Using Supervised learning model to determine radiation levels around uranium mines.

By :
-Amrit Krishn
-Mahfooz Alam

INTRODUCTION :

- The motivation behind this project is the life of the people living near mine areas and unaware whether it's safe to live there just because it's not possible to do survey in their area .

- To overcome this we created a supervised model that accepts data from other surveys to predict the radiation level of remote areas.

- It will ensure the safety of hundreds of residents who live close to the uranium mine.

METHODOLGY :

->The program is using supervised learning algorithm to train it's model.

->The algorithm predicts the value of Th, U, K by fitting the multiple linear regressions models based on pH, organic matter and texture.

->After getting the values for Th, U, K a formula if applied that takes in conc of Th, U, K and outputs the gamma dose rate .

->At the end the program checks whether the resultant gamma dose rate is within the standard values or not.

ABOUT DATASETS :

-->Synthetic data was generated due to lack of availability of feasible data.

```
n_samples = 100
latitudes = np.random.uniform(low=30, high=35, size=n_samples)
longitudes = np.random.uniform(low=-120, high=-115, size=n_samples)
elevations = np.random.uniform(low=0, high=2000, size=n_samples)
depths = np.random.uniform(low=0, high=2, size=n_samples)
ph_values = np.random.uniform(low=4, high=8, size=n_samples)
om_values = np.random.uniform(low=0, high=20, size=n_samples)
texture_values = np.random.choice(['Sandy Loam', 'Clay Loam', 'Silt Loam'], size=n_samples)
```

->The values for lat , long ,elevations ,depths ,pH ,organic matter ,texture were generated
Using **Numpy's 'random funciton'.**

```
# Calculate Th, U, and K
Th = 0.2*survey_data['pH'] + 0.4*survey_data['organic_matter'] + 0.1*(survey_data['texture']==0)
U = 0.5*survey_data['pH'] + 0.1*survey_data['organic_matter'] + 0.2*(survey_data['texture'] == 1)
K = 0.3*survey_data['pH'] - 0.1*survey_data['organic_matter'] + 0.2*(survey_data['texture'] == 2)
```

->The calculation for the values for Th, U ,K using some hypothetical relations between the params.

Units:
pH: dimensionless ;conc: ppm ;**o_m**: % ;
Texture: categorical ;
gama dose rate: miuSv/h;

ALGORITHMS :

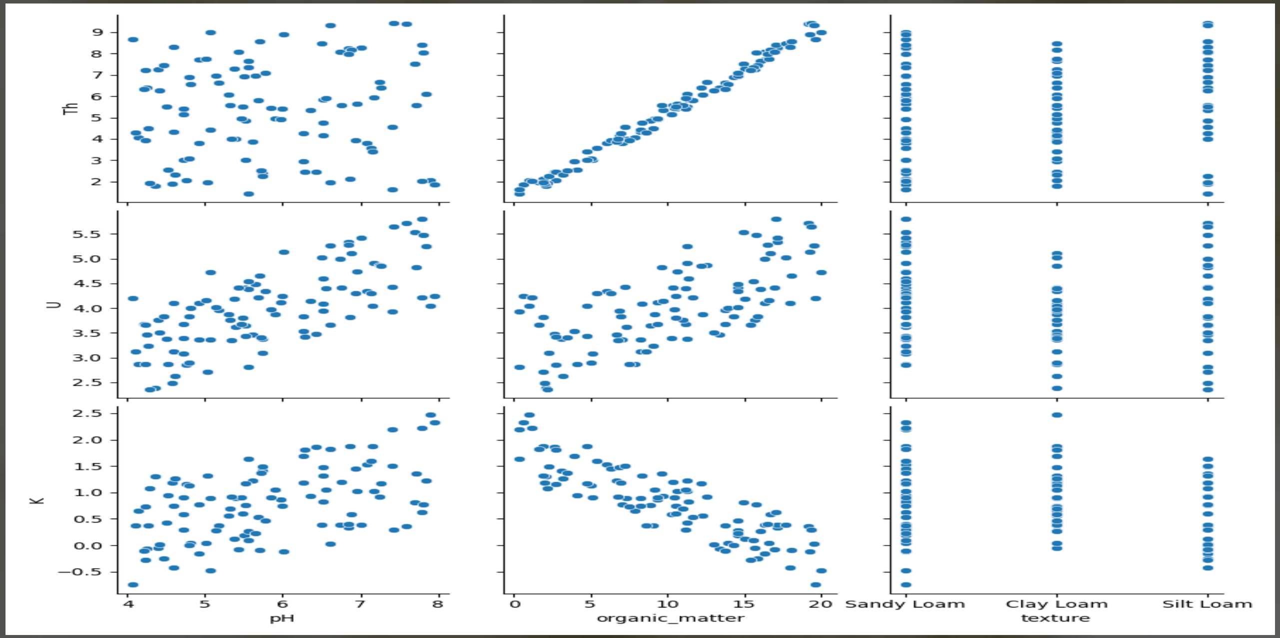
Multivariate statistical analysis is a technique that can be used to identify patterns and relationships between multiple parameters in a dataset. By analyzing the correlations between different parameters, it may be possible to identify relationships between these parameters and the concentration of Th. U, and K. This technique is often used in exploratory data analysis and can be performed using various statistical software packages, including Python.

