Intelligence Analytics Challenge 5.0

We believe the first priority for kids is to survive. In the dataset from UNICEF, we choose under-5 mortality rate to evaluate the quality of life of the kids. After performing Regression analysis, we found sanitation services, drinking water services, polio vaccine, and number of internet users have a strong negative relationship with under-5 mortality rate. It means increasing any one of these features could decrease under-5 mortality rate. (Figure 1)

In the Demographic Indicators category, we notice a trend between the ratio of under-5 population and under-5 mortality rate. The higher ratio of under-5 population tends to have higher under-5 mortality rate, and the trend is exponential (Figure 2). So, we further inspect the countries that are at the top right corner. Most of those countries are in Africa with a low percent of sanitation services and drinking water services, and are among the list of least developed countries, which is the main reason they have a higher under-5 mortality rate. (Figure 3)

Figure 1

OLS Regression Results

Dep. Variable:	Under5_mortality_rate	R-squared:	0.814
Model:	OLS	Adj. R-squared:	0.810
Method:	Least Squares	F-statistic:	207.4
Date:	Mon, 04 May 2020	Prob (F-statistic):	3.76e-68
Time:	10:47:34	Log-Likelihood:	174.55
No. Observations:	195	AIC:	-339.1
Df Residuals:	190	BIC:	-322.7
Df Model:	4		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.9233	0.042	22.040	0.000	0.841	1.006
sanitation_service	-0.3318	0.053	-6.313	0.000	-0.435	-0.228
drinkingwater_services	-0.1873	0.070	-2.670	0.008	-0.326	-0.049
internet	-0.1476	0.044	-3.358	0.001	-0.234	-0.061
polio3	-0.2785	0.047	-5.901	0.000	-0.372	-0.185

Omnibus:	14.070	Durbin-Watson:	2.035
Prob(Omnibus):	0.001	Jarque-Bera (JB):	30.628
Skew:	0.273	Prob(JB):	2.23e-07
Kurtosis:	4.863	Cond. No.	21.3

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 2

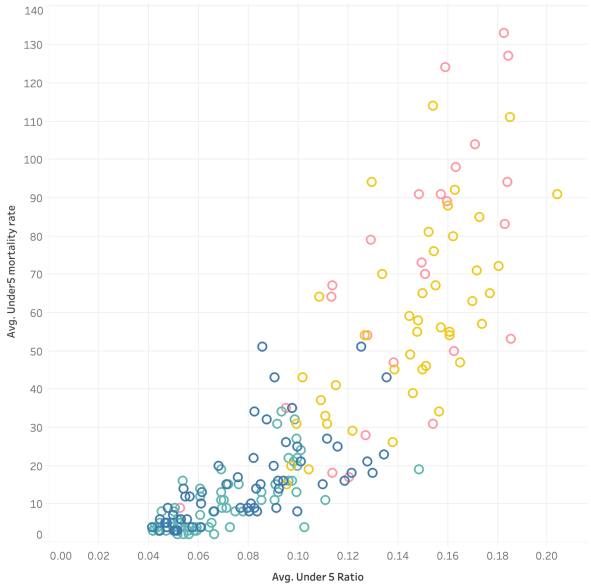
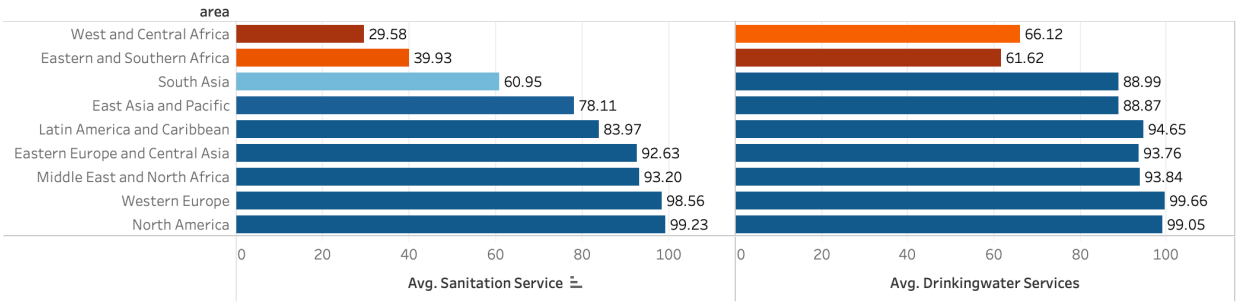


