

Neil Nordeje

Harnessing Big Data and High Performance Compute
to Understand Commercial Energy Usage Dynamics

Newfoundland, Canada
23 July, 2024

Background

#1

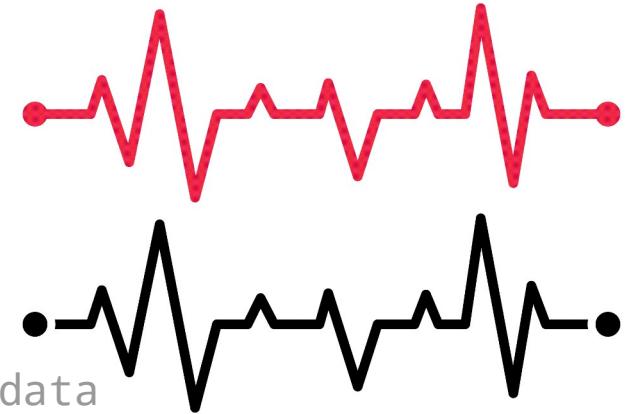
Most of us pay our power bill without looking twice. The same happens for most businesses, but what if they had a tool to empower them?

NEWFOUNDLAND POWER A FORTIS COMPANY	Buddy whaisname Service Address: <i>7 Turnip Lane</i>	Account Number 55123
PO Box 8910, St. John's, NL A1B 3P6 Power Outages & Emergencies: 1-800-474-5711 Customer Relations: 1-800-553-2803 Fax: 737-2903 Email: customerservice@newfoundlandpower.com Visit us online at newfoundlandpower.com	Rate: Domestic Service Your meter was read on: July 16 Next meter reading on or about: August 15	Billing Date July 16, 2024
Previous Balance Payments to June 17 - Thank you Previous balance owing		CR \$0.00
This Month's Electric Charges Basic Customer Charge Energy Charge: 1,231 kWh @ \$0.13256 /kWh Discount: -1.5% Subtotal electric charges Harmonized Sales Tax: 15% (10386 4831 RT0001) Total Charges		CR \$420.00
Meter Reading Information Meter number: 55123		
Date Jul 16 Jun 13	Reading 2233944	33 Days 202
Your Past Energy Usage		
Total Amount Due on or before July 30, 2024		
Total kWh 202	This Month 33	Same Month 258
Billing Days 33	Last Year 33	Average kWh/day 6.12
Please keep this portion for your records.		
Please return this portion with your payment.		
Account Number 55123	Amount Due \$420.00	Discount Date
Amount Due After Discount Date		The Power of Life Project Donation
Enter Amount Paid		

Research question

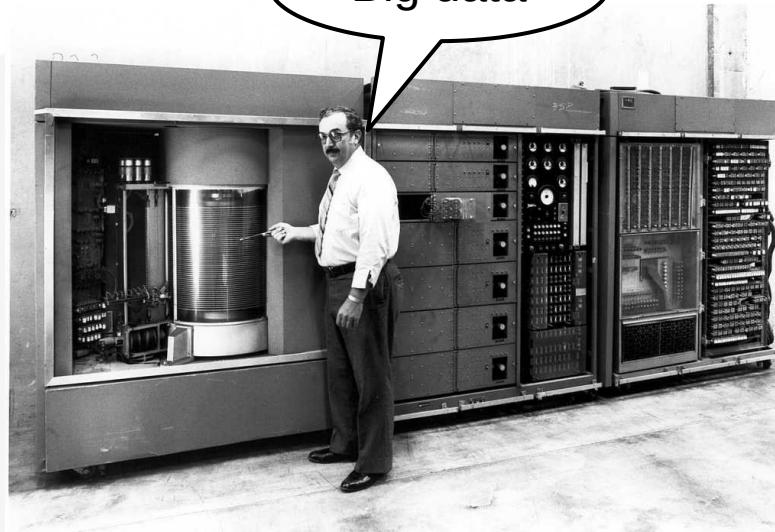
#2

"How can analyzing electricity consumption data across different sectors reveal trends and insights that can lead to more informed energy usage decisions?"



Data source

#3

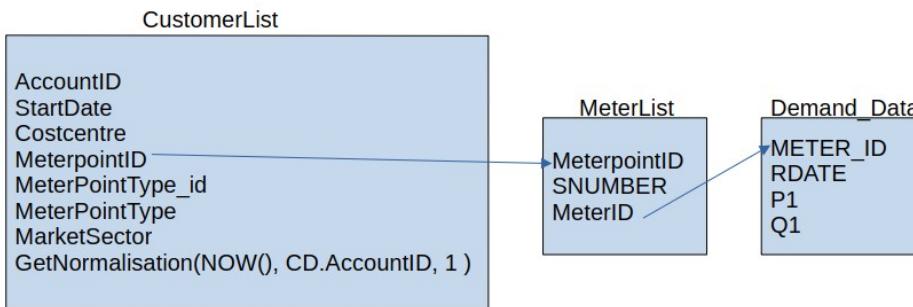


Data source

About half a billion rows of data from a SQL database

Data Structure

The database consisted of 3 tables which was inner joined to create an exportable CSV file.