BREAKDOWN OF THE CODE:

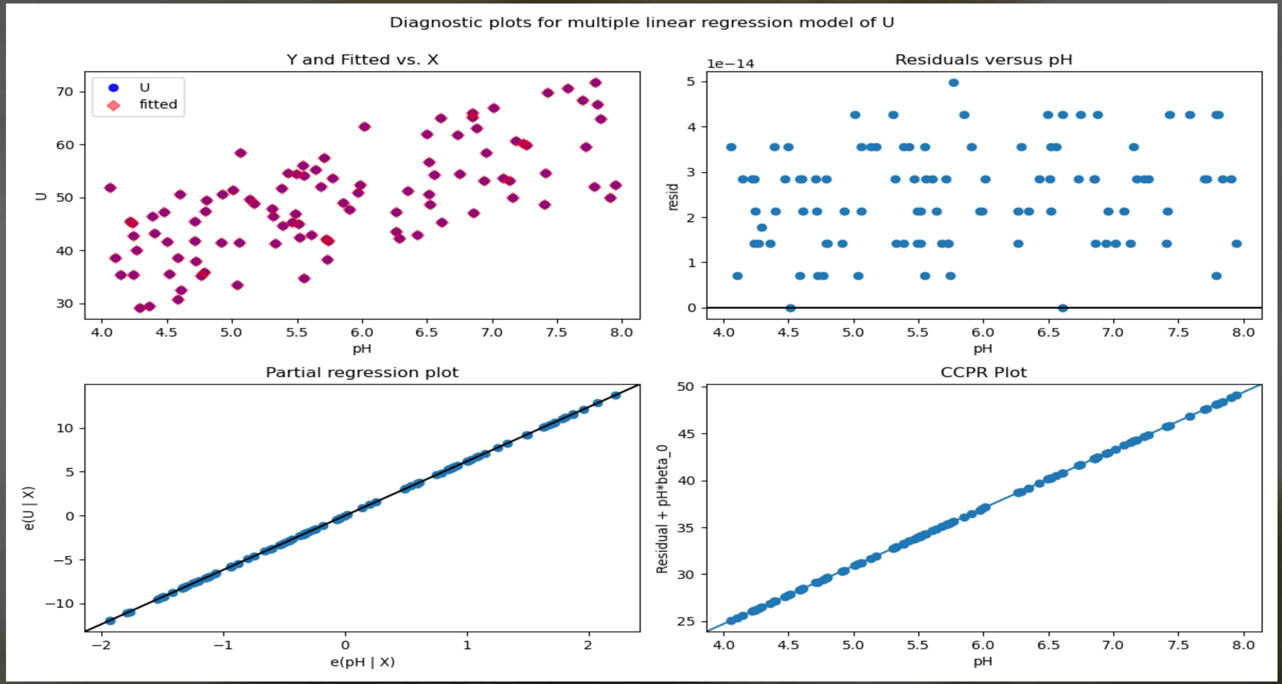
```
# Load the survey data into a pandas dataframe
survey_data = pd.read_csv('./synthetic_survey_data.csv')

# Define the predictor variables and the response variables
X = survey_data[['pH', 'organic_matter', 'texture']]
y_th = survey_data['Th']*4.06
y_u = survey_data['U']*12.35
y_k = survey_data['K']*302 #these numbers are conversions factors from ppm to bq per kg
```

