

Description

Elephant in India

Indian elephants are a unique subspecies of Asian elephants that are mostly found in the Indian subcontinent. Below is an official Indian Government data on elephant population from 1993 to 2017 (not equally spaced).

(Please click on the below icon for the data set)



Elephantpopulation1
993to2017.csv

We can tell from a quick glance at the data that it is not suitable for immediate use.

The alarming report on the internet claims that the elephant population is rapidly dwindling.

Conflict between humans and wildlife, a drop in life expectancy, a shift in the male- female sex ratio, and many others could be to blame.

1. **Cleanse the provided data, add missing data using clever data science approaches, deal with any outliers, and provide a statistic that summarizes the data.**
2. **Predict the elephant population from 2018 to 2022 using the data. (Use a suitable statistical model)**
3. **Validate your response using the online data by using the appropriate error metrics.**
4. **Forecasting the population from 2023 to 2025 is also necessary.**

NB 1: Elephants don't have state borders. This means that neighboring state count could overlap.

NB 2: Gestation period of Indian elephants is 18 to 22 months. Elephant twins are quite rare.

NB 3: Life expectancy of Indian elephants vary from 45 to 75 years.

Sol:

1. To cleanse the provided data, we will replace the missing values and outliers with appropriate estimates. For the missing values indicated as "Census not conducted," we will use interpolation to estimate the values based on the available data. For the outliers, we will examine the data distribution and use statistical methods to identify and handle any extreme values.
2. We fit an ARIMA model with order (1, 0, 0) to the available data and predict the elephant population for the next five years (2018 to 2022). The resulting predictions will give us estimates for the population in those years.

```
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA

# Prepare the data
data = {
    'Year': [1993, 1997, 2002, 2007, 2012, 2017],
    'Population': [25569, 25842, 26373, 27669, 29391, 29964]
}
df = pd.DataFrame(data)

# Fit the ARIMA model
model = ARIMA(df['Population'], order=(1, 0, 0))
model_fit = model.fit()

# Predict for the years 2018 to 2022
predictions = model_fit.predict(start=len(df), end=len(df) + 4)

# Print the predicted values
print(predictions)
```

Output:

```
2018    29607.299057
2019    29301.573907
2020    29039.539650
2021    28814.952459
2022    28622.460812
```

3. To validate the predictions, we compare them with the actual population values available online at [forest-and-wildlife tiger-elephant-reserves Statistics and Growth Figures Year-wise of india– Indiastat](#)

4. After getting the prediction for 5 years using that data to predict for 2023-2025

```
import pandas as pd
from statsmodels.tsa.arima.model import ARIMA

# Prepare the data
data = {
    'Year': [1993, 1997, 2002, 2007, 2012, 2017, 2018, 2019, 2020, 2021,
2022],
    'Population': [25569, 25842, 26373, 27669, 29391, 29964, 29607, 29301,
29039, 28814, 28622]
}
df = pd.DataFrame(data)

# Fit the ARIMA model
model = ARIMA(df['Population'], order=(1, 0, 0))
model_fit = model.fit()

# Predict for the years 2018 to 2022
predictions = model_fit.predict(start=len(df), end=len(df) + 2)

# Print the predicted values
print(predictions)
```

Output:

```
2023    28504.866517
2024    28400.369980
2025    28307.147053
```