

# EDA

Frequency: monthly

Rows: 223560 | Series: 1428 | Domains: 2

Date ranges per domain:

	domain	min	max	count
0	m1_monthly_dataset	1900-01-01	1991-01-01	55998
1	m3_monthly_dataset	1857-01-01	2005-02-01	167562

Series length summary:

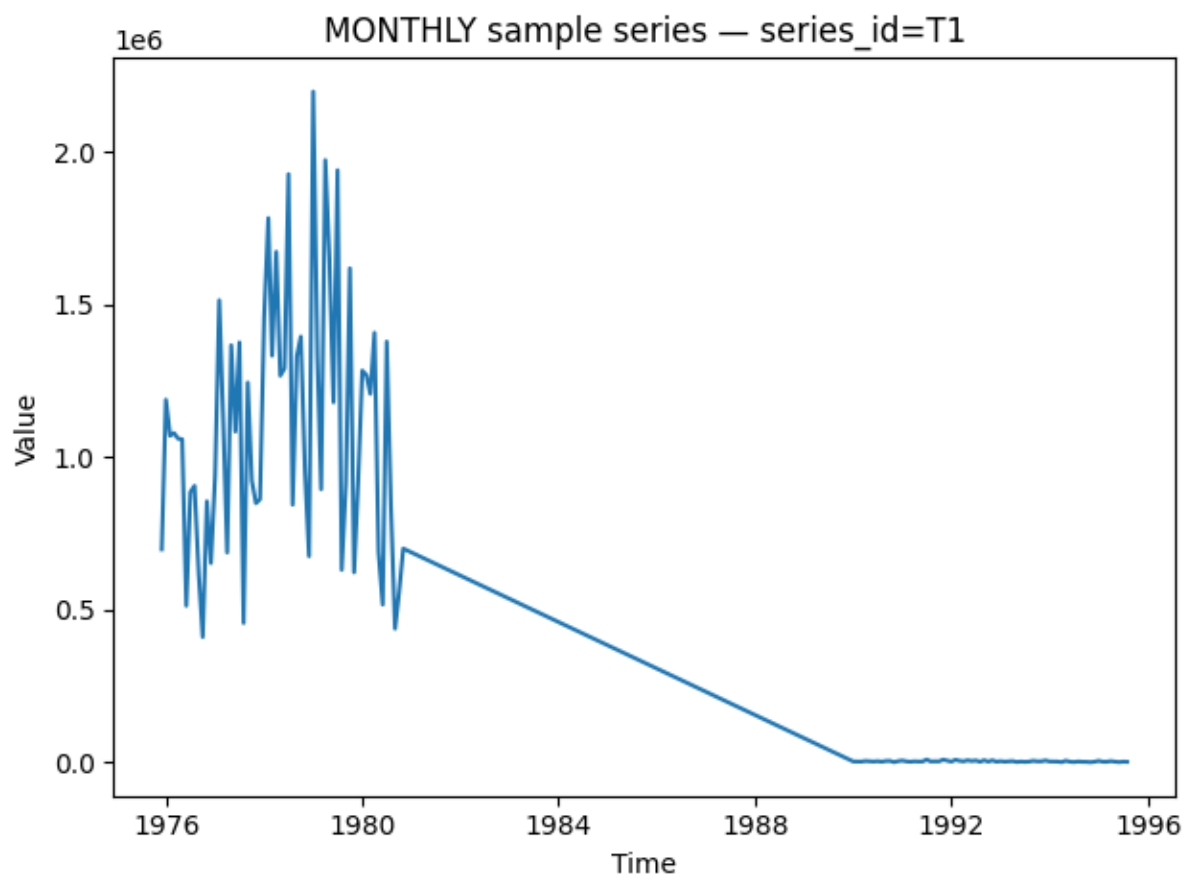
	length
count	1428.000000
mean	156.554622
std	44.403709
min	66.000000
25%	134.000000
50%	142.000000
75%	192.000000
max	288.000000

dtype: float64

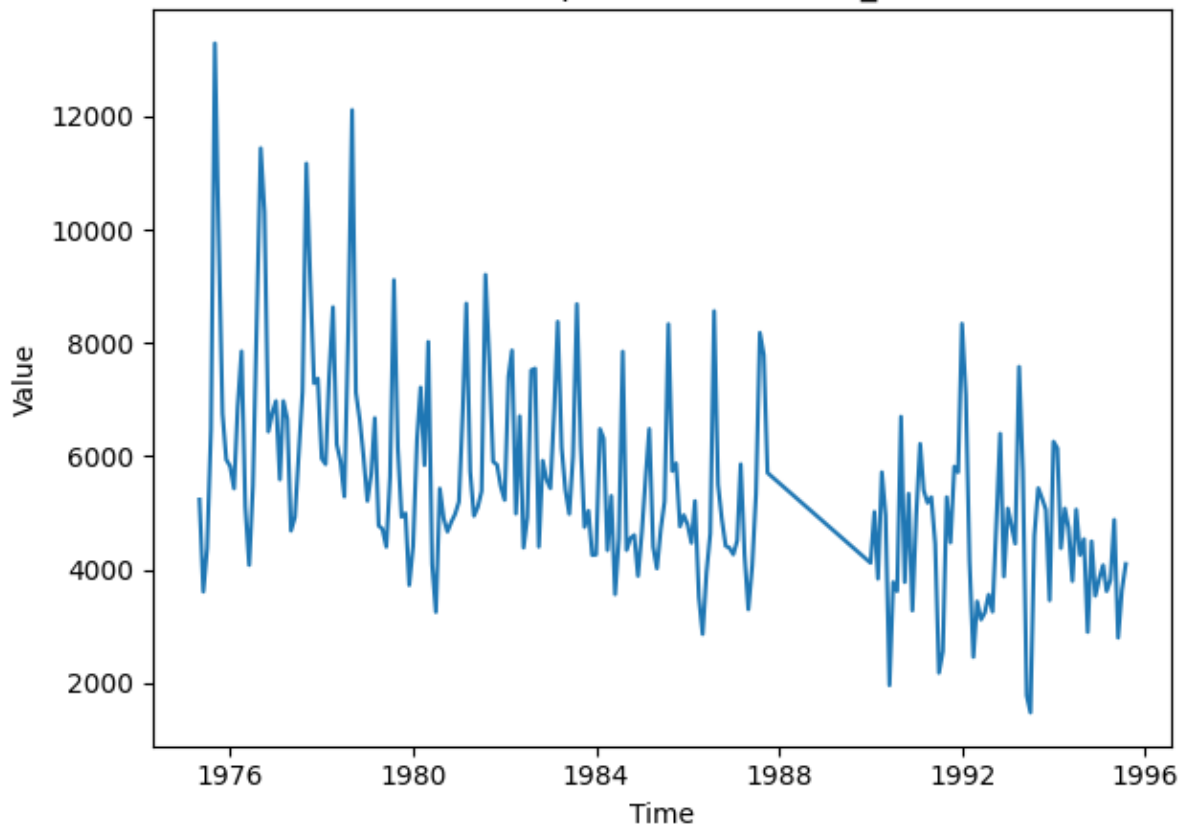
Value stats per domain:

	domain	mean	std	min	max	median	count
0	m1_monthly_data et	9635.442974	49670.448459	0.0	2196870.0	371.0	55998
1	m3_monthly_data et	5000.309346	2296.394263	-1200.0	86730.0	4850.0	167562

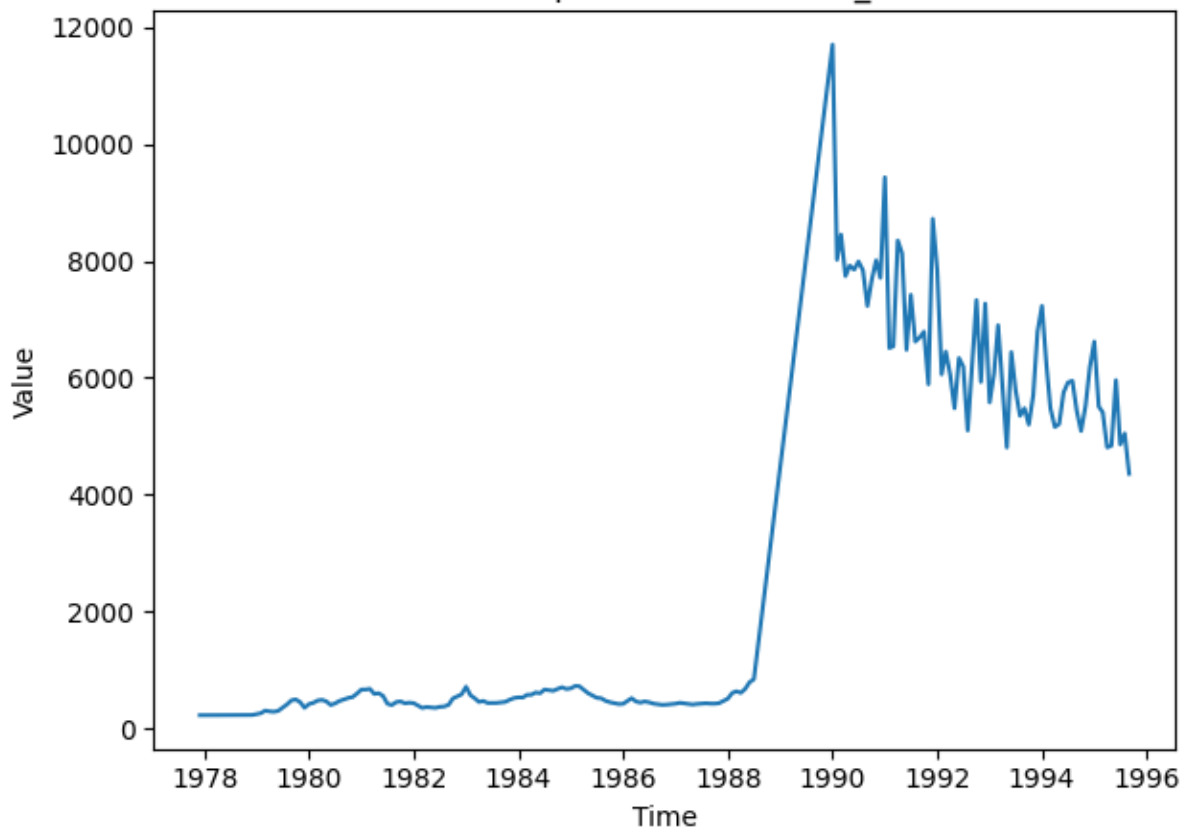
Sample plots (up to 3 series):

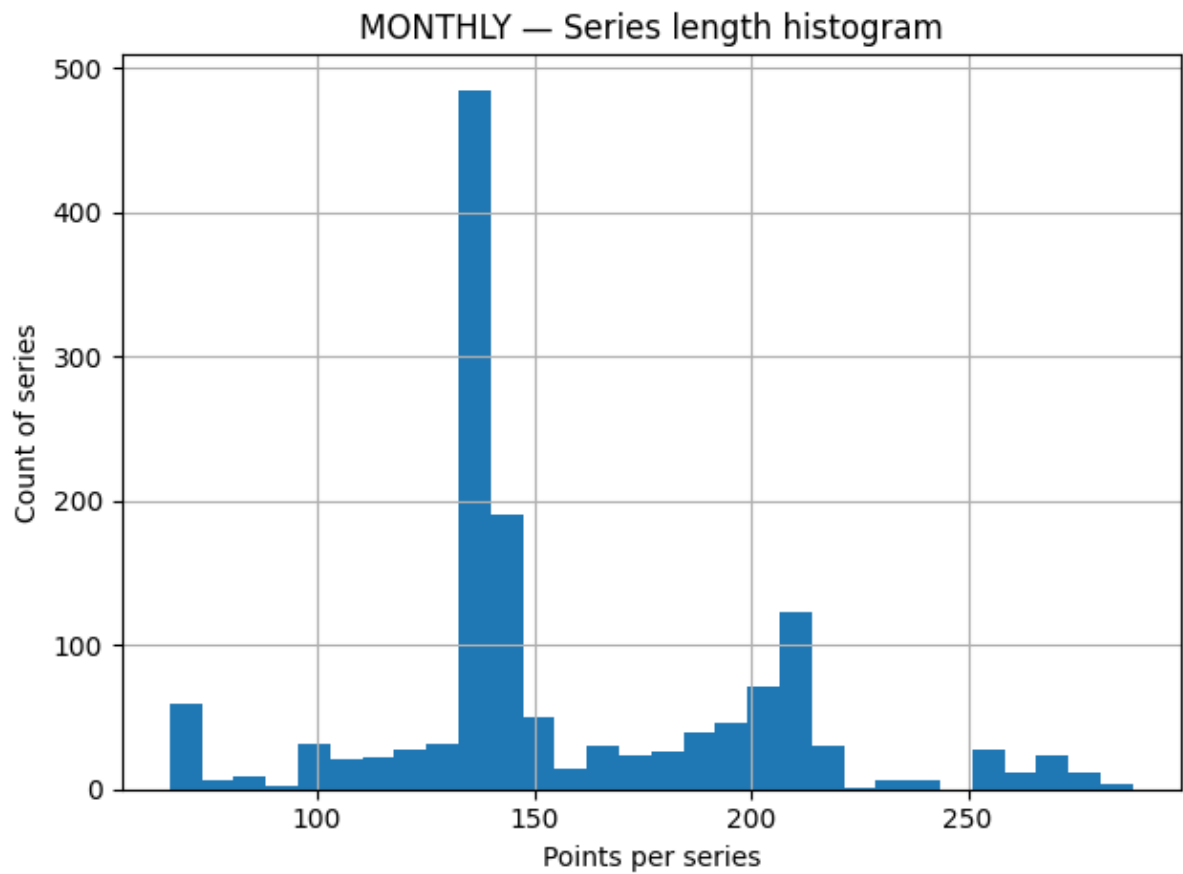


MONTHLY sample series — series\_id=T10



MONTHLY sample series — series\_id=T100





-----  
Frequency: quarterly

Rows: 46948 | Series: 756 | Domains: 2

Date ranges per domain:

	domain	min	max	count
0	m1_quarterly_dataset	1900-01-01	2005-10-01	9944
1	m3_quarterly_dataset	1946-01-01	1996-07-01	37004

Series length summary:

	length
count	756.000000

**length**

<b>mean</b>	62.100529
<b>std</b>	23.063642
<b>min</b>	24.000000
<b>25%</b>	52.000000
<b>50%</b>	52.000000
<b>75%</b>	72.000000
<b>max</b>	158.000000

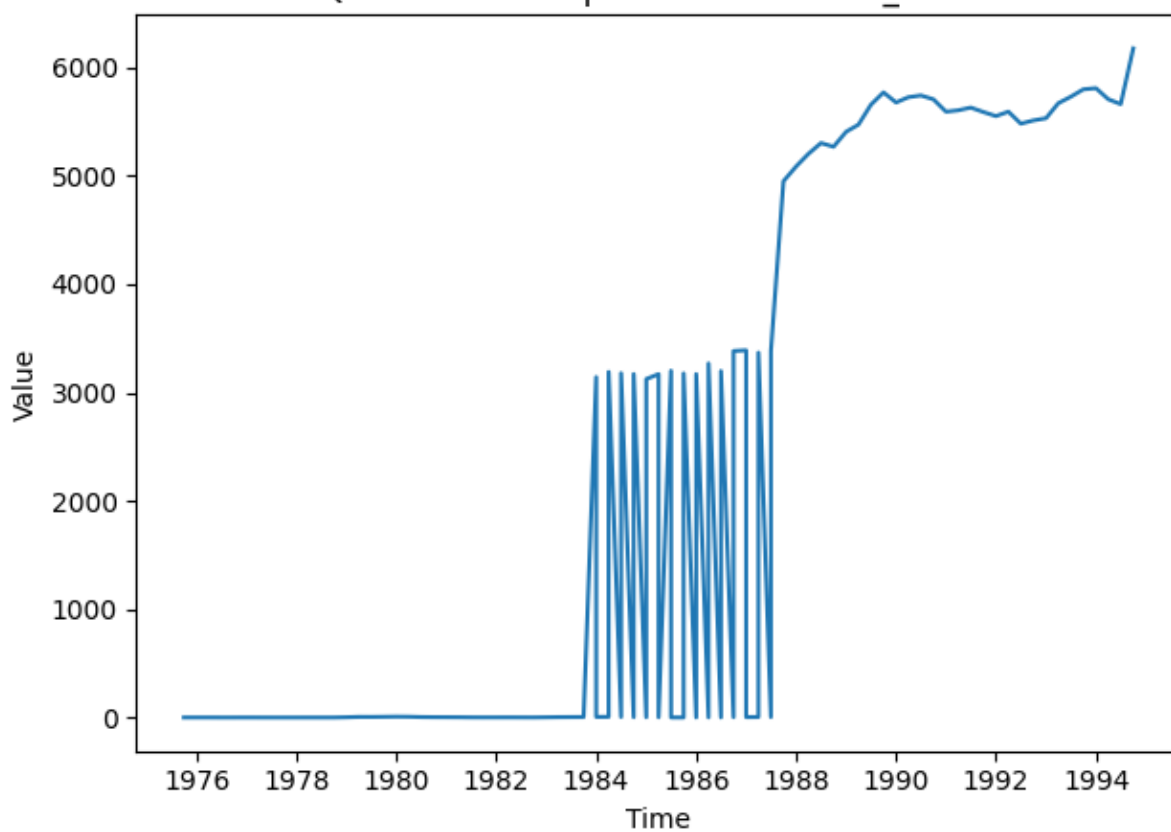
**dtype:** float64

Value stats per domain:

