The Effects of Local Demographics on Black Lives Matter Protester Crowd Size

Makayla Bibby (mdb9vku)

Kathia Crawford (klc6myc)

Hannan Mumtaz (hm9tac)

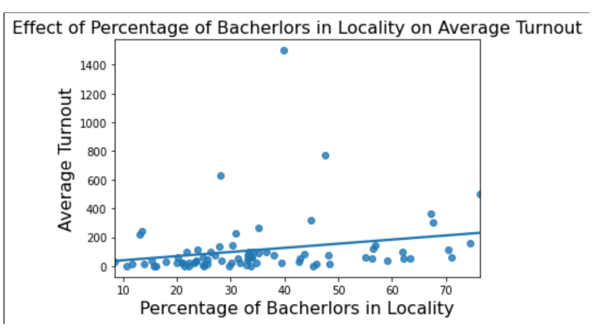
Tyler Vo (tkv9zd)

University of Virginia

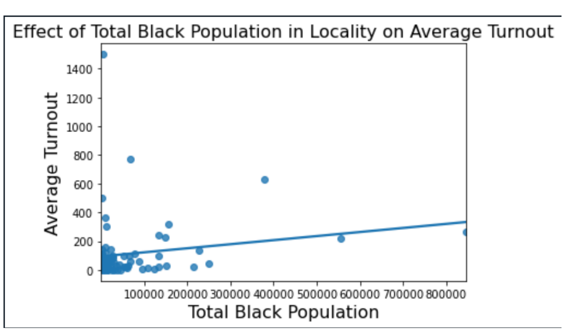
STAT 2120

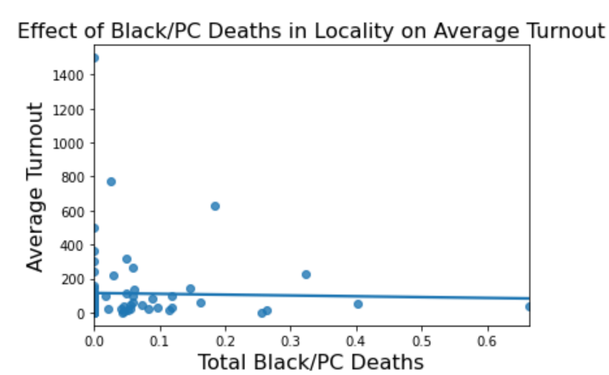
On our honor, we did not give nor receive aid on this assignment.

Introduction — We are attempting to predict protester crowd size for Black Lives Matter (BLM) protests. The three explanatory variables we are utilizing are: the number of Black people in the total population of a given city, the number police related Black and People of Color (POC) deaths in a given city, and the percentage of people who possess Bachelor's degrees in a given city. We chose the total Black population because the larger this demographic population is the larger the personal impact by black death is on each individual, influencing BLM protest turnout. We chose the number of police-related Black and POC deaths in a given city for the same reason. Bachelor's degrees were chosen because cities with a greater proportion of people that have higher education are more likely to be educated about racial equity and crime reform, which would lead to more protest turnout. Our response variable is average turnout per protest, as we are trying to predict the number of people that will attend a protest. Protest turnout is a quantitative variable that we will be predicting based on other quantitative data. We are predicting a positive relationship between all three variables with our response variable, which we will further explore in this study.

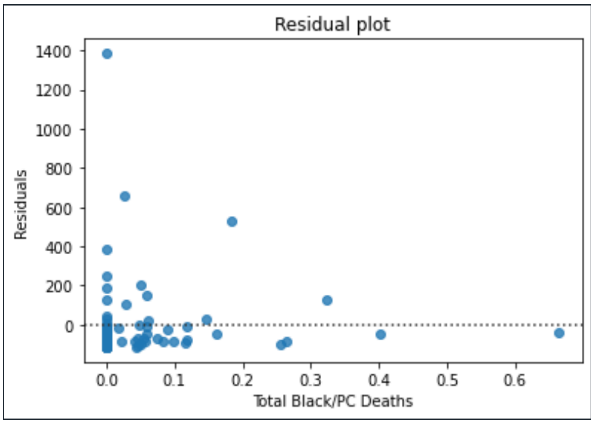


Analysis With Simple Linear Regression (SLR) — We used SLR to analyze the relationship between average protester turnout and all three explanatory variables. Graphical summaries for our three explanatory variables are to the left and below.

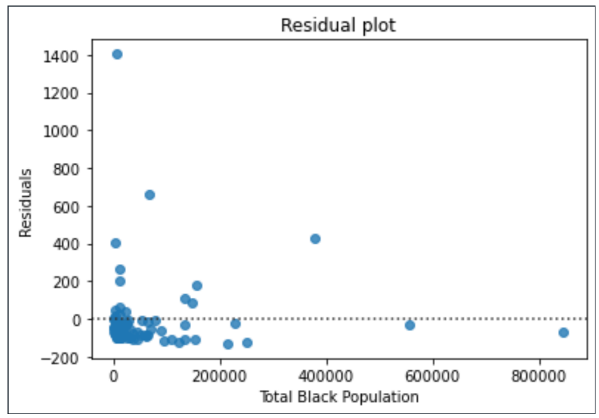




Residual Graph for The Effect of Total Black Population in Locality on Average Turnout:

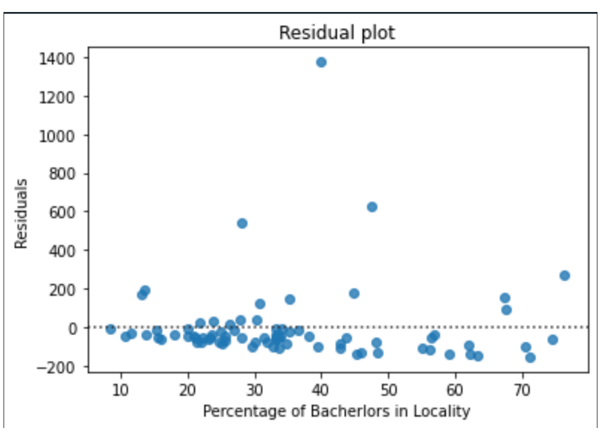


SLR models are robust against deviations from the Normal distribution, so the normality assumption is satisfied. Linearity, independence, and constant variance are not satisfied because although the graph is scattered randomly and around zero, it does not have uniform variation. Thus, a linear regression plot is not a reliable way to analyze this set of data.