Data format

The data was exported to CSV format and was almost 70GB

Data points

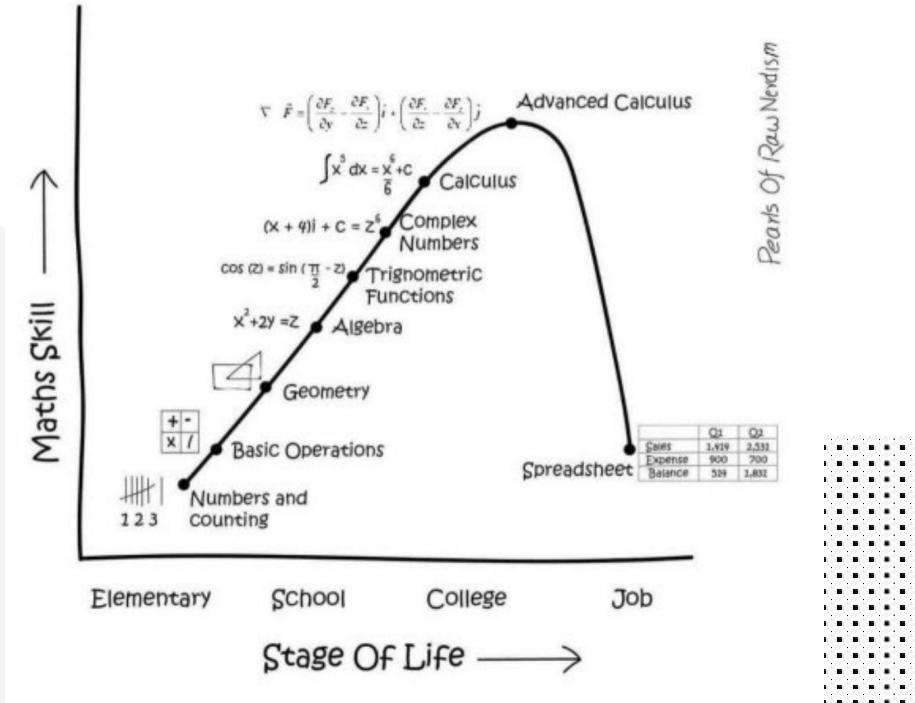
Unique numbers in the tables:
AccountID: 4301
MeterpointID: 10,848
MarketSector: 79

Important Values

Demand_Data
RDATE: 2024:07:23T18:00:00 or
Unix:1721757600
P1: kWh (Real Power)
Q1: kVAR (Reactive Power)
GetNorm: m^2 (Surface Area)

Methodology

#4



01

Pre-processing

Removed NaN rows
Removed extremes by applying IQR

03

HPC

Data intensive vs. compute intensive
Sample dataset: 2 CPUs @ 8GB RAM
Full dataset: 16 CPUs @ 24GB RAM

02

Analysis method

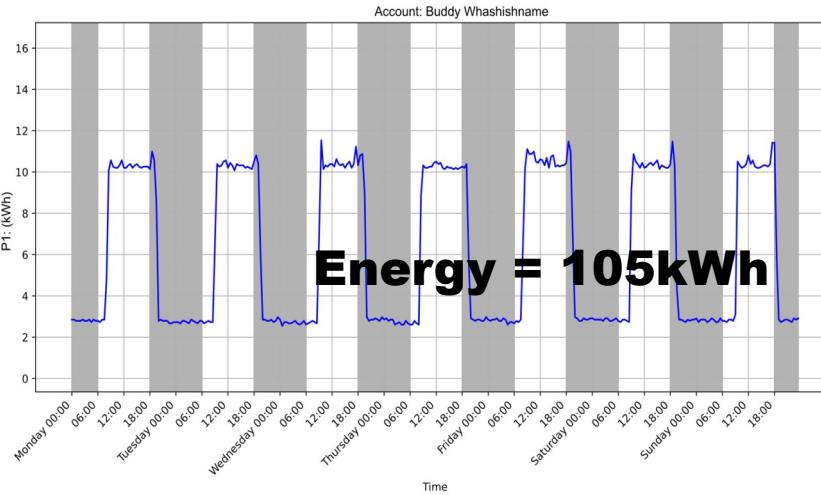
A combination of statistical histograms and time-series plots to present the energy density (Wh/m²)

