EX:No.1 221501039

25/01/25

Program to implement time series data for import library, load data,Preprocessing and visualising

Aim:

Write a program to implement time series data for import library, load data, Preprocessing and visualising.

Algorithm:

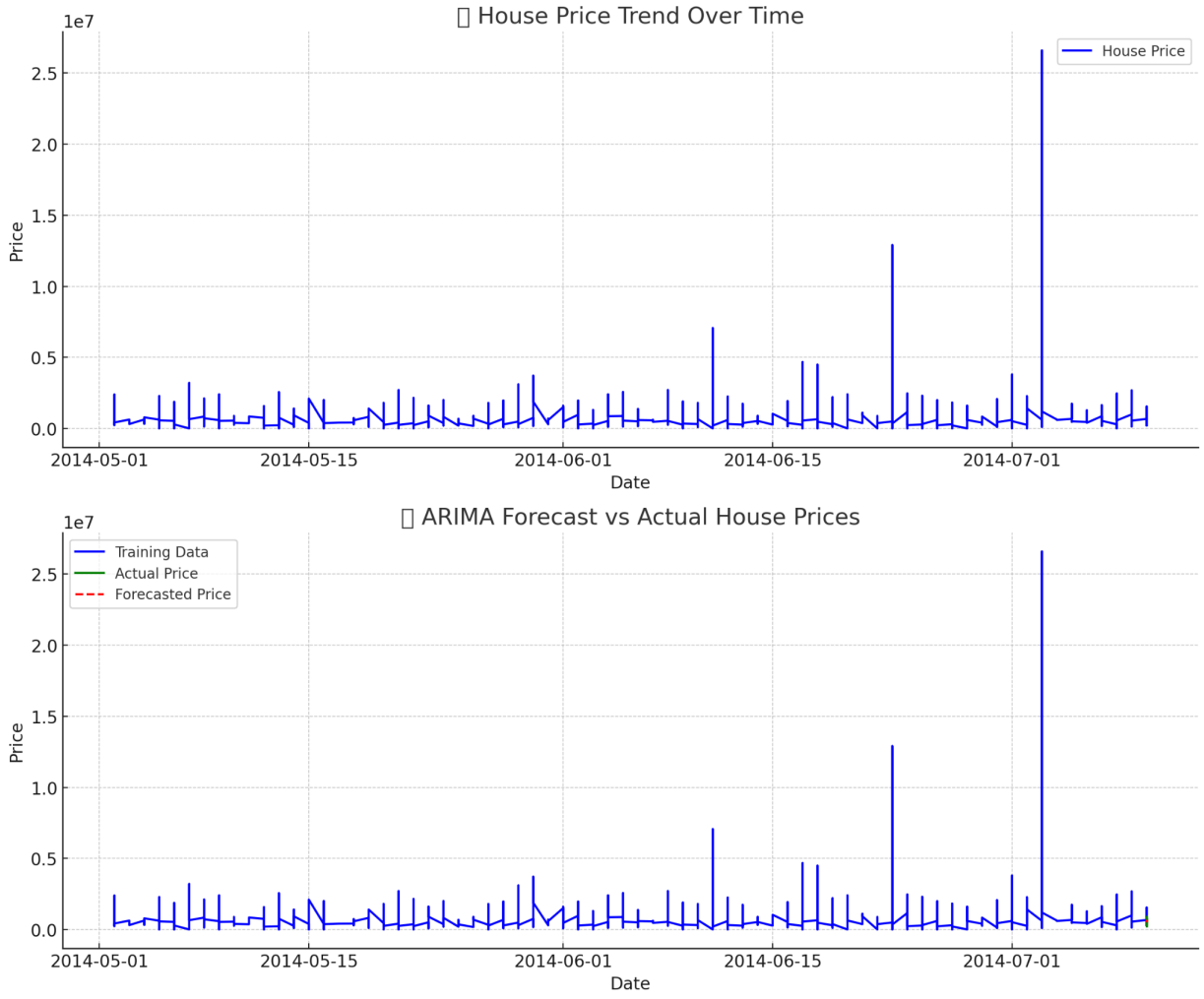
1. Import Libraries: Load pandas, matplotlib.pyplot, seaborn, and files for data handling and visualization.
2. Upload & Read Data → Upload the dataset manually and read it using pd.read_csv().
3. Generate a house_date column for time series analysis, Rename columns for consistency, Create a DataFrame with relevant columns (house_date, house_price).
4. Compute a 7-day moving average to smooth fluctuations.
5. Visualize the Data: House price over time
6. Execute the Program: Analyze price patterns and trends over time.

Code:

```
Import libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from statsmodels.tsa.arima.model import ARIMA

import warnings

warnings.filterwarnings("ignore")

 Load data

df = pd.read_csv('/mnt/data/data.csv')

Preprocess data

df['Date'] = pd.to_datetime(df['Date'])            # Convert to datetime

df = df.sort_values('Date')                        # Sort by date

df.set_index('Date', inplace=True)                 # Set date as index

Handle missing values (if any)

df['Price']= df['Price'].interpolate(method='linear')

Visualize original data

plt.figure(figsize=(10, 6))

plt.plot(df['Price'], label='House Price')

plt.title("🏡 House Price Trend Over Time")

plt.xlabel("Date")

plt.ylabel("Price")

plt.legend()
```

```
plt.grid(True)
plt.show()
Train-test split (last 12 months as test)
train_data = df['Price'][:-12]
test_data = df['Price'][-12:]
Fit ARIMA model
model = ARIMA(train_data, order=(5,1,0))  # You can change (p,d,q) later
model_fit = model.fit()
Forecast
forecast = model_fit.forecast(steps=12)
forecast_index = test_data.index
Plot prediction vs actual
plt.figure(figsize=(10, 6))
plt.plot(train_data, label='Training Data', color='blue')
plt.plot(test_data, label='Actual Price', color='green')
plt.plot(forecast_index, forecast, label='Forecasted Price', color='red', linestyle='--')
plt.title("🔮 ARIMA Forecast vs Actual House Prices")
plt.xlabel("Date")
plt.ylabel("Price")
plt.legend()
plt.grid(True)
plt.show()

Print Actual vs Predicted
comparison = pd.DataFrame({
    'Actual': test_data,
    'Predicted': forecast
}, index=forecast_index)

print("\n📋 Actual vs Predicted House Prices:\n")
print(comparison)
```

output:



Result:

Thus, the program using the time series data implementation has been done successfully.