Residual Graph for the Effect of Black/PC Deaths in Locality on Average Turnout: SLR models are robust against deviations from the Normal distribution, so the normality assumption is satisfied. Linearity, independence, and constant variance are not satisfied because although the graph is scattered randomly and around zero, it does not have uniform variation. Thus, a linear regression plot is not a reliable way to analyze this set of data.

Residual Graph for the Effect of Percentage of Bachelors in Locality on Average Turnout:



SLR models are robust against deviations from the Normal distribution, so the normality assumption for this residual is satisfied. Linearity, independence, and constant variance are not satisfied. Although the graph is scattered randomly and around zero and it does not have uniform variation. Thus, a linear regression plot may not be a reliable way to analyze this set of data .

The SLR Model for each of our three variables is: $y_\iota = \beta_0 + \beta_1 X + \epsilon_\iota$ where y_ι is the predicted value, β_0 is the population intercept, β_1 is the population slope and ϵ_ι are the deviations.

We chose the Percentage of Bachelor's Degrees as the "best" option to perform a test on because the r-squared value for Percentage of Bachelor's is the largest (0.052) compared to the Total Black Population (0.029) and Total Black/PC Deaths (0.001). The square root of r-squared values allows us to find the correlation. The strongest correlation was between Percentage Bachelors (where r=0.228) and Average Turnout. The linear regression model for the this is: $y_\iota = \beta_0 + \beta_1 X + \epsilon_\iota$ where y_ι is the predicted value, β_0 is the population intercept, β_1 is the population slope and ϵ_ι are the deviations. The least squares regression equation for this is: y = 11.17 + 2.87x. This means that the predicted average turnout will increase by 2.87 members when the Percentage of the population with Bachelor's Degrees increases by one unit. When it is zero, the predicted value is 11.17 (the intercept).

The following is our t-test for slope and conclusion for the slope in context: $H_0: \beta_1 = 0$, *There is no relationship between the percentage of people with Bachelor Degrees and average turnout.* $H_a: \beta_1 \neq 0$, *There is a relationship between the percentage of people with Bachelor Degrees and average turnout*

The following table gives us our test-statistic and p-value for this variable:

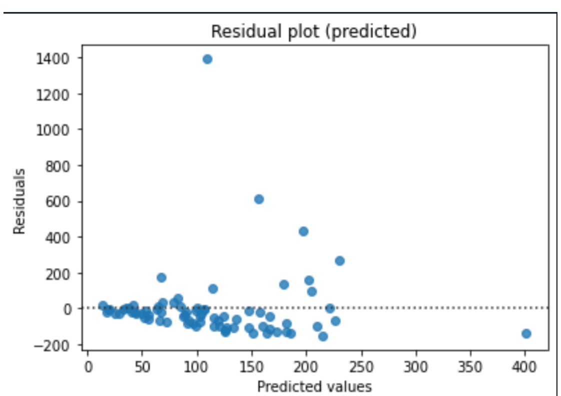
	coef	std err	t	P>|t|	[0.025	0.975]
const	11.1698	55.082	0.203	0.840	-98.559	120.899
PercentBachelor.s	2.8696	1.413	2.031	0.046	0.055	5.684

The resulting test-statistic is 2.031. At a significance level $\alpha = 0.05$ (**note: this significance level will be used for all subsequent tests**) we reject our null hypothesis given our p-value of 0.046. There is sufficient evidence to suggest that there is a linear relationship between the percentage of people with a Bachelor's degree in each locality and average protester turnout.

In our original dataset that included Null/Nan in the data, rows 95-99 had no values for the average protester turnout. We used the least squares regression equation from the Percent Bachelor's variable to predict these values. We used 35.1923 as our value from the Percent Bachelor's column to predict an average turnout value of 112.157, 24.52 to predict 81.535, 26.289 to predict 86.608, 48.248 to predict 149.624, and 74.5203 to predict 255.012.

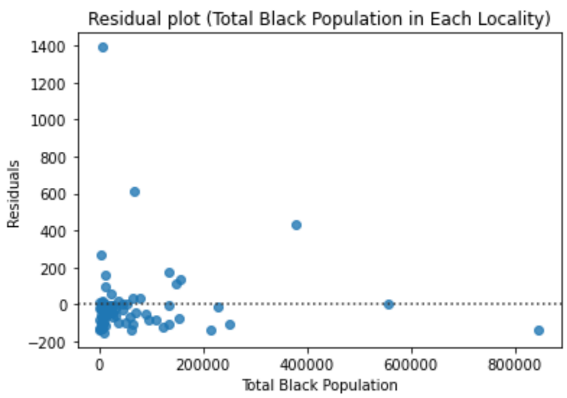
From these estimates, we can be confident in using Percent Bachelor's as a predictor of turnout because these predicted values are within the higher range of real averages for protester turnout from the dataset. Also, the prediction intervals (a predicted range of estimates for each missing value) demonstrate a significant overlap where the real value for each of these missing values is. Thus, we are also comfortable with ranking these predictions from highest to lowest protester turnout as follows: 225.012, 149.624, 112.157, 86.608, 81.535.

