UNIVERSITY OF MUMBAI

DEPARTMENT OF COMPUTER SCIENCE

M.Sc. Computer Science with Spl. in Data Science – Semester IV

MARKETING ANALYSIS

JOURNAL

2022-2023

Seat No. _____

मुंबई विद्यापीठ
University of Mumbai
Re-accredited with A++ Grade
(CGPA 3.65) by NAAC (3rd Cycle 2021)

UNIVERSITY OF MUMBAI

DEPARTMENT OF COMPUTER SCIENCE

CERTIFICATE

This is to certify that the work entered in this journal was done in the University Department of Computer Science laboratory by

Mr./Ms._____ Seat No. _____

for the course of M.Sc. Computer Science with Spl. in Data Science - Semester IV (CBCS) (Revised) during the academic year 2022-2023 in a satisfactory manner.

_____ _____

Subject In-charge **Head of Department**

_____

External Examiner

Index

Sr. no.	Name of the practical	Page No.	Date	Sign
1	Slicing and Dicing, Using Excel Charts and Functions to Summarize Marketing Data	4-13		
2	Estimating demand curves and using solver to optimize price	14-17		
3	Regression: Simple and Multiple	18-25		
4	Forecasting	26-37		
5	Conjoint Analysis & Logistic Regression	38-46		
6	Calculating Lifetime Customer value	47-49		
7	Market Segmentation	50-65		
8	Forecasting New Product sales	66-75		
9	Retailing and Advertising	76-78		
10	Research Methodology Tools	79-86		

Practical – 1

Slicing and Dicing Marketing Data with PivotTable and Summarizing with Charts

1. The Makeup2007.xlsx file gives sales data for a small makeup company. Each row lists the salesperson, product sold, location of the sale, units sold, and revenue generated. Use this file to perform the following exercises:

a. Summarize the total revenue and units sold by each person of each product

Step 1: Insert Pivot Table.

Step 2: Drag the Name & Product fields into Rows field.

Step 3: Drag the Units and Dollars into Value filed.

		Sum of Units	Sum of Dollars
Ashley	eye liner	1920	5844.948744
	foundation	1373	4186.058628
	lip gloss	1985	6053.684565
	lipstick	1066	3245.442978
	mascara	2172	6617.100349
Ashley Total		**8516**	**25947.23526**
Betsy	eye liner	1987	6046.534282
	foundation	2649	8043.486462
	lip gloss	1855	5675.650045
	lipstick	1305	3968.605496
	mascara	1582	4827.253996
Betsy Total		**9378**	**28561.53028**
Cici	eye liner	1960	5982.823291
	foundation	2031	6198.248632
	lip gloss	1701	5199.949201
	lipstick	1035	3148.84065
	mascara	2317	7060.711397
Cici Total		**9044**	**27590.57317**
Colleen	eye liner	1107	3389.625314
	foundation	2242	6834.767608
	lip gloss	1831	5573.323725
	lipstick	765	2346.413777
	mascara	2215	6746.525368
Colleen Total		**8160**	**24890.65579**
Cristina	eye liner	1770	5397.273636
	foundation	1729	5290.989935

	lip gloss	1734	5297.97981
	lipstick	788	2401.668343
	mascara	1790	5461.646997
Cristina Total		**7811**	**23849.55872**
Emilee	eye liner	2490	7587.38898
	foundation	1744	5313.787561
	lip gloss	1725	5270.250313
	lipstick	720	2189.137568
	mascara	1545	4719.299731
Emilee Total		**8224**	**25079.86415**
Hallagan	eye liner	2288	6964.621074
	foundation	2298	6985.734333
	lip gloss	1840	5603.119378
	lipstick	1045	3177.871325
	mascara	1873	5703.34667
Hallagan Total		**9344**	**28434.69278**
Jen	eye liner	2302	7010.440514
	foundation	1844	5628.648036
	lip gloss	1792	5461.61479
	lipstick	1299	3953.300132
	mascara	2265	6887.17495
Jen Total		**9502**	**28941.17842**
Zaret	eye liner	2681	8166.749063
	foundation	2117	6451.650057
	lip gloss	1862	5670.329329
	lipstick	800	2448.707163
	mascara	1268	3879.949944
Zaret Total		**8728**	**26617.38556**
Grand Total		**78707**	**239912.6741**

b. Summarize the percentage of each person's sales that came from each location. Create a PivotChart to summarize this information.

Step 1: Insert Pivot Table.

Step 2: Drag the Name & Location into Rows field.

Step 3: Drag the Dollars column into Value filed and select % of Column Total from Value Field Setting

Row Labels	Percentage of Dollars
Ashley	**10.82%**
east	3.24%
midwest	2.08%

south	3.08%
west	2.41%
Betsy	**11.90%**
east	3.65%
midwest	2.03%
south	3.22%
west	2.99%
Cici	**11.50%**
east	2.48%
midwest	3.39%
south	2.99%
west	2.64%
Colleen	**10.37%**
east	2.38%
midwest	2.75%
south	3.25%
west	2.00%
Cristina	**9.94%**
east	1.72%
midwest	2.45%
south	2.49%
west	3.29%
Emilee	**10.45%**
east	2.62%
midwest	2.35%
south	2.52%
west	2.96%
Hallagan	**11.85%**
east	2.07%
midwest	3.08%
south	3.42%
west	3.28%
Jen	**12.06%**
east	2.90%
midwest	2.66%
south	2.97%
west	3.54%
Zaret	**11.09%**
east	2.85%
midwest	2.06%
south	2.86%
west	3.32%
Grand Total	**100.00%**

c. Summarize each girl's sales by location and use the Report Filter to change the calculations to include any subset of products.

Step 1: Insert Pivot Table.

Step 2: Drag the Name in Rows filed & Location into Columns field.

Step 3: Drag the Units column into Values and drag the Product to Filter field.

Product	(All)

Sum of Dollars	**Column Labels**				
Row Labels	**east**	**midwest**	**south**	**west**	**Grand Total**
Ashley	7772.704761	4985.896509	7398.565792	5790.068203	25947.23526
Betsy	8767.431725	4878.085848	7732.05698	7183.955727	28561.53028
Cici	5956.320446	8129.619289	7174.448975	6330.184462	27590.57317
Colleen	5713.069445	6586.142169	7785.632708	4805.811471	24890.65579
Cristina	4126.268644	5870.034488	5964.158473	7889.097115	23849.55872
Emilee	6295.472056	5642.196163	6050.594346	7091.601589	25079.86415
Hallagan	4965.615813	7378.321391	8210.814251	7879.941325	28434.69278
Jen	6949.209483	6381.320681	7116.016774	8494.631484	28941.17842
Zaret	6825.995148	4953.797616	6864.065862	7973.52693	26617.38556
Grand Total	**57372.08752**	**54805.41416**	**64296.35416**	**63438.81831**	**239912.6741**

2. The Station.xlsx file contains data for each family including the family size (large or small), income (high or low), and whether the family bought a station wagon. Use this file to perform the following exercises:

a. Does it appear that family size or income is a more important determinant of station wagon purchases?

Step 1: Generate correlation of family size & station wagon purchases.

Step 2 : Generate correlation of income & station wagon purchases.

Step 3: Compare both of them.

Conclusion : family size has higher impact on station wagon purchases, as correlation is 0.68

Family Size & Station Wagon	0.682248
Salary & Station Wagon	0.024236

b. Compute the percentage of station wagon purchasers that are high or low income.

Step 1: Insert Pivot Table.

Step 2: Drag the Salary column to Rows value field.

Step 3: Drag Station Wagon? to Values field. Then select drop down and select the Value field setting and summarize value field by Count. Show values as % of Column Total.

Salary	Percentage of Station Wagon?
High	71.72%
Low	28.28%
Grand Total	**100.00%**

c. Compute the percentage of station wagon purchasers that come from each of the following four categories: High Income Large Family, High Income Small Family, Low Income Large Family, and Low Income Small Family.

Step 1: Insert Pivot Table.

Step 2: Drag the Salary & Family Size to Rows field.

Step 3: Drag Station Wagon? to Values field. Then select drop down and select the Value field setting and summarize value field by Count. Show values as % of Grand Total

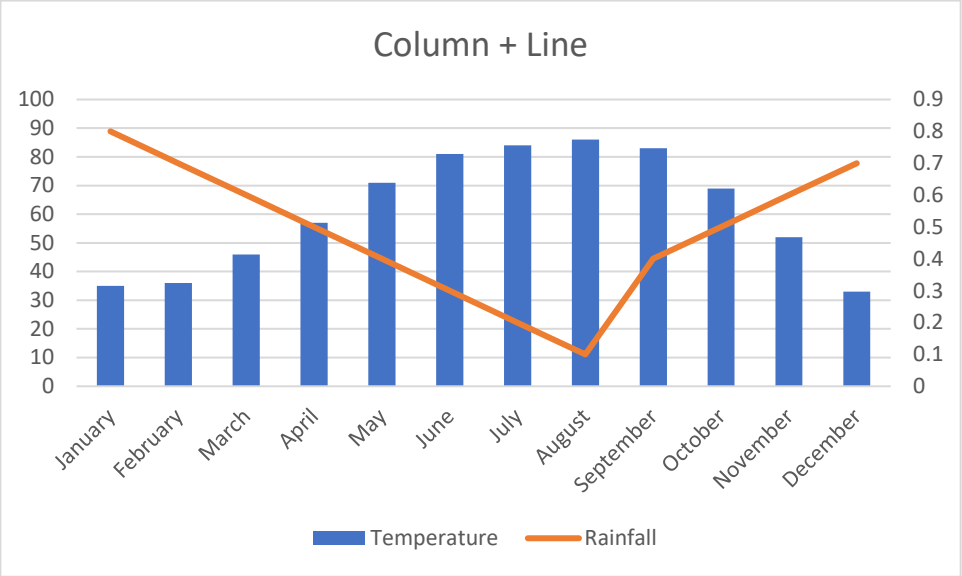
Row Labels	Count of Station Wagon?
High	**71.72%**
Large	39.07%
Small	32.65%
Low	**28.28%**
Large	15.16%
Small	13.12%
Grand Total	**100.00%**

3. The Weather worksheet includes monthly average temperature and rainfall in Bloomington, Indiana. Create a combination chart involving a column and line graph with a secondary axis to chart the temperature and rainfall data.

Step 1: Insert Chart.

Step 2: Right click on one of the data series and select change data series type.

Step 3: Select the Column and line.

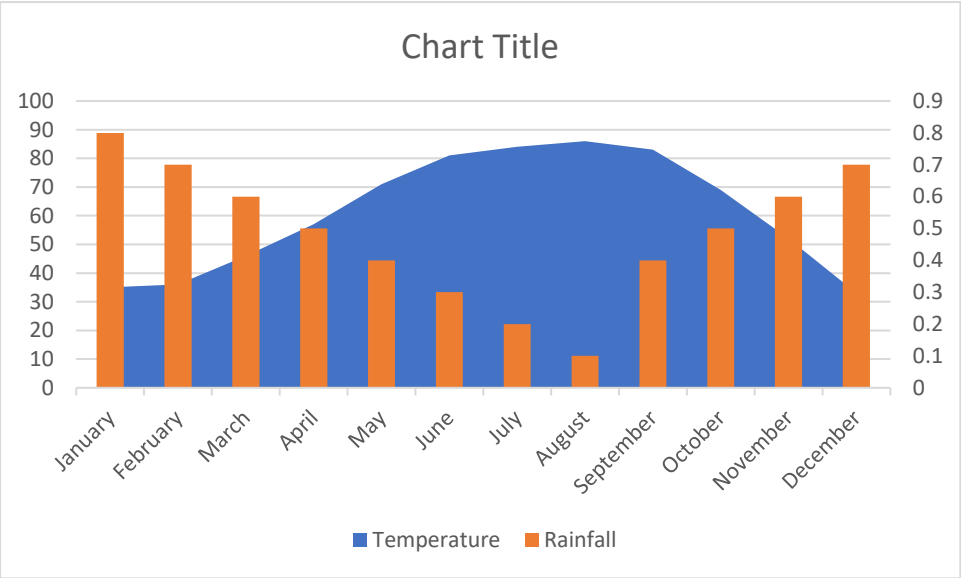


4. The Weather worksheet includes monthly average temperature and rainfall in Bloomington, Indiana. Create a combination chart involving a column and area graph with a secondary axis to chart the temperature and rainfall data.

Step 1: Insert Chart.

Step 2: Right click on one of the data series and select change data series type.

Step 3: Select the Column and area



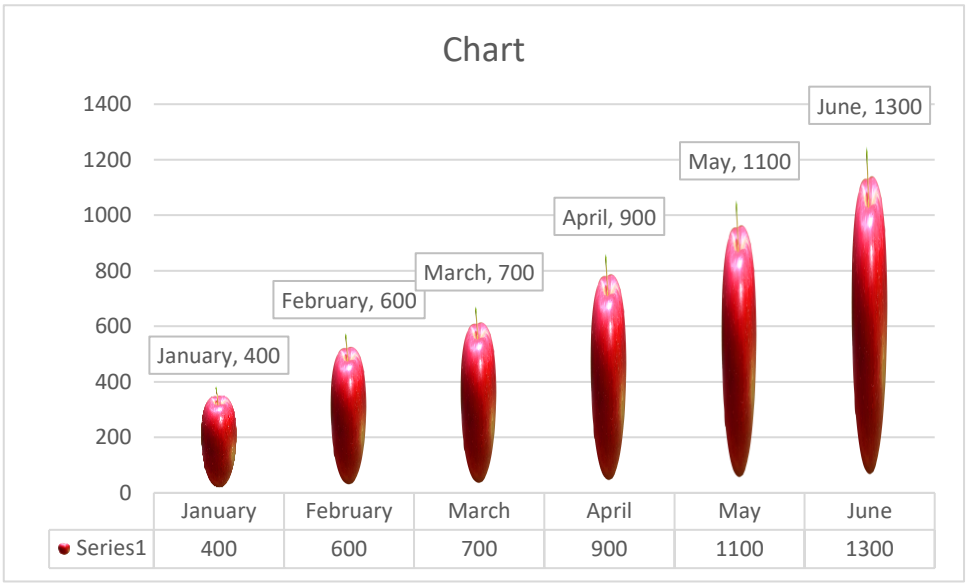
5. The Pictures and Labels worksheet includes monthly tomato sales on Farmer Smith's farm. Summarize this data with pictures of tomatoes, data labels, and a data table.

Step 1: Insert Chart. Select Column chart.

Step 2: Click on the + symbol present beside the graph and select data table and data labels.

Step 3: In data labels select more options, then select category and values. Then select Separator as new line.

Step 4: In Fill, select Picture or texture fill and browse an image and insert

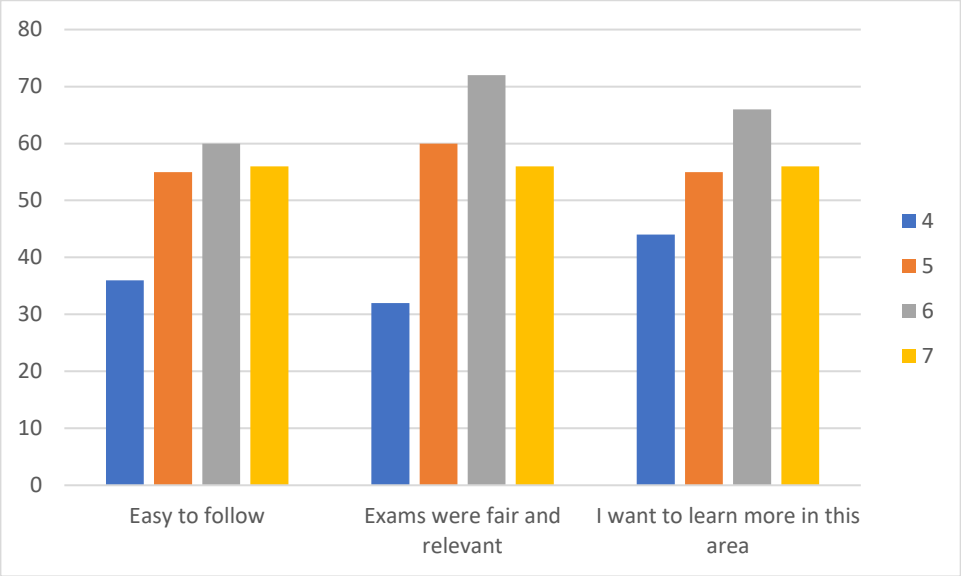


6. The Survey worksheet contains results evaluating a training seminar for salespeople. Use a PivotChart to summarize the evaluation data.

Step 1: Insert Pivot table.

Step 2: Drag the Question to Rows value field and Response to Columns and Values.

Step 3: Click on any cell in pivot table. Insert pivot chart.



7. The data in the check-boxes worksheet contains monthly sales during 2010and2011. Use check boxes to set up a chart in which the user can choose which series are charted.

Step 1: To place a check box in a worksheet, select the Developer tab and chose Insert.

Step 2: From the Form Controls (Not ActiveX) select the check box that is third from the left in the top row.

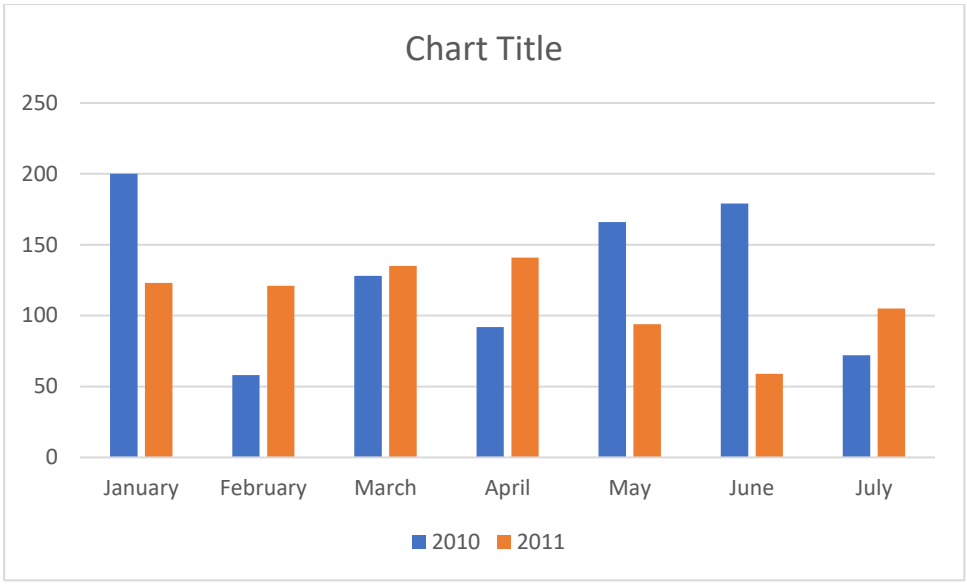
Step 3: Hold down the left mouse to size the check box. Make checkboxes for years 2010 and 2011.

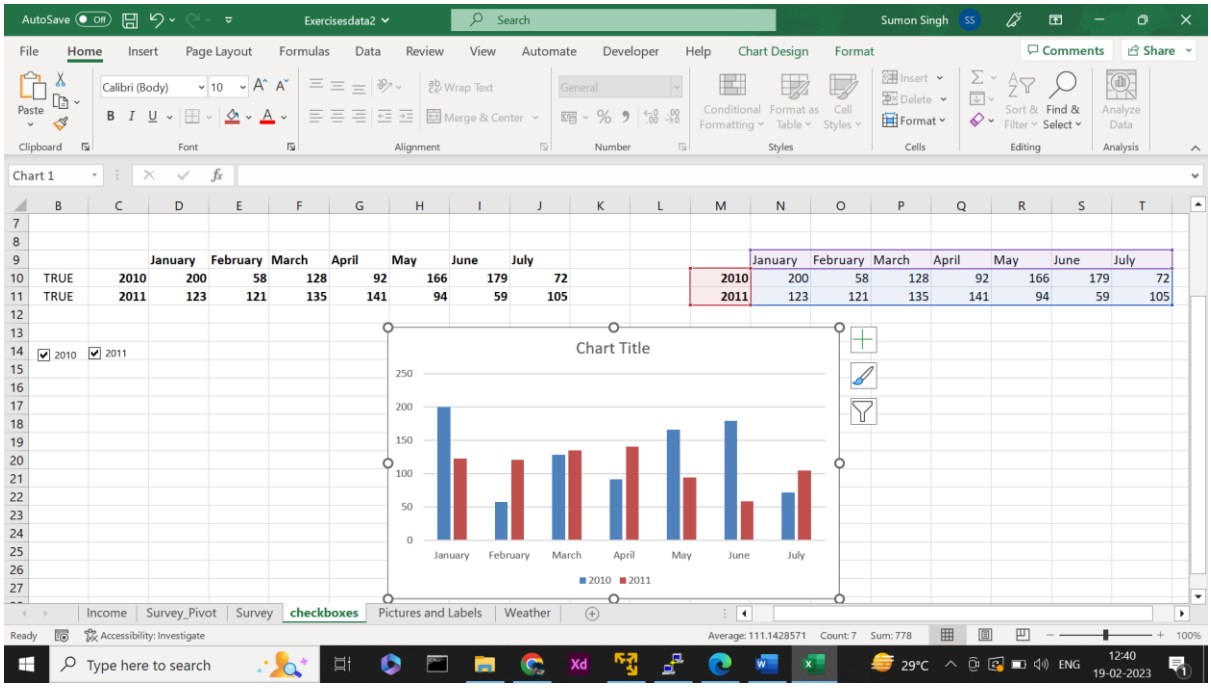
Step 4: Next, select the cell controlled by the check box by moving the cursor to the checkbox until you see the Pointer (a hand with a pointing finger).

Step 5: Right-click, select the Format Control dialog box, and select cell B10 as the cell link for year 2010. Click OK. Similarly, do for year 2011. 0 10 20 30 40 50 60 70 80 Easy to follow Exams were fair and relevant I want to learn more in this area 4 5 6 7

Step 6: To create the table for data to be charted, copy the column and row names. For values insert formula =IF($B$10,D10,NA()) for year 2010 and column January. Do the same for remaining columns and year 2011.

Step 7: Select the charted data table and insert chart.



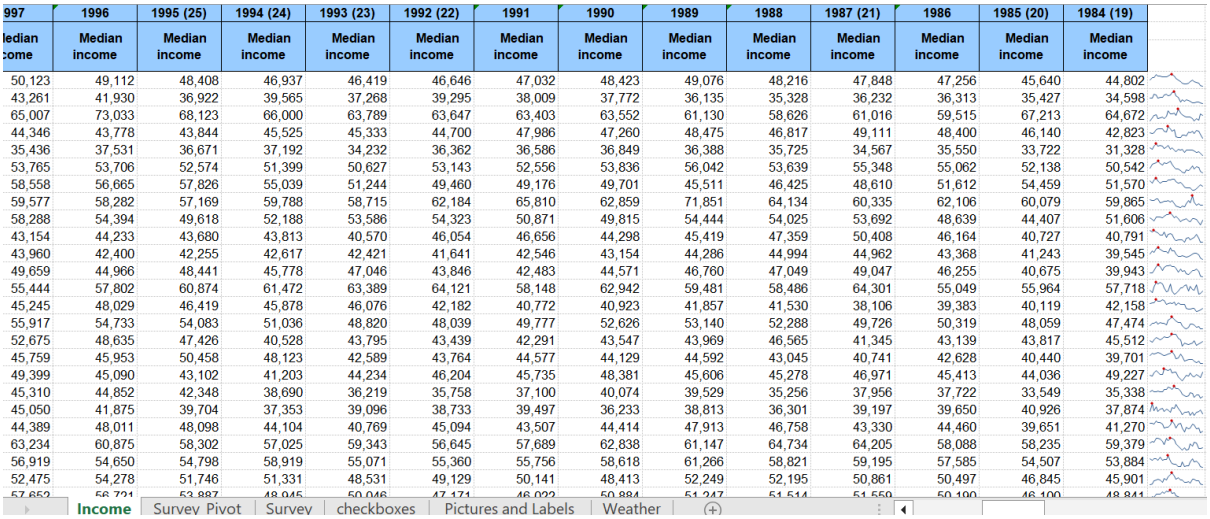


8. The Income worksheet contains annual data on median income in each state for the years 1984–2010. Use Sparklines to summarize this data.

Step 1: Select the numerical data.

Step 2: Insert Sparkline.

Step 3: Give the output location range as AD6:AD57



997	1996	1995 (25)	1994 (24)	1993 (23)	1992 (22)	1991	1990	1989	1988	1987 (21)	1986	1985 (20)	1984 (19)
edian come	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income	Median income
50,123	49,112	48,408	46,937	46,419	46,646	47,032	48,423	49,076	48,216	47,848	47,256	45,640	44,802
43,261	41,930	36,922	39,565	37,268	39,295	38,009	37,772	36,135	35,328	36,232	36,313	35,427	34,598
65,007	73,033	68,123	66,000	63,789	63,647	63,403	63,552	61,130	58,626	61,016	59,515	67,213	64,672
44,346	43,778	43,844	45,525	45,333	44,700	47,986	47,260	48,475	46,817	49,111	48,400	46,140	42,823
35,436	37,531	36,671	37,192	34,232	36,362	36,586	36,849	36,388	35,725	34,567	35,550	33,722	31,328
53,765	53,706	52,574	51,399	50,627	53,143	52,556	53,836	56,042	53,639	55,348	55,062	52,138	50,542
58,558	56,665	57,826	55,039	51,244	49,460	49,176	49,701	45,511	46,425	48,610	51,612	54,459	51,570
59,577	58,282	57,169	59,788	58,715	62,184	65,810	62,859	71,851	64,134	60,335	62,106	60,079	59,865
58,288	54,394	49,618	52,188	53,586	54,323	50,871	49,815	54,444	54,025	53,692	48,639	44,407	51,606
43,154	44,233	43,680	43,813	40,570	46,054	46,656	44,298	45,419	47,359	50,408	46,164	40,727	40,791
43,960	42,400	42,255	42,617	42,421	41,641	42,546	43,154	44,286	44,994	44,962	43,368	41,243	39,545
49,659	44,966	48,441	45,778	47,046	43,846	42,483	44,571	46,760	47,049	49,047	46,255	40,675	39,943
55,444	57,802	60,874	61,472	63,389	64,121	58,148	62,942	59,481	58,486	64,301	55,049	55,964	57,718
45,245	48,029	46,419	45,878	46,076	42,182	40,772	40,923	41,857	41,530	38,106	39,383	40,119	42,158
55,917	54,733	54,083	51,036	48,820	48,039	49,777	52,626	53,140	52,288	49,726	50,319	48,059	47,474
52,675	48,635	47,426	40,528	43,795	43,439	42,291	43,547	43,969	46,565	41,345	43,139	43,817	45,512
45,759	45,953	50,458	48,123	42,589	43,764	44,577	44,129	44,592	43,045	40,741	42,628	40,440	39,701
49,399	45,090	43,102	41,203	44,234	46,204	45,735	48,381	45,606	45,278	46,971	45,413	44,036	49,227
45,310	44,852	42,348	38,690	36,219	35,758	37,100	40,074	39,529	35,256	37,956	37,722	33,549	35,338
45,050	41,875	39,704	37,353	39,096	38,733	39,497	36,233	38,813	36,301	39,197	39,650	40,926	37,874
44,389	48,011	48,098	44,104	40,769	45,094	43,507	44,414	47,913	46,758	43,330	44,460	39,651	41,270
63,234	60,875	58,302	57,025	59,343	56,645	57,689	62,838	61,147	64,734	64,205	58,088	58,235	59,379
56,919	54,650	54,798	58,919	55,071	55,360	55,756	58,618	61,266	58,821	59,195	57,585	54,507	53,884
52,475	54,278	51,746	51,331	48,531	49,129	50,141	48,413	52,249	52,195	50,861	50,497	46,845	45,901
57,652	56,721	53,887	48,945	50,046	47,171	46,022	50,884	51,247	51,514	51,559	50,199	46,199	48,841

Income | Survey Pivot | Survey | checkboxes | Pictures and Labels | Weather

Practical – 2

Estimating demand curves and using solver to optimize price

1. Your company charges $60 for a board game it invented and has sold 3,000 copies during the last year. Elasticity for board games is known to equal 3. Use this information to determine a linear and power demand curve.

Linear Demand Curve :

Step 1 : Calculate Number of sold copies for 1% increase on price $60 based on the formula

Number of sold copies - Number of sold copies*Elasticity%

Step 2 : Insert a X-Y Scatter plot chart.

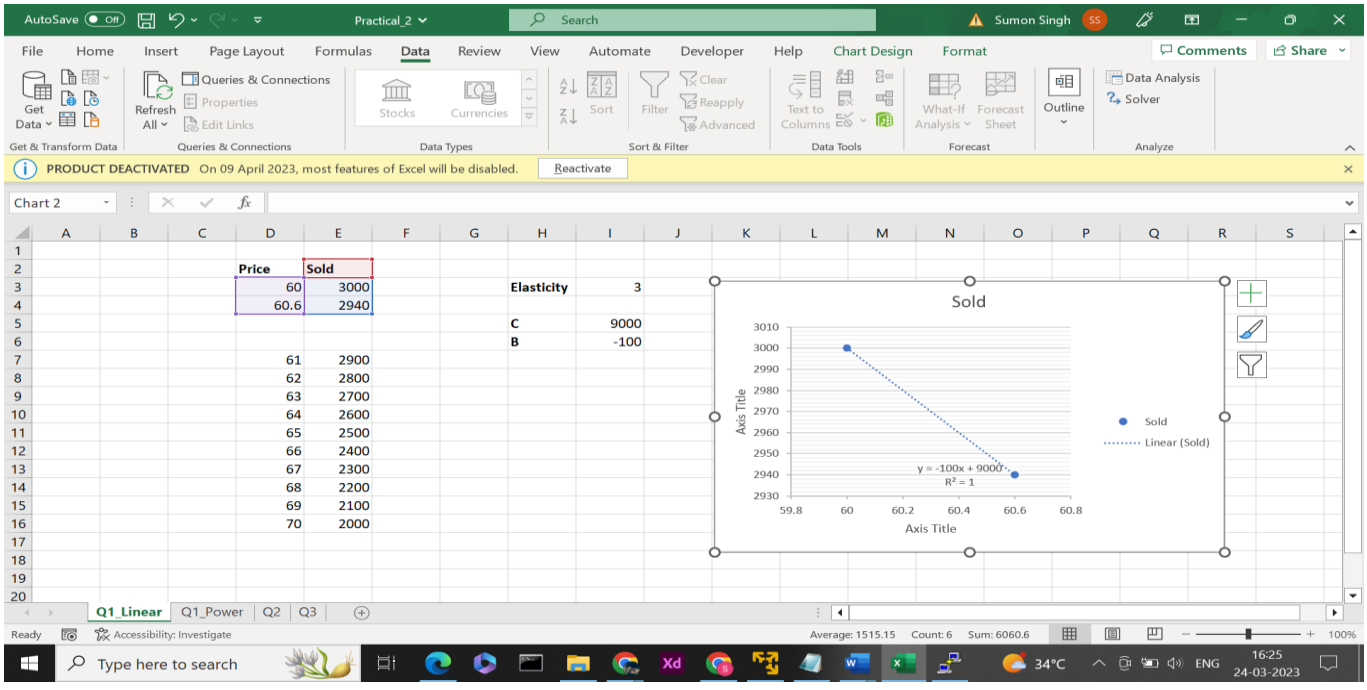
Step 3 : From Chart Design section select Quick Layout (layout 9) .

Step 4 : Get the straight line equation.

Step 5 : Get intercept and slope from the straight line.

Step 6 : Generate Number of sold copies based on the following formula -------

Number of copies sold = intercept + Price*slope



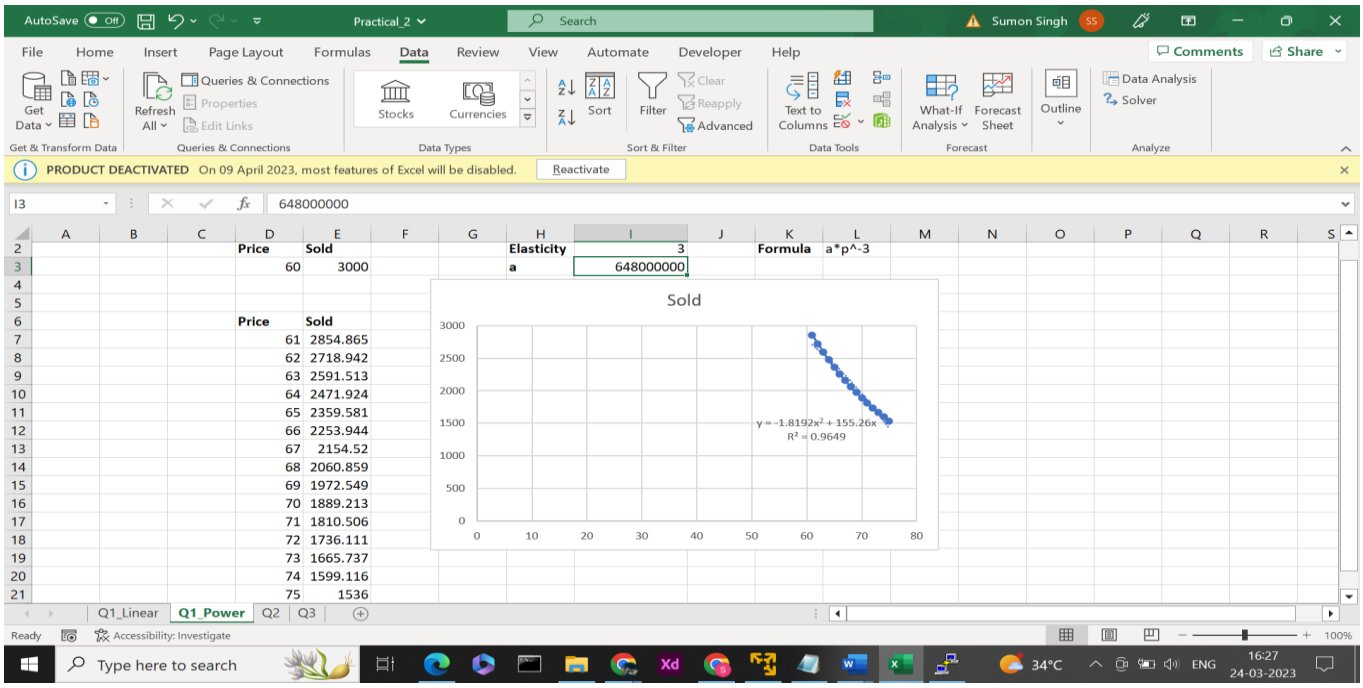
Power Demand Curve :

Step 1 : Using Goal Seek tool get the correct coefficient value of '**a**' from the formula a*p^-e

a -> coefficient, p -> price, e -> elasticity

Step 2 : Based on the correct coefficient generate Number of copies sold based on a*p^-e

Step 3 : Insert chart for new generated points and select trendline for polynomial of 2.



2. In the razorsandblades.xls file example, suppose the cost to produce a blade is $0.20. If you charge $0.35 for a blade, a customer buys an average of 50 blades from you. Assume the price elasticity of demand for blades is 3. What price should you charge for a razor and for a blade?

Step 1 : Calculate demand for 1% increase on price based on the formula

Demand -Demand*Elasticity%

Step 2 : Insert a X-Y Scatter plot chart.

Step 3 : From Chart Design section select Quick Layout (layout 9) .

Step 4 : Get the straight line equation.

Step 5 : Get intercept and slope from the straight line.

Step 6 : Take any random price value.

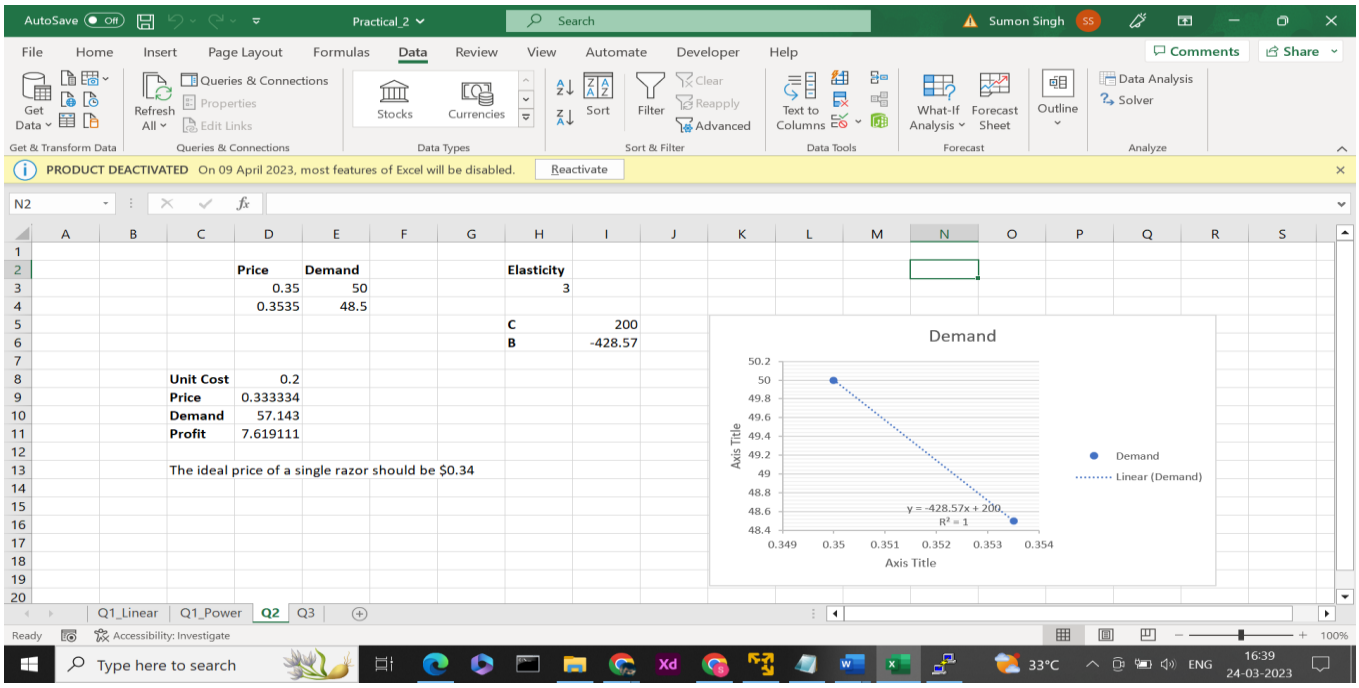
Step 7 : Generate Demand based on the following formula ----

Demand = intercept + Price*slope

Step 8 : Get profit based on following formula ------

Profit = Demand * (Price – Unit Cost)

Step 9 : Using solver set profit for max value by changing the price value.



