

DATA ANALYSIS: HAPPINESS DETERMINANTS PER COUNTRY - 2023

The 2023 data resulted from the merger of the 3 csv files below:

1. Crime.csv (Crime Rate by Country 2023 retrieved from <https://worldpopulationreview.com/country-rankings/crime-rate-by-country> (<https://worldpopulationreview.com/country-rankings/crime-rate-by-country>)),
2. Happiness.csv (World Happiness Report 2023 retrieved from <https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023> (<https://www.kaggle.com/datasets/ajaypalsinghlo/world-happiness-report-2023>)), and
3. Freedom.csv (2023 Heritage Foundation Economic Freedom data retrieved from <https://www.heritage.org/index/explore?view=by-region-country-year&countryids=®ionids=&yearids=24> (<https://www.heritage.org/index/explore?view=by-region-country-year&countryids=®ionids=&yearids=24>)).

The .csv files were imported into SQL where the data was housed and merged to produce one database and a corresponding.csv file with 111 unique records: Happiness_Factors.sql and Happiness_Factors.csv

The analysis consisted of Scatterplots, Multivariate Regressions, and Interactive Bar Charts using the following variables:

- y variable: Happiness Score
- x variables: GDP per Capita, Social Support, Healthy Life Expectancy, Corruption, Freedom_Score, Property_Rights, Government_Spending, Business, Trade, and Crime_Index

DATA SETUP

```
In [1]: # Import Dependencies
import csv
import pandas as pd
import statsmodels.api as sm
import matplotlib.pyplot as plt
```

```
In [2]: # Read the CSV file and create pandas dataframe
df = pd.read_csv('Happiness_Factors.csv')

# Print the dataframe
df.head()
```

Out[2]:

	Country	Happiness_Score	GDP_per_capita	Social_support	Healthy_life_expectancy	Corruption	Id	Freedom_Score	Pro
0	Afghanistan	1.859	7.324	0.341	54.712	0.847	1	0.0	
1	Albania	5.277	9.567	0.718	69.150	0.878	2	65.3	
2	Algeria	5.329	9.300	0.855	66.549	0.717	3	43.2	
3	Argentina	6.024	9.959	0.891	67.200	0.814	5	51.0	
4	Armenia	5.342	9.615	0.790	67.789	0.705	6	65.1	

The link below leads to a Github page with a Line Graph of Happiness Scores by Country:

https://jcndongo.github.io/happinessline.github.io/Happiness_LINE_index.html
(https://jcndongo.github.io/happinessline.github.io/Happiness_LINE_index.html)

