Analyzing the Relationship Between GDP and Happiness in 2021

Introduction

This analysis explores the relationship between **GDP per Capita** and **Happiness Scores** across countries in 2021. The data is sourced from the **World Bank** (GDP per capita) and the **World Happiness Report** (Happiness scores). By visualizing this relationship, we can identify trends and potentially uncover insights about how economic performance impacts the well-being of citizens worldwide.

Data Sources

1. **World Bank GDP Data**: The dataset contains GDP per capita values for countries from various years. For this analysis, we focused on **2021** data.
2. **World Happiness Report**: The report ranks countries based on subjective well-being measures, with the **Ladder score** representing happiness.

```
In [9]:    import pandas as pd
           import matplotlib.pyplot as plt
           import seaborn as sns
```

```
In [11]:   happiness = pd.read_excel("Happiness.xlsx")
           gdp = pd.read_csv("GDP.csv", skiprows=4)
```

```
In [13]:   happiness_clean = happiness[['Country name', 'Ladder score']]
           happiness_clean = happiness_clean.rename(columns={'Country name': 'Country', 'Ladder score': 'Happiness Score'})
```

```
In [15]:   gdp_clean = gdp[['Country Name', '2021']]
           gdp_clean = gdp_clean.rename(columns={'Country Name': 'Country', '2021': 'GDP per Capita'})
           gdp_clean = gdp_clean.dropna()
```

```
In [17]:   merged_df = pd.merge(happiness_clean, gdp_clean, on='Country', how='inner')
           merged_df = merged_df.dropna()
```

Data Cleaning and Preparation

Before merging the datasets, we performed the following cleaning steps:

- **GDP Data**: We kept only the 2021 GDP per capita values, removed unnecessary rows, and renamed columns for clarity.
- **Happiness Data**: We kept only the country names and happiness scores, then renamed the columns for consistency with the GDP data.

Finally, both datasets were merged on the **Country** column, resulting in a single dataset with the necessary information for analysis.

```
In [19]:   plt.figure(figsize=(10,6))
           sns.scatterplot(data=merged_df, x='GDP per Capita', y='Happiness Score')
           sns.regplot(data=merged_df, x='GDP per Capita', y='Happiness Score', scatter=False, color='red')

           plt.title("Happiness vs GDP per Capita (2021)")
           plt.xlabel("GDP per Capita (USD)")
           plt.ylabel("Happiness Score")
           plt.grid(True)

           plt.show()
```
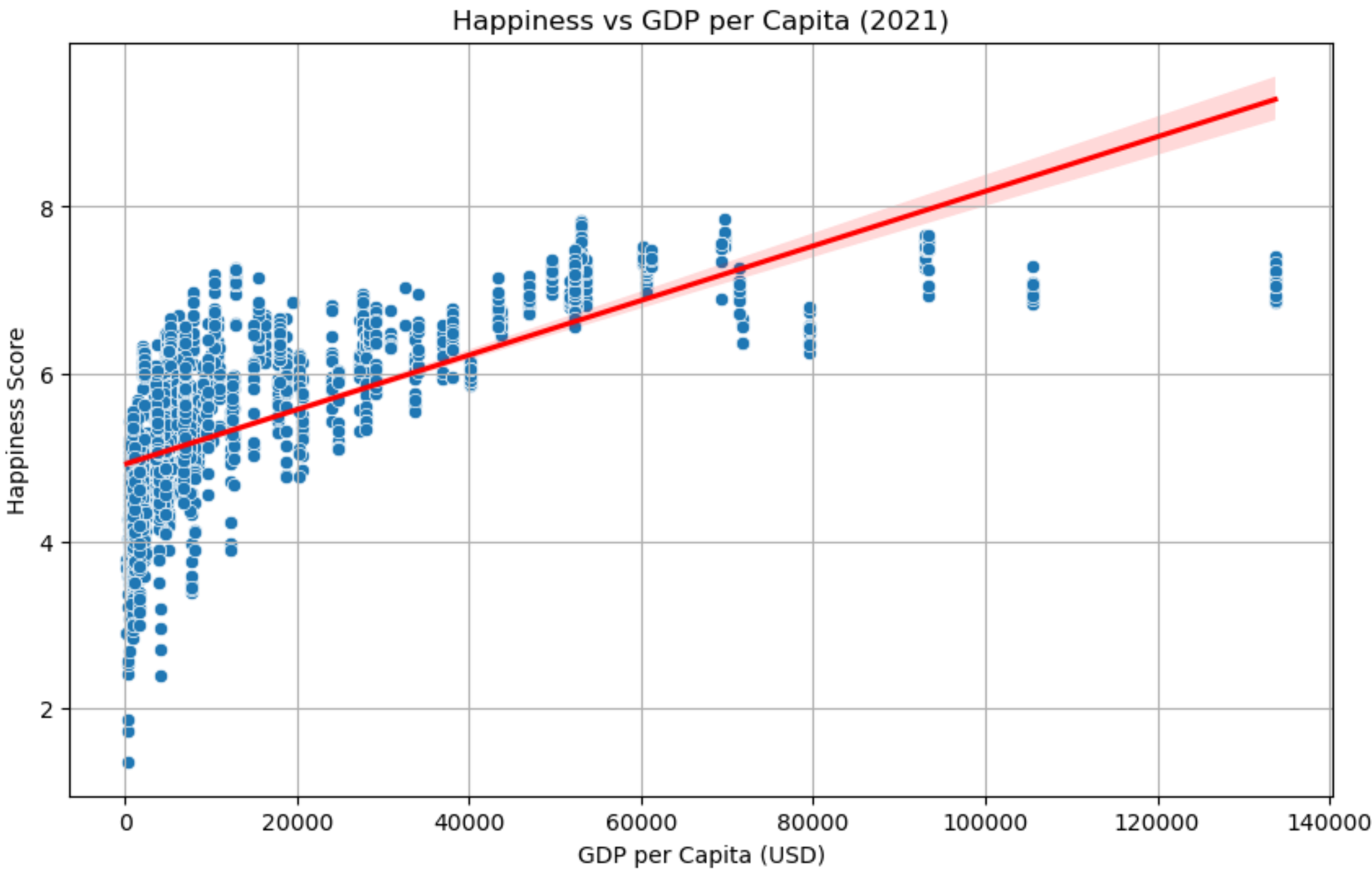

Key Insights

- **Positive Correlation**: Our scatter plot shows a positive correlation between GDP per capita and happiness scores. This suggests that wealthier countries tend to report higher happiness levels.
- **Outliers**: Some countries like **Finland** and **Denmark** have high happiness scores despite not being the wealthiest, which suggests other factors (such as social support, healthcare, and education) may play a key role in happiness.
- **Data Gaps**: A few countries are missing data for either GDP or happiness scores. Future work could focus on obtaining missing data or performing imputation for a more complete analysis.

Conclusion

This analysis demonstrates a clear link between economic wealth and happiness, but it also suggests that GDP is not the only factor influencing well-being. Further research could investigate other indicators that contribute to happiness.