3. Given the following information in the table for four products, find the profit maximizing the price for each product.

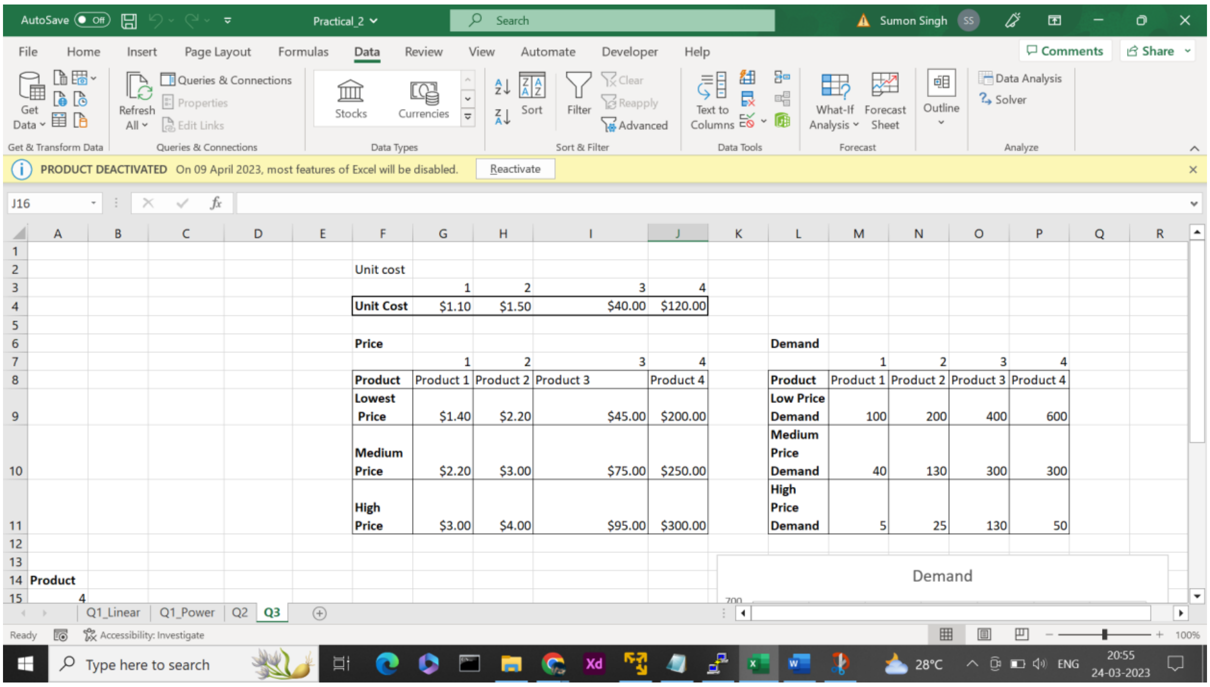
Product	Lowest Price	Medium Price	High Price	Low Price Demand	Medium Price Demand	High Price Demand	Unit cost
Product 1	$1.40	$2.20	$3.00	100	40	5	$1.10
Product 2	$2.20	$3.00	$4.00	200	130	25	$1.50
Product 3	$45	$75	$95	400	300	130	$40
Product 4	$200	$250	$300	600	300	50	$120

Step 1 : Create a HLOOKUP function so that for each product we can get power curve and curve equation by just manipulating the number of the product.

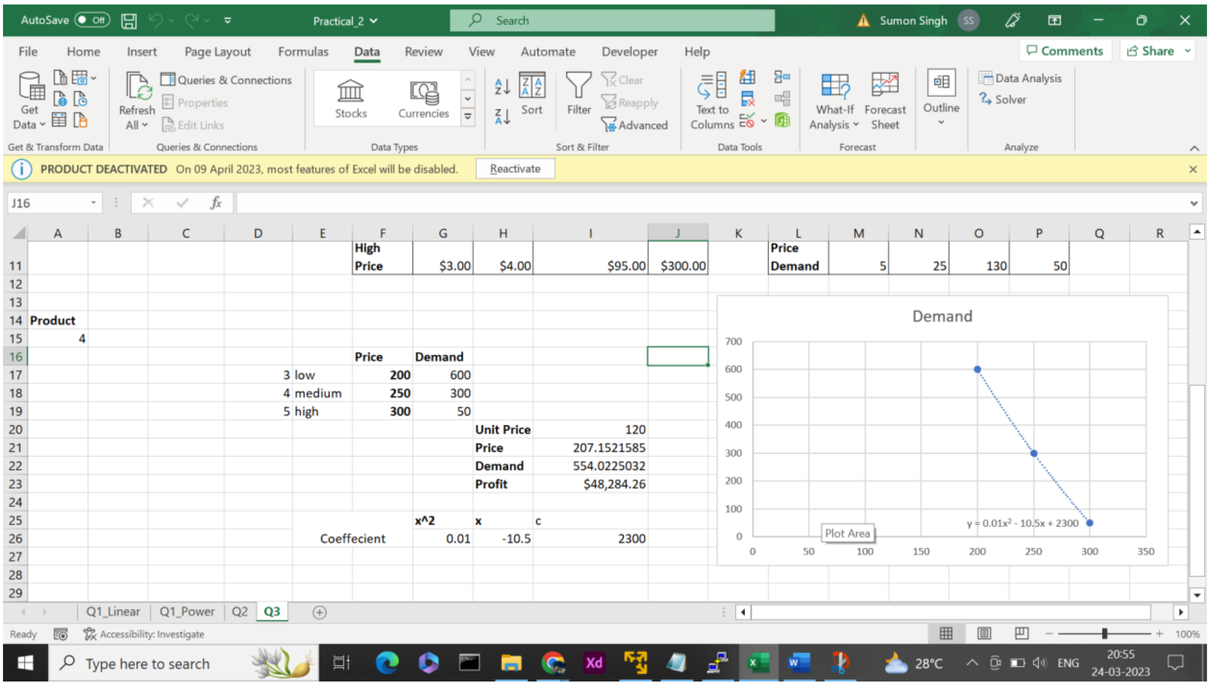
Step 2 : Choose a random price. (Cell I21)

Step 3 : Run the Solver. (Don’t change any field)

Step 4 : Maximum profit will be generated for the correct price.



Unit cost
Unit Cost $1.10 $1.50 $40.00 $120.00
Price
Product Product 1 Product 2 Product 3 Product 4
Lowest Price $1.40 $2.20 $45.00 $200.00
Medium Price $2.20 $3.00 $75.00 $250.00
High Price $3.00 $4.00 $95.00 $300.00
Demand
Product Product 1 Product 2 Product 3 Product 4
Low Price Demand 100 200 400 600
Medium Price Demand 40 130 300 300
High Price Demand 5 25 130 50
Product
4
Demand



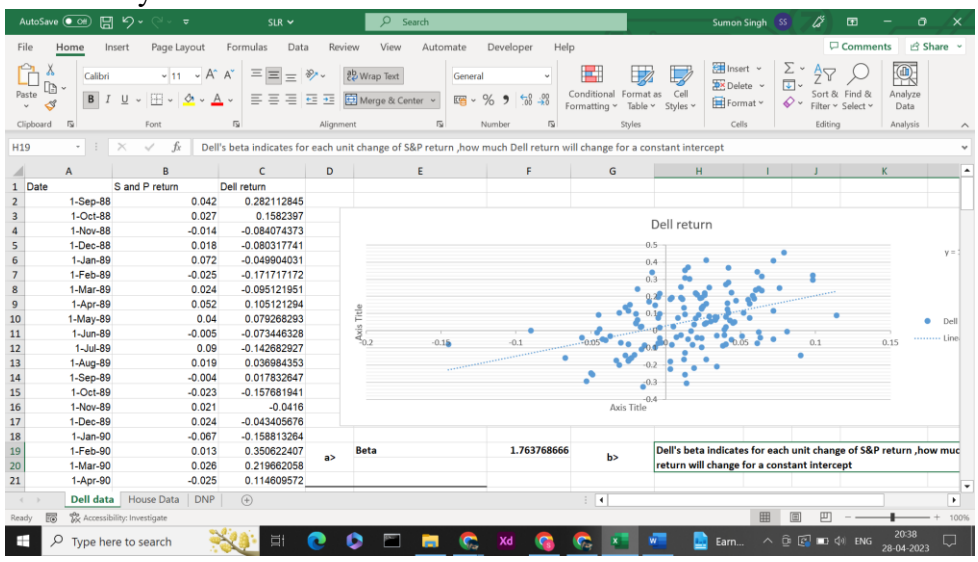
Product
4
Price Demand
3 low 200 600
4 medium 250 300
5 high 300 50
Unit Price 120
Price 207.1521585
Demand 554.0225032
Profit $48,284.26
x^2 x c
Coeffecient 0.01 -10.5 2300
Demand
y = 0.01x² - 10.5x + 2300

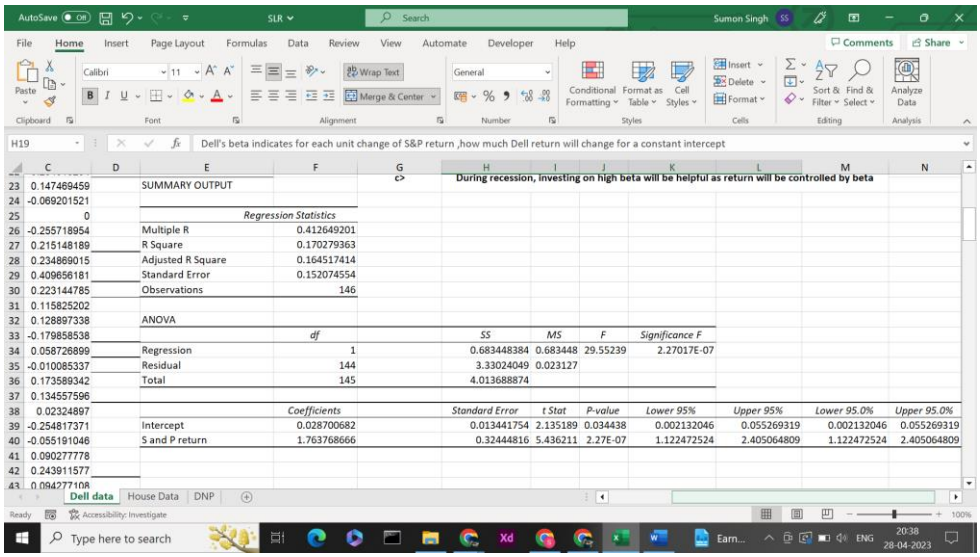
Practical – 3

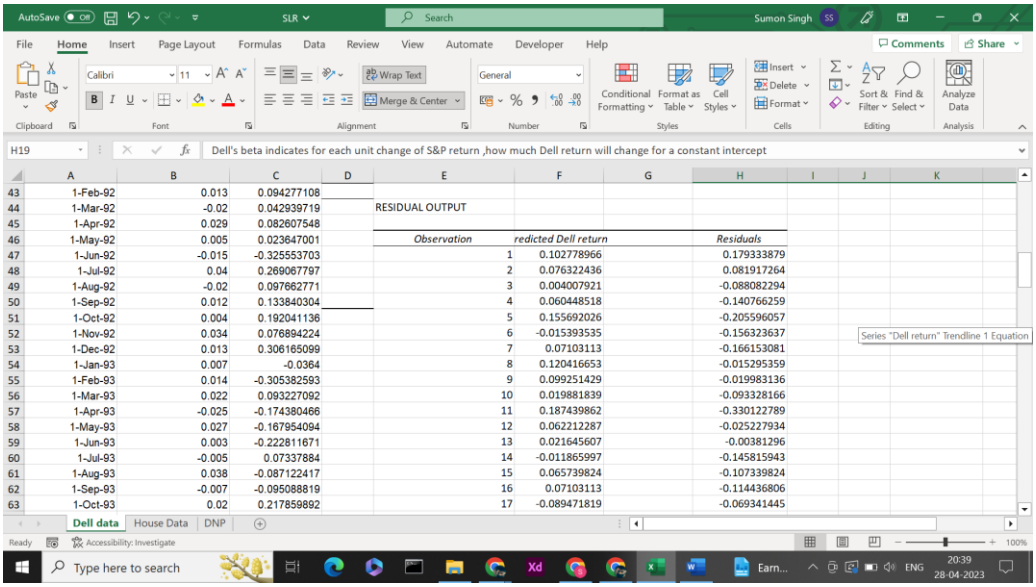
Regression: Simple and Multiple

Q 1] The file Delldata.xlsx contains monthly returns for the Standard &Poor's stock index and for Dell stock. The beta of a stock is defined as the slope of the least- squares line used to predict the monthly return for a stock from the monthly return for the market. Use this file to perform the following exercises:

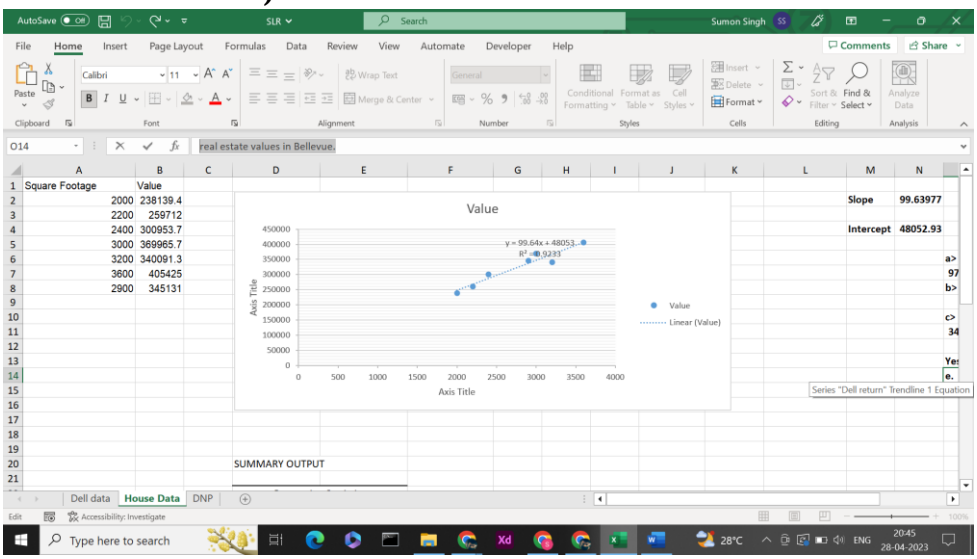
1. Estimate the beta of Dell - **1.763768666**
2. Interpret the meaning of Dell's beta - Dell's beta indicates for each unit change of S&P return, how much Dell return will change for a constant intercept
3. If you believe a recession is coming, would you rather invest in a high-beta or low-beta stock? - During recession, investing on high beta will be helpful as return will be controlled by beta

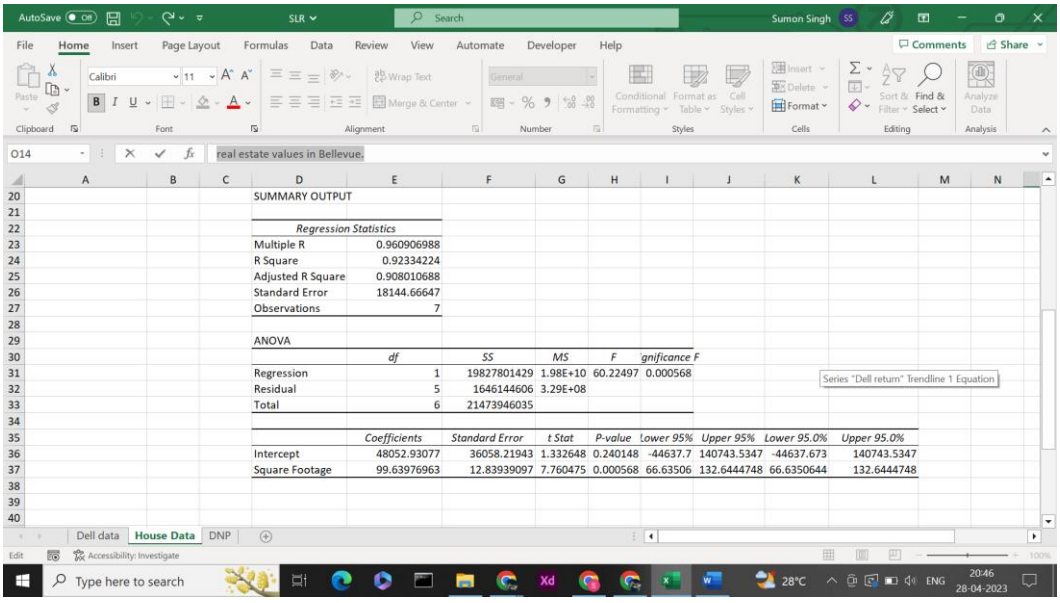






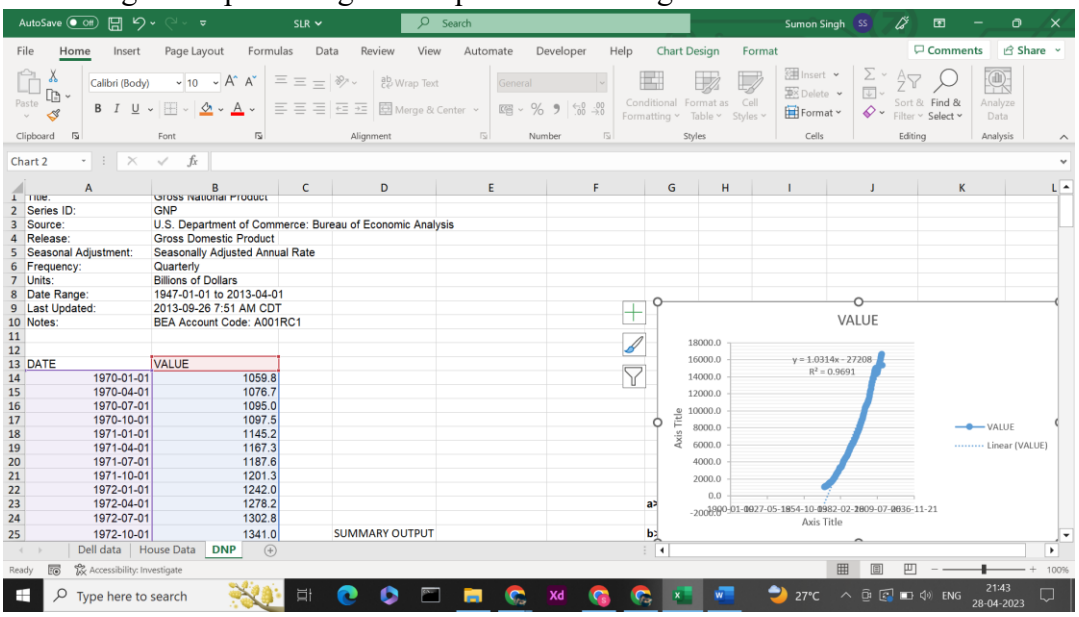
Q 2] The file Housedata.xlsx gives the square footage and sales prices for several houses in Bellevue, Washington. Use this file to answer the following questions:

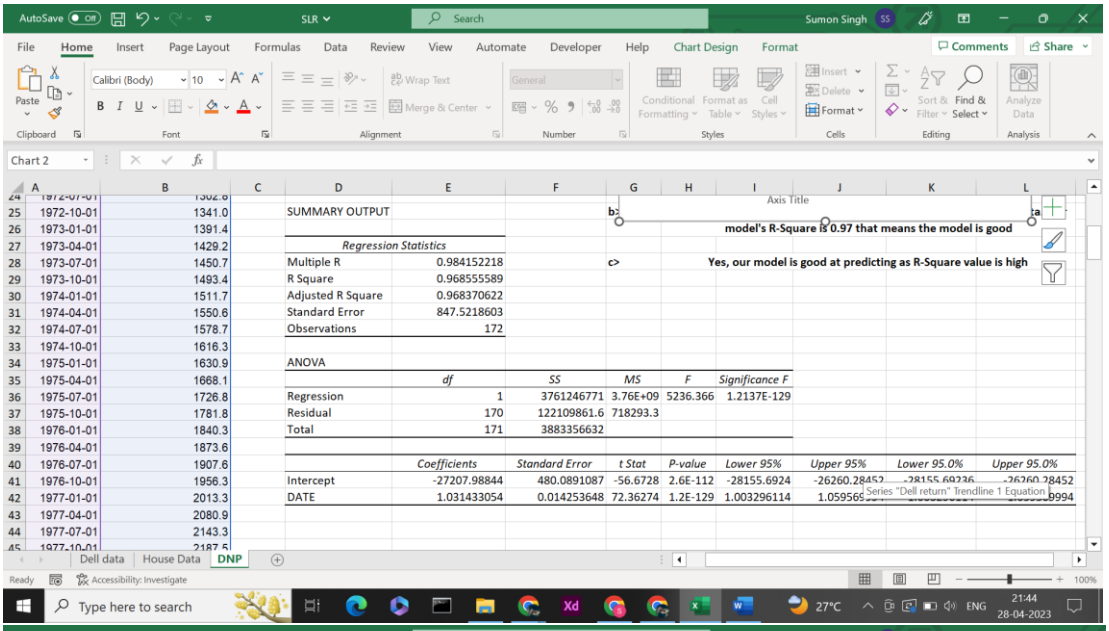
1. You plan to build a 500-square-foot addition to your house. How much do you think your home value will increase as a result? - **97872.82**
2. What percentage of the variation in home value is explained by the variation interhouse size? - **0.92**
3. A 3,000-square-foot house is listed for $500,000. Is this price out of line with typical real estate values in Bellevue? What might cause this discrepancy? - **346972.2 (Yes, the price of $500,000 is out of line with typical real estate values in Bellevue.)**
4. 

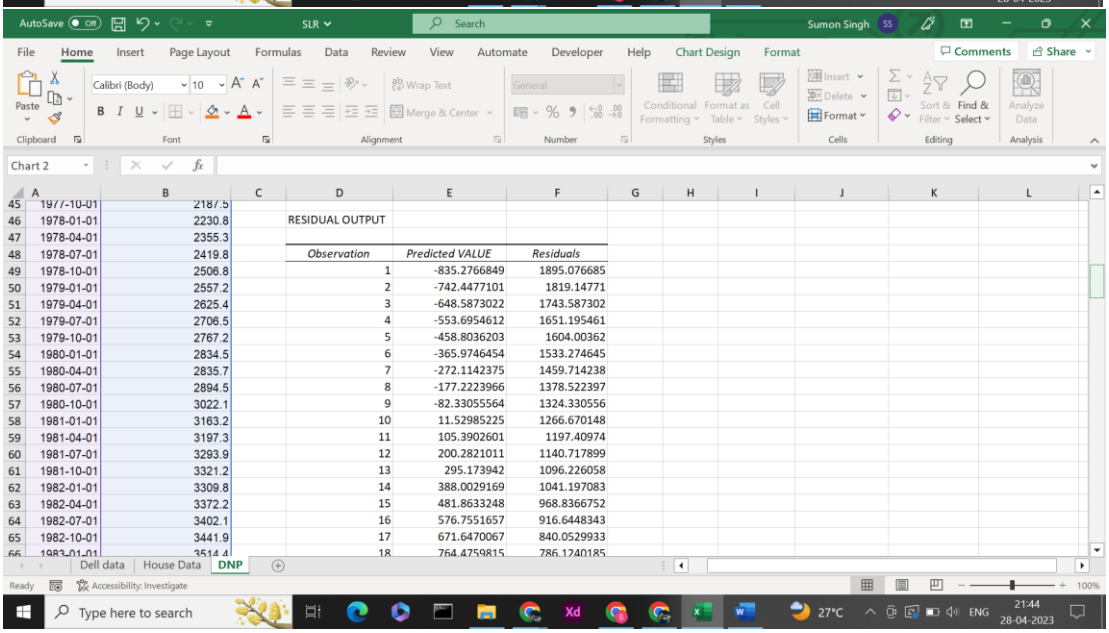
5. 

Q 3] The file GNP.xls contains quarterly GNP data for the United States in the years 1970–2012.

1. Try to predict next quarter's GNP from last quarter's GNP - The next quarter's value is 15364.4
2. What is the R^2? - R-square indicates how much good the model is fitted on training data. Our model's R-Square is 0.97 that means the model is good
3. Does this mean you are good at predicting next quarter's GNP? - Yes, our model is good at predicting as R-Square value is high

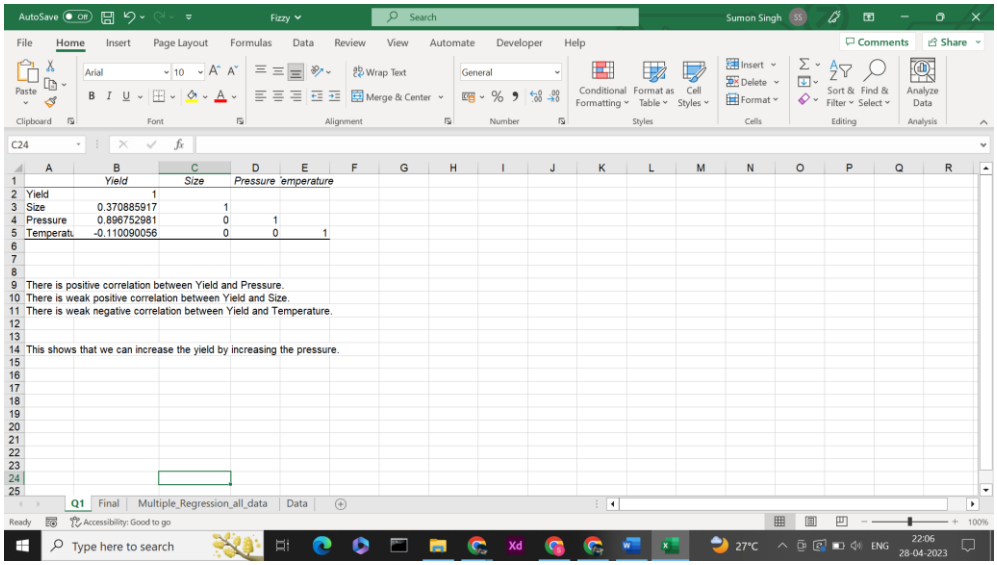
4. 

5. 

6. 

Q 4] Fizzy Drugs wants to optimize the yield from an important chemical process. The company thinks that the number of pounds produced each time the process runs depends on the size of the container used, the pressure, and the temperature. The scientists involved believe the effect to change one variable might depend on the values of other variables. The size of the process container must be between 1.3and 1.5 cubic meters; pressure must be between 4 and 4.5 mm; and temperature must be between 22 and 30 degrees Celsius. The scientists patiently set up experiments at the lower and upper levels of the three control variables and obtain the data shown in the file Fizzy.xlsx

1. Determine the relationship between yield, size, temperature, and pressure –

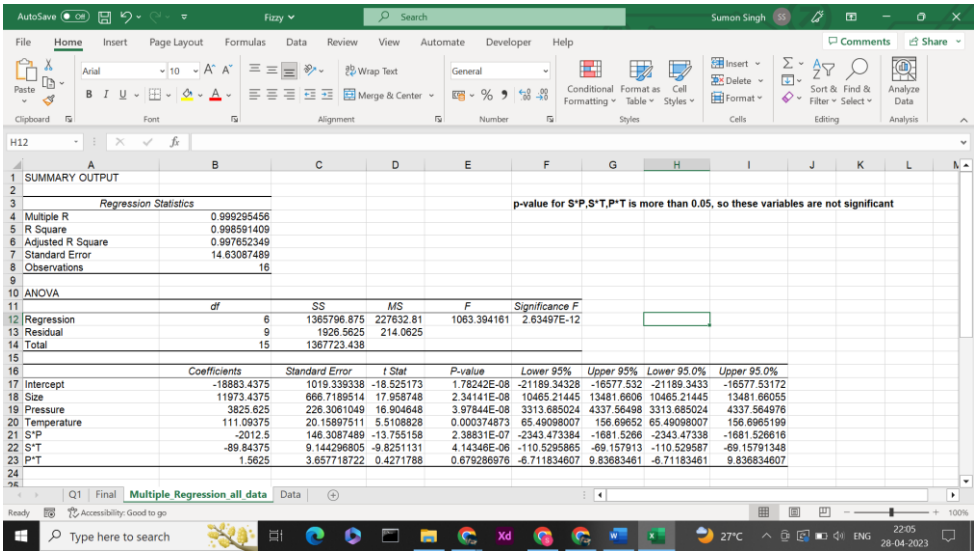


	A	B	C	D	E
1		*Yield*	*Size*	*Pressure*	*emperature*
2	Yield	1			
3	Size	0.370885917	1		
4	Pressure	0.896752981	0	1	
5	Temperatu	-0.110090056	0	0	1

There is positive correlation between Yield and Pressure.
There is weak positive correlation between Yield and Size.
There is weak negative correlation between Yield and Temperature.

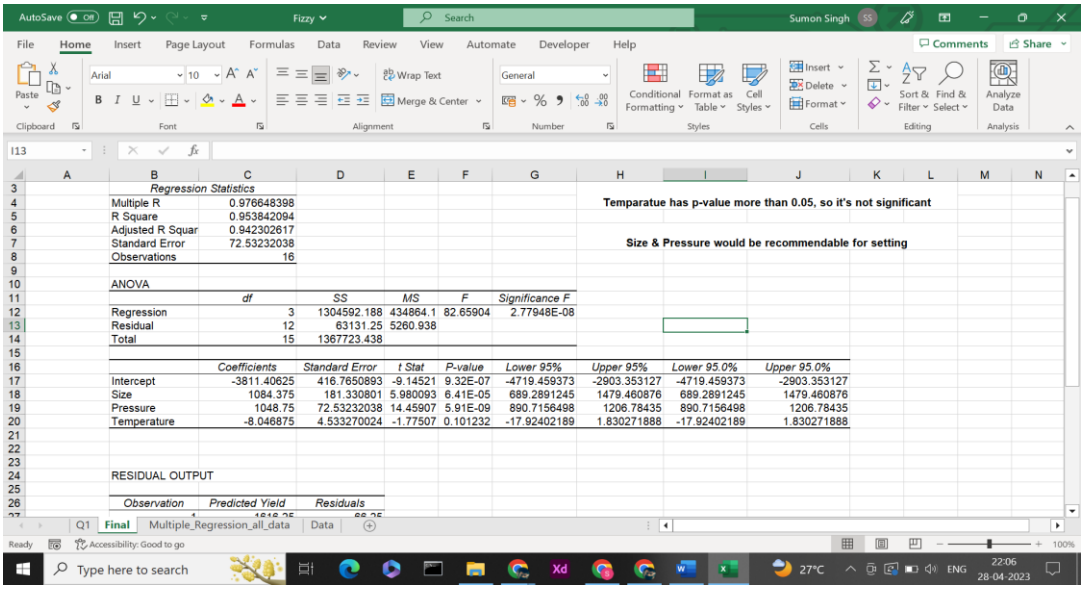
This shows that we can increase the yield by increasing the pressure.

2. Discuss the interactions between pressure, size, and temperature –



	A	B	C	D	E	F	G	H	I	J	K	L
1	SUMMARY OUTPUT											
2												
3	*Regression Statistics*					**p-value for S*P,S*T,P*T is more than 0.05, so these variables are not significant**						
4	Multiple R	0.999295456										
5	R Square	0.998591409										
6	Adjusted R Square	0.997652349										
7	Standard Error	14.63087489										
8	Observations	16										
9												
10	ANOVA											
11		*df*	*SS*	*MS*	*F*	*Significance F*						
12	Regression	6	1365796.875	227632.81	1063.394161	2.63497E-12						
13	Residual	9	1926.5625	214.0625								
14	Total	15	1367723.438									
15												
16		*Coefficients*	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*			
17	Intercept	-18883.4375	1019.339338	-18.525173	1.78242E-08	-21189.34328	-16577.532	-21189.3433	-16577.53172			
18	Size	11973.4375	666.7189514	17.958748	2.34141E-08	10465.21445	13481.6606	10465.21445	13481.66055			
19	Pressure	3825.625	226.3061049	16.904648	3.97844E-08	3313.685024	4337.56498	3313.685024	4337.564976			
20	Temperature	111.09375	20.15897511	5.5108828	0.000374873	65.49098007	156.69652	65.49098007	156.6965199			
21	S*P	-2012.5	146.3087489	-13.755158	2.38831E-07	-2343.473384	-1681.5266	-2343.47338	-1681.526616			
22	S*T	-89.84375	9.144296805	-9.8251131	4.14346E-06	-110.5295865	-69.157913	-110.529587	-69.15791348			
23	P*T	1.5625	3.657718722	0.4271788	0.679286976	-6.711834607	9.83683461	-6.71183461	9.836834607			

3. What settings for temperature, size, and pressure would you recommend?



	Regression Statistics	
Multiple R		0.976648398
R Square		0.953842094
Adjusted R Squar		0.942302617
Standard Error		72.53232038
Observations		16

Temparatue has p-value more than 0.05, so it's not significant

Size & Pressure would be recommendable for setting

ANOVA

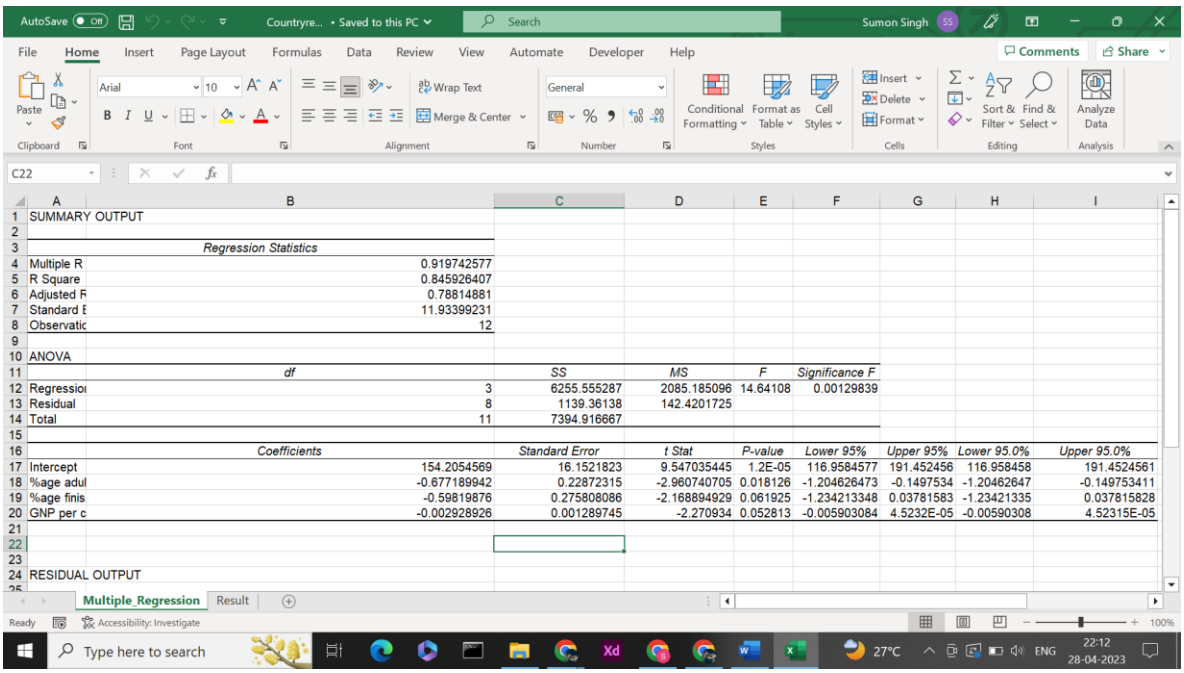
	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1304592.188	434864.1	82.65904	2.77948E-08
Residual	12	63131.25	5260.938		
Total	15	1367723.438			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-3811.40625	416.7650893	-9.14521	9.32E-07	-4719.459373	-2903.353127	-4719.459373	-2903.353127
Size	1084.375	181.330801	5.980093	6.41E-05	689.2891245	1479.460876	689.2891245	1479.460876
Pressure	1048.75	72.53232038	14.45907	5.91E-09	890.7156498	1206.78435	890.7156498	1206.78435
Temperature	-8.046875	4.533270024	-1.77507	0.101232	-17.92402189	1.830271888	-17.92402189	1.830271888

RESIDUAL OUTPUT

Observation	*Predicted Yield*	*Residuals*

Q 5] The file Countryregion.xlsx contains the following data for several underdeveloped countries: · Infant mortality rate · Adult literacy rate · Percentage of students finishing primary school · Per capita GNP Use this data to develop an equation that can be used to predict infant mortality. Are there any outliers in this set of data? Interpret the coefficients in your equation. Within what value should 95 percent of your predictions for infant mortality be accurate?