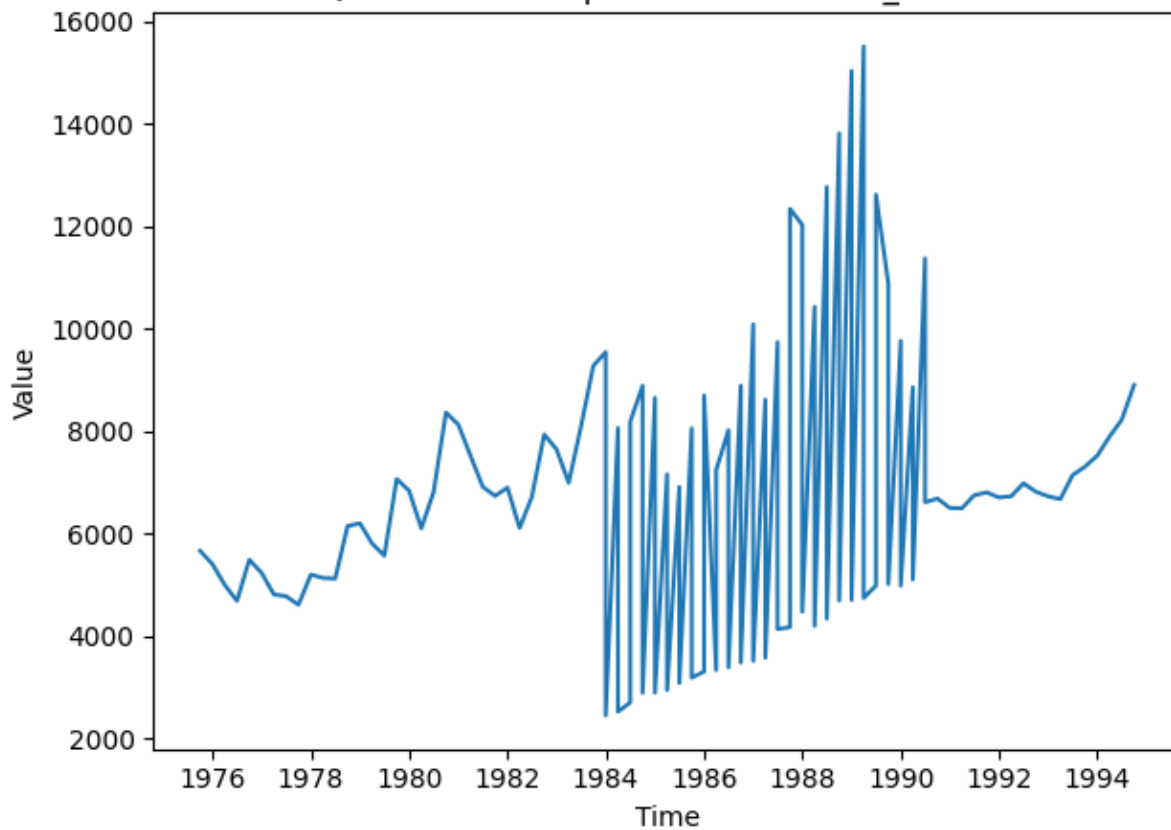
	<b>domain</b>	<b>mean</b>	<b>std</b>	<b>min</b>	<b>max</b>	<b>media n</b>	<b>coun t</b>
<b>0</b>	m1_quarterly_daset	13718.814578	66621.971370	0.34	818797.0	132.16	9944
<b>1</b>	m3_quarterly_daset	4981.028521	1793.966855	121.00	20375.0	4905.31	37004

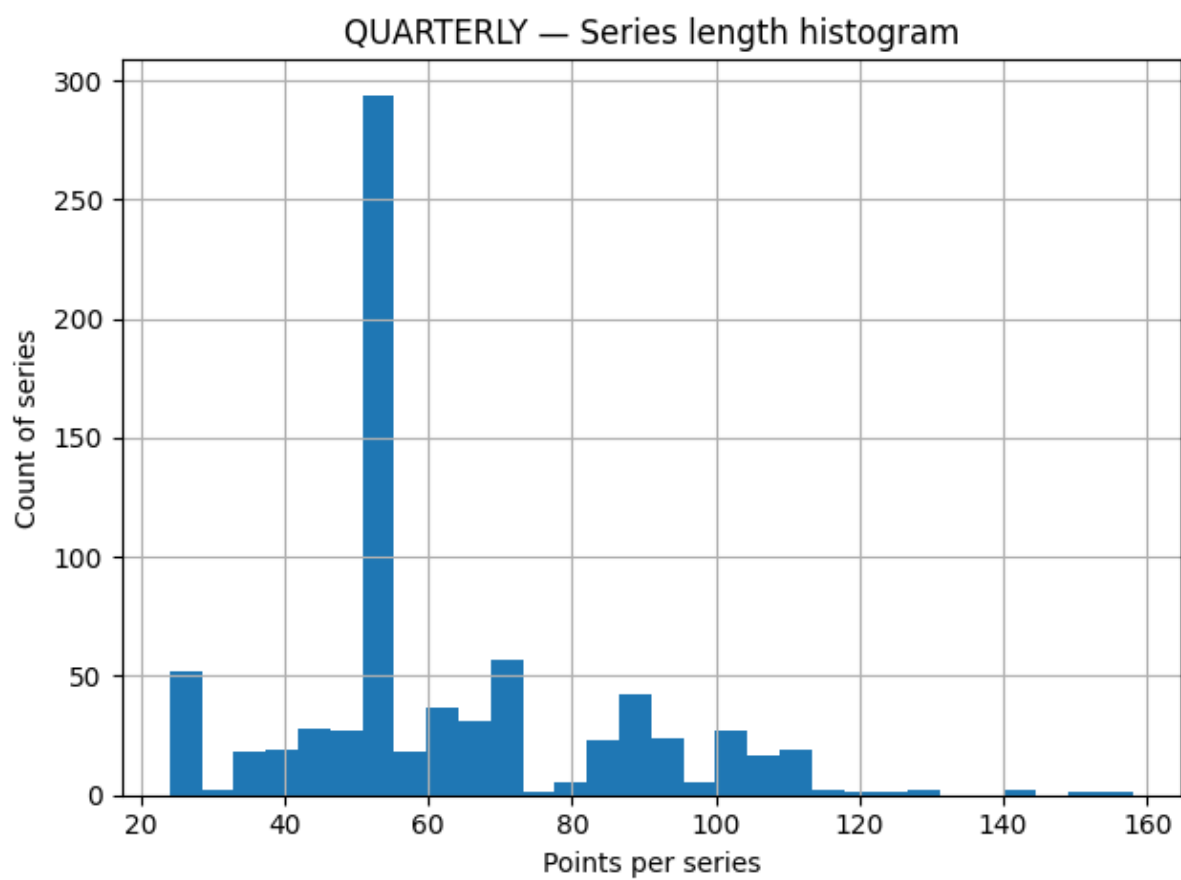
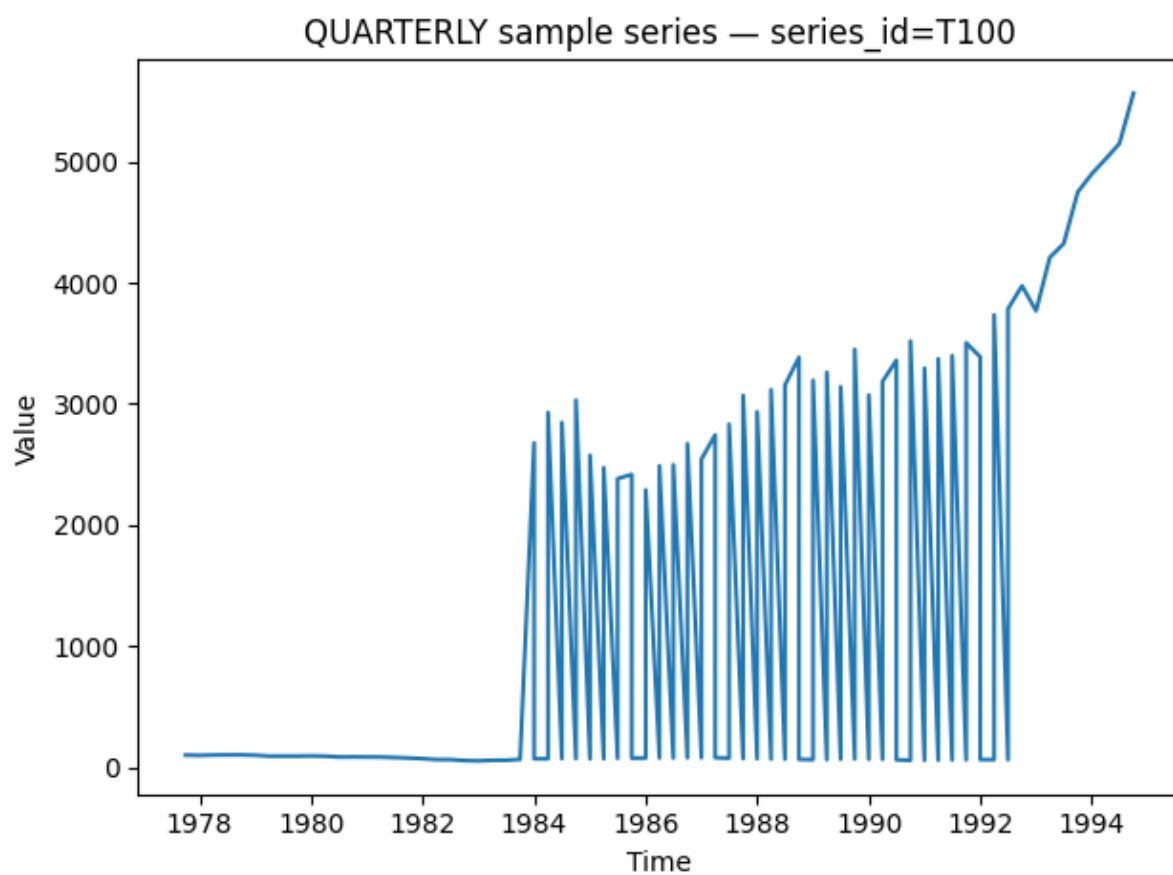
Sample plots (up to 3 series):

QUARTERLY sample series — series\_id=T1



QUARTERLY sample series — series\_id=T10





-----  
Frequency: yearly

Rows: 22834 | Series: 645 | Domains: 2

Date ranges per domain:

	domain	min	max	count
0	m1_yearly_dataset	1865-01-01	2033-01-01	4515
1	m3_yearly_dataset	1811-01-01	2001-01-01	18319

Series length summary:

	length
count	645.000000
mean	35.401550
std	13.806279
min	20.000000
25%	23.000000
50%	33.000000
75%	46.000000
max	92.000000

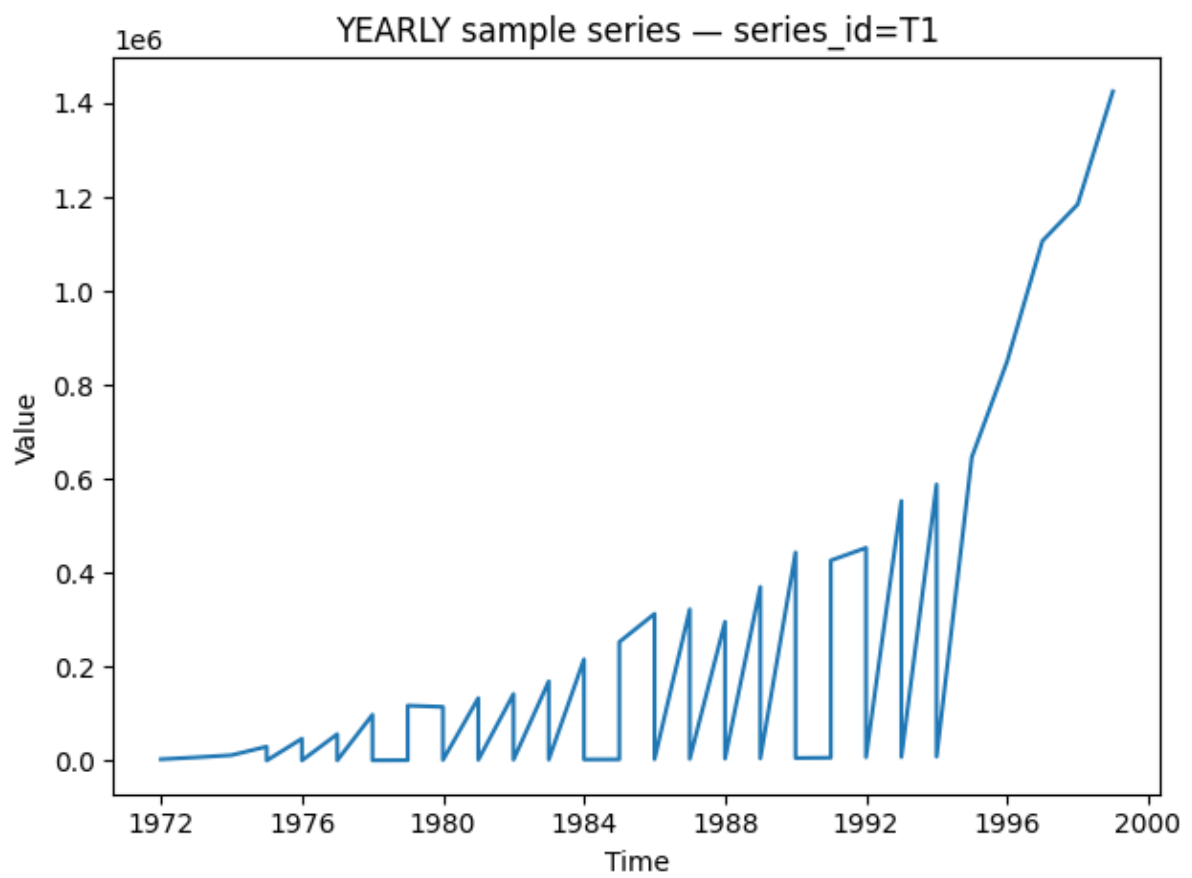
**dtype:** float64

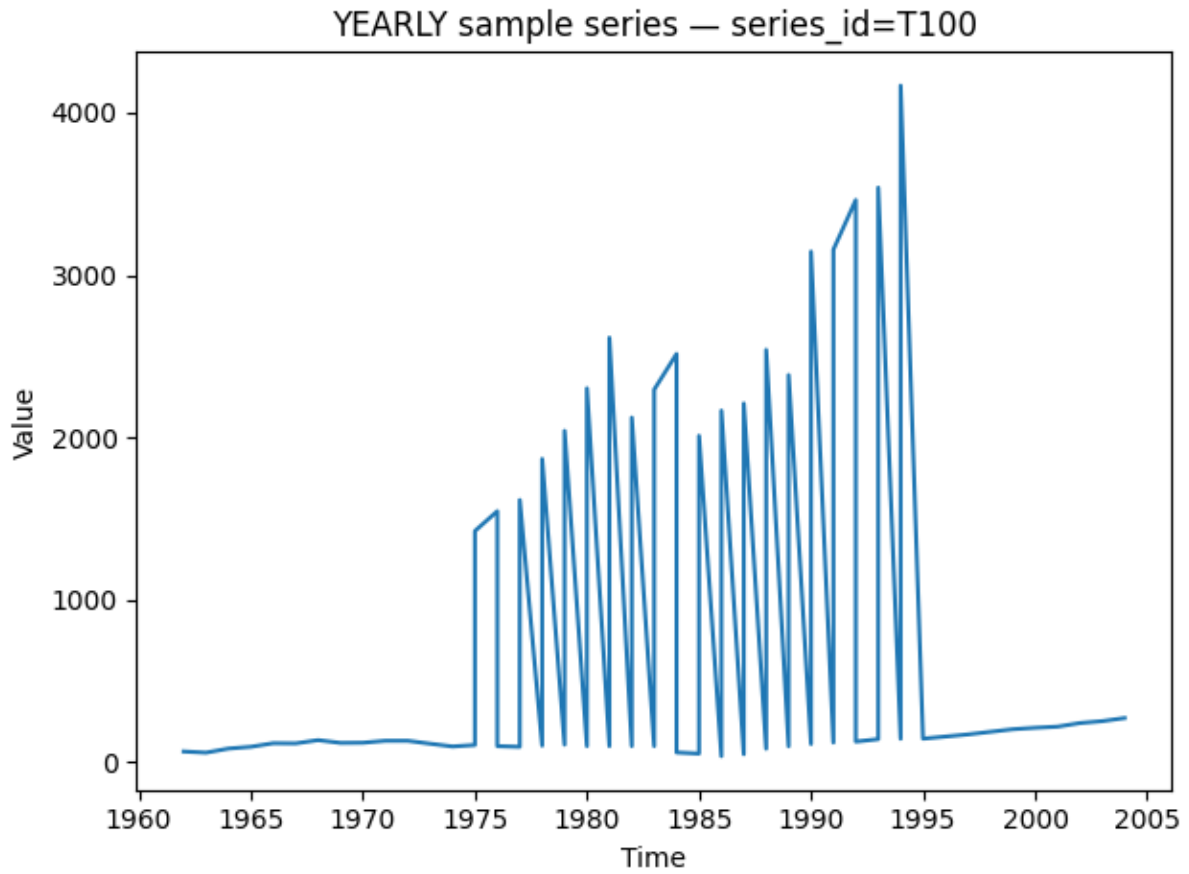
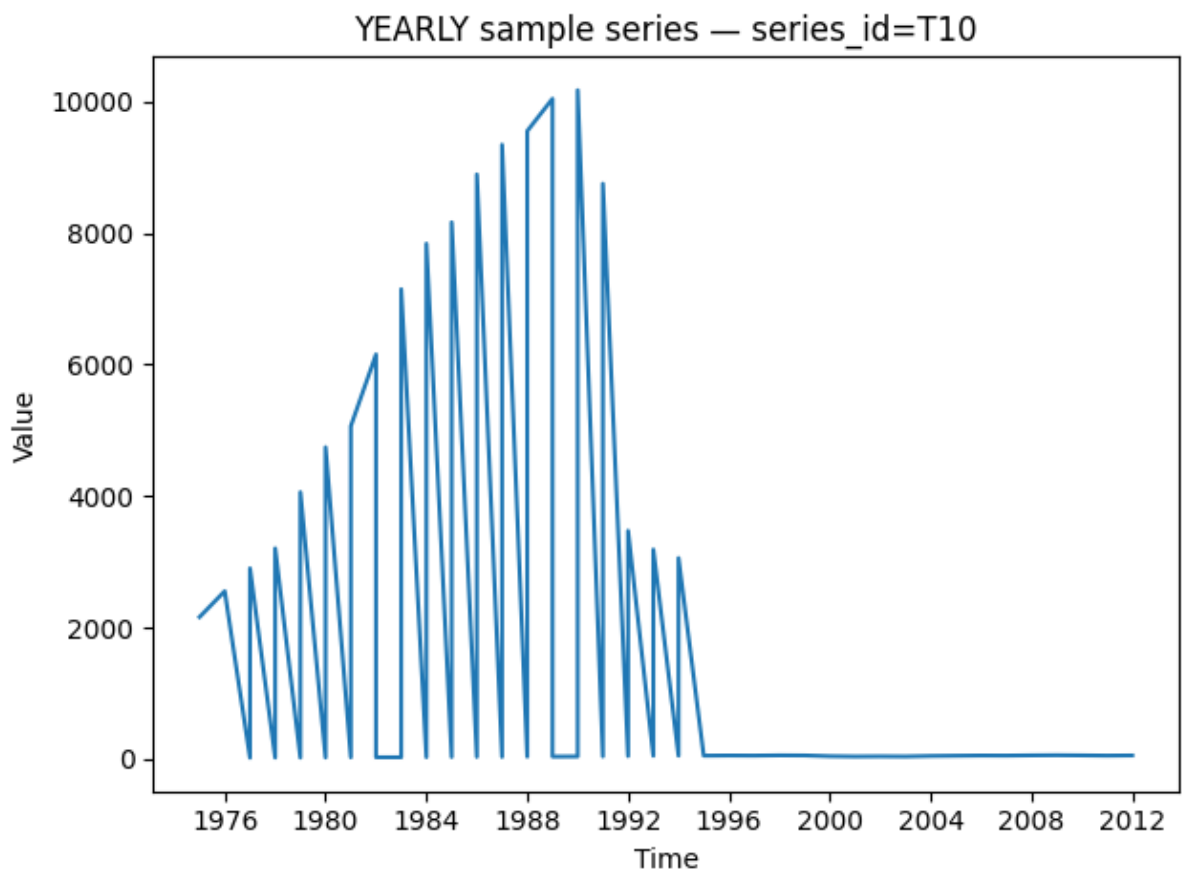
Value stats per domain:

