and stating assumptions, and, if necessary, addressing any issues related to violation of some assumptions through re-analysis (10 - 15 minutes)
- Finally, making a case for explicit (additional) consideration of ***geography*** as it impacts the results of the analysis. This involves comparing the results of Step 3 and Step 4 as they pertain to the problem statement in Step 1 (15 – 25 minutes)

One potential solution to this step

The analysis results for 100 observations in the data set is provided below (Minitab 19 was used for this analysis). The hold-out sample consists of the last 11 observations and the predictions were done for observations 101 through 111. The important parts/points of the results are highlighted, and the instructor should discuss these in the class. Interpretation of the predictions as they pertain to the stated problem (in Step 1 of the case) is critical for this step – addressing the "Prediction interval (PI)" in terms of its precision and accuracy and in terms of its "usefulness" for solving the problem at hand. A sample "executive report" (developed as per the instruction in Appendix A), developed by the instructor, should be discussed in class.

In addition, an explicit discussion evaluating the results from Step 3 and Step 4 is necessary here (as we conclude the case discussion). As we can see from the results, geography did impact the results positively (i.e., the usefulness of the results is better from Step 4) – the prediction intervals (PIs) are smaller (i.e., about $2,000 in Step4 as compared to about $4,000 in Step3) and hence a bit more valuable. For analysis results for the holdout sample of 11 stores' data, please see the table titled "Evaluating the two models -- without and with geography-influenced data" at the end of the analysis section for some measures that can be used in evaluating the two models (highlighted values – a. R2 (results from part B and C); b. variability in "actual sales minus predicted sales" – range and standard deviation; and c. variability in "Prediction interval as a % of point estimate of predicted sales."

A final concluding discussion that discusses all the four steps of the case and the relationships/integration between them would enhance the value of the case. Also, a brief discussion of the two final questions of the data-driven decision-making model would be appropriate here – "do we have the right results" and "are we interpreting the results right" (see Ramakrishna, Sarkar, and Vijayaraman, 2022).

Analysis results: Geography-Influenced data (CoffeePlus-Step 4)
Best Subsets Regression: Sales versus Traffic, Competitors, Population, AverageInc, MedianInc, Ease
Response is Sales

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	Traffic	Competitors	Population	AverageInc	MedianInc	Ease
1	89.6	89.5	89.2	863.4	1517.4	X					
1	14.3	13.5	10.4	7842.5	4364.7		X				
2	98.3	98.2	98.1	66.5	623.74	X			X		
2	98.1	98.0	97.9	86.3	661.22	X				X	
3	98.7	98.7	98.6	25.9	537.42	X	X			X	
3	98.4	98.4	98.2	54.7	599.40	X	X		X		
4	98.9	98.9	98.7	10.9	499.72	X	X	X		X	
4	98.7	98.7	98.5	27.8	540.03	X	X			X	X
5	99.0	98.9	98.8	5.2	482.77	X	X	X	X	X	
5	98.9	98.9	98.7	12.8	502.29	X	X	X		X	X
6	99.0	98.9	98.8	7.0	484.95	X	X	X	X	X	X

COFFEEPLUS-REGULAR-MINITAB.MWX

Regression Analysis: Sales versus Traffic, Competitors, Population, AverageInc, MedianInc

Regression Equation

Sales = -6689 + 1259.5 Traffic + 14.36 Competitors - 0.002387 Population
- 0.0481 AverageInc + 0.1632 MedianInc

Coefficients

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	-6689	524	-12.77	0.000	
Traffic	1259.5	14.6	86.37	0.000	1.08
Competitors	14.36	1.95	7.38	0.000	3.30
Population	-0.002387	0.000478	-5.00	0.000	1.68
AverageInc	-0.0481	0.0173	-2.79	0.006	33.17
MedianInc	0.1632	0.0223	7.31	0.000	26.09

