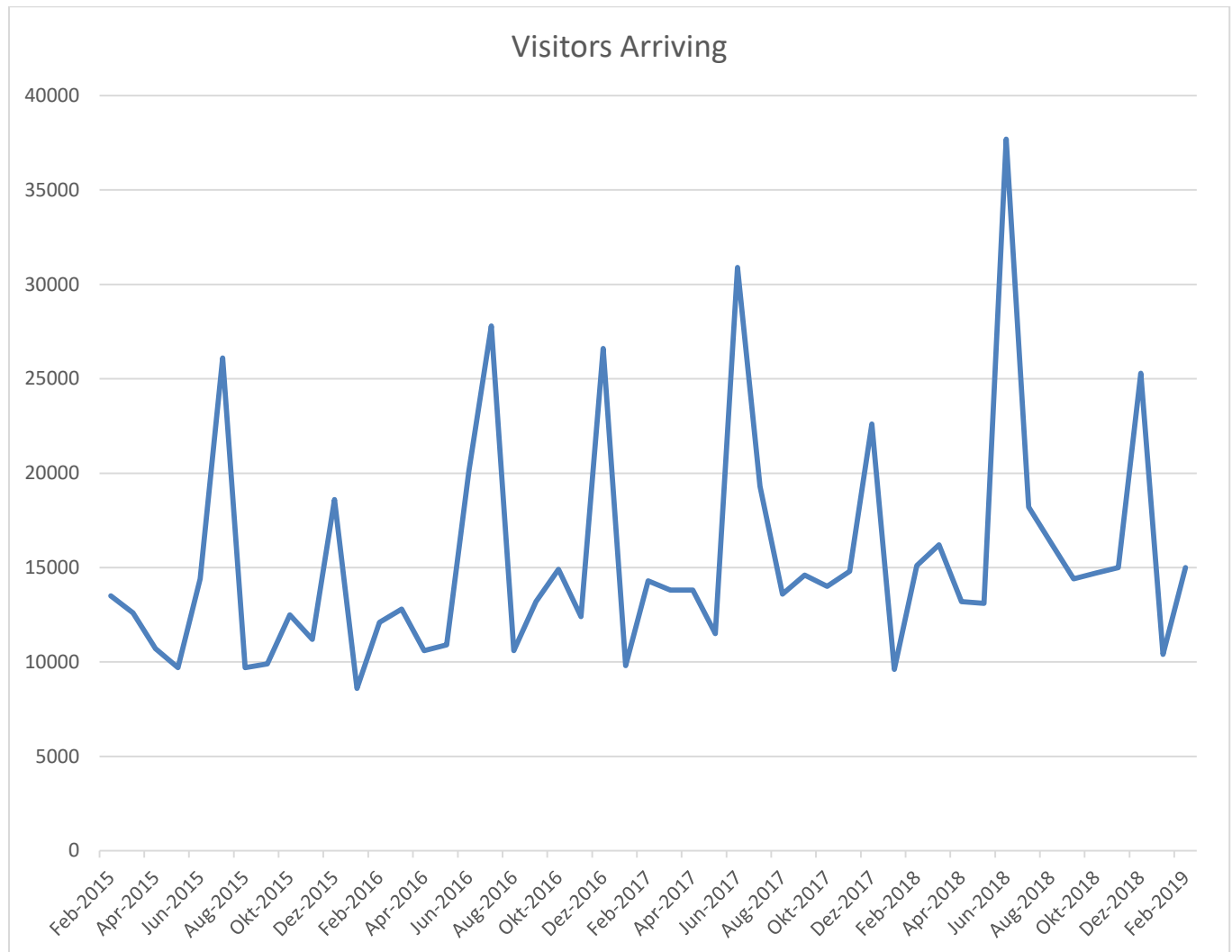


“Question Zero – referring to your last case study report.

Refer to your data that you used for Question #5 in Case Study Report #1.

Provide a line chart of the relevant time series for the last 4 years.

Include a relevant heading for your line chart – clearly identifying the country.”