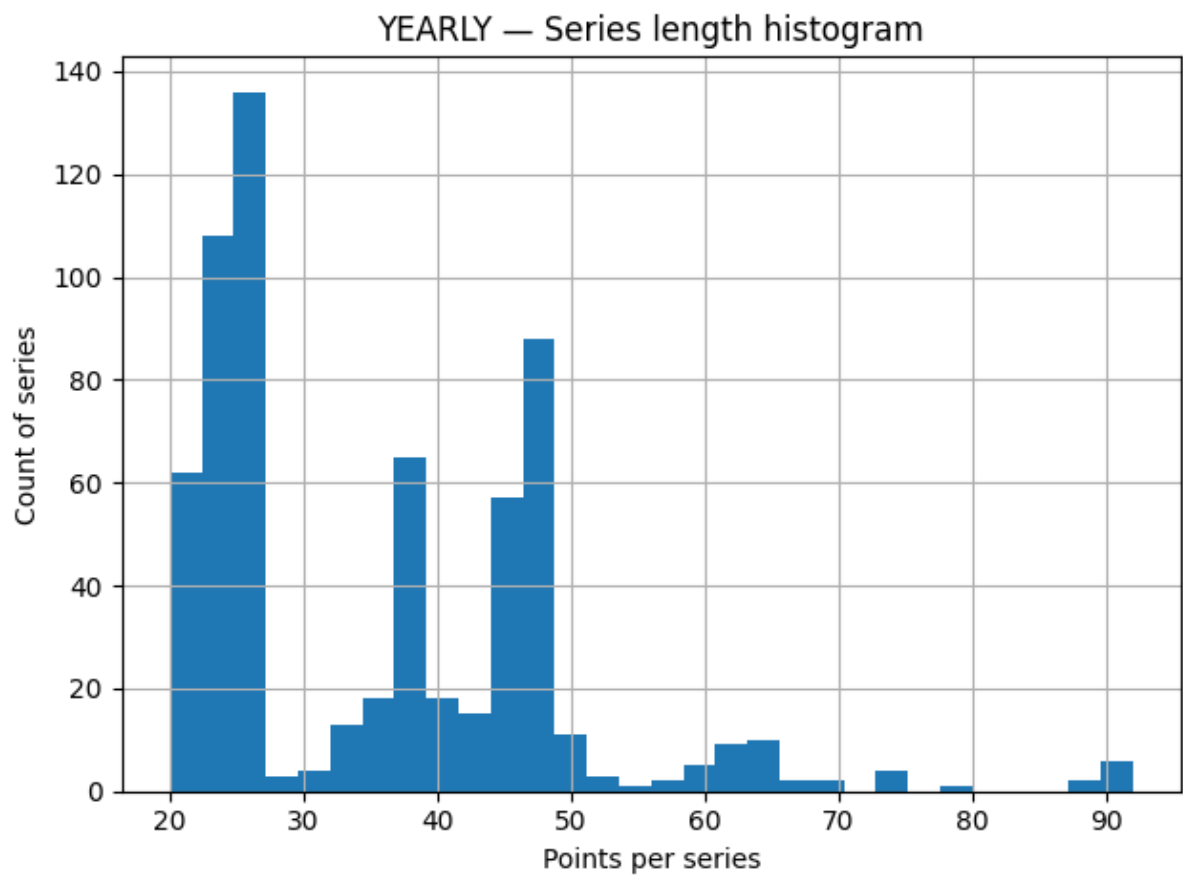


	domain	mean	std	min	max	median	count
0	m1_yearly_data set	592652.0425	3.210858e+06	0.1	47180900.0	870.0	4515
1	m3_yearly_data set	4749.213704	2.510016e+03	30.0	45525.66	4639.5	18319

Sample plots (up to 3 series):







Frequency: weekly

[INFO] Empty corpus — skipping EDA.

Frequency: hourly

[INFO] Empty corpus — skipping EDA.

# Extended EDA -

=====

[EXTENDED EDA] Frequency = monthly

Domain coverage:

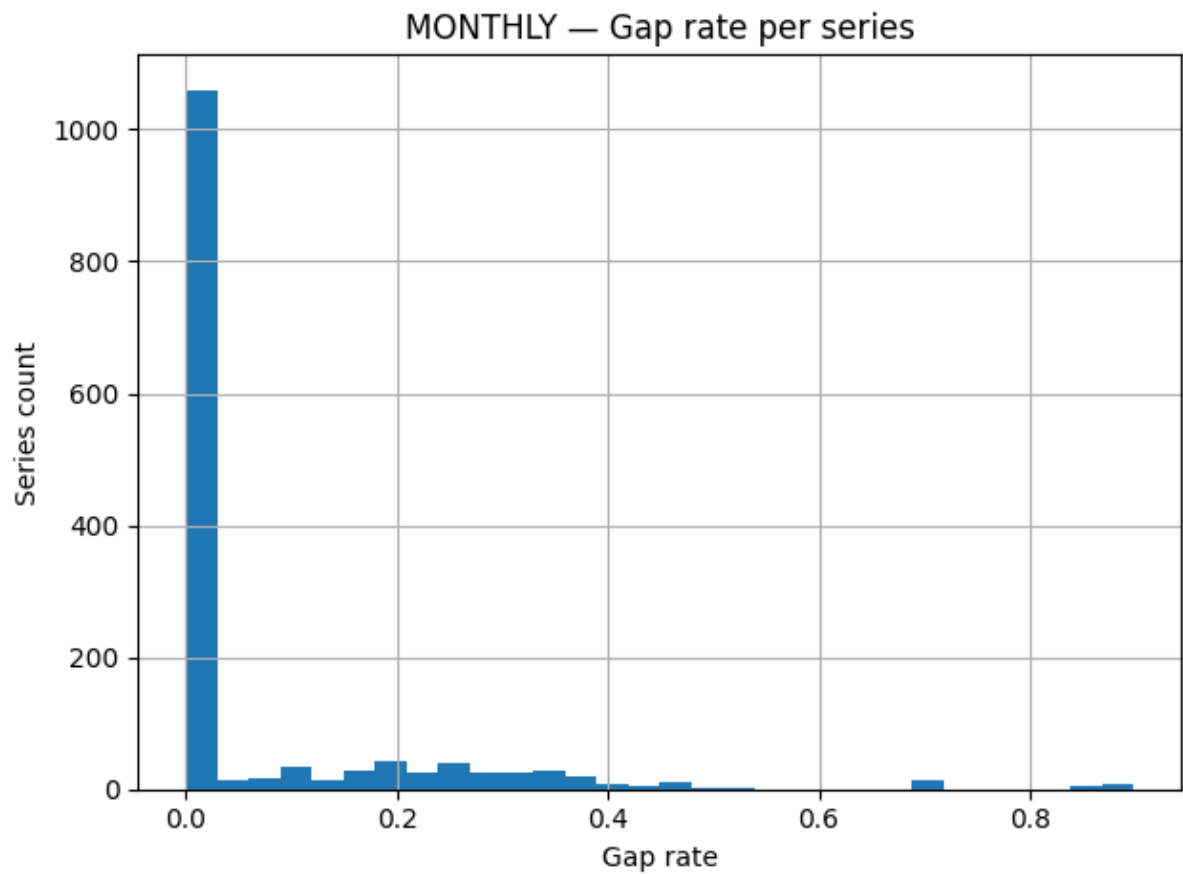
	domain	min	max	count	coverage_days
0	m1_monthly_dataset	1900-01-01	1991-01-01	55998	33237
1	m3_monthly_dataset	1857-01-01	2005-02-01	167562	54087

Overlap window: start=1900-01-01 00:00:00, end=1991-01-01 00:00:00, days=33237

Gaps summary (per series):

	gap_rate
count	1428.000000
mean	0.072216
std	0.154393
min	0.000000
25%	0.000000
50%	0.000000
75%	0.057914
max	0.898172

dtype: float64



Negatives/outliers summary:

	neg_rate	outlier_rate
count	1428.000000	1428.000000
mean	0.000007	0.012458
std	0.000276	0.028727
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	0.000000	0.009524

	<b>neg_rate</b>	<b>outlier_rate</b>
<b>max</b>	0.010417	0.315789

Volatility & growth summary:

	<b>cv</b>	<b>median_abs_diff</b>	<b>median_pct_growth</b>
<b>count</b>	1428.000000	1428.000000	1428.000000
<b>mean</b>	0.467604	368.445802	0.003471
<b>std</b>	0.379294	738.876829	0.019560
<b>min</b>	0.012345	0.310000	-0.245639
<b>25%</b>	0.138571	50.990000	-0.000371
<b>50%</b>	0.335031	170.000000	0.002861
<b>75%</b>	0.758113	381.625000	0.007993
<b>max</b>	2.065382	9616.000000	0.380579

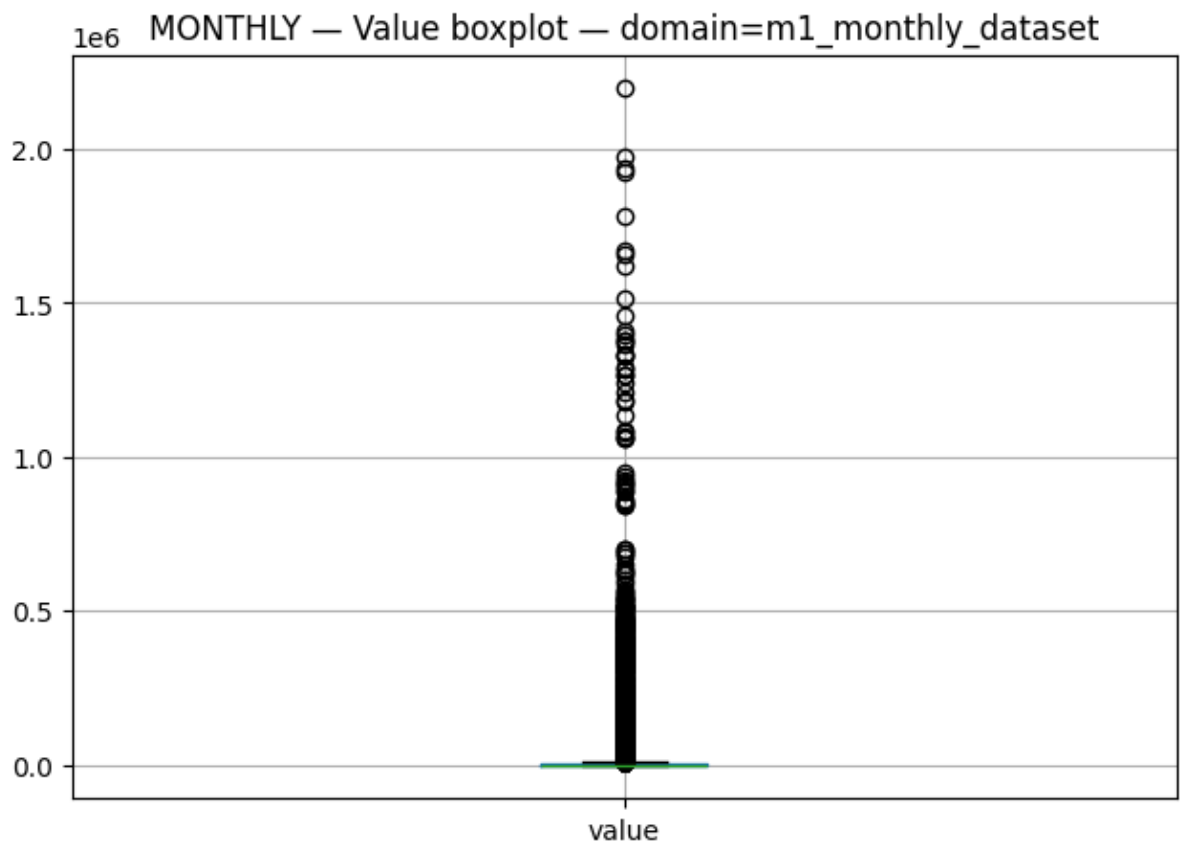
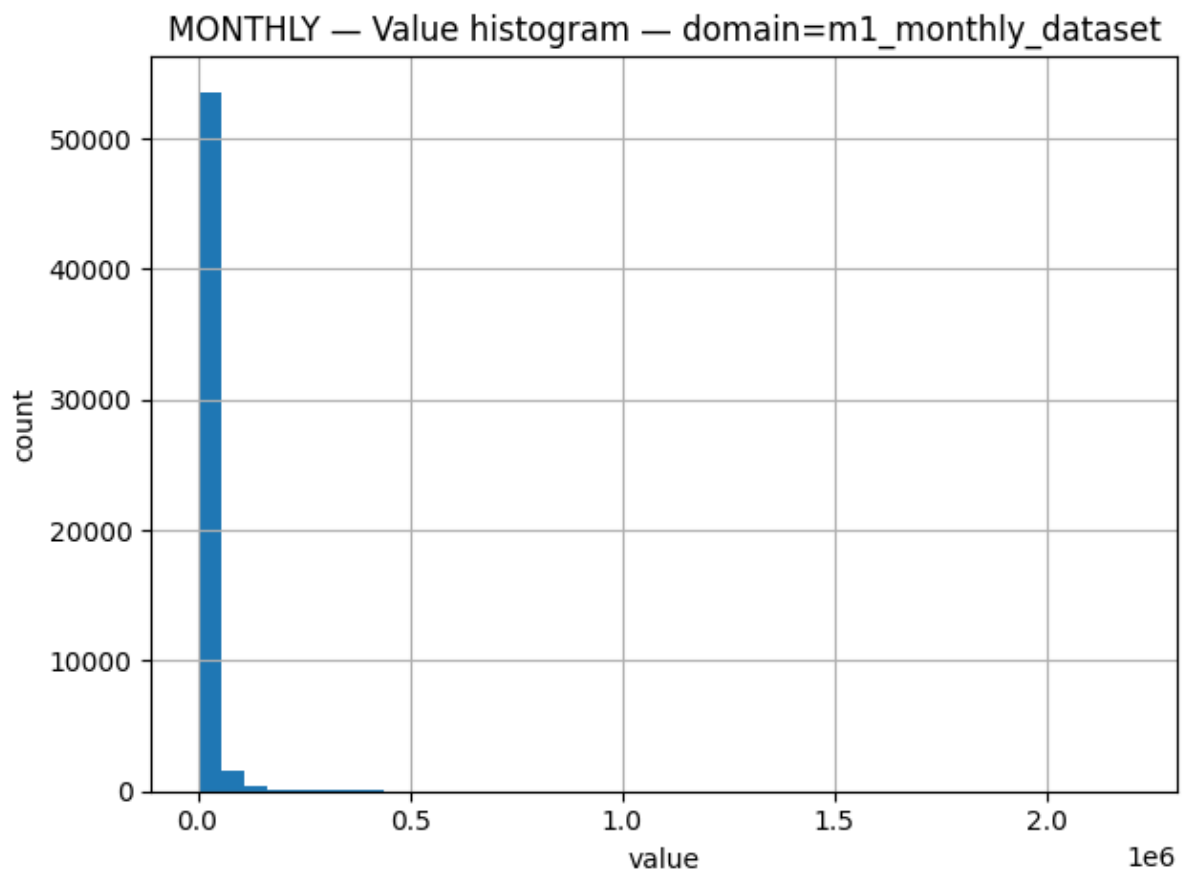
Spike summary:

	<b>n_spikes</b>	<b>spike_rate</b>
<b>count</b>	1428.000000	1428.000000
<b>mean</b>	2.316527	0.014183
<b>std</b>	2.685178	0.013747
<b>min</b>	0.000000	0.000000
<b>25%</b>	1.000000	0.004908

	<b>n_spikes</b>	<b>spike_rate</b>
<b>50%</b>	2.000000	0.013378
<b>75%</b>	3.000000	0.021053
<b>max</b>	20.000000	0.098039

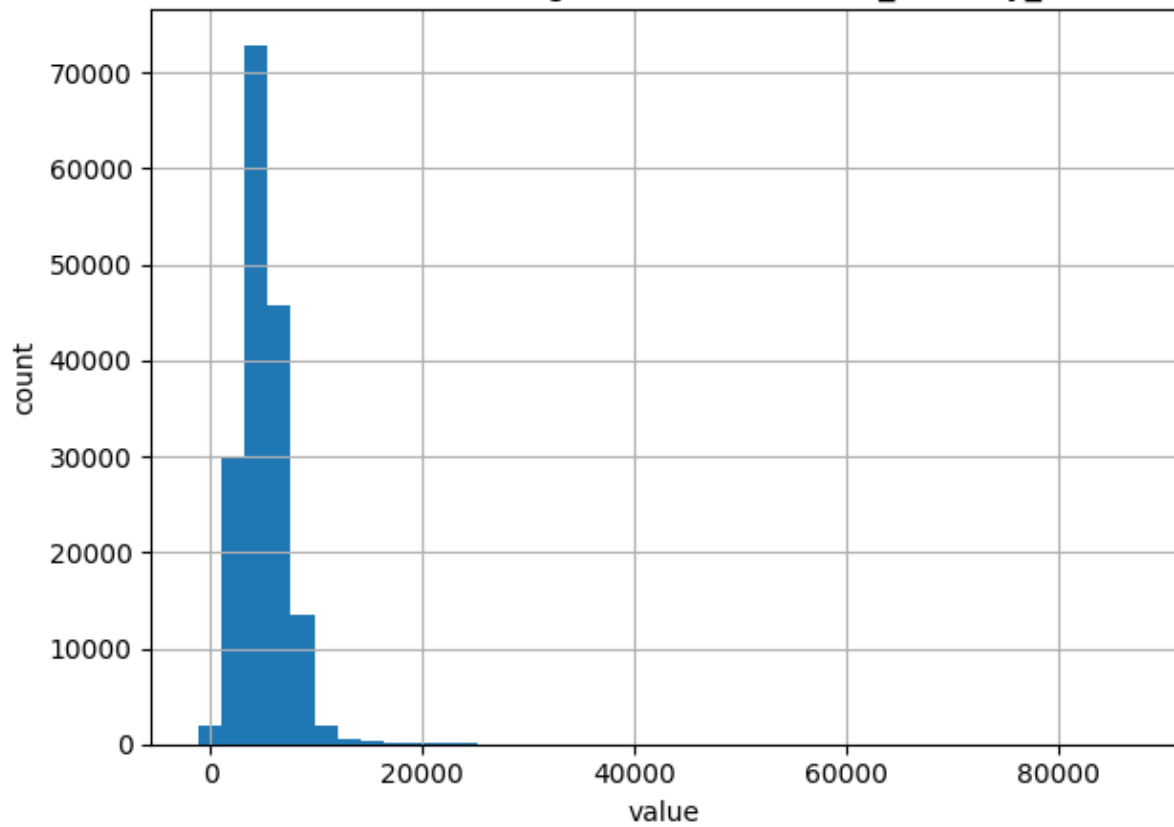
ADF (sample series):