SUMMARY OUTPUT

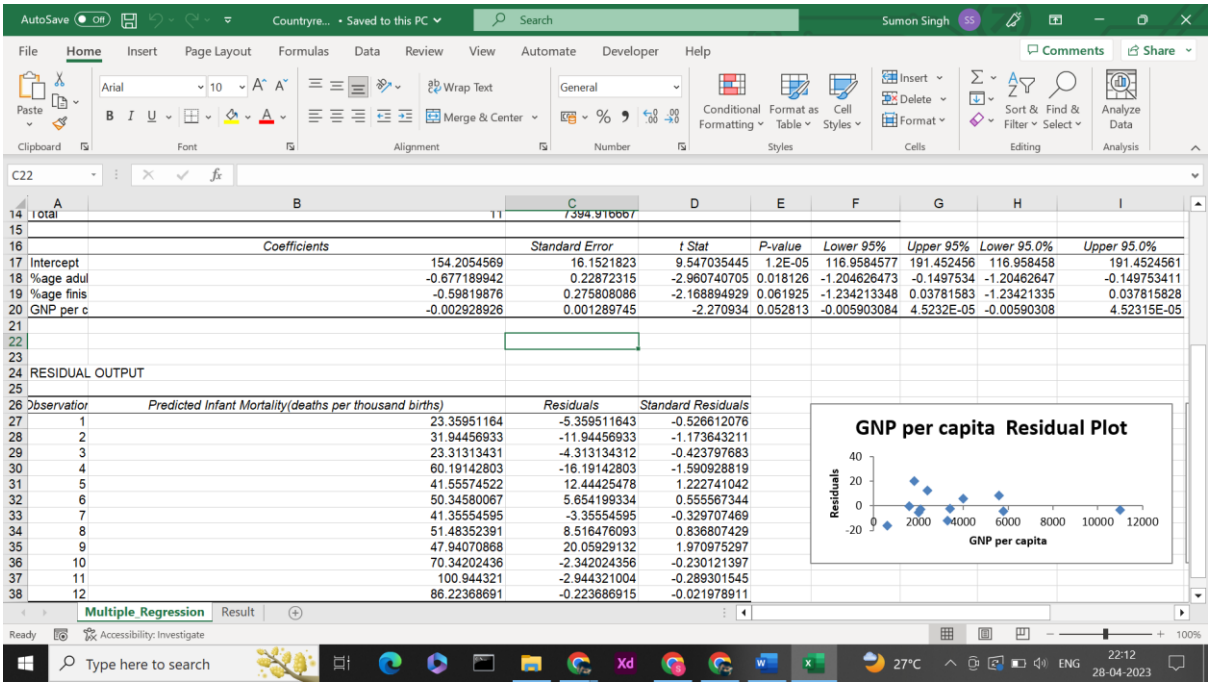
Regression Statistics	
Multiple R	0.919742577
R Square	0.845926407
Adjusted F	0.78814881
Standard E	11.93399231
Observatic	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regressio	3	6255.555287	2085.185096	14.64108	0.00129839
Residual	8	1139.36138	142.4201725		
Total	11	7394.916667			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	154.2054569	16.1521823	9.547035445	1.2E-05	116.9584577	191.452456	116.958458	191.4524561
%age adul	-0.677189942	0.22872315	-2.960740705	0.018126	-1.204626473	-0.1497534	-1.20462647	-0.149753411
%age finis	-0.59819876	0.275808086	-2.168894929	0.061925	-1.234213348	0.03781583	-1.23421335	0.037815828
GNP per c	-0.002928926	0.001289745	-2.270934	0.052813	-0.005903084	4.5232E-05	-0.00590308	4.52315E-05

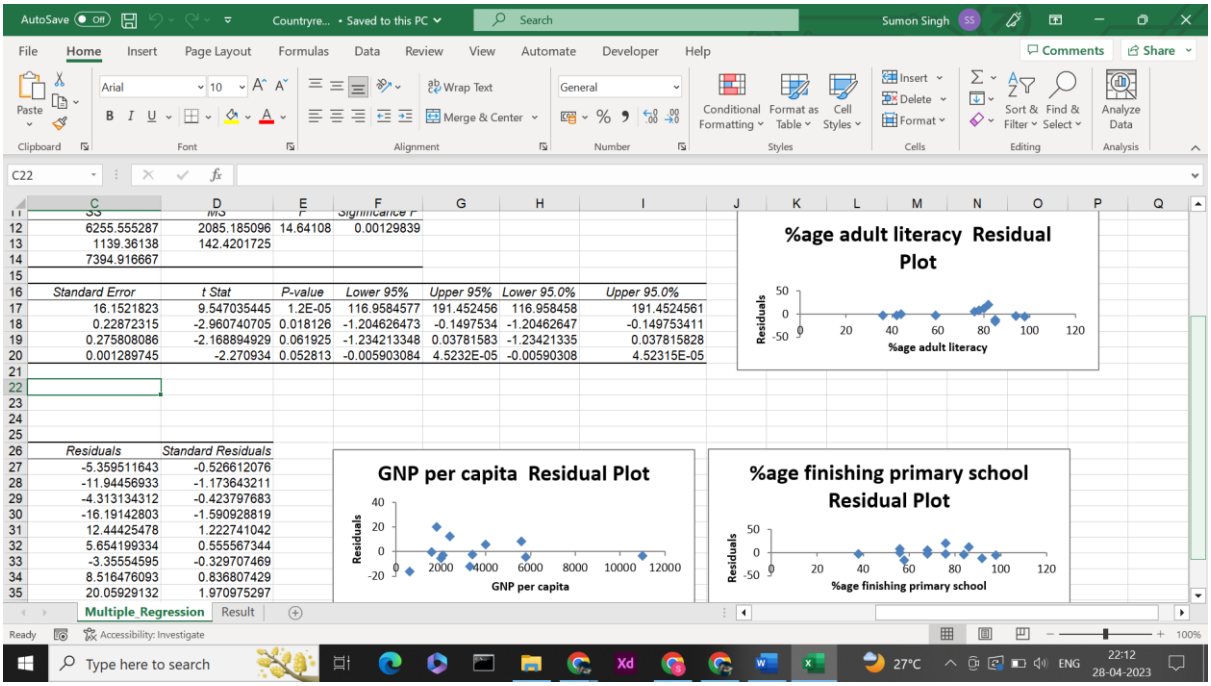
RESIDUAL OUTPUT



	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	154.2054569	16.1521823	9.547035445	1.2E-05	116.9584577	191.452456	116.958458	191.4524561
%age adul	-0.677189942	0.22872315	-2.960740705	0.018126	-1.204626473	-0.1497534	-1.20462647	-0.149753411
%age finis	-0.59819876	0.275808086	-2.168894929	0.061925	-1.234213348	0.03781583	-1.23421335	0.037815828
GNP per c	-0.002928926	0.001289745	-2.270934	0.052813	-0.005903084	4.5232E-05	-0.00590308	4.52315E-05

RESIDUAL OUTPUT

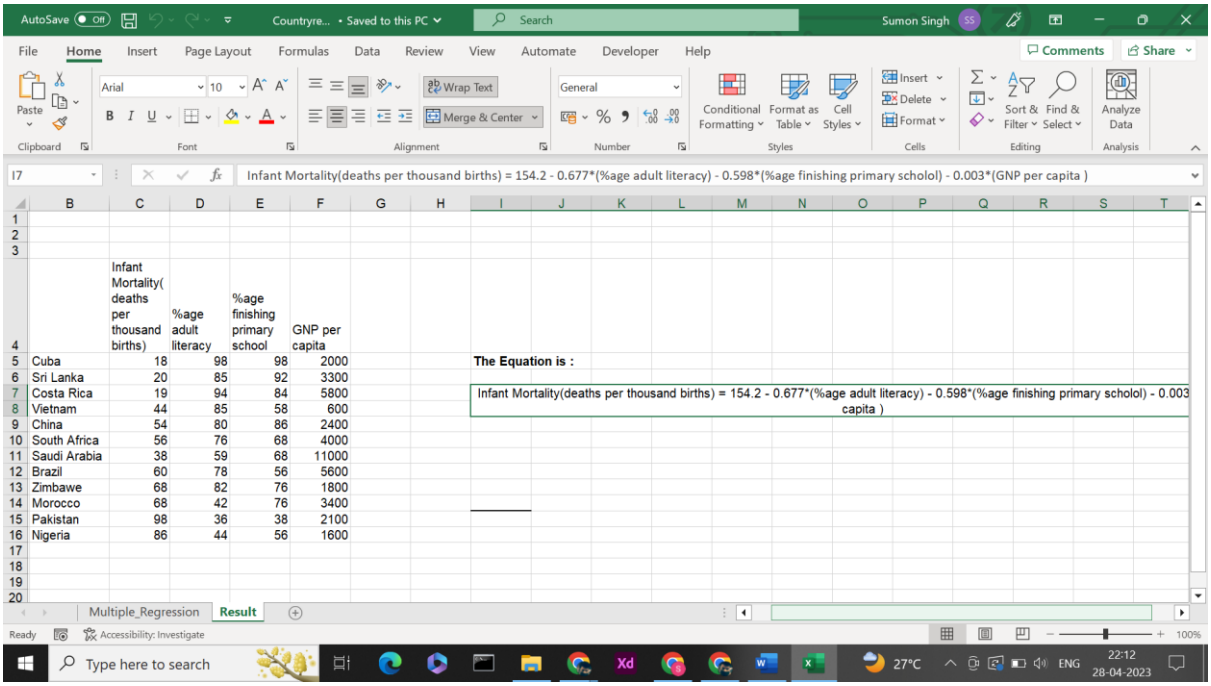
Observation	Predicted Infant Mortality(deaths per thousand births)	Residuals	Standard Residuals
1	23.35951164	-5.359511643	-0.526612076
2	31.94456933	-11.94456933	-1.173643211
3	23.31313431	-4.313134312	-0.423797683
4	60.19142803	-16.19142803	-1.590928819
5	41.55574522	12.44425478	1.222741042
6	50.34580067	5.654199334	0.555567344
7	41.35554595	-3.35554595	-0.329707469
8	51.48352391	8.516476093	0.836807429
9	47.94070868	20.05929132	1.970975297
10	70.34202436	-2.342024356	-0.230121397
11	100.944321	-2.944321004	-0.289301545
12	86.22368691	-0.223686915	-0.021978911



	SS	MS	F	Significance F
	6255.555287	2085.185096	14.64108	0.00129839
	1139.36138	142.4201725		
	7394.916667			

	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	16.1521823	9.547035445	1.2E-05	116.9584577	191.452456	116.958458	191.4524561
%age adul	0.22872315	-2.960740705	0.018126	-1.204626473	-0.1497534	-1.20462647	-0.149753411
%age finis	0.275808086	-2.168894929	0.061925	-1.234213348	0.03781583	-1.23421335	0.037815828
GNP per c	0.001289745	-2.270934	0.052813	-0.005903084	4.5232E-05	-0.00590308	4.52315E-05

Residuals	Standard Residuals
-5.359511643	-0.526612076
-11.94456933	-1.173643211
-4.313134312	-0.423797683
-16.19142803	-1.590928819
12.44425478	1.222741042
5.654199334	0.555567344
-3.35554595	-0.329707469
8.516476093	0.836807429
20.05929132	1.970975297


AutoSave
Countryre... • Saved to this PC
Search
Sumon Singh
File Home Insert Page Layout Formulas Data Review View Automate Developer Help
Comments Share
Paste Arial 10 Wrap Text General
Merge & Center
Conditional Formatting Format as Table Cell Styles
Insert Delete Format
Sort & Filter Find & Select Analyze Data
Clipboard Font Alignment Number Styles Cells Editing Analysis
I7
Infant Mortality(deaths per thousand births) = 154.2 - 0.677*(%age adult literacy) - 0.598*(%age finishing primary schoolol) - 0.003*(GNP per capita)
Infant Mortality(deaths per thousand births)
%age adult literacy
%age finishing primary school
GNP per capita
Cuba 18 98 98 2000
Sri Lanka 20 85 92 3300
Costa Rica 19 94 84 5800
Vietnam 44 85 58 600
China 54 80 86 2400
South Africa 56 76 68 4000
Saudi Arabia 38 59 68 11000
Brazil 60 78 56 5600
Zimbawe 68 82 76 1800
Morocco 68 42 76 3400
Pakistan 98 36 38 2100
Nigeria 86 44 56 1600
The Equation is :
Infant Mortality(deaths per thousand births) = 154.2 - 0.677*(%age adult literacy) - 0.598*(%age finishing primary schoolol) - 0.003 capita)
Multiple_Regression
Result
Ready
Accessibility: Investigate
100%
Type here to search
27°C
ENG
22:12
28-04-2023

PRACTICAL – 4

Forecasting

Q 1] Build a model to forecast daily customer count at the Indiana University Credit Union.

Use creditunion data

Step 1. First, create indicator variables (in columns G through J) for whether the day is a staff payday (SP), faculty payday (FAC), before a holiday (BH), or after a holiday (AH). (Refer to Figure 11.1). For example, cells G4, H4, and J4 use 1 to indicate that January 2 was a staff payday, faculty payday, and after a holiday. Cell I4 contains 0 to indicate that January 2 was not before a holiday.

Step 2. The forecast is defined by a constant (which helps to center the forecasts so that they will be more accurate), and effects for each day of the week, each month, a staff payday, a faculty payday, a day occurring before a holiday, and a day occurring after a holiday. Insert Trial values for all these parameters (the Solver changing cells) in the cell range O4:O26. Solver can then choose values that make the model best fit the data. For each day, the forecast of customer count will be generated by the following equation:
Predicted customer count=Constant+(Month effect)+(Day of week effect)+(Staff payday effect, if any)+(Faculty payday effect, if any)+(Before holiday effect, if any)+(After holiday effect, if any)

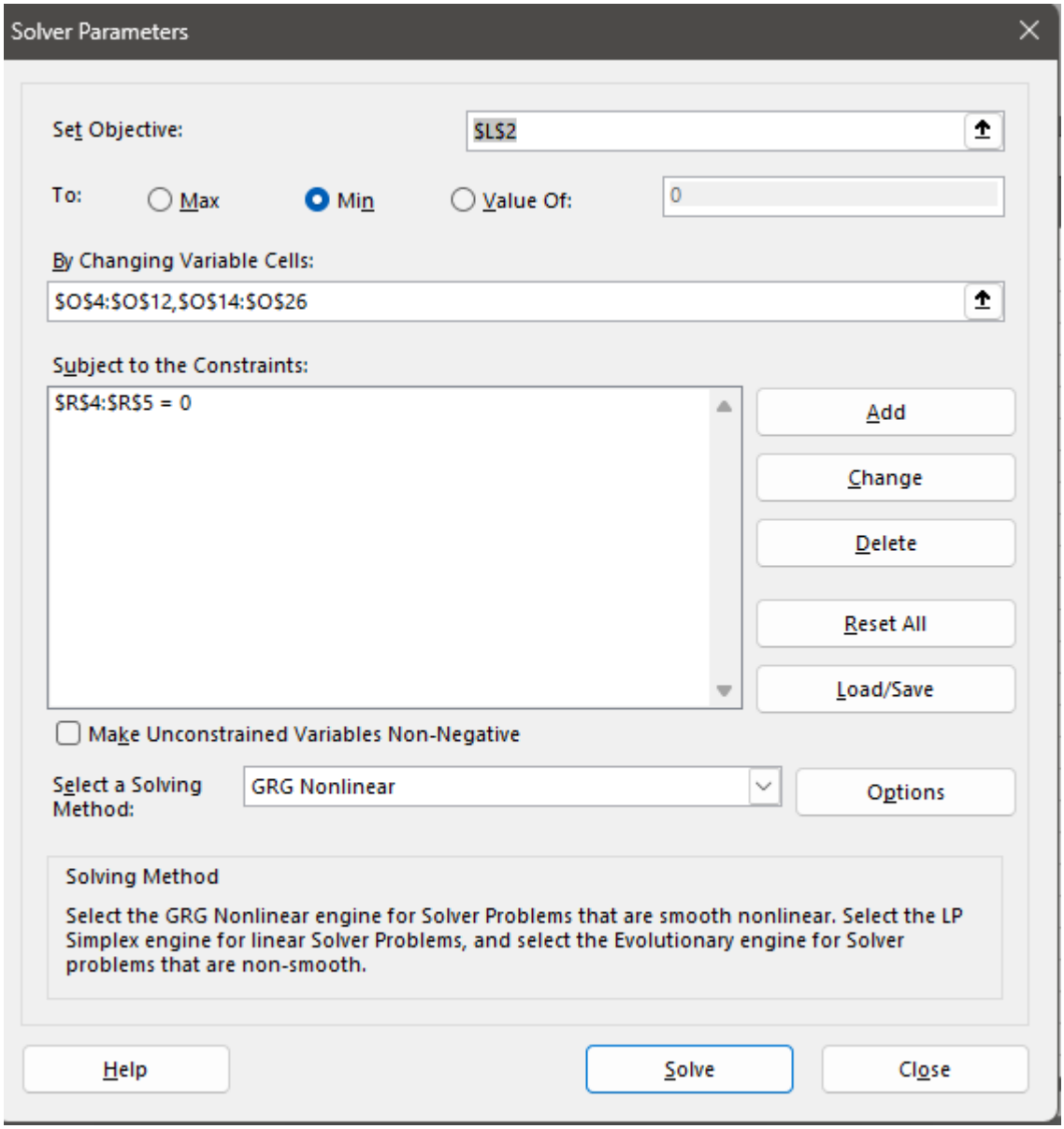
Step 3. Using this model, compute a forecast for each day's customer count by copying the following formula from K4 to K5:K257:
$O$26+VLOOKUP(B4,$N$14:$O$25,2)+VLOOKUP(D4,$N$4:$O$8,2)
+G4*$O$9+H4*$O$10+I4*$O$11+J4*$O$12.
Cell O26 picks up the constant term. VLOOKUP(B4,$N$14:$O$25,2) picks up the month coefficient for the current month, and VLOOKUP(D4,$N$4:$O$8,2) picks up the day of the week coefficient for the current week.
=G4*$O$9+H4*$O$10+I4*$O$11+J4*$O$12 picks up the effects (if any) when the current day is SP, FAC, BH, or AH.

Step 4. Copy the formula =(E4-K4)^2, from L4 to L5:L257 to compute the squared error for each day. Then, in cell L2, compute the sum of squared errors with the formula =SUM(L4:L257).

Step 5. In cell R4, average the day of the week changing cells with the formula =AVERAGE(O4:O8), and in cell R5, average the month changing cells with the formula =AVERAGE(O14:O25). Later in this section you will add constraints to your Solver model which constrain the average month and day of the week effects to equal 0. These constraints ensure that a month or day of the week with a positive effect has a higher than average customer count, and a month or day of the week with a negative effect has a lower than average customer count.

Step 6. Use the Solver settings to choose the forecast parameters to minimize the sum of squared errors.

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
							RSQ	0.771		stdeverr	163.1772					
									SSE	6736582						
MONTH	DAYMON	DAYWEEK	CUST	SPECIAL	SP	FAC	BH	AH	Forecast	Sq Err	Error	Day of Week			average	
1	2	2	1825	SP,FAC,AH	1	1	0	1	1766.78	3389.56	58.21993	1	103.357		dayweek	2.0464E-13
1	3	3	1257	0	0	0	0	0	709.603	299643	547.3965	2	-139.19		month	-3.8571E-09
1	4	4	969	0	0	0	0	0	745.698	49863.78	223.302	3	-150.34			
1	5	5	1672	SP	1	0	0	0	1557.22	13174.18	114.7788	4	-114.25			
1	8	1	1098	0	0	0	0	0	963.303	18143.29	134.697	5	300.424			
1	9	2	691	0	0	0	0	0	720.753	885.2568	-29.7533	SP	396.851			
1	10	3	672	0	0	0	0	0	709.603	1414.02	-37.6035	FAC	394.894			
1	11	4	754	0	0	0	0	0	745.698	68.92317	8.301998	BH	205.293			
1	12	5	972	0	0	0	0	0	1160.37	35483.19	-188.37	AH	254.281			
1	15	1	816	0	0	0	0	0	963.303	21698.16	-147.303	Month				
1	16	2	717	0	0	0	0	0	720.753	14.08701	-3.75327	1	-110.69			
1	17	3	728	0	0	0	0	0	709.603	338.4326	18.39654	2	-75.715			
1	18	4	711	0	0	0	0	0	745.698	1203.951	-34.698	3	-40.341			
1	19	5	1545	SP	1	0	0	0	1557.22	149.3565	-12.2212	4	0.02839			
1	22	1	873	0	0	0	0	0	963.303	8154.623	-90.303	5	87.8157			
1	23	2	713	0	0	0	0	0	720.753	60.11313	-7.75327	6	133.341			
1	24	3	626	0	0	0	0	0	709.603	6989.539	-83.6035	7	115.803			
1	25	4	653	0	0	0	0	0	745.698	8592.92	-92.698	8	28.7743			
1	26	5	1080	0	0	0	0	0	1160.37	6459.308	-80.3698	9	-87.563			
1	29	1	650	0	0	0	0	0	963.303	98158.74	-313.303	10	-53.002			
1	30	2	644	0	0	0	0	0	720.753	5891.064	-76.7533	11	-42.761			



The Solver model changes the coefficients for the month, day of the week, BH, AH, SP, FAC, and the constant to minimize the sum of square errors. It also constrains the average day of the week and month effect to equal 0. These show that Friday is the busiest day of the week and June is the busiest month. A staff payday raises the forecast (all else being equal—in the Latin, ceteris paribus) by 397 customers.

Step 7. Select the range M4:M257, and then click Conditional Formatting on the Home tab.

Step 8. Select New Rule and in the New Formatting Rule dialog box, choose Use a Formula to Determine Which Cells to Format.

Step 9. Fill in the rule description in the dialog box,

I	J	K	L	M	N	O	P	Q	R
RSQ	0.771		stdeverr	163.1772					
		SSE	6736582						
BH	AH	Forecast	Sq Err	Error	Day of Week			average	
0	1	1766.78	3389.56	58.21993	1	103.357		dayweek	2.0464E-13
0	0	709.603	299643	547.3965	2	-139.19		month	-3.8571E-09
0	0	745.698	49863.78	223.302	3	-150.34			
0	0	1557.22	13174.18	114.7788	4	-114.25			
0	0	963.303	18143.29	134.697	5	300.424			
0	0	720.753	885.2568	-29.7533	SP	396.851			
0	0	709.603	1414.02	-37.6035	FAC	394.894			
0	0	745.698	68.92317	8.301998	BH	205.293			
0	0	1160.37	35483.19	-188.37	AH	254.281			
0	0	963.303	21698.16	-147.303	Month				
0	0	720.753	14.08701	-3.75327	1	-110.69			
0	0	709.603	338.4326	18.39654	2	-75.715			
0	0	745.698	1203.951	-34.698	3	-40.341			
0	0	1557.22	149.3565	-12.2212	4	0.02839			
0	0	963.303	8154.623	-90.303	5	87.8157			
0	0	720.753	60.11313	-7.75327	6	133.341			
0	0	709.603	6989.539	-83.6035	7	115.803			
0	0	745.698	8592.92	-92.698	8	28.7743			
0	0	1160.37	6459.308	-80.3698	9	-87.563			
0	0	963.303	98158.74	-313.303	10	-53.002			
0	0	720.753	5891.064	-76.7533	11	-42.761			
0	0	709.603	8722.913	93.39654	12	44.3091			

Q 2] Use the Solver to develop an additive or multiplicative model to estimate trends and seasonality.

Use arilinemiles data

Step 1. Enter trial values of the base and trend in cells B2 and B3. Name cell B2 baseadd and cell B3 trend.

Step 2. Enter trial seasonal indices in the range B5:B16.

Step 3. In cell B18, average the seasonal indices with the formula =AVERAGE(B5:B16).The Solver model can set this average to 0 to ensure the seasonal indices average to 0.

Step 4. Copy the formula =baseadd+trend*D9+VLOOKUP(F9,$A$5:$B$16,2) from H9 to H10:H42 to compute the forecast for each month.

Step 5. Copy the formula =G9-H9 from I9 to I10:I42 to compute each month's forecast error.

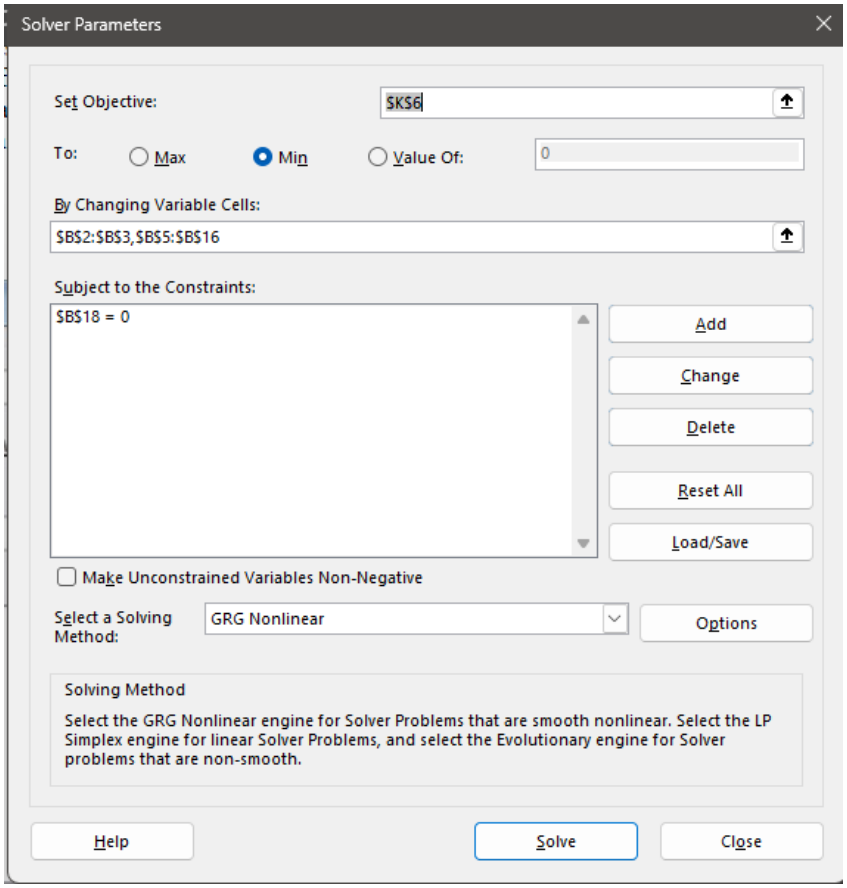
Step 6. Copy the formula =(I9^2) from J9 to J10:J42 to compute each month's squared error.

Step 7. In cell K6, compute the Sum of Squared Errors (SSE) using the formula =SUM(J9:J42).

Step 8. Now set up the Solver model. Change the parameters to minimize SSE and constrain the average of the seasonal indices to 0. Do not check the non-negative box because some seasonal indices must be negative. The forecasting model is a linear forecasting model because each unknown parameter is multiplied by a constant. When the forecasts are created by adding together terms that multiply changing cells by constants, the GRG Solver Engine always finds a unique solution to the least square minimizing parameter estimates for a forecasting model.

Additive trend Solver model

	A	B	C	D	E	F	G	H	I	J	K
1											
2	baseadd	37.37856									
3	trendadd	0.059026									
4										stddeverr	0.386323
5	1	-4.45733								RSQ	0.988934
6	2	-6.62334								SSE	4.9251
7	3	1.601041									
8	4	-0.319		MonthNumber	Month	Month	Airline Miles (billion	Forecast	Error	Sq Error	
9	5	1.274636		1	01-07-2009	7	44.22	43.7288	0.49	0.236896	
10	6	3.795057		2	01-08-2009	8	42.40	41.95583	0.44	0.194662	
11	7	6.291206		3	01-09-2009	9	34.68	35.12698	-0.45	0.203932	
12	8	4.459215		4	01-10-2009	10	37.32	37.69339	-0.38	0.140881	
13	9	-2.42866		5	01-11-2009	11	34.58	35.31697	-0.74	0.54817	
14	10	0.078726		6	01-12-2009	12	36.46	36.41789	0.04	0.001696	
15	11	-2.35673		7	01-01-2010	1	33.49	33.33441	0.15	0.023327	
16	12	-1.31482		8	01-02-2010	2	30.72	31.22743	-0.51	0.259421	
17				9	01-03-2010	3	39.37	39.51084	-0.14	0.019948	
18	mean	0		10	01-04-2010	4	37.76	37.64982	0.11	0.012653	
19				11	01-05-2010	5	38.88	39.30248	-0.42	0.175395	
20				12	01-06-2010	6	41.90	41.88193	0.02	0.000401	
21				13	01-07-2010	7	44.02	44.43711	-0.42	0.172428	



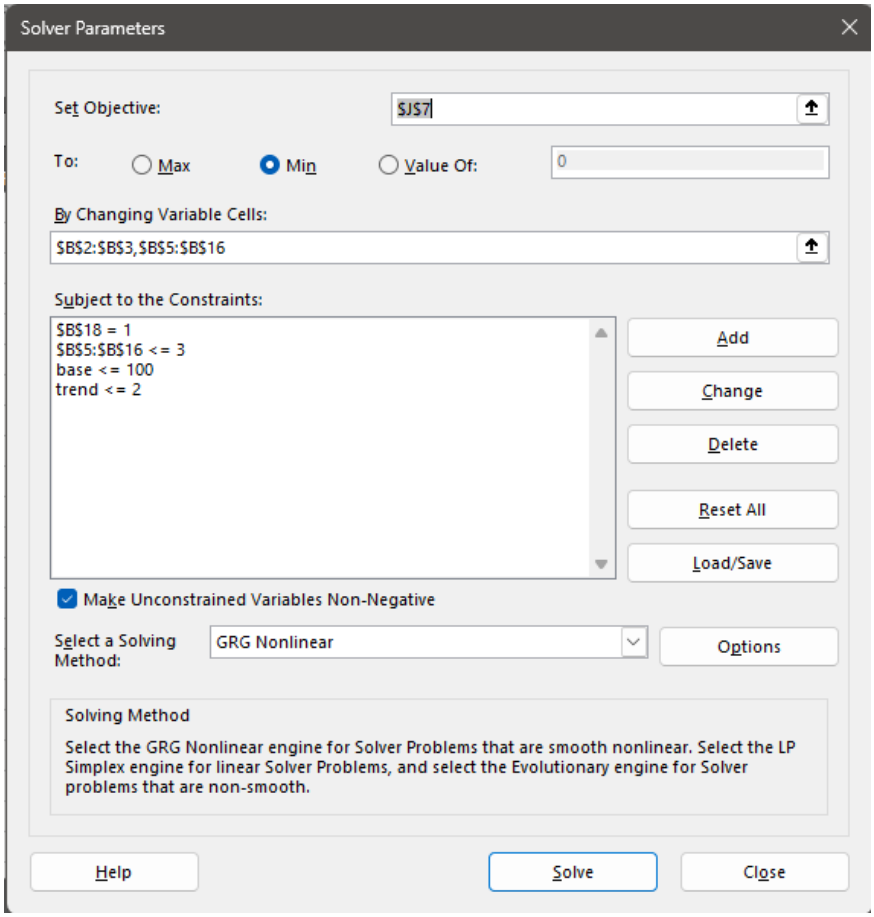
Multiplicative Trend model

Step 9. In Solver select Options, and from the GRG tab, select Multistart. This ensures the Solver will try many (between 50 and 200) starting solutions and find the optimal solution from each starting solution.

Step 10. To use the Multistart option, input lower and upper bounds on the changing cells. To speed up solutions, these bounds should approximate sensible values for the estimated parameters. For example, a seasonal index will probably be between 0 and 3, so an upper bound of 100 would be unreasonable. You can choose an upper bound of 3 for each seasonal index and an upper bound of 2 for the trend. For this example, choose an upper bound of 100 for the base.

Step 11. Cell B18 averages the seasonal indices, so in the Solver window add the constraint $B$18 =1 to ensure that the seasonal indices average to 1.

Step 12. Select Solve, and the Solver will then find the optimal solution



	A	B	C	D	E	F	G	H	I	J	K
1											
2	base	3.74E+01									
3	trend	1.001493569					1.00E+06				
4										stddeverrors	0.411756002
5	1	0.884049011								RSQ	0.987429707
6	2	0.82837254								SSE	
7	3	1.041400111								5.59	
8	4	0.991684904		MonthNumber	Month	Month	AirlineMiles (b	Forecast	Error	Sq Error	
9	5	1.03315296		1	01-07-2009	7	44.22	43.63945105	0.58	0.33	
10	6	1.098599337		2	01-08-2009	8	42.40	41.89570886	0.50	0.25	
11	7	1.164327334		3	01-09-2009	9	34.68	35.19980697	-0.52	0.28	
12	8	1.116136195		4	01-10-2009	10	37.32	37.71609694	-0.40	0.16	
13	9	0.936353344		5	01-11-2009	11	34.58	35.38768125	-0.81	0.66	
14	10	1.00179316		6	01-12-2009	12	36.46	36.46161088	0.00	0.00	
15	11	0.938545346		7	01-01-2010	1	33.49	33.43255124	0.05	0.00	
16	12	0.965585734		8	01-02-2010	2	30.72	31.37379361	-0.66	0.43	
17				9	01-03-2010	3	39.37	39.50091231	-0.13	0.02	
18	mean	0.999999998		10	01-04-2010	4	37.76	37.67136647	0.09	0.01	
19				11	01-05-2010	5	38.88	39.30524076	-0.42	0.18	
20				12	01-06-2010	6	41.90	41.85750467	0.04	0.00	
21				13	01-07-2010	7	44.02	44.4280508	-0.41	0.16	
22				14	01-08-2010	8	42.81	42.65279779	0.16	0.03	

Q 3] Using the first 20 quarters for the data Ratioma, forecast sales for the following four quarters.

Step 1. In cell F1 use the formula =SLOPE(G7:G22,B7:B22) to find the slope of the trend line.

Step 2. In cell F2 use the formula =INTERCEPT(G7:G22,B7:B22) to find the intercept of the trend line.

Step 3. Estimate the level of the series during Quarter t to be 6.94t c + 30.17.

Step 4. Copy the formula =intercept + slope*B25 down from cell G25 to G26:G28 to compute the estimated level (excluding seasonality) of the series from Quarter 21 onward.

Step 5. Compute the Seasonal Indexes
Copy the formula
=E7/G7 down from cell H7 to H8:H22. You find, for example, that during Quarter 1 sales were 77 percent, 71 percent, 90 percent and 89 percent of average, so you could estimate the seasonal index for Quarter 1 as the average of these four numbers (82 percent). To calculate the initial seasonal index estimates, you can copy the formula =AVERAGEIF($D$7:$D$22,J3,$H$7:$H$22) from cell K3 to K4:K6. This formula averages the four estimates you have for Q1 seasonality.

Step 6. To ensure that your final seasonal indices average to 1, copy the formula =K3/AVERAGE($K$3:$K$6) from cell L3 to L4:L6.

Step 7. To create your sales forecast for each future quarter, simply multiply the trend line estimate for the quarter's level (from Column G) by the appropriate seasonal index. Copy the formula =VLOOKUP(D25,season,3)*G25 from cell G25 to G26:G28 to compute the final forecast for Quarters 21–24. This forecast includes estimates of trend and seasonality.

Step 8. Use the centered moving averages for Quarters 13–18 to get a more recent trend estimate by using the formula =SLOPE(G17:G22,B17:B22).

	A	B	C	D	E	F	G	H	I	J	K	L
1					slope	6.9387868						
2					intercept	30.166176				quarter	seasonal index	normalized
3										1	0.818547	0.81373678
4		Quarter#	Year	Quarter	Sales	4 period MA	Centered MA	Actual/CMA	Forecast	2	0.93934	0.9338196
5		1	1	1	24					3	1.067364	1.06109143
6		2	1	2	44	52				4	1.198394	1.19135219
7		3	1	3	61	58	55.00	1.11				
8		4	1	4	79	63.5	60.75	1.30				
9		5	2	1	48	71	67.25	0.71				
10		6	2	2	66	77.5	74.25	0.89				
11		7	2	3	91	82.5	80.00	1.14				
12		8	2	4	105	87.25	84.88	1.24				
13		9	3	1	68	89.5	88.38	0.77				
14		10	3	2	85	94.5	92.00	0.92				
15		11	3	3	100	104.25	99.38	1.01				
16		12	3	4	125	114.25	109.25	1.14				
17		13	4	1	107	123.75	119.00	0.90				
18		14	4	2	125	132.25	128.00	0.98				
19		15	4	3	138	139.25	135.75	1.02				
20		16	4	4	159	146.75	143.00	1.11				
21		17	5	1	135	156	151.38	0.89				
22		18	5	2	155	164.25	160.13	0.97				
23		19	5	3	175							
24		20	5	4	192							

23	19	5	3	175			
24	20	5	4	192			
25	21	6	1		175.880699	143.121	
26	22	6	2		182.819485	170.72	
27	23	6	3		189.758272	201.351	
28	24	6	4		196.697059	234.335	

Q 4] Demonstrate Winters method usin the airline winters data.

Step 1. In H11:J11, enter trial values (between 0 and 1) for the smoothing constants.

Step 2. In C26:C113, compute the updated series level with Equation 1 by copying the formula =alp*(B26/H14)+(1–alp)*(C25*D25) from cell C26 to C27:C113.

Step 3. In D26:D113, update the series trend. Copy the formula =bet*(C26/C25)+(1-bet)*D25 cell from D26 to D27:D113.

Step 4. In H26:H113, update the seasonal indices. Copy the formula =gam*(B26/C26)+(1-gam)*H14 from cell H26 to H27:H113.

Step 5. In E26:E113, compute the forecast for the current month by copying the formula =(C25*D25)*H14 from cell E26 to E27:E113.

Step 6. In F26:F113 compute each month's error by copying the formula =(B26-E26) from cell E26 to E27:E113.

