Panel Data Regression to Determine the Factors that Affect the Number of Wisatawan Nusantara (Wisnus) Trips in Indonesia

Jasons Januard Bongtari[a], Kania Agatha[a], Delvina Wongsono[a]

[a]*Computer Science and Statistics Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia 11480*

Abstract

The number of Wisatawan Nusantara (Wisnus) trips in Indonesia varies by province and is increasing every year. Each province in Indonesia has a variety of interesting tourist destinations, ranging from nature tourism, cultural tourism, and other types of tourism. The purpose of this paper is to determine if the number of tourist destinations affects the number of Wisatawan Nusantara (Wisnus) trips in Indonesia. Using the panel data regression method, the author use the data on the number of Wisatawan Nusantara (Wisnus) trips in each province in Indonesia in between 2018 and 2019 as well as the data on the number of various types of tourist destinations in 2018 and 2019 in all the provinces in Indonesia to conclude that there are several factors which affects the number of Wisatawan Nusantara (Wisnus) trips in Indonesia.

Keywords: Wisatawan Nusantara; Tourist Destinations; Tourism; Panel Data Regression

1. Introduction

Indonesia is a country that has a lot of tourist sites to visit, where they are categorized into two types of sites, such as natural and artificial sites. Indonesia is well-known with its thriving tourism industry, therefore, Indonesia should be able to maximize its potential for the welfare of society. In addition, tourism is an industry that is more eco-friendly when compared to other industries.

There are multiple tourist destinations, such as educational tours, attractions, and natural tours. Aside from shopping at the mall, there are other accessible tourist destinations that can provide alternative options for entertainment in Indonesia. Furthermore, the existence of these domestic tourists can not be underestimated because it can help the country's economy.

Wisatawan Nusantara (Wisnus) are Indonesian residents who travel within Indonesia's territory, for a period of less than 6 months, and not for a working purpose (earning wages/salaries). Wisnus do not generate foreign exchange for the country directly, but winus are able to revive other sectors such as transportation, hotels, creative industries, and so on.

According to data from the BPS (Badan Pusat Statistik), the number of Wisnus trips in 2018 grew 12.37% to 303.4 million times compared to the previous year. Later on, the number of Wisnus trips in 2019 increased to 722.1 million. However, in 2020, the number decreased by 28,2% from the previous year, where it only reached 518.59 million.

There are several places that became a destination for domestic tourists and those places are split into 6 groups. Those would be the Natural, Cultural, Man-made tourist attraction, Amusement and Recreation Parks, Tourism Areas, Tirta Tourism. Natural tourist attractions are seen for its natural resources that have a potential and have an appeal for the visitors, both in their natural state and post-cultivation state. Cultural tourist attraction is an attraction that takes a form of creativity, and the sense of humane initiative as cultural beings. A man-made tourist attraction is an attraction that is made intentionally by human hands. An Amusement Park is an attraction that offers entertainment in the form of games and rides. In addition, the shops, stalls, restaurants, and other additional facilities also plays a part in the amusement park. A tourism area is an area that is built with a certain length and/or distance to provide and meet the needs of the tourists and its tourism services. Tirta Tourism is a type of tourist attraction that is related to water sports and/or activities that are done on beaches, lakes, rivers, bays.

The reason why we choose to use this data set is because of the lack of time we need to collect new data. We decided to use the existing dataset entitled as Jumlah Wisatawan Nusantara that was collected by a government organization called Badan Pusat Statistik (BPS). Seeing as the data was collected by the government, the accuracy of the dataset can be trusted. Moreover, this dataset is still considered as recent data.

Therefore, for the sake of this research, we use the “panel data regression” method. Panel data regression analysis is a combination of cross-section data and time series data, where the same cross-section unit is measured at different times. In other words, panel data is a data from the same variables that were observed over a certain period of time. The advantage of this method is, the panel data that can provide a larger number of data, increasing the degree of freedom, and reducing the collinearity between explanatory variables, in order to obtain the efficient econometric estimates.

The goal of this paper is to determine the factors that affected the Number of Wisatawan Nusantara’s (Wisnus) Trips using the “panel data regression” method. This method includes testing the variable affects from the attractiveness of the natural, cultural, and man-made attraction, as well as the amusement and recreation park, main tourism sites, and the tirta tourism that are available within the 34 province in Indonesia. We have examined the data that is available in BPS (Badan Pusat Statistik) regarding the number of domestic tourist trips. The purpose of our research is to find Indonesians' welfare rate. In addition, the welfare rate can be seen through the prediction results regarding the number of Wisnus trips in Indonesia.