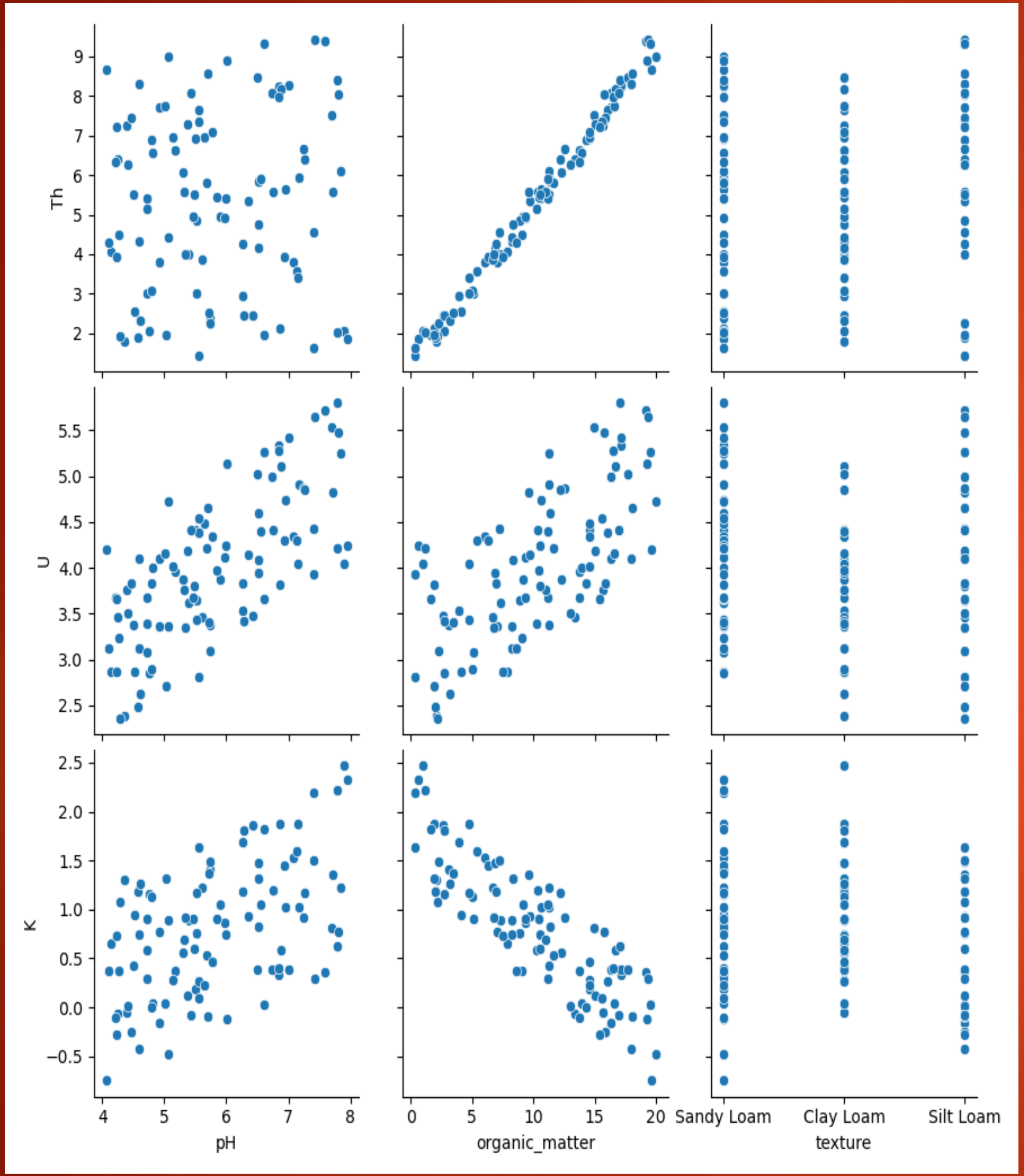
```
# Fit multiple linear regression models to predict Th, U, and K based on pH, organic matter, and texture
model_th = sm.OLS(y_th, X).fit()
model_u = sm.OLS(y_u, X).fit()
model_k = sm.OLS(y_k, X).fit()
```

