Impact of Park Spaces and the Health of the Community

With heart disease being the leading cause of death in the United States, it's crucial to explore ways to improve public health and prevent the onset of this and other health ailments. Physical activity, as recommended by the CDC, is a proven way to improve health and lower the risk of heart disease and other health problems. The suggestion of physical activity is somewhat vague. There are those who want to be more physically active, however, there may be a barrier of finding accessible and affordable ways to engage in physical activity can be a challenge, particularly for those living in urban areas.

Parks provide a unique solution to this problem, offering a wide range of opportunities for physical activity such as walking paths, trails, basketball courts, and tennis courts - all for free or low cost. This presents an opportunity to investigate the potential impact of increasing the number of parks in a city on the health of its citizens. By exploring this relationship, we hope to inform policy decisions and contribute to a healthier, more active community.

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

```
#'Run this cell to mount your drive'
from google.colab import drive
drive.mount('/content/drive')
```

```
Mounted at /content/drive
```

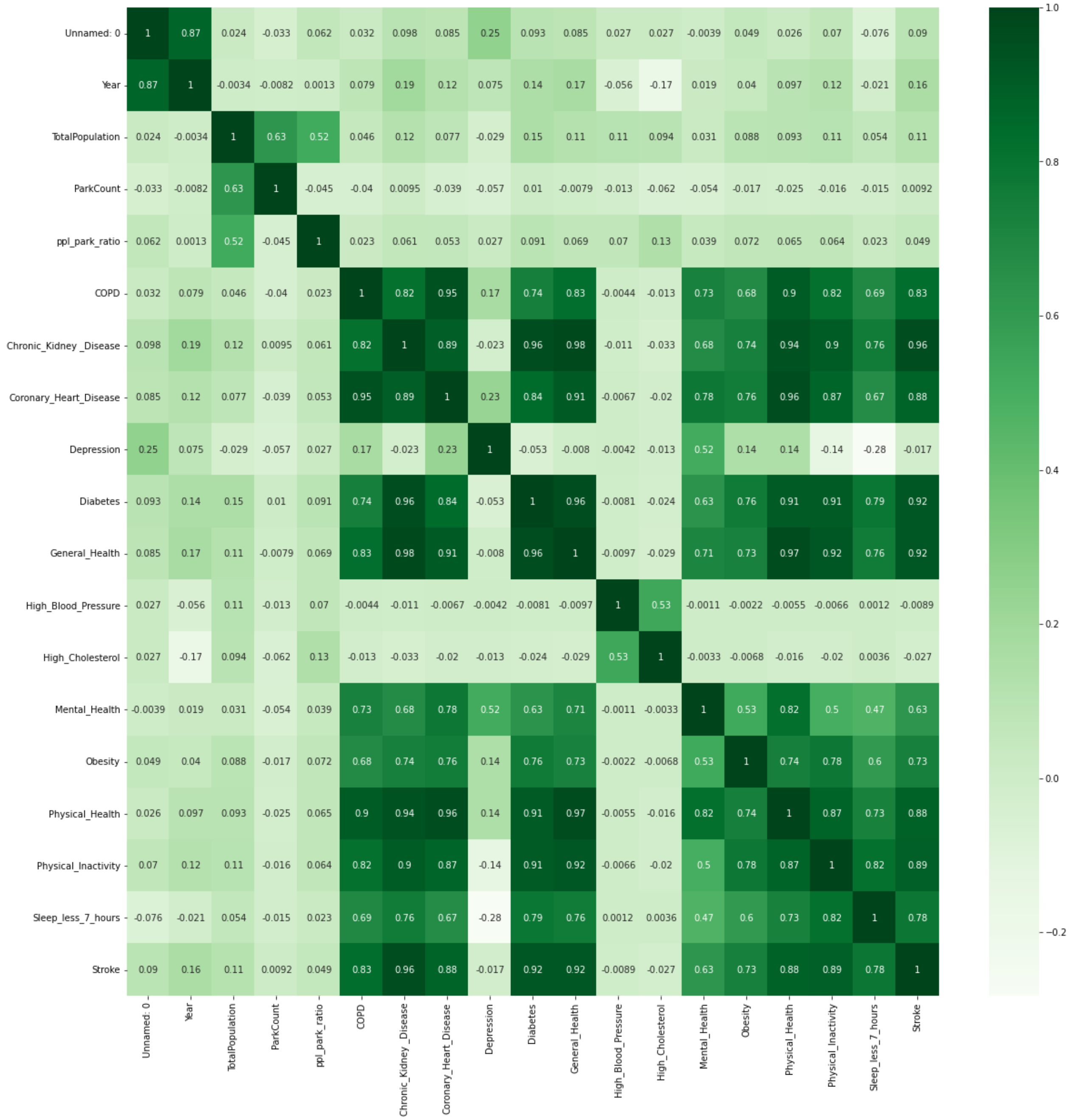
```
park_cdc = pd.read_csv('/content/drive/MyDrive/Coding/More_Park_Datasets/park_cdc.csv')
park_cdc_pivot = pd.read_csv('/content/drive/MyDrive/Coding/More_Park_Datasets/park_cdc_pivot.csv')
```

```
park_cdc_pivot.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 575 entries, 0 to 574