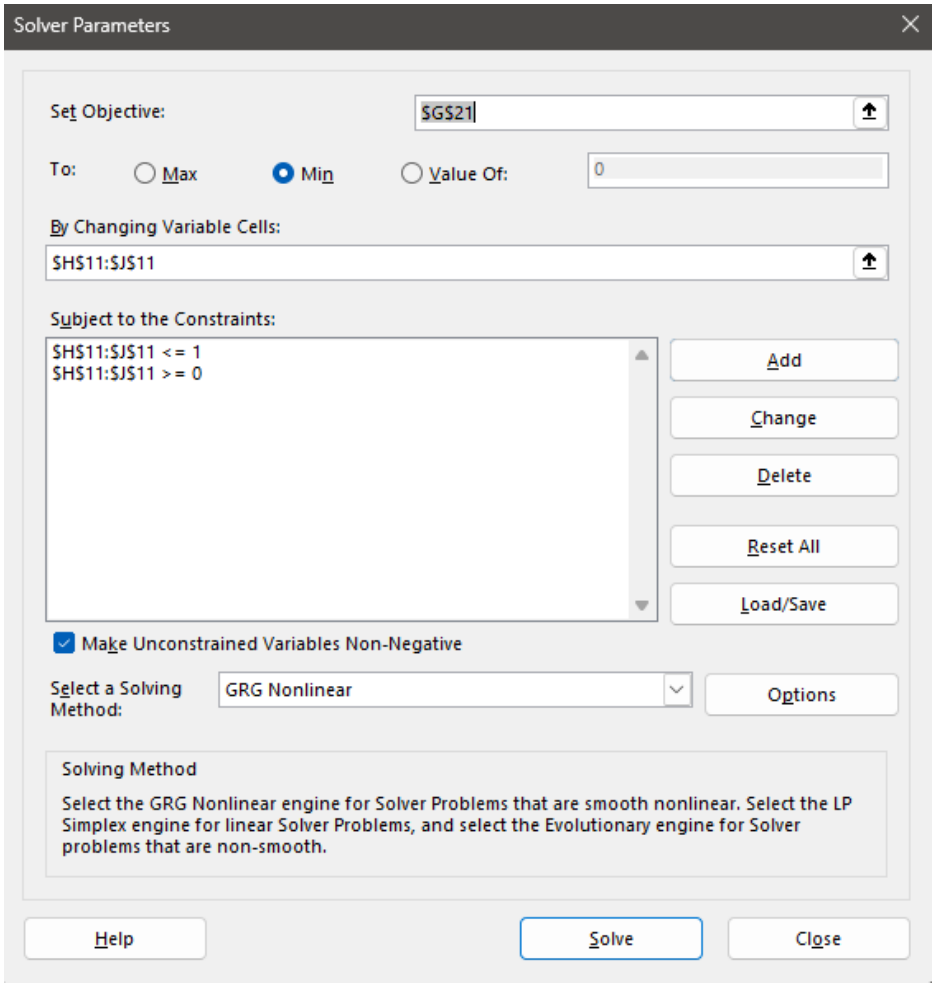
Step 7. In G26:G113, compute the squared error for each month by copying the formula =F26^2 from cell F26 to F27:F113. In cell G21 compute the Sum of Squared Errors (SSE) using the formula =SUM(G26:G113).

Step 8. Now use the Solver to determine smoothing parameter values that minimize SSE.

Step 9. Choose the smoothing parameters (H11:J11) to minimize SSE (cell G21). The Excel Solver ensures you can find the best combination of smoothing constants. Smoothing constants must be α The Solver finds that alp = 0.55, bet = 0.05, and gamma = 0.59.

	A	B	C	D	E	F	G	H	I	J	K	L
1	DATE	Airline Miles(billions)										
2	Jan2003	32.85										
3	Feb2003	30.81										
4	Mar2003	37.59										
5	Apr2003	35.23										
6	May2003	36.57										
7	Jun2003	39.75										
8	Jul2003	43.37										
9	Aug2003	42.09										
10	Sep2003	32.55						alp	bet	gam		
11	Oct2003	36.44						0.548512014	0.049142462	0.5888		
12	Nov2003	34.35										
13	Dec2003	37.39						Seasonal Indices				
14	Jan2004	33.54						0.903049602				
15	Feb2004	33.91						0.875947455				
16	Mar2004	40.81						1.053814727				
17	Apr2004	40.17						1.00800602				
18	May2004	39.67						1.011705575				
19	Jun2004	43.65						1.099865309				
20	Jul2004	46.26						1.173705527				
21	Aug2004	44.70				SSE	77.8196	1.129148564				
22	Sep2004	35.47				stdeverrors	0.9369659	0.879643964		34 sign changes of 87		
23	Oct2004	39.63						0.977358616		MAPE		
24	Nov2004	37.57	Base	Trend	Forecast	Error	Sq Eerror	0.918146041	34	0.0206		

	A	B	C	D	E	F	G	H	I	J	K
1											
2	base	3.51E+01							SSE		
3	trend	1.006491							5.63		
4			MonthNu	Month	Month	AirlineMiles (billions)	Forecast	Error	Sq Error		
5	1	0.90305	1	Jan2003	1	32.85	31.87289	0.98	0.96		
6	2	0.875947	2	Feb2003	2	30.81	31.11699	-0.30	0.09		
7	3	1.053815	3	Mar2003	3	37.59	37.6785	-0.09	0.01		
8	4	1.008006	4	Apr2003	4	35.23	36.27456	-1.05	1.10		
9	5	1.011706	5	May2003	5	36.57	36.644	-0.07	0.01		
10	6	1.099865	6	Jun2003	6	39.75	40.09572	-0.35	0.12		
11	7	1.173706	7	Jul2003	7	43.37	43.06529	0.30	0.09		
12	8	1.129149	8	Aug2003	8	42.09	41.69932	0.39	0.15		
13	9	0.879644	9	Sep2003	9	32.55	32.696	-0.15	0.02		
14	10	0.977359	10	Oct2003	10	36.44	36.5638	-0.12	0.01		
15	11	0.918146	11	Nov2003	11	34.35	34.57155	-0.22	0.05		
16	12	0.969609	12	Dec2003	12	37.39	36.74627	0.64	0.41		
17			13	Jan2004	1	33.54	34.44595	-0.91	0.83		
18	mean	1	14	Feb2004	2	33.91	33.62903	0.28	0.08		
19			15	Mar2004	3	40.81	40.72023	0.08	0.01		
20			16	Apr2004	4	40.17	39.20296	0.97	0.94		
21			17	May2004	5	39.67	39.60222	0.07	0.00		
22			18	Jun2004	6	43.65	43.33259	0.32	0.10		
23			19	Jul2004	7	46.26	46.54189	-0.28	0.08		



Step 10. Copying this formula down to cells D117:D123 enables you to forecast sales for the months of May through December of 2012.

Step 11. By copying the formula =($C$113*$D$113^B116)*H102 from cell D116 to D117:D123. Cell D124 adds up these forecasts and predicts the rest of 2012 to see 314.17 billion airline miles travelled.

	A	B	C	D	E	F	G	H	I	J	K
99	Feb2011	31.83	38.26068	0.999343	33.14054	-1.32	1.7304	0.838179411	0	0.0413	
100	Mar2011	40.51	38.18969	0.999285	40.59559	-0.09	0.0079	1.061105008	0	0.0022	
101	Apr2011	38.51	38.09331	0.999196	38.63321	-0.13	0.0162	1.011448481	0	0.0033	
102	May2011	40.43	38.7357	1.000064	39.16698	1.26	1.5942	1.03767775	1	0.0312	
103	Jun2011	42.57	38.84163	1.000195	42.36397	0.21	0.0425	1.095009094	0	0.0048	
104	Jul2011	45.07	39.00168	1.000388	44.75388	0.32	0.1025	1.154171797	0	0.0071	
105	Aug2011	42.78	38.66139	0.99994	43.50483	-0.72	0.5220	1.110060016	1	0.0169	
106	Sep2011	36.70	39.65425	1.001205	35.05387	1.65	2.7064	0.917771587	1	0.0448	
107	Oct2011	38.70	39.72586	1.001235	38.66143	0.04	0.0018	0.974072493	0	0.0011	
108	Nov2011	36.83	40.16433	1.001716	36.182	0.65	0.4171	0.913943328	0	0.0175	
109	Dec2011	37.49	39.82658	1.001219	38.1972	-0.70	0.4955	0.944695094	1	0.0188	
110	Jan2012	34.31	39.32558	1.000541	35.19792	-0.88	0.7821	0.876725748	0	0.0258	
111	Feb2012	33.26	39.53299	1.000773	32.97971	0.28	0.0809	0.840092107	1	0.0086	
112	Mar2012	40.78	38.94334	1.000002	41.9811	-1.20	1.4396	1.05291499	1	0.0294	
113	Apr2012	38.81	38.6274	0.999604	39.38927	-0.58	0.3396	1.00743814	0	0.015	
114			Base	Trend	Forecast	Error	Sq Eerror	Seasonal Indices			
115			forecasts								
116		1	May-12	40.06691							
117		2	Jun-12	42.26383							
118		3	Jul-12	44.52966							
119		4	Aug-12	42.81079							
120		5	Sep-12	35.38093							
121		6	Oct-12	37.53649							
122		7	Nov-12	35.20542							
123		8	Dec-12	36.37556							
124			total	314.1696							

Practical – 5

Conjoint Analysis & Logistic Regression

Q 1] Determine how various attributes impact the purchase of a car. There are four attributes, each with three levels:

· Brand: Ford = 0, Chrysler = 1, GM = 2

· MPG: 15 MPG = 0, 20 MPG = 1, 25 MPG = 2

· Horsepower (HP): 100 HP = 0, 150 HP = 1, 200 HP = 2

· Price: $18,000 = 0, $21,000 = 1, $24,000 = 2

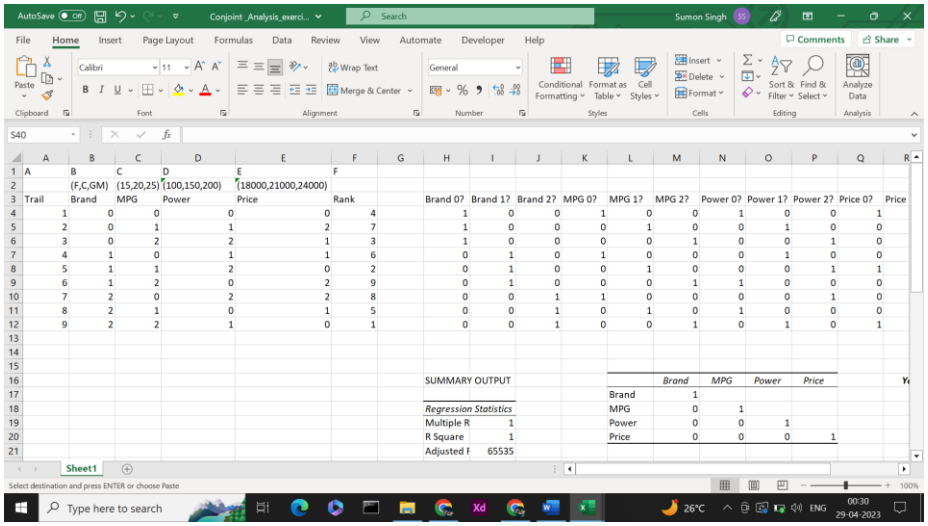
The nine product profiles ranked in Figure below were evaluated by a consumer.

	A	B	C	D	E	F
1		(F,C,GM)	(15,20,25)	(100,150,200)	(18000,21000,24000)	
2	Trial	Brand	MPG	Power	Price	Rank
3	1	0	0	0	0	4
4	2	0	1	1	2	7
5	3	0	2	2	1	3
6	4	1	0	1	1	6
7	5	1	1	2	0	2
8	6	1	2	0	2	9
9	7	2	0	2	2	8
10	8	2	1	0	1	5
11	9	2	2	1	0	1

For this market segment, rank the product attributes from most important to least important –

Answer :

		Ranking
Brand	5.1368E-16	3rd
MPG	1.66666667	2nd
Power	1.66666667	2nd
Price	5.66666667	1st



A	B	C	D	E	F
	(F,C,GM)	(15,20,25)	(100,150,200)	(18000,21000,24000)	
Trail	Brand	MPG	Power	Price	Rank
1	0	0	0	0	4
2	0	1	1	2	7
3	0	2	2	1	3
4	1	0	1	1	6
5	1	1	2	0	2
6	1	2	0	2	9
7	2	0	2	2	8
8	2	1	0	1	5
9	2	2	1	0	1

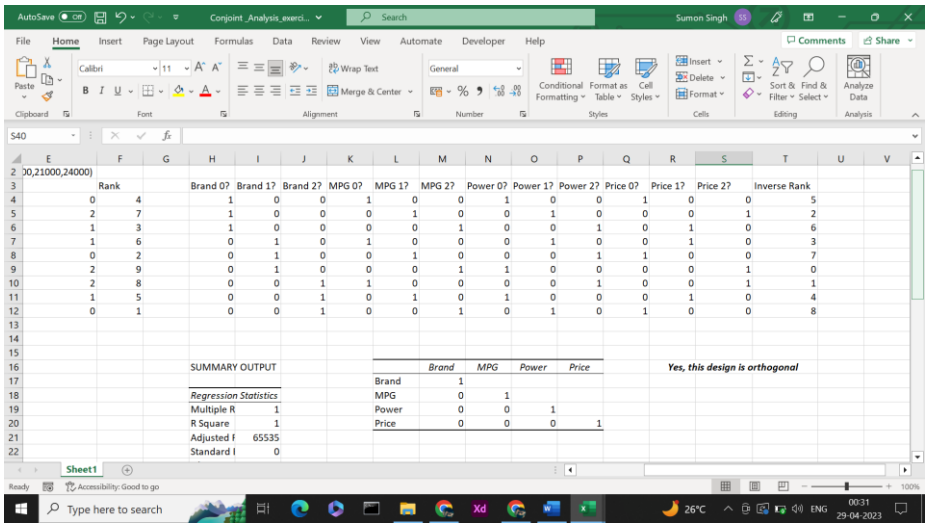
Brand 0?	Brand 1?	Brand 2?	MPG 0?	MPG 1?	MPG 2?	Power 0?	Power 1?	Power 2?	Price 0?	Price
1	0	0	1	0	0	1	0	0	1	
1	0	0	0	1	0	0	1	0	0	
1	0	0	0	0	1	0	0	1	0	
0	1	0	1	0	0	0	0	1	0	
0	1	0	0	0	1	0	0	0	1	
0	1	0	0	1	0	1	0	0	0	
0	0	1	0	0	1	0	1	0	0	
0	0	1	0	1	0	0	0	1	0	
0	0	1	0	0	1	0	1	0	1	

SUMMARY OUTPUT

Regression Statistics

Multiple R	1
R Square	1
Adjusted F	65535

	Brand	*MPG*	*Power*	*Price*
Brand	1			
MPG	0	1		
Power	0	0	1	
Price	0	0	0	1



Rank	Brand 0?	Brand 1?	Brand 2?	MPG 0?	MPG 1?	MPG 2?	Power 0?	Power 1?	Power 2?	Price 0?	Price 1?	Price 2?	Inverse Rank
4	1	0	0	1	0	0	1	0	0	1	0	0	5
7	1	0	0	0	1	0	0	1	0	0	1	0	2
3	1	0	0	0	0	1	0	0	1	0	1	0	6
6	0	1	0	1	0	0	0	0	1	0	1	0	3
2	0	1	0	0	1	0	0	0	1	1	0	0	7
9	0	1	0	0	0	1	1	0	0	0	0	1	0
8	0	0	1	1	0	0	0	0	1	0	0	1	1
5	0	0	1	0	1	0	1	0	0	0	1	0	4
1	0	0	1	0	0	1	0	1	0	1	0	0	8

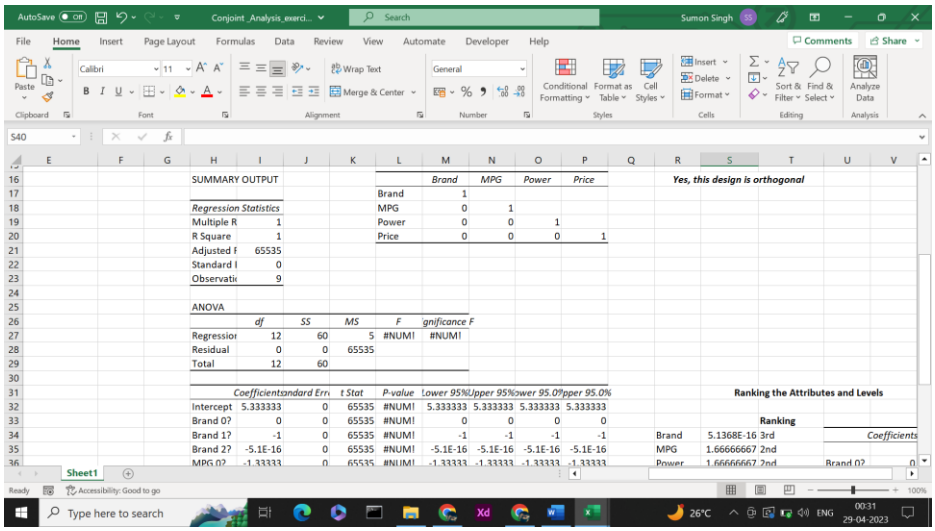
SUMMARY OUTPUT

Regression Statistics

Multiple R	1
R Square	1
Adjusted F	65535
Standard E	0

	Brand	*MPG*	*Power*	*Price*
Brand	1			
MPG	0	1		
Power	0	0	1	
Price	0	0	0	1

Yes, this design is orthogonal



SUMMARY OUTPUT

Regression Statistics

Multiple R	1
R Square	1
Adjusted F	65535
Standard E	0
Observatio	9

ANOVA

	df	*SS*	*MS*	*F*	*gnificance F*
Regressior	12	60	5	#NUM!	#NUM!
Residual	0	0	65535		
Total	12	60			

	Coefficients	*tandard Erro*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	5.333333	0	65535	#NUM!	5.333333	5.333333	5.333333	5.333333
Brand 0?	0	0	65535	#NUM!	0	0	0	0
Brand 1?	-1	0	65535	#NUM!	-1	-1	-1	-1
Brand 2?	-5.1E-16	0	65535	#NUM!	-5.1E-16	-5.1E-16	-5.1E-16	-5.1E-16
MPG 0?	-1.33333	0	65535	#NUM!	-1.33333	-1.33333	-1.33333	-1.33333

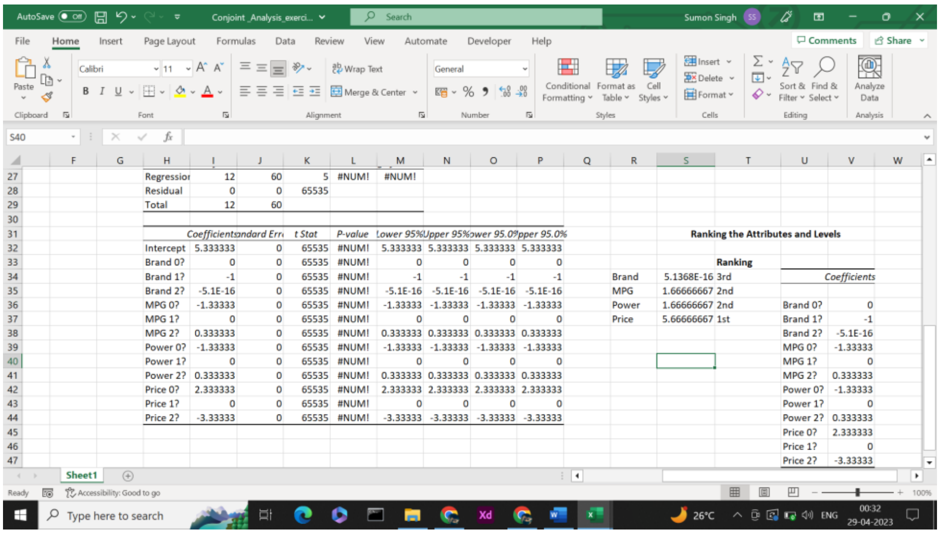
	Brand	*MPG*	*Power*	*Price*
Brand	1			
MPG	0	1		
Power	0	0	1	
Price	0	0	0	1

Yes, this design is orthogonal

Ranking the Attributes and Levels

Ranking

			Coefficients
Brand	5.1368E-16	3rd	
MPG	1.66666667	2nd	
Power	1.66666667	2nd	
Brand 0?			0



Regressior	12	60	5	#NUM!	#NUM!
Residual	0	0	65535		
Total	12	60			

	Coefficients	tandard Err	t Stat	P-value	Lower 95%	Upper 95%	ower 95.0%	pper 95.0%
Intercept	5.333333	0	65535	#NUM!	5.333333	5.333333	5.333333	5.333333
Brand 0?	0	0	65535	#NUM!	0	0	0	0
Brand 1?	-1	0	65535	#NUM!	-1	-1	-1	-1
Brand 2?	-5.1E-16	0	65535	#NUM!	-5.1E-16	-5.1E-16	-5.1E-16	-5.1E-16
MPG 0?	-1.33333	0	65535	#NUM!	-1.33333	-1.33333	-1.33333	-1.33333
MPG 1?	0	0	65535	#NUM!	0	0	0	0
MPG 2?	0.333333	0	65535	#NUM!	0.333333	0.333333	0.333333	0.333333
Power 0?	-1.33333	0	65535	#NUM!	-1.33333	-1.33333	-1.33333	-1.33333
Power 1?	0	0	65535	#NUM!	0	0	0	0
Power 2?	0.333333	0	65535	#NUM!	0.333333	0.333333	0.333333	0.333333
Price 0?	2.333333	0	65535	#NUM!	2.333333	2.333333	2.333333	2.333333
Price 1?	0	0	65535	#NUM!	0	0	0	0
Price 2?	-3.33333	0	65535	#NUM!	-3.33333	-3.33333	-3.33333	-3.33333

Ranking the Attributes and Levels

		Ranking
Brand	5.1368E-16	3rd
MPG	1.66666667	2nd
Power	1.66666667	2nd
Price	5.66666667	1st

	Coefficients
Brand 0?	0
Brand 1?	-1
Brand 2?	-5.1E-16
MPG 0?	-1.33333
MPG 1?	0
MPG 2?	0.333333
Power 0?	-1.33333
Power 1?	0
Power 2?	0.333333
Price 0?	2.333333
Price 1?	0
Price 2?	-3.33333

Is this design orthogonal?

	Brand	*MPG*	*Power*	*Price*
Brand	1			
MPG	0	1		
Power	0	0	1	
Price	0	0	0	1

Yes, this design is orthogonal

Q 2] The soda.xlsx file (see Figure below) gives a consumer's ranking on an orthogonal design with 12 product profiles involving a comparison of Coke and Pepsi. The attributes and levels are as follows:

· **Brand: Coke or Pepsi**

· **Packaging: 12 oz. can or 16 oz. bottle**

· **Price per ounce: 8 cents, 10 cents, 12 cents**

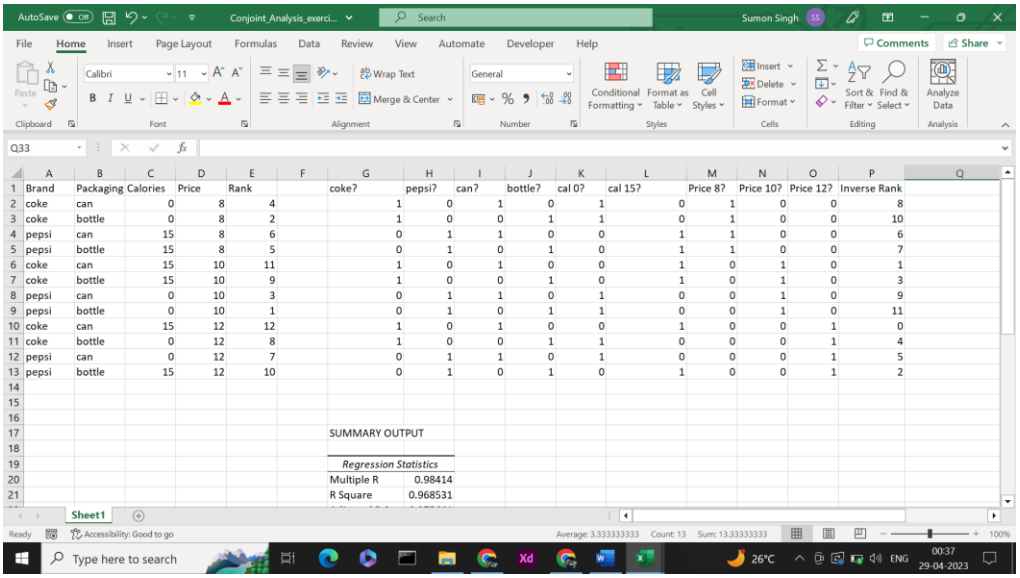
· **Calories per ounce: 0 or 15**

	F	G	H	I	J	K
1						
2				per oz	per oz	
3	Profile	Brand	Packaging	Calories	Price	Rank
4	1	coke	can	0	8	4
5	2	coke	bottle	0	8	2
6	3	pepsi	can	15	8	6
7	4	pepsi	bottle	15	8	5
8	5	coke	can	15	10	11
9	6	coke	bottle	15	10	9
10	7	pepsi	can	0	10	3
11	8	pepsi	bottle	0	10	1
12	9	coke	can	15	12	12
13	10	coke	bottle	0	12	8
14	11	pepsi	can	0	12	7
15	12	pepsi	bottle	15	12	10

Determine the ranking of the attributes' importance and the ranking of all attribute levels.

Answer :

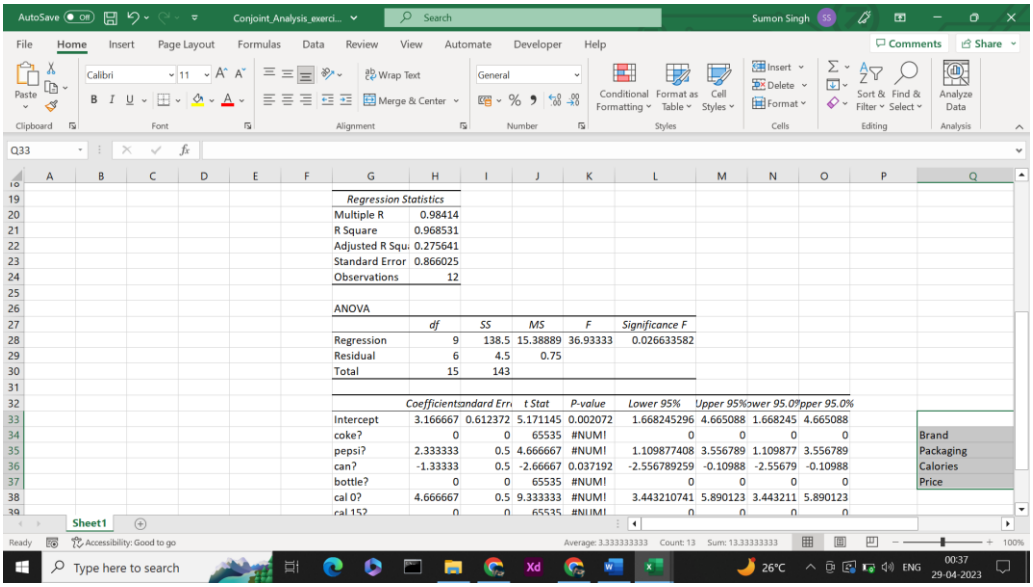
		Ranking
Brand	2.333333	3rd
Packaging	1.333333	4th
Calories	4.666667	2nd
Price	5	1st



Brand	Packaging	Calories	Price	Rank	coke?	pepsi?	can?	bottle?	cal 0?	cal 15?	Price 8?	Price 10?	Price 12?	Inverse Rank
coke	can	0	8	4	1	0	1	0	1	0	1	0	0	8
coke	bottle	0	8	2	1	0	0	1	1	0	1	0	0	10
pepsi	can	15	8	6	0	1	1	0	0	1	1	0	0	6
pepsi	bottle	15	8	5	0	1	0	1	0	1	1	0	0	7
coke	can	15	10	11	1	0	1	0	0	1	0	1	0	1
coke	bottle	15	10	9	1	0	0	1	0	1	0	1	0	3
pepsi	can	0	10	3	0	1	1	0	1	0	0	1	0	9
pepsi	bottle	0	10	1	0	1	0	1	1	0	0	1	0	11
coke	can	15	12	12	1	0	1	0	0	1	0	0	1	0
coke	bottle	0	12	8	1	0	0	1	1	0	0	0	1	4
pepsi	can	0	12	7	0	1	1	0	1	0	0	0	1	5
pepsi	bottle	15	12	10	0	1	0	1	0	1	0	0	1	2

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.98414
R Square	0.968531



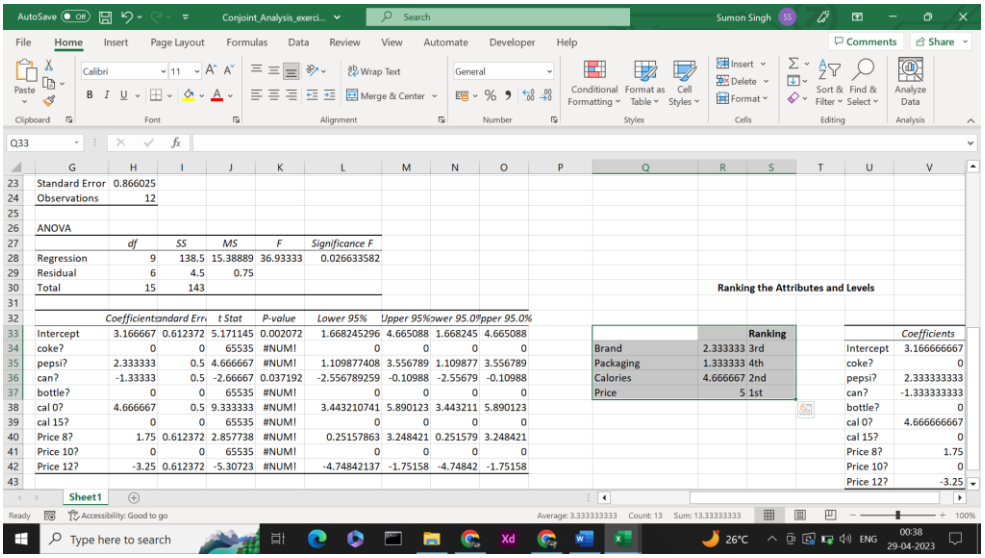
Regression Statistics	
Multiple R	0.98414
R Square	0.968531
Adjusted R Squ	0.275641
Standard Error	0.866025
Observations	12

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	138.5	15.38889	36.93333	0.026633582
Residual	6	4.5	0.75		
Total	15	143			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	3.166667	0.612372	5.171145	0.002072	1.668245296	4.665088	1.668245	4.665088
coke?	0	0	65535	#NUM!	0	0	0	0
pepsi?	2.333333	0.5	4.666667	#NUM!	1.109877408	3.556789	1.109877	3.556789
can?	-1.33333	0.5	-2.66667	0.037192	-2.556789259	-0.10988	-2.55679	-0.10988
bottle?	0	0	65535	#NUM!	0	0	0	0
cal 0?	4.666667	0.5	9.333333	#NUM!	3.443210741	5.890123	3.443211	5.890123
cal 15?	0	0	65535	#NUM!	0	0	0	0

Brand
Packaging
Calories
Price



	G	H	I	J	K	L	M	N	O
23	Standard Error	0.866025							
24	Observations	12							
25									
26	ANOVA								
27		*df*	*SS*	*MS*	*F*	*Significance F*			
28	Regression	9	138.5	15.38889	36.93333	0.026633582			
29	Residual	6	4.5	0.75					
30	Total	15	143						
31									
32		*Coefficients*	*tandard Erri*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
33	Intercept	3.166667	0.612372	5.171145	0.002072	1.668245296	4.665088	1.668245	4.665088
34	coke?	0	0	65535	#NUM!	0	0	0	0
35	pepsi?	2.333333	0.5	4.666667	#NUM!	1.109877408	3.556789	1.109877	3.556789
36	can?	-1.33333	0.5	-2.66667	0.037192	-2.556789259	-0.10988	-2.55679	-0.10988
37	bottle?	0	0	65535	#NUM!	0	0	0	0
38	cal 0?	4.666667	0.5	9.333333	#NUM!	3.443210741	5.890123	3.443211	5.890123
39	cal 15?	0	0	65535	#NUM!	0	0	0	0
40	Price 8?	1.75	0.612372	2.857738	#NUM!	0.25157863	3.248421	0.251579	3.248421
41	Price 10?	0	0	65535	#NUM!	0	0	0	0
42	Price 12?	-3.25	0.612372	-5.30723	#NUM!	-4.74842137	-1.75158	-4.74842	-1.75158
43									

Ranking the Attributes and Levels

	Ranking
Brand	2.333333 3rd
Packaging	1.333333 4th
Calories	4.666667 2nd
Price	5 1st

	Coefficients
Intercept	3.166666667
coke?	0
pepsi?	2.333333333
can?	-1.333333333
bottle?	0
cal 0?	4.666666667
cal 15?	0
Price 8?	1.75
Price 10?	0
Price 12?	-3.25

Q 3] For 1,022 NFL field goal attempts the file FGdata.xlsx contains the distance of the field goal and whether the field goal was good. How does the length of the field goal attempt affect the chance to make the field goal? Estimate the chance of making a 30, 40, or 50 yard field goal.

Step 1. Select trail intercept (cell F1) & slope (cell F2) values

Step 2. Create a "score" column based on the formula intercept + slope * length (cell B) column

Step 3. Create a "prob" column based on the formula exp(score)/(1+exp(score))

Step 4. Create a column "1-prob"

Step 5. Create a "likelihood" column based on the formula if Good = 1 then select prob else select 1-prob

Step 6. Create "ln likelihood" using formula LN(likelihood) [ln function]

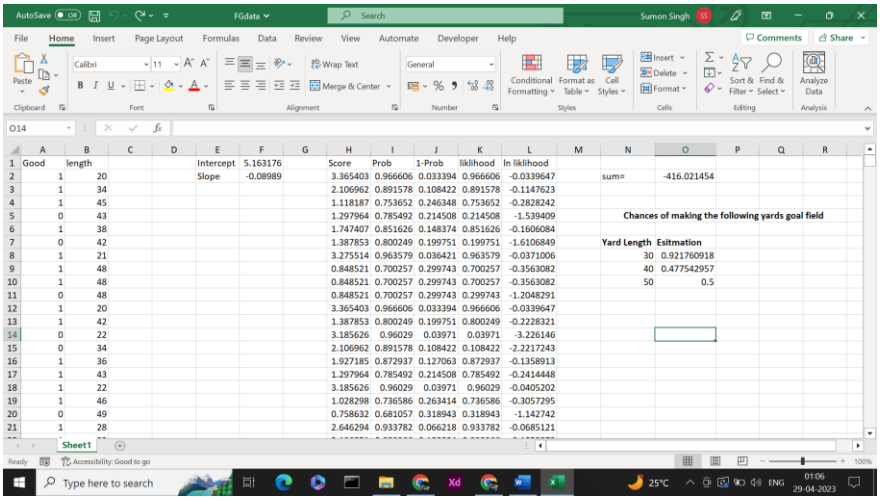
Step 7. Calculate sum of all the values of "ln likelihood" column to compute the Log Likelihood.

Step 8. Use the Solver window, to determine the slope and intercept that maximize the Log Likelihood.

Step 9. The maximum likelihood estimates of Intercept = 5.163176 and slope = -0.08989

Step 10. Using the formula : EXP(F1+F2*Yard Length)/(1+EXP(F1+F2*Yard Length)) estimate the chance of making the field goal

Chances of making the following yards goal field				
Yard Length	**Esitmation**			
30	0.921760918			
40	0.477542957			
50	0.5			



Q 4] The file Logitsubscribedata.xls gives the number of people in each age group who subscribe and do not subscribe to a magazine. How does age influence the chance of subscribing to the magazine?

Step 1. Select trail intercept (cell F1) Age (cell F2) and Gender (cell F3) values

Step 2. Create a "score" column based on the formula intercept + Age(F2) * Age column(A) + Gender(F3)*Gender column (B)

Step 3. Create a "prob" column based on the formula exp(score)/(1+exp(score))

Step 4. Create a column "1-prob"

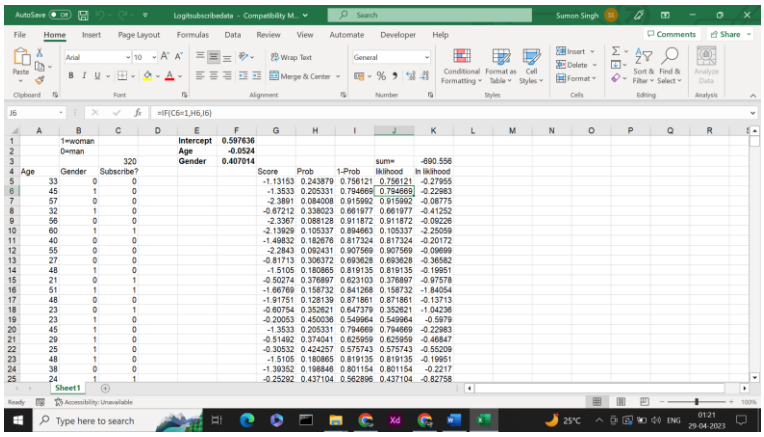
Step 5. Create a "likelihood" column based on the formula if Subscribe? (column C) = 1 then select prob else select 1-prob

Step 6. Create "ln likelihood" using formula LN(likelihood) [ln function]

Step 7. Calculate sum of all the values of "ln likelihood" column to compute the Log Likelihood.

Step 8. Use the Solver window, to determine the slope and intercept that maximize the Log Likelihood.

Step 9. The maximum likelihood estimates of Age = -0.0524 , Gender = 0.407014 and Intercept = 0.597636



	A	B	C	D	E	F	G	H	I	J	K
1		1=woman			Intercept	0.597636					
2		0=man			Age	-0.0524					
3			320		Gender	0.407014				sum=	-690.556
4	Age	Gender	Subscribe?				Score	Prob	1-Prob	likihood	ln likihood
5	33	0	0				-1.13153	0.243879	0.756121	0.756121	-0.27955
6	45	1	0				-1.3533	0.205331	0.794669	0.794669	-0.22983
7	57	0	0				-2.3891	0.084008	0.915992	0.915992	-0.08775
8	32	1	0				-0.67212	0.338023	0.661977	0.661977	-0.41252
9	56	0	0				-2.3367	0.088128	0.911872	0.911872	-0.09226
10	60	1	1				-2.13929	0.105337	0.894663	0.105337	-2.25059
11	40	0	0				-1.49832	0.182676	0.817324	0.817324	-0.20172
12	55	0	0				-2.2843	0.092431	0.907569	0.907569	-0.09699
13	27	0	0				-0.81713	0.306372	0.693628	0.693628	-0.36582
14	48	1	0				-1.5105	0.180865	0.819135	0.819135	-0.19951
15	21	0	1				-0.50274	0.376897	0.623103	0.376897	-0.97578
16	51	1	1				-1.66769	0.158732	0.841268	0.158732	-1.84054
17	48	0	0				-1.91751	0.128139	0.871861	0.871861	-0.13713
18	23	0	1				-0.60754	0.352621	0.647379	0.352621	-1.04236
19	23	1	0				-0.20053	0.450036	0.549964	0.549964	-0.5979
20	45	1	0				-1.3533	0.205331	0.794669	0.794669	-0.22983
21	29	1	0				-0.51492	0.374041	0.625959	0.625959	-0.46847
22	25	1	0				-0.30532	0.424257	0.575743	0.575743	-0.55209
23	48	1	0				-1.5105	0.180865	0.819135	0.819135	-0.19951
24	38	0	0				-1.39352	0.198846	0.801154	0.801154	-0.2217
25	24	1	1				-0.25292	0.437104	0.562896	0.437104	-0.82758

Q 5] The file Healthcaredata.xlsx gives the age, party affiliation, and income of 300people. You are also told whether they favour Obamacare. Develop a model to predict the chance that a person favours Obamacare. For each person generate a prediction of whether the person favours Obamacare. Interpret the coefficients of the independent variables in your logistic regression.