Analysis with Multiple Linear Regression (MLR) — Next, we used MLR to analyze the relationship between average protester turnout and all the explanatory variables to determine if adding additional variables to our regression test would be more useful in predicting protester turnout. To conduct this analysis we used the MLR model, $y_i = \beta_0 + \beta_1 x_{\iota 1} + \beta_2 x_{\iota 2} + \beta_3 x_{\iota 3} + \epsilon_\iota$, where β_0 is the population intercept, and the predicted protester turnout when all the other

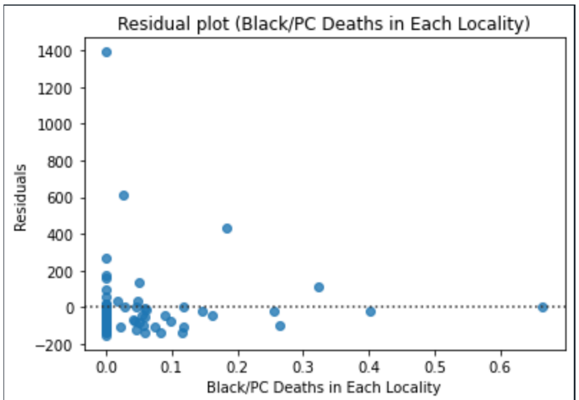


variables in this model have a value of 0. $\beta_1 x_{\iota 2}$ is each variable population slope (β_1..) multiplied by each additional variables value ($x_{\iota p}$). This model helps us determine how dependent protester turnout is on the variables. Using the MLR model, we conducted the following residual plots to assess the linear regression assumptions for each of our variables:

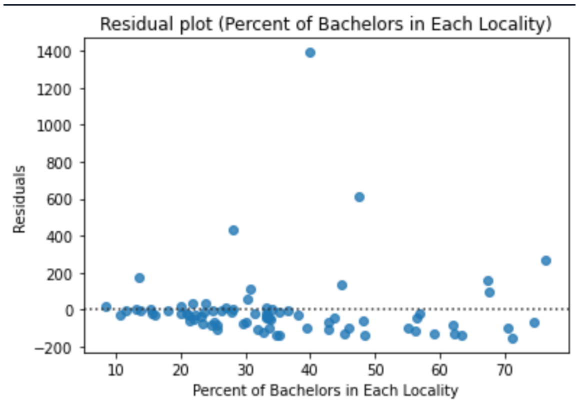
This residual plot (previous page) is a summation of all the chosen variables as predictors for protest crowd size. The MLR model is robust against deviations from the Normal distribution, so the normality assumption for this residual is satisfied. Linearity, independence, and constant variance assumptions are not satisfied. While the residuals are clustered around 0, there is no random scattering, and all the points seem to follow a negative value, with some points above 0. The variation is not uniform.



For the residual for Total Black Population (left), it is a MLR model, robust against deviations from the Normal distribution and satisfies the normality assumption. The linearity, independence, and constant variance assumptions are not satisfied. The residuals are clustered under 0 in the same location, but there is no random scattering.



For the residual for Black/PC Deaths(left), it is a MLR model, robust against deviations from the Normal distribution satisfying the normality assumption. The linearity, independence, and constant variance assumptions are not satisfied. The residuals are clustered under 0 in the same location, there is no random scattering, and it is not uniform.



For the residual for the Percentage of Bachelor's, it is a MLR model, robust against deviations from the Normal distribution, which satisfies the normality assumption. The linearity, independence, and constant variance assumptions are not satisfied. The residuals are clustered under 0 in the same location, there is no random scattering, and it is not uniform.

We conducted MLR t-tests to determine whether each explanatory variable added to the model will be useful in predicting the number of people who will attend a protest.

Total Black Population:

H_0:β_j= 0, Adding the total Black population in each locality to the model will have no effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest.H_a:β_j $\neq$ 0 , Adding the total Black population in each locality to the model will have an effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest.

The following table gives us our test-statistic and p-value for this variable (highlighted):

	coef	std err	t	P>|t|	[0.025	0.975]
const	-28.1480	60.010	-0.469	0.640	-147.748	91.452
BlackPop	0.0004	0.000	1.975	0.052	-3.35e-06	0.001

The resulting test-statistic for this hypothesis test is 1.975. We fail to reject our null hypothesis given a p-value of 0.052. We do not have sufficient evidence to suggest that the model with this added variable is useful at predicting the number of protesters.

Deaths of Black/PC:

H_0:β_j= 0, Adding the proportion of Black and POC deaths in each locality to the model will have no effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest. H_a:$\beta_j$$\neq$ 0 ,Adding the proportion of Black and POC deaths in each locality to the model will have an effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest.

	coef	std err	t	P>|t|	[0.025	0.975]
const	-28.1480	60.010	-0.469	0.640	-147.748	91.452
BlackPop	0.0004	0.000	1.975	0.052	-3.35e-06	0.001
deaths_black_pc	-53.4169	223.574	-0.239	0.812	-498.999	392.166

Once again, to conduct the t-test for slope, the table above gives us our test-statistic and p-value for this variable. The resulting test-statistic for this hypothesis test is -0.239. We fail to reject our null hypothesis given a p-value of 0.812. We do not have sufficient evidence to suggest that the model with this added variable is useful at predicting the number of protesters.

Percent Bachelor's Degrees:

H_0:$\beta_j = 0$, *Adding the Percentage of people with a Bachelor's degree in each locality to the model will have no effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest.* H_a:$\beta_j \neq 0$, *Adding the Percentage of people with a Bachelor's degree in each locality to the model will have an effect on the model's usefulness in predicting the number of protesters that will attend a BLM protest.*

	coef	std err	t	P>|t|	[0.025	0.975]
const	-28.1480	60.010	-0.469	0.640	-147.748	91.452
BlackPop	0.0004	0.000	1.975	0.052	-3.35e-06	0.001
deaths_black_pc	-53.4169	223.574	-0.239	0.812	-498.999	392.166
PercentBachelor.s	3.3755	1.424	2.371	0.020	0.538	6.213

The table (above) gives us the test-statistic and p-value for this variable. The resulting test-statistic for this hypothesis test is 2.371. We reject our null hypothesis given a p-value of 0.020. We do have sufficient evidence to suggest that the model with this variable is useful at predicting the number of protesters. This is consistent with our findings from the SLR analysis conducted in Part II. Based on our results from each t-test, two out of the three variables (total Black population and deaths of Black citizens/POC) demonstrate no linear relationship with average turnout. This indicates that these two variables are not appropriate to use in a linear regression analysis. The Percent Bachelor's variable indicated some statistical linear relationship with average turnout. Therefore, it may be an appropriate variable to use in regression analysis.