Model Summary

S	R-sq	R-sq(adj)	R-sq(pred)
482.770	98.99%	98.94%	98.80%

Analysis of Variance

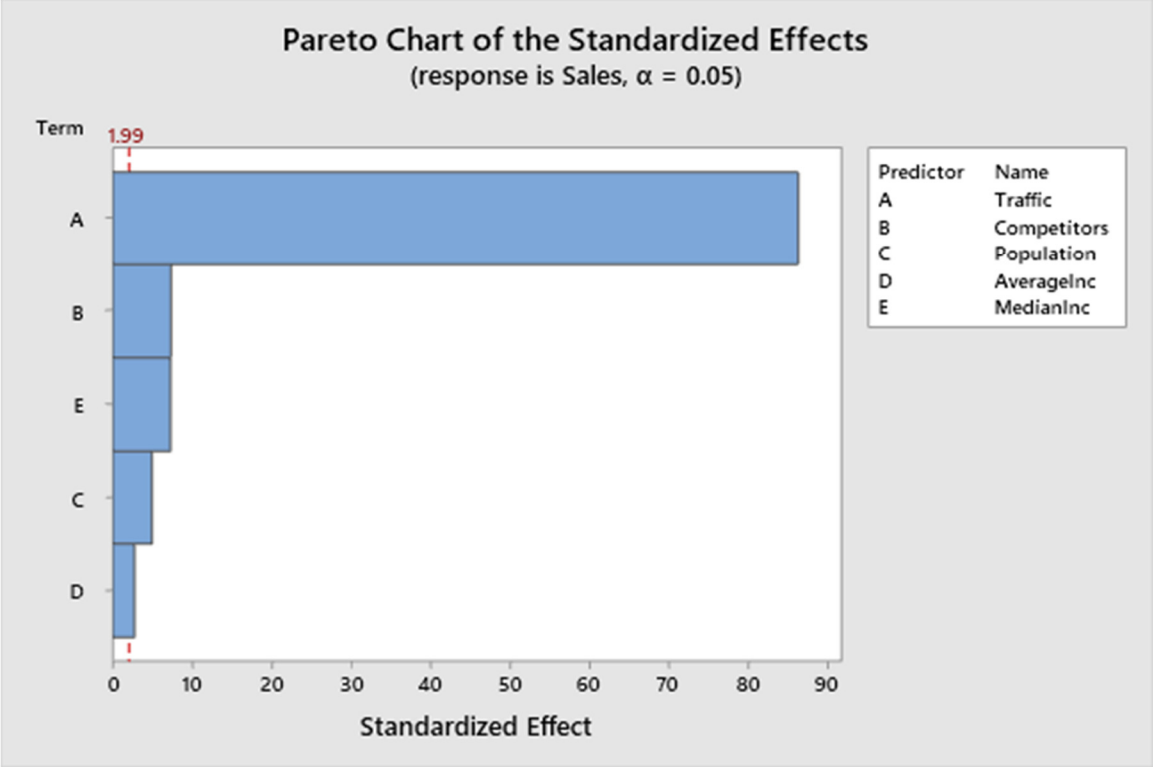
Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	5	2157838213	431567643	1851.69	0.000

Traffic	1	1738577692	1738577692	7459.55	0.000
Competitors	1	12703041	12703041	54.50	0.000
Population	1	5817509	5817509	24.96	0.000
AverageInc	1	1815272	1815272	7.79	0.006
MedianInc	1	12457209	12457209	53.45	0.000
Error	94	21908326	233067		
Total	99	2179746538			

Fits and Diagnostics for Unusual Observations

Obs	Sales	Fit	Resid	Std Resid	
54	21904	20960	944	2.02	R
65	3808	2555	1253	2.78	R
66	27385	26393	992	2.11	R
82	9596	10539	-943	-2.13	R
93	22788	23777	-989	-2.15	R

R Large residual



COFFEEPLUS-REGULAR-MINITAB.MWX

Prediction for Sales

Regression Equation

Sales = -6689 + 1259.5 Traffic + 14.36 Competitors - 0.002387 Population
- 0.0481 AverageInc + 0.1632 MedianInc

Settings

Variable	Setting
Traffic	18
Competitors	131
Population	338545
AverageInc	122707
MedianInc	79710

Prediction

Fit	SE Fit	95% CI	95% PI
24156.0	103.164	(23951.2, 24360.9)	(23175.8, 25136.2)

Settings

Variable	Setting
Traffic	16
Competitors	149
Population	428672
AverageInc	128837
MedianInc	83670

Prediction

Fit	SE Fit	95% CI	95% PI
22031.6	78.8700	(21875.0, 22188.2)	(21060.4, 23002.9)

Settings

Variable	Setting
Traffic	8
Competitors	61
Population	428422
AverageInc	99552
MedianInc	62758

Prediction

Fit	SE Fit	95% CI	95% PI
8689.54	132.128	(8427.20, 8951.89)	(7695.74, 9683.35)