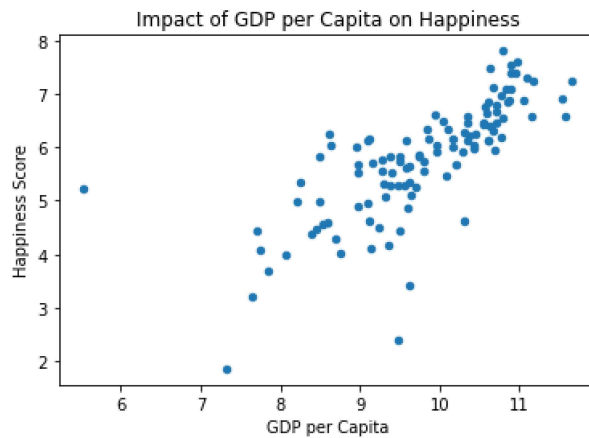
SCATTERPLOTS

The Section Below uses Scatterplots to Illustrate the Impact of GDP per Capita, Social Support, Healthy Life Expectancy, Corruption, Freedom_Score, Property_Rights, Government_Spending, Business, Trade, and Crime_Index on Happiness Scores

```
In [3]: # Create a scatter plot: Happiness & GDP per Capita
df.plot.scatter(x='GDP_per_capita', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('GDP per Capita')
plt.ylabel('Happiness Score')
plt.title('Impact of GDP per Capita on Happiness')

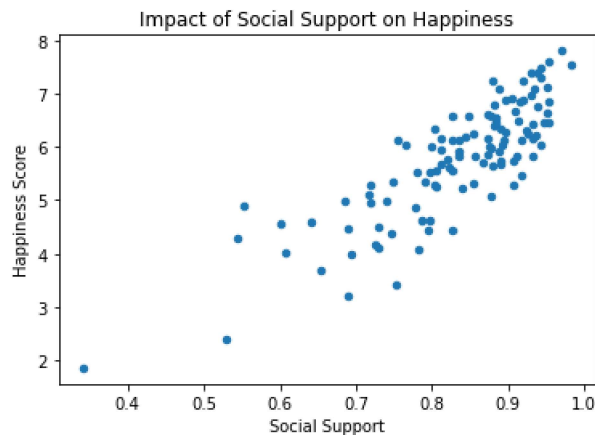
# Display the plot
plt.show()
```



```
In [4]: # Create a scatter plot: Happiness & Social Support
df.plot.scatter(x='Social_support', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Social Support')
plt.ylabel('Happiness Score')
plt.title('Impact of Social Support on Happiness')

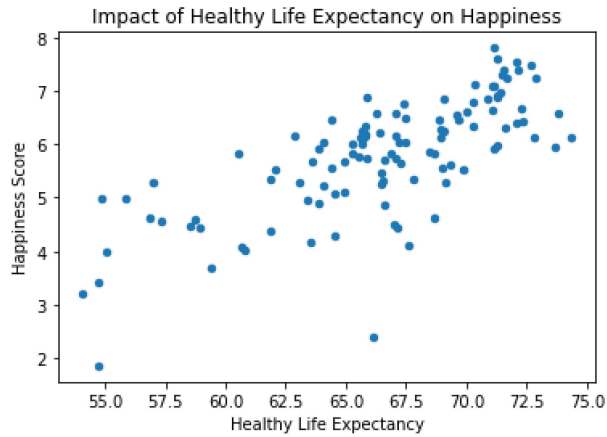
# Display the plot
plt.show()
```



```
In [5]: # Create a scatter plot: Happiness & Healthy Life Expectancy
df.plot.scatter(x='Healthy_life_expectancy', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Healthy Life Expectancy')
plt.ylabel('Happiness Score')
plt.title('Impact of Healthy Life Expectancy on Happiness')

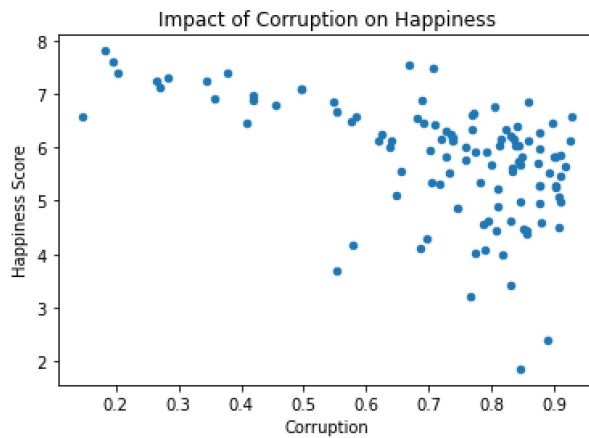
# Display the plot
plt.show()
```



```
In [6]: # Create a scatter plot: Happiness & Corruption
df.plot.scatter(x='Corruption', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Corruption')
plt.ylabel('Happiness Score')
plt.title('Impact of Corruption on Happiness')

# Display the plot
plt.show()
```