Individually, our variables may not be useful to our model. To determine if these variables will be useful together, we conducted the ANOVA F test. This helps us assess how much of the variation in the response variable is a product of our model and how much of the variation in the response variable is not caused by the model.

H_0:$\beta_1 = \beta_2 = \beta_3$ *= 0, the model is not useful at predicting protester turnout.* H_a:$\beta_j \neq 0$ *(at least one variable is useful at predicting turnout), the model is useful at predicting protester turnout.*

	SS	df	MS	F	pvalue
Regression	3.359392e+05	1.0	111979.721883	2.711404	0.051146
Residual	3.014866e+06	75.0	41299.536850	NaN	NaN
Total	3.350805e+06	76.0	NaN	NaN	NaN

We used the F-statistic to conduct this test. Using the ANOVA F model table, (above) to determine our F-statistic and p-value for this variable we received 2.711 as the F-statistic. We fail to reject our null hypothesis given a p-value of 0.051. We do not have sufficient evidence to suggest that the model is useful at predicting the number of protesters.. This is consistent with our findings from the ANOVA F model, which indicated an r-squared value of 0.100. This is almost a 0.05 increase from the r-square value of just the Percent Bachelor's variable conducted in our SLR analysis in Part II. If an r-squared value increases with the addition of variables to the MLR model, then that variable(s) provides no useful information to the model, thereby making it less useful. To support this finding, we conducted a Partial F-test, to evaluate if adding additional variables to the Percent Bachelor's variable chosen in the SLR analysis in Part II made our model less useful.

H_0*:all* β_{a_j} *= 0, the set of the extending variables in the model is not useful at predicting protester turnout.* H_a: $\beta_{a_j} \neq 0$ *(at least one extending variable is useful at predicting turnout) the set of extending variables in the model is useful in predicting protester turnout*

For the Partial F test, we coded a comparison (displayed below) using the SLR model for Percent Bachelor's and an extended variable model using MLR with two additional variables (total Black population and proportion of deaths for Black citizens and POC).

```
# Partial F test
r2_2 = model.rsquared
r2_1 = model_ext.rsquared

p = 3 #total number of pred variables
q = 2 # total number of additional variables
n = model.nobs

# Test statistic this is the F-statistic
test_stat = ( (n-p-1) / q )*( (r2_1-r2_2) / (1-r2_1))
test_stat = round(test_stat, 2)
print(test_stat)

#%%
# p-value
pval = 1 - stats.f.cdf(test_stat, q, n-p-1)
pval = round(pval, 4)
print(pval)
```

```
ProjectWo
1.95

In [66]:
Downloads
ProjectWo
0.1496
```

The resulting F-statistic for this hypothesis test was 1.95, and the p-value was 0.1496, which is larger than the significance level. As a result, we fail to reject the null hypothesis. This indicates that we do not have sufficient evidence to suggest that adding additional variables to our model is useful at predicting the number of protesters.

Conclusion — Our prediction before analyzing this data was a positive relationship between each of the explanatory variables and the response, meaning that, as each variable increased, the response variable would as well. Utilizing a SLR Model and conducting tests on all explanatory variables, we discovered that each has a relatively weak relationship with the average turnout per protest. However, the best explanatory variable, percentage of people with a Bachelor's degree in each locality, demonstrated sufficiently that there is a linear relationship between the percentage of people with a Bachelor's degree and average protester turnout. Utilizing MLR and adding the total Black population and Deaths of Black/POC as explanatory variables to the model, we found no sufficient evidence to suggest that the model with either of these added variables is useful at predicting the number of protesters that will attend. Conducting this test only for Percent

Bachelor’s, however, demonstrated that this variable might be useful to predict protest turnout *on its own.* In order to accomplish our goal of accurately estimating protester crowd size, the best option would be using the Percentage of Bachelors to predict, even if it is not incredibly strong.

Reflection — If we were to conduct this analysis again, we would have worked to address other variables that might have demonstrated a stronger linear relationship with the response. In earlier deliberations, we noticed that the Poverty Rate, both for the general population, and for the Black community, demonstrated a linear relationship. We found it incredibly interesting that the variables we suspected to be the most influential in garnering protester turnout (the Black population, proportion of deaths for Black/POC), actually had no real effect when making predictions. This was shocking, since this suggests that the magnitude of the Black population in any given locality, is not a central factor in garnering protester turnout. Also, we expected the most “best variable” (Percent Bachelor’s) to yield one of the least influential effects. It was surprising to see that this had some degree of influence as a predictor variable. However, it is probably more significant to note that the linear relationship between the percentage of people in the locality with a Bachelor’s degree and average turnout was very weak (about 0.2) and the p-value from this test was not *significantly* smaller than the significance level. Despite the linear relationship, we would not suggest that this an immediate factor in predicting protester turnout. Instead, it is likely that exogenous factors have a significant effect on whether or not people participate in protests. Finally, in conducting the models for the ANOVA F test, the coding output made a note that our model displays some multicollinearity between variables. This suggests that with one or more of these variables, the results from the model are unreliable.

Appendix A

Milestone Checklist

Milestone 1:

Sergey Levshin <sl7tcu@virginia.edu> Tue, Apr 27, 2:12 PM (2 days ago)

to me

This email serves as a confirmation that I have reviewed your work on Milestone 1. Please note that this check does not guarantee any project. If you have additional questions, please feel free to reach out to a LTA or Professor Ross.