1. “Estimate the parameters of a multiple linear regression model for visitor arrivals using an intercept, a time variable and a suitable set of dummy variables for the month effect. Explain what you have done, provide relevant screenshots, and write out the overall regression equation.”
2. “Explain the meaning of the intercept, the coefficient of the time variable, as well as the coefficient for the June dummy variable. (If June is your baseline, then change June to September).”

Answer 1: Following are the estimated parameters of multiple linear regression model:

Date: Monthly date for past 4 years

Time: A time index

Visitors Arriving: Number of visitors arriving from Indonesia

Dum March: Dummy of March, Dum April: Dummy of April, Dum May: Dummy of May, Dum June: Dummy of June, Dum July: Dummy of July, Dum August: Dummy of August, Dum Sept: Dummy of September, Dum Oct: Dummy of October, Dum Nov: Dummy of November, Dum Dec: Dummy of December, Dum Jan: Dummy of January.

From the equation: $Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_{12}$.

When we put our results, we get:

$Y = 11238.78 + 110.448 * \text{Time} + 402.24 * \text{Dm} - 1483.20 * \text{Da} - 2368.65 * \text{Dm} + 11995.89 * \text{Dj} + 8960.44 * \text{Dj} - 1450 * \text{Da} - 1085.45 * \text{Ds} - 195.89 * \text{Do} - 981.35 * \text{Dn} + 8833.21 * \text{Dd} - 4952.24 * \text{Dj}$.

SUMMARY OUTPUT							
Regression Statistics							
Multiple R	0.876993443						
R Square	0.769117499						
Adjusted R Square	0.692156665						
Standard Error	3367.061096						
Observations	49						
ANOVA							
	df	SS	MS	F	Significance F		
Regression	12	1359584385	113298698.7	9.993622218	3.4118E-08		
Residual	36	408135615.4	11337100.43				
Total	48	1767720000					
	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%
Intercept	11238.78205	1739.091145	6.462445678	1.68229E-07	7711.741734	14765.82237	7711.741734
Time	110.4487179	34.80271079	3.173566525	0.00307841	39.86554898	181.0318869	39.86554898
Dum March	402.2435897	2265.386481	0.177560691	0.86006332	-4192.173143	4996.660322	-4192.173143
Dum April	-1483.205128	2262.979198	-0.655421459	0.516363244	-6072.739664	3106.329408	-6072.739664
Dum May	-2368.653846	2261.105095	-1.047564685	0.301821791	-6954.387525	2217.079833	-6954.387525
Dum June	11995.89744	2259.765498	5.308470036	5.83689E-06	7412.880585	16578.91429	7412.880585
Dum July	8960.448718	2258.961359	3.966623281	0.000332293	4379.062738	13541.8347	4379.062738
Dum August	-1450	2258.693249	-0.641964109	0.524963524	-6030.842228	3130.842228	-6030.842228
Dum Sept	-1085.448718	2258.961359	-0.480507873	0.633773535	-5666.834698	3495.937262	-5666.834698
Dum Oct	-195.8974359	2259.765498	-0.086689276	0.931398889	-4778.914286	4387.119415	-4778.914286
Dum Nov	-981.3461538	2261.105095	-0.434011739	0.666870002	-5567.079833	3604.387525	-5567.079833
Dum Dec	8833.205128	2262.979198	3.903352331	0.000399332	4243.670592	13422.73966	4243.670592
Dum Jan	-4952.24359	2265.386481	-2.186048001	0.03539679	-9546.660322	-357.8268573	-9546.660322

Answer 2: The intercept basically means that the number of visitors arriving in the base month or when all the variables are zero. While the coefficient of time is explaining the variation in number of visitors. The June coefficient which has a value of 11995.897 represents more visitors in the month of February as compared to June.

3. “What is the R-squared for your regression? Explain what this represents, and comment on the importance (or not) of this statistic.”
4. “Write out the regression equation for each of the following months: March, June, September, and December. Draw a graph that represents the four different regression lines for each of the following months: March, June, September, and December.”

Answer:3. The value for R sq. from the output of Regression Statistics is: 0.769117.

