

Introducción

El objetivo de este trabajo es analizar la relación entre la felicidad y diversas variables socioeconómicas mediante modelos de regresión lineal simple y múltiple, con el fin de comparar su capacidad explicativa y comprender la importancia de considerar múltiples dimensiones del bienestar.

Descripción de datos

Para el análisis se utilizó una base de datos obtenida del *World Happiness Report*, la cual contiene información de 154 países.

Las variables incluidas fueron:

- Happiness: nivel promedio de felicidad del país (variable dependiente).
- GDP per capita: indicador económico.
- Family: apoyo social.
- Health (Life Expectancy): esperanza de vida.
- Trust (Government Corruption): percepción de corrupción gubernamental.

Preparación y limpieza de los datos

Antes de realizar el análisis estadístico, fue necesario preparar la base de datos con el fin de asegurar la calidad de la información.

Código utilizado:

```
df = df.rename(columns={  
    "Happiness.Score": "Happiness",  
    "Economy..GDP.per.Capita.": "GDP",  
    "Family": "Family",  
    "Health..Life.Expectancy.": "Health",  
    "Trust..Government.Corruption.": "Trust"
```

)

Este paso permitió trabajar con nombres de variables más claros y facilitar la lectura del análisis.

Posteriormente, se identificaron valores iguales a cero, los cuales representan datos faltantes y no valores reales. Por esta razón, se decidió trabajar únicamente con los países que contaban con información completa:

```
df_clean = df[  
    (df["GDP"] > 0) &  
    (df["Health"] > 0) &  
    (df["Family"] > 0)  
]
```

Esta decisión redujo el tamaño de la muestra, pero permitió garantizar la consistencia de los resultados.

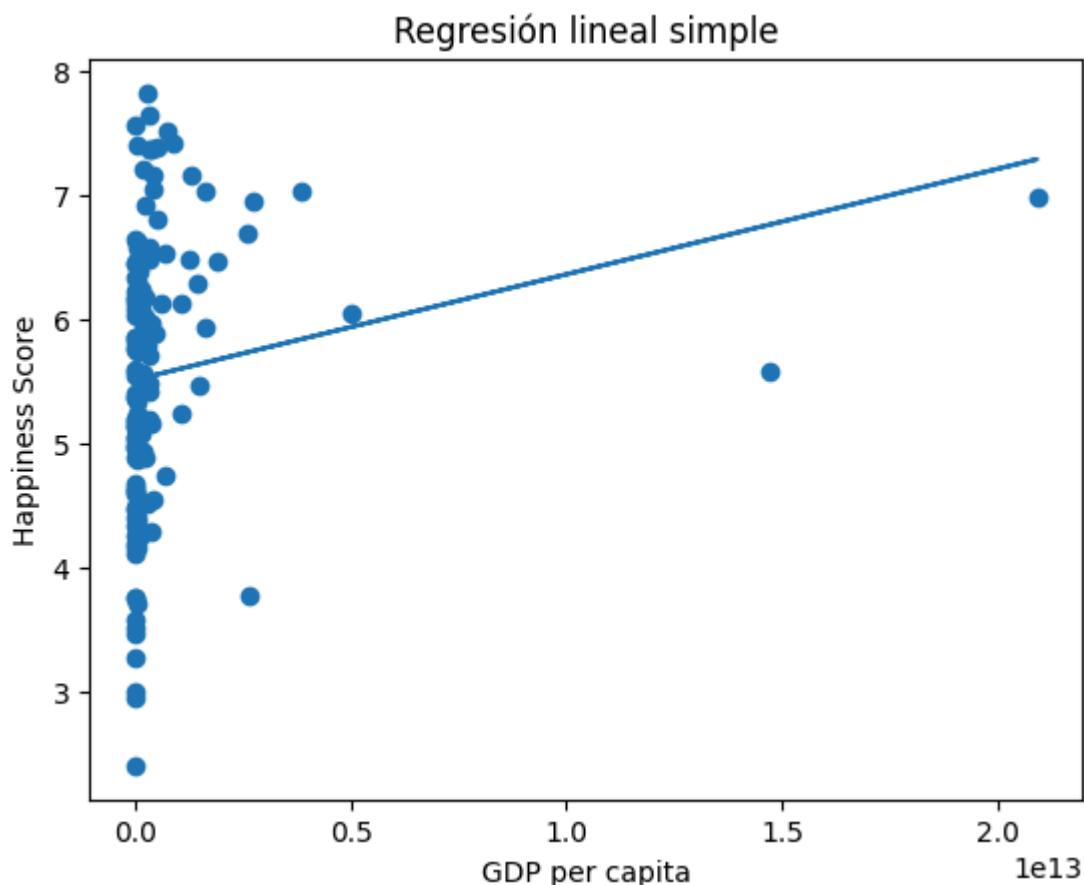
Modelo de regresión lineal simple

Como primer acercamiento, se construyó un modelo de regresión lineal simple utilizando únicamente el GDP per cápita como variable explicativa, sin aplicar ninguna transformación.

Código del modelo simple:

```
X_simple = sm.add_constant(df_clean["GDP"])  
  
modelo_simple = sm.OLS(df_clean["Happiness"], X_simple).fit()  
  
modelo_simple.summary()
```

Este modelo permite analizar si el ingreso económico, por sí solo, es suficiente para explicar las diferencias en el nivel de felicidad entre países.



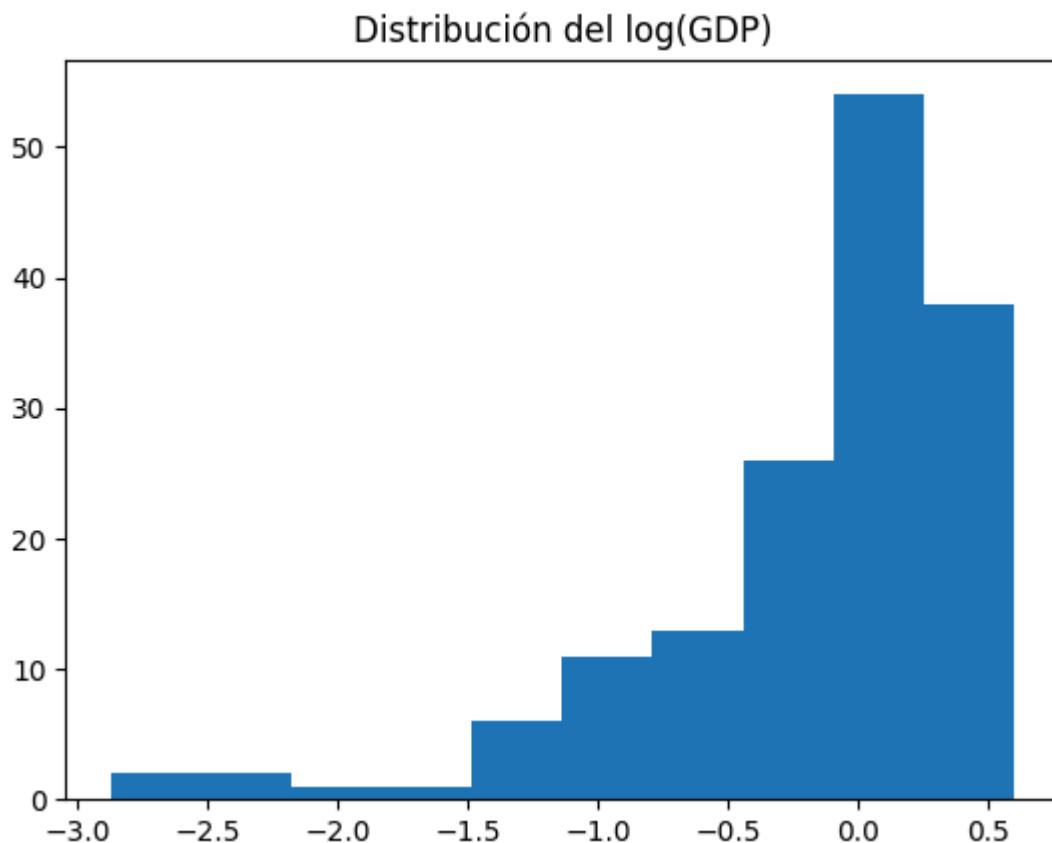
La gráfica de dispersión muestra una relación positiva entre el GDP per cápita y el nivel de felicidad. Sin embargo, se observa una dispersión considerable, lo cual indica que el ingreso económico no explica completamente las diferencias en felicidad, sugiriendo la influencia de otros factores adicionales.

Transformación del GDP para el modelo múltiple

Durante la exploración de los datos se observó que el GDP per cápita presenta una distribución altamente asimétrica, debido a las grandes diferencias económicas entre países. Por esta razón, se aplicó una transformación logarítmica únicamente para el modelo de regresión múltiple.

```
df_clean["log_GDP"] = np.log(df_clean["GDP"])
```

La transformación logarítmica permite reducir la influencia de valores extremos y analizar una relación más cercana a la linealidad, lo cual mejora el comportamiento estadístico del modelo.



La distribución del log(GDP) muestra una forma más equilibrada que la variable original, lo que justifica el uso de esta transformación para el análisis multivariable.

Modelo de regresión lineal múltiple

Con el objetivo de construir un análisis más realista, se desarrolló un modelo de regresión lineal múltiple incorporando variables económicas, sociales, de salud e institucionales.

Código del modelo múltiple:

```
X = df_clean[["log_GDP", "Family", "Health", "Trust"]]
```

```
X = sm.add_constant(X)
```

```
modelo_multiple = sm.OLS(df_clean["Happiness"], X).fit()
```

```
modelo_multiple.summary()
```

Este modelo permite evaluar el efecto conjunto de distintas dimensiones del bienestar sobre el nivel de felicidad.

Resultados del modelo de regresión múltiple

Tabla 1. Coeficientes estimados

Variable	Coeficiente	Valor p	Interpretación
Intercepto	2.6969	0.000	Nivel base de felicidad
log(GDP)	0.2706	0.018	Relación económica positiva
Family	1.6770	0.000	Fuerte influencia del apoyo social
Health	2.0397	0.000	Alta influencia del bienestar en salud
Trust	1.8167	0.000	Importancia institucional

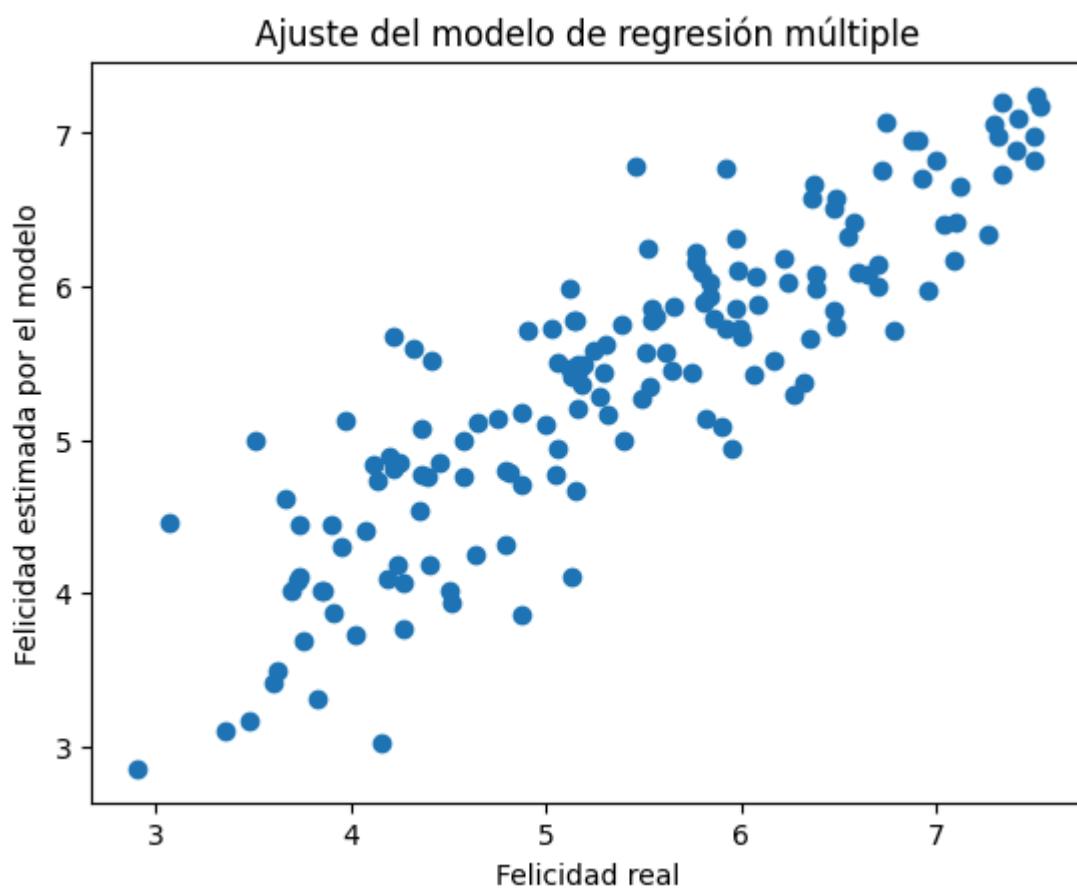
El modelo presenta un $R^2 = 0.770$, lo que indica que aproximadamente el 77% de la variabilidad del nivel de felicidad es explicada por las variables incluidas.

Evaluación del ajuste del modelo

Para evaluar el ajuste del modelo múltiple se utilizaron los valores estimados obtenidos a partir del propio modelo:

```
y_est = modelo_multiple.fittedvalues
```

Estos valores corresponden a los niveles de felicidad estimados por el modelo para cada país, y no representan predicciones futuras, sino una evaluación del ajuste del modelo sobre los datos observados.



La concentración de los puntos alrededor de la línea diagonal indica que el modelo de regresión múltiple presenta un buen ajuste general y una mayor capacidad explicativa en comparación con el modelo simple.

Uso de inteligencia artificial como apoyo académico

El desarrollo del presente trabajo contó con el apoyo de herramientas de inteligencia artificial como recurso complementario en el desarrollo, para entender el uso de logaritmos, porque y como esto ayuda con la implementación. Además de la estimación de valores para la creación de la representación gráfica.

Conclusiones

Los resultados obtenidos muestran que el GDP per cápita presenta una relación positiva con el nivel de felicidad, pero no es suficiente para explicarlo de manera completa. La inclusión de variables sociales, de salud e institucionales permitió construir un modelo significativamente más robusto.

El modelo de regresión múltiple explicó aproximadamente el 77% de la variabilidad en el nivel de felicidad, lo cual evidencia que el bienestar de una población depende de múltiples dimensiones y no únicamente del ingreso económico.