Data columns (total 23 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   Unnamed: 0               575 non-null    int64
 1   Year                     575 non-null    int64
 2   State_abbr               575 non-null    object
 3   State                    575 non-null    object
 4   City                     575 non-null    object
 5   TotalPopulation          575 non-null    int64
 6   County                   575 non-null    object
 7   ParkCount                575 non-null    float64
 8   ppl_park_ratio           575 non-null    float64
 9   COPD                     575 non-null    float64
 10  Chronic_Kidney _Disease  575 non-null    float64
 11  Coronary_Heart_Disease   575 non-null    float64
 12  Depression               575 non-null    float64
 13  Diabetes                 575 non-null    float64
 14  General_Health           575 non-null    float64
 15  High_Blood_Pressure      575 non-null    float64
 16  High_Cholesterol         575 non-null    float64
 17  Mental_Health            575 non-null    float64
 18  Obesity                  575 non-null    float64
 19  Physical_Health          575 non-null    float64
 20  Physical_Inactivity      575 non-null    float64
 21  Sleep_less_7_hours       575 non-null    float64
 22  Stroke                   575 non-null    float64
dtypes: float64(16), int64(3), object(4)
memory usage: 103.4+ KB
```

```
park_cdc_pivot_matrix = park_cdc_pivot.corr()
fig, ax = plt.subplots(figsize=(20, 20))
sns.heatmap(park_cdc_pivot_matrix, annot=True, cmap='Greens')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f32c2ae0790>
```




```
1 park_cdc_pivot.describe()
```

