

ARIMA: Descifrando el Enigma de la Felicidad en las Ciudades

Somos una Asociación civil que busca prevenir los su[REDACTED]os en la población y queremos predecir el comportamiento de las ciudades en donde nos encontramos.

Tokyo, Delhi, Shanghai, Sao Paulo, Mumbai, Beijing, Estambul, Karachi, Dhaka y Ciudad de Mexico

Tenemos indices de Felicidad desde el año 2015 hasta el 2019 y queremos proyectar el comportamiento desde 2020 al 2025 para crear campañas preventivas cuando se necesiten

H0: La población de las ciudades está relacionada con el nivel de Felicidad

H1: La población de las ciudades no está relacionada con el nivel de Felicidad

¿Donde necesitamos mayor inversion para nuestras campañas en base a los datos que ya tenemos?

ARIMA:

Autoregressive (AR): Es la relación actual con las observaciones anteriores

Integrated (I): Hacemos la serie temporal estacionaria al diferenciarla

Moving Average (MA): Implica la relación entre una observación actual y los errores de predicción anteriores

Datos:

- Ciudades (10 ciudades)
- Población
- Año del 2015 al 2019
- Nivel de felicidad entre el 20 y el 150

	Ciudad	Poblacion	Anio	Felicidad
0	Tokyo	37797531	2015	110
1	Delhi	29399141	2016	113
2	Shanghai	26317104	2017	115
3	Sao Paulo	21192735	2018	27
4	Mumbai	20876485	2019	20

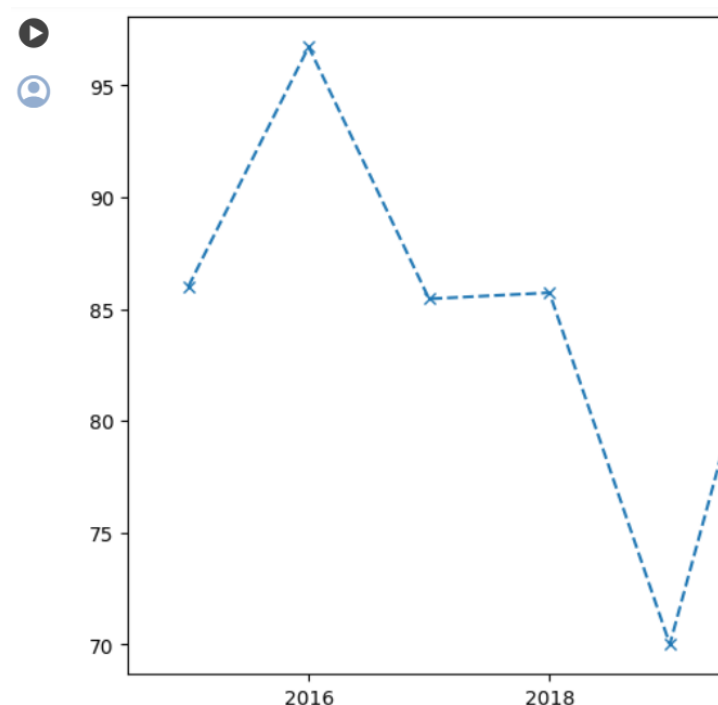
Paso 1:

Transformamos los datos en promedios (tendencia_promedio) para visualizar todas las tendencias juntas.

Código:

```
tendencia_promedio=datos_ciudades_felicidad.  
groupby('Anio')['Felicidad'].mean()
```

Resultado: La tendencia promedio se visualiza en el gráfico con una línea punteada y marcadores 'x'.



Paso 2:

Verificación de Estacionariedad:

Aseguramos que la serie temporal sea estacionaria para garantizar la robustez de nuestras predicciones.

Utilizamos la prueba de Dickey-Fuller antes 'result' y después 'result_diff' de la diferenciación para confirmar la estacionariedad.