Step 1. Select trail intercept (cell L2) , Age (cell L4) , Party (cell L3) and Income (cell L5) values

Step 2. Create a "score" column based on the formula intercept + Age(L4) * Age column(D) + Income(L5)*Income column (E) + Party(L3)*Party_Num column(C)

Step 3. Create a "prob" column based on the formula exp(score)/(1+exp(score))

Step 4. Create a column "1-prob"

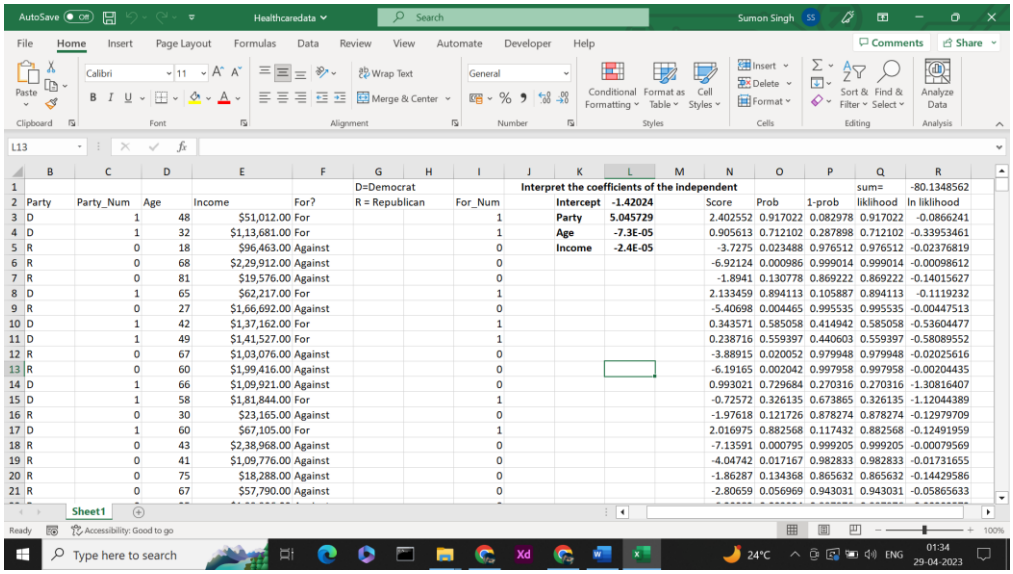
Step 5. Create a "likelihood" column based on the formula if For_Num? (column C) = 1 then select prob else select 1-prob

Step 6. Create "ln likelihood" using formula LN(likelihood) [ln function]

Step 7. Calculate sum of all the values of "ln likelihood" column to compute the Log Likelihood.

Step 8. Use the Solver window, to determine the slope and intercept that maximize the Log Likelihood.

Step 9. The maximum likelihood estimates of Age = -7.3E-05 , Income = -2.4E-0.5, Party = 5.0457289721412 and Intercept = -1.42023816338202



	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1						D=Democrat			Interpret the coefficients of the independent							sum=	-80.1348562
2	Party	Party_Num	Age	Income	For?	R = Republican		For_Num		Intercept	-1.42024		Score	Prob	1-prob	liklihood	ln liklihood
3	D	1	48	$51,012.00	For			1		Party	5.045729		2.402552	0.917022	0.082978	0.917022	-0.0866241
4	D	1	32	$1,13,681.00	For			1		Age	-7.3E-05		0.905613	0.712102	0.287898	0.712102	-0.33953461
5	R	0	18	$96,463.00	Against			0		Income	-2.4E-05		-3.7275	0.023488	0.976512	0.976512	-0.02376819
6	R	0	68	$2,29,912.00	Against			0					-6.92124	0.000986	0.999014	0.999014	-0.00098612
7	R	0	81	$19,576.00	Against			0					-1.8941	0.130778	0.869222	0.869222	-0.14015627
8	D	1	65	$62,217.00	For			1					2.133459	0.894113	0.105887	0.894113	-0.1119232
9	R	0	27	$1,66,692.00	Against			0					-5.40698	0.004465	0.995535	0.995535	-0.00447513
10	D	1	42	$1,37,162.00	For			1					0.343571	0.585058	0.414942	0.585058	-0.53604477
11	D	1	49	$1,41,527.00	For			1					0.238716	0.559397	0.440603	0.559397	-0.58089552
12	R	0	67	$1,03,076.00	Against			0					-3.88915	0.020052	0.979948	0.979948	-0.02025616
13	R	0	60	$1,99,416.00	Against			0					-6.19165	0.002042	0.997958	0.997958	-0.00204435
14	D	1	66	$1,09,921.00	Against			0					0.993021	0.729684	0.270316	0.270316	-1.30816407
15	D	1	58	$1,81,844.00	For			1					-0.72572	0.326135	0.673865	0.326135	-1.12044389
16	R	0	30	$23,165.00	Against			0					-1.97618	0.121726	0.878274	0.878274	-0.12979709
17	D	1	60	$67,105.00	For			1					2.016975	0.882568	0.117432	0.882568	-0.12491959
18	R	0	43	$2,38,968.00	Against			0					-7.13591	0.000795	0.999205	0.999205	-0.00079569
19	R	0	41	$1,09,776.00	Against			0					-4.04742	0.017167	0.982833	0.982833	-0.01731655
20	R	0	75	$18,288.00	Against			0					-1.86287	0.134368	0.865632	0.865632	-0.14429586
21	R	0	67	$57,790.00	Against			0					-2.80659	0.056969	0.943031	0.943031	-0.05865633

Practical – 6

Calculating Lifetime Customer Value

1. **AOL bought customers from CD Now for $60 per customer. CD Now's annual retention rate was 60 percent, and customers generated $15 of profit per year. Assuming an annual discount rate of 12 percent, evaluate AOL's purchase of CD Now customers**.

Retention Rate	0.6
Discount Rate	0.12
Profit	15

The formula to calculate the CLV is given below CLV = Gross margin * (Retention rate / (1+ Rate of discount - Retention rate))
CLV = $15 * (0.60 / (1+0.12 - 0.60))
CLV = $15 * (0.60 / {1.12 - 0.60}
CLV = $15 * (0.60 / 0.52)
CLV = $15 * 1.15385
CLV=$17.31
The customer lifetime value is $17.31
The purchase cost of customers is $60.00
Net Profit/ Loss on customer acquisition = customer lifetime value – purchase cost of customer
= 17.31 – 60.00 = -42.69 (Loss)

2. **It costs Ameritrade $203 to acquire a customer. Ameritrade earned $200 per year from a customer and had an annual retention rate of 95 percent. Assuming cash flows are discounted at 12 percent, estimate the value of a customer and the net of acquisition costs**.

Retention Rate	0.95
Discount Rate	0.12
Profit	-3

The formula to calculate the CLV is given below CLV = Gross margin * (Retention rate / (1+ Rate of discount - Retention rate))

CLV = $(-3*(0.95/(1+0.12-0.95)))
CLV = -$16.7647

The customer lifetime value is -$16.7647
The purchase cost of customers is $203
Net Profit/ Loss on customer acquisition = customer lifetime value – purchase cost of customer
= -$16.7647 - $203
= -$219.765 (Loss)

3. **Assume the annual retention rate for a cell phone subscriber is 70 percent and the customer generates $300 per year in profit. Assuming an annual discount rate of 8 percent, compute the value of a customer**

Retention Rate	0.7
Discount Rate	0.8
Profit	300

The formula to calculate the CLV is given below CLV = Gross margin * (Retention rate / (1+ Rate of discount - Retention rate))

CLV = $(300*(0.7/(1+0.8-0.7)))
CLV = $190.9091
The customer lifetime value is $190.9091

4. **Assume a customer has been with a company for 10 years and has made purchases at times 0.2, 1.2, 0.8, and 3. Estimate the probability the customer is still active.**

t (Time Of Last Purchase)	**3**
T (Time elapsed between acquisition of customer and present time)	**10**
N (Number Of Purchases)	**4**

Formula : T* = t/T = 0.3
Customer is still active probability : (T*)^N = 0.3^4 =0.0081

5. **Assume a customer has been with a company for 25 years and has made purchases at times 9, 11, and 15. Estimate the probability the customer is still active.**

t (Time Of Last Purchase)	**15**
T (Time elapsed between acquisition of customer and present time)	**25**
N (Number Of Purchases)	**3**

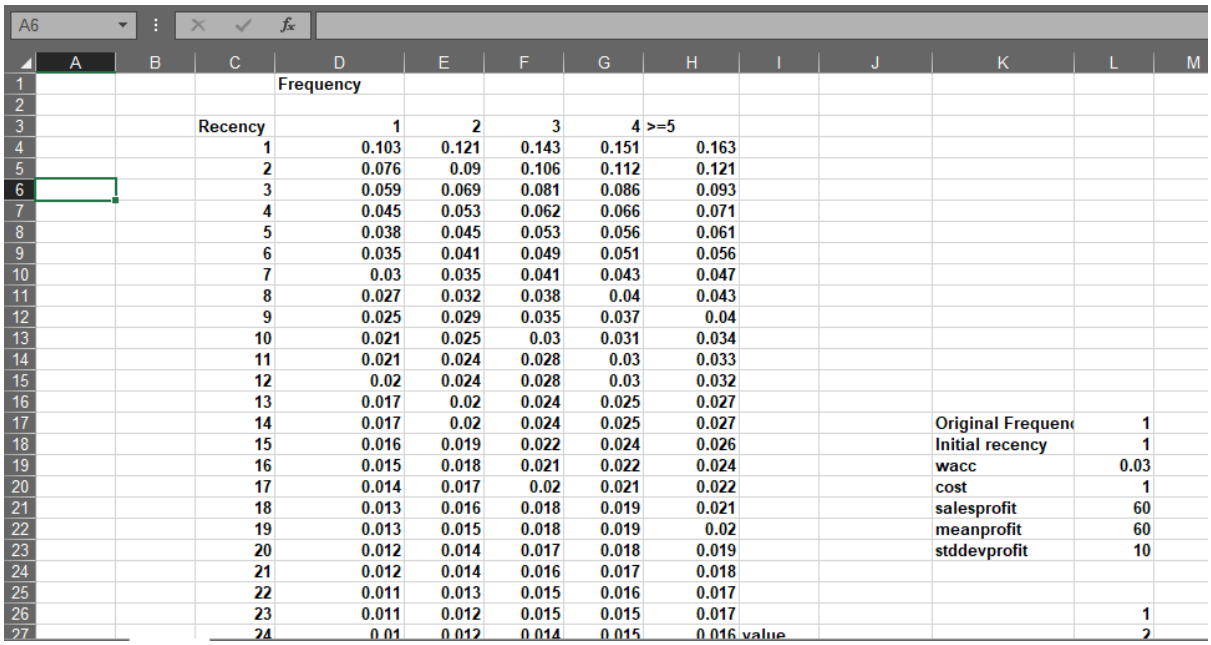
Formula : $T^* = t/T = 0.6$
Customer is still active probability : $(T^*)^N = 0.6^3 = 0.216$

PRACTICAL 7

Market Segmentation

Q 1] Shows how you can use Monte Carlo simulation to model customer value in situations in which customers pass through stages. [Dataset: Markov.xls]

The following steps describe how to determine the customer value for a customer who has bought one time (frequency = 1) and bought from the last received catalog (recency = 1).



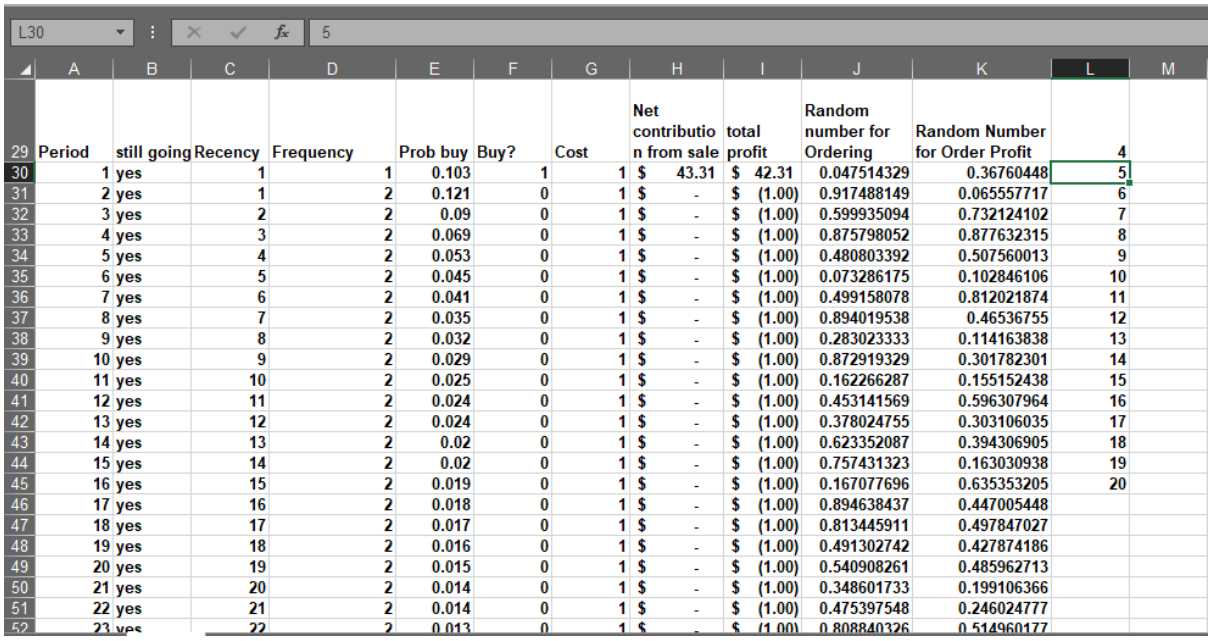
		Frequency									
	Recency	1	2	3	4	>=5					
	1	0.103	0.121	0.143	0.151	0.163					
	2	0.076	0.09	0.106	0.112	0.121					
	3	0.059	0.069	0.081	0.086	0.093					
	4	0.045	0.053	0.062	0.066	0.071					
	5	0.038	0.045	0.053	0.056	0.061					
	6	0.035	0.041	0.049	0.051	0.056					
	7	0.03	0.035	0.041	0.043	0.047					
	8	0.027	0.032	0.038	0.04	0.043					
	9	0.025	0.029	0.035	0.037	0.04					
	10	0.021	0.025	0.03	0.031	0.034					
	11	0.021	0.024	0.028	0.03	0.033					
	12	0.02	0.024	0.028	0.03	0.032					
	13	0.017	0.02	0.024	0.025	0.027					
	14	0.017	0.02	0.024	0.025	0.027			Original Frequen	1	
	15	0.016	0.019	0.022	0.024	0.026			Initial recency	1	
	16	0.015	0.018	0.021	0.022	0.024			wacc	0.03	
	17	0.014	0.017	0.02	0.021	0.022			cost	1	
	18	0.013	0.016	0.018	0.019	0.021			salesprofit	60	
	19	0.013	0.015	0.018	0.019	0.02			meanprofit	60	
	20	0.012	0.014	0.017	0.018	0.019			stddevprofit	10	
	21	0.012	0.014	0.016	0.017	0.018					
	22	0.011	0.013	0.015	0.016	0.017					
	23	0.011	0.012	0.015	0.015	0.017				1	
	24	0.01	0.012	0.014	0.015	0.016	value			2	

Step 1. In cell C30 enter the recency level (in this case 1) with the formula =Initial_recency.

Step 2. In cell D30 enter the initial frequency with the formula =Original_Frequency.

Step 3. In cell E30 determine the probability that the customer orders in with the formula

=IF(B30="no",0,INDEX(probs,C30,D30)). If you have ended the relationship with the customer, the order probability is 0.



	A	B	C	D	E	F	G	H	I	J	K	L	M
29	Period	still going	Recency	Frequency	Prob buy	Buy?	Cost	Net contribution from sale	total profit	Random number for Ordering	Random Number for Order Profit	4	
30	1	yes	1	1	0.103	1	1	$ 43.31	$ 42.31	0.047514329	0.36760448	5	
31	2	yes	1	2	0.121	0	1	$ -	$ (1.00)	0.917488149	0.065557717	6	
32	3	yes	2	2	0.09	0	1	$ -	$ (1.00)	0.599935094	0.732124102	7	
33	4	yes	3	2	0.069	0	1	$ -	$ (1.00)	0.875798052	0.877632315	8	
34	5	yes	4	2	0.053	0	1	$ -	$ (1.00)	0.480803392	0.507560013	9	
35	6	yes	5	2	0.045	0	1	$ -	$ (1.00)	0.073286175	0.102846106	10	
36	7	yes	6	2	0.041	0	1	$ -	$ (1.00)	0.499158078	0.812021874	11	
37	8	yes	7	2	0.035	0	1	$ -	$ (1.00)	0.894019538	0.46536755	12	
38	9	yes	8	2	0.032	0	1	$ -	$ (1.00)	0.283023333	0.114163838	13	
39	10	yes	9	2	0.029	0	1	$ -	$ (1.00)	0.872919329	0.301782301	14	
40	11	yes	10	2	0.025	0	1	$ -	$ (1.00)	0.162266287	0.155152438	15	
41	12	yes	11	2	0.024	0	1	$ -	$ (1.00)	0.453141569	0.596307964	16	
42	13	yes	12	2	0.024	0	1	$ -	$ (1.00)	0.378024755	0.303106035	17	
43	14	yes	13	2	0.02	0	1	$ -	$ (1.00)	0.623352087	0.394306905	18	
44	15	yes	14	2	0.02	0	1	$ -	$ (1.00)	0.757431323	0.163030938	19	
45	16	yes	15	2	0.019	0	1	$ -	$ (1.00)	0.167077696	0.635353205	20	
46	17	yes	16	2	0.018	0	1	$ -	$ (1.00)	0.894638437	0.447005448		
47	18	yes	17	2	0.017	0	1	$ -	$ (1.00)	0.813445911	0.497847027		
48	19	yes	18	2	0.016	0	1	$ -	$ (1.00)	0.491302742	0.427874186		
49	20	yes	19	2	0.015	0	1	$ -	$ (1.00)	0.540908261	0.485962713		
50	21	yes	20	2	0.014	0	1	$ -	$ (1.00)	0.348601733	0.199106366		
51	22	yes	21	2	0.014	0	1	$ -	$ (1.00)	0.475397548	0.246024777		
52	23	yes	22	2	0.013	0	1	$ -	$ (1.00)	0.808840326	0.514960177		

Step 4. In cell F30 you can determine if the customer orders during period 1 by using the formula =IF(B30="no",0,IF(J30<E30,1,0)). If you mailed a catalog and the random number in Column J is less than or equal to the chance of the customer placing an order, then an order is placed. Because the RAND() value is equally likely to be any number between 0 and 1, this gives a probability of E30 that an order is placed.

Step 5. Book the cost of mailing the catalog (if the customer is still with you) by using the formula =IF(B30="yes",$L$20,0) in cell G30.

Step 6. Book the profit from an order by using the formula =IF(AND(B30="yes",F30=1),NORMINV(J30,meanprofit,stddevprofit),0) in cell H30. If an order is received, then the profit of the order is generated with the NORMINV(J30,meanprofit,stddevprofit,0) portion of the formula. If the random number in Column J equals x, then this formula returns the xth percentile of a normal random variable with the given mean and standard deviation. For example, if J30 contains a 0.5, you generate a profit equal to the mean, and if J30 contains a 0.841, you generate a profit equal to one standard deviation above the mean.

Step 7. Copy the formulas from E30:K30 to E31:K109 and copy the formulas from B31:D31 to B32:D109 to arbitrarily cut off profits after 80 quarters (20 years).

Step 8. In cell I28 compute the present value of all profits (assuming end-of-period profits) with the formula =NPV(wacc,I30:I109).

Step 9. Go to the Formulas tab, and from Calculation Options select Automatic Except for Data Tables. After you choose this option, data tables recalculate only when you select the F9 key. This option enables you to modify the spreadsheet without waiting for data tables to recalculate.

Step 10. Enter the possible recency values (1–24) in the range Q5:AN5.

Step 11. Enter the integers 1 through 10,000 (corresponding to the 10,000 "iterations" of recalculating the spreadsheet) in the range P6:P10005. To accomplish this enter a 1 in P6, and from the Home tab, select Fill and then Series. Then complete the dialog box, as shown in Figure 3.

Step 12. Enter the output formula =I28 in the upper-left corner (cell P5) of the table range.

Step 13. From the Data tab select What-If Analysis and then choose Data Table…

Step 14. Fill in the Row input cell as L18 (Initial Recency Level).

Step 15. For the Column input cell choose any blank cell (such as AD2). In each column Excel sequentially places 1, 2, …10,000 in the blank cell and recalculates the RAND() values for the column's recency level. After a few minutes you have "played out" 10,000 customers for each recency level. (Recall you have fixed Frequency = 1.)

Step 16. Copy the formula =AVERAGE(Q6:Q10005) from Q4 to R4:AN4 to compute an estimate of the average profit for each recency level.

AB4 =AVERAGE(AB6:AB1005)

	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ
1																									
2					frequency =1																				
3																									
4					mean profit	$14.48	$8.76	$4.57	$2.42	$0.79	$0.21	-$0.29	-$0.85	-$1.79	-$2.80	-$2.44	-$2.69	-$3.84	-$1.98	-$3.38	-$1.89	-$2.53	-$1.08	-$1.96	-$0.9
5					$55.45	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	2
6					1	$99.31	$4.64	$20.04	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	$6.57	-$9.95	-$9.25	$40.05	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
7					2	$94.40	-$16.44	$23.48	-$15.42	$1.25	-$14.32	-$13.75	$100.87	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	$11.24	-$4.5
8					3	-$16.94	$44.92	-$15.94	-$15.42	$186.04	$21.98	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
9					4	$7.93	-$16.44	-$15.94	-$15.42	$17.19	-$14.32	$18.94	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
10					5	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	$20.43	-$13.75	-$13.17	$79.70	-$11.94	-$11.30	$37.95	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
11					6	$18.90	-$16.44	$29.76	$6.27	-$14.88	-$14.32	$36.04	-$13.17	-$12.56	-$11.94	$10.66	-$10.63	$16.91	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
12					7	$1.29	-$16.44	-$15.94	-$15.42	-$14.88	$18.18	$4.27	-$13.17	-$12.56	$6.38	-$11.30	-$10.63	$73.47	-$9.25	-$8.53	$12.96	-$7.02	-$6.23	-$5.42	-$4.5
13					8	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	$96.57	-$10.63	-$2.74	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
14					9	$122.50	-$16.44	-$15.94	-$0.83	-$14.88	-$14.32	-$13.75	-$13.17	$17.88	-$11.94	$24.29	$8.08	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
15					10	$125.81	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	$52.19	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
16					11	$117.76	$52.02	-$15.94	$46.97	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	$54.46	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	$18.18	-$5.42	-$4.5
17	1				12	$54.03	$8.29	-$15.94	-$15.42	$66.78	$7.03	$33.38	-$13.17	-$12.56	$75.49	-$11.30	-$10.63	$35.69	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
18	1				13	$68.79	-$16.44	-$15.94	$12.29	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	$16.28	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
19	0.03				14	$9.72	-$16.44	-$15.94	-$15.42	$2.91	$10.70	-$13.75	-$1.52	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
20	1				15	$40.58	$30.75	-$15.94	$115.56	-$14.88	$9.68	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	$63.0
21	60				16	$51.17	-$16.44	-$15.94	-$15.42	-$14.88	$123.73	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
22	60				17	$4.02	$19.84	$51.08	$36.85	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
23	10				18	-$16.94	$43.52	$44.90	-$15.42	-$14.88	-$14.32	-$13.75	$21.54	$8.26	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
24					19	-$16.94	-$16.44	-$15.94	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
25					20	$18.15	-$16.44	$52.36	-$15.42	-$14.88	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	$29.71	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
26	1				21	-$16.94	-$3.41	$76.47	-$15.42	$28.66	-$14.32	-$13.75	-$13.17	-$12.56	-$11.94	-$0.64	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	$124.38	-$6.23	-$5.42	-$4.5
27	2				22	-$16.94	-$16.44	-$15.94	-$15.42	$80.05	-$14.32	-$13.75	-$0.01	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	$6.52	-$7.02	-$6.23	-$5.42	$15.9
28	3				23	-$16.94	-$16.44	$25.66	$34.71	-$14.88	-$14.32	-$13.75	$14.06	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	$2.95	$6.15	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
29	4				24	-$16.94	-$16.44	-$15.94	-$9.82	-$14.88	-$14.32	-$13.75	$144.61	-$12.56	-$11.94	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5
30	5				25	-$16.94	-$16.44	$24.60	-$15.42	$32.27	-$14.32	-$13.75	-$13.17	$77.28	$39.43	-$11.30	-$10.63	-$9.95	-$9.25	-$8.53	-$7.79	-$7.02	-$6.23	-$5.42	-$4.5

Q 2] Use Monte Carlo Simulation to Predict Success of a Marketing Initiative. Analyse whether a pizza parlour will benefit from a Groupon offer.

	A	B	C	D	E	F	G	H
1			Groupon					
2								
3			margin	0.5				
4			2pizzas	$26.00				
5			weget	$5.00				
6			cost	$13.00	Low	High		
7			probnewcustomer	0.65	0.65	0.85		
8			newspendmorethandeal	0.38	0.3	0.42		
9			newpeoplewhoreturn	0.1	0.1	0.3		
10								
11			Look at 100 people who take deal					
12		new	65					
13		returnees	35					
14		howmanyspendmore	38	Low	High			
15		howmuchmore	$4.00	$3.00	$17.00			
16		valuenewcustomer	$31.87	Low	High			
17		anncustomerprofit	$20.00	$20.00	$40.00			
18		retentionrate	0.58	0.55	0.85			
19		newcomeback	6.5					
20								
21								
22								

I9 | 1

	A	B	C	D	E	F	G	H
22								
23								
24		Loss						
25		fromnew	$520.00					
26		fromreturning	$735.00	assume they would have come anyway				
27		total loss	$1,255.00					
28		Benefits						
29		extraprofittoday	$76.00					
30		valuenewcustomers	$207.14					
31		total	$283.14					
32								
33		Net Gain	-$971.86					
34								
35								

Step 1. In D3:D5 enter Carrie's profit margin, the price of two pizzas without Groupon, and what Carrie receives from Groupon for two pizzas.

Step 2. In D6 compute the cost of producing two pizzas with the formula =(1-margin)*_2pizzas.

Step 3. In cell D7 enter the fraction of Groupon offer takers who are new customers with

the formula =RANDBETWEEN(100*E7,100*F7)/100. This is equally likely to enter a .65, .66, …, .84, .85 in cell D7.

Step 4. In cell D8 generate the fraction of new customers who spend more than $26 with the formula =RANDBETWEEN(100*E8,100*F8)/100.

Step 5. Copy this formula from cell D8 to D9 to generate the fraction of new Groupon customers who return.

Step 6. Assume without loss of generality that 100 customers take the Groupon offer and you generate the random net gain (or loss) in profit from these 100 offer takers. Include gains or losses today and gains from added new customers.

Step 7. In cell C12 compute the number of your 100 offer takers who will be new customers with the formula =100*probnewcustomer.

Step 8. In cell C13 compute the number of offer takers who are returnees with the formula =100*(1-probnewcustomer).

Step 9. In cell C14 compute the number of offer takers who spend more than the deal with the formula =100*newspendmorethandeal.

Step 10. In cell C15 determine the average amount spent in excess of $26 by those spending more than $26 with the formula =RANDBETWEEN(D15,E15). Copy this formula to C17 to determine the average level of annual customer profit for new customers created by the Groupon offer.

Step 11. In the worksheet basic model, attach your Customer Value template, as discussed in “Calculating Lifetime Customer Value,”. Then in C16 use the formula ='basic model'!E5*C17 to compute the lifetime value of a customer based on mid-year cash flows.

Step 12. In cell C18 compute the average retention rate for new customers with the formula =RANDBETWEEN(100*D18,100*E18)/100.

Step 13. In cell C19 compute the number of the 100 offer takers who are returning new customers with the formula =C12*newpeoplewhoreturn.

Step 14. In the range C25:C33 you can compute your gain or loss from the 100 offer takers: To begin in cell C25, use the formula =(cost-weget)*C12 to compute your loss today from the new customers among the offer takers as $8 * number of new customers.

Step 15. To simplify your work assume all offer takers who were previous customers would have shown up anyway. Because each of these returning customers would have paid $26, you lose $26 – $5 = $21 on each of these customers.

Step 16. Then in cell C26 use the formula =C13*(_2pizzas-weget) to compute your loss on these customers.

Step 17. In cell C27 the formula =SUM(C25:C26) computes your total loss today on the 100 people who took the Groupon offer.

Step 18. In C29 with the formula =margin*C15*C14, compute the extra profit earned today by multiplying your 50 percent profit margin by the amount in excess of $26 spent today by offer takers.

Step 19. In C30 use the formula =C19*C16 to compute the value of the new customers by multiplying the number of returning new customers times the average value for each new customer.

Step 20. In cell C31 use the formula =C29+C30 to compute the total benefits created by the Groupon offer.

Step 21. In cell C33 use the formula =C31-C27 to compute the total benefits less today's losses.

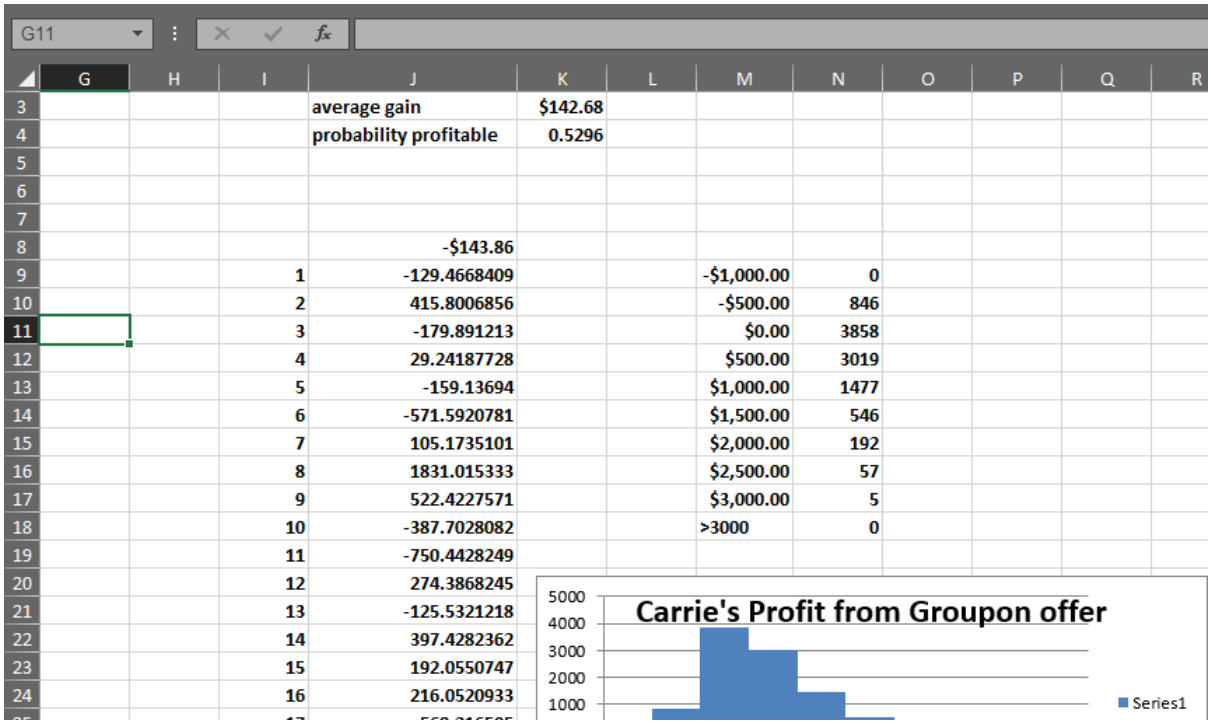
Step 22. Use FILL SERIES (from the Home tab) to enter the iteration numbers (1, 2, …, 10,000) in the range I9:I10008.

Step 23. Use a one-way data table to "trick" Excel into replaying your spreadsheet 10,000 times. Recalculate your total gain on the 100 offer takers, so enter the total gain in cell J8 with the formula =C33.

Step 24. Select the data table range (I8:J10008), and from the Data tab, choose What-If Analysis and select Data Table… In a one-way data table, there is no row input cell, so all you need to do is choose any blank cell (such as N7) as the column input cell. Then Excel places 1, 2, …, 10,000 in N7 and each time recalculates Carrie's net gain. During each recalculation each RANDBETWEEN function recalculates, so you play out the modeled uncertainty 10,000 times. The resulting simulated profits are shown in the range J9:J10008 of Figure 6.

Step 25. In cell K3 compute the average profit over your 10,000 iterations earned from the 100 deal takers with the formula =AVERAGE(J9:J10008). You find an average gain of $144.11, which indicates that on average the Groupon deal can improve Carrie's bottom line.

Step 26. In cell K4 compute the probability that the deal increases profits with the formula =COUNTIF(J9:J10008,">0")/10000. There is a 53.8 percent chance the deal yields a favorable result.

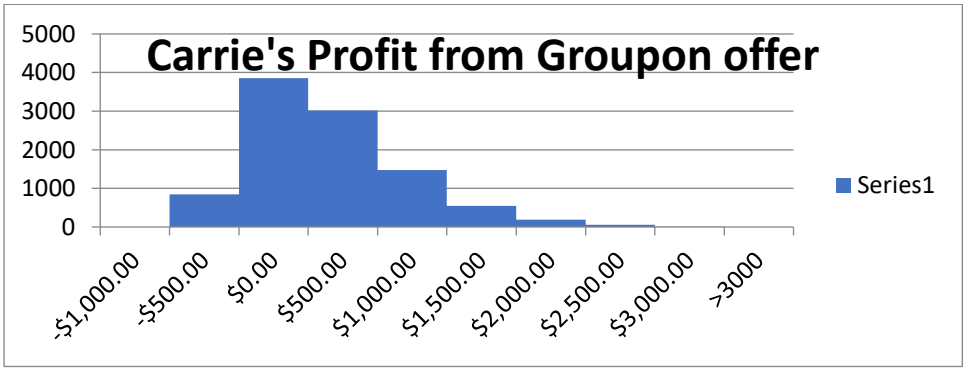


	J	K	M	N
3	average gain	$142.68		
4	probability profitable	0.5296		
8	-$143.86			
9	-129.4668409		-$1,000.00	0
10	415.8006856		-$500.00	846
11	-179.891213		$0.00	3858
12	29.24187728		$500.00	3019
13	-159.13694		$1,000.00	1477
14	-571.5920781		$1,500.00	546
15	105.1735101		$2,000.00	192
16	1831.015333		$2,500.00	57
17	522.4227571		$3,000.00	5
18	-387.7028082		>3000	0
19	-750.4428249			
20	274.3868245			
21	-125.5321218			
22	397.4282362			
23	192.0550747			
24	216.0520933			

This finding indicates that the deal is of marginal value to the pizza parlor. From cell K3, you find that the average profit per customer equals $144.11. This again indicates that on average, the Groupon deal does just a little better than breaking even.