It represents that approximately 76 % of variability of Y (dependent variable) can be explained by the entire set of independent variables. Further we may say that rest of the 24% variation in dependent variable is explained by some other factors which we could not include in our model.

These factors could be some explanatory variable such as easy visa facilities and the tourist rating of the country. This 76% tells the model fitness. We may also say that all independent variable determining the dependent variable by 76 %. In addition, it used to compare between the models. Higher indicates a better model.

Answer:4.

March: $Y = \text{Intercept} + \text{time} + \text{Dum March}$

$$Y = (11238.78 + 110.45) + 402.24 * 1$$

June: $Y = \text{Intercept} + \text{time} + \text{Dum June}$

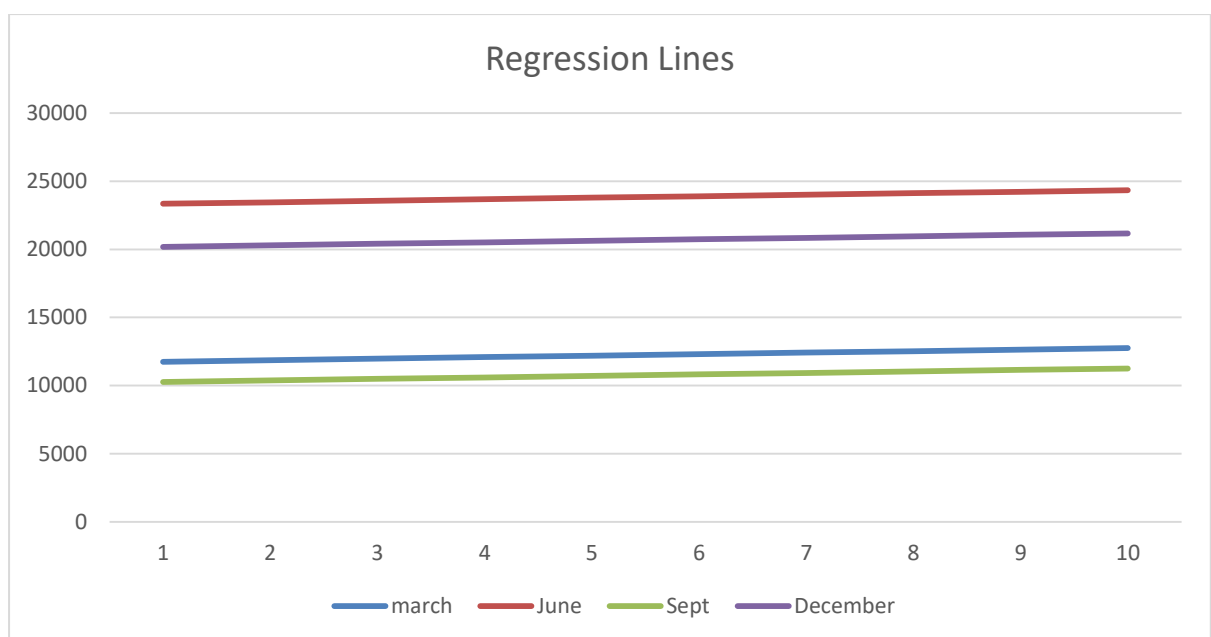
$$Y = (11238.78 + 110.45) + 11995.89 * 1$$

September: $Y = \text{Intercept} + \text{time} + \text{Dum September}$

$$Y = (11238.78 + 110.45) + (-1085.45) * 1$$

December: $Y = \text{Intercept} + \text{time} + \text{Dum December}$

$$Y = (11238.78 + 110.45) + 8833.21 * 1$$



5. “Conduct a test for the overall significance of the model. Explain and analyze your results.”

Hypothesis:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_{12} = 0$$

$$H_1: \text{at least one } \beta_i \neq 0$$

Test Statistic:

To test this hypothesis F-test is the most appropriate.

P-value:

If P-value ≤ 0.05 we accept H_1 .

Decision:

As P-value is $3.4118E08 \leq 0.05$.