2. Methods

2.1 Panel Data Regression

Panel data regression analysis is a regression analysis with panel data structure. Generally, the parameter estimation in regression analysis with cross-section data is done using the least square estimation method or, in other words, the Ordinary Least Square (OLS). Panel Data Regression is a combination of the cross-section data and the time series data, where the same cross-section unit is measured at different times. In other words, panel data is data from the same variables who are observed over a certain period of time. If we have time periods [T] (t = 1,2,...,T) and number of individuals [N] (i=1,2,...,N), then with panel data, we will have a total of N*T units of observation. If the number of time units is the same for each individual, then the data is called a balanced panel. Otherwise, if the number of units of time is different for each individual, it will be called an unbalanced panel. As for other types of data, such as series data for example, one or more variables will be observed in one unit of observation within a certain period of time. Meanwhile, cross-sectional data are observations from several units of observation at one point in time.

There are 2 kinds of panel data regression equations which are One Way Model and Two Way Model.

A One Way Model only considers individual effects (α_i) in the model. The equations are shown in the following:

$$\boldsymbol{y_{it} = \alpha + \alpha_i + X'_{it}\beta + \varepsilon_{it}}$$

Where:
α = constant number
β = p*1 sized vector, which is the parameter of the estimation result
X_{it} = it-th observation of the independent variable P
α_i = different individual effects for each i-th individual
ε_{it} = regression error (the same as in the classical regression model)

A Two Way Model is a model that considers the effect of time or includes a time variable. The equations are shown in the following:

$$\boldsymbol{y_{it} = \alpha + \alpha_i + \delta_i + X'_{it}\beta + \varepsilon_{it}}$$

The equation above shows where there is an additional time effect denoted by deltha which can be seen as fixed or random times between the years.

There are three approaches in calculating panel data regression models, which are:

1. Common Effect model

    This is the simplest data model approach because it only combines time series and cross-section data. This model does not pay attention to the dimensions of time and individuals, so it is assumed that the behavior of the data is the same in various periods of time. Estimation for this model can be done using the OLS (Ordinary Least Square) method.

2. Fixed Effect Model

    This model assumes that differences between individuals can be accommodated from the differences it intercepts. To estimate the panel data of the Fixed Effects model, we are using a dummy variable technique to capture the differences in intercepting between companies, differences in intercepting can occur due to the differences in work culture, managerial, and incentives. However, the slope is the same between companies. This estimation model is often called the Least Squares Dummy Variable (LSDV) technique.

3. Random Effect Model

    This model will estimate the panel data, where the disturbance variables may be interrelated over time and between individuals. In the Random Effect model, the differences in intercepting are accommodated by the error terms of each company. The advantage of using the Random Effect model is that, it eliminates heteroscedasticity. This model is also called the Error Component Model (ECM) or the Generalized Least Square (GLS) technique.

    When determining the panel data regression model, there are several tests that can be done:

    1. Chow test

        Chow test is a test to determine whether the fixed effect method is better to use than the common effect method.
        H0: Use Common Effect
        H1: Use Fixed Effect

    2. Hausman test

        Hausman test is a test to determine whether the random effect method is a better method to use when compared to the fixed effect method.
        H0: Use Random Effect
        H1: Use Fixed Effect

    3. Lagrange Multiplier / Breusch Pagan test

        Breusch Pagan test is a test to find out whether there are individual and/or time effects in the data panel.

2.2 Weighted Least Squares Regression (WLS)

Weighted Least Squares in this research will be used to handle heteroscedastic errors. Weighted Least Squares is a regression method where parameter estimates are obtained through minimizing a weighted sum of squares of residuals, where the weights are inversely proportional to the variance of the errors. The WLS estimates for regression models are obtained by minimizing

$$\sum_{i=1}^{n} \omega_i (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip})^2$$

where [ω_i] are weights inversely proportional to the variances of the residuals, [x_1], [x_2], …, [x_p] are independent variables, [y_i] are dependent variables, and [β_0], [β_1], …, [β_p] are the parameters.

