

# World Population

De acuerdo con los resultados obtenidos del módulo de Python se crearon las mismas gráficas en Tableau y fueron presentados en dos dashboards publicados en [este enlace](#).

El dashboard World Population Analysis presenta los países más poblados, los países más grandes y densos, así como un mapa interactivo donde se puede revisar la población de cada país y una gráfica que muestra la población por continente.

El dashboard Yearly Change Regression presenta los parámetros utilizados para hacer una regresión lineal. Estos son el cambio anual, la tasa de fertilidad, edad media y población urbana de cada país. Se hace una comparación de fertilidad y edad media ya que están inversamente relacionados con el cambio anual.

Este modelo se utiliza para responder a la pregunta inicial que es si la población se está expandiendo o contrayendo. Gracias a este análisis se logran relacionar estos parámetros para una predicción de expansión o contracción de una población.

Modelo de regresión lineal usado en Python:

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Dataset already loaded df_world_population

df_world_population.dropna(inplace=True)

# Define the target variable (Yearly Change %) and the predictors
target_column = 'Yearly Change %'
predictor_columns = ['Fert. Rate', 'Med. Age', 'Urban Pop %']

# Split the data into target variable (y) and predictors (X)
X = df_world_population[predictor_columns]
y = df_world_population[target_column]

# Split the data into training and testing sets
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Create and train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
mae = mean_absolute_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

print("Mean Squared Error:", mse)
print("Mean Absolute Error:", mae)
print("R-squared:", r2)

# Analyze the coefficients of the linear regression model
coefficients = pd.DataFrame({'Feature': predictor_columns, 'Coefficient': model.coef_})
print(coefficients)
```

Los resultados son los siguientes:

Mean Squared Error: 1.6592343948762598e-05

Mean Absolute Error: 0.00287971461948536

R-squared: 0.8587902372102856

	Feature	Coefficient
0	Fert. Rate	0.003031
1	Med. Age	-0.000744
2	Urban Pop %	0.007532

Los resultados indican lo siguiente:

Error cuadrático medio (MSE): El MSE es una medida de qué tan bien las predicciones del modelo coinciden con los valores reales objetivo. Un valor de MSE más bajo indica un mejor rendimiento del modelo.

Error absoluto medio (MAE): El MAE representa la diferencia absoluta promedio entre las predicciones del modelo y los valores reales objetivo. Un valor de MAE más bajo indica un mejor rendimiento del modelo.

R-cuadrado ( $R^2$ ): El valor de R-cuadrado mide la proporción de la varianza en la variable objetivo que es predecible a partir de las variables predictoras. Varía de 0 a 1, donde 0 significa que el modelo no explica la variabilidad de la variable objetivo y 1 significa que el modelo explica perfectamente la variabilidad. En este caso, un valor de R-cuadrado de aproximadamente 0.86 sugiere que el modelo explica alrededor del 86% de la variación en el cambio anual de población, lo cual es un ajuste razonablemente bueno.

Coefficientes: Los coeficientes representan el impacto de cada predictor en la variable objetivo. Un coeficiente positivo sugiere una relación positiva, mientras que un coeficiente negativo sugiere una relación negativa. En este caso, "Fert. Rate" y "Urban Pop %" tienen coeficientes positivos, lo que indica que un aumento en estas características se asocia con un aumento en el cambio anual de población. Por otro lado, "Med. Age" tiene un coeficiente negativo, lo que sugiere que un aumento en la edad mediana se asocia con una disminución en el cambio anual de población.

Variable objetivo: "Cambio anual %" - Esta es la variable que estamos tratando de predecir utilizando las variables predictoras.

Variables predictoras: "Tasa de fertilidad" - Tasa de fertilidad, que representa el número promedio de hijos nacidos de mujeres en edad fértil en un país o región. "Edad mediana" - Edad mediana, que representa la edad en la que la mitad de la población es mayor y la mitad es más joven. "Porcentaje de población urbana" - Porcentaje de población urbana, que representa la proporción de la población total que vive en áreas urbanas. El modelo de regresión lineal utiliza los valores de las variables predictoras para estimar la variable objetivo, es decir, el "Cambio anual %" en la población. Al analizar los coeficientes del modelo, podemos comprender las relaciones entre las variables predictoras y la variable objetivo. Por ejemplo, el coeficiente positivo para "Tasa de fertilidad" sugiere que mayores tasas de fertilidad se asocian con un mayor cambio anual en la población, mientras que el coeficiente negativo para "Edad mediana" indica que mayores edades medianas se asocian con un menor cambio anual en la población. El coeficiente para "Porcentaje de población urbana" sugiere una relación positiva, lo que significa que mayores porcentajes de población urbana se asocian con un mayor cambio anual en la población.

Tenemos un buen modelo que luego podemos entrenar con nuevos datos.