	Unnamed: 0	Year	TotalPopulation	ParkCount	ppl_park_ratio	COPD	Chronic_Kidney_Disease	Coronary_Heart_
count	575.000000	575.000000	5.750000e+02	575.000000	575.000000	575.000000	575.000000	57
min	0.000000	2019.000000	1.420000e+02	1.000000	11.833333	2.600000	1.800000	

ANOVA

```
from statsmodels.formula.api import ols
import statsmodels.api as sm

# Model for Obesity
model1Obesity = ols('Obesity ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_table1 = sm.stats.anova_lm(model1Obesity, typ=2)
#modelobe = sm.OLS(park_cdc_pivot['Obesity'], park_cdc_pivot['ppl_park_ratio']).fit()
#aov_table1b = sm.stats.anova_lm(modelobe, typ=2)
print('Obesity')
print(aov_table1)
#print(aov_table1b)

# Model for High Blood Pressure
model2High_Blood_Pressure = ols('High_Blood_Pressure ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_table2 = sm.stats.anova_lm(model2High_Blood_Pressure, typ=2)
print('High_Blood_Pressure')
print(aov_table2)

# Model for High Cholesterol
model3High_Cholesterol = ols('High_Cholesterol ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_table3 = sm.stats.anova_lm(model3High_Cholesterol, typ=2)
print('High_Cholesterol')
print(aov_table3)

# Model for Stroke
model3Stroke = ols('Stroke ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_table4= sm.stats.anova_lm(model3Stroke, typ=2)
print('Stroke')
print(aov_table4)

modeldepression= ols('Depression ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tabledepression= sm.stats.anova_lm(modeldepression, typ=2)
print('Depression')
print(aov_tabledepression)

modelsleep= ols('Sleep_less_7_hours~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tablesleep= sm.stats.anova_lm(modelsleep, typ=2)
print('Sleep')
print(aov_tablesleep)

modelDiabetes= ols('Diabetes ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tabledia= sm.stats.anova_lm(modelDiabetes, typ=2)
print('Diabetes')
print(aov_tabledia)

modelPhysical_Health= ols('Physical_Health ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_table5ph= sm.stats.anova_lm(modelPhysical_Health, typ=2)
print('Physical Health')
print(aov_table5ph)

modelPhysical_Inactivity = ols('Physical_Inactivity ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tablePhysical_Inactivity= sm.stats.anova_lm(modelPhysical_Inactivity, typ=2)
print('Physical Inactivity')
print(aov_tablePhysical_Inactivity)

modelMental_Health = ols('Mental_Health ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tablemodelMental_Health= sm.stats.anova_lm(modelMental_Health, typ=2)
print('Mental health')
print(aov_tablemodelMental_Health)

modelCOPD= ols('COPD ~ ppl_park_ratio', data=park_cdc_pivot).fit()
aov_tableCOPD= sm.stats.anova_lm(modelCOPD, typ=2)
print('COPD')
print(aov_tableCOPD)
```

```
Obesity
                   sum_sq    df         F   PR(>F)
ppl_park_ratio  43.719778   1.0  2.993646  0.08413
```

```
Residual        8368.202587  573.0       NaN      NaN
High_Blood_Pressure
                    sum_sq     df         F   PR(>F)
ppl_park_ratio    26.844297    1.0  2.809544  0.09425
Residual        5474.832225  573.0       NaN      NaN
High_Cholesterol
                    sum_sq     df          F    PR(>F)
ppl_park_ratio    29.401923    1.0  10.398478  0.001333
Residual        1620.169972  573.0        NaN       NaN
Stroke
                   sum_sq     df         F   PR(>F)
ppl_park_ratio    0.361175    1.0  1.400263  0.23717
Residual        147.795973  573.0       NaN      NaN
Depression
                    sum_sq     df         F    PR(>F)
ppl_park_ratio     1.765989    1.0  0.415242  0.519579
Residual        2436.921559  573.0       NaN       NaN
Sleep
                    sum_sq     df         F    PR(>F)
ppl_park_ratio     2.954428    1.0  0.306751  0.579897
Residual        5518.770894  573.0       NaN       NaN
Diabetes
                    sum_sq     df         F    PR(>F)
ppl_park_ratio    15.409043    1.0  4.809891  0.028698
Residual        1835.671931  573.0       NaN       NaN
Physical Health
                    sum_sq     df         F    PR(>F)
ppl_park_ratio     6.725746    1.0  2.444214  0.118511
Residual        1576.724584  573.0       NaN       NaN
Physical Inactivity
                     sum_sq     df         F    PR(>F)
ppl_park_ratio     46.506992    1.0  2.36705  0.124473
Residual        11258.107199  573.0      NaN       NaN
Mental health
                    sum_sq     df         F    PR(>F)
ppl_park_ratio     1.574723    1.0  0.895418  0.344412
Residual        1007.703973  573.0       NaN       NaN
COPD
                   sum_sq     df         F    PR(>F)
ppl_park_ratio    0.312641    1.0  0.295233  0.587098
Residual        606.786420  573.0       NaN       NaN
```

Notable Results:

High Cholesterol & Diabetes p-value are under 0.05, making the results statistically significant. The next closest are High Blood Pressure & Obesity, but not significant since they are above 0.05.

OLS

```
# Model for Obesity
model1Obesity = ols('Obesity ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(model1Obesity.summary())

# Model for High Blood Pressure
model2High_Blood_Pressure = ols('High_Blood_Pressure ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(model2High_Blood_Pressure.summary())

# Model for High Cholesterol
model3High_Cholesterol = ols('High_Cholesterol ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(model3High_Cholesterol.summary())

# Model for Stroke
model3Stroke = ols('Stroke ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(model3Stroke.summary())

modeldepression= ols('Depression ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modeldepression.summary())

modelSleep_less_7_hours= ols('Sleep_less_7_hours~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelSleep_less_7_hours.summary())

modelPhysical_Inactivity = ols('Physical_Inactivity ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelPhysical_Inactivity.summary())

modelDiabetes= ols('Diabetes ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelDiabetes.summary())

```

```
modelPhysical_Health= ols('Physical_Health ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelPhysical_Health.summary())

modelPhysical_Inactivity = ols('Physical_Inactivity ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelPhysical_Inactivity.summary())

modelMental_Health = ols('Mental_Health ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelMental_Health.summary())

modelCOPD= ols('COPD ~ ppl_park_ratio', data=park_cdc_pivot).fit()
print(modelCOPD.summary())
```