2.3 Residual Assumptions

It is necessary to test the residual assumptions to ensure that the regression model that was obtained is the best model in terms of estimation accuracy, unbiased, and consistency. The errors [e1], [e2], …, [en] are assumed to have an independent and an identical normal random distributions with zero mean and a common variance, for each of the said errors. Residual assumptions are fulfilled if:

1. The errors have a normal distribution or normality assumption
2. The errors have mean zero
3. The errors have the same variance or homoscedasticity assumption
4. The errors are independent or auto-correlation assumption

However, the normality assumption is not a part of the BLUE (Best Linear Unbiased Estimator) requirement, therefore it can be concluded that in panel data regression, not all residual assumptions are used, only multicollinearity and heteroscedasticity are needed.

2.4 Variance Inflation Factor (VIF)

Multicollinearity exists when there is a linear relationship, or correlation, between one or more of the independent variables. The multicollinearity occurs when independent variables in a regression model are correlated. Multicollinearity can cause problems when fitting the model and interpreting the results because all the inputs are influencing each other. Therefore, they are not actually independent, and it is difficult to test how much the combination of the independent variables affects the dependent variable, or outcome, within the regression model.

VIF test can determine if there is any multicollinearity between the variables in the dataset. VIF value above 10 indicates that the associated independent variable is highly collinear with the other variables in the model. By using VIF, it helps to identify the severity of any multicollinearity issues so that the model can be adjusted. VIF measures how much the behavior (variance) of an independent variable is influenced, or inflated, by its interaction/correlation with the other independent variables. VIF allows a quick measure of how much a variable is contributing to the standard error in the regression. After these variables are identified, several approaches can be used to eliminate or combine collinear variables, resolving the multicollinearity issue.

2.5 Backward Elimination

Backward elimination method is a computer-based iterative variable-selection procedure. It begins with a model containing all the independent variables of interest. Then, at each step, the variable with the smallest F-statistic is deleted (if the F is not higher than the chosen cutoff level).

2.6 Dataset

In this paper, the dataset is obtained from BPS with a total of 68 cases, including 34 provinces and 1 country. The dataset we used is a combination of 3 datasets, which are the Number of Wisatawan Nusantara dataset in year 2018 and year 2019, the Number Of Commercial Tourist Attraction Objects by Province and Kind of Attraction dataset in 2018, and the Number Of Commercial Tourist Attraction Objects by Province and Kind of Attraction dataset in 2019. We combine those 3 datasets into one data panel dataset so that we can use data panel regression methods. We also altered and removed some columns and rows that are not useful for the research.

The data contains the Jumlah_Wisnus as dependent variable and Daya_Tarik Wisata_Alam, Daya_Tarik_Wisata_Budaya, Daya_Tarik_Wisata_Buatan, Taman_Hiburan_dan_Rekreasi, Kawasan_Pariwisata, Wisata_Tirta as independent variable. The variable “Provinsi” contains the name of the province, “Tahun” contains the year of the case, “Jumlah_Wisnus” is the number of Wisnus in the certain year and province, “Daya_Tarik Wisata_Alam” is the number of Natural Tourist Attractions in the certain year and province, “Daya_Tarik_Wisata_Budaya” is the number of Cultural Tourist Attractions in the certain year and province, “Daya_Tarik_Wisata_Buatan” is the number of Artificial tourist Attractions in the certain year and province, “Taman_Hiburan_dan_Rekreasi” is the number of Amusement and Recreation Parks in the certain year and province, “Kawasan_Pariwisata” is the number of Tourism Areas in the certain year and province, and “Wisata_Tirta” is the number of Water Tourism in the certain year and province.

3. Result and Discussion

We will use the [R] software to determine the panel data regression model using Chow test, Hausman test, and Breusch Pagan test.

3.1 Chow Test

Table 1. F Statistic for Chow Test

F-Value	P-Value
2.2018	0.01805

Based on Table 1. Chow Test, with a significant level value of 0.05, the p-value obtained is 0.01805 which is lower than 0.05. It means reject H_0, so with a 95% confidence level, we believe that the fixed effect method is better than the common effect method, in terms of usage.

3.2 Hausman test

Table 2. Chi-Squared Statistic for Hausman Test

F-Value	P-Value
6.5995	0.3595

Based on Table 2. Hausman Test, with a significant level value of 0.05, the p-value obtained is 0.3595 which is greater than 0.05. It means fail to reject H_0, so with a 95% confidence level, we believe that the Random effect method is better than the fixed effect method, in terms of usage.