Conclusion:

Hence, we accept H_1 and results are significant. It indicates all independent variables significantly explaining the variation of dependent variable. In this way, we can apply sample results on entire population.

6. “Conduct a test for the individual significance of the coefficient for the time variable. Explain and analyze your results.”

“Conduct a test for the individual significance of the coefficient for the December dummy variable. (If December is your baseline, then change December to March). Explain and analyze your results.”

FOR TIME:

Hypothesis:

$$H_0: \beta_{\text{time}} = 0$$

$$H_1: \text{at least one } \beta_{\text{time}} \neq 0$$

Test Statistic:

To test this hypothesis t-test is the most appropriate.

P-value:

If P-value ≤ 0.05 we accept H_1 .

Decision:

As P-value is $0.00307841 \leq 0.05$.

Conclusion:

Hence, we reject H_0 , the model is significant. It indicates time variable significantly explaining the variation of dependent variable.

FOR DECEMBER:

Hypothesis:

$$H_0: \beta_{12} = 0$$

$$H_1: \text{at least one } \beta_i \neq 0$$

Test Statistic:

To test this hypothesis t-test is the most appropriate.

P-value:

If P-value ≤ 0.05 we accept H_1 .

Decision:

As P-value is $0.000399332 \leq 0.05$.

Conclusion:

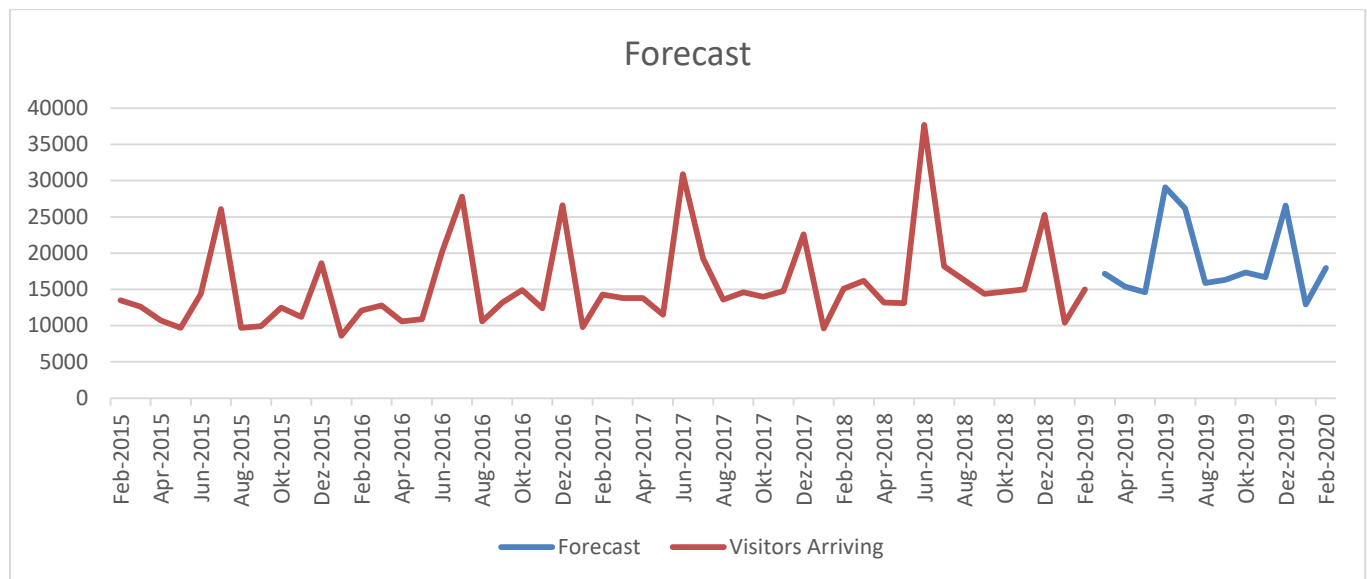
Hence, we reject H_0 , the model is significant. It indicates December dummy variable significantly explaining the variation of dependent variable Y.