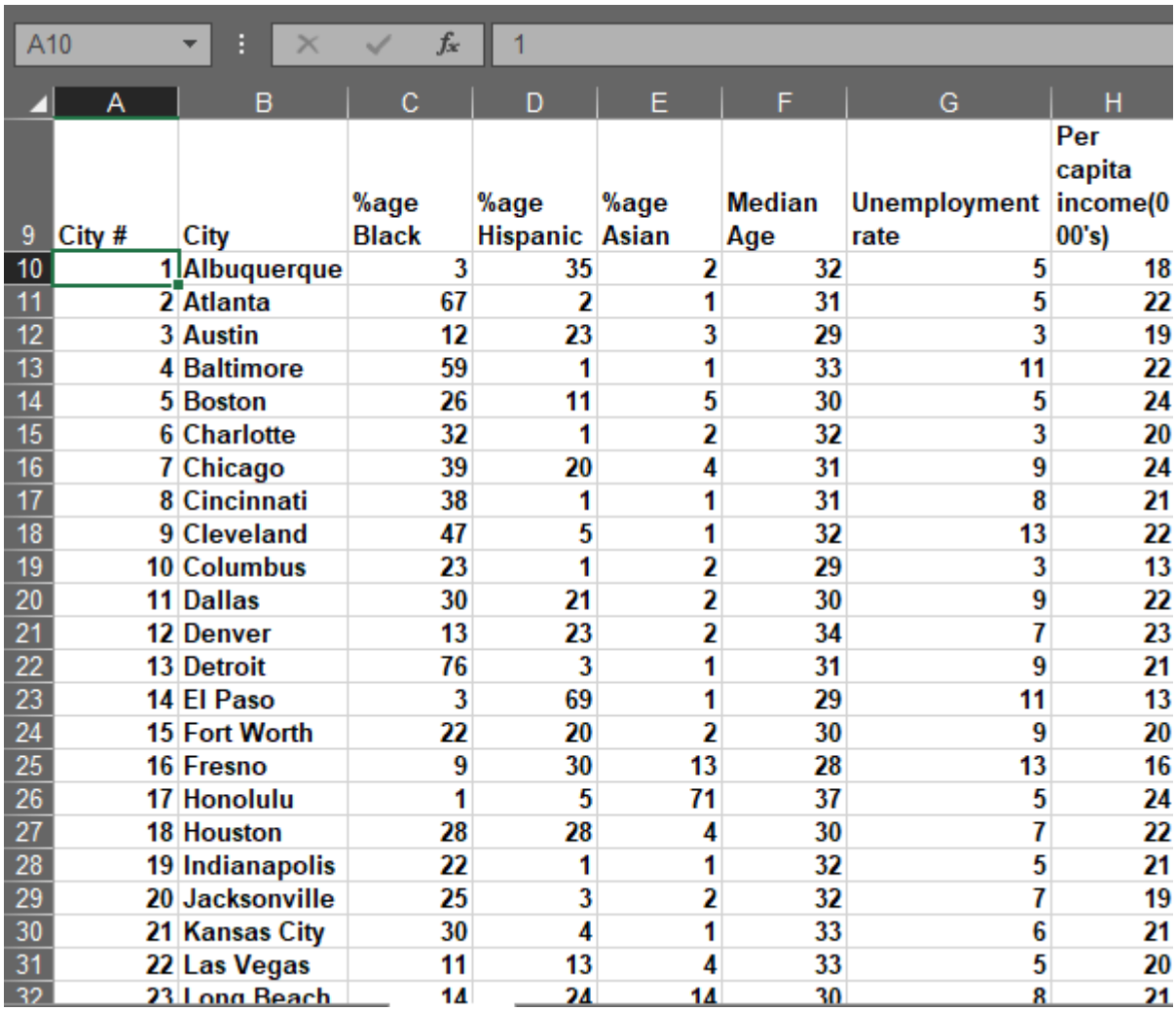
Step 26. Using a Histogram to Summarize the Simulation Results
Enter the boundaries of bin ranges (–$1000, –$500, …, $3000) in M9:M17. Append a label >3000 for all iterations in which profit is more than $3000.

Step 27. Select the range N9:N18 and array enter by selecting Ctrl+Shift+Enter (see "Using Excel Charts to Summarize Marketing Data") the formula =FREQUENCY(J9:J10008,M9:M17). In N9 this computes the number of iterations in which profit is <=$1000; in N10 this array formula computes the number of iterations in which profit is >–$1000 and <=–$500 (802); in N18 this array formula computes the number of iterations (1) in which profit is >$3000.



Q 3] Use Evolutionary version of Excel Solver to perform cluster analysis using the US city dataset.

Use the Solver to identify a given number of clusters.



	A	B	C	D	E	F	G	H
9	City #	City	%age Black	%age Hispanic	%age Asian	Median Age	Unemployment rate	Per capita income(000's)
10	1	Albuquerque	3	35	2	32	5	18
11	2	Atlanta	67	2	1	31	5	22
12	3	Austin	12	23	3	29	3	19
13	4	Baltimore	59	1	1	33	11	22
14	5	Boston	26	11	5	30	5	24
15	6	Charlotte	32	1	2	32	3	20
16	7	Chicago	39	20	4	31	9	24
17	8	Cincinnati	38	1	1	31	8	21
18	9	Cleveland	47	5	1	32	13	22
19	10	Columbus	23	1	2	29	3	13
20	11	Dallas	30	21	2	30	9	22
21	12	Denver	13	23	2	34	7	23
22	13	Detroit	76	3	1	31	9	21
23	14	El Paso	3	69	1	29	11	13
24	15	Fort Worth	22	20	2	30	9	20
25	16	Fresno	9	30	13	28	13	16
26	17	Honolulu	1	5	71	37	5	24
27	18	Houston	28	28	4	30	7	22
28	19	Indianapolis	22	1	1	32	5	21
29	20	Jacksonville	25	3	2	32	7	19
30	21	Kansas City	30	4	1	33	6	21
31	22	Las Vegas	11	13	4	33	5	20
32	23	Long Beach	14	24	14	30	8	21

Step 1. Compute the Black mean percentage in C1 with the formula =AVERAGE(C10:C58).

Step 2. In C2 compute the standard deviation of the Black percentages with the formula =STDEV(C10:C58).

Step 3. Copy these formulas to D1:G2 to compute the mean and standard deviation for each attribute.

Step 4. In cell I10 (see Figure 3) compute the standardized percentage of Blacks in Albuquerque (often called a z-score) with the formula

=STANDARDIZE(C10,C$1,C$2). This formula is equivalent, of course, to (C10 – C$1)/C$2. The reader can verify that for each demographic attribute the z-scores have a mean of 0 and a standard deviation of 1.

Step 5. Copy this formula from I10 to N58 to compute z-scores for all cities and attributes.

Step 6. Look up z-scores for cluster anchors.
In H5:H8 enter "trial values" for cluster anchors. Each of these values can be any integer between 1 and 49. For simplicity you can let the four trial anchors be cities 1–4.

Step 7. After naming A9:N58 as the range lookup in G5, look up the name of the first cluster anchor with the formula =VLOOKUP(H5,Lookup,2).

Step 8. Copy this formula to G6:G8 to identify the name of each cluster center candidate.

Step 9. In I5:N8 identify the z-scores for each cluster anchor candidate by copying from I5 to I5:N8 the formula =VLOOKUP($H5,Lookup,I$3).

	Column	9	10	11	12	13	14
City	Cluster	z *Black*	z Hispanic	z Asian	z Age	z Unemp	z income
Los Angeles	24	-0.57133	1.542497	0.355249	-0.43962	1.480155	0.024482
Omaha	34	-0.62655	-0.70373	-0.4523	0.061342	-0.75146	-0.27542
Memphis	25	1.69258	-0.82514	-0.4523	0.061342	0.736282	-0.27542
San Francisco	43	-0.73698	-0.03593	2.06008	2.06518	-0.37953	3.023526

Step 10. You can now compute the squared distance from each city to each cluster candidate. To compute the distance from city 1 (Albuquerque) to cluster candidate anchor 1, enter in O10 the formula =SUMXMY2($I$5:$N$5,$I10:$N10). This cool Excel function computes the following:
$(I5-I10)^2+(J5-J10)^2+(K5-K10)^2+(L5-L10)^2+(M5-M10)^2+(N5-N10)^2$

Step 11. To compute the squared distance of Albuquerque from the second cluster anchor, change each 5 in O10 to a 6. Similarly, in Q10 change each 5 to a 7. Finally, in R10 we change each 5 to an 8.

Step 12. Copy from O10:R10 to O11:R58 to compute the squared distance of each city from each cluster anchor.

Step 13. In S10:S58 compute the distance from each city to the "closest" cluster anchor by entering the formula =MIN(O10:R10) in cell S10 and copying it to the cell range S10:S59.

Step 14. In S8 compute the sum of squared distances of all cities from their cluster anchor with the formula =SUM(S10:S58).

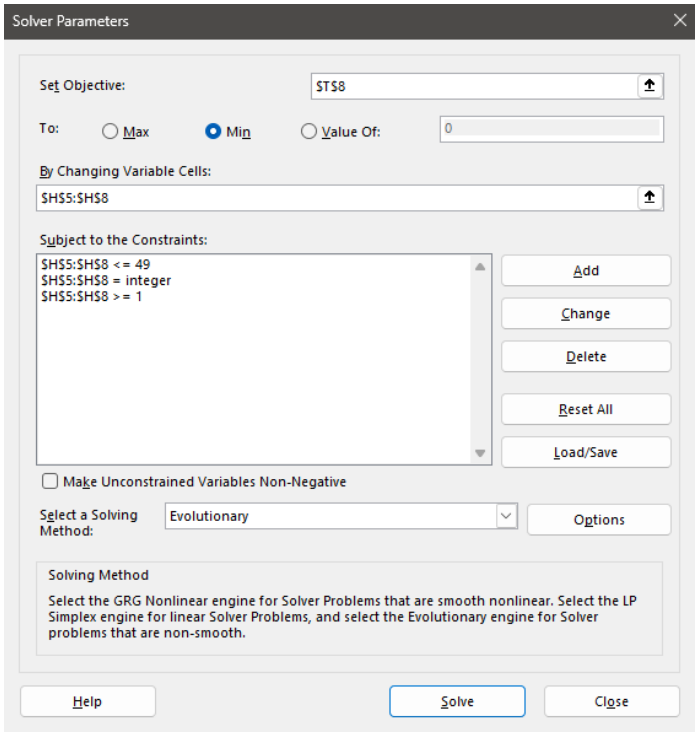
Step 15. In T10:T58 compute the cluster to which each city is assigned by entering in T10 the formula =MATCH(S10,O10:R10,0) and copying this formula to T11:T58. This

formula identifies which element in columns O:R gives the smallest squared distance to the city.

Step 16. Use the Solver window, as shown in Figure 7, to find the optimal cluster anchors for the four clusters.

Step 17. Cell S8 (sum of squared distances) is minimized in the example. The cluster anchors (H5:H8) are the changing cells. They must be integers between 1 and 49. Choose the Evolutionary Solver. Select Options from the Solver window, navigate to the Evolutionary tab, and increase the Mutation rate to 0.5. This setting of the Mutation rate usually improves the performance of the Evolutionary Solver.

9	Distance ^2 to 1	Distance ^2 to 2	Distance ^2 to 3	Distance ^2 to 4	Min Distance	Assigned to	City
10	7.016897	4.44672	15.08608	27.04372	4.44672	2	Albuquerque
11	19.60865	9.505167	3.266853	30.102	3.266853	3	Atlanta
12	11.68898	4.411405	14.78223	32.23795	4.411405	2	Austin
13	13.52578	12.05718	1.212861	26.96212	1.212861	3	Baltimore
14	9.780438	3.32289	7.717853	18.93672	3.32289	2	Boston
15	16.3033	1.676812	6.601068	23.98015	1.676812	2	Charlotte
16	5.032486	7.102106	3.873518	19.48122	3.873518	3	Chicago
17	9.259088	3.506272	1.360387	24.97923	1.360387	3	Cincinnati
18	9.381499	12.75265	2.827259	28.64125	2.827259	3	Cleveland
19	21.98167	7.546867	14.77613	49.61466	7.546867	2	Columbus
20	3.520536	5.660316	4.751479	24.71547	3.520536	1	Dallas
21	6.415243	3.848943	9.536884	13.07848	3.848943	2	Denver
22	17.97118	14.65559	1.700234	36.15315	1.700234	3	Detroit
23	10.88078	28.0051	32.50553	62.55287	10.88078	1	El Paso
24	3.078873	4.537368	5.662687	27.53352	3.078873	1	Fort Worth
25	5.577794	18.2029	18.37833	46.09383	5.577794	1	Fresno
26	49.81239	47.61727	58.32659	19.60205	19.60205	4	Honolulu
27	3.972443	4.978902	6.898872	23.09371	3.972443	1	Houston
28	11.68407	0.35165	5.623641	20.45181	0.35165	2	Indianapolis
29	8.753829	1.09039	3.410142	24.01824	1.09039	2	Jacksonville
30	10.6714	1.364073	3.524699	19.0342	1.364073	2	Kansas City
31	9.07795	0.704171	8.970302	18.31566	0.704171	2	Las Vegas



Solver window for cluster anchors

W	X	Y
Assigned to	City	Anchor
1	Dallas	LA
1	El Paso	LA
1	Fort Worth	LA
1	Fresno	LA
1	Houston	LA
1	Long Beach	LA
1	Los Angeles	LA
1	Miami	LA
1	NY	LA
1	San Antonio	LA
1	San Diego	LA
1	San Jose	LA
2	Albuquerque	Omaha
2	Austin	Omaha
2	Boston	Omaha
2	Charlotte	Omaha
2	Columbus	Omaha
2	Denver	Omaha
2	Indianapolis	Omaha
2	Jacksonville	Omaha
2	Kansas City	Omaha
2	Las Vegas	Omaha
2	Milwaukee	Omaha

Interpretation of Clusters

The z-scores of the anchors represent a typical member of a cluster. Therefore, examining the z-scores for each anchor enables you to easily interpret your clusters. You can find that the San Francisco cluster consists of rich, older, and highly Asian cities. The Memphis cluster consists of highly Black cities with high unemployment rates. The Omaha cluster consists of average income cities with few minorities. The Los Angeles cluster consists of highly Hispanic cities with high unemployment rates. From your clustering of U.S. cities a company like Procter & Gamble that often engages in test marketing of a new product could now predict with confidence that if a new product were successfully marketed in the San Francisco, Memphis, Los Angeles, and Omaha areas, the product would succeed in all 49 cities. This is because the demographics of each city in the data set are fairly similar to the demographics of one of these four cities.

Q 4] Using Conjoint Analysis to Segment a Market.

Use the Cokepepsi dataset

D	E	F	G
Choice	Price	Coke or Pepsi	Diet or Regular
1	$3.50	C	D
2	$3.75	C	D
3	$4.00	C	D
4	$4.25	C	D
5	$4.50	C	D
6	$3.50	C	R
7	$3.75	C	R
8	$4.00	C	R
9	$4.25	C	R
10	$4.50	C	R
11	$3.50	P	D
12	$3.75	P	D
13	$4.00	P	D
14	$4.25	P	D
15	$4.50	P	D
16	$3.50	P	R
17	$3.75	P	R
18	$4.00	P	R
19	$4.25	P	R
20	$4.50	P	R

Step 1. Determine the regression equation for each customer. Let each row of your spreadsheet be the regression coefficients the customer gives to each attribute in the regression. Then do a cluster analysis on these regression coefficients.

Step 2. Next, determine the regression coefficients for each customer where the dependent variable is the customer's rank and the independent variables describe the product profile. Figure 10 shows the data needed to run the regression for customer 1. Coke = 1 indicates a product profile was Coca-Cola (so Coke = 0 indicates Pepsi) and Diet = 1 to indicate that a product profile was diet (so Diet = 0 indicates the product profile was regular soda).

Step 3. Enter the customer number in cell J3. Then copy the formula =INDEX(ranks,$J$3,D6) from cell J6 to J7:J25 to pick off the customer's ranking of the product profiles. (The range ranks refers to the range AD29:AW160.)

To get the regression coefficients, you need to run a regression based on each customer's rankings. Therefore you need to run 132 regressions

Step 4. To run a regression with LINEST when there are m independent variables, select a blank range with five rows and m + 1 columns. The syntax of LINEST is LINEST(knowny's,knownx's,const,stats).

Step 5. To make any array function work you must use the Control+Shift+Enter key sequence. After selecting the cell range R12:U16 with the cursor in R12, enter =LINEST(J6:J25,K6:M25,TRUE,TRUE) and the array enters the formula with the Control+Shift+Enter key sequence.

Step 6. Now in R12:U12 you can find the least-squares regression equation; with the coefficients read right to left, starting with the Intercept. For example, for customer 1 the best fit to her rankings is 47.5 − 3.8*Diet + 6.6*Coke − 9.6*Price. This indicates that customer 1 prefers Regular to Diet and Coke to Pepsi. You do not need to be concerned with the remainder of the LINEST output.

AY	AZ	BA	BB	BC	BD	BE
	Diet	Coke	Price			
	-3.8	6.6	-9.6			
1	-3.8	6.6	-9.6			
2	1	-1.6	-16			
3	4.6	-6	-11.2			
4	5.8	6	-10			
5	4.6	-6	-11.2			
6	-5.8	5.8	-8.8			
7	4.4	-4.8	-12.8			
8	-8.4	-2.8	-9.6			
9	6.8	-5.2	-10			
10	3.97E-16	1.6	-16			
11	-6.8	-3.4	-11.2			
12	7.6	-5.2	-8.8			
13	-1	1.6	-16			
14	4.4	4.8	-12.8			
15	8.4	3.4	-9.6			
16	5.8	6.2	-9.4			
17	4.6	6	-11.2			
18	3.8	-6.6	-10.8			
19	1	-1.6	-16			
20	-4.4	-5.2	-11.6			

Step 7. Enter the customer numbers in AY11:AY130. The customer numbers are the input into a one-way data table (so called because the table has only one input cell: customer number.) As you vary the customer number, the one-way data table uses the LINEST function to compute the coefficients of each customer's regression equation.

Step 8. Copy the formula =R12 from AZ10 to BA10:BB10 to create output cells, which pick up each regression coefficient.

Step 9. Select the Table range AY10:BB130, and select What-If Analysis from the Data Tools Group on the Data tab.

Step 10. Next select Data Table… and as shown in Figure 23.12, choose J3 as the column input cell. This causes Excel to loop through each customer and use LINEST to run a regression based on each customer's rank.

Step 11. Copy the regression results from your data table to the worksheet cluster, and run a cluster analysis with five clusters on the regression coefficients. Use customers 1–5 as the initial set of anchors.

J	K	L	M	N	O
	5	6	7		
Anchor	z Diet	z Coke	z Price		
27	1.092289	1.287635545	0.633988	Diet Coke	
117	1.273207	-1.091955351	0.402277	Diet Pepsi	
72	-0.21032	-0.328690346	-1.68313	Just Price	
40	-1.09681	-1.159302263	0.633988	Regular Pepsi	
98	-1.00635	1.10804378	0.633988	Regular Coke	

Q 5] Show how you can use decision trees to develop simple rules that can be used to predict the value of a binary dependent variable from several independent variables.

Use the greekyougurt dataset

	A	B	C	D	E	F
1						
2		Person	Gender	Marital Status	Income Level	Bought?
3		1	Female	Single	High	No
4		2	Male	Married	Average	No
5		3	Male	Single	Low	No
6		4	Female	Married	High	No
7		5	Male	Divorced	Average	Yes
8		6	Male	Married	Low	No
9		7	Female	Divorced	High	No
10		8	Male	Single	Average	Yes
11		9	Male	Married	Low	No
12		10	Male	Single	Average	Yes

Step 1. Copy the formula =COUNTIFS($C$3:$C$12,$C15,$F$3:$F$12,D$14) from D15 to D15:E16 to compute the number of females and males who buy and do not buy Greek yogurt. For example, you find four males do not buy Greek yogurt.

Step 2. Copy the formula =COUNTIFS($E$3:$E$12,$C19,$F$3:$F$12,D$14) from D19 to D19:E21 to count the number of people for each income level that buy and do not buy Greek yogurt. For example, three average income people bought Greek yogurt.

Step 3. Copy the formula =COUNTIFS($D$3:$D$12,$C24,$F$3:$F$12,D$14) from D24 to the range D24:E26 to count how many people for each marital status buy or do not buy Greek yogurt. For example, two single people buy Greek yogurt.

Step 4. Copy the formula =SUM(D15:E15) from F15 to the range F15:F26 to compute the number of people for the given attribute value. For example, cell F25 tells you there are four married people in the population.

Step 5. Copy the formula =F15/SUM($F$15:$F$16) from G15 to G15:G26, to compute the fraction of observations having each possible attribute value. For example, from G16 you can find that 70 percent of the observations involve males.

Step 6. Copy the formula =IFERROR((D15/$F15)*LOG(D15/$F15,2),0) from H15 to H15:I26 to compute for each attribute value category level combination the term $P(i|X=a)*Log_2(P(i|X=a)$. You need IFERROR to ensure that when $P(i|X=a)=0$ Log2(0) the undefined value is replaced by 0. In general entering IFERROR(formula, anything) will enter the value computed by the formula as long as the formula does not return an error. If the formula does return an error, IFERROR returns whatever is entered after the comma (in this case 0.)

Step 7. Copy the formula =SUM(H15:I15) from J15 to J15:J26 to compute via Equation 1 the entropy for each possible node split.

Step 8. Copy the formula =−SUMPRODUCT(G15:G16,J15:J16) from K15 to K16:K24 to compute via Equation 2 the impurity for each split.

Step 9. The impurity for income of 0.325 is smaller than the impurities for gender (0.69) and marital status (0.60), so begin the tree by splitting the parent node on income.

Step 10. The Income = Low and Income = High nodes are pure, so no further splitting is necessary. The Income = Average is not pure, so you need to consider splitting this node on either gender or marital status. Splitting on gender yields an impurity of 0.811, whereas splitting on marital status yields an impurity of 0. Therefore, split the Income = Average node on marital status. Because all terminal nodes are pure (that is, each respondent for a terminal node is in the same class), no further splitting is needed, and you obtain the decision tree.

	B	C	D	E	F	G	H	I	J	K
13			Buy							
14			Yes	No	Total	Fraction			Entropy	Impurity
15	Gender	Female	0	3	3	0.3	0	0	0	0.68966
16		Male	3	4	7	0.7	-0.524	-0.461	-0.98523	
17			Buy							
18			Yes	No						
19	Income Level	High	0	3	3	0.3	0	0	0	0.324511
20		Average	3	1	4	0.4	-0.311	-0.5	-0.81128	
21		Low	0	3	3	0.3	0	0	0	
22			Buy							
23			Yes	No						
24	Marital Status	Single	2	2	4	0.4	-0.5	-0.5	-1	0.6
25		Married	0	4	4	0.4	0	0	0	
26		Divorced	1	1	2	0.2	-0.5	-0.5	-1	
27										

PRACTICAL 8

Forecasting New Product Sales

Q 1] Predicting Movie Revenues

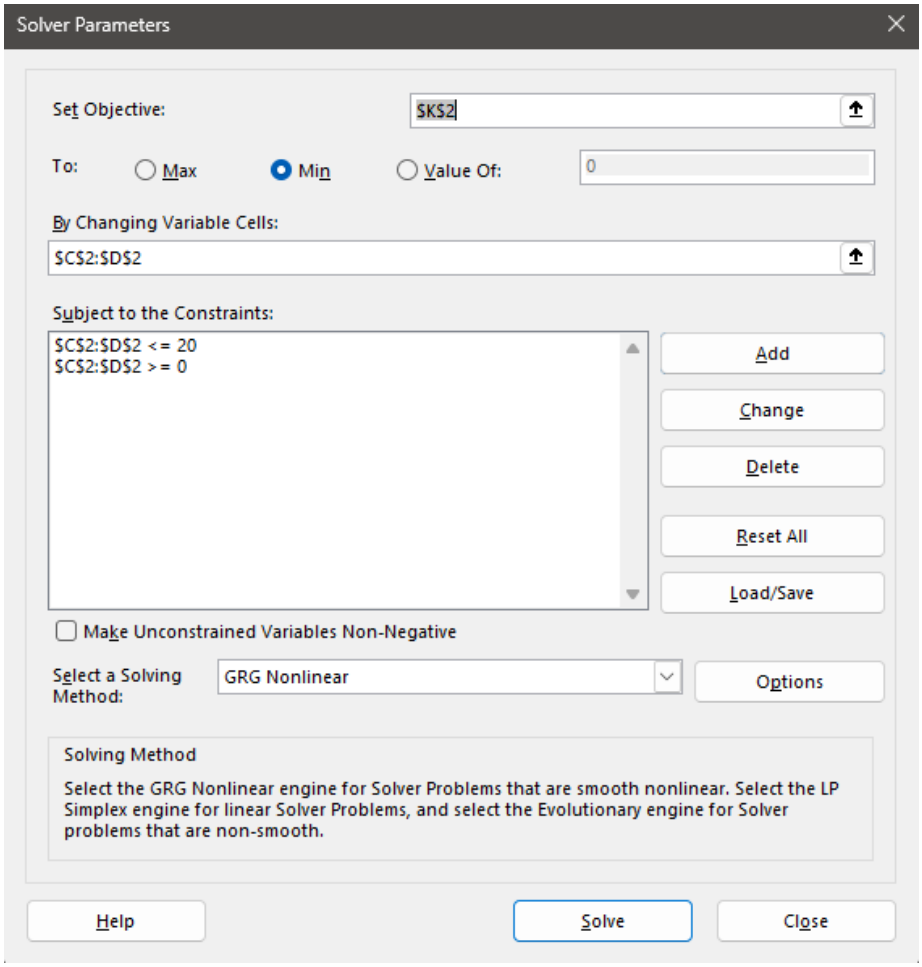
To find the values of a and alpha that minimize MAD,

Step 1. Copy the formula =a*SUM(D4:E4)*(H4^alpha) from J4 to J5:J79 to generate the forecast for each movie's total revenue.

Step 2. Copy the formula =ABS(G4-J4) from K4 to K5:K79 to compute the absolute error for each movie's prediction.

Step 3. In cell K2 compute the average absolute error (the target cell) with the formula =AVERAGE(K4:K79).

Step 4. Using the Solver window, find values of alpha and a that minimize MAD.

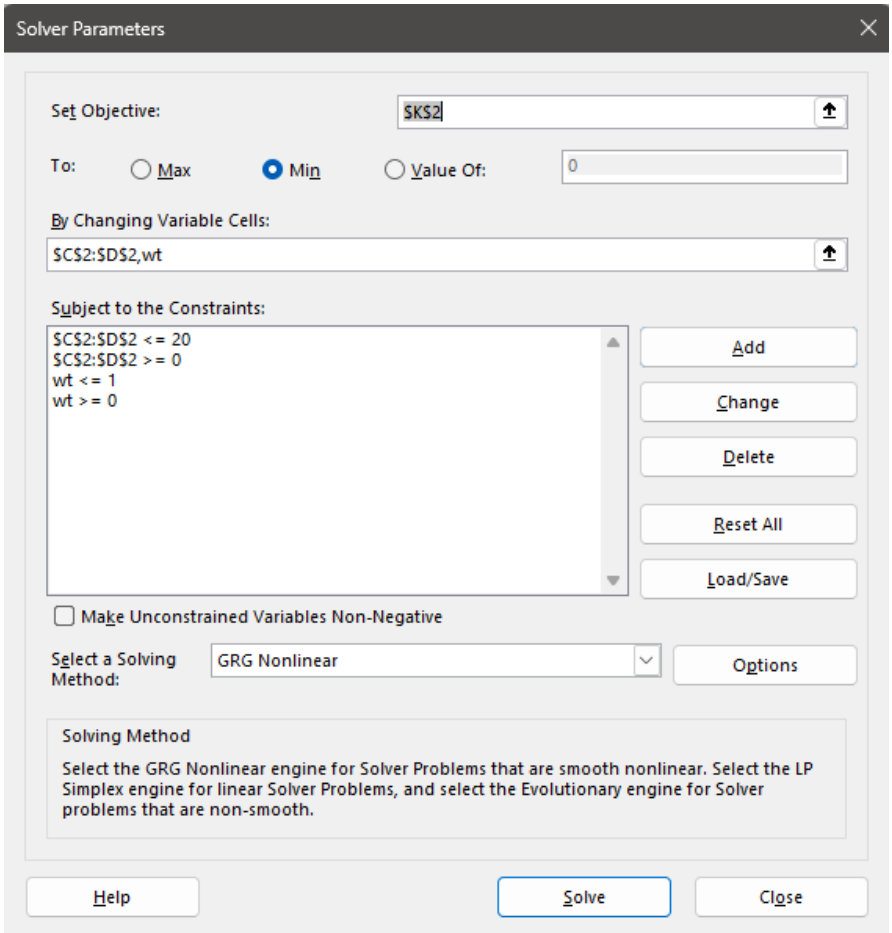


	A	B	C	D	E	F	G	H	I	J	K	L	M
1			a	alpha		Total=a*1st 2 wks*Legs^alpha					MAPE		stdeverro
2			1.783118312	0.304894							5558539.958		9676753
3			Movie	Week 1	Week 2	Week 3	Total	Legs	1st 2 weeks	Forecast	Abs Err	Error	
4		1	Monster's Ball	2321246	3217185	2097738	19964720	1.385973	5538431	10909054	9055666.019	9055666.019	
5		2	Beauty & The Beast - SE	2073437	2443280	1470492	8527183	1.178372	4516717	8467137	60045.7153	60045.7153	
6		3	About A Boy	8557630	9821030	4126600	29370520	1.147634	18378660	34176511	4805991.454	-4805991.454	
7		4	Gosford Park	3684621	4151226	2782555	27265504	1.126636	7835847	14489540	12775964.18	12775964.18	
8		5	Jimmy Neutron: Boy Genius	13832786	15035649	9015854	50699555	1.086957	28868435	52801271	2101715.682	-2101715.682	
9		6	Brotherhood of the Wolf	1533927	1608920	1904085	6219382	1.04889	3142847	5686222	533160.4853	533160.4853	
10		7	In the Bedroom	2853430	2859733	1941677	21949644	1.002209	5713163	10194101	11755542.73	11755542.73	
11		8	Big Fat Liar	11554015	11428335	6324015	39453765	0.989122	22982350	40843821	1390055.556	-1390055.556	
12		9	The Lord of the Rings	47211490	38695582	23006447	1.85E+08	0.819622	85907072	1.44E+08	41261479.1	41261479.1	
13		10	I Am Sam	8315581	6303148	4619148	28593343	0.757992	14618729	23955230	4638112.71	4638112.71	
14		11	The Royal Tenenbaums	8514122	6408153	5358838	27476757	0.75265	14922275	24399963	3076793.873	3076793.873	
15		12	Star Wars Episode II	80027814	60003949	21002876	1.99E+08	0.749789	140031763	2.29E+08	29876881.41	-29876881.41	
16		13	The Rookie	16021684	11703657	8076763	56652477	0.730489	27725341	44923505	11728971.96	11728971.96	
17		14	A Walk to Remember	12177488	8836201	5542525	32043262	0.725618	21013689	33979200	1935938.356	-1935938.356	
18		15	Clockstoppers	10108333	7284214	4652393	27816421	0.720615	17392547	28064539	248118.3936	-248118.3936	
19		16	The New Guy	9007833	6478078	5426096	22933345	0.719161	15485911	24972614	2039269.1	-2039269.1	
20		17	Unfaithful	14065277	10013104	7554263	37456921	0.711902	24078381	38708946	1252025.094	-1252025.094	
21		18	We Were Soldiers	20212543	14208525	8488331	57218634	0.702956	34421068	55123122	2095511.827	2095511.827	
22		19	Orange County	15053226	10529180	4418401	33993105	0.699463	25582406	40906408	6913303.117	-6913303.117	
23		20	Life as a House	3818623	2646422	2121159	10596087	0.69303	6465045	10308562	287525.2052	287525.2052	
24		21	Murder By Numbers	9307394	6362457	3624487	23920680	0.683592	15669851	24881444	960764.0644	-960764.0644	
25		22	The Mothman Prophecies	11208851	7364011	4870476	26370560	0.656982	18572862	29136148	2765587.863	-2765587.863	

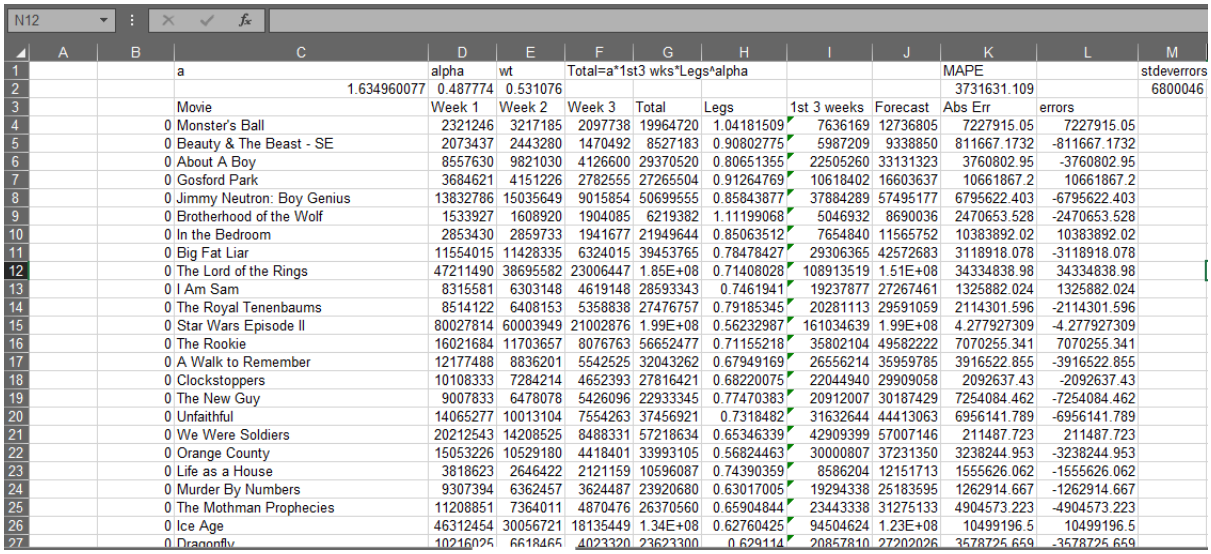
Using 3 Weeks of Revenue to Forecast Movie Revenues

Step 5. Copy the formula =wt*(E4/D4)+(1-wt)*(F4/E4) from H4 to H5:H79 to compute the weighted estimate of Legs for each movie.

Step 6. Copy the formula =a*I4*(H4^alpha) from J4 to J5:J79 to compute the forecast for each movie. The absolute error and MAD formulas are as before, and the new Solver window (reflecting the fact that wt is a changing cell)



Step 7. Use the following equation to predict Total Revenue:
Total revenue = 1.63 * (0.54* Week 2 Legs + 0.46 * Week 3 Legs) ^ 0.49
The MAD has been reduced to $3.7 million.



		a	alpha	wt	Total=a*1st3 wks*Legs^alpha					MAPE		stdeverrors
		1.634960077	0.487774	0.531076						3731631.109		6800046
		Movie	Week 1	Week 2	Week 3	Total	Legs	1st 3 weeks	Forecast	Abs Err	errors	
4	0	Monster's Ball	2321246	3217185	2097738	19964720	1.04181509	7636169	12736805	7227915.05	7227915.05	
5	0	Beauty & The Beast - SE	2073437	2443280	1470492	8527183	0.90802775	5987209	9338850	811667.1732	-811667.1732	
6	0	About A Boy	8557630	9821030	4126600	29370520	0.80651355	22505260	33131323	3760802.95	-3760802.95	
7	0	Gosford Park	3684621	4151226	2782555	27265504	0.91264769	10618402	16603637	10661867.2	10661867.2	
8	0	Jimmy Neutron: Boy Genius	13832786	15035649	9015854	50699555	0.85843877	37884289	57495177	6795622.403	-6795622.403	
9	0	Brotherhood of the Wolf	1533927	1608920	1904085	6219382	1.11199068	5046932	8690036	2470653.528	-2470653.528	
10	0	In the Bedroom	2853430	2859733	1941677	21949644	0.85063512	7654840	11565752	10383892.02	10383892.02	
11	0	Big Fat Liar	11554015	11428335	6324015	39453765	0.78478427	29306365	42572683	3118918.078	-3118918.078	
12	0	The Lord of the Rings	47211490	38695582	23006447	1.85E+08	0.71408028	108913519	1.51E+08	34334838.98	34334838.98	
13	0	I Am Sam	8315581	6303148	4619148	28593343	0.7461941	19237877	27267461	1325882.024	1325882.024	
14	0	The Royal Tenenbaums	8514122	6408153	5358838	27476757	0.79185345	20281113	29591059	2114301.596	-2114301.596	
15	0	Star Wars Episode II	80027814	60003949	21002876	1.99E+08	0.56232987	161034639	1.99E+08	4.277927309	-4.277927309	
16	0	The Rookie	16021684	11703657	8076763	56652477	0.71155218	35802104	49582222	7070255.341	7070255.341	
17	0	A Walk to Remember	12177488	8836201	5542525	32043262	0.67949169	26556214	35959785	3916522.855	-3916522.855	
18	0	Clockstoppers	10108333	7284214	4652393	27816421	0.68220075	22044940	29909058	2092637.43	-2092637.43	
19	0	The New Guy	9007833	6478078	5426096	22933345	0.77470383	20912007	30187429	7254084.462	-7254084.462	
20	0	Unfaithful	14065277	10013104	7554263	37456921	0.7318482	31632644	44413063	6956141.789	-6956141.789	
21	0	We Were Soldiers	20212543	14208525	8488331	57218634	0.65346339	42909399	57007146	211487.723	211487.723	
22	0	Orange County	15053226	10529180	4418401	33993105	0.56824463	30000807	37231350	3238244.953	-3238244.953	
23	0	Life as a House	3818623	2646422	2121159	10596087	0.74390359	8586204	12151713	1555626.062	-1555626.062	
24	0	Murder By Numbers	9307394	6362457	3624487	23920680	0.63017005	19294338	25183595	1262914.667	-1262914.667	
25	0	The Mothman Prophecies	11208851	7364011	4870476	26370560	0.65904844	23443338	31275133	4904573.223	-4904573.223	
26	0	Ice Age	46312454	30056721	18135449	1.34E+08	0.62760425	94504624	1.23E+08	10499196.5	10499196.5	
27	0	Dragonfly	10216025	6618465	4023320	23623300	0.629114	20857810	27202026	3578725.659	-3578725.659	

Q 2] Show how an S- curve arises.

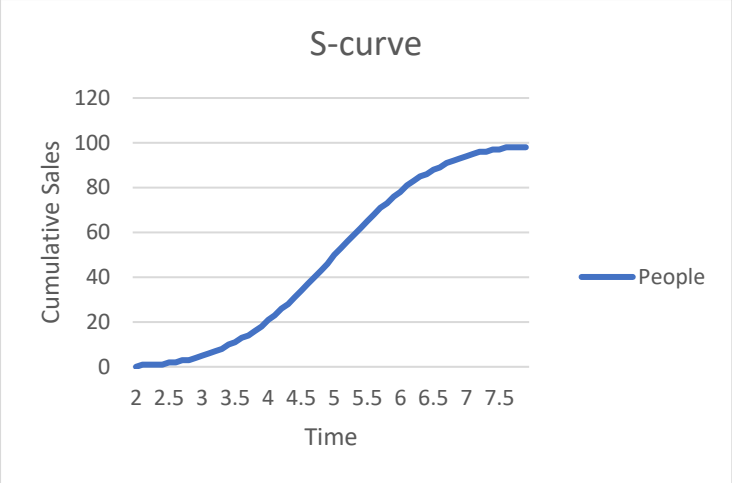
Use the scurvenormal dataset

Step 1. In H9 through H10:H107 compute an “average” time at which the nth person adopts the product. For example, in H18 compute the 10th percentile of the time to adoption and assume that the 10th person will adopt at this time.