Resultado: Después de la diferenciación, la estadística de prueba de Dickey-Fuller es -3.26 con un valor p de 0.0167. Esto indica que la serie temporal diferenciada es estacionaria.

```
# Diferenciación y nuevo test de Dickey-Fuller
serie_temporal_diff = serie_temporal.diff().dropna()
result_diff = adfuller(serie_temporal_diff)
print('Estadística de prueba de Dickey-Fuller después de diferenciación:', result_diff[0])
print('Valor p después de diferenciación:', result_diff[1])
```

Estadística de prueba de Dickey-Fuller después de diferenciación: -3.260407557486053
Valor p después de diferenciación: 0.01674371913423467

Paso 3:

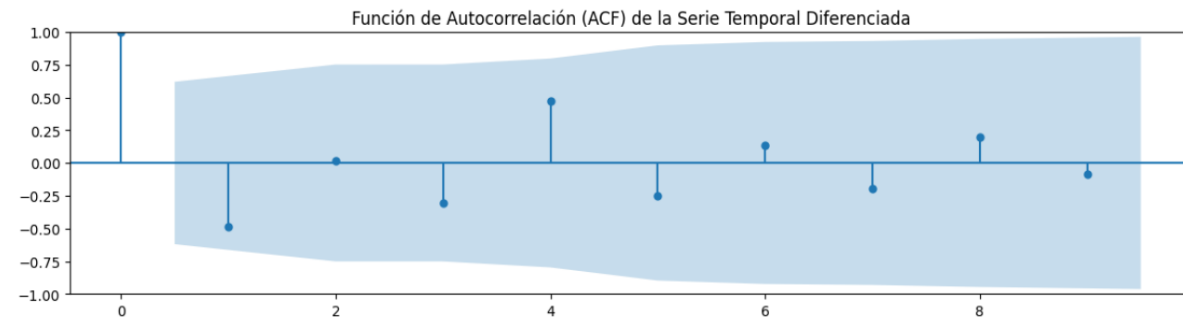
Análisis de Autocorrelación (ACF):

Analizamos la autocorrelación para descubrir relaciones temporales.

Una ACF fuerte implica una relación entre los niveles de felicidad de un año y los de años anteriores.

```
plot_acf(serie_temporal_diff.values, ax=plt.gca())
```

Resultado: La función de autocorrelación (ACF) muestra patrones significativos, lo que sugiere relaciones temporales entre los niveles de felicidad de diferentes años.



Paso 4:

Modelo ARIMA:

Utilizamos el modelo ARIMA para navegar a través del tiempo y entender cómo los patrones pasados afectan la felicidad futura.

```
orden_auto_arima = modelo_auto_arima.order
```

Resultado: La propuesta de orden del modelo ARIMA es (1, 0, 1), lo que significa que el modelo tiene una componente autorregresiva (AR) de orden 1, una componente de diferenciación (I) de orden 0 y una componente de media móvil (MA) de orden 1.

```
# Propuestas de modelos ARIMA
modelo_auto_arima = auto_arima(serie_temporal, seasonal=False, suppress_warnings=True)
orden_auto_arima = modelo_auto_arima.order
print('Propuesta de orden de modelo auto_arima:', orden_auto_arima)
```

```
Propuesta de orden de modelo auto_arima: (1, 0, 1)
```