	<b>series_id</b>	<b>adf_stat</b>	<b>p_value</b>	<b>n</b>
<b>0</b>	T1	-1.111719	0.710376	128
<b>1</b>	T10	-1.147632	0.695708	218
<b>2</b>	T100	-1.466712	0.549909	197
<b>3</b>	T1000	-0.256704	0.931416	134
<b>4</b>	T1001	-1.104724	0.713185	134

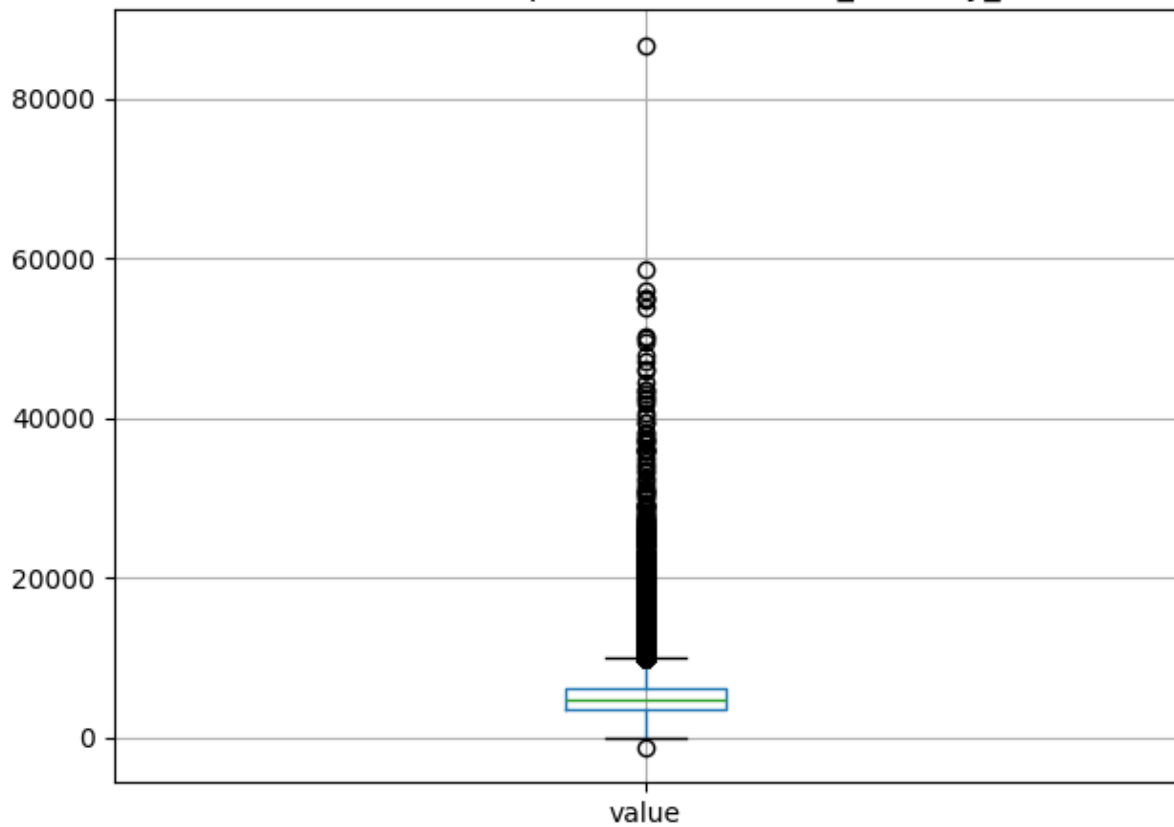




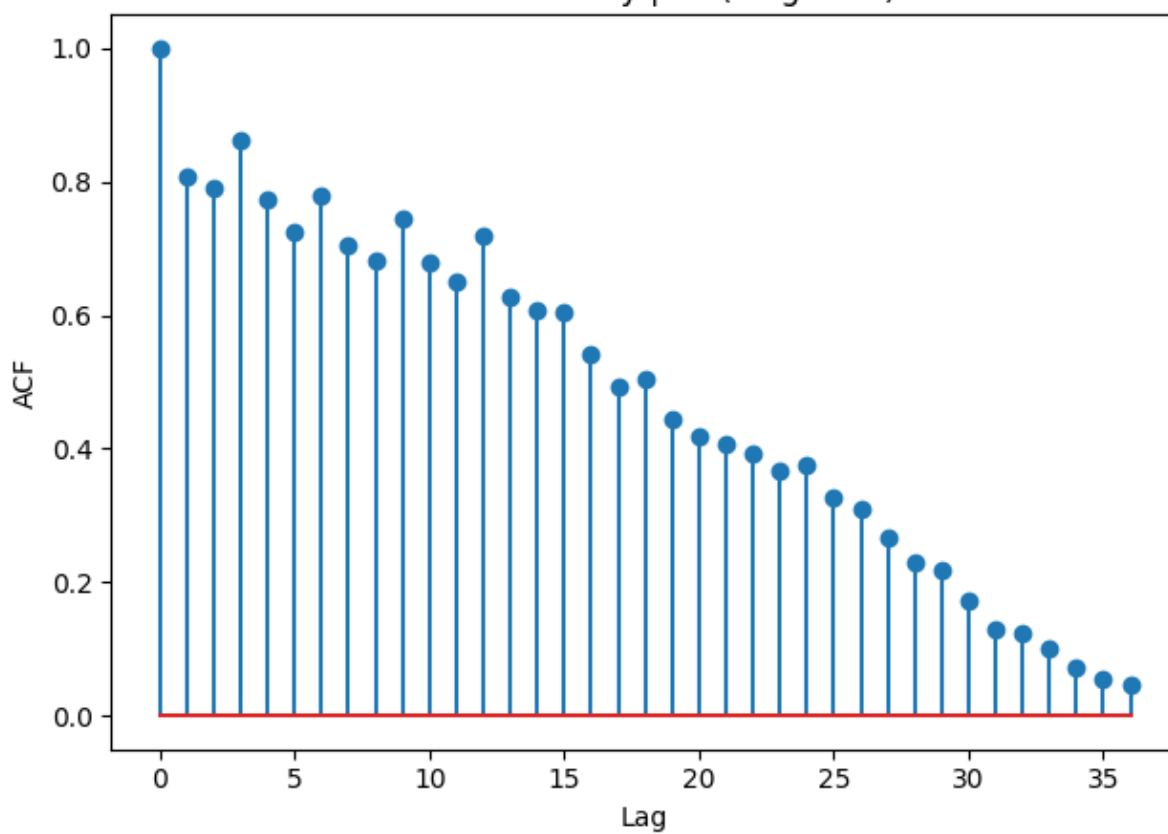
MONTHLY — Value histogram — domain=m3\_monthly\_dataset



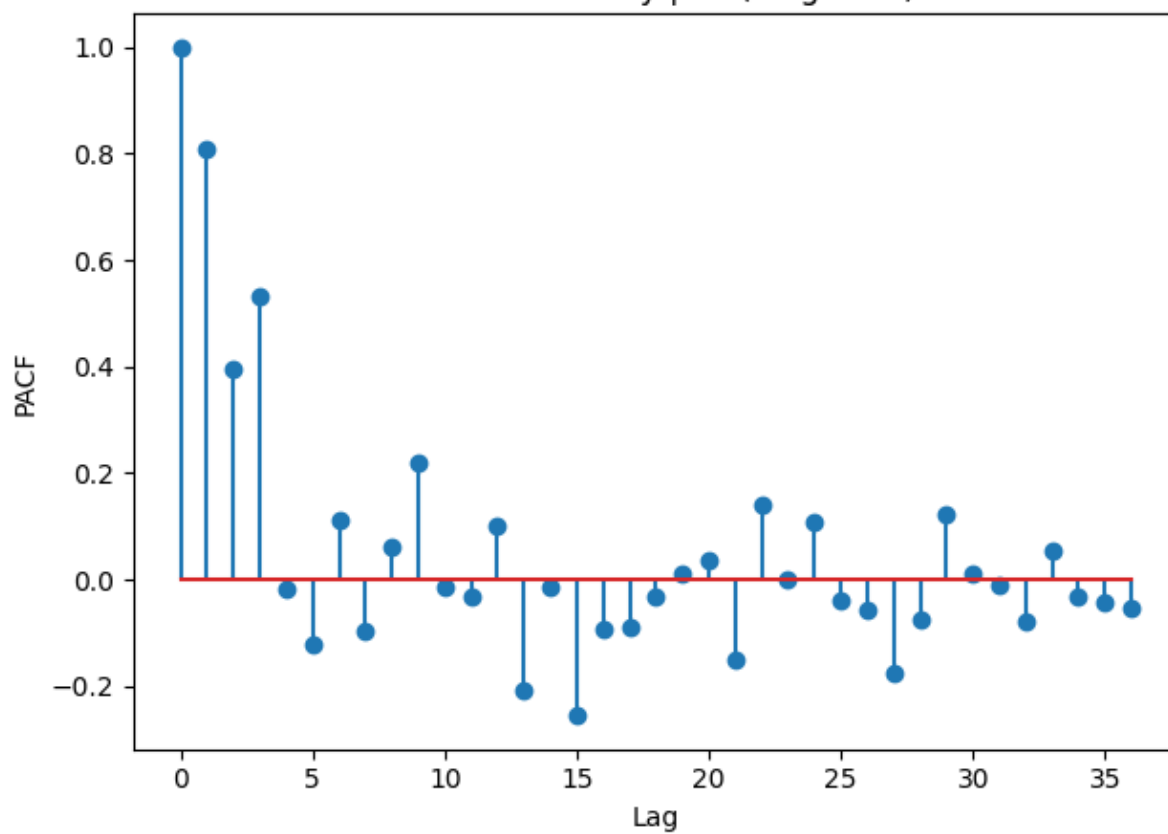
MONTHLY — Value boxplot — domain=m3\_monthly\_dataset



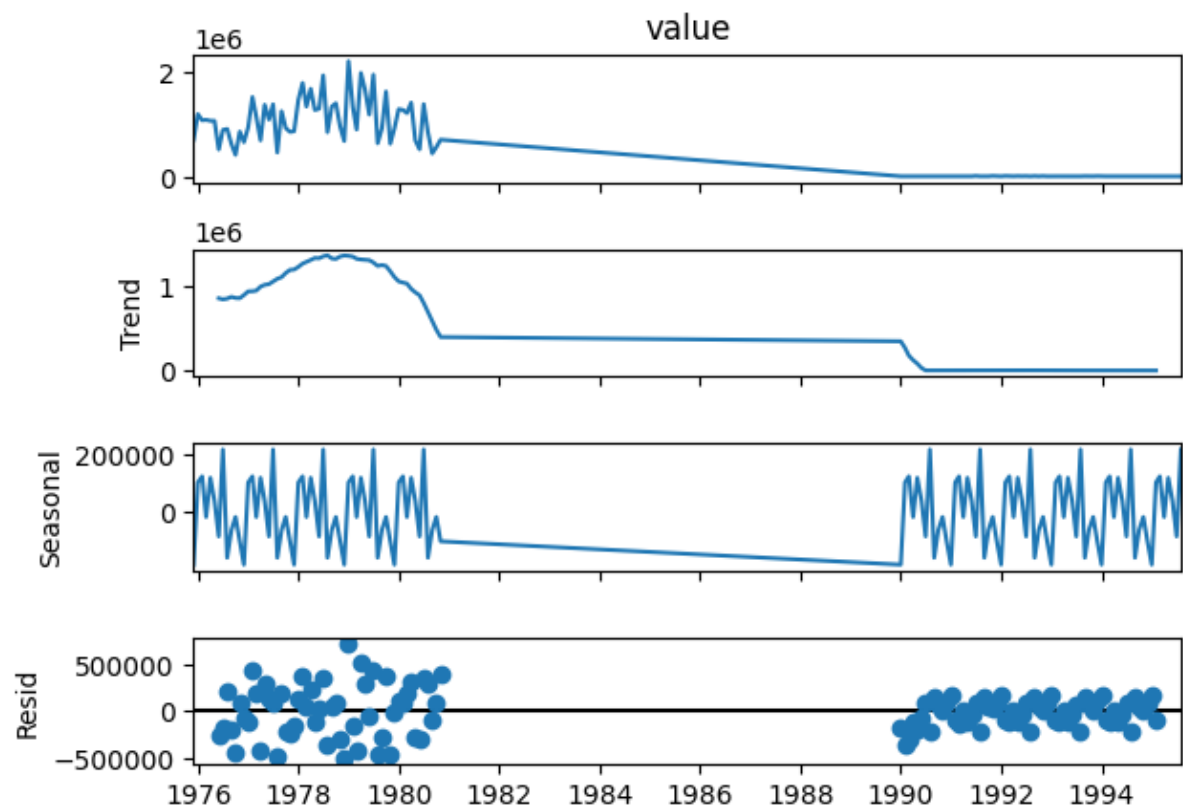
ACF — monthly | T1 (nlags=36)



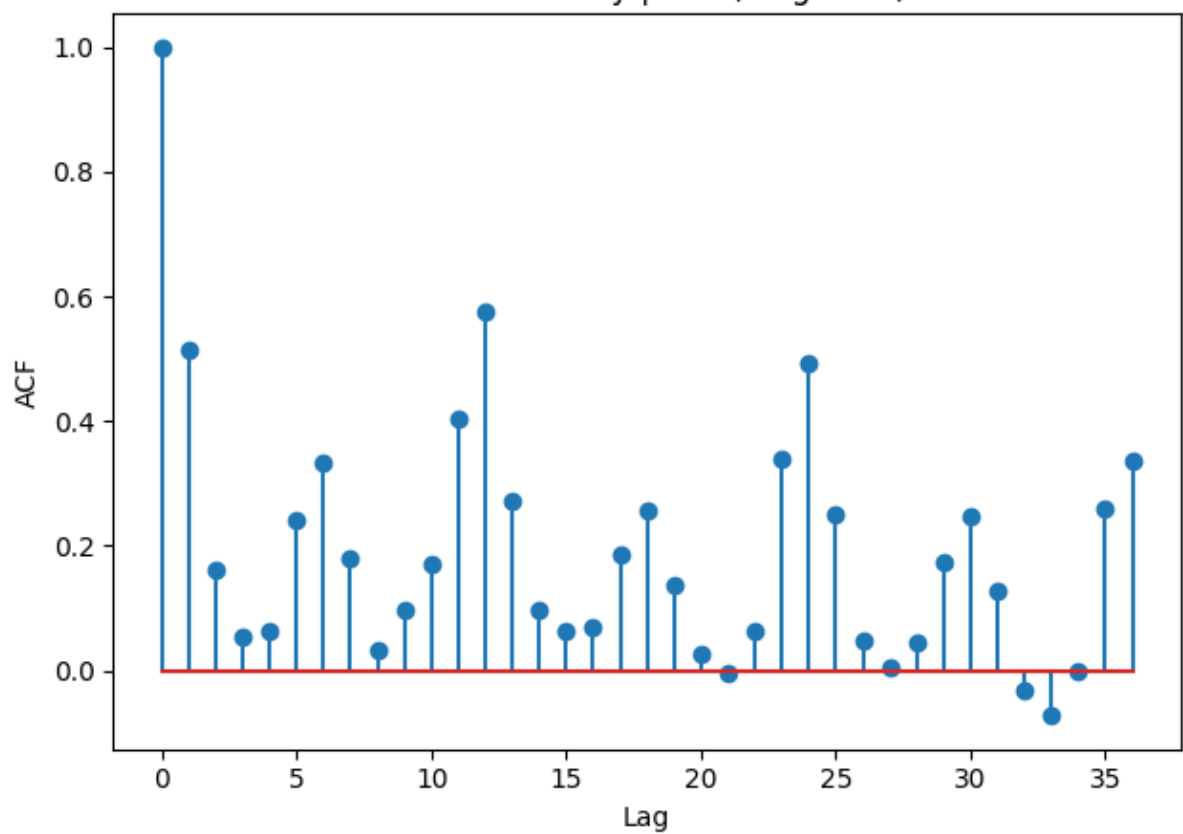
PACF — monthly | T1 (nlags=36)