Evaluating the two models -- without and with geography-influenced data

Without geography-influenced data

Diff.	*Actual*	*Predicted*	*Prediction interval (95% conf.)*		*Prediction interval as a % of the point estimate of the predicted value*
872.9	24635	23762.1	21830.4	25693.8	16%
-634.8	21904	22538.8	20652.7	24424.9	17%
-232.96	9115	9347.96	7436.44	11259.5	41%
812.4	21904	21091.6	19233.4	22949.8	18%
173.6	23269	23095.4	21201.9	24988.9	16%
-366.8	20538	20904.8	18929.1	22880.5	19%
-407.5	20538	20945.5	18996.9	22894.1	19%
-1034.1	17096	18130.1	16263.8	19996.4	21%
-1114.75	6846	7960.75	5944.52	9976.98	51%
1196.6	27385	26188.4	24250.5	28126.3	15%
1793.7	20538	18744.3	16881.7	20606.9	20%

Range 2908.45
St. Dev. 952.33

With geography-influenced data

Diff.	*Actual*	*Predicted*	*Prediction interval (95% conf.)*		*Prediction interval as a % of the point Estimate of the predicted value*
479	24635	24156	23175.8	25136.2	8%
-127.6	21904	22031.6	21060.4	23002.9	9%
425.5	9115	8689.5	7695.7	9683.3	23%
345.5	21904	21558.5	20569.7	22547.3	9%
323.4	23269	22945.6	21973.7	23917.4	8%
-674.7	20538	21212.7	20239.7	22185.7	9%
-612.3	20538	21150.3	20178.7	22122	9%
-219.9	17096	17315.9	16309.3	18322.4	12%
-1696.5	6846	8542.5	7521.4	9563.5	24%
810.4	27385	26574.6	25552.2	27597.1	8%
484.3	20538	20053.7	19035.2	21072.3	10%

Range 2506.9
St. Dev. 726.58

REFERENCES

Ramakrishna, H, Sarkar, A, and Vijayaraman, B. (2022) Development of an Introductory MBA Course in Business Analytics Using Data-Driven Decision-Making (DDDM) Model, Journal of Higher Education Theory and Practice, Vol. 22(12) 2022

Appendix A
A sample guide for writing "Executive summaries"
Please use the instructions in this section to write executive summaries for the case. The case presents a business scenario which you will address via statistical methods applied to data. You will use Excel and/or any of the recommended statistical software (like Minitab, Minitab Express, PHStat, or MegaStat), and/or Tableau, as necessary, to perform descriptive and/or predictive analytics.

A case begins with the presentation of a business scenario followed by the description of data. The case provides varying amounts of guidance as to what to do. Your task is to consider the situation, analyze the data, and report on your findings.

Executive Summary of Case

Reports involving *data analysis cases* often follow a standard format.

Organization: A Template

Below is a recommended configuration for your executive summary. Italics are intended only as commentary—avoid using italics in your own report. Details follow the template.

To: A specific person is often mentioned in the case. If so, use that person as your audience. Otherwise, improvise as appropriate to the case.
From: Improvise a name here (Do not use real names of students in the class.)
Subject: In 3-7 words briefly state the topic
Date:

A brief statement of the problem goes here. Use no section heading for the problem statement.

EXECUTIVE SUMMARY

Major Findings
Provide here a concise statement of your major findings. What did you discover as a result of your data analysis as it relates to the problem statement and tasks assigned to you?
Recommendations for Action
Given your major findings above, state what recommendations for action follow. This section may be optional depending on the specific guidance given with the case
Analytical Overview
This is where you briefly describe in nontechnical language the approach or method you used in your analysis.

Problem Statement

Begin with a concise restatement of the problem. Do not re-explain the business problem in great detail; assume that the reader is aware of the task you faced. For example, you might write, "This summary reports on the link between employee absenteeism and the

new on-site health care facilities." Do not use a section heading for the problem statement. Strive for one sentence; do not exceed two sentences.

Major Finding(s)

For this section (ideally with a bold-font section heading, not underlined or italicized and the same for all other subsection headings), briefly state the major findings of your data analysis in nontechnical language *as they relate to the problem statement*. For example, you might write, "Average monthly employee absenteeism declined by 3 hours per month after the institution of the on-site wellness program." Alternatively, this might be the place to provide a nontechnical interpretation of regression coefficients. Or you might indicate that temperature is statistically related to the firm's revenues (and by how much) whereas customer satisfaction is not.