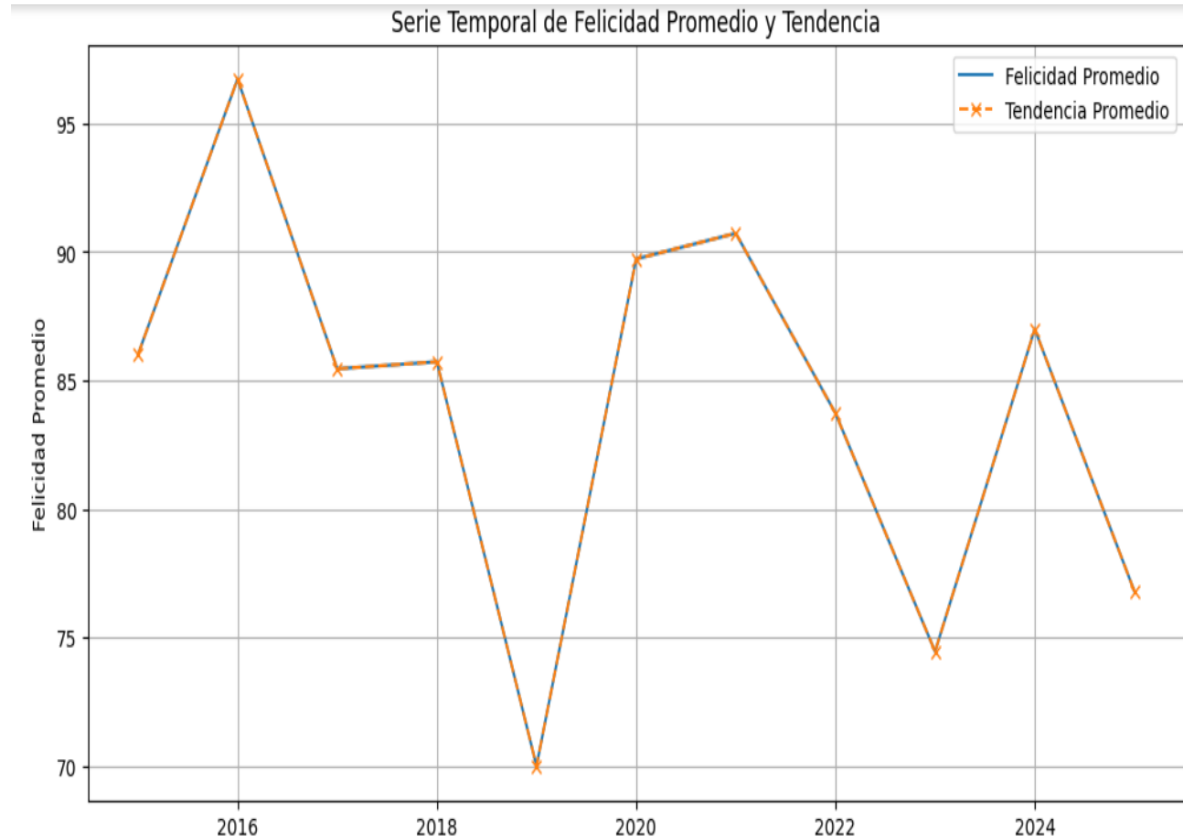
Paso 5:

Predicciones ARIMA:

Comparadas con predicciones anteriores, las predicciones ARIMA, basadas en conectividad temporal y estacionariedad, ofrecen una visión más precisa del futuro emocional.

predicciones_df

El coeficiente de correlación se calcula y se interpreta como una correlación positiva entre las predicciones ARIMA y la felicidad.



Paso 6 (Doble check):

Coeficiente de correlación:

Una vez realizadas las pruebas ARIMA, Comparamos si se tiene o no correlación entre las variables felicidad o no.

Observamos en el código 'correlacion_arima_felicidad'

Resultado: 0.7056 indica que existe una relación positiva entre los resultados generados por ARIMA y el nivel de felicidad

```
Coeficiente de correlación entre predicciones ARIMA y felicidad: 0.7056087688340265
```

```
Hay una correlación positiva entre las predicciones ARIMA y la felicidad.
```

Conclusión:

Existe una relación entre los niveles de población y la felicidad, por lo tanto podemos asegurarnos en base a los cálculos de cuando deberíamos invertir, es decir, no se tiene evidencia necesaria para rechazar nuestra H_0 .

H_0 : La población de las ciudades está relacionada con el nivel de Felicidad

GRACIAS