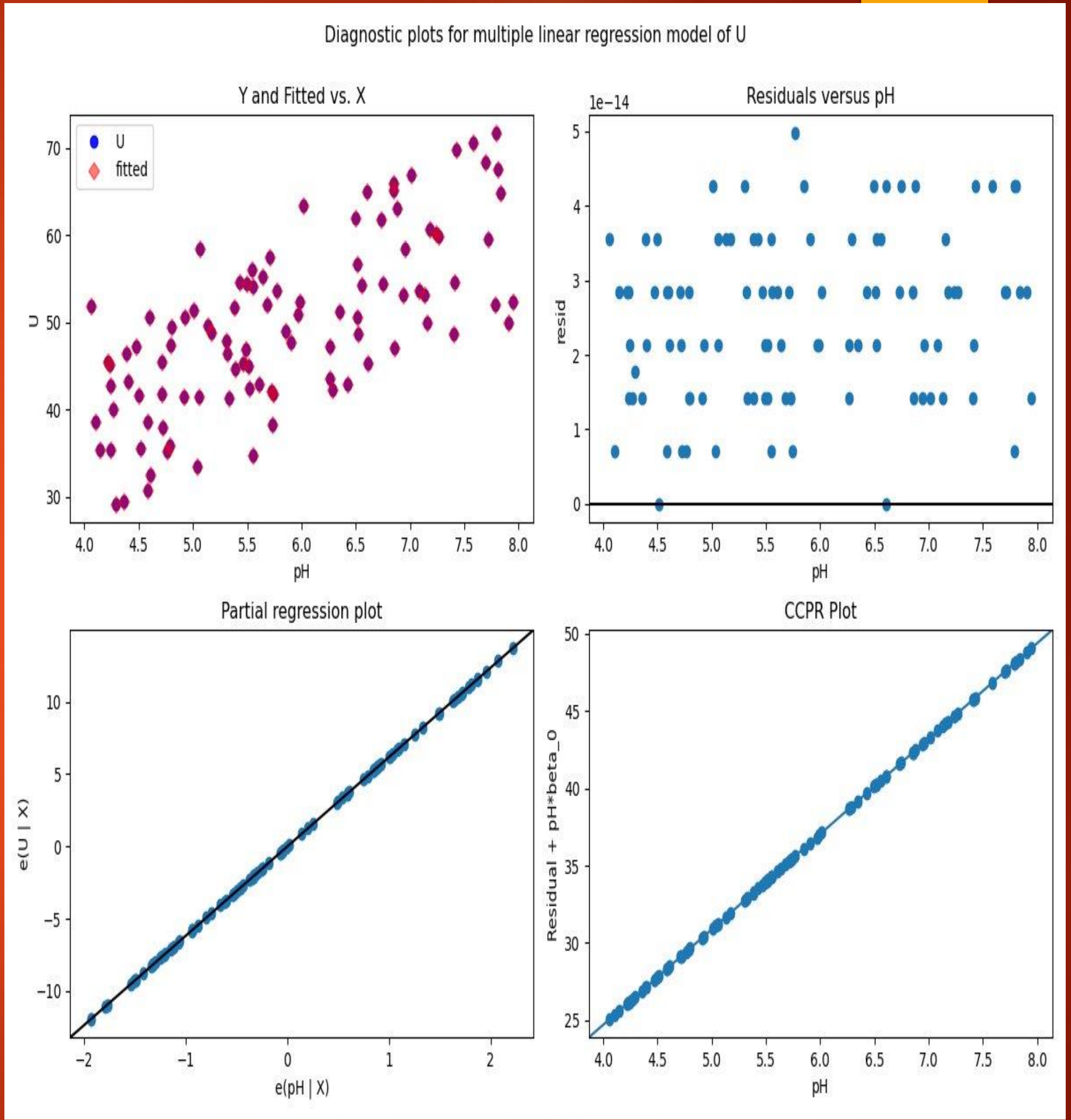


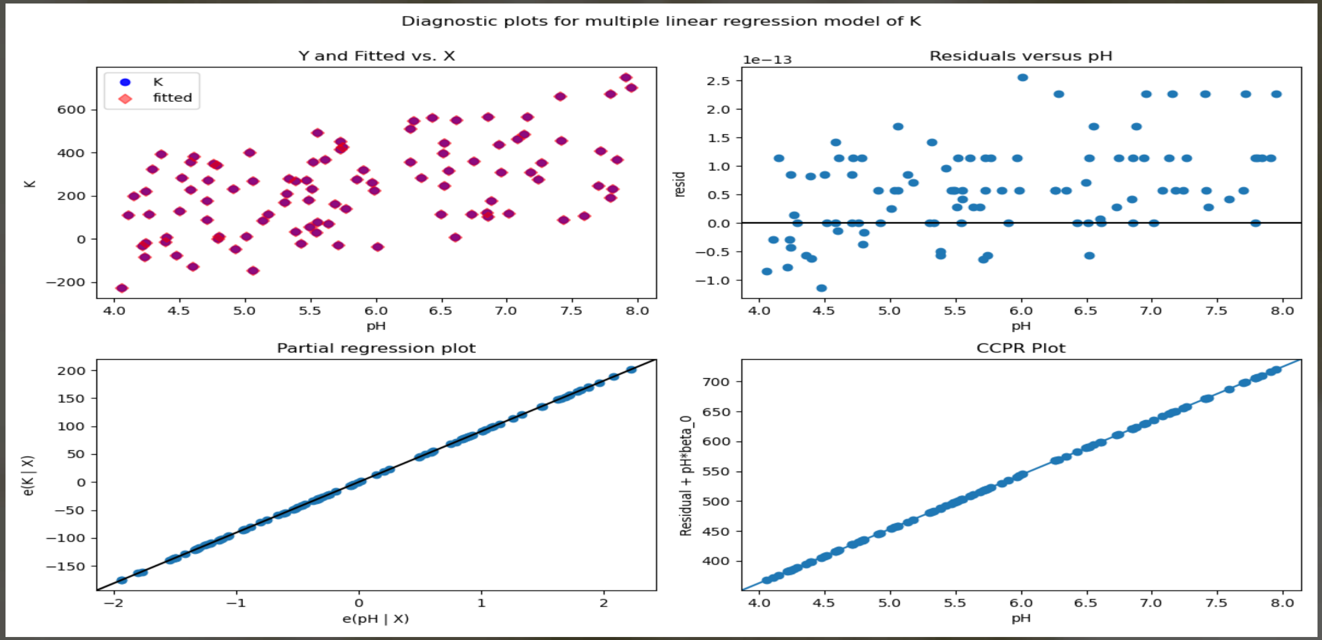

```python
# Scatter plots of predictor variables against response variables
sns.pairplot(survey_data, x_vars=['pH', 'organic_matter', 'texture'], y_vars=['Th', 'U', 'K'], height=3)
```