3.3 Lagrange Multiplier test

3.3.1 Two Ways Effect

Table 3. Chi-Squared Statistic for Lagrange Multiplier Test Two Ways Effect

Chi-Squared	P-Value
5.0459	0.08022

Based on Table 3. Breusch Pagan Test Two Ways Effect, with a significant level value of 0.05. The p-value obtained is 0.08022 which is greater than 0.05. Because the p-value is greater than alpha, it can be concluded that there is no Two Ways Effect in the model.

3.3.2 Individual / Cross-Section Effect

Table 4. Chi-Squared Statistic for Lagrange Multiplier Test Individual Effect

Chi-Squared	P-Value
4.3958	0.03603

Based on Table 4. Breusch Pagan Test Cross-Section Effect, with a significant level value of 0.05. The p-value obtained is 0.03603 which is less than 0.05. Because the p-value is less than alpha, it can be concluded that there is an Individual Effect in the model.

3.3.3 Time Effect

Table 5. Chi-Squared Statistic for Lagrange Multiplier Test Time Effect

Chi-Squared	P-Value
0.65011	0.4201

Based on Table 5. Breusch Pagan Test Time Effect, with a significant level value of 0.05. The p-value obtained is 0.4201 which is greater than 0.05. Because the p-value is greater than alpha, it can be concluded that there is no Time Effect in the model.

From the Chow Test, Hausman Test, and Lagrange Multiplier Test, it can be concluded that the model formed is a random effect model which has an Individual effect.

3.4 *Variable Selection*

We will use the R software to calculate the results of further calculations. Table 6 shows the p-value of each independent variable in the t-test for the model1. Let, our significance value, be 0.05. Thus, the independent variable with P-Value greater than α is eliminated, which is "Daya_Tarik_Wisata_Budaya".

Table 6. P-values of Each Regressor Variable for model1

Regressor Variable	P-Value
Daya_Tarik_Wisata_Alam	0.16934
Daya_Tarik_Wisata_Budaya	0.94385
Daya_Tarik_Wisata_Buatan	< 2e-16
Taman_Hiburan_dan_Rekreasi	0.92327
Kawasan_Pariwisata	0.04808
Wisata_Tirta	0.01656

Table 7 states the p-values of each predictor variable in the t-test from model2 after we omitted "Daya_Tarik_Wisata_Budaya" from our equation. After that, it can be seen that the P-Value of "Taman_Hiburan_dan_Rekreasi" is greater than α.

Table 7. P-values of Each Regressor Variable for model2

Regressor Variable	P-Value
Daya_Tarik_Wisata_Alam	0.12337
Daya_Tarik_Wisata_Buatan	< 2e-16
Taman_Hiburan_dan_Rekreasi	0.91237
Kawasan_Pariwisata	0.04328
Wisata_Tirta	0.01551

Table 8 states the p-values of each predictor variable in the t-test from model3 after we omitted "Taman_Hiburan_dan_Rekreasi" from our equation. Because none of the predictor variables has a p-value greater than α, we halt the process and account for these two variables in our equation.

Table 8. P-values of Each Regressor Variable for model3

Regressor Variable	P-Value
Daya_Tarik_Wisata_Alam	0.01690
Daya_Tarik_Wisata_Buatan	< 2e-16
Kawasan_Pariwisata	0.04424
Wisata_Tirta	0.01597

After getting the final model, all variables are significant, which means we get 3 models. Then a diagnostic test will be carried out for each model. After the diagnostic test has been carried out, it is continued to select the best model with an additional criteria, namely adjusted R-squared and residual sum of squares (Tabel 9).

Tabel 9. Adj. R-squared, Residual Sum of Squared, and P-value of Autocorrelation test and Homoscedastic test from all model

Model	Adj. R-squared	Residual Sum of Squares	Autocorelation (P-Value)	Homoscedastic (P-Value)
1	0.8009	3,10E+19	0.03558	1,12E-03
2	0.80429	3,11E+19	0.05788	0.01031
3	0.80621	3,05E+19	0.06209	0.006074

First seen on serial correlation and heteroscedasticity, then the candidate model is between model 2 and model 3. In the selection of this model, we should look for the model with the largest Adjusted R-squared, the smallest Residual Sum of Squares, no serial correlation, and homoscedasticity. Then seen from the smallest Residual Sum of Squares and from the largest Adjusted R-squared, therefore, model 3 is chosen.

Based on our backward elimination method, our final independent variables are “Daya_Tarik_Wisata_Alam”, "Daya_Tarik_Wisata_Buatan", "Kawasan_Pariwisata" and “Wisata_Tirta”. This means that we can estimate our dependent variable, “Jumlah_Wisnus”.

Table 10. Coefficient from the model

	Coefficient
Intercept	2951551
Daya_Tarik_Wisata_Alam	193235
Daya_Tarik_Wisata_Buatan	425605
Kawasan_Pariwisata	900059
Wisata_Tirta	-94653

Jumlah_Wisnus = 2951551 + 193235(Daya_Tarik_Wisata_Alam) + 425605(Daya_Tarik_Wisata_Buatan) + 900059(Kawasan_Pariwisata) - 94653(Wisata_Tirta)

We must test this new regression equation for its residual assumption using tests such as the Durbin-Watson test and the Breusch-Pagan test, as we have explained in the Methods section. Table 9 shows the p-value results of our regression equation in each test. For each test, we want the p-value to be greater than our α, which is 0.05.

Table 11. P-Value Of Regression Equation in Residual Assumption Tests.

Test	P-Value
Durbin-Watson	0.06209
Breusch-Pagan	0.006074

Table 11 shows that our model fulfills the Durbin-Watson residual assumption. However, it does not fulfill the residual assumptions for the Breusch-Pagan test. This means that:

1. The residuals of our model are not autocorrelated
2. The variance of residuals from our model is not constant (heteroscedastic)

Regarding this issue, we need to develop an optimal regression model that can fulfill the residual assumptions for the Breusch-Pagan test.

Then, from the new model (model3), we would like to test if there is any multicollinearity from our independent variable with Variance Inflation Factor (VIF) Test using R Software.

Tabel 12. Value of Variance Inflation Factor Test

Predictor Variable	VIF
Daya_Tarik_Wisata_Alam	3.606917
Daya_Tarik_Wisata_Buatan	1.652414
Kawasan_Pariwisata	2.585666
Wisata_Tirta	2.172530

Based on Table 12. There are no VIF values obtained greater than 10. It means there is no multicollinearity between all the independent variables.

3.5 Weighted Least Squared (WLS)

To fulfill the heteroscedastic assumption, we develop a regression model using Weighted Least Squares method using R Software to account for heteroscedasticity assumption.

Table 13. Coefficient from the new developed model using Weighted Least Squared method

	Coefficient
Intercept	2614431
Daya_Tarik_Wisata_Alam	161506
Daya_Tarik_Wisata_Buatan	440032
Kawasan_Pariwisata	1077000
Wisata_Tirta	-108825

Based on Table 13. We can estimate our dependent variable, “Jumlah_Wisnus”.
Jumlah_Wisnus = 2614431 + 161506(Daya_Tarik_Wisata_Alam) + 440032(Daya_Tarik_Wisata_Buatan) + 1077000(Kawasan_Pariwisata) - 108825(Wisata_Tirta)

Table 14. P-Value Of Regression Equation in Residual Assumption Tests for new model.

Test	P-Value
Durbin-Watson	0.05892
Breusch-Pagan	1

Table 14. Shows that our model fulfills the Durbin-Watson residual assumption. However, it does not fulfill the residual assumptions for the Breusch-Pagan test. This means that:

1. The residuals of our model are not autocorrelated
2. The variance of residuals from our model is constant (homoscedastic)

Regarding this issue, we have fulfilled all the residual assumptions for our new regression model.

Table 15. Cross Section Value for Each Provinces

Aceh	6504543.17
Gorontalo	-1743162.21
Kalimantan Selatan	1325303.21
Lampung	3052323.04
Papua Barat	-3211567.6
Sulawesi Utara	72498.09
Bali	-3117701.86
Jambi	733656.91
Kalimantan Tengah	-110112.35
Maluku	-591760.53
Riau	3799817.14