04

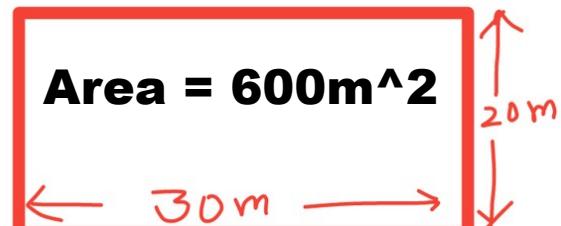
Results

Exported files:
Statistical time-series benchmark
Skewed distribution benchmark

Normalizing power data



$$= 33 \frac{\text{Wh}}{\text{m}^2}$$



Issues faced

#5



Issues faced

Speed of Dask vs SQL queries
Jupyter notebook vs Visual Studio Code

Small sample dataset
VS Code debugging
and intellisense



Scaling up

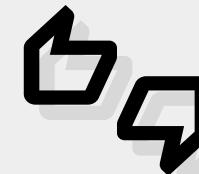
Static vs live model

Siku and Dask

Modules not available
Certain versions available



Recommendations



First Siku
environment

Second local
environment

Results

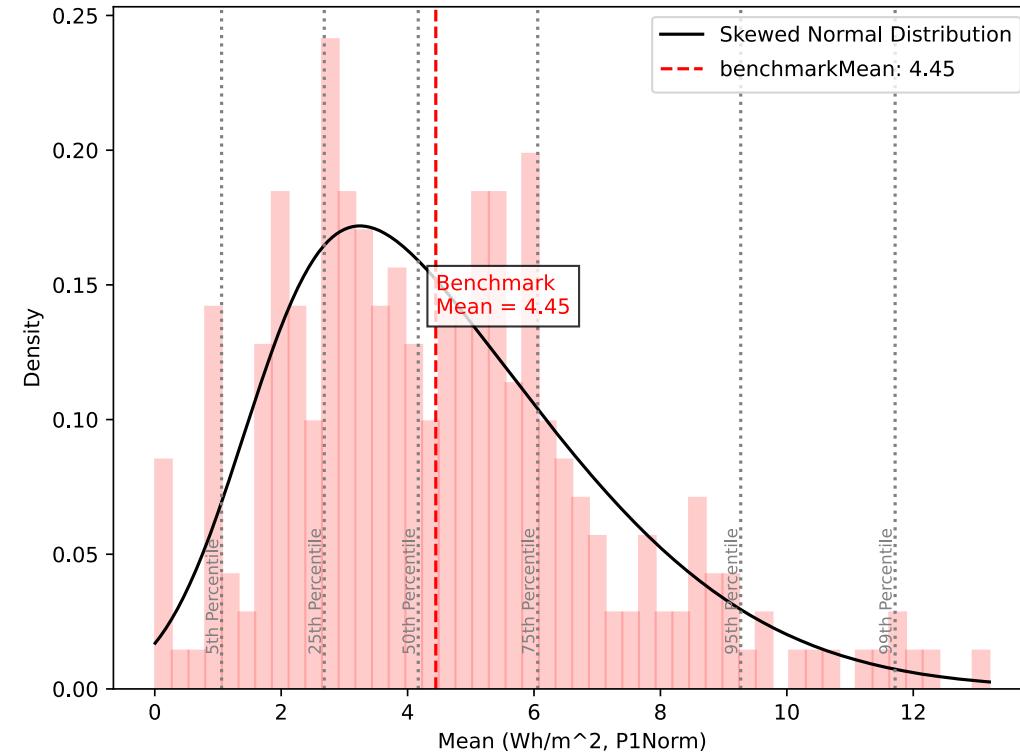
#6



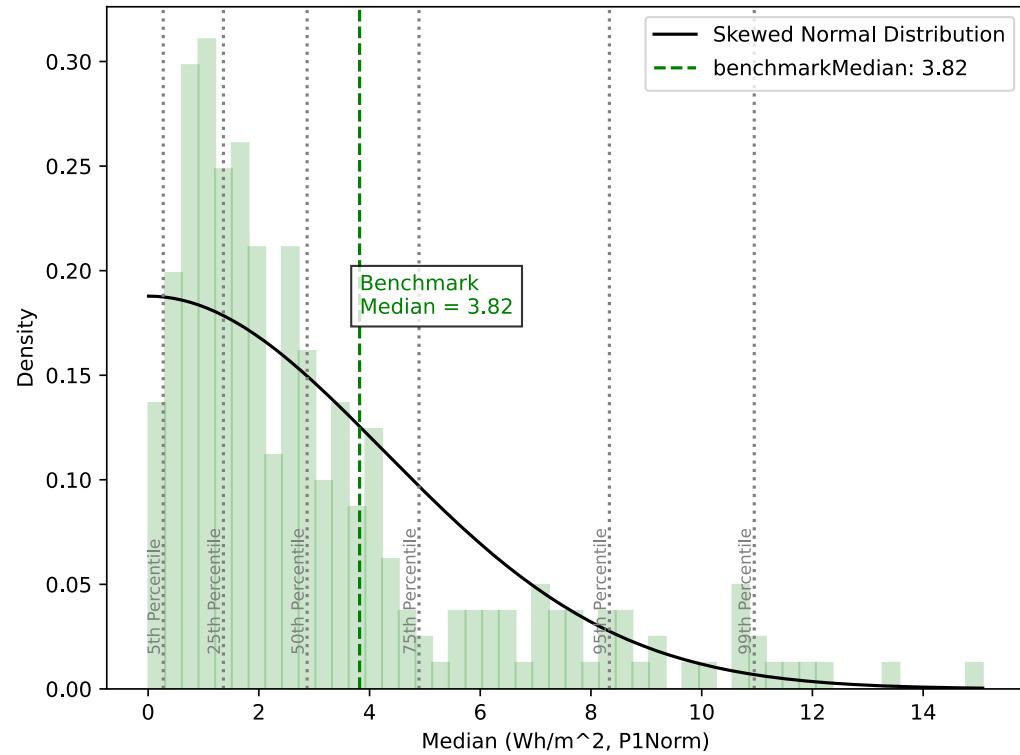
Benchmark data

Histogram with skewed Probability density

P1 Mean histogram from 307 sampled meters of sector:
Fashion: Unisex Wear
Total area of bars in histogram equals 1 (Density=True)
Constructed with 7,938,828 samples

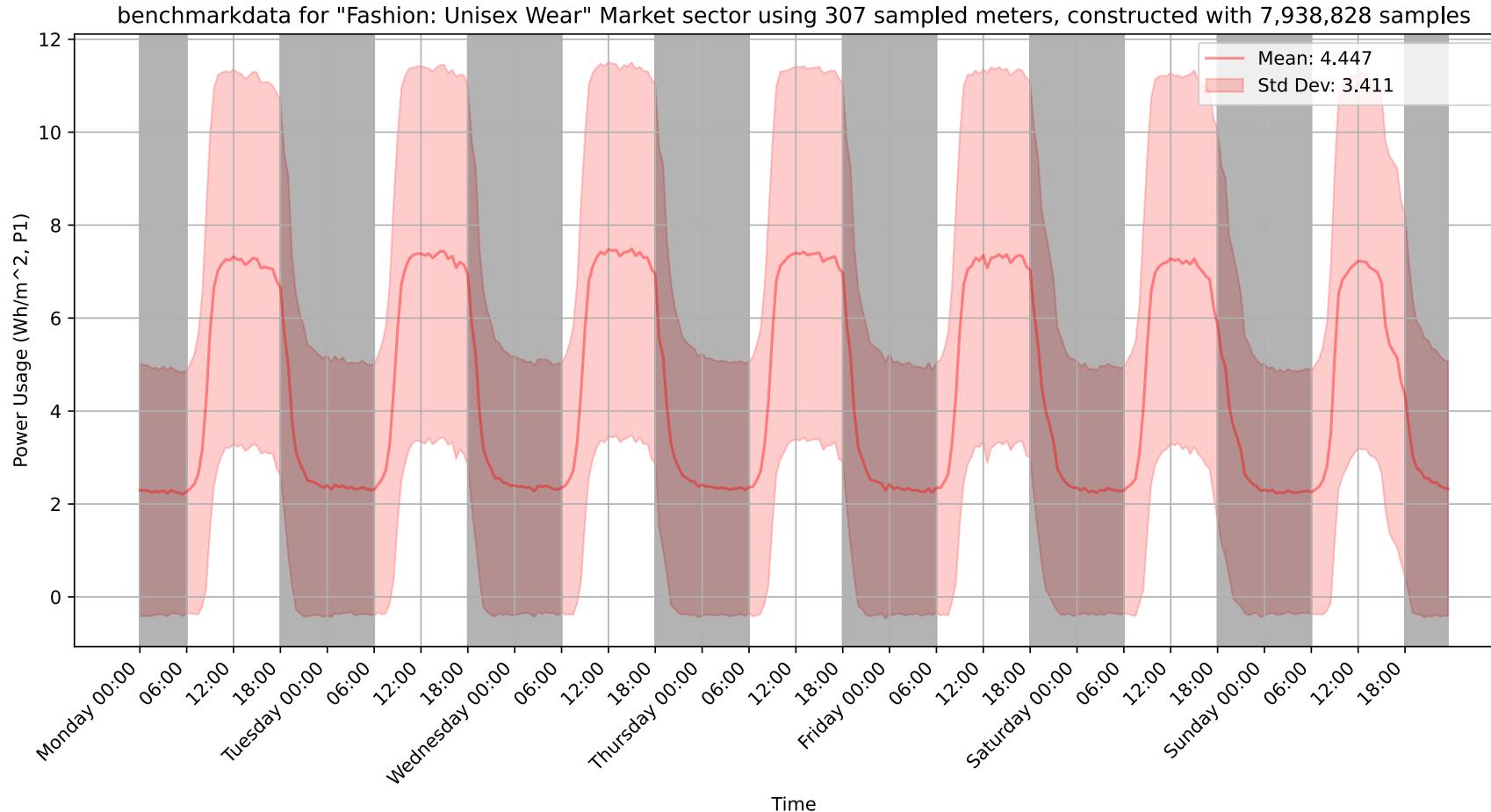


P1 Median histogram from 307 sampled meters of sector:
Fashion: Unisex Wear
Total area of bars in histogram equals 1 (Density=True)
Constructed with 7,938,828 samples



Benchmark data

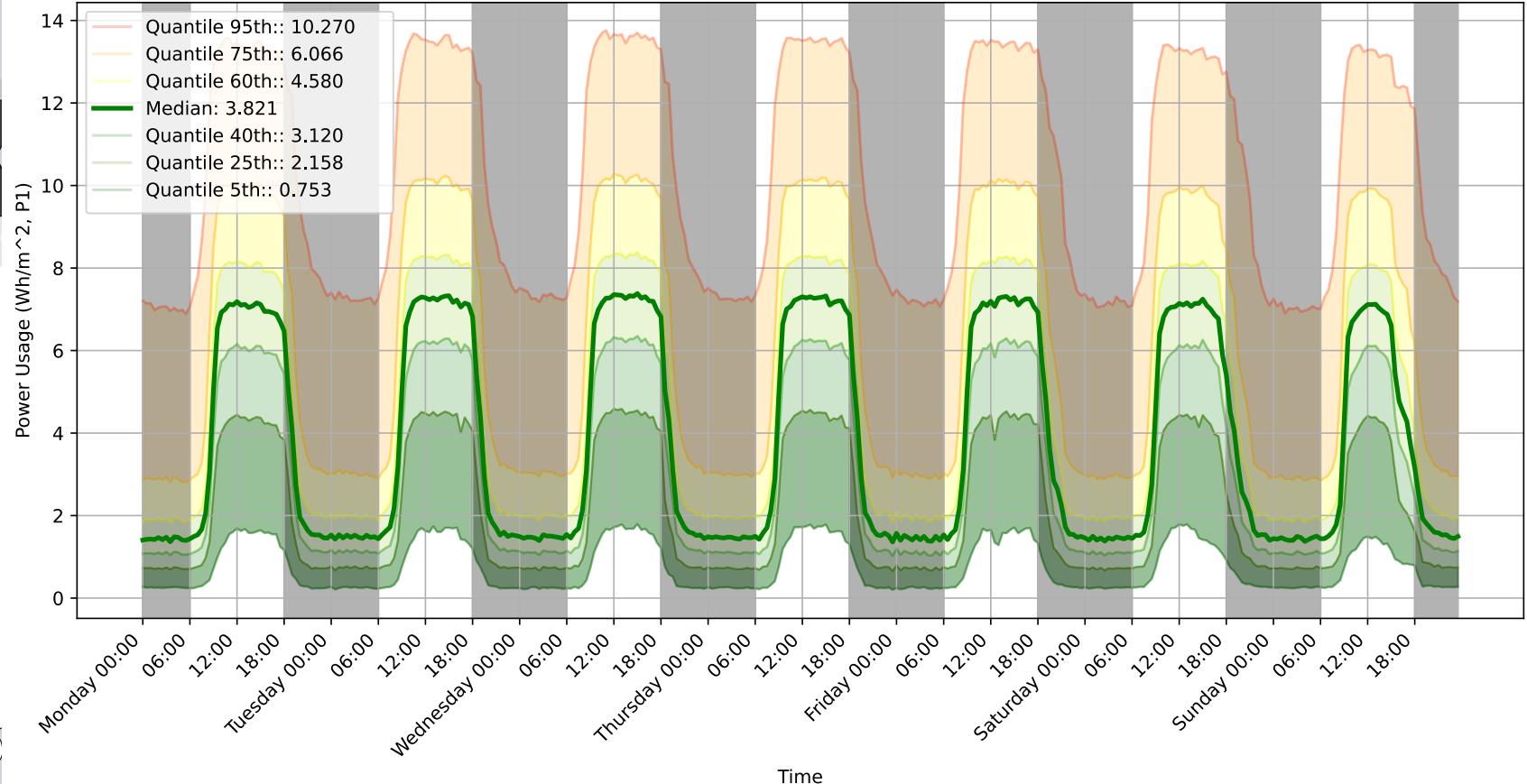
Time-series w mean and standard deviation



Benchmark data

Time-series w median and quantile range

benchmarkdata for "Fashion: Unisex Wear" Market sector using 307 sampled meters, constructed with 7,938,828 samples

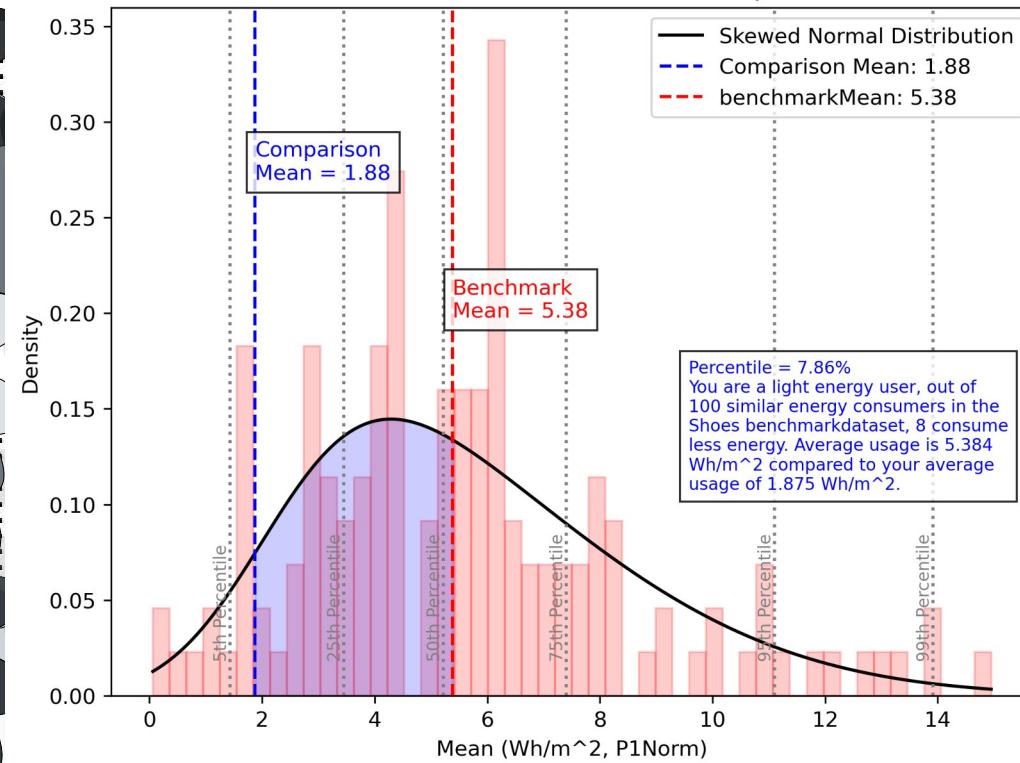


Test-Case #1

Histogram with skewed Probability density

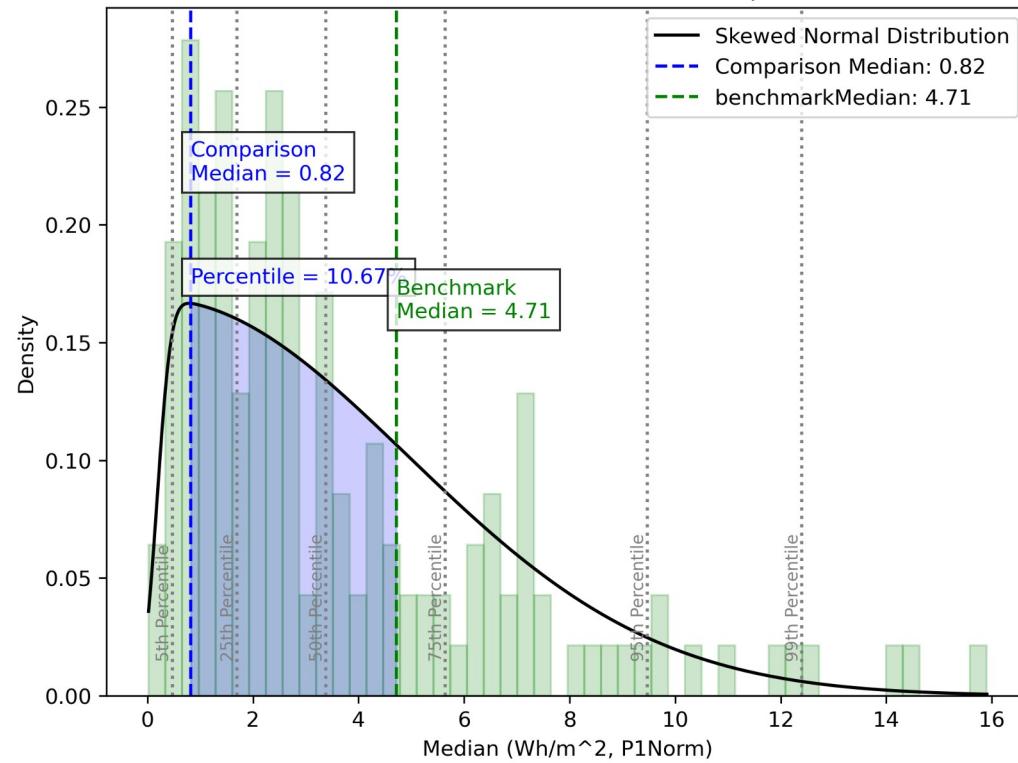
P1 Mean histogram from 170 sampled meters of sector:
Shoes