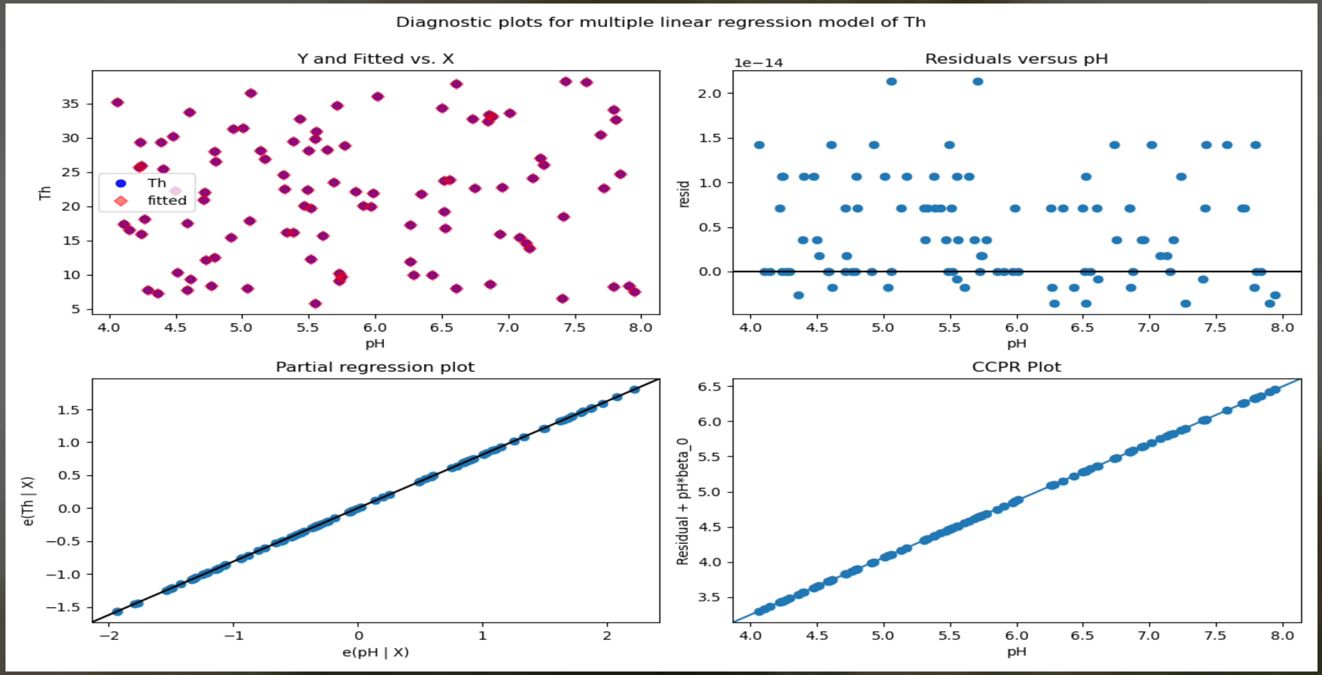

```python
# Generate diagnostic plots for model_u
fig_u = sm.graphics.plot_regress_exog(model_u, 'pH', fig=plt.figure(figsize=(12, 8)))
fig_u.suptitle('Diagnostic plots for multiple linear regression model of U')
plt.show()
```


Th
U
K
pH
organic_matter
Sandy Loam
Clay Loam
Silt Loam
texture


Diagnostic plots for multiple linear regression model of U
Y and Fitted vs. X
U
fitted
pH
Residuals versus pH
1e−14
resid
Partial regression plot
e(U | X)
e(pH | X)
CCPR Plot
Residual + pH*beta_0




```python
# Generate diagnostic plots for the Th model
fig_Th = sm.graphics.plot_regress_exog(model_th, 'pH', fig=plt.figure(figsize=(12, 8)))
fig_Th.suptitle('Diagnostic plots for multiple linear regression model of K')
plt.show()
```




```python
# Generate diagnostic plots for model_k
fig_k = sm.graphics.plot_regress_exog(model_k, 'pH', fig=plt.figure(figsize=(12, 8)))
fig_k.suptitle('Diagnostic plots for multiple linear regression model of K')
plt.show()
```

```python
# Use the models to predict the concentrations of Th, U, and K in unsampled areas
new_data = pd.DataFrame({'pH': [6.5, 7.0, 7.5],
                         'organic_matter': [2.5, 3.0, 3.5],
                         'texture_Cl': [0, 1, 0],
                         'texture_Sandy Loam': [1, 0, 0],
                         'texture_Silt Loam': [0, 0, 1]})
predictions_th = model_th.predict(new_data)
predictions_u = model_u.predict(new_data)
predictions_k = model_k.predict(new_data)
```

```python
# Output the predicted concentrations of Th, U, and K and the air absorbed gamma dose rate
print('Predicted activity concentrations of Th (Bq/kg):')
print(predictions_th)
print('Predicted activity concentrations of U (Bq/kg):')
print(predictions_u)
print('Predicted activity concentrations of K (Bq/kg):')
print(predictions_k)
print('Air absorbed gamma dose rate (nGy/h):')
print(gamma_dose_rate)
if gamma_dose_rate.max() <= 1/0.0126:
    print('Airborne gamma dose rate is within limits.')
else:
    print('Airborne gamma dose rate is above the limit of 1 microsievert per hour.')
```