2pm Lab, 4/27/2021

--

Milestone 2:

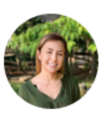
Tyler Lapointe 2:00 PM (

to me

Hi Hannan,

This is confirmation of your Milestone 2 on 4/30/2021.

Milestone 3:



Tyler Lapointe

to me

This is your confirmation of Milestone 3 on 4/30/2021.

•••

--

Tyler LaPointe

Appendix B

Data

TotalPop	PovertyRate	BlackPop	BlackPoverty	AsianPop	AsianPoverty	HispanicPop	HispanicPo	WhitePop	WhitePove	PercentBa	SqMiles	tot.protests	tot.attend	AverageTurnout	NAACPyears	MayorPartisanship	dem	rep	pop.density	collegeenrollpc	mayorrep	unarmed_deaths_pc	deaths_pc	unarmed_deaths_black_pc	deaths_black_pc	anyprotests
205048	31	152536	34.1	1999	14.3	7239	38.4	41665	19.1	23.42096	149.236	4	112	28	10	D	88787	17486	1373.984829	8.534587024	0	0	0.0975381	0	0.097538137	1
599620	18.2	300482	26.5	20585	16.1	60874	17.1	208411	7.4	54.01759	61.095	22	48983	2226.5	26	D	NA	NA	9814.551109	15.84336747	0	0	0.1334178	0	0.083386145	1
44493	8.4	1289	20.3	1005	11.8	25727	8.8	16498	7	62.11579	12.927	1	50	50	0	R	10910	11787	3441.865862	0	1	0	0	0	0	1
169310	21.2	53268	34.2	2115	12.9	26568	25	84838	11.9	34.073	34.789	1	100	100	2	D	48885	23769	4866.768231	0	0	0.059063257	0.3543795	0.059063257	0.118126513	1
112696	35.8	27084	40.5	7926	39.6	11046	46.5	63942	31.3	43.65177	60.134	3	250	83.33333333	3	D	40958	16299	1874.081219	39.93043231	0	0	0.1774686	0	0.088734294	1
818636	17.8	249884	28.8	36151	12.5	67982	20.9	443077	11.7	25.7012	746.987	4	170	42.5	13	R	186447	189137	1095.917332	4.15325004	1	0.012215441	0.1343699	0.012215441	0.073292648	1
408781	29.9	78753	44.4	3662	16.6	290886	28.8	45211	14.6	23.8293	35.924	4	450	112.5	3	R	67685	43725	11379.05022	7.338892952	1	0.07338893	0.3914076	0	0.048925953	1
245928	19.8	69483	29.8	9038	16	66017	25	98276	9.9	33.73409	101.773	5	294	58.8	0	D	61026	30485	2416.436579	0	0	0	0.4066231	0	0.162649231	1
243375	17.2	58636	30.3	8110	17.4	16774	15.2	153090	12	30.15731	61.739	1	24	24	7	D	74777	39574	3941.997765	0	0	0	0.3697997	0	0.041088855	1
53988	21	15317	31.9	1395	14.2	11240	19.7	25162	15.5	22.11825	22.766	1	0	0	0	R	10122	5297	2371.431081	0	1	0	0.3704527	0	0	1
172761	31.1	61651	41.9	6860	22.7	10490	35.1	90844	24.1	48.40661	99.99	3	50	16.66666667	4	D	62727	27949	1727.782778	34.7300606	0	0	0.1157669	0	0.115766869	1
338153	22	88793	37.5	12527	10.1	79412	28.4	153517	11.5	33.88104	113.468	4	248	62	13	D	86428	47781	2980.161808	11.08965468	0	0	0.0887172	0	0.059144825	1
99511	20.2	31332	24.2	2503	16.1	23888	28.1	40819	12.3	31.37057	55.383	1	50	50	0	D	27077	10849	1796.778795	0	0	0.100491403	0.4019656	0.100491403	0.401965612	1
107933	36	29880	37.8	4278	38.8	12139	45.3	59556	33.5	39.79699	116.357	1	30	30	9	D	NA	NA	927.6021211	30.11127273	0	0	0	0	0	1
412162	25.2	221102	35.8	13048	25	22319	32	150249	9.1	47.7806	133.094	15	1722	114.8	12	D	NA	NA	3096.773709	16.37705562	0	0.024262305	0.0970492	0	0	1
82388	6.3	12256	8.3	4748	5.7	5788	35	58391	3.2	56.26351	47.478	1	55	55	0	R	5265	17807	1735.287923	0	1	0	0	0	0	1
132491	26.1	72735	31.4	2933	28.1	6133	37.8	48743	16.7	27.59603	102.685	1	60	60	0	D	NA	NA	1290.266349	7.547682484	0	0	0.1509537	0	0	1
177552	17.6	52091	29.6	4294	21	10233	31.8	106197	9.8	39.46439	206.808	1	24	24	0	R	39642	40247	858.5354532	5.632152834	1	0.056321528	0.1689646	0.056321528	0.056321528	1
52558	13.5	16871	16.6	3667	4.2	7601	34.7	24080	5.8	51.71881	15.335	2	324	162	0	R	NA	NA	3427.323117	0	1	0	0	0	0	1
54074	33.3	27516	42.9	940	39.5	2663	44	21653	21.3	24.98125	35.48	1	20	20	0	R	NA	NA	1524.069899	27.73976403	1	0	0	0	0	1
334016	12	6153	8.2	182668	8.9	19297	14.8	54665	12	35.5753	NA	3	54	18	1	D	79891	29047	NA	14.96934279	0	0	0.0598774	0	0	1
2660712	22.7	845577	34.2	151738	20.1	777985	24.2	851266	10.6	35.1757	227.498	27	7106	263.1851852	18	D	930923	149420	11695.54018	4.942286125	0	0.018791962	0.1014766	0.011275177	0.06013428	1
68427	13.9	12236	18	5449	31.9	7032	18.1	41815	9.7	67.52879	7.795	1	300	300	13	D	31046	4158	8778.319436	65.7635144	0	0	0.1461411	0	0	1
51594	8.8	10706	15.9	2727	13.2	3707	15	32721	5.7	67.26267	4.7	2	726	363	0	D	24580	4271	10977.44681	0	0	0	0	0	0	1
147382	25.4	31206	45	4311	21.3	25513	34.7	82640	15.1	21.30459	61.028	1	0	0	6	I	37762	22438	2414.989841	1.696272272	0	0	0	0	0	1
36199	22	6926	37	579	45.9	4128	23.3	23497	16.9	22.26122	16.817	1	24	24	0	D	11550	6473	2152.524231	0	0	0	0	0	0	1
113589	18.6	22134	37.7	2563	11.6	2709	26.9	83389	13	35.15672	59.392	2	175	87.5	0	D	32964	24594	1912.530307	0	0	0	0	0	0	1
34454	36	6156	42.9	5937	56.8	1462	38.9	19971	27.8	54.97696	11.617	1	60	60	4	D	12089	3447	2965.825945	87.07261856	0	0	0	0	0	1
50301	6.8	1870	22.8	3088	18.1	2398	6.5	42135	5.5	62.00581	11.252	1	100	100	2	NP	14459	14899	4470.405261	4.970080118	0	0	0	0	0	1
67839	39	2741	47	6177	55.8	2833	43.6	53684	36.4	56.90877	23.168	1	140	140	0	D	27170	9470	2928.133633	44.22234998	0	0	0.1474078	0	0.147407833	1
816687	21.4	227483	29.2	19738	24	79037	40.3	469578	14.2	27.89766	361.528	2	270	135	18	R	220161	117892	2258.986856	7.407978822	1	0.024489186	0.1469351	0.024489186	0.061222965	1
34297	27.9	2332	34.1	526	0	1405	46.8	28474	25.7	18.01671	23.915	2	56	28	4	D	7832	6442	1434.120845	7.289267283	0	0	0	0	0	1
52203	25.5	3786	46.2	864	48	1304	44	44497	21.7	21.07528	35.106	1	20	20	0	R	13731	7850	1487.010767	0	1	0	0	0	0	1
53029	30	1955	34.8	5113	41.1	1600	41.8	42992	28	63.31134	24.149	1	50	50	0	D	10469	7472	2195.908733	56.57281865	0	0	0.1885761	0	0	1
201343	19.9	22023	39.6	10291	18.6	25099	29.9	138158	14.1	24.88409	81.124	4	240	60	15	D	36783	21818	2481.916572	6.20831119	0	0	0.099333	0	0	1
63887	27.6	3866	47.4	4984	33	3351	23.4	49849	26.1	59.18597	24.903	1	40	40	0	D	9909	5112	2565.433883	46.95791006	0	0	0	0	0	1
291067	8.3	16864	15.1	25066	15.2	23812	9.7	178423	4.9	33.24485	1704.718	2	145	72.5	0	D	31897	52323	170.742023	0	0	0	0.0343564	0	0	1
53982	29.5	7293	39.2	2296	43.5	3828	44.1	39275	24.7	26.84421	38.27	1	40	40	2	R	NA	NA	1410.556572	55.57408025	1	0	0	0	0	1
291027	19.3	41436	32.2	10488	15.9	20294	36.1	211320	15	41.02211	283.77	2	88	44	2	D	NA	NA	1025.573528	11.16734873	0	0	0	0	0	1
592342	18.4	134071	33	13453	18.2	27727	28.7	402201	12.3	27.67549	325.325	3	40	13.33333333	0	D	NA	NA	1820.769999	0	0	0	0	0	0	1
220776	25.5	123273	31	7848	22.9	7191	29.8	78848	16.7	32.85976	76.952	2	10	5	4	D	46870	32939	2869.009253	1.13236946	0	0	0.0452948	0	0.045294778	1
33510	20.9	8355	38.2	351	0	1690	30.3	21066	12.8	16.1454	14.408	1	0	0	1	NP	3607	7098	2325.791227	0	0	0	0	0	0	1
119433	19.4	36950	34.3	2398	20.4	5601	20.5	72852	11.5	33.77402	49.199	1	0	0	2	R	17575	28575	2427.54934	15.07121147	1	0.083728953	0.1674579	0	0	1
357411	27.7	214584	35.6	10456	22.4	19242	26	108478	13	34.74465	169.423	3	75	25	0	D	90964	24390	2109.577802	0	0	0.027978993	0.083937	0.027978993	0.08393698	1
196172	23.1	108826	33.7	2677	12.5	5379	16.2	76471	9.1	25.7094	105.39	1	14	14	16	D	43547	30498	1861.391024	1.529270232	0	0.050975674	0.152927	0.050975674	0.050975674	1
30857	25	499	41.5	705	20.9	634	32.8	28392	24.8	28.68452	34.396	1	0	0	6	D	NA	NA	897.1101291	9.722267233	0	0	0	0	0	1
64828	21	4674	59.9	2495	31	2160	29.4	53691	16.3	46.19217	21.331	2	109	54.5	0	D	NA	NA	3039.144906	11.56907509	0	0	0	0	0	1
37841	10.8	9070	24.2	673	8.9	6830	15.7	21215	4.3	45.94779	7.201	2	30	15	0	R	11394	5271	5254.964588	0	1	0	0.2642636	0	0.264263629	1
598360	24.2	379557	28.3	14496	22.9	26420	25.5	166696	14.4	28.07632	80.948	16	10124	632.75	19	D	197653	26137	7391.905915	9.191790895	0	0.100274082	0.2841099	0.050137041	0.183835818	1
65073	10.8	12194	17.8	4129	8.9	10776	12.9	35473	7.4	38.03203	20.764	1	75	75	5	R	14858	8984	3133.933731	15.36735666	1	0	0.1536736	0	0	1
30109	30.7	10729	36.1	969	36.5	1977	39	15368	25	28.20743	13.333	1	40	40	0	D	5758	3843	2258.231456	0	0	0	0.6642532	0	0.664253213	1
49545	11.4	9676	12.4	6728	8.2	20244	15.3	11818	6.5	34.3134	NA	1	20	20	0	NA	9771	3131	NA	0	0	0	0	0	0	1
43482	4.4	873	7.6	3927	7.1	1646	3.9	35989	4.1	67.48304	5.16	1	400	400	0	NA	14553	2280	8426.744186	0	0	0	0	0	0	1
597942	21.9	156752	25.4	53299	30.4	113622	33.2	266988	14.1	44.89895	48.266	16	5089	318.0625	0	D	200362	24487	12388.47222	2.592224664	0	0	0.0668961	0	0.05017209	1
57493	11.8	1392	25.2	9508	15.1	3199	13.6	42317	10.7	81.57751	6.766	3	625	208.3333333	0	NA	20286	3230	8497.339639	6.087697633	0	0	0	0	0	1
90989	15	10667	28.5	12367	18	6498	28.4	58334	10.2	74.42161	6.385	5	810	162	0	D	35814	2827	14250.4307	131.8840739	0	0	0	0	0	1
87370	23.3	3459	36.1	1893	26.1	7548	51	71460	19.5	13.90748	33.133	1	12	12	0	D	27903	3222	2636.948058	8.584182214	0	0	0	0	0	1
31887	4.4	350	16.6	7075	4.1	577	0.3	22947	4.2	77.71054	16.371	1	150	150	0	NA	8527	2261	1947.773502	0	0	0	0	0	0	1
54651	10.5	5075	13.2	3904	6.6	2479	18	41695	9.9	45.22462	8.089	1	0	0	0	D	16326	2538	6756.21214	27.44689027	0	0	0	0	0	1
79162	5.6	2497	22.6	10697	6.7	4006	11.9	59954	4.4	76.33716	17.836	1	500	500	0	D	25847	4541	4438.326979	6.316161795	0	0	0	0	0	1
54906	10.4	3275	19.2	5951	19.1	8111	19.7	36249	6	48.2283	12.741	1	75	75	0	R	12667	3439	4309.394867	27.31941864	1	0	0	0	0	1
171864	22	21873	23.1	12110	17.4	36584	40.9	98986	15.3	30.24974	37.388	3	420	140	8	D	41976	8400	4596.769017	18.91030117	0	0	0	0	0	1
105161	22.6	8159	26.2	15565	29.8	4978	25.2	72609	20.6	71.04776	27.806	4	250	62.5	8	D	51539	10664	3781.953535	29.00314755	0	0	0	0	0	1
685058	39.8	556491	39.5	8049	54.9	50450	40.4	57608	38.9	13.12972	138.751	4	885	221.25	0	D	251266	6394	4937.319371	4.379191251	0	0	0.1167784	0	0.029194608	1
32245	42.9	2321	47.4	2765	50.5	1126	28.5	24618	43.5	70.44021	13.594	2	230	115	0	D	19099	6226	2372.002354	93.03768026	0	0	0	0	0	1
68112	35	14616	46.6	1856	42.1	4717	46.9	44249	29.7	33.25913	24.679	1	50	50	10	D	25833	8051	2759.917339	3.670425182	0	0	0.293634	0	0	1
113788	29.4	24825	36.5	4697	30	14389	35.9	62901	23.2	25.13633	35.996	1	0	0	5	D	39937	12400	3161.129014	13.18240939	0	0	0	0	0	1
49564	35.5	22115	44.9	250	8.8	7197	34.1	18605	25.6	11.63563	17.338	1	13	13	5	D	18571	4401	2858.691891	0	0	0	0	0	0	1
84204	9	6464	26.8	5442	7.5	6088	23	63823	6	39.87894	34.678	1	1500	1500	0	R	28570	21703	2428.167714	8.906940288	1	0	0	0	0	1
80521	22.7	2132	65.9	1202	37.9	1265	21.5	72248	19.9	33.29145	67.858	1	100	100	11	D	35614	15255	1186.610274	31.04780119	0	0	0.1241912	0	0	1
378923	22.6	68165	48.1	21994	29.4	38013	28.2	229918	13	47.56101	54.783	9	6940	771.1111111	15	D	169204	34958	6916.799007	11.21599903	0	0	0.0527812	0	0.026390586	1
106974	9.7	6777	39.8	7743	6.3	5684	16.7	83780	7.1	42.8631	53.65	1	50	50	0	D	29297	24255	1993.923579	7.478452708	0	0	0	0	0	1
167412	29.9	134540	33.6	448	12.9	2536	43.4	28961	12.2	26.2064	108.692	1	100	100	5	D	58489	13259	1540.242152	6.271951831	0	0.059732875	0.0597329	0.059732875	0.059732875	1
104086	24.9	11283	35.5	6035	31.8	3312	24.1	79587	22.2	56.42311	62.547	6	720	120	6	R	33114	20779	1664.124578	43.71385201	1	0	0	0	0	1
51372	8.2	14505	9.1	528	55.1	694	19.5	33506	6.7	23.29332	12.438	1	20	20	0	NP	20562	9586	4130.24602	0	0	0	0	0	0	1
457040	19.4	133421	31.7	11204	21.5	46800	29.1	251269	10.7	31.82266	314.385	3	75	25	17	D	146838	61344	1453.758926	7.657973044	0	0.043759846	0.1531595	0	0.021879923	1
73038	19.5	3801	30.3	736	5.3	4605	24.6	61321	17.6	20.0183	43.987	1	21	21	8	R	15199	14529	1660.445132	10.26862729	1	0	0.136915	0	0	1
308873	27.8	149191	39.2	8223	28.7	11296	27.7	132002	14.7	30.843	61.876	32	7249	226.53125	15	D	132925	24662	4991.806193	36.42273685	0	0.032375766	0.3561334	0.032375766	0.323757661	1
151492	26.4	5761	43.7	2970	39.7	6203	39.2	131175	24.4	27.06321	81.357	1	75	75	12	R	35295	40456	1862.064727	33.3350936	1	0.066010086	0.1980303	0	0	1
65714	19.8	385	75.8	767	31.9	2183	24.6	58815	18.8	45.98488	26.625	1	50	50	0	D	NA	NA	2468.131455	22.8261862	0	0	0	0	0	1
52598	17.1	572	0.5	1274	7.6	11791	30.4	36754	13	21.06555	144.663	1	25	25	0	D	11611	11416	363.5898606	4.753032435	0	0	0	0	0	1
589447	17.7	65913	27.8	37521	13.1	190055	24.5	273424	10.8	21.74777	135.041	1	100	100	0	D	118169	78789	4364.948423	0	0	0.016965054	0.1866156	0.016965054	0.016965054	1
220378	16.9	44157	21.1	13305	8.5	86384	23.7	65888	7.7	15.34254	81.686	1	40	40	0	D	42915	17976	2697.867444	0	0	0.04537658	0.0907532	0	0.04537658	1
226490	19.1	6594	42.6	14564	14.2	56311	28.1	139998	14.8	29.62574	101.644	1	0	0	0	D	53324	33971	2228.267286	0	0	0	0.0883041	0	0.04415206	1
39180	35.8	15812	38.3	6383	26.8	9814	43.3	7358	29.8	15.69836	10.746	1	0	0	8	R	9850	1990	3646.007817	0	1	0	0.2552323	0	0.255232261	1
74540	39.3	35866	38.5	1853	23.6	36604	42.2	3000	31.4	8.454274	8.921	3	95	31.66666667	12	D	19934	1127	8355.56552	10.06171183	0	0	0	0	0	1
253412	19	64451	25.9	63365	11.6	69773	25.6	53878	13	42.73842	14.794	1	30	30	13	D	49346	10337	17129.37677	6.90575032	0	0.03946143	0.1183843	0.03946143	0.118384291	1
267061	29.9	134746	33.1	4267	31.1	95182	30.3	28608	17.6	13.53595	24.186	5	1210	242	15	D	64281	3688	11041.96643	5.616694313	0	0	0.1123339	0	0	1
47935	34.9	6330	25.8	2735	44.9	29889	32	8864	48.6	20.0869	5.227	2	120	60	9	D	10046	1796	9170.652382	62.58475018	0	0	0.2086158	0	0	1
144799	28.4	45654	29.6	5475	39.9	84843	26.1	12040	31.3	10.71372	8.428	1	0	0	4	D	35511	3841	17180.70717	5.179593782	0	0	0.0690613	0	0	1
548804	18.5	17985	26.5	14476	11.4	259976	23.5	227061	11.5	33.24664	188.043	4	169	42.25	15	R	141814	88880	2918.502683	10.93286492	1	0.018221442	0.2368787	0	0.054664325	1
88947	26.7	29123	30.9	5654	49	8216	37	44319	18.2	36.66358	21.386	2	200	100	0	D	31019	6467	4159.122791	20.23677021	0	0	0	0	0	1
45361	33.9	5839	50.2	2334	58.7	3035	57.1	32935	27.4	23.62227	10.42	1	40	40	6	R	10940	5993	4353.262956	0	1	0	0	0	0	1
251561	30.9	95775	39	9918	49.6	25382	50.6	113878	17.5	24.82029	40.39	2	20	10	15	D	83990	18183	6228.299084	7.950357965	0	0	0	0	0	1
8219906	20.6	2028266	23.2	1095334	20.3	2376798	28.9	2682684	12.5	35.19234	303.374				54	D	2058903	510502	27094.9587	5.066967919	0	0.008515912	0.0328471	0.008515912	0.019464943	1
201739	33.8	84385	41.5	6444	34.9	35683	43.8	71731	19.5	24.52114	35.809				4	D	61622	11909	5633.751292	7.435349635	0	0	0.049569	0	0.049568998	1