Recommendation(s) for Action

- The need for a recommendations section depends on the tasks assigned to you. For example, if you were asked which employees should be fired, the answer would be in the recommendations section. If, on the other hand, answers to all of your tasks are addressed in the Major Findings section, then there is no need for a Recommendations section.
- Recommendations, as needed, must relate specifically to the problem statement and major findings.
- Recommendations, as needed, must be specific and actionable, e.g., "Continue funding the on-site wellness center," or "To achieve the goal of decreasing employee turnover, focus on our employee benefits package."

Analytical Overview

This is where you briefly describe your analytical approach. Using only *nontechnical language*, provide a *succinct* overview of the procedure(s) you used. For example, you might write, "The data were sorted into two groups: before and after the implementation of the tax cuts. Averages were then computed for each group."

Features of an Effective Executive Summary

Here are some recommendations for preparing an executive summary that has a crisp, professional appearance.

1. Executive summaries must be focused and concise. Imagine that you are writing to a top executive who wants only to know how to act as a result of your analysis. In this course, brevity will be enforced via a word count:
    - Your word count cannot exceed 300 words (approximately one page of text). In most cases, you can and should strive for fewer.
    - The word count includes all words beginning with the problem statement through the last word in the "Analytical Overview" section.
    - To perform a word count, first highlight (select) the relevant material. Then, depending on your version of Word, the word count may be displayed on the lower left-hand corner of your screen, or you may need to use Tools > Word Count.

2. Consider the information needs of your intended audience and avoid technical jargon. If you need to explain a statistical concept, do so without statistical verbiage (include statistical verbiage in documentation pages). The only exception to the no-technical-jargon rule is when you use the phrase "statistically significant" in your executive summary (if used appropriately and with supporting documentation), since most people have a sense that it implies achievement of an important statistical hurdle.
3. Make only data-driven statements. Report only on the results of the data analysis. Make no statements, no matter how logical or obvious, unless they can be traced directly to a statistical result from your analysis.
4. Executive summaries must contain no spelling, punctuation, grammatical, or other presentation errors.
5. Use bulleted lists as needed, but sparingly. Never use a bullet when there is only one item in the list.
6. Document every statistical claim or number in your executive summary. Documentation involves a reference to an Excel, Minitab, Minitab Express, PHStat, MegaStat or Tableau statistical exhibit. For example, you might write, "Sales are higher by $52k in December than January (see Exhibit A.)" The exhibit may be a reference to an exhibit in the documentation page of your report (also described below).
7. Round numbers to an appropriate number of decimals.
8. Include specific numerical information whenever possible. Our hypothetical reader is a business professional comfortable working with numbers. S/he needs numbers to inform his/her decision-making. For example, suppose you write: "...there is a statistically significant difference in the average salaries of men and women." This is of almost no value to the Director of Human Resources since s/he will react quite differently depending on whether that difference is $12 or $20,000. Instead, write: "there is a difference of $20,000 in the average salary of men and women."
9. Plagiarism is prohibited. Individual assignments must be completed individually. Be sure that anything obtained from any outside source has proper attribution.

Place the exhibits needed to document your statistical analysis in a Documentation page following the executive summary.

10. Remember every statistical number or claim must be documented. Begin the second page of your report with this title: "Documentation Page." On this page, include exhibits needed to document either:
    - Excel, Minitab, Minitab Express, PHStat, or MegaStat displays that are too technical for the main body of the executive summary, or
    - items of secondary importance to the main body of the executive summary.

They can be copied and pasted from Excel, Minitab, Minitab Express, PHStat, MegaStat, or Tableau. Here are some guidelines for the documentation page.

11. Use consecutive numbering of exhibits beginning with the first exhibit in the executive summary (if any) and ending with the last one on the documentations page. Use either 1, 2, 3... or A, B, C... Because ease-of-use is the goal, label everything as an "exhibit" (as opposed to, for example, Exhibit A, Table A, Exhibit B, Graph A, Table B...).

12. Do not include Excel, Minitab, or Tableau exhibits in the documentation page unless there is a specific reference to it in the main body of the executive summary. Other than exhibit titles, no commentary is allowed on the documentation page.
13. Crop and size exhibits on the documentation page to ensure readability.