The working code can be found here :
https://colab.research.google.com/drive/10mHrKR55EJPUe72NVNL2SPDHNcUzKqi9#scrollTo=MhE1OEU0wVo7
That's all for the code part.
All the data are in their SI units !

RESULTS:

--> Based on the data feeded as the sample check input it gives :

```
Predicted activity concentrations of Th (Bq/kg):
0      9.338
1     10.962
2     12.586
dtype: float64
Predicted activity concentrations of U (Bq/kg):
0     45.695
1     46.930
2     50.635
dtype: float64
Predicted activity concentrations of K (Bq/kg):
0     513.4
1     604.0
2     573.8
dtype: float64
Air absorbed gamma dose rate (nGy/h):
0     37.2384
1     42.4274
2     43.0368
dtype: float64
```

--> since the maximum gamma dose rate is 43.0368 and the max cricitical value is 79.36 nGy/h ;
So our area is within the safe limit of the radiation.

DISCUSSION :

->The research paper that was followed gave us the idea to devise this model as by just knowing the conc of Th, U ,K one can devise the radioactivity level using the formula :

$$D = 0.463A_{\mathrm{U}} + 0.604A_{\mathrm{Th}} + 0.0417A_{\mathrm{K}}$$

->Also from research paper one more relation was used to convert the conc. In ppm to Bq/kg:

$$1\%\ \mathrm{K} = 302\ \mathrm{Bq\ kg^{-1} of\ {}^{40}K},$$
$$1\ \mathrm{ppm\ U} = 12.35\ \mathrm{Bq\ kg^{-1} of\ {}^{238}U},$$
$$1\ \mathrm{ppm\ Th} = 4.06\ \mathrm{Bq\ kg^{-1} of\ {}^{232}Th}$$

->It should be noted that the accuracy of the multivariate statistical analysis depends on the quality and representativeness of the survey data, as well as the selection of appropriate predictors. Therefore, caution should be exercised when interpreting the results obtained from this technique.

CONCLUSION :

-->The beauty of this program is that it just need the value of organic matter ,pH ,texture which are easily available from other surveys data such as agricultural soil testing surveys ,ground water testing surveys ,etc.
So to devise the radioactivity level from this program we actually do not need to conduct any physical survey rather than we can ask for the data from other survey institutions to perform our work.

-->If the values are under limit , no worries but if against the limit than preventive measure can be taken to reduce the limit and the same survey can be done to verify it .

REFRENCES:

-->https://www.geeksforgeeks.org/linear-regression-in-python-using-statsmodels/

-->https://www.sciencedirect.com/science/article/pii/S0265931X11002955

-->https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwixnf-b_6H-AhUjzjgGHclkCN8QFnoECAwQAQ&url=https%3A%2F%2Fwww.section.io%2Fengineering-education%2Fseaborn-tutorial%2F&usg=AOvVaw2gBfM2lJ4YIOekb5I1k7BJ

-->https://www.researchgate.net/publication/51865999_An_assessment_of_the_radiological_scenario_around_uranium_mines_in_Singhbhum_East_District_Jharkhand_India