Step 2. Copy the formula =NORMINV(G9/100,$I$4,$I$5) from H9 to H10:H107 to compute your estimate of the average time that each person adopts the product. For example, in H18 this computes the 10th percentile of the time to adoption.

Step 3. Copy the formula =COUNTIF($H$9:$H$107,"<="&J11) from K11 to K12:K70 to count the number of people that adopted the product by time t. For example, the formula in cell K21 computes the number of people (five) that have adopted the product by time 3.

Step 4. Graph the range J10:K70 with a scatter chart to create the chart.
Note the graph's inflection point appears to occur near t = 5.



F1

	F	G	H	I	J	K	L	M
1			mean	5				
2			sigma	1.25				
3								
4								
5		Person	Time					
6		1	2.092065					
7		2	2.432814		Time	People		
8		3	2.649008		2	0		
9		4	2.811642		2.1	1		
10		5	2.943933		2.2	1		
11		6	3.056533		2.3	1		
12		7	3.155261		2.4	1		
13		8	3.243661		2.5	2		
14		9	3.324056		2.6	2		
15		10	3.398061		2.7	3		
16		11	3.46684		2.8	3		
17		12	3.531267		2.9	4		
18		13	3.592011		3	5		
19		14	3.649601		3.1	6		
20		15	3.704458		3.2	7		
21		16	3.756928		3.3	8		
22		17	3.807293		3.4	10		

Q 3] Use Excel to fit an S curve to the number of worldwide cell phone subscribers per 100 people.

Step 1. In F2:H2 enter trial values of L, a, and b.

Step 2. Copy the formula =L/(1+a*EXP(−b*C5)) from F5 to F6:F15 and generate an estimate of cell phones per 100 people for the given parameters.

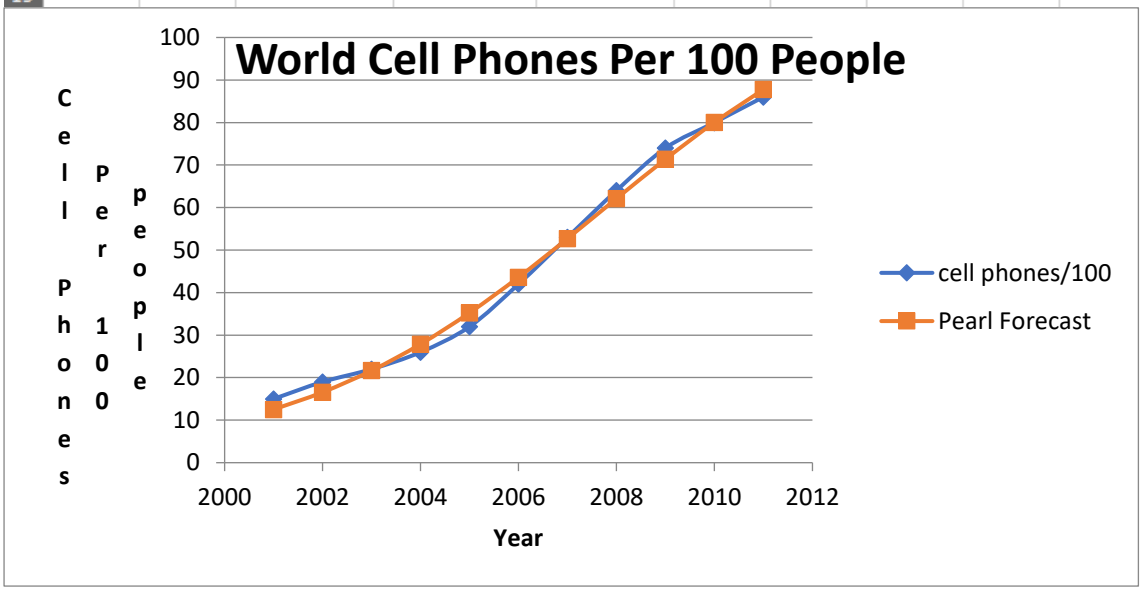
Step 3. Copy the formula =(E5-F5)^2 from G5 to G6:G15 to compute the squared error for each observation.

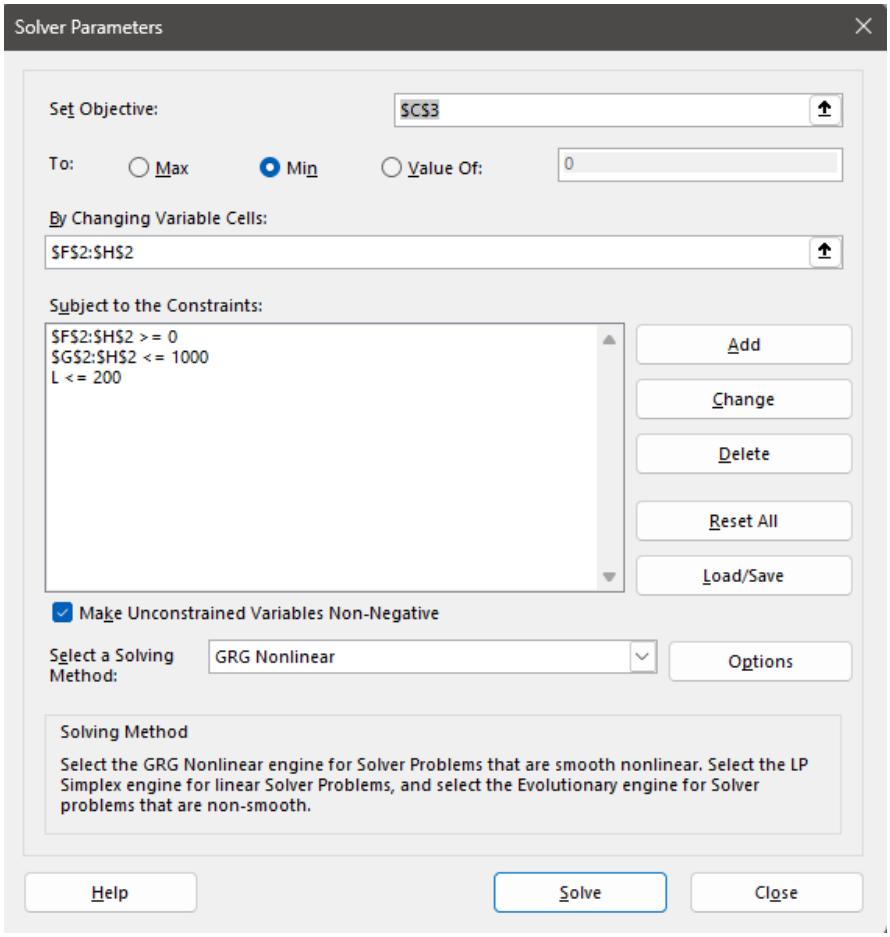
Step 4. In cell C3 compute the Sum of Squared Errors with the formula =SUM(G5:G15).

Step 5. Using the Solver window in Figure 4, use the GRG MultiStart Engine to find your estimated Logistic curve to be the following:

Cell phones per 100 people during year t = 118.17/ (1 + 11.618 e^-0.319t).

	C	D	E	F	G	H	I	J	K	L
1				L	a	b				
2	SSE			118.1673298	11.61785575	0.31943				
3	42.6364									
4		Year	cell phones/100	Pearl Forecast	Squared error					
5	1	2001	15	12.51625031	6.169012515					
6	2	2002	19	16.56627278	5.923028158					Inflection Point
7	3	2003	22	21.65815532	0.116857788					7.67789616
8	4	2004	26	27.88555203	3.555306453					
9	5	2005	32	35.24952225	10.55939484					
10	6	2006	42	43.61861468	2.619913498					
11	7	2007	53	52.71156351	0.08319561					
12	8	2008	64	62.12052561	3.532423988					
13	9	2009	74	71.37754434	6.877273696					
14	10	2010	80	80.04393295	0.001930104					
15	11	2011	86	87.78831929	3.198085888					
16	12	2012		94.42612404						
17	13	2013		99.9151168						
18	14	2014		104.3211374						
19										





Q 4] Create a Scan*Pro model to predict weekly sales of Snickers.

Use the Snickers.xlsx

Step 1. Raise the price to an unknown power. (Call this power OWNELAS.) This creates a term of the form (Our Price)^{OWNELAS}. The value of OWNELAS can estimate the price elasticity. You would expect OWNELAS to be negative. For example, if OWNELAS = −3, you can estimate that for any price you charge, a 1 percent price increase reduces demand 3 percent.

Step 2. Raise the competitor's price to an unknown power (COMPELAS). This creates a term of the form (Comp Price)^{COMPELAS}. The value of COMPELAS estimates a crosselasticity of demand. You would expect COMPELAS to be positive and smaller in magnitude than OWNELAS. For example, if COMPELAS = 0.4, then a 1 percent increase in the competitor's price increases demand for Snickers (for any set of prices) by 0.4 percent.

Step 3. Model the effect of a display by a term that raises an unknown parameter (Call it DISPLAYEFFECT.) to the power $^{DISPLAY\#}$ (which is 1 if there is a display and 0 if there is no display). This term is of the form (DISPLAYEFFECT?)$^{DISPLAY\#}$. When there is a display, this term equals DISPLAYEFFECT, and when there is no display,

this term equals 1. Therefore, a value of, say, DISPLAYEFFECT=1.2 indicates that after adjusting for prices, a display increases weekly sales by 20 percent.

Now use the GRG multistart Solver Engine to determine values of CONSTANT, OWNELAS, COMPELAS, and DISPLAYEFFECT that minimize the sum of the squared weekly prediction errors. Proceed as follows:

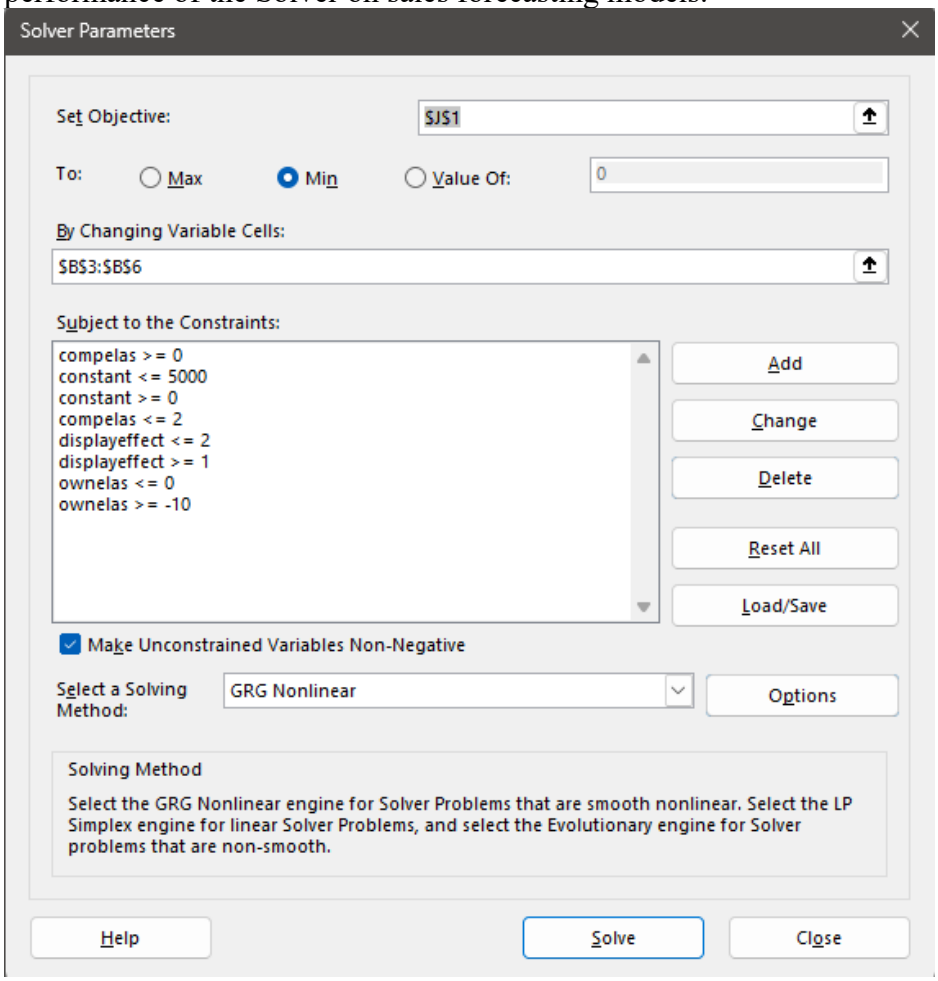
Step 4. Copy the formula =constant*(D4^ownelas)*(E4^compelas)*(displayeffect^F4) from H4 to H5:H45 to create a forecast for each week's demand.

Step 5. Copy the formula =(G4-H4)^2 from I4 to I5:I45 to compute the squared error for each week's forecast.

Step 6. In cell J1 compute the sum of the weekly squared errors with the formula =SUM(I4:I45).

Step 7. In cell J2 compute the R-squared value between your predictions and actual sales with the formula =RSQ(G4:G45,H4:H45).

Step 8. Using the Solver window, find the parameter estimates that minimize the sum of the squared forecast errors. Check Use Automatic Scaling from Options. This improves the performance of the Solver on sales forecasting models.



J1 =SUM(I4:I45)

	A	B	C	D	E	F	G	H	I	J
1									SSE	753677.4
2									rsq	0.921893
3	displayeffect	1.198152	Week	Our price	Comp price	Display?	Sales	Prediction	Sq Error	
4	ownelas	-3.19015	1	1.04	0.81	1	986	868.45425	13817	
5	compelas	0.400499	2	1.09	1.17	1	788	866.26289	6125.079	
6	constant	893.7695	3	1.16	1.04	1	580	677.52998	9512.097	
7			4	1.07	1.02	0	660	725.99079	4354.785	
8			5	0.9	0.94	0	1263	1220.2161	1830.465	
9			6	0.8	0.89	0	1972	1738.2579	54635.39	
10			7	0.84	0.83	0	1522	1446.6956	5670.754	
11			8	1.06	1.02	0	755	748.06664	48.07155	
12			9	1.06	0.96	1	904	874.79772	852.773	
13			10	0.86	1.14	1	1751	1825.9465	5616.983	
14			11	0.97	1.14	1	1104	1243.736	19526.15	
15			12	1.09	1.13	0	739	712.99603	676.2063	
16			13	1.16	0.88	1	707	633.68294	5375.391	
17			14	1.17	1.17	0	540	576.78041	1352.798	
18			15	0.96	1.03	1	1108	1234.3446	15962.97	
19			16	0.82	0.83	1	1596	1871.8696	76104.04	
20			17	0.81	1.09	1	2050	2171.0702	14657.99	
21			18	0.95	1.2	1	1361	1356.7968	17.66681	
22			19	1.08	1.06	1	862	857.52277	20.04555	

Q 5] Build a Scan*Pro model to predict the weekly Software Sales.

Use the softwaresales data

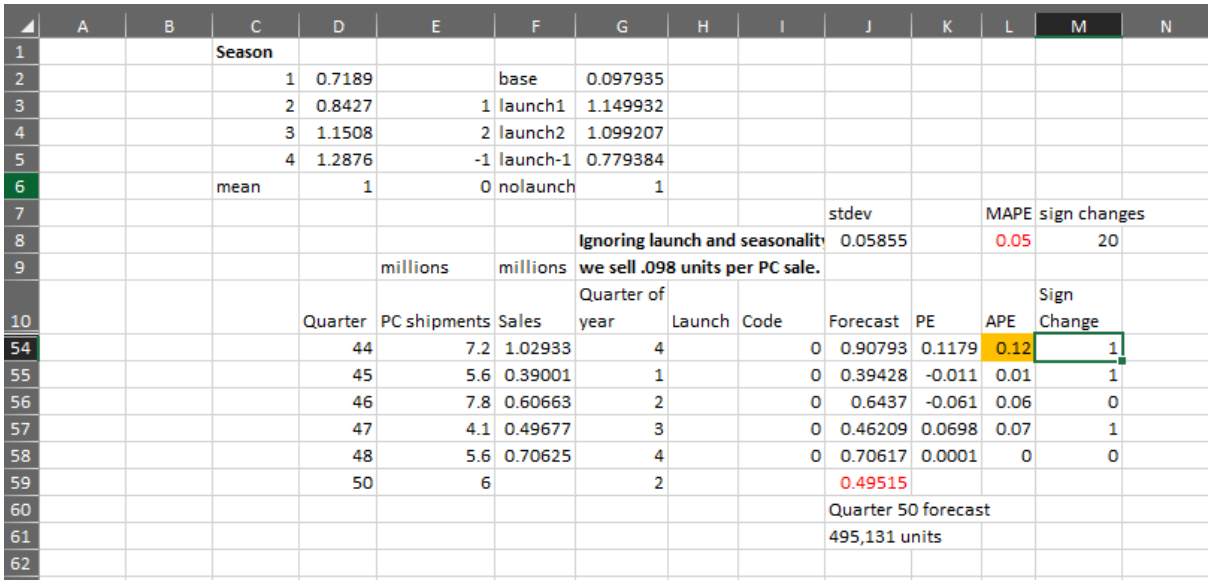
Step 1. Copy the formula =base*E11*VLOOKUP(G11,lookup,2)*VLOOKUP(I11,launch,3,FALSE) from J11 to J12:J58 to compute your forecasts for all 48 quarters of data.

Step 2. Copy the formula =(F11-J11)/F11 from K11 to K12:K58 to compute the percentage error for each quarter. For example, in Quarter 1 actual software sales were 3.0 percent higher than your prediction.

Step 3. Copy the formula =ABS(K11) from L11 to L12:L58 to compute each week's absolute percentage error.

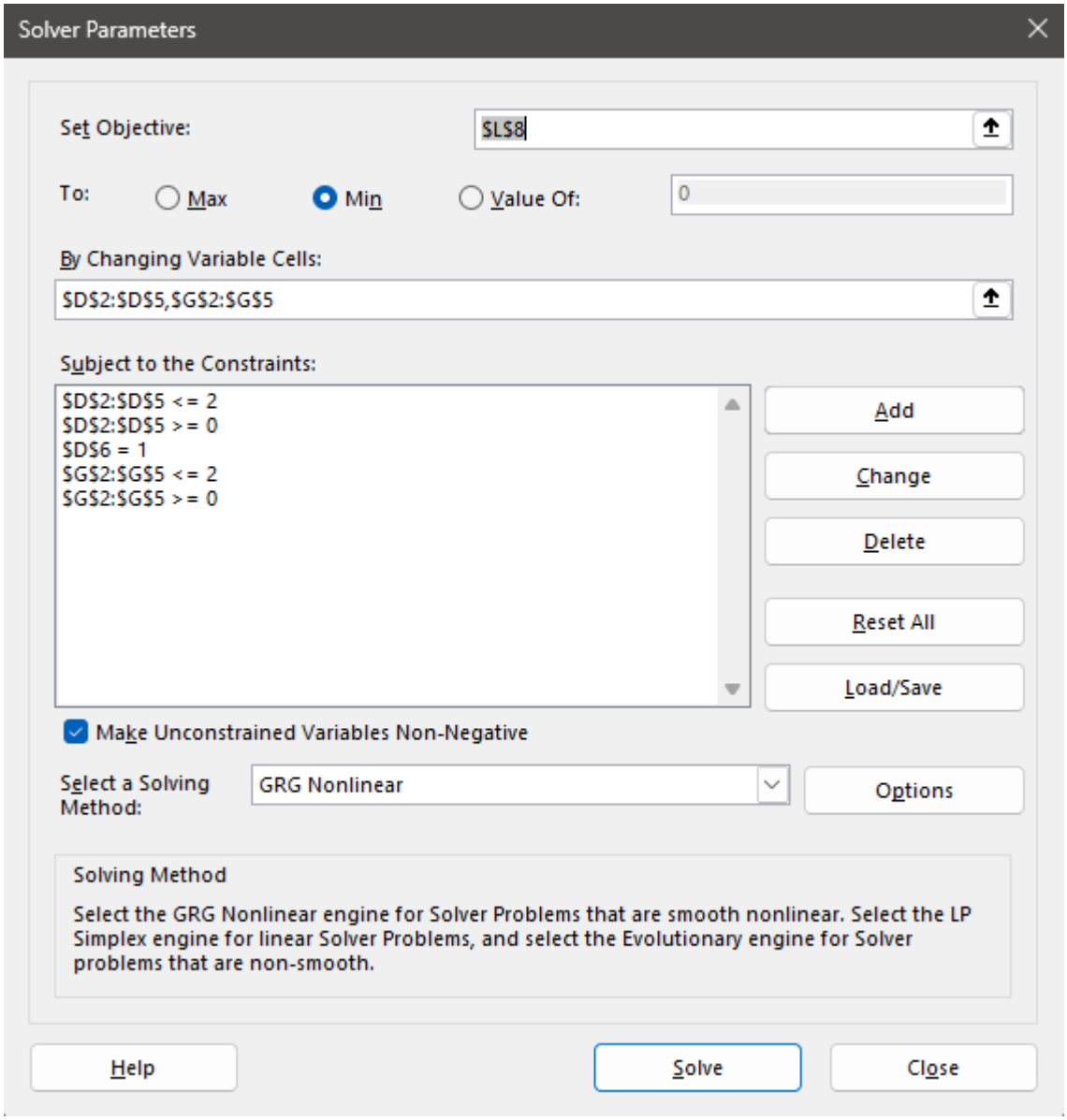
Step 4. Compute the MAPE in cell L8 with the formula =AVERAGE(L11:L58).

Step 5. In cell J8 you can compute the standard deviation of percentage errors with the formula =STDEV(K11:K58).



	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1			**Season**											
2			1	0.7189		base	0.097935							
3			2	0.8427	1	launch1	1.149932							
4			3	1.1508	2	launch2	1.099207							
5			4	1.2876	-1	launch-1	0.779384							
6			mean	1	0	nolaunch	1							
7										stdev		MAPE	sign changes	
8							**Ignoring launch and seasonality**			0.05855		0.05	20	
9					millions	millions	**we sell .098 units per PC sale.**							
10				Quarter	PC shipments	Sales	Quarter of year	Launch	Code	Forecast	PE	APE	Sign Change	
54				44	7.2	1.02933	4		0	0.90793	0.1179	0.12	1	
55				45	5.6	0.39001	1		0	0.39428	-0.011	0.01	1	
56				46	7.8	0.60663	2		0	0.6437	-0.061	0.06	0	
57				47	4.1	0.49677	3		0	0.46209	0.0698	0.07	1	
58				48	5.6	0.70625	4		0	0.70617	0.0001	0	0	
59				50	6		2			0.49515				
60										Quarter 50 forecast				
61										495,131 units				
62														

Step 6. Using the GRG multi-start Solver Engine.


Solver Parameters
Set Objective:
$L$8
To:
Max
Min
Value Of:
0
By Changing Variable Cells:
$D$2:$D$5,$G$2:$G$5
Subject to the Constraints:
$D$2:$D$5 <= 2
$D$2:$D$5 >= 0
$D$6 = 1
$G$2:$G$5 <= 2
$G$2:$G$5 >= 0
Add
Change
Delete
Reset All
Load/Save
Make Unconstrained Variables Non-Negative
Select a Solving Method:
GRG Nonlinear
Options
Solving Method
Select the GRG Nonlinear engine for Solver Problems that are smooth nonlinear. Select the LP Simplex engine for linear Solver Problems, and select the Evolutionary engine for Solver problems that are non-smooth.
Help
Solve
Close

Practical – 9

Retailing and Advertising

Q 1] Show Adstock Model

Step 1. In cell F6 compute quarter 1's Adstock level with the formula=E6+intialadstock*lambda using Quarter 1 ADSTOCK = LAMBDA * INITIALADSTOCK *QUARTER 1 ADS

Step 2. Copy the formula =E7+lambda*F6 from F7 to F8:F29 to use Quarter t ADSTOCK = LAMBDA * QUARTER T-1 ADSTOCK *QUARTER T ADS to compute the Adstock level for the remaining quarters.

Step 3. Copy the formula=(const*(trend)^D6+adeffect*F6)*VLOOKUP(C6,season,2)*(G6)^(-elasticity) fromH6 to H7:H29 to compute a forecast for each quarter's sales.

Step 4. Copy the formula =ABS(I6-H6)/I6 from J6 to J7:J29 to compute each quarter's absolute percentage error

Step 5. In cell I4 compute the MAPE with the formula =AVERAGE(J6:J29).

Step 6. Use the Solver window in Figure 1 with the GRG (Generalized Reduced Gradient) Multistart engine to find the parameter values that minimize the MAPE associatedwith Equation 3. Most of the upper bounds on the changing cells are "intelligent guesses." Of course, if the Solver set a changing cell value near its upper bound, you need to relax the bound. The constraint $M$11 = 1 ensures that the seasonal indices average to 1. The solution found by Solver is shown in Figure 2.

Fig : 1

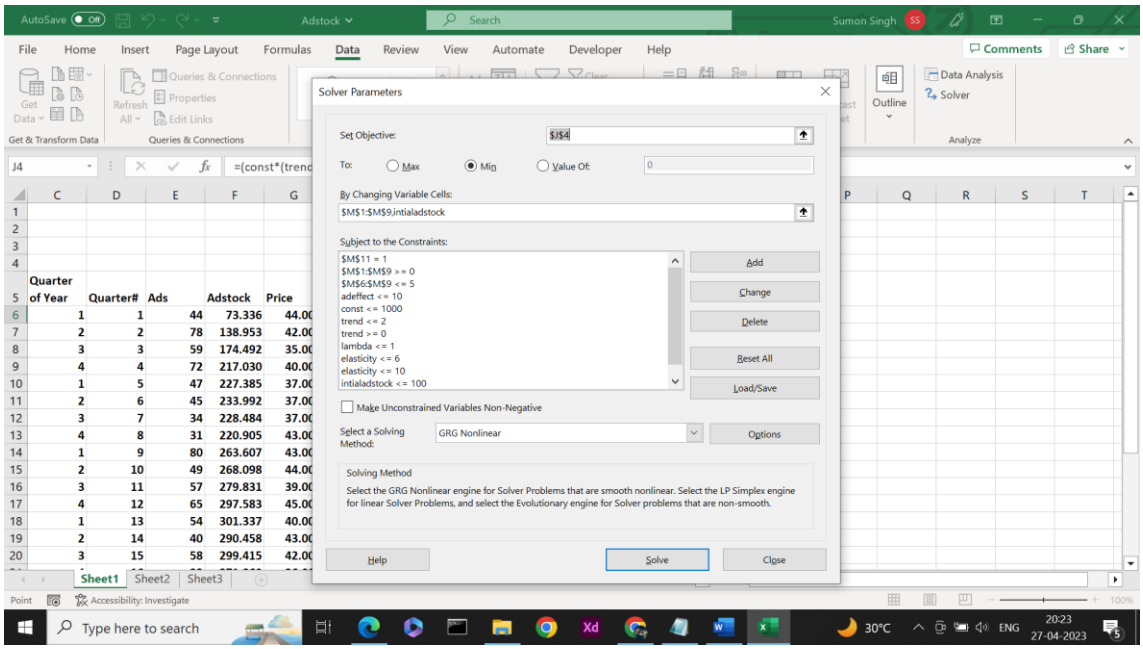
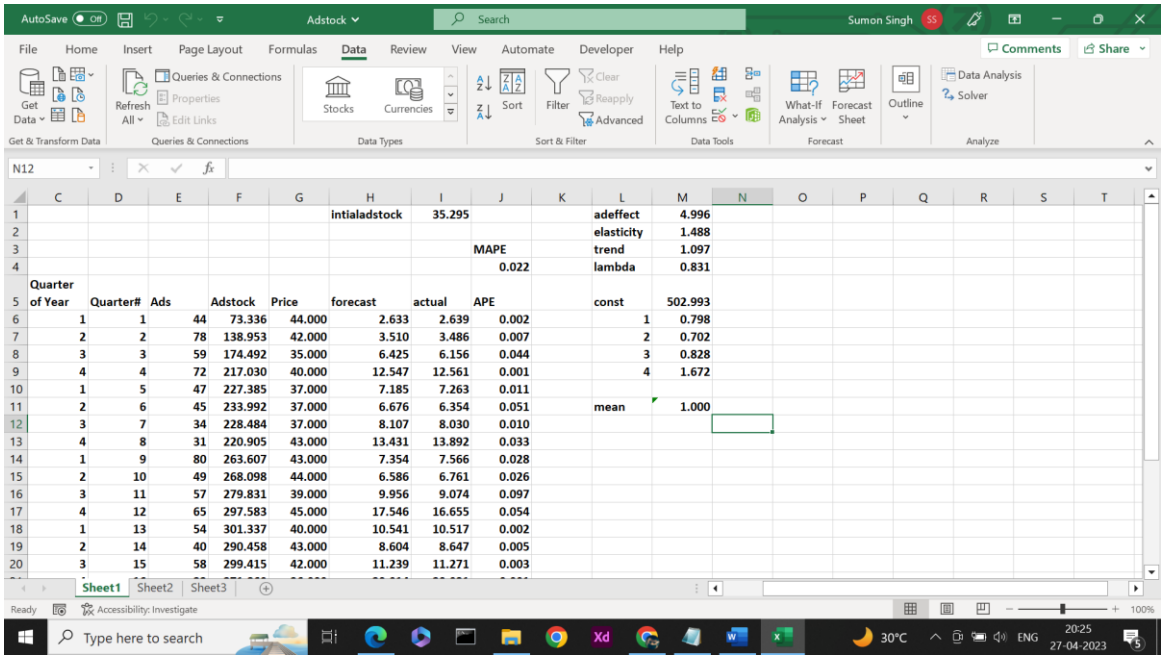


Fig : 2



	C	D	E	F	G	H	I	J	K	L	M
1						intialadstock	35.295			adeffect	4.996
2										elasticity	1.488
3								MAPE		trend	1.097
4								0.022		lambda	0.831
5	Quarter of Year	Quarter#	Ads	Adstock	Price	forecast	actual	APE		const	502.993
6	1	1	44	73.336	44.000	2.633	2.639	0.002		1	0.798
7	2	2	78	138.953	42.000	3.510	3.486	0.007		2	0.702
8	3	3	59	174.492	35.000	6.425	6.156	0.044		3	0.828
9	4	4	72	217.030	40.000	12.547	12.561	0.001		4	1.672
10	1	5	47	227.385	37.000	7.185	7.263	0.011			
11	2	6	45	233.992	37.000	6.676	6.354	0.051		mean	1.000
12	3	7	34	228.484	37.000	8.107	8.030	0.010			
13	4	8	31	220.905	43.000	13.431	13.892	0.033			
14	1	9	80	263.607	43.000	7.354	7.566	0.028			
15	2	10	49	268.098	44.000	6.586	6.761	0.026			
16	3	11	57	279.831	39.000	9.956	9.074	0.097			
17	4	12	65	297.583	45.000	17.546	16.655	0.054			
18	1	13	54	301.337	40.000	10.541	10.517	0.002			
19	2	14	40	290.458	43.000	8.604	8.647	0.005			
20	3	15	58	299.415	42.000	11.239	11.271	0.003			

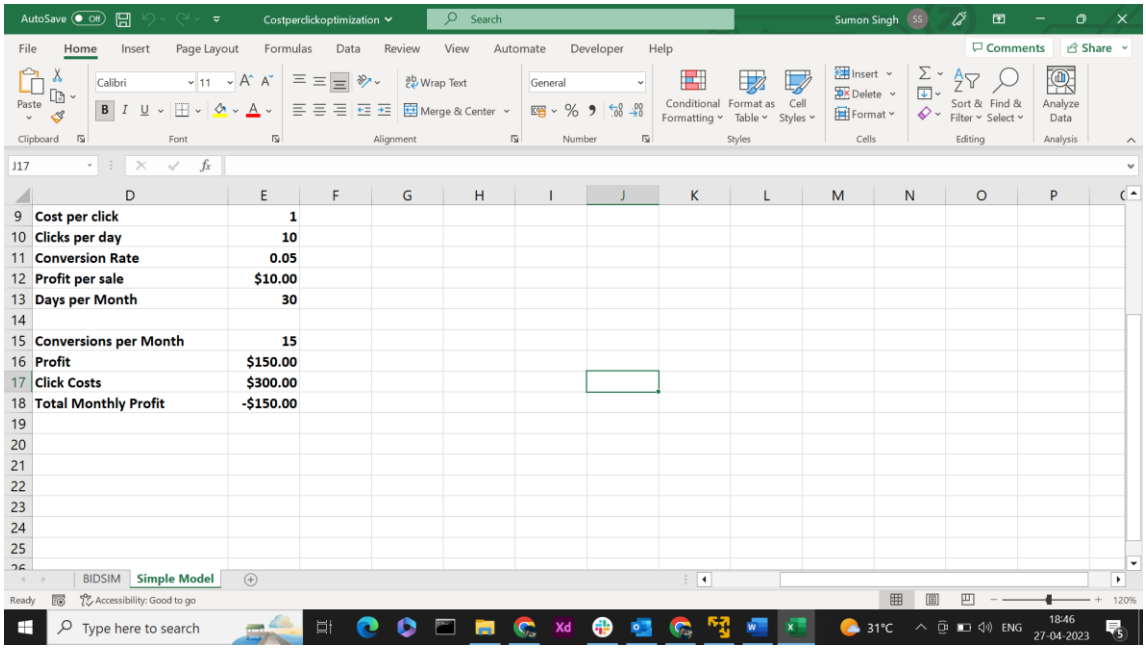
Q 2] Show the calculations needed to determine if PPC advertising can help company's bottom line.

Step 1. In E15 compute conversions per month by multiplying the conversion rate by the number of clicks per month. The exact formula is = Conversion Rate * Clicks_per_day * Days_per_Month.

Step 2. In cell E16 compute the monthly profit by multiplying the expected profit per conversion times the monthly number of conversions. The exact formula is = Conversions_per_Month * Profit_per_sale.

Step 3. In cell E17 compute the monthly click costs by multiplying the cost per click times the number of monthly clicks. The exact formula is = Clicks_per_day * Days_per_Month * Cost_per_click.

Step 4. In cell E18 compute the monthly profit with the formula = Profit-Click_Costs.



	D	E
9	Cost per click	1
10	Clicks per day	10
11	Conversion Rate	0.05
12	Profit per sale	$10.00
13	Days per Month	30
14		
15	Conversions per Month	15
16	Profit	$150.00
17	Click Costs	$300.00
18	Total Monthly Profit	-$150.00

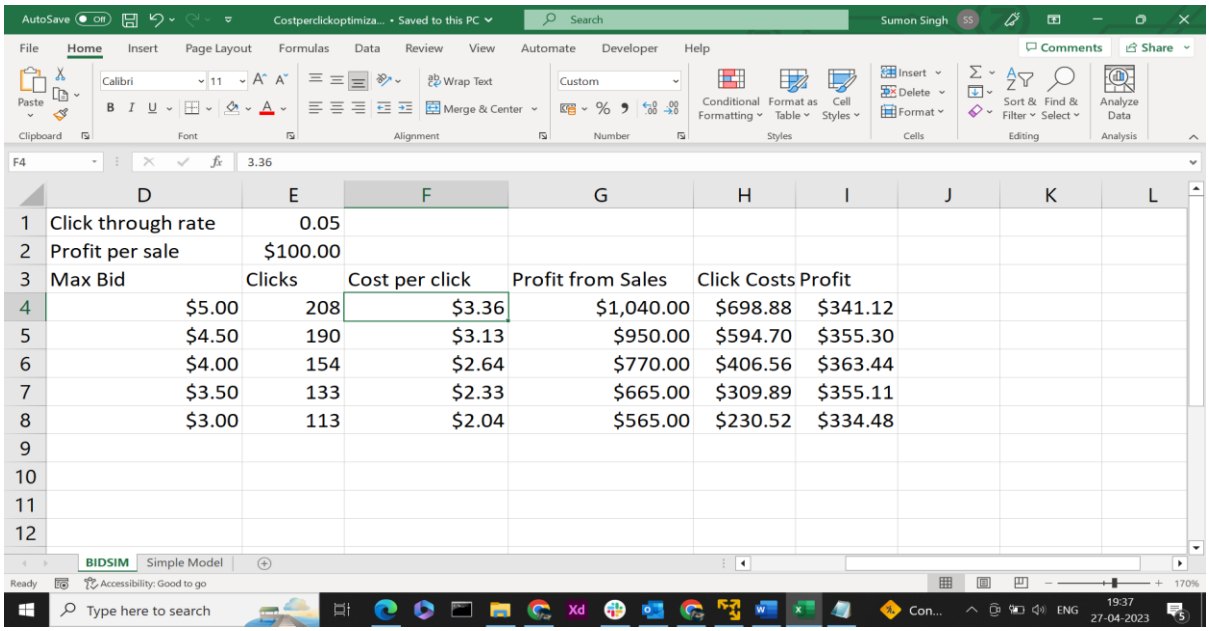
Conclusion : Monthly profit for PPC advertising is -$150.00, so it does not appear that PPC ads would be profitable.