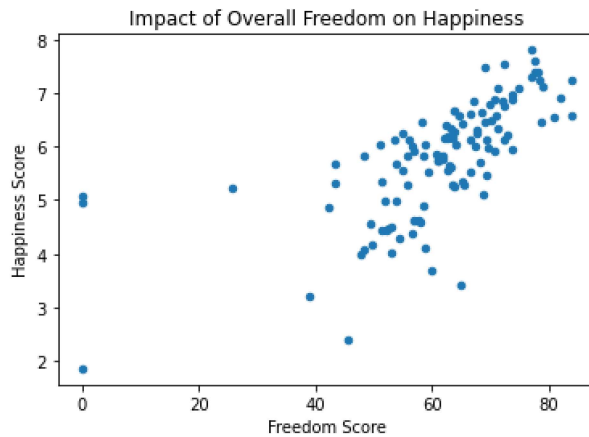


```
In [7]: # Create a scatter plot: Happiness & Overall Freedom

df.plot.scatter(x='Freedom_Score', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Freedom Score')
plt.ylabel('Happiness Score')
plt.title('Impact of Overall Freedom on Happiness')

# Display the plot
plt.show()
```

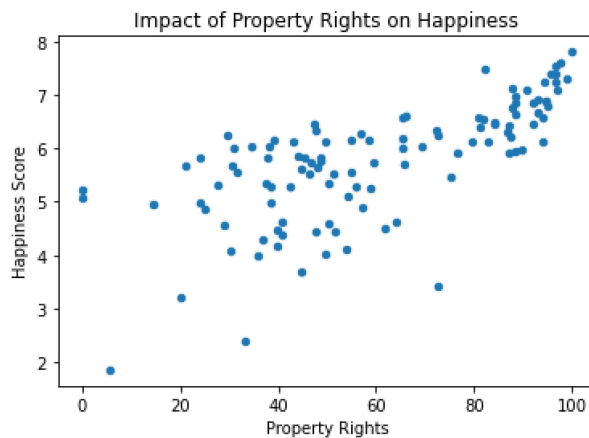


```
In [8]: # Create a scatter plot: Happiness & Property Rights

df.plot.scatter(x='Property_Rights', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Property Rights')
plt.ylabel('Happiness Score')
plt.title('Impact of Property Rights on Happiness')

# Display the plot
plt.show()
```

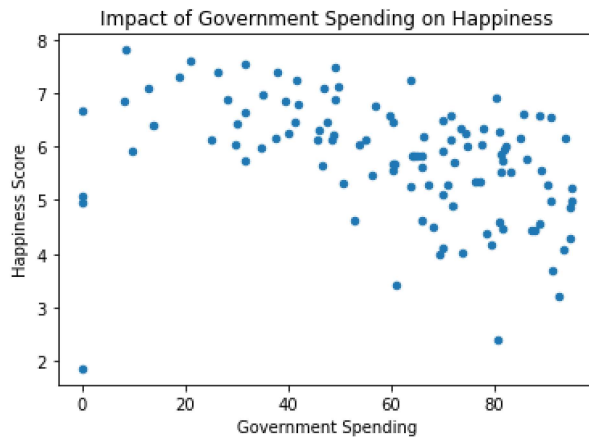


```
In [9]: # Create a scatter plot: Happiness & Government Spending

df.plot.scatter(x='Government_Spending', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Government Spending')
plt.ylabel('Happiness Score')
plt.title('Impact of Government Spending on Happiness')

# Display the plot
plt.show()
```

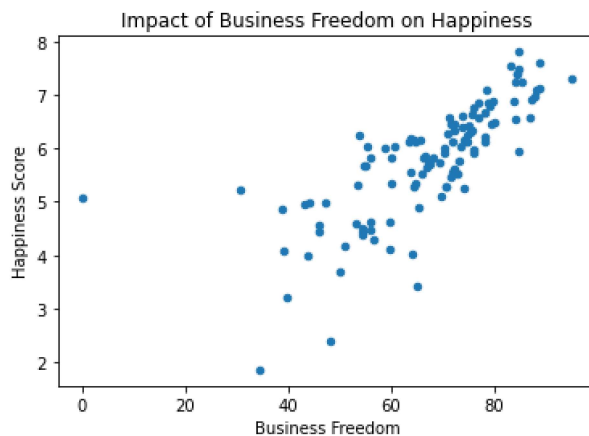


```
In [10]: # Create a scatter plot: Happiness & Business Freedom

df.plot.scatter(x='Business_Freedom', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Business Freedom')
plt.ylabel('Happiness Score')
plt.title('Impact of Business Freedom on Happiness')

# Display the plot
plt.show()
```