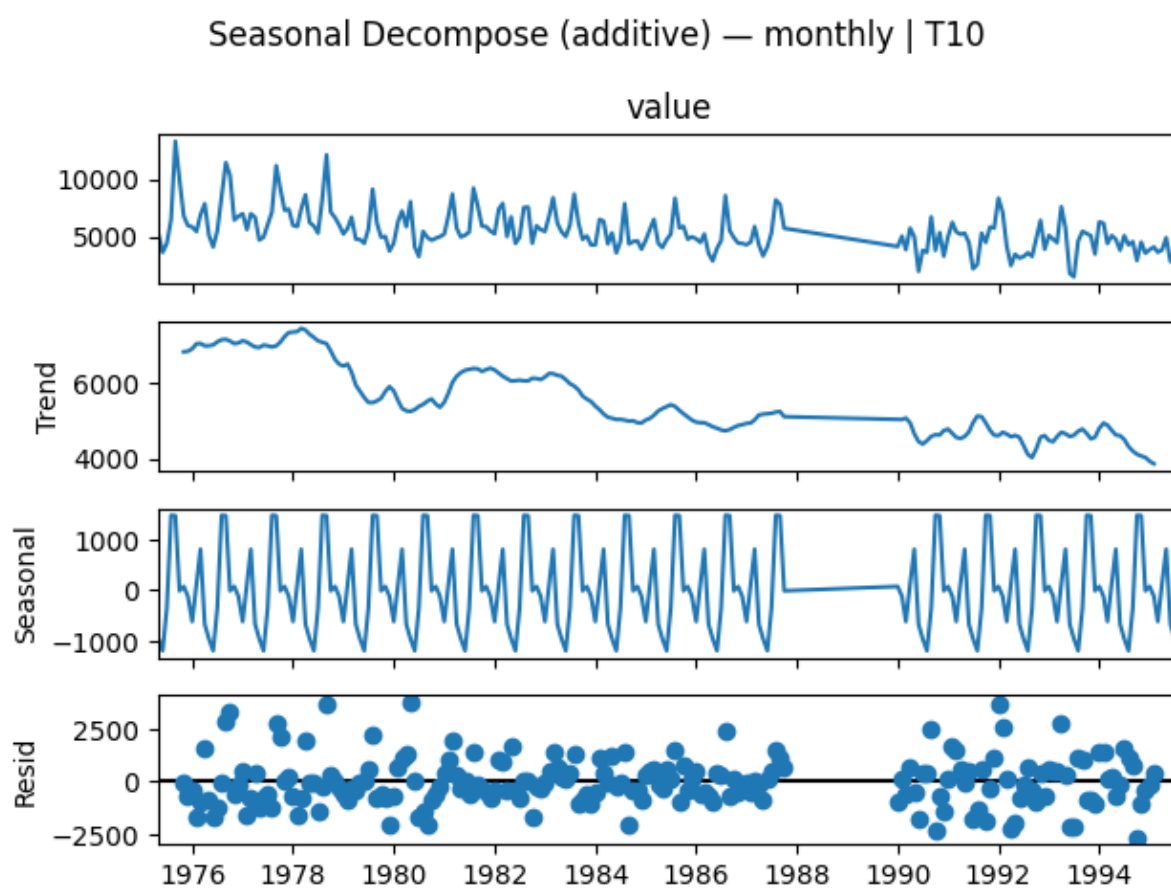
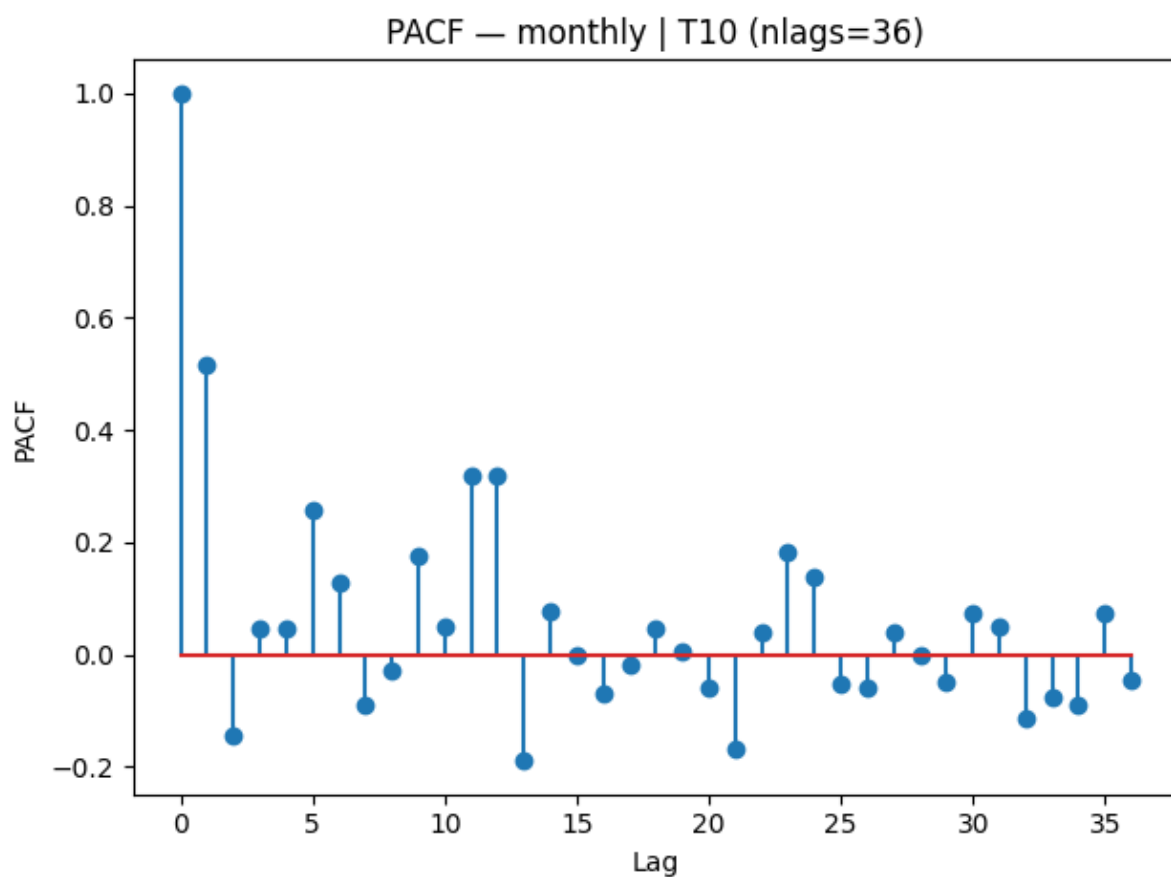


### Seasonal Decompose (additive) — monthly | T1



### ACF — monthly | T10 (nlags=36)





=====

[EXTENDED EDA] Frequency = quarterly

Domain coverage:

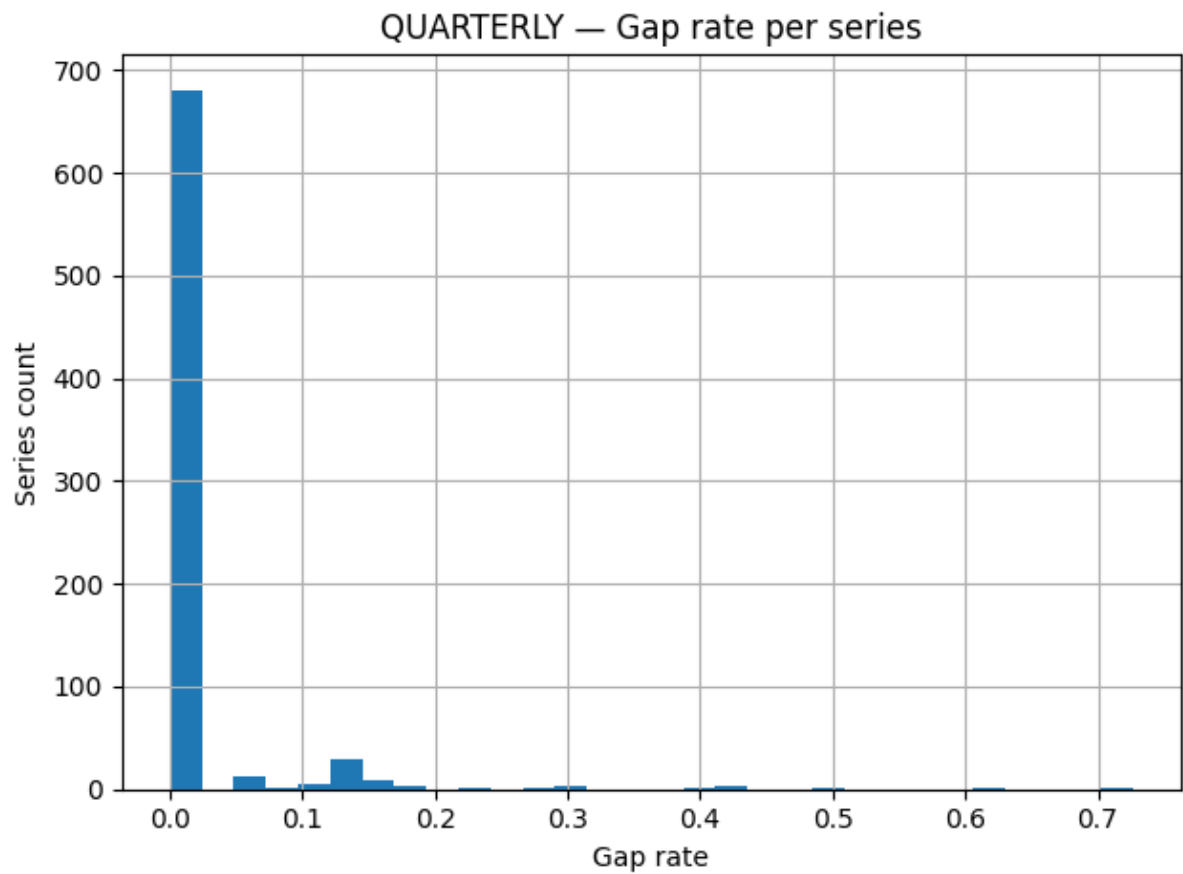
	domain	min	max	count	coverage_days
0	m1_quarterly_dataset	1900-01-01	2005-10-01	9944	38624
1	m3_quarterly_dataset	1946-01-01	1996-07-01	37004	18444

Overlap window: start=1946-01-01 00:00:00, end=1996-07-01 00:00:00, days=18444

Gaps summary (per series):

	gap_rate
count	756.000000
mean	0.017798
std	0.068906
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	0.726316

**dtype:** float64



Negatives/outliers summary:

	neg_rate	outlier_rate
count	756.0	756.000000
mean	0.0	0.008007
std	0.0	0.025515
min	0.0	0.000000
25%	0.0	0.000000
50%	0.0	0.000000
75%	0.0	0.000000

	<b>neg_rate</b>	<b>outlier_rate</b>
<b>max</b>	0.0	0.200000

Volatility & growth summary:

	<b>cv</b>	<b>median_abs_diff</b>	<b>median_pct_growth</b>
<b>count</b>	756.000000	756.000000	756.000000
<b>mean</b>	0.400810	541.625701	0.010128
<b>std</b>	0.382956	2103.980840	0.037930
<b>min</b>	0.007597	1.400000	-0.473615
<b>25%</b>	0.112139	71.650000	0.002936
<b>50%</b>	0.228764	157.000000	0.009734
<b>75%</b>	0.628536	334.130000	0.019762
<b>max</b>	1.443269	40025.000000	0.218751

Spike summary:

	<b>n_spikes</b>	<b>spike_rate</b>
<b>count</b>	756.000000	756.000000
<b>mean</b>	0.378307	0.005934
<b>std</b>	0.725874	0.010843
<b>min</b>	0.000000	0.000000
<b>25%</b>	0.000000	0.000000

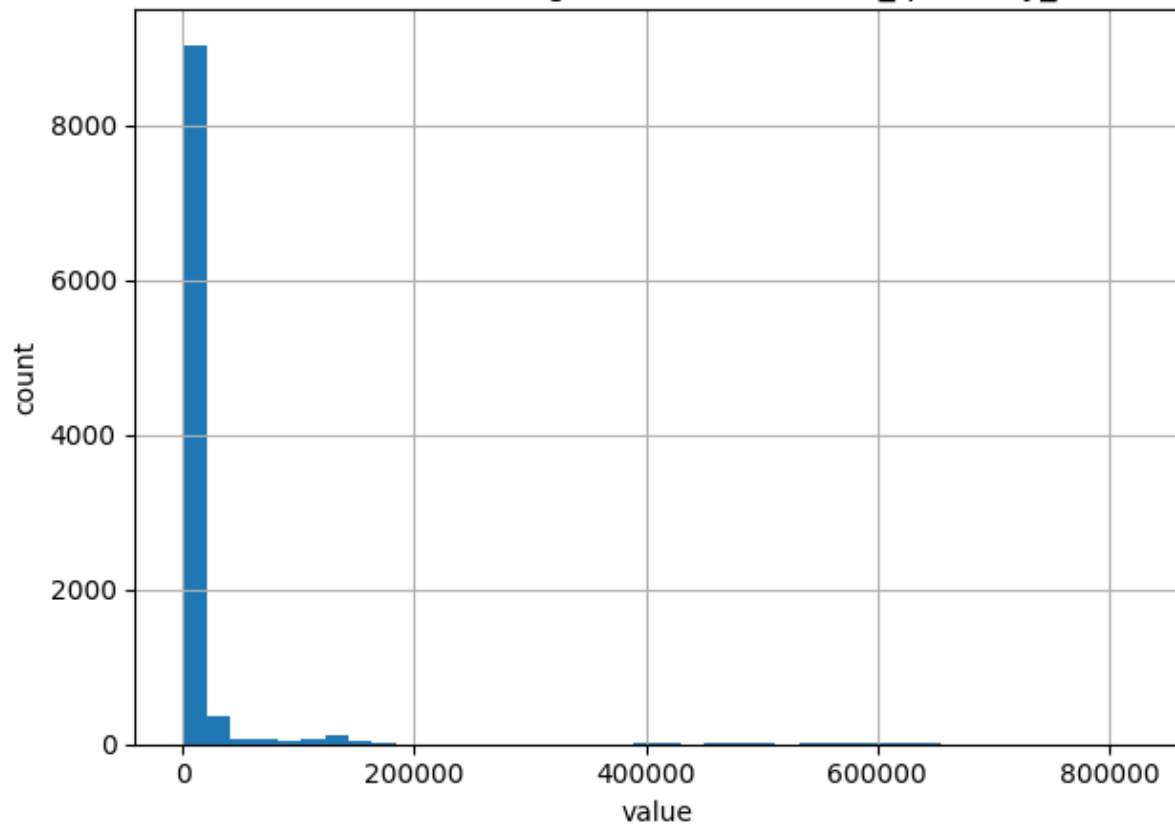
	<b>n_spikes</b>	<b>spike_rate</b>
<b>50%</b>	0.000000	0.000000
<b>75%</b>	1.000000	0.011494
<b>max</b>	5.000000	0.070423

ADF (sample series):

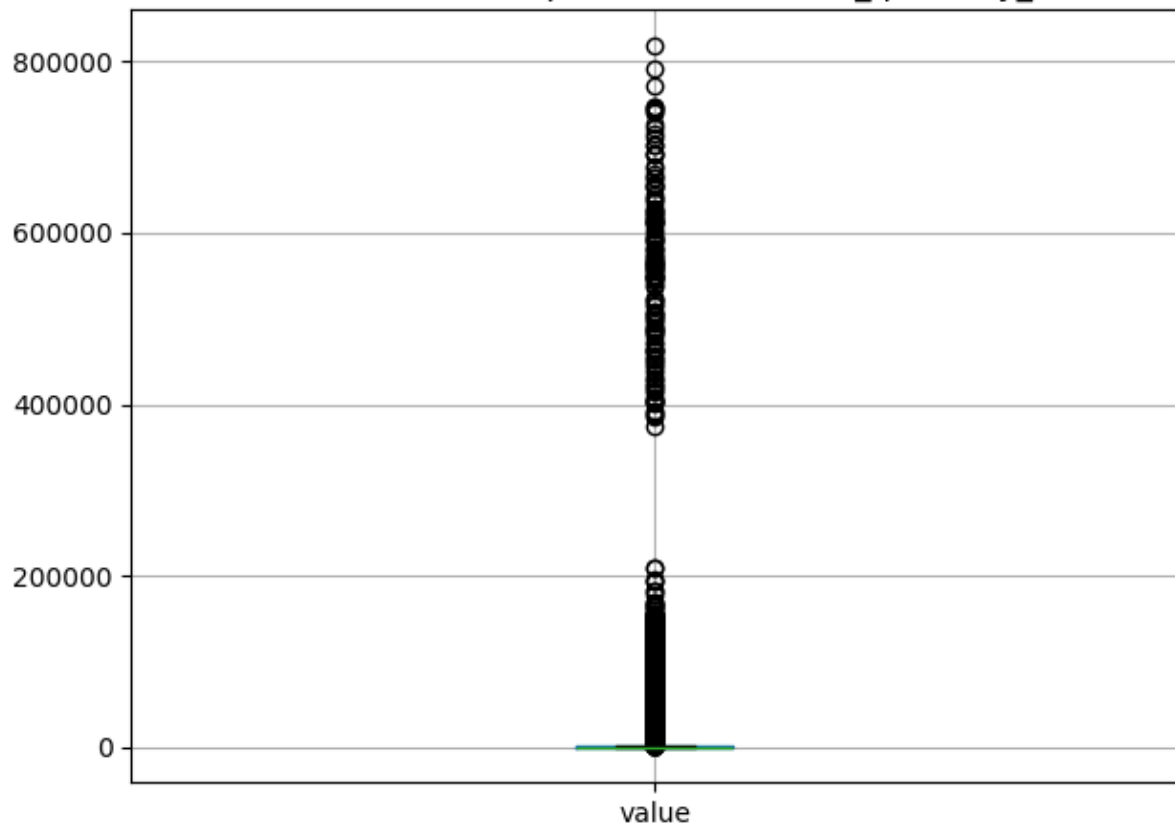
	<b>series_id</b>	<b>adf_stat</b>	<b>p_value</b>	<b>n</b>
<b>0</b>	T1	-0.454433	0.900606	92
<b>1</b>	T10	-2.779194	0.061308	104
<b>2</b>	T100	0.718775	0.990205	104
<b>3</b>	T101	-0.614836	0.867594	70
<b>4</b>	T102	-1.702203	0.429978	128

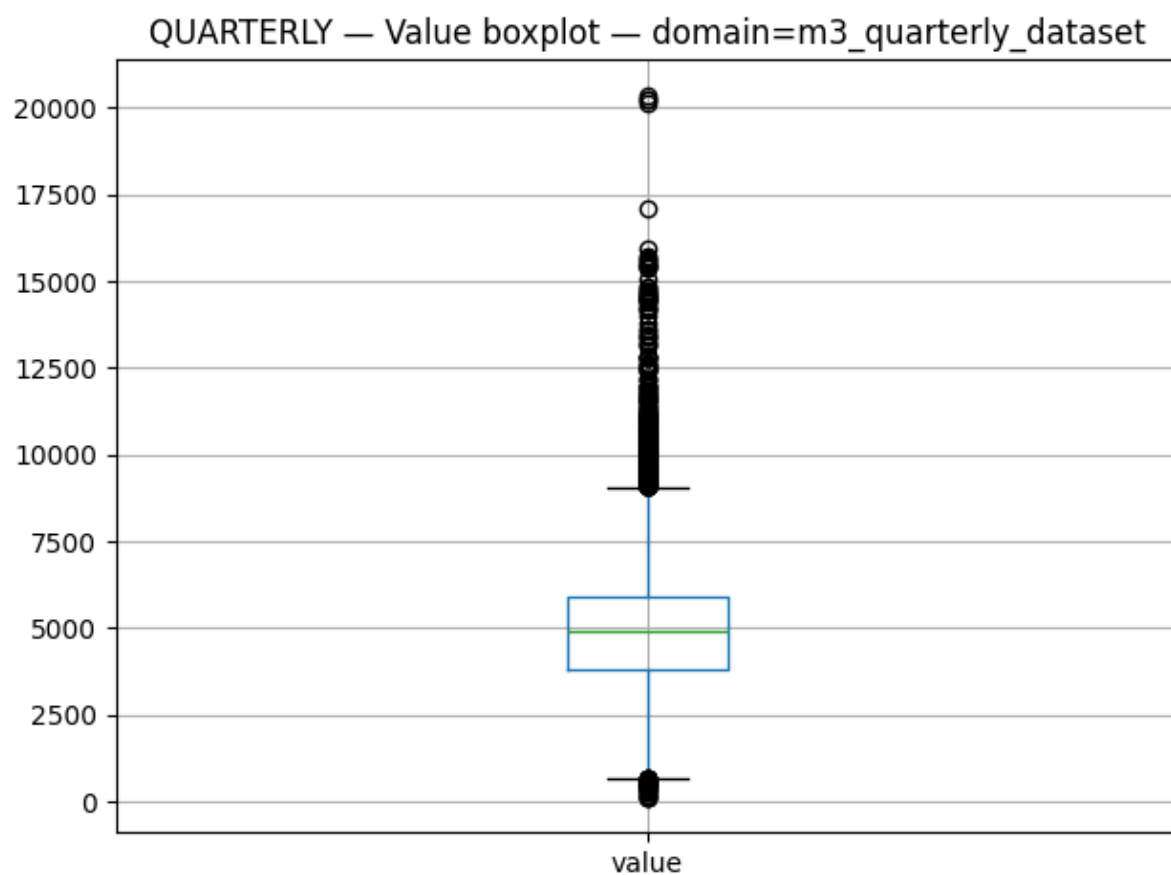
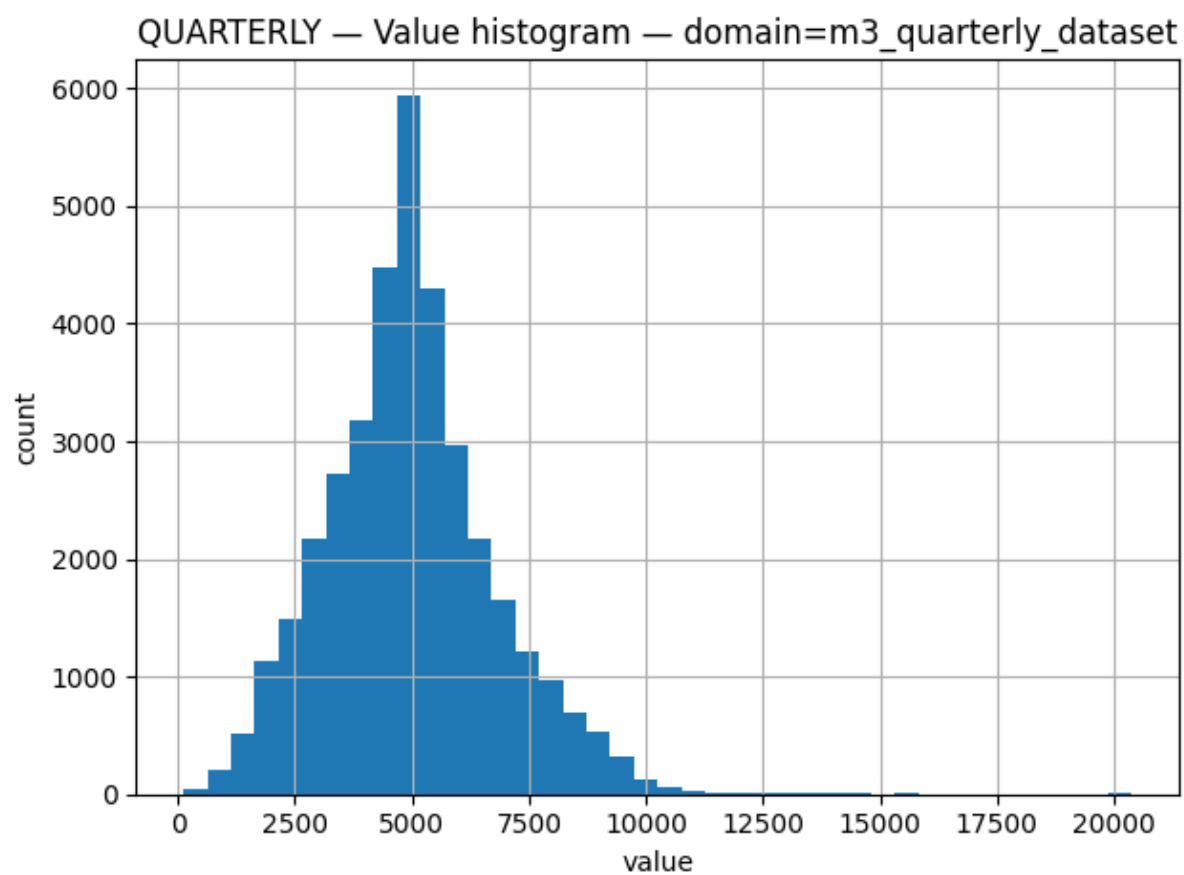


QUARTERLY — Value histogram — domain=m1\_quarterly\_dataset

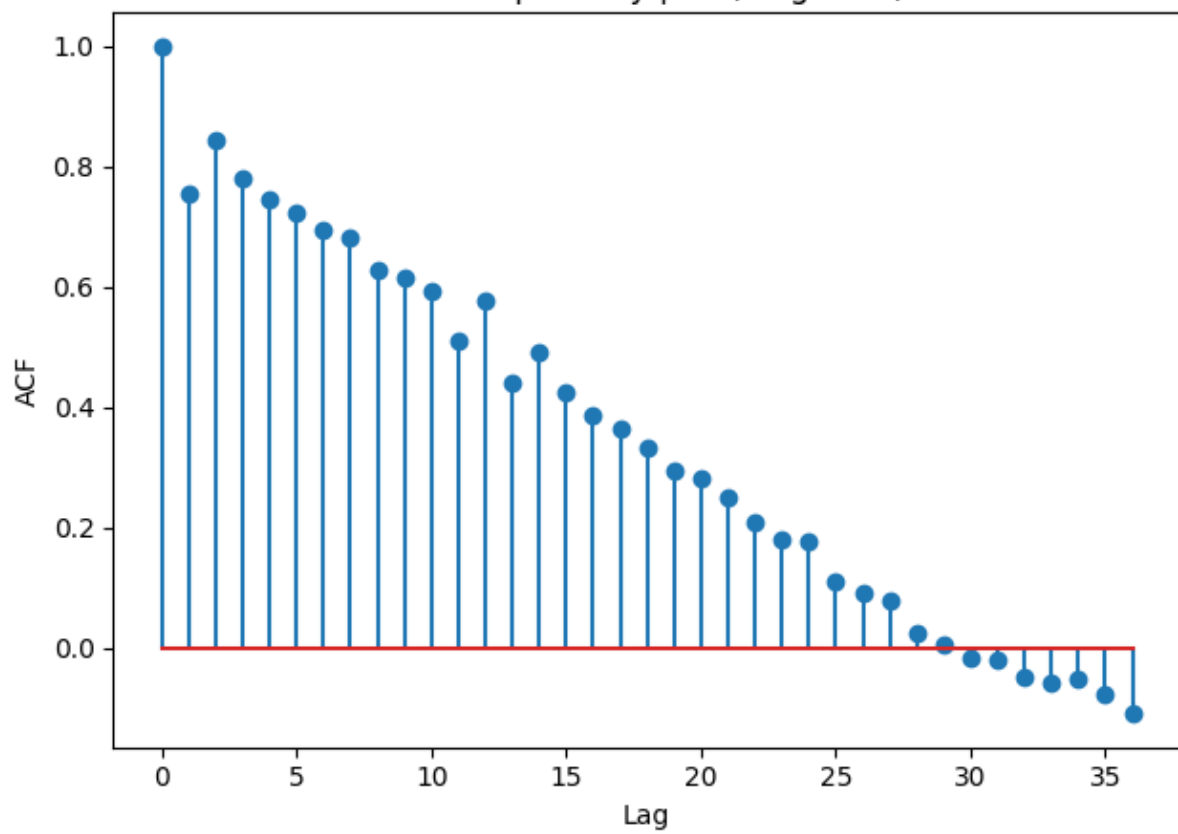


QUARTERLY — Value boxplot — domain=m1\_quarterly\_dataset

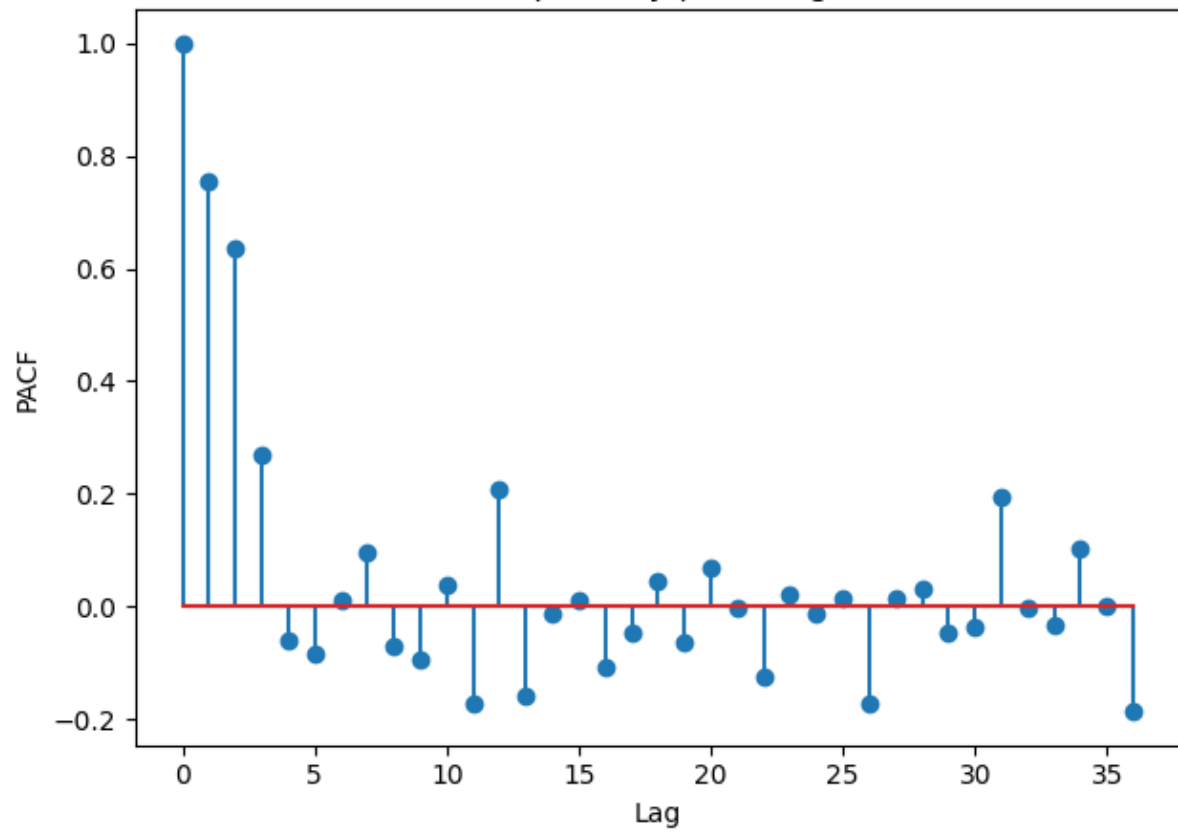




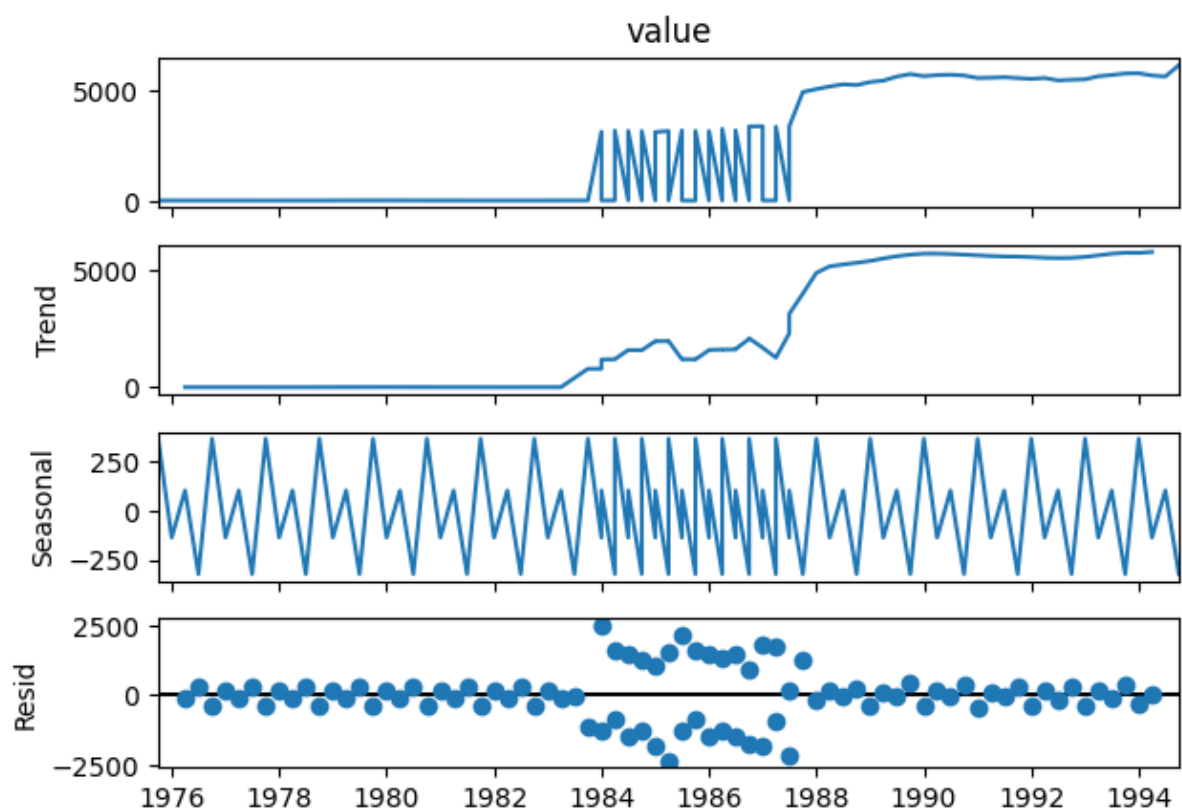
ACF — quarterly | T1 (nlags=36)



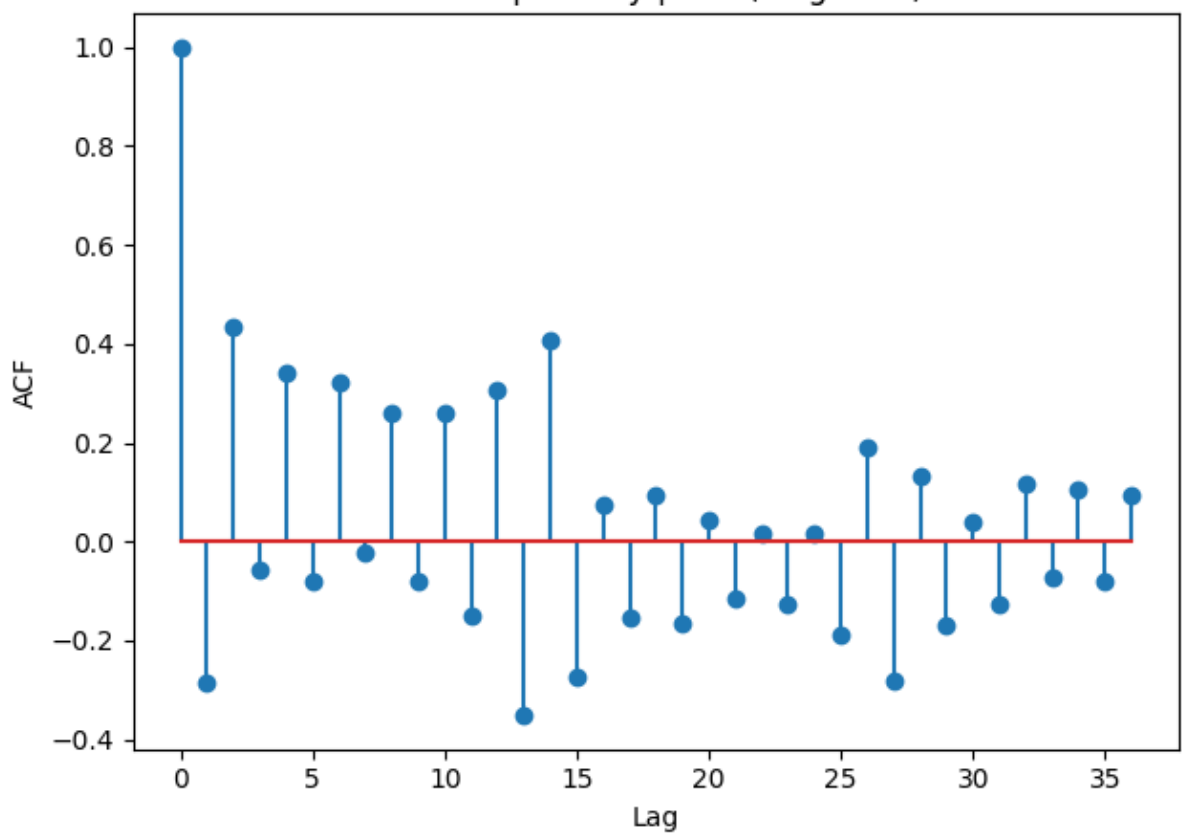
PACF — quarterly | T1 (nlags=36)

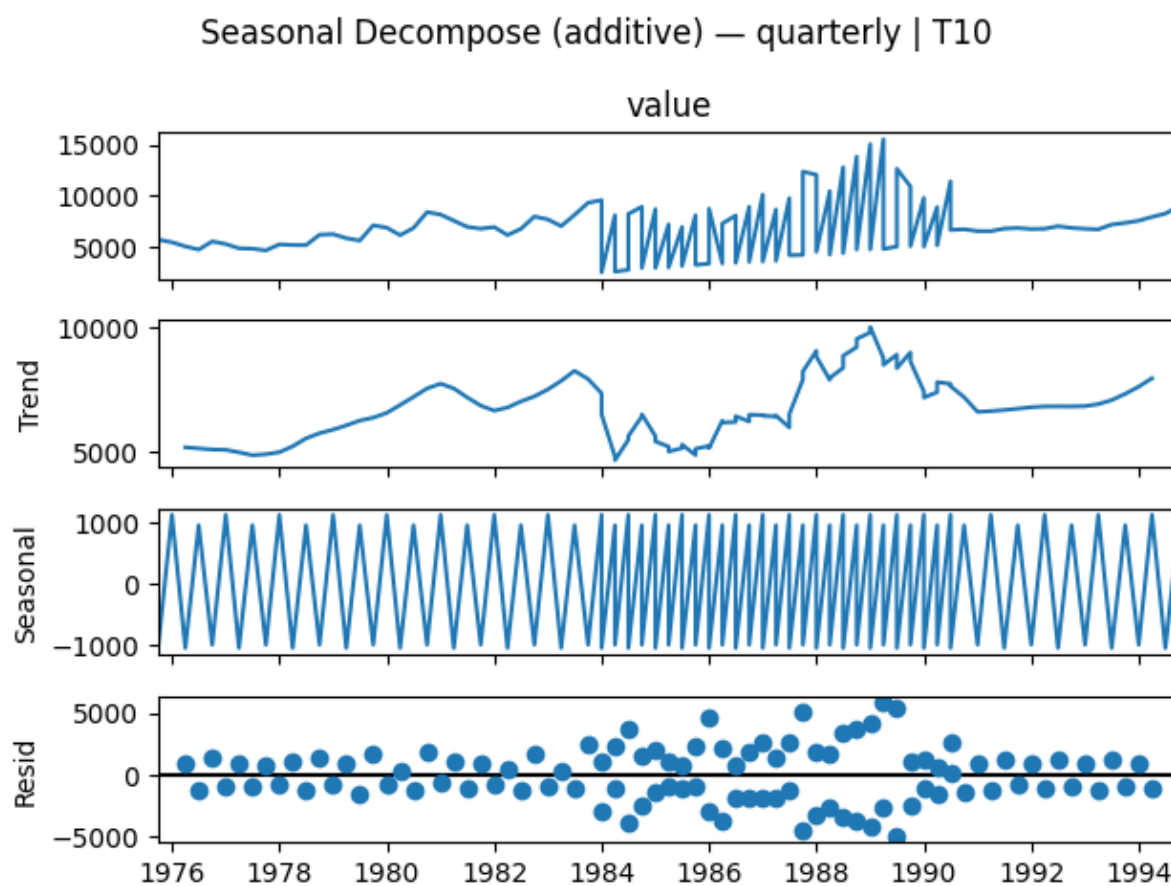
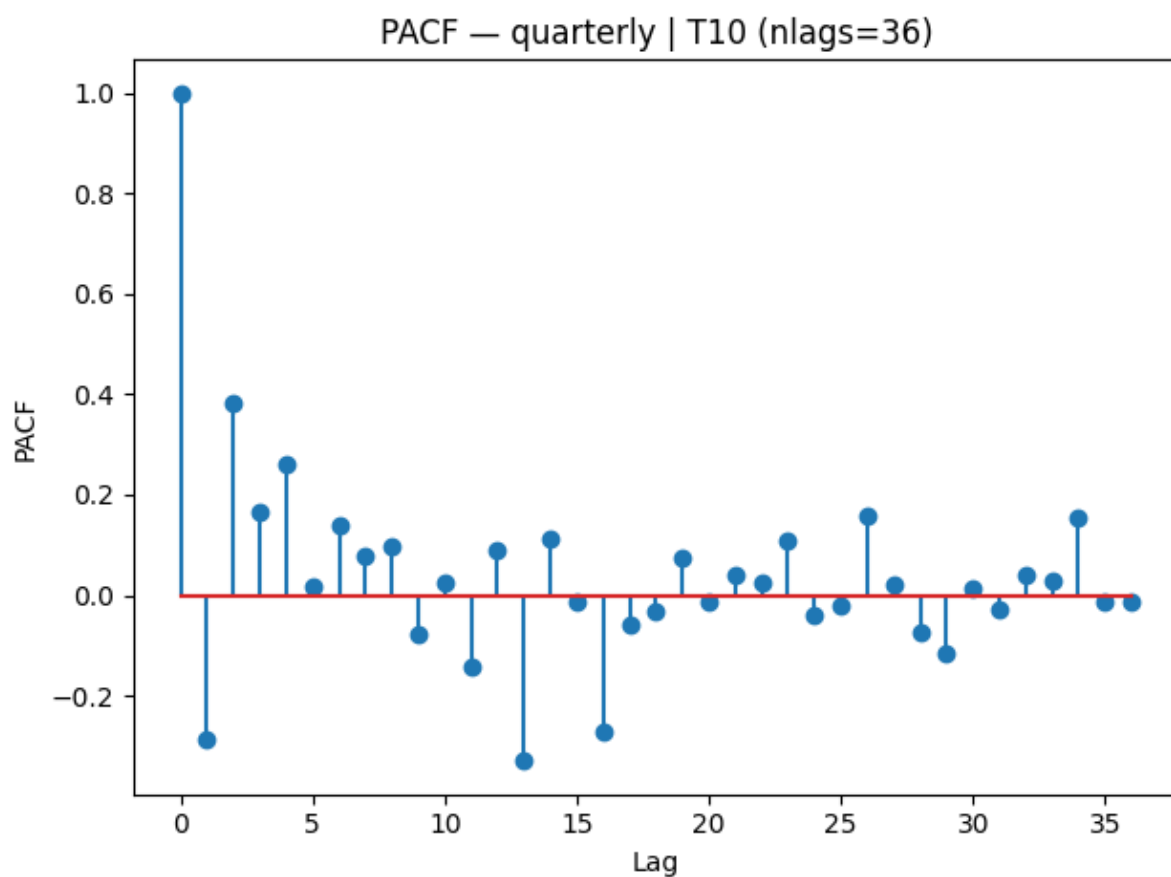


Seasonal Decompose (additive) — quarterly | T1



ACF — quarterly | T10 (nlags=36)





=====

[EXTENDED EDA] Frequency = yearly

Domain coverage:

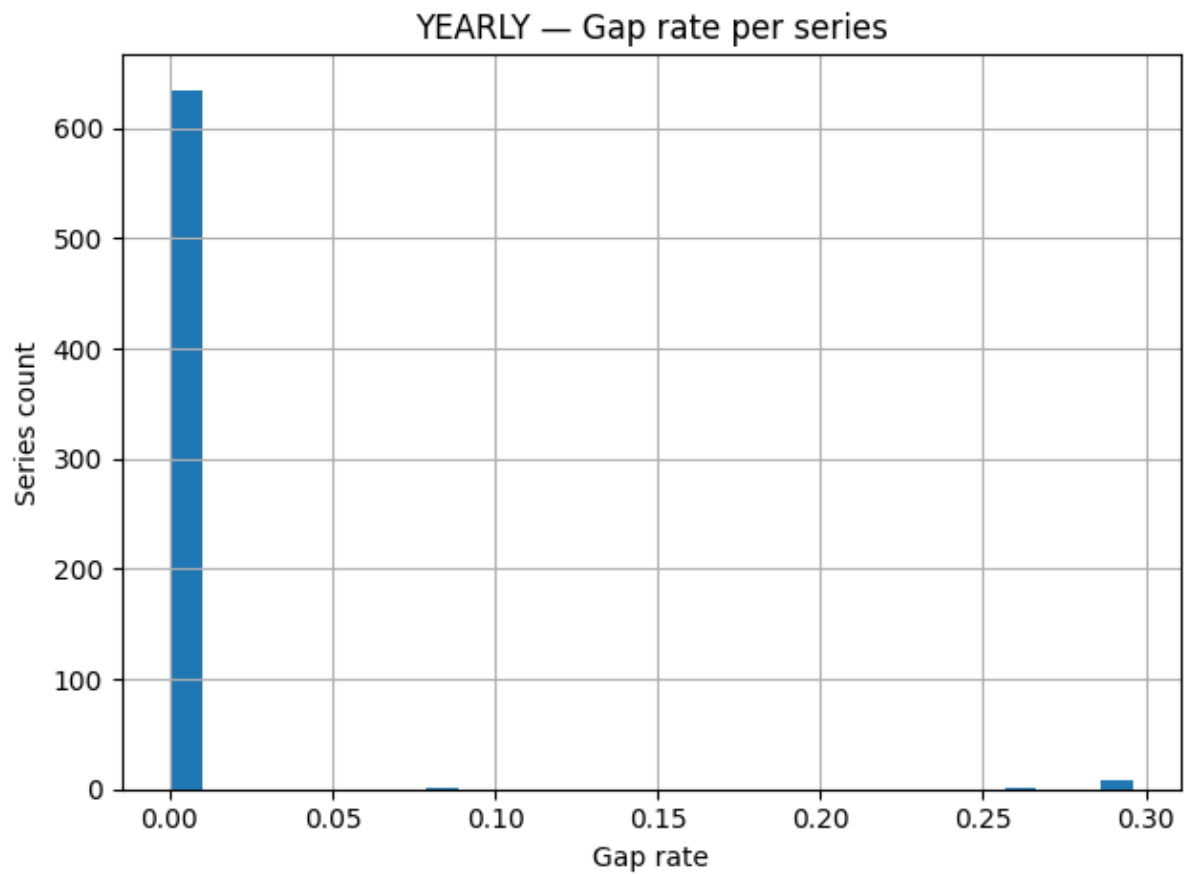
	domain	min	max	count	coverage_days
0	m1_yearly_dataset	1865-01-01	2033-01-01	4515	61361
1	m3_yearly_dataset	1811-01-01	2001-01-01	18319	69397

Overlap window: start=1865-01-01 00:00:00, end=2001-01-01 00:00:00, days=49673

Gaps summary (per series):

	gap_rate
count	645.000000
mean	0.004130
std	0.033776
min	0.000000
25%	0.000000
50%	0.000000
75%	0.000000
max	0.296000

**dtype:** float64



Negatives/outliers summary:

	neg_rate	outlier_rate
count	645.0	645.000000
mean	0.0	0.022467
std	0.0	0.048454
min	0.0	0.000000
25%	0.0	0.000000
50%	0.0	0.000000
75%	0.0	0.021277

	<b>neg_rate</b>	<b>outlier_rate</b>
<b>max</b>	0.0	0.280000

Volatility & growth summary:

	<b>cv</b>	<b>median_abs_diff</b>	<b>median_pct_growth</b>
<b>count</b>	645.000000	6.450000e+02	645.000000
<b>mean</b>	0.516226	8.677084e+04	0.033755
<b>std</b>	0.426956	1.018585e+06	0.045592
<b>min</b>	0.019624	1.550000e+01	-0.131351
<b>25%</b>	0.186188	1.687500e+02	0.007074
<b>50%</b>	0.335052	2.790000e+02	0.027619
<b>75%</b>	0.792793	9.104850e+02	0.054768
<b>max</b>	2.094457	1.602049e+07	0.277555

Spike summary:

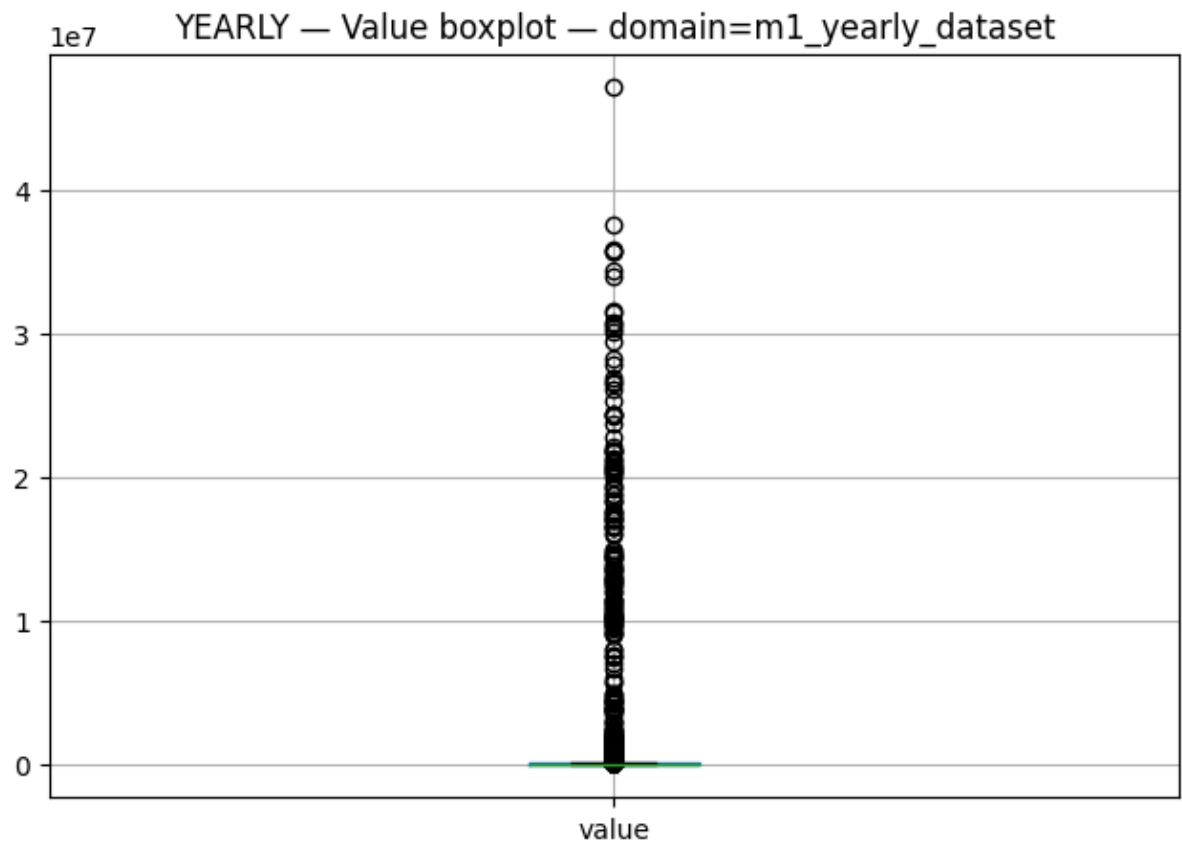
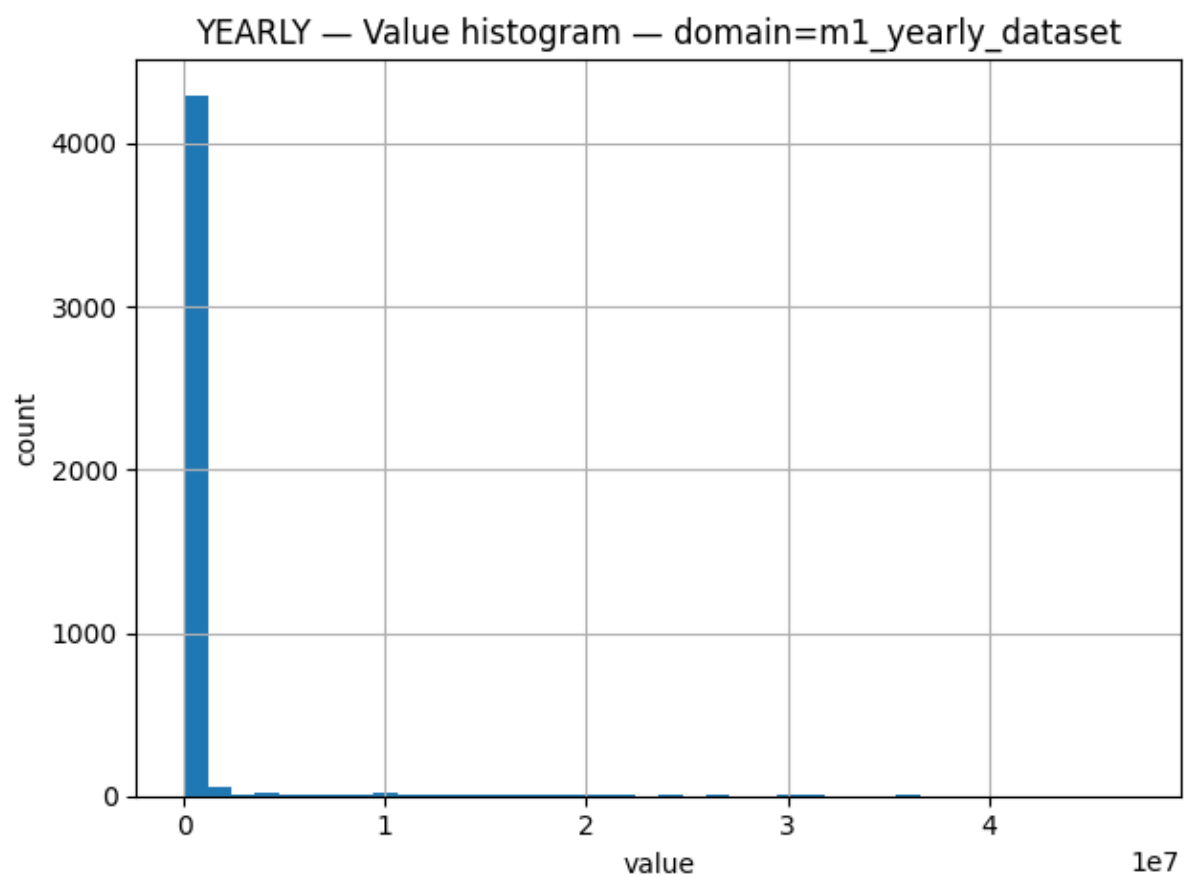
	<b>n_spikes</b>	<b>spike_rate</b>
<b>count</b>	645.000000	645.000000
<b>mean</b>	0.237209	0.006874
<b>std</b>	0.532641	0.014751
<b>min</b>	0.000000	0.000000
<b>25%</b>	0.000000	0.000000



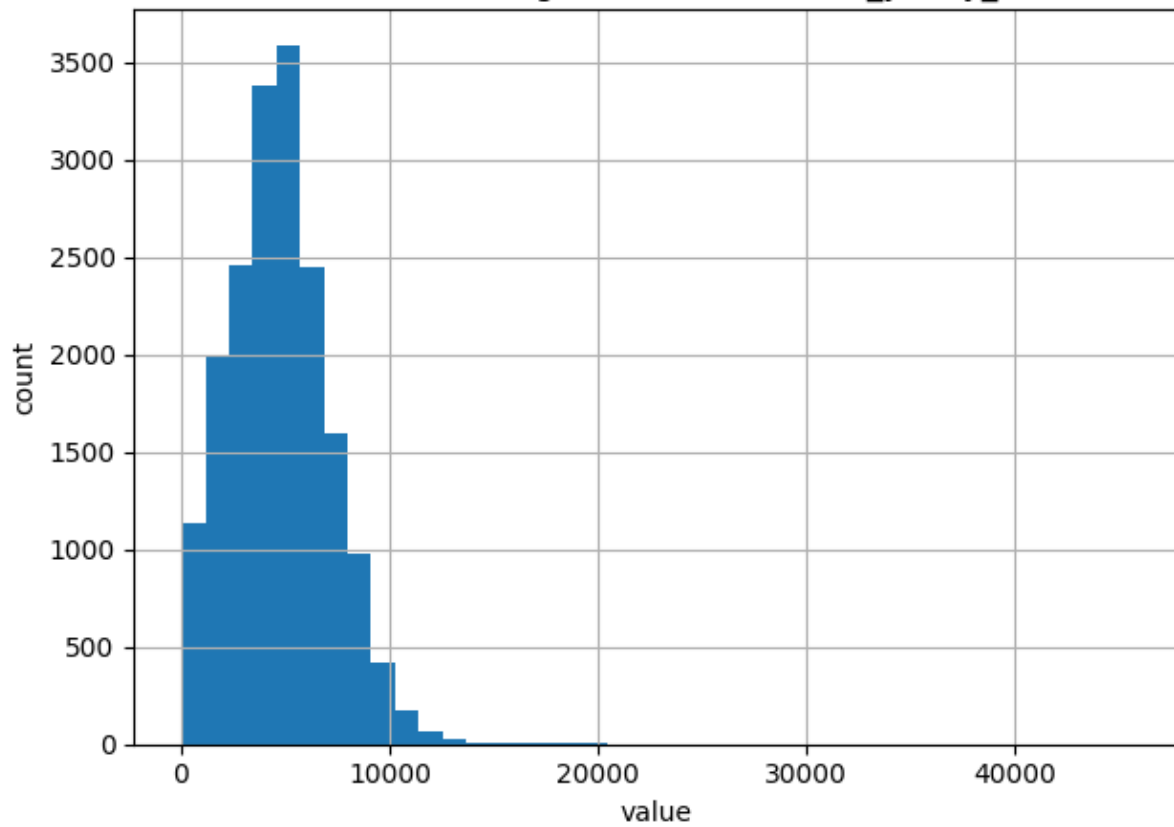
	<b>n_spikes</b>	<b>spike_rate</b>
<b>50%</b>	0.000000	0.000000
<b>75%</b>	0.000000	0.000000
<b>max</b>	4.000000	0.083333

ADF (sample series):

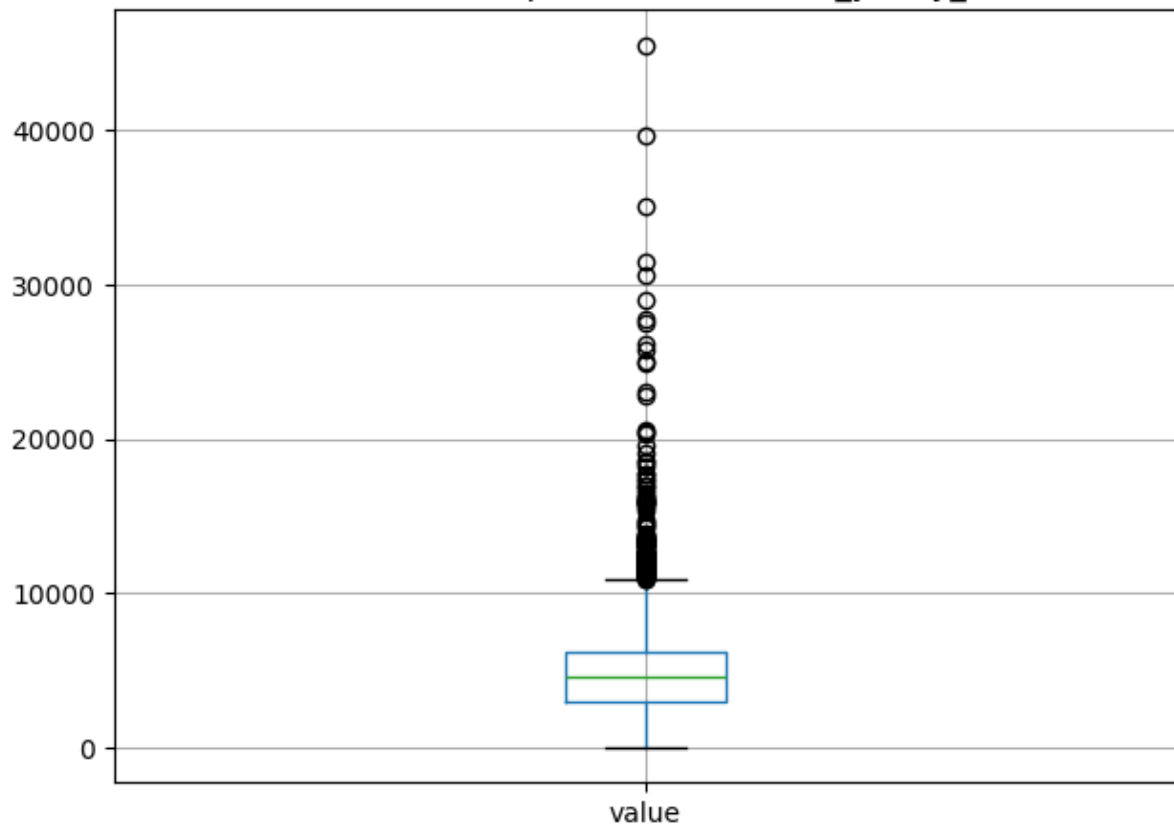
	<b>series_id</b>	<b>adf_stat</b>	<b>p_value</b>	<b>n</b>
<b>0</b>	T1	0.731148	9.904394e-01	48
<b>1</b>	T10	-0.515362	8.889837e-01	56
<b>2</b>	T100	-3.902410	2.017182e-03	63
<b>3</b>	T101	-2.250351	1.884806e-01	60
<b>4</b>	T102	-11.198655	2.272774e-20	46



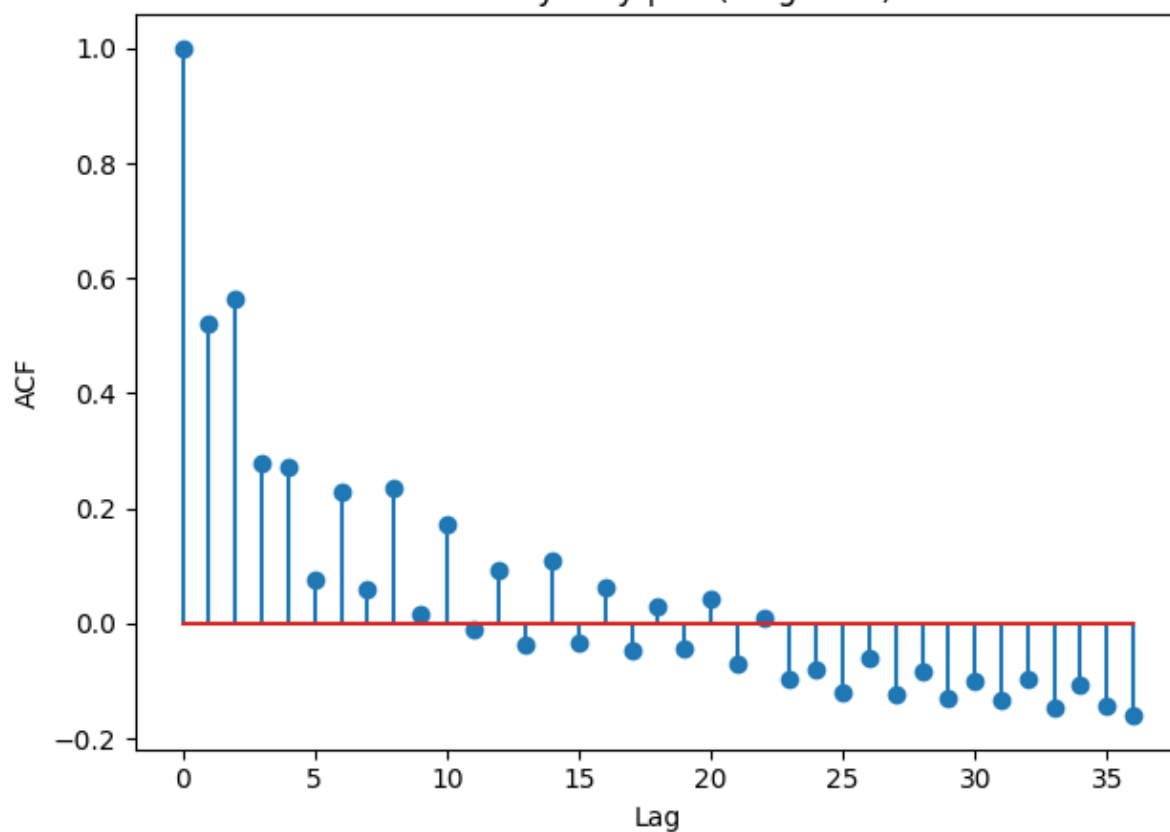
YEARLY — Value histogram — domain=m3\_yearly\_dataset



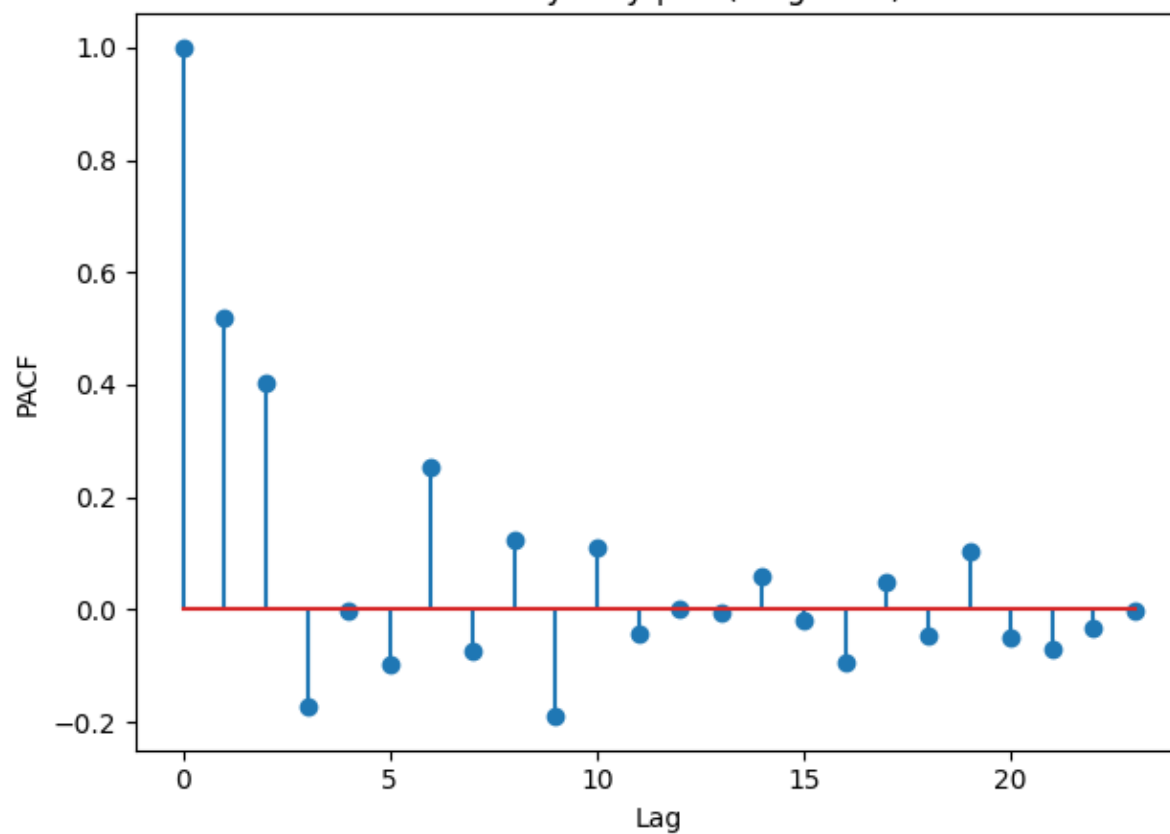
YEARLY — Value boxplot — domain=m3\_yearly\_dataset



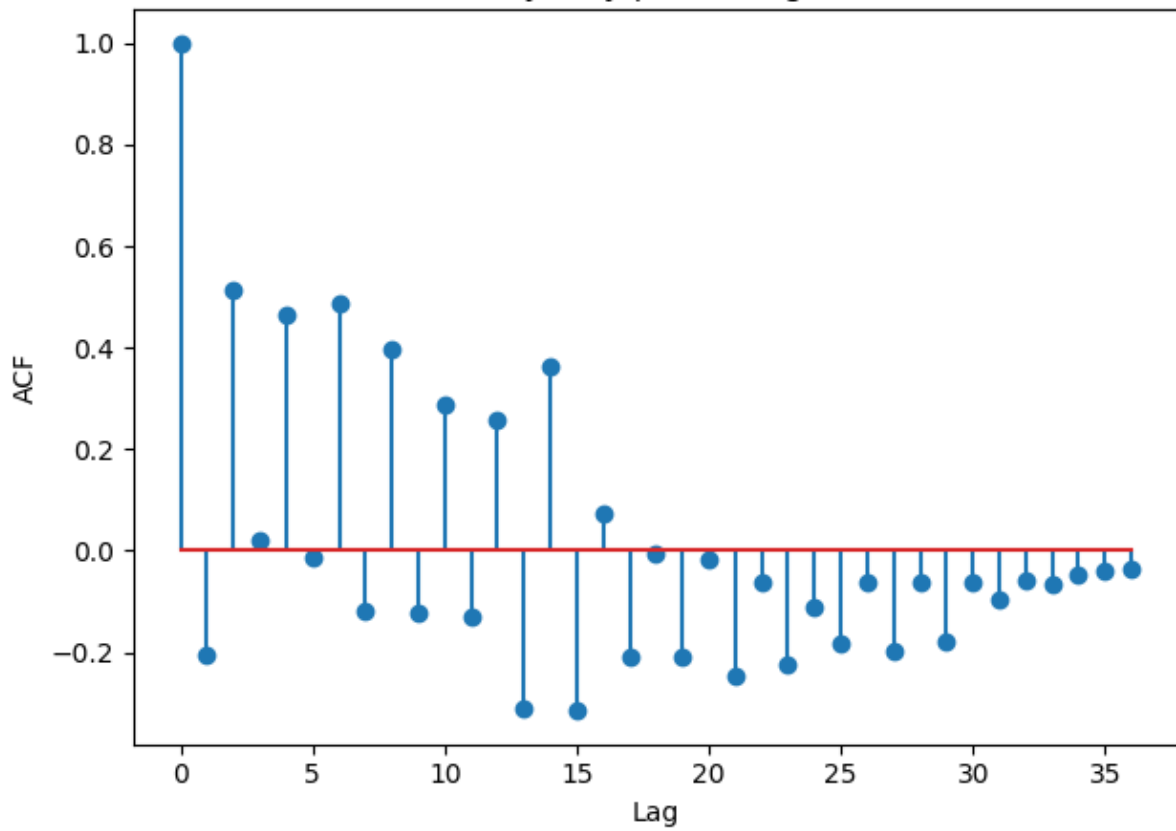
ACF — yearly | T1 (nlags=36)



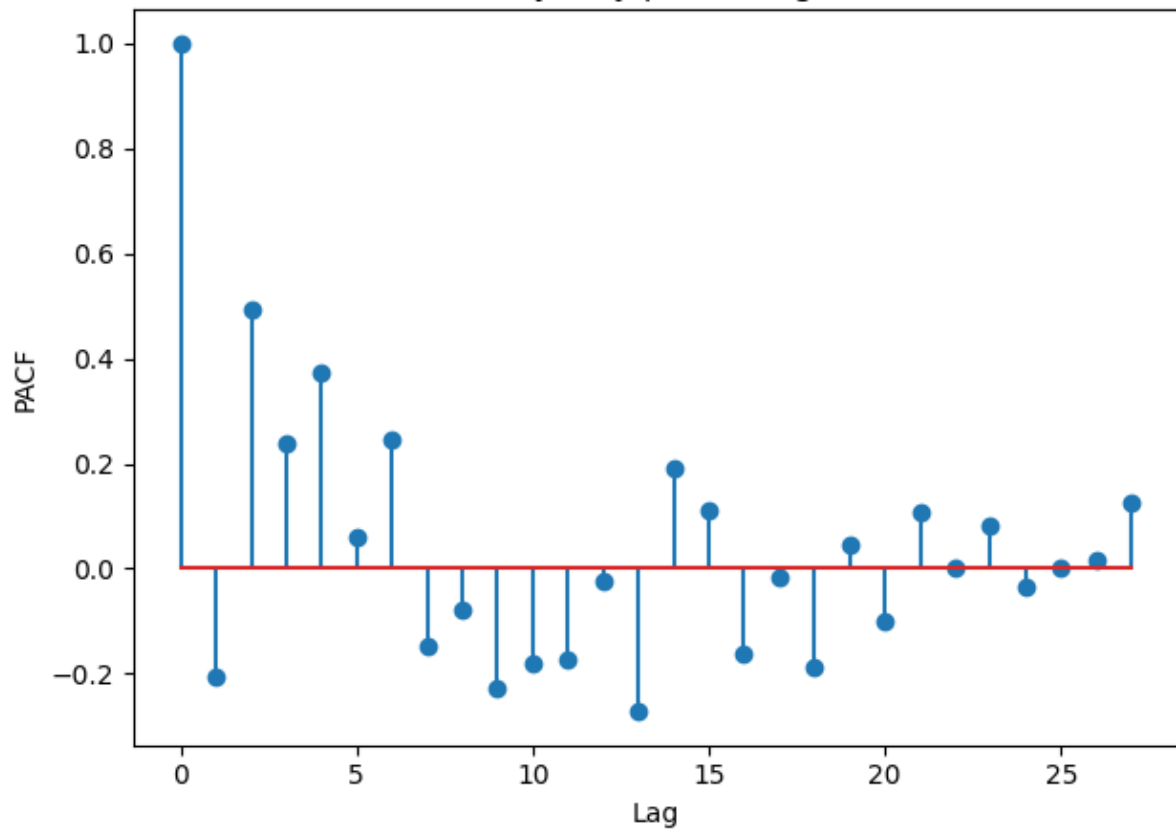
PACF — yearly | T1 (nlags=23)



ACF — yearly | T10 (nlags=36)



PACF — yearly | T10 (nlags=27)



=====

[EXTENDED EDA] Frequency = weekly

[INFO] Empty corpus — skipping.

=====

[EXTENDED EDA] Frequency = hourly

[INFO] Empty corpus — skipping.