131280	35.1	40904	43.7	7777	47.9	10868	55.6	65954	23.8	26.28899	25.043	12	D	39116	10709	5242.183444	25.51797684	0	0	0.0761731	0	0	1
56225	10.1	6494	24.2	3847	2.9	19324	14.9	25226	4.9	48.24901	9.761	5	D	16325	6897	5760.168016	0	0	0	0	0	0	1
48950	21.7	4125	40.4	6794	9.5	2756	19.6	33976	22	74.52035	21.119	1	D	20473	5069	2317.818079	61.28702758	0	0	0	0	0	1
762199																							
230992																							
191807																							
265381																							
403167																							
225142																							
45369																							
285416																							
381531																							
790894																							
129185																							
276336																							
49914																							
152380																							
589292																							
48719																							
1503859																							
282694																							
70547																							
37902																							
164833																							
119805																							
103544																							
36988																							
56534																							
96711																							
165147																							
61219																							
172824																							
640637																							
601882																							
503216																							
844908																							
87649																							
1224026																							
112193																							
661852																							
45084																							
2136523																							
223341																							
226852																							
133552																							
142353																							
1360649																							
95796																							
185340																							
35859																							
145129																							
41829																							
40971																							
131988																							
43328																							
57729																							
172654																							
220995																							
200784																							
96176																							
434326																							
131842																							
35437																							
46758																							
619248																							
196389																							
193311																							
581847																							
37836																							
355537																							
105054																							
63969																							
222380																							
57559																							
3793097																							
396963																							
65580																							
136770																							
303080																							
468960																							
1306746																							
815670																							
976164																							
99596																							
336686																							
93365																							
621206																							
117804																							
121638																							
124235																							

Appendix C

Sources Used

We used one source from JournalDev to learn how to drop the Null/NA values from our DataFrame, which greatly impacted our ability to conduct the linear analyses. The citation for this source is included below.

Pankaj. (2021). *"Pandas Dropna() - Drop Null/NA Values from DataFrame."* www.journaldev.com/33492/pandas-dropna-drop-null-na-values-from-dataframe. Accessed 27 April 2021.