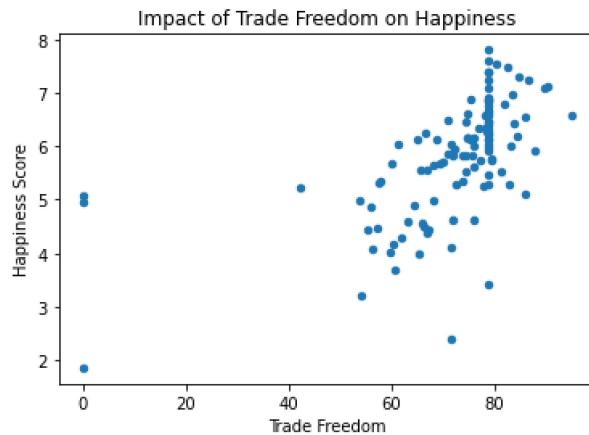


```
In [11]: # Create a scatter plot: Happiness & Trade Freedom

df.plot.scatter(x='Trade', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Trade Freedom')
plt.ylabel('Happiness Score')
plt.title('Impact of Trade Freedom on Happiness')

# Display the plot
plt.show()
```

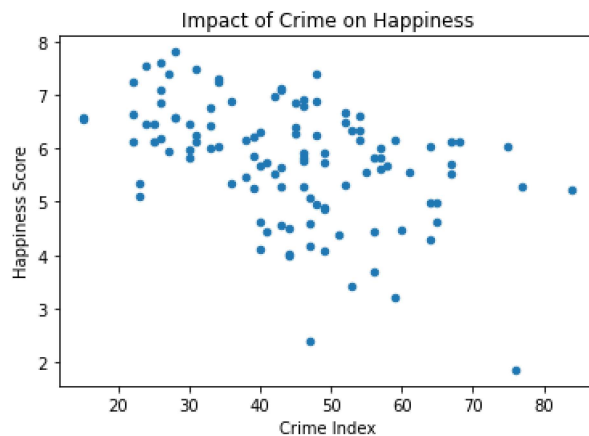


```
In [12]: # Create a scatter plot: Happiness & Crime

df.plot.scatter(x='Crime_Index', y='Happiness_Score')

# Set the axis labels and plot title
plt.xlabel('Crime Index')
plt.ylabel('Happiness Score')
plt.title('Impact of Crime on Happiness')

# Display the plot
plt.show()
```



MULTIVARIATE REGRESSION ANALYSIS

The Section below represents a multivariate regression analysis of the Impact of GDP per Capita, Social Support, Healthy Life Expectancy, Corruption, Freedom_Score, Property_Rights, Government_Spending, Business, Trade, Investment, Crime_Index, and Population on Happiness Scores

```
In [13]: # Regression analysis to test model

# Define the dependent variable (y) and independent variables (x)
y = df['Happiness_Score']
x = df[['GDP_per_capita', 'Social_support', 'Healthy_life_expectancy', 'Corruption', 'Freedom_Score',
'Property_Rights', 'Government_Spending', 'Business', 'Trade', 'Investment', 'Crime_Index', 'Population']]

# Add a constant term to the independent variables
x = sm.add_constant(x)

# Perform the multivariate regression analysis
model = sm.OLS(y, x).fit()

# Print the summary statistics of the regression model
print(model.summary())
```

```

=====
                        OLS Regression Results
=====
Dep. Variable:      Happiness_Score  R-squared:      0.850
Model:              OLS              Adj. R-squared:  0.831
Method:             Least Squares    F-statistic:    46.23
Date:               Tue, 30 May 2023  Prob (F-statistic): 6.33e-35
Time:               16:42:50         Log-Likelihood:  -61.850
No. Observations:   111             AIC:            149.7
Df Residuals:       98              BIC:            184.9
Df Model:           12
Covariance Type:    nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	-2.2913	1.102	-2.080	0.040	-4.478	-0.105
GDP_per_capita	0.0597	0.094	0.637	0.526	-0.126	0.246
Social_support	5.6560	0.596	9.489	0.000	4.473	6.839
Healthy_life_expectancy	0.0325	0.015	2.168	0.033	0.003	0.062
Corruption	-0.7455	0.292	-2.549	0.012	-1.326	-0.165
Freedom_Score	0.0124	0.015	0.811	0.419	-0.018	0.043
Property_Rights	-0.0101	0.005	-1.988	0.050	-0.020	-1.87e-05
Government_Spending	-0.0018	0.003	-0.563	0.575	-0.008	0.005
Business	0.0291	0.008	3.431	0.001	0.012	0.046
Trade	-0.0167	0.008	-2.052	0.043	-0.033	-0.001
Investment	0.0044	0.004	1.067	0.289	-0.004	0.013
Crime_Index	0.0018	0.004	0.437	0.663	-0.006	0.010
Population	9.045e-11	2.41e-10	0.375	0.708	-3.88e-10	5.69e-10