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.38e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
                            OLS Regression Results
==============================================================================
Dep. Variable:          Mental_Health   R-squared:                       0.002
Model:                            OLS   Adj. R-squared:                 -0.000
Method:                 Least Squares   F-statistic:                    0.8954
Date:                Fri, 10 Feb 2023   Prob (F-statistic):              0.344
Time:                        06:28:53   Log-Likelihood:                -977.19
No. Observations:                 575   AIC:                             1958.
Df Residuals:                     573   BIC:                             1967.
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept         13.6664      0.056    242.956      0.000      13.556      13.777
ppl_park_ratio  1.602e-06   1.69e-06      0.946      0.344   -1.72e-06    4.93e-06
==============================================================================
Omnibus:                       54.354   Durbin-Watson:                   1.068
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              231.960
Skew:                           0.288   Prob(JB):                     4.27e-51
Kurtosis:                       6.058   Cond. No.                     3.38e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.38e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
                            OLS Regression Results
==============================================================================
Dep. Variable:                   COPD   R-squared:                       0.001
Model:                            OLS   Adj. R-squared:                 -0.001
Method:                 Least Squares   F-statistic:                    0.2952
Date:                Fri, 10 Feb 2023   Prob (F-statistic):              0.587
Time:                        06:28:53   Log-Likelihood:                -831.36
No. Observations:                 575   AIC:                             1667.
Df Residuals:                     573   BIC:                             1675.
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
Intercept          5.0779      0.044    116.335      0.000       4.992       5.164
ppl_park_ratio  7.137e-07   1.31e-06      0.543      0.587   -1.87e-06    3.29e-06
==============================================================================
Omnibus:                      279.219   Durbin-Watson:                   1.401
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             3528.179
Skew:                           1.811   Prob(JB):                         0.00
Kurtosis:                      14.582   Cond. No.                     3.38e+04
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.38e+04. This might indicate that there are
strong multicollinearity or other numerical problems.
```

Notable Results:

Diabetes & High Cholesterol have p-values below 0.05. With R-squared values being 0.018 or less, they seem to have a smaller impact over all.

```
# Fit the linear regression model
modelHC = sm.OLS(park_cdc_pivot['High_Cholesterol'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modelHC.summary())

# Fit the linear regression model
modelHBP = sm.OLS(park_cdc_pivot['High_Blood_Pressure'], park_cdc_pivot['ppl_park_ratio']).fit()
```

```
print(modelHBP.summary())

# Fit the linear regression model
modelS = sm.OLS(park_cdc_pivot['Stroke'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modelS.summary())

# Fit the linear regression model
modelPI = sm.OLS(park_cdc_pivot['Physical_Inactivity'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modelPI.summary())

# Fit the linear regression model
modelobe = sm.OLS(park_cdc_pivot['Obesity'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modelobe.summary())

# Fit the linear regression model
modeldia = sm.OLS(park_cdc_pivot['Diabetes'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modeldia.summary())

# Fit the linear regression model
modeldep = sm.OLS(park_cdc_pivot['Depression'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modeldep.summary())

# Fit the linear regression model
modelCHD = sm.OLS(park_cdc_pivot['Coronary_Heart_Disease'], park_cdc_pivot['ppl_park_ratio']).fit()
print(modelCHD.summary())
```

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:       High_Cholesterol   R-squared (uncentered):                   0.036
Model:                            OLS   Adj. R-squared (uncentered):              0.035
Method:                 Least Squares   F-statistic:                              21.55
Date:                Fri, 10 Feb 2023   Prob (F-statistic):                    4.27e-06
Time:                        06:33:56   Log-Likelihood:                         -2724.0
No. Observations:                 575   AIC:                                      5450.
Df Residuals:                     574   BIC:                                      5454.
Df Model:                           1
Covariance Type:            nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
ppl_park_ratio     0.0002   3.47e-05      4.643      0.000    9.29e-05       0.000
==============================================================================
Omnibus:                     1096.689   Durbin-Watson:                   0.072
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           971454.093
Skew:                         -13.068   Prob(JB):                         0.00
Kurtosis:                     202.661   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                                 OLS Regression Results
=======================================================================================
Dep. Variable:     High_Blood_Pressure   R-squared (uncentered):                  0.036
Model:                             OLS   Adj. R-squared (uncentered):             0.034
Method:                  Least Squares   F-statistic:                             21.28
Date:                 Fri, 10 Feb 2023   Prob (F-statistic):                   4.91e-06
Time:                         06:33:56   Log-Likelihood:                        -2730.4
No. Observations:                  575   AIC:                                     5463.
Df Residuals:                      574   BIC:                                     5467.
Df Model:                            1
Covariance Type:             nonrobust
==================================================================================
                     coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------
ppl_park_ratio     0.0002   3.51e-05      4.613      0.000    9.29e-05       0.000
==============================================================================
Omnibus:                      913.242   Durbin-Watson:                   0.085
Prob(Omnibus):                  0.000   Jarque-Bera (JB):           390912.531
Skew:                          -9.068   Prob(JB):                         0.00
Kurtosis:                     129.441   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                 Stroke   R-squared (uncentered):                   0.036
Model:                            OLS   Adj. R-squared (uncentered):              0.034
Method:                 Least Squares   F-statistic:                              21.19
Date:                Fri, 10 Feb 2023   Prob (F-statistic):                    5.12e-06
Time:                        06:33:56   Log-Likelihood:                         -1339.8
```

```
No. Observations:                 575   AIC:                              2682.
```