Total area of bars in histogram equals 1 (Density=True)
Constructed with 4,295,863 samples



P1 Median histogram from 170 sampled meters of sector:
Shoes

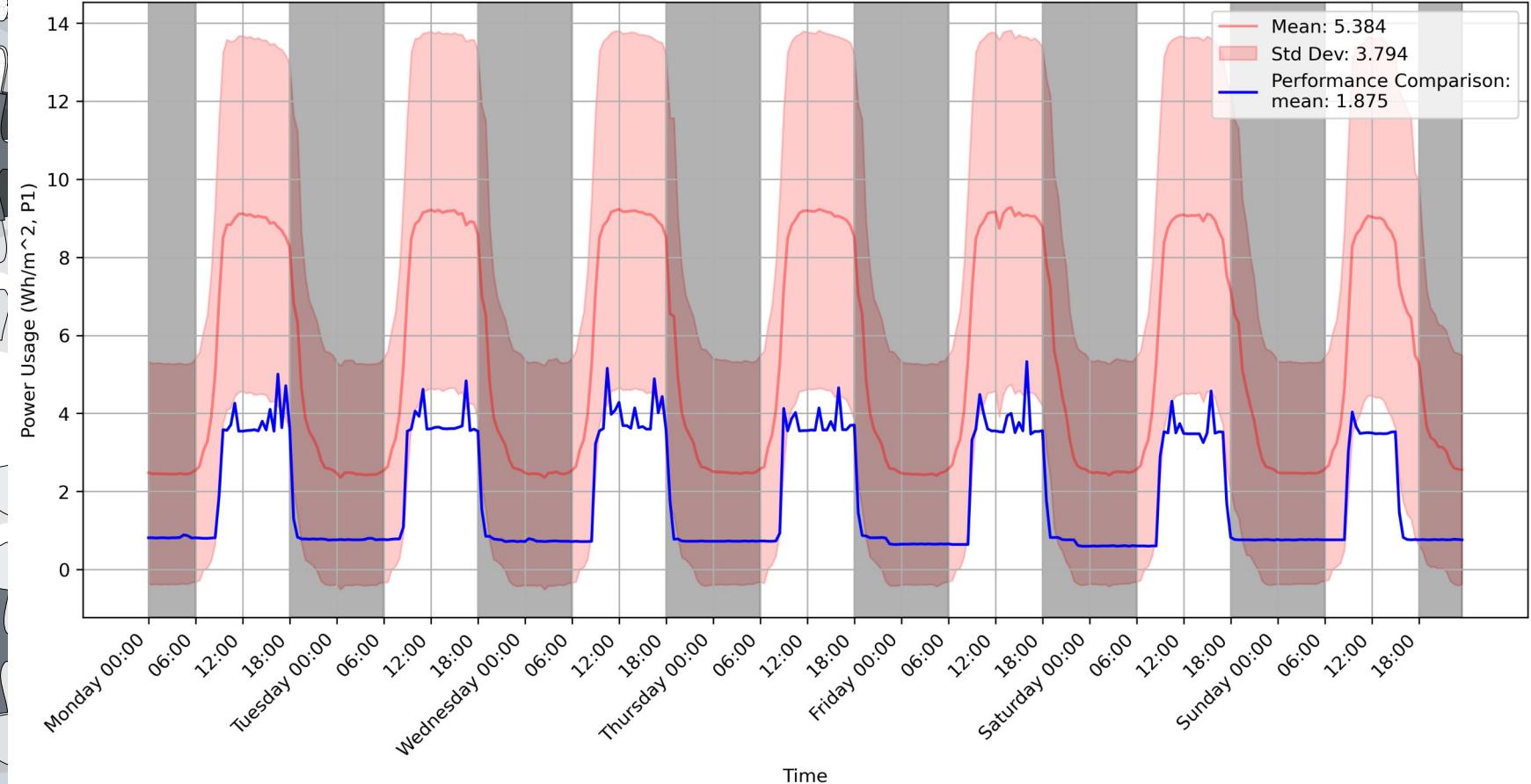
Total area of bars in histogram equals 1 (Density=True)
Constructed with 4,295,863 samples