Q 3] Show how bid simulator optimize the bid

Step 1. In cell G4:G8 calculate Profit from sales using the formula Click_through_rate*Profit_per_sale*clicks

Step 2. In cells H4:H8 calculate Click Costs using the formula click*Cost Per Clicks

Step 3. In cells I4:I8 calculate profit using the formula Profit_per_sale-Click Costs



	D	E	F	G	H	I
1	Click through rate	0.05				
2	Profit per sale	$100.00				
3	Max Bid	Clicks	Cost per click	Profit from Sales	Click Costs	Profit
4	$5.00	208	$3.36	$1,040.00	$698.88	$341.12
5	$4.50	190	$3.13	$950.00	$594.70	$355.30
6	$4.00	154	$2.64	$770.00	$406.56	$363.44
7	$3.50	133	$2.33	$665.00	$309.89	$355.11
8	$3.00	113	$2.04	$565.00	$230.52	$334.48
9						
10						
11						
12						

Practical – 10

Research Methodology Tools

Q 1] Show Principal Components Analysis

Finding 1st PCA

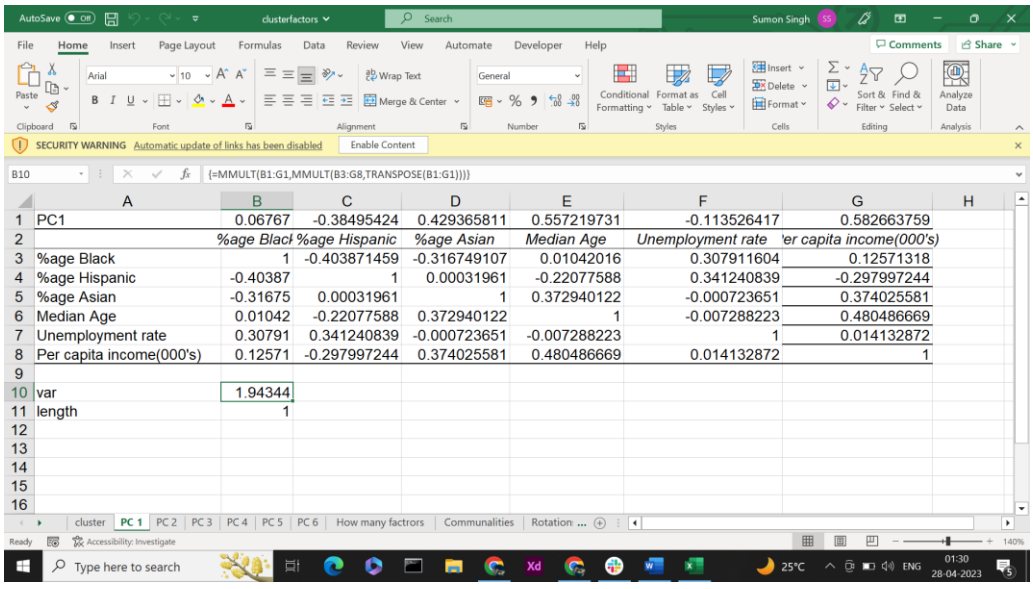
Step 1. In the cell range B1:G1, enter trial values for the first principal component's weights.

Step 2. In cell B11 compute the length of the first principal component with the formula=SUMPRODUCT(B1:G1,B1:G1).

Step 3. Using Equation 5, the variance of the first principal component is computed in cell B10 with the array entered formula=MMULT(B1:G1,MMULT(B3:G8,TRANSPOSE(B1:G1))).

Step 4. Use the Solver window, as shown in Figure 9, to determine the first principal component.

Step 5. Maximize the variance of PC1 subject to the constraint that the length of PC1equals 1. Use the GRG Multistart Engine, so bounds on the changing cells are required. Because the length of PC1 equals 1, each coefficient in PC must be less than1 in absolute value, so that provides the needed bounds.



	A	B	C	D	E	F	G	H
1	PC1	0.06767	-0.38495424	0.429365811	0.557219731	-0.113526417	0.582663759	
2		*%age Black*	*%age Hispanic*	*%age Asian*	*Median Age*	*Unemployment rate*	*er capita income(000's)*	
3	%age Black	1	-0.403871459	-0.316749107	0.01042016	0.307911604	0.12571318	
4	%age Hispanic	-0.40387	1	0.00031961	-0.22077588	0.341240839	-0.297997244	
5	%age Asian	-0.31675	0.00031961	1	0.372940122	-0.000723651	0.374025581	
6	Median Age	0.01042	-0.22077588	0.372940122	1	-0.007288223	0.480486669	
7	Unemployment rate	0.30791	0.341240839	-0.000723651	-0.007288223	1	0.014132872	
8	Per capita income(000's)	0.12571	-0.297997244	0.374025581	0.480486669	0.014132872	1	
9								
10	var	1.94344						
11	length	1						
12								
13								
14								
15								
16								

Q 2] MDS Analysis of U.S. City Distances

Step 1. In the range G3:H31 enter trial values for the x and y coordinates of each city in two-dimensional space. Arbitrarily restrict each city's x and y coordinate to be between 0 and 10

Step 2. Copy the formula =RANK(K3,distances,1) from K34 to the range K34:AM62tocompute the ranking of the distances between each pair of cities. The last

argument of 1 in this formula ensures that the smallest distance (New York to Brooklyn) receives a rank of 1, and so on. All diagonal entries in the RANK matrix contain 813 because we assigned a large distance to diagonal entries

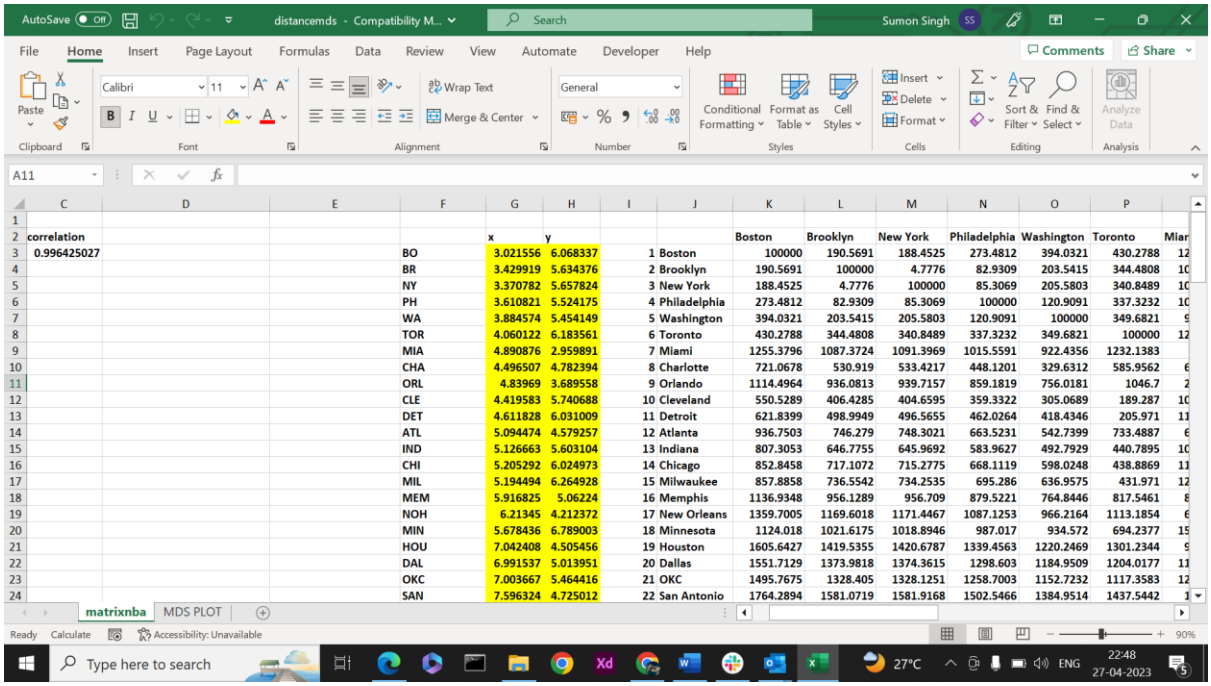
Step 3. Copy the formula =IF($I66=K$64,10000000,(OFFSET($G$2,$I66,0,1,1)-OFFSET($G$2,K$64,0,1,1))^2+(OFFSET($H$2,$I66,0,1,1)-OFFSET($H$2,K$64,0,1,1))^2) from K66 to the range K66:AM94 to compute for each pair of different cities the square of the two-dimensional distances between each pair of cities. The term OFFSET($G$2,$I66,0,1,1) in the formula pulls the x coordinate of the city in the current row; the term OFFSET($G$2,K$64,0,1,1) pulls the x coordinate of the city in the current column; the term OFFSET($G$2,$I66,0,1,1) in the formula pulls the y coordinate of the city in the current row; and the term OFFSET($H$2,K$64,0,1,1) pulls the y coordinate of the city in the current column. For distances corresponding to the same city twice, assign a huge distance (say 10 million miles).

Step 4. Our strategy is to have Solver choose the two-dimensional locations of the cities so that the ranking of the distances in the two-dimensional space closely matches the actual rankings of the distances. To accomplish this goal compute the ranking of the distances in two-dimensional space.

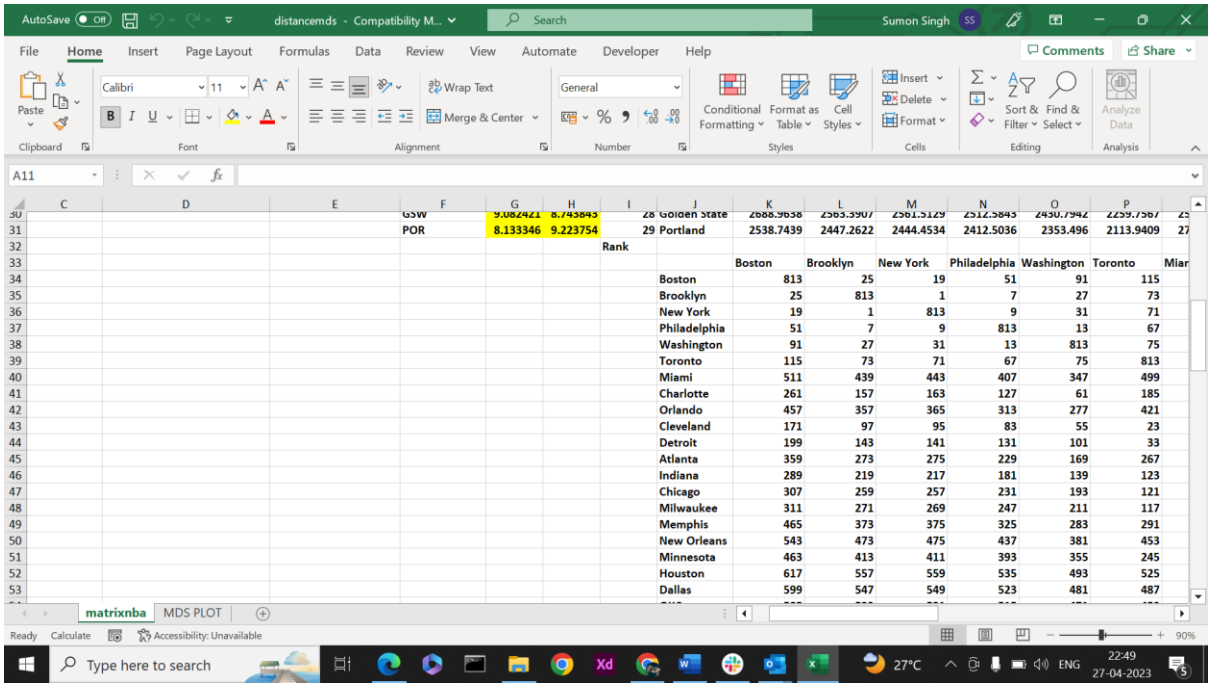
Step 5. Copy the formula =RANK(K66,twoddistances,1) from K98 to K98:AM126 to compute the rankings of the distances in two-dimensional space. For example, for the two-dimensional locations in G3:H31, Brooklyn and New York are the closest pair of cities.

Step 6. In cell C3 compute the correlation between the original similarity ranks and the two-dimensional ranks with the formula =CORREL(originalranks,twodranks). The range K34:AM62 is named original ranks and the range K98:AM126 is namedtwodranks

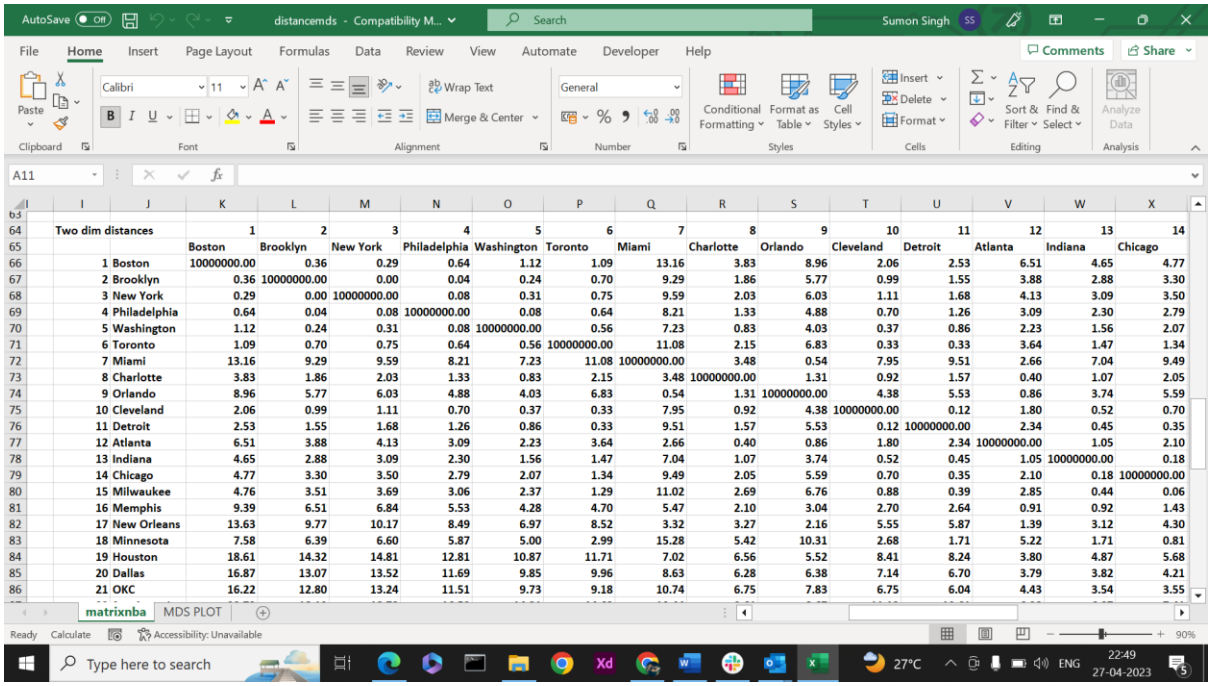
Step 7. Use the Evolutionary Solver to locate each city in two-dimensional space to maximize the correlation between the original ranks and the ranks in two dimensions. This should ensure that cities that are actually close together will be close in two- dimensional space.



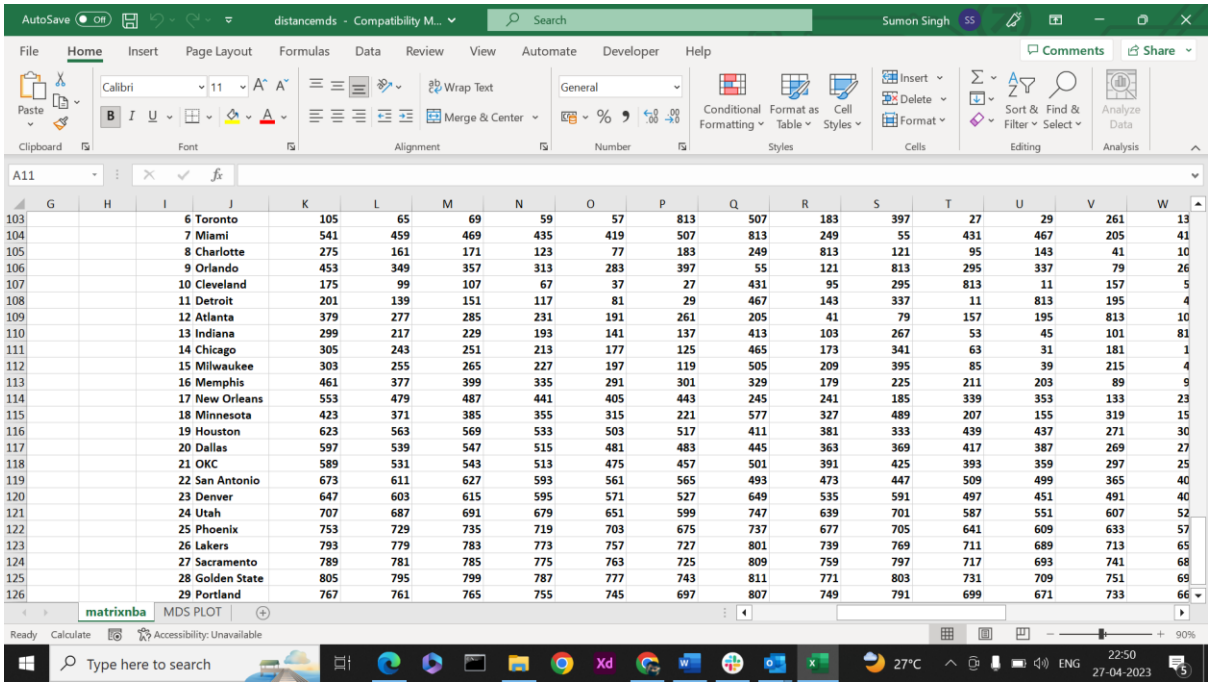
	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1														
2	correlation				x	y			Boston	Brooklyn	New York	Philadelphia	Washington	Toronto
3	0.996425027			BO	3.021556	6.068337	1	Boston	100000	190.5691	188.4525	273.4812	394.0321	430.2788
4				BR	3.429919	5.634376	2	Brooklyn	190.5691	100000	4.7776	82.9309	203.5415	344.4808
5				NY	3.370782	5.657824	3	New York	188.4525	4.7776	100000	85.3069	205.5803	340.8489
6				PH	3.610821	5.524175	4	Philadelphia	273.4812	82.9309	85.3069	100000	120.9091	337.3232
7				WA	3.884574	5.454149	5	Washington	394.0321	203.5415	205.5803	120.9091	100000	349.6821
8				TOR	4.060122	6.183561	6	Toronto	430.2788	344.4808	340.8489	337.3232	349.6821	100000
9				MIA	4.890876	2.959891	7	Miami	1255.3796	1087.3724	1091.3969	1015.5591	922.4356	1232.1383
10				CHA	4.496507	4.782394	8	Charlotte	721.0678	530.919	533.4217	448.1201	329.6312	585.9562
11				ORL	4.83969	3.689558	9	Orlando	1114.4964	936.0813	939.7157	859.1819	756.0181	1046.7
12				CLE	4.419583	5.740688	10	Cleveland	550.5289	406.4285	404.6595	359.3322	305.0689	189.287
13				DET	4.611828	6.031009	11	Detroit	621.8399	498.9949	496.5655	462.0264	418.4346	205.971
14				ATL	5.094474	4.579257	12	Atlanta	936.7503	746.279	748.3021	663.5231	542.7399	733.4887
15				IND	5.126663	5.603104	13	Indiana	807.3053	646.7755	645.9692	583.9627	492.7929	440.7895
16				CHI	5.205292	6.024973	14	Chicago	852.8458	717.1072	715.2775	668.1119	598.0248	438.8869
17				MIL	5.194494	6.264928	15	Milwaukee	857.8858	736.5542	734.2535	695.286	636.9575	431.971
18				MEM	5.916825	5.06224	16	Memphis	1136.9348	956.1289	956.709	879.5221	764.8446	817.5461
19				NOH	6.21345	4.212372	17	New Orleans	1359.7005	1169.6018	1171.4467	1087.1253	966.2164	1113.1854
20				MIN	5.678436	6.789003	18	Minnesota	1124.018	1021.6175	1018.8946	987.017	934.572	694.2377
21				HOU	7.042408	4.505456	19	Houston	1605.6427	1419.5355	1420.6787	1339.4563	1220.2469	1301.2344
22				DAL	6.991537	5.013951	20	Dallas	1551.7129	1373.9818	1374.3615	1298.603	1184.9509	1204.0177
23				OKC	7.003667	5.464416	21	OKC	1495.7675	1328.405	1328.1251	1258.7003	1152.7232	1117.3583
24				SAN	7.596324	4.725012	22	San Antonio	1764.2894	1581.0719	1581.9168	1502.5466	1384.9514	1437.5442



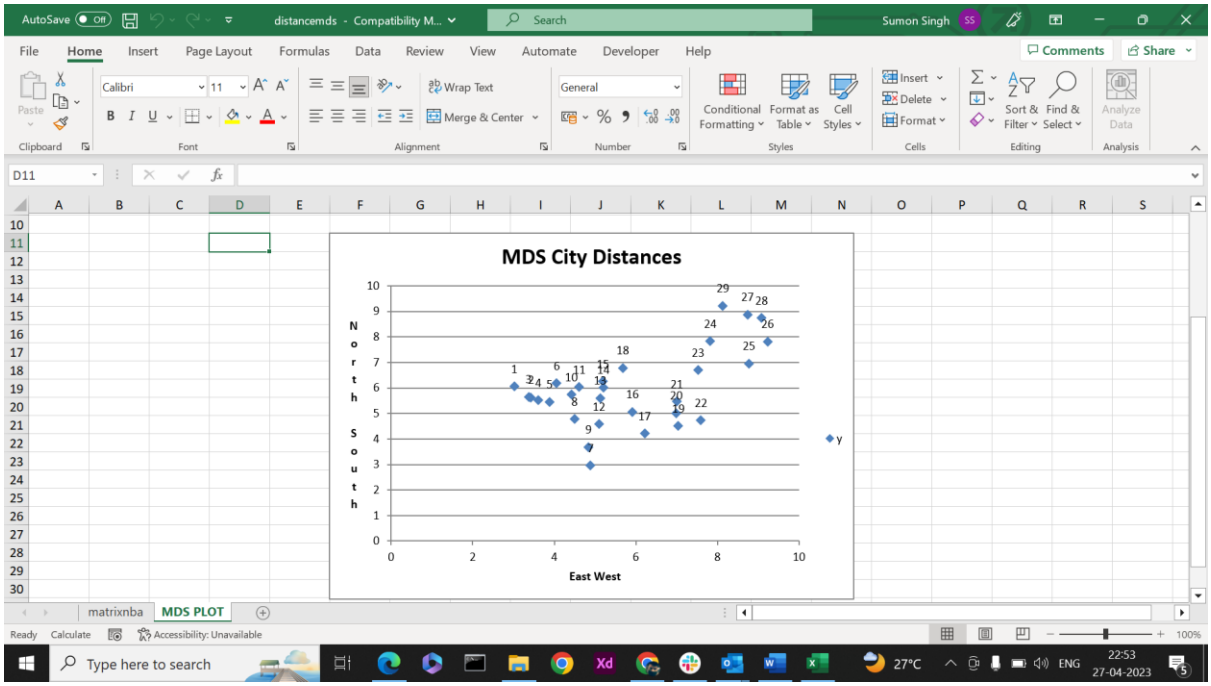
	C	D	E	F	G	H	I	J	K	L	M	N	O	P
31				POR	8.133346	9.223754	29	Portland	2538.7439	2447.2622	2444.4534	2412.5036	2353.496	2113.9409
32							Rank							
33									Boston	Brooklyn	New York	Philadelphia	Washington	Toronto
34								Boston	813	25	19	51	91	115
35								Brooklyn	25	813	1	7	27	73
36								New York	19	1	813	9	31	71
37								Philadelphia	51	7	9	813	13	67
38								Washington	91	27	31	13	813	75
39								Toronto	115	73	71	67	75	813
40								Miami	511	439	443	407	347	499
41								Charlotte	261	157	163	127	61	185
42								Orlando	457	357	365	313	277	421
43								Cleveland	171	97	95	83	55	23
44								Detroit	199	143	141	131	101	33
45								Atlanta	359	273	275	229	169	267
46								Indiana	289	219	217	181	139	123
47								Chicago	307	259	257	231	193	121
48								Milwaukee	311	271	269	247	211	117
49								Memphis	465	373	375	325	283	291
50								New Orleans	543	473	475	437	381	453
51								Minnesota	463	413	411	393	355	245
52								Houston	617	557	559	535	493	525
53								Dallas	599	547	549	523	481	487



Two dim distances		1	2	3	4	5	6	7	8	9	10	11	12	13	14
		Boston	Brooklyn	New York	Philadelphia	Washington	Toronto	Miami	Charlotte	Orlando	Cleveland	Detroit	Atlanta	Indiana	Chicago
1	Boston	10000000.00	0.36	0.29	0.64	1.12	1.09	13.16	3.83	8.96	2.06	2.53	6.51	4.65	4.77
2	Brooklyn	0.36	10000000.00	0.00	0.04	0.24	0.70	9.29	1.86	5.77	0.99	1.55	3.88	2.88	3.30
3	New York	0.29	0.00	10000000.00	0.08	0.31	0.75	9.59	2.03	6.03	1.11	1.68	4.13	3.09	3.50
4	Philadelphia	0.64	0.04	0.08	10000000.00	0.08	0.64	8.21	1.33	4.88	0.70	1.26	3.09	2.30	2.79
5	Washington	1.12	0.24	0.31	0.08	10000000.00	0.56	7.23	0.83	4.03	0.37	0.86	2.23	1.56	2.07
6	Toronto	1.09	0.70	0.75	0.64	0.56	10000000.00	11.08	2.15	6.83	0.33	0.33	3.64	1.47	1.34
7	Miami	13.16	9.29	9.59	8.21	7.23	11.08	10000000.00	3.48	0.54	7.95	9.51	2.66	7.04	9.49
8	Charlotte	3.83	1.86	2.03	1.33	0.83	2.15	3.48	10000000.00	1.31	0.92	1.57	0.40	1.07	2.05
9	Orlando	8.96	5.77	6.03	4.88	4.03	6.83	0.54	1.31	10000000.00	4.38	5.53	0.86	3.74	5.59
10	Cleveland	2.06	0.99	1.11	0.70	0.37	0.33	7.95	0.92	4.38	10000000.00	0.12	1.80	0.52	0.70
11	Detroit	2.53	1.55	1.68	1.26	0.86	0.33	9.51	1.57	5.53	0.12	10000000.00	2.34	0.45	0.35
12	Atlanta	6.51	3.88	4.13	3.09	2.23	3.64	2.66	0.40	0.86	1.80	2.34	10000000.00	1.05	2.10
13	Indiana	4.65	2.88	3.09	2.30	1.56	1.47	7.04	1.07	3.74	0.52	0.45	1.05	10000000.00	0.18
14	Chicago	4.77	3.30	3.50	2.79	2.07	1.34	9.49	2.05	5.59	0.70	0.35	2.10	0.18	10000000.00
15	Milwaukee	4.76	3.51	3.69	3.06	2.37	1.29	11.02	2.69	6.76	0.88	0.39	2.85	0.44	0.06
16	Memphis	9.39	6.51	6.84	5.53	4.28	4.70	5.47	2.10	3.04	2.70	2.64	0.91	0.92	1.43
17	New Orleans	13.63	9.77	10.17	8.49	6.97	8.52	3.32	3.27	2.16	5.55	5.87	1.39	3.12	4.30
18	Minnesota	7.58	6.39	6.60	5.87	5.00	2.99	15.28	5.42	10.31	2.68	1.71	5.22	1.71	0.81
19	Houston	18.61	14.32	14.81	12.81	10.87	11.71	7.02	6.56	5.52	8.41	8.24	3.80	4.87	5.68
20	Dallas	16.87	13.07	13.52	11.69	9.85	9.96	8.63	6.28	6.38	7.14	6.70	3.79	3.82	4.21
21	OKC	16.22	12.80	13.24	11.51	9.73	9.18	10.74	6.75	7.83	6.75	6.04	4.43	3.54	3.55



6	Toronto	105	65	69	59	57	813	507	183	397	27	29	261	13
7	Miami	541	459	469	435	419	507	813	249	55	431	467	205	41
8	Charlotte	275	161	171	123	77	183	249	813	121	95	143	41	10
9	Orlando	453	349	357	313	283	397	55	121	813	295	337	79	26
10	Cleveland	175	99	107	67	37	27	431	95	295	813	11	157	5
11	Detroit	201	139	151	117	81	29	467	143	337	11	813	195	4
12	Atlanta	379	277	285	231	191	261	205	41	79	157	195	813	10
13	Indiana	299	217	229	193	141	137	413	103	267	53	45	101	81
14	Chicago	305	243	251	213	177	125	465	173	341	63	31	181	1
15	Milwaukee	303	255	265	227	197	119	505	209	395	85	39	215	4
16	Memphis	461	377	399	335	291	301	329	179	225	211	203	89	9
17	New Orleans	553	479	487	441	405	443	245	241	185	339	353	133	23
18	Minnesota	423	371	385	355	315	221	577	327	489	207	155	319	15
19	Houston	623	563	569	533	503	517	411	381	333	439	437	271	30
20	Dallas	597	539	547	515	481	483	445	363	369	417	387	269	27
21	OKC	589	531	543	513	475	457	501	391	425	393	359	297	25
22	San Antonio	673	611	627	593	561	565	493	473	447	509	499	365	40
23	Denver	647	603	615	595	571	527	649	535	591	497	451	491	40
24	Utah	707	687	691	679	651	599	747	639	701	587	551	607	52
25	Phoenix	753	729	735	719	703	675	737	677	705	641	609	633	57
26	Lakers	793	779	783	773	757	727	801	739	769	711	689	713	65
27	Sacramento	789	781	785	775	763	725	809	759	797	717	693	741	68
28	Golden State	805	795	799	787	777	743	811	771	803	731	709	751	69
29	Portland	767	761	765	755	745	697	807	749	791	699	671	733	66



Q 3] Show One-way ANOVA

Step 1. On the Data tab, click Data Analysis, and then select ANOVA: Single Factor

Step 2. Select input range, including labels, is in cells B3:D8

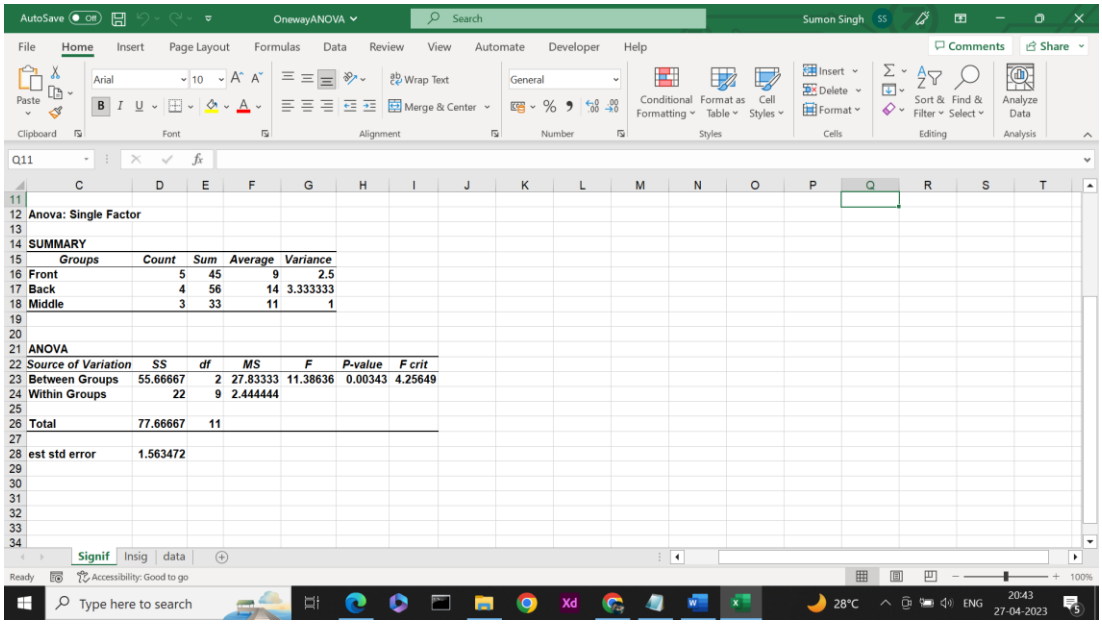
Step 3. Select the Labels option because the first row of input range contains labels.

Step 4. Select the Columns option because the data is organized in columns.

Step 5. Select C12 as the upper-left cell of the output range.

Step 6. The selected Alpha value is not important. Just use the default value.

Step 7. Click OK, to obtain the results.



Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
Front	5	45	9	2.5
Back	4	56	14	3.333333
Middle	3	33	11	1

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	55.66667	2	27.83333	11.38636	0.00343	4.25649
Within Groups	22	9	2.444444			
Total	77.66667	11				

est std error 1.563472

Conclusion : the p-value of 0.003 (in cell H23) is less than 0.05, we can conclude that these means are significantly different, so the Null Hypothesis of identical group means is rejected.

Q 4] Two-way ANOVA without Replication

Step 1. To get two way anova first go to Microsoft Office Excel, click Data Analysis on the Data tab, and then select the Anova: Two-Factor With Replication in the Data Analysis dialog box.

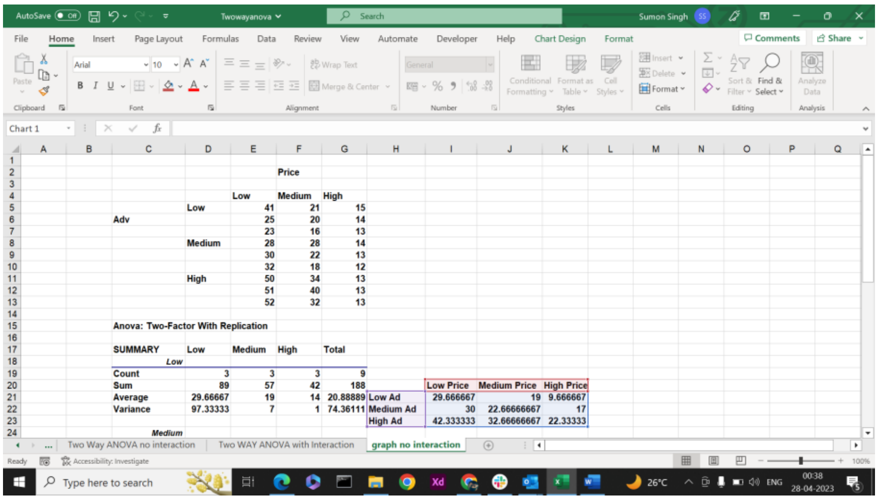
Step 2. The input range data, including labels, is in C3:F12

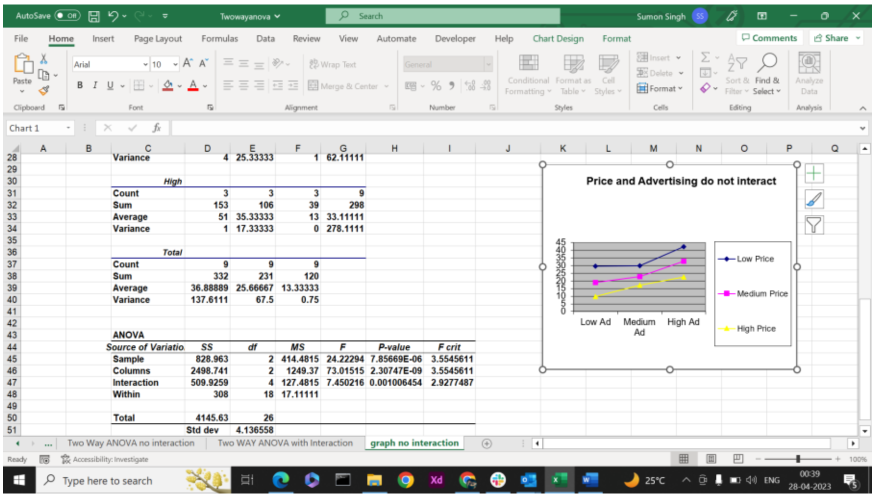
Step 3. In two-way ANOVA with replication, Excel requires a label for each level of the column effect in the first row of each column in the input range. Thus, enter low, medium, and high in cells D3:F3 to indicate the possible price levels.

Step 4. Excel also requires a label for each level of the row effect in the first column of the input range. These labels must appear in the row that marks the beginning of the data for each level. Thus place labels corresponding to low, medium, and high levels of advertising in cells C4, C7, and C10

Step 5. In the Rows Per Sample box, enter 3 because we have three replications for each combination of price and advertising level.

Step 6. Enter B14 in the upper-left cell of the output range.





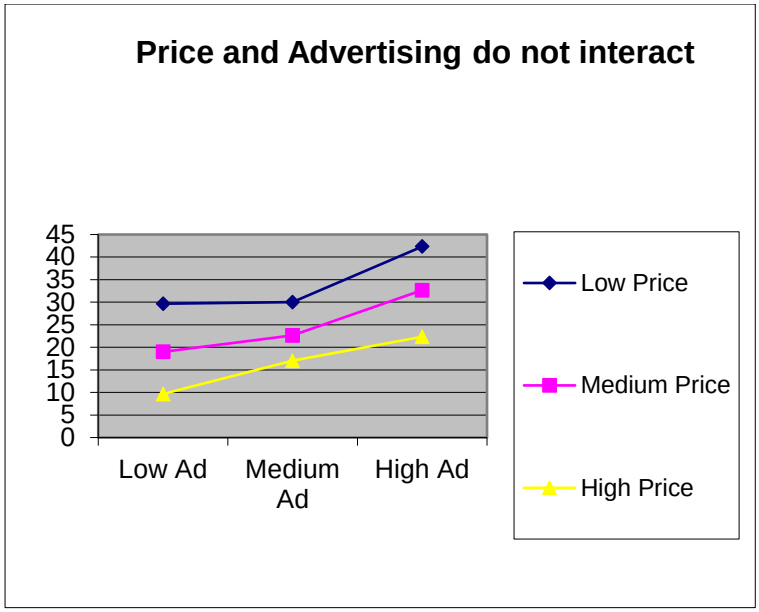
An effect (including interactions) is significant if it has a p-value that's less than 0.05. Sample (this is the row for advertising effect) and Price(shown in the row labelled Columns) are highly significant and there is no significant interaction. (The interaction p-value is 0.79!) Therefore, we can conclude that price and advertising influence sales and that the effect of advertising on sales does not depend on the price level.

Step 7. Following graph demonstrates that price and advertising do not exhibit a significant interaction. To create this chart, complete the following steps –

1. In the cell range I20:K22 compute the average sales for each price and advertising combination.

2. Select the range I19:K22 and select the first line chart option.

3. From the Design tab choose Switch Row/Column to place advertising categories on the x-axis.


Price and Advertising do not interact
45
40
35
30
25
20
15
10
5
0
Low Ad
Medium Ad
High Ad
Low Price
Medium Price
High Price