7. “Use your regression equation to forecast tourist arrivals for each month beyond your sample period. Show your working for **only** the December forecast as an example. Plot your original data against time and include your forecasts. Clearly differentiate between the original data and the forecasts.”

Answer:7: Regression equation $Y = a + b_1x_1 + b_2x_2 + \dots + b_nx_n$.

For December: $t = 59$

So, No. of Visitors: $(11238.78 + 110.45 \times 59 + 8833.21)$; No. of Visitors = 26588



8. “Based on your analyses above, are there any modifications that you would suggest to arrive at an improved regression model? Compare and contrast the performance of your model from Case Study Report #1 and your regression model above. Based on your analysis, which model would you utilise for your specific business case?”

Answer:8: From the output result of regression analysis we can see that the P-value is not significant hence we can use some modifications to improve our regression model which are firstly we need to add more dummies. Another way is to be changing the data to quarterly basis and add more variables.

Following table shows the comparison of WES model with Regression:

	MAE	MSE	MAPE
WES(Additive)	1896	3043.5	10.18
WES(Multiplicative)	1951	3187	10.90
Regression	3676.9	5999.10	20.5

Form the above results we can see that errors from regression model are higher so, we can suggest that WES is better to perform our business case.

9. “Research question (refer to the material on iLearn): In the context of judgmental forecasting, what are *subjective assessment methods*? Critically evaluate how subjective assessment methods may assist in addressing your business problem.”

Subjective assessment is the methods use relevant (subjective) information and then subjectively assess and weight the information to produce forecasts. (Karunaratne, P 2019)

Subjective methods are common in typical business forecasting situations. They are typically used in place of quantitative forecast methods (possibly in conjunction) Managers or relevant forecasters make subjective assessments of relevant information and determine necessary forecasts.

(Karunaratne, P 2019)

Following includes Subjective Methods;

1. Sales Force Composite Forecasting
2. Jury of Executive Opinion
3. Subjective Probability assessments

The Subjective methods would assist in forecasting as we cannot totally rely on numerical data because their maybe rapid changes in future which cannot be predicted by past numerical data as we have witnessed in the data that there is an uncertain increase in over few years’ time in number of tourist.

10. “Research question (refer to the material on iLearn): In the context of judgmental forecasting, what is *scenario analysis*? Critically evaluate how scenario analysis may assist in addressing your business problem.”

Scenarios can be described as “stories of possible futures that the organisation might encounter”

They are graphic and dynamic revealing the flow of an evolving future. They are holistic, combining social, technological, economic, environmental and political trends, events and other data. They focus on the ‘branching points’ of the future, the potential contingencies and discontinuities.

Scenarios allow us to better prepare for uncertainty. (Karunaratne, P 2019).

So, scenario analysis is basically an in-depth analysis of all the scenarios an organisation might face.

Scenario analysis can be of great importance as it can help us analyse all the different scenarios we might face and doesn’t just rely on what happened in the past it rather focusses on what could happen in the future. And specifically, to our problem for example there might be an increase in tax by the government which may increase the fares and could affect the number of visitors. Or maybe other economic, social, political factors can affect the number of visitors and this cannot be forecasted just by using the numerical data.

References

Karunaratne, P 2019, Seminar 11: Judgmental Forecasting, lecture notes, Business Forecasting BBA315, Macquarie University, Delivered 24 May 2019

Appendices

Regression Lines data:

Time	march	June	Sept	December
1	11751.4744	23345.1282	10263.7821	20182.4359
2	11861.9231	23455.5769	10374.2308	20292.8846
3	11972.3718	23566.0256	10484.6795	20403.3333
4	12082.8205	23676.4744	10595.1282	20513.7821
5	12193.2692	23786.9231	10705.5769	20624.2308
6	12303.7179	23897.3718	10816.0256	20734.6795
7	12414.1667	24007.8205	10926.4744	20845.1282
8	12524.6154	24118.2692	11036.9231	20955.5769
9	12635.0641	24228.7179	11147.3718	21066.0256
10	12745.5128	24339.1667	11257.8205	21176.4744