Sumatera Barat	-1666347.12
Banten	-2876260.22
Jawa Barat	8339058.91
Kalimantan Timur	2003714.73
Maluku Utara	-1323498.75
Sulawesi Barat	-2918352.81
Sumatera Selatan	4162399.99
Bengkulu	-968367.99
Jawa Tengah	29200.23
Kalimantan Utara	-1723968.36
Nusa Tenggara Barat	-2752259.2
Sulawesi Selatan	3307077.59
Sumatera Utara	4714216.85
DI Yogyakarta	-15195066.8
Jawa Timur	77684.29
Kepulauan Bangka Belitung	-879237.8
Nusa Tenggara Timur	1223943.94
Sulawesi Tengah	40969.04
DKI Jakarta	11803947.93
Kalimantan Barat	1059149.21
Kepulauan Riau	-1929119.63
Papua	-891349.48
Sulawesi Tenggara	103502.82

Table 16. F-statistics test from the new developed model using Weighted Least Squares method

P-value	**< 2.22e-16**

Based on Table 16. With a significant level value of 0.05. The p-value obtained is < 2.22e-16, which is less than 0.05. Because the p-value is less than alpha, it can be concluded that there is a linear relation between the independent variable and the dependent variable.

Table 17. P-value from the new developed model using Weighted Least Squared method

	P-value
Intercept	0.023929
Daya_Tarik_Wisata_Alam	0.065310
Daya_Tarik_Wisata_Buatan	1,13E-10
Kawasan_Pariwisata	0.005206
Wisata_Tirta	0.004358

Based on table 17, it is found that the p-value on the variables Daya_Tarik_Wisata_Buatan, Kawasan_Pariwisata, and Wisata_Tirta is less than alpha 0.05, so it can be concluded that Daya_Tarik_Wisata_Buatan, Kawasan_Pariwisata, and Wisata_Tirta has a significant effect on Jumlah_Wisnus.

Table 18. Adj. R-Squared from the new developed model using Weighted Least Squared method

Adjusted R-Squared	**0.80364**

Based on table 18, it can be seen that the adjusted R-Squared value is 0.80364. This means that the model is a good fit and the accuracy is 80.36%.

4. Conclusion

The final model formed after the backward elimination method has four independent variables, which are Daya_Tarik_Wisata_Alam, Daya_Tarik_Wisata_Buatan, Kawasan_Pariwisata, and Wisata_Tirta. Two independent variables, which are Wisata_Tirta and Taman_Hiburan_dan_Rekreasi, are not used because their p-value is greater than the significance level, that would be 5%. There is a possibility that these two variables have no effect or no correlation with the dependent variable. Since our goal is to determine the factors affecting the dependent variable, we have decided not to use these two independent variables.

From the model, after the residual assumption test is done, we find that the residual variance is constant. To overcome this violation, we develop our model using the weighted least square method. From the new developed model, there is no residual assumption violation found, and the model is a good fit with 80% accuracy. This means that the independent variables which are Daya_Tarik_Wisata_Alam, Daya_Tarik_Wisata_Buatan, Kawasan_Pariwisata, and Wisata_Tirta are able to explain the dependent variable from the Jumlah_Wisnus of 80%, while the remaining 20% is explained by the other factors that are not mentioned in the model.

5. References

G. Hutcheson and N. Sofroniou. 2015. The multivariate social scientist. [Place of publication not identified]: SAGE Publications Ltd.

Chatterjee, S. and Hadi, A. S. 2006. Regression Analysis by Example. Canada: John Wiley & Sons, Inc.

Iqbal, Muhammad. 2015. “Regresi Data Panel (2) ‘Tahap Analisis’”, https://dosen.perbanas.id/regresi-data-panel-2-tahap-analisis/ , diakses pada 12 Juni 2022 pukul 22.30.

Mubarok, Muhammad Ilham. 2018. “ANALISIS DATA PANEL [IN R]”, https://muhammadilhammubarok.wordpress.com/2018/09/02/analisis-data-panel-in-r/ , diakses pada 13 Juni 2022 pukul 11.35.

The Investopedia Team. 2021. “Variance Inflation Factor (VIF)”. https://www.investopedia.com/terms/v/variance-inflation-factor.asp , diakses pada 13 Juni 2022 pujul 13.50.

Clarissa, Annebel Diestya. 2021. “Analisis Data Panel”. https://rpubs.com/annebel/analisisdatapanel , diakses pada 14 Juni 2022 pukul 15.45.

Putri, Dwi Septiandini. 2020. “Penerapan Metode Regresi Data Panel dengan R”, https://medium.com/@17611055/penerapan-metode-regresi-data-panel-dengan-r-ad31d220b87a , diakses pada 14 Juni 2022 pukul 18.00.