```

=====
Omnibus:      0.060  Durbin-Watson:      1.990
Prob(Omnibus): 0.970  Jarque-Bera (JB):      0.040
Skew:         -0.038  Prob(JB):              0.980
Kurtosis:     2.944  Cond. No.              5.31e+09
=====

```

Notes:

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

[2] The condition number is large, 5.31e+09. This might indicate that there are strong multicollinearity or other numerical problems.

CONCLUSION: based on the model above, social support, corruption, business freedom, and trade freedom are the only statistically significant determinants of happiness for all the countries in our dataset.

TABLES & INTERACTIVE BAR CHARTS: Social Support and Corruption

Bottom 6 - The table illustrates the Bottom 6 Countries based on Happiness Score and the link leads to a Github page with an interactive bar chart of social support and corruption for those countries:

https://jcndongo.github.io/bottom6countries.github.io/BOTTOM6_index.html

(https://jcndongo.github.io/bottom6countries.github.io/BOTTOM6_index.html)

```
In [14]: # Print df by ascending Happiness Score - Least Happy
Bottom6 = df.sort_values('Happiness_Score', ascending=True)

# Print the sorted dataframe
Bottom6.head(6)
```

Out[14]:

	Country	Happiness_Score	GDP_per_capita	Social_support	Healthy_life_expectancy	Corruption	Id	Freedom_Score
0	Afghanistan	1.859	7.324	0.341	54.712	0.847	1	0.0
57	Lebanon	2.392	9.478	0.530	66.149	0.891	93	45.6
110	Zimbabwe	3.204	7.641	0.690	54.050	0.766	183	39.0
12	Botswana	3.435	9.629	0.753	54.725	0.830	21	64.9
96	Tanzania	3.694	7.857	0.653	59.401	0.554	162	60.0
109	Zambia	3.982	8.074	0.694	55.032	0.818	182	47.8

Top 6 - The table illustrates the Top 6 Countries based on Happiness Score and the link leads to a Github page with an interactive bar chart of social support and corruption for those countries:

https://jcndongo.github.io/top6countries.github.io/TOP6_index.html (https://jcndongo.github.io/top6countries.github.io/TOP6_index.html)

```
In [15]: # Print df by descending Happiness Score - Most Happy
Top6 = df.sort_values('Happiness_Score', ascending=False)

# Print the sorted dataframe
Top6.head(6)
```

Out[15]:

	Country	Happiness_Score	GDP_per_capita	Social_support	Healthy_life_expectancy	Corruption	Id	Freedom_Score	F
32	Finland	7.804	10.792	0.969	71.150	0.182	57	77.1	
25	Denmark	7.586	10.962	0.954	71.250	0.196	45	77.6	
41	Iceland	7.530	10.896	0.983	72.050	0.668	73	72.2	
47	Israel	7.473	10.639	0.943	72.697	0.708	79	68.9	
71	Netherlands	7.403	10.942	0.930	71.550	0.379	119	78.0	
93	Sweden	7.395	10.883	0.939	72.150	0.202	157	77.5	

In []: