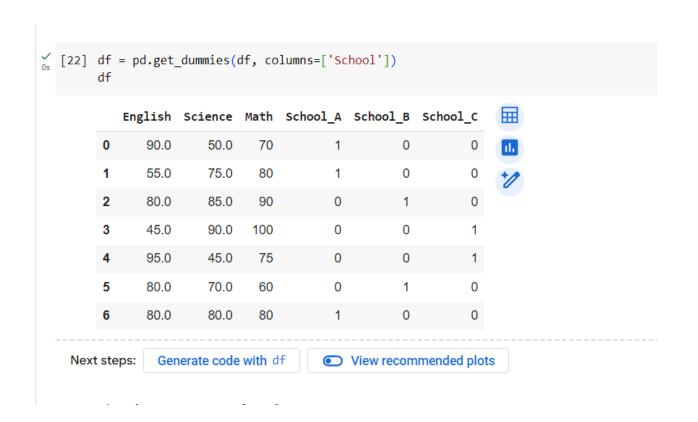
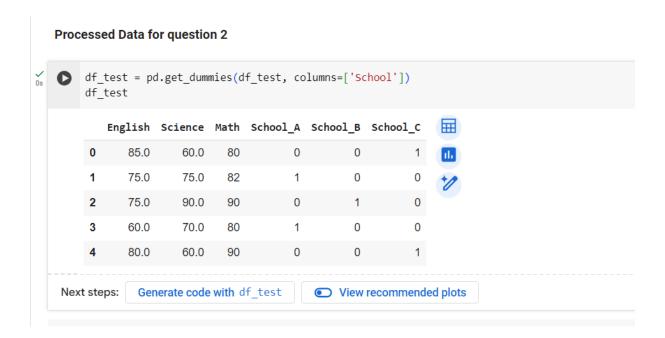
Sabura Khanam

KSU ID: 001057990

1. Train Data Set after applying one-hot encoder and imputing with the median values.



2. In addition, you have a trained linear regression model with the intercept and coefficients: $\beta 0=14,\beta Sch_A=-1,\beta Sch_B=-10,\beta Sch_C=10,\beta Eng=0.2$, and $\beta Sci=0.8$. Apply the same data preprocessing techniques as in Q1 and manually calculate the predictions using the regression model. You can either type or scan manually written results. Note that imputation should use the median from the training set.



Prediction of the Data Set:

```
    a = -1*0 + -10*0 + 10*1 + 0.2*85 + 0.8*60 + 14

    b = -1*1 + -10*0 + 10*0 + 0.2*75 + 0.8*75 + 14

    c = -1*0 + -10*1 + 10*0 + 0.2*75 + 0.8*90 + 14

    d = -1*1 + -10*0 + 10*0 + 0.2*60 + 0.8*70 + 14

    e = -1*0 + -10*0 + 10*1 + 0.2*80 + 0.8*60 + 14

    print(a)
    print(b)
    print(c)
    print(d)
    print(e)

    89.0
    88.0
    91.0
    81.0
    88.0
```

3. Evaluate the model provided in Q2 by computing the following metrics on the test dataset: Mean Squared Error (MSE), Mean Absolute Error (MAE), and R2 score.

Datuset: -			
	Achal man (y)	Redicted With (5)	
4	80	89	
	82	88	
	90	91	
	80	81	
	90	88	
	$\sum_{i=0}^{n-1} y - \hat{y}$		
	\$ (180-89)+	+ 190-801)	+ 80-8]
	$\frac{1}{5} = 3.8$	72)	
MSE = = = ((80-89)~+(8	2-88) + (90-91 + (90-88)))~+(80-81)2
$= \frac{1}{5}$ $= \frac{123}{5}$ $= \frac{29}{29}$ CS Scanned with C			

$$R^{\gamma} = 1 - \frac{\tilde{E}}{\frac{\tilde{E}}{2}} (\gamma - \tilde{\gamma})^{\gamma}$$

$$y = \frac{80+82+90+80+90}{5} = \frac{422}{5} = 84.4$$

$$(y-\hat{y})^{2} = (80-89)^{2} + (82-88)^{2} + (90-91)^{2} + (80-81)^{2}$$

$$+ (90-88)^{2}$$

$$= 81 + 36 + 1 + 1 + 4$$

$$= 123$$

$$(\gamma - \gamma)^{2} = (80 - 89.9)^{2} + (82 - 89.9)^{2} + (90 - 89.9)^{2} + (80 - 89.9)^{2} + (90 - 84.9)^{2} + (80 - 89.9)^{2} + (90 - 84.9)^{2}$$

$$= 13.6 + 9.6 + 30.6 + 13.6 + 30.6$$

$$= 105.2$$

$$(R^{2} = 1 - \frac{123}{105.2})$$

$$= 1 - 1.168$$

$$= -0.168$$

4. Suppose you have another model trained with a different dataset, which consists of the following intercept and coefficients: $\beta 0=14$, $\beta Sch_A=-2$, $\beta Sch_B=-11$, $\beta Sch_C=12$, $\beta Eng=0.2$, and $\beta Sci=0.8$. Compute the predictions with the test dataset and its MSE and MAE. Compare them with the results in Q3 to explain which metric penalizes the model generating more extreme errors.

```
m = 0*-2 + 0*-11 + 1*12 + 85*0.2 + 60*0.8 + 14
n = 1*-2 + 0*-11 + 0*12 + 75*0.2 + 75*0.8 + 14
o = 0*-2 + 1*-11 + 0*12 + 75*0.2 + 90*0.8 + 14
p = 1*-2 + 0*-11 + 0*12 + 60*0.2 + 70*0.8 + 14
q = 0*-2 + 0*-11 + 1*12 + 80*0.2 + 60*0.8 + 14

print(m)
print(n)
print(o)
print(p)
print(q)

31.0
87.0
90.0
80.0
90.0
80.0
90.0
```

Actual (7)
80
82
90
80
90

$$MAE = \frac{1}{5} (|80-91| + |82-82| + |90-90| + |80-80| + |90-90|)$$

$$= \frac{1}{5} (|11+5+0+0+0|)$$

$$= \frac{1}{5} \times 16$$

$$= |3\cdot 2|$$

$$MSE = \frac{1}{5} (|80-91|)^{2} + (82-82)^{2} + (90-90)^{2} + (80-80)^{2} + (90-90)^{2})$$

$$= \frac{1}{5} (|121+25+0+0+0|)$$

$$= \frac{1}{5} \times 196$$

$$= |29\cdot 2|$$

From dataset 3, the higher MSE is penalized more for generating extreme errors compared to dataset from 4. Therefore, MSE is the metric that penalizes the model generating more extreme errors.

$$MAE = \frac{1}{5} \sum_{i=0}^{n=1} | y - \hat{y} |$$

$$= \frac{1}{5} x \left(80 - 91 \right) + \left(82 - 83 \right) + \left(90 - 90 \right) + \left(80 - 80 \right) + \left(90 - 90 \right) + \left(80 - 80 \right) + \left(90 - 90 \right) + \left(80 - 80 \right) + \left(90 - 90 \right) + \left(80 - 80 \right) + \left(90 - 90 \right) + \left(80 - 80 \right) + \left(90 - 90 \right) + \left(80 - 90 \right) + \left(90 - 90 \right) + \left(80 - 90 \right) + \left(90 - 90 \right) + \left(80 - 90 \right) + \left(90 - 90 \right) + \left(90$$