Explain Analysis

Interesting results when using another algorithm for OLS. All the p-values became less than 0.06. I would note that if these are statistically significantly - I would say they only affect the overall health ailments a small percentage.

The results presented are from an analysis of the relationship between the ratio of people to park area and various health outcomes. The analysis includes both an ANOVA test and an OLS Linear Regression model.

ANOVA Results:

The ANOVA results show the sum of squares, degrees of freedom, F-statistic, and the p-value for each health outcome. The F-statistic measures the ratio of explained variation to unexplained variation and is used to test the hypothesis that the population means are equal for all groups. The p-value is the probability of observing a test statistic as extreme or more extreme than the one observed, under the assumption that the null hypothesis is true.

The results indicate that for some health outcomes, there is a significant relationship between the ratio of people to park area and the outcome. For example, for High Cholesterol, the p-value is 0.001333, which is less than the commonly used significance level of 0.05, suggesting that the relationship between the ratio of people to park area and High Cholesterol is statistically significant.

However, for other health outcomes such as Stroke, Depression, and Sleep, the p-value is larger than 0.05, suggesting that there is not a statistically significant relationship between the ratio of people to park area and these health outcomes.

OLS Linear Regression Results:

The OLS Linear Regression results provide additional information about the strength and direction of the relationship between the ratio of people to park area and the health outcomes. The coefficients represent the change in the outcome for a unit change in the predictor variable. The p-values for the coefficients indicate the level of significance of each predictor variable in the model.

Based on the results, it appears that the ratio of people to park area has a positive and statistically significant relationship with some health outcomes such as Diabetes, Physical Health, and Physical Inactivity, while it has a negative but statistically insignificant relationship with other health outcomes such as Mental Health and COPD.

In conclusion, the results suggest that there is a complex relationship between the ratio of people to park area and various health outcomes, and that some health outcomes may be more influenced by this factor than others. Further research is needed to fully understand these relationships and their potential implications for public health.

Summary

The results of the OLS regression and ANOVA analysis suggest that the relationship between "ppl_park_ratio" and various health outcomes is complex and not consistently significant across all outcomes.

Starting with the OLS regression results, the R-squared value provides an indication of the proportion of variation in the dependent variable that is explained by the independent variable(s). A high R-squared value (close to 1) indicates that the independent variable is a good predictor of the dependent variable, while a low R-squared value (close to 0) indicates the opposite.

The coefficients in the regression equation represent the change in the dependent variable for a unit change in the independent variable while controlling for the other variables. Positive coefficients indicate a positive relationship between the independent and dependent variables, while negative coefficients indicate a negative relationship. The p-values associated with each coefficient indicate the significance of the relationship, with smaller p-values indicating stronger evidence against the null hypothesis that the coefficient is equal to zero.

In the ANOVA results, the F-statistic measures the strength of the relationship between the independent variable and the dependent variable. The p-value associated with the F-statistic tests the hypothesis that the independent variable has no effect on the dependent variable. Smaller p-values indicate stronger evidence against the null hypothesis and support the conclusion that the independent variable has an effect on the dependent variable.

Overall, the results of the OLS regression and ANOVA analysis indicate that "ppl_park_ratio" has a significant impact on High Cholesterol, Diabetes and Physical Inactivity but no impact on other health outcomes like Depression, Sleep, Mental health, COPD and so on.

It is important to keep in mind that these results are based on a sample of the population and may not generalize to the population as a whole. In addition, the results are based on a number of assumptions and limitations, including linearity, independence of observations, homoscedasticity, and normality of residuals, among others. Further research is needed to confirm these findings and to explore the relationship between "ppl_park_ratio" and health outcomes in more detail.