PyCaret Time Series EDA Feature Contribution

📍 Repository

Project: PyCaret — Open Source Low-Code Machine Learning Library
Your Fork: https://github.com/farhann-saleem/pycaret
PR: [ENH] Time-series EDA: Series/DataFrame input, ADF, ACF, decomposition + tests
Linked Issue: #3696 – EDA for Time Series module

🚀 Objective

Add a **Time Series EDA (Exploratory Data Analysis)** utility to PyCaret's time_series module, inspired by existing regression EDA functionality.

The goal was to help users quickly:

- Inspect data quality,
- Check for stationarity,
- Visualize trends, seasonality, and autocorrelation before modeling.

🧩 Implementation Summary

🏗 Environment Setup

Step	Tool / Version	Notes
Python	**3.10.10**	PyCaret supports <3.13
Virtual Environment	venv	py -3.10 -m venv .venv
Install editable	pip install -e .[full]	Editable mode lets local changes reflect instantly
IDE	VS Code	Interpreter set to .venv

🧱 Code Additions

📁 File Added

pycaret/time_series/eda.py

🔍 Features Implemented

Feature	**Description**
✅ Numeric Summary	Prints length, missing values, mean, min, max
✅ ADF Stationarity Test	Uses statsmodels.tsa.stattools.adfuller
✅ Line Plot	Matplotlib and optional Plotly
✅ ACF Plot	Matplotlib (plot_acf) + optional Plotly
✅ Seasonal Decomposition	Uses seasonal_decompose for trend/seasonal/residual
✅ DataFrame + Target	API aligned with PyCaret style
✅ Plotly Mode	Interactive mode via plotly=True
✅ Test Suite	Added under tests/test_time_series_eda.py
✅ CI Safe	Uses Agg backend (no GUI dependency)

🧪 Unit Tests

File: tests/test_time_series_eda.py

- Tested Series and DataFrame inputs.
- Verified ADF output and decomposition logic.
- Checked behavior on differenced data (stationarity improves).
- Confirmed proper error when target missing.
- **All tests passed** ✅

```
pytest -q tests/test_time_series_eda.py
```

```
3 passed, 1 warning (distutils deprecation)
```

🖼️ Visual Results

Generated and attached in PR:

Time Series Line Plot
Autocorrelation (ACF) Plot
Seasonal Decomposition Plot

(You can keep the PNGs in a folder named /pr_screens/ for portfolio use.)

🧠 Key Learnings

Concept	Takeaway
Open-Source Workflow	Fork → Branch → Commit → Push → PR
Editable Installs	Enables local dev in large repos
EDA for Time Series	Trend, seasonality, autocorrelation, stationarity
Statistical Testing	Interpreting ADF test p-values properly
CI-safe Plotting	Using Agg backend to prevent GUI errors
Professional PR Writing	Structured description, visuals, checklists

🗂️ Current PR Status

Field	Status
PR merged	❌ (not merged yet)
Repo activity	Low (last owner commit ~3 years ago)
Tests passing	✅ locally
Conflicts	None
Functionality	Complete, working feature