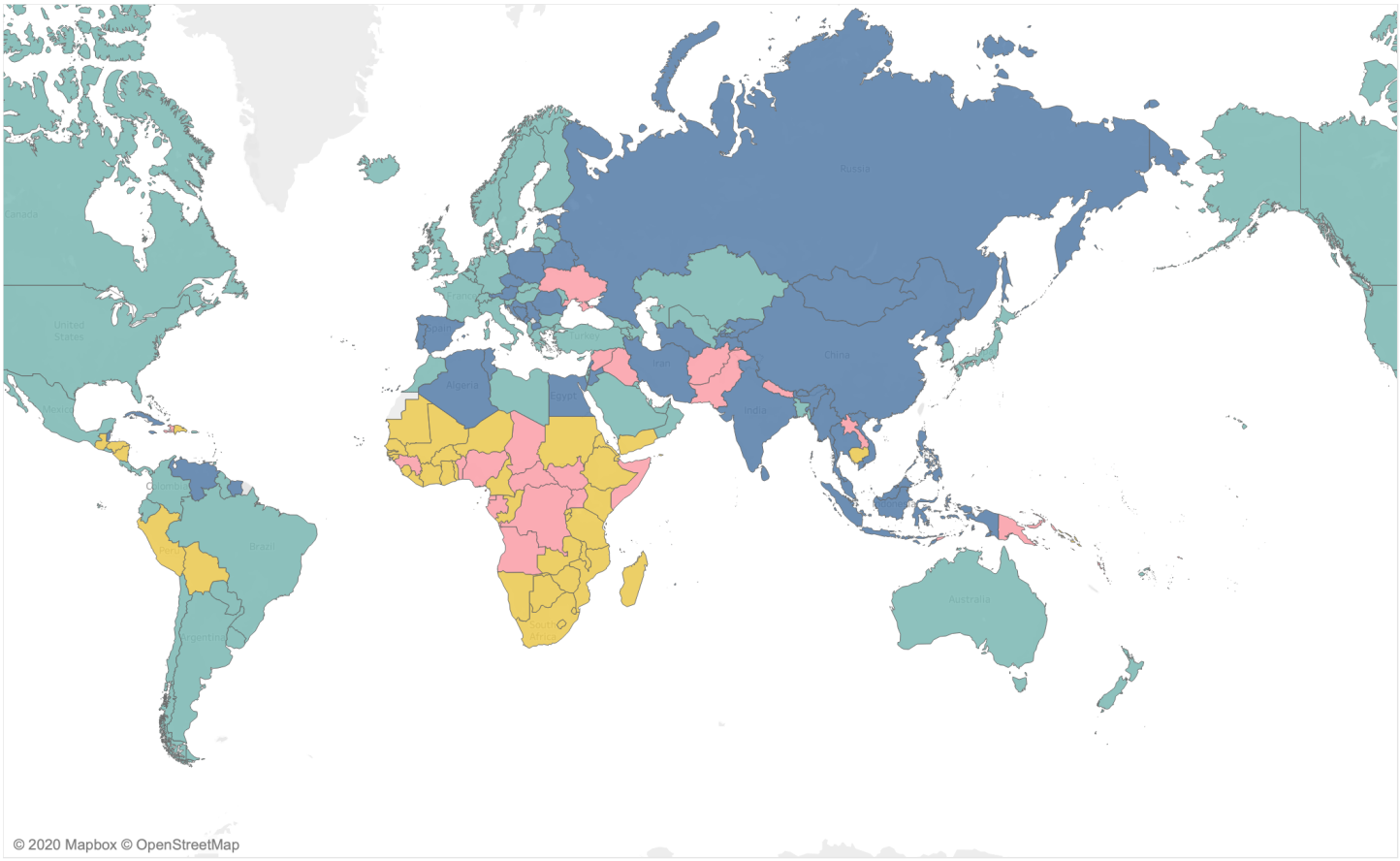
Figure 3



To decrease under-5 mortality rate, according to our model, we suggest people to enhance sanitation service and drinking water services, increase injection of Polio vaccine, and expand the number of internet users (Figure 1). Besides an overall advice, we provide more specific recommendations for each cluster. We classify all countries into 4 clusters based on their under-5 mortality rate by K-Means modeling (Figure 4). Although we complete analysis for all 4 clusters,

our recommendations would mainly focus on clusters with over 20% of under-5 mortality rate, which is the cluster pink and yellow in Figure 4. That is because their situation is more urgent.

Figure 4



Based on our regression model, for both these 2 clusters (Figure 5 and 6), increasing 1% in sanitation services or polio vaccine injection would help the under-5 mortality rate decrease by around 0.4 thousandths. In addition, increasing 100 internet users would help the under-5 mortality rate decrease by around 0.8 thousandths for cluster pink (Figure 6).

Figure 5

OLS Regression Results

Dep. Variable:	Under5_mortality_rate	R-squared:	0.407
Model:	OLS	Adj. R-squared:	0.378
Method:	Least Squares	F-statistic:	14.09
Date:	Mon, 04 May 2020	Prob (F-statistic):	2.21e-05
Time:	11:42:37	Log-Likelihood:	24.420
No. Observations:	44	AIC:	-42.84
Df Residuals:	41	BIC:	-37.49
Df Model:	2		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	0.9357	0.148	6.318	0.000	0.637	1.235
sanitation_service	-0.4282	0.095	-4.525	0.000	-0.619	-0.237
polio3	-0.4307	0.174	-2.475	0.018	-0.782	-0.079

Omnibus:	1.397	Durbin-Watson:	2.250
Prob(Omnibus):	0.497	Jarque-Bera (JB):	1.309
Skew:	0.396	Prob(JB):	0.520
Kurtosis:	2.704	Cond. No.	14.1

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

Figure 6

OLS Regression Results

Dep. Variable:	Under5_mortality_rate	R-squared:	0.770
Model:	OLS	Adj. R-squared:	0.740
Method:	Least Squares	F-statistic:	25.72
Date:	Mon, 04 May 2020	Prob (F-statistic):	1.57e-07
Time:	11:42:37	Log-Likelihood:	16.591
No. Observations:	27	AIC:	-25.18
Df Residuals:	23	BIC:	-20.00
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	1.0386	0.083	12.520	0.000	0.867	1.210
sanitation_service	-0.4023	0.124	-3.243	0.004	-0.659	-0.146
polio3	-0.3949	0.126	-3.134	0.005	-0.656	-0.134
internet	-0.8106	0.291	-2.788	0.010	-1.412	-0.209

Omnibus:	3.888	Durbin-Watson:	1.679
Prob(Omnibus):	0.143	Jarque-Bera (JB):	1.574
Skew:	-0.140	Prob(JB):	0.455
Kurtosis:	1.851	Cond. No.	13.9

Warnings:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.