Test-Case #1

Time-series w mean and standard deviation

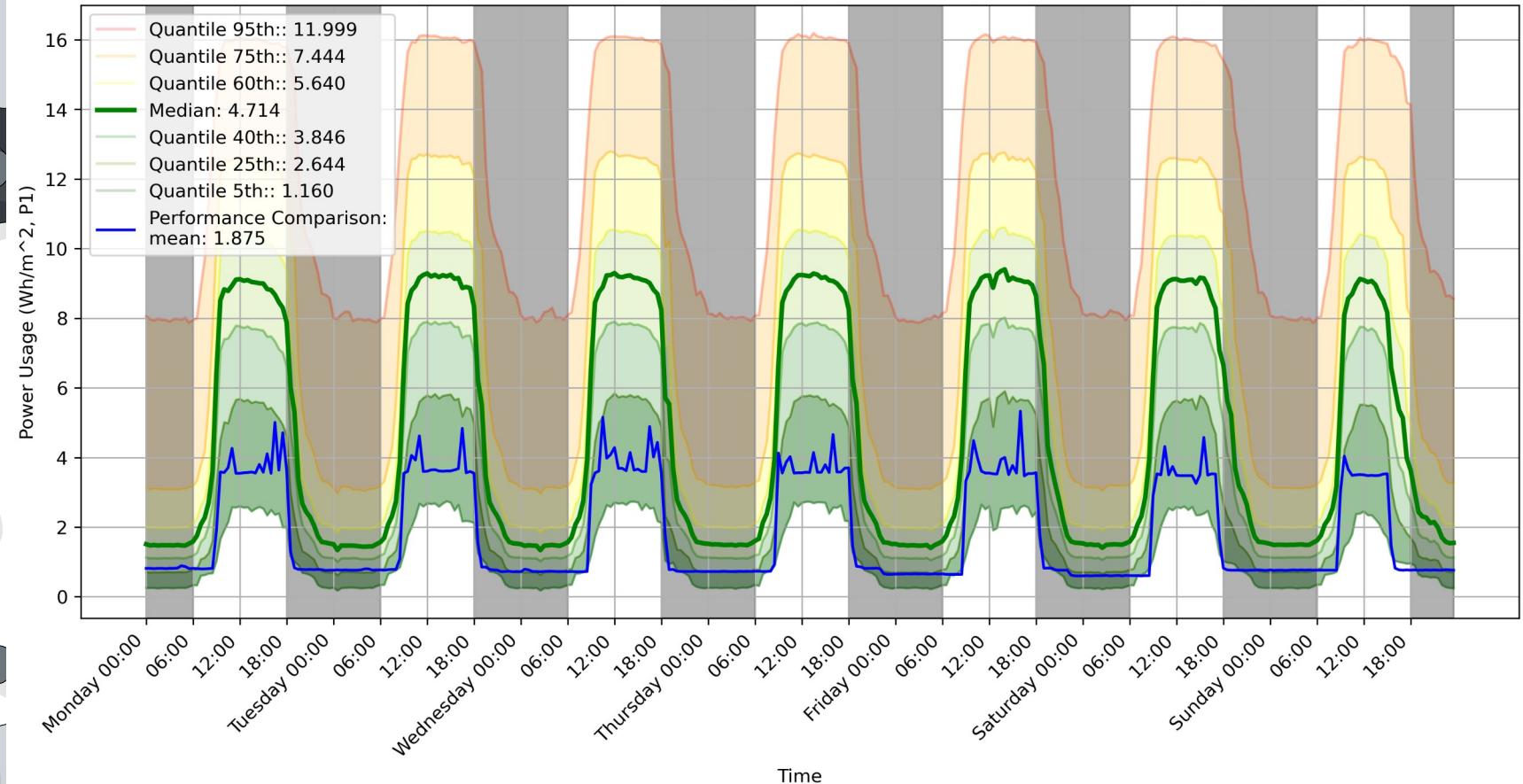
benchmarkdata for "Shoes" Market sector using 170 sampled meters, constructed with 4,295,863 samples



Test-Case #1

Time-series w median and quantile range

benchmarkdata for "Shoes" Market sector using 170 sampled meters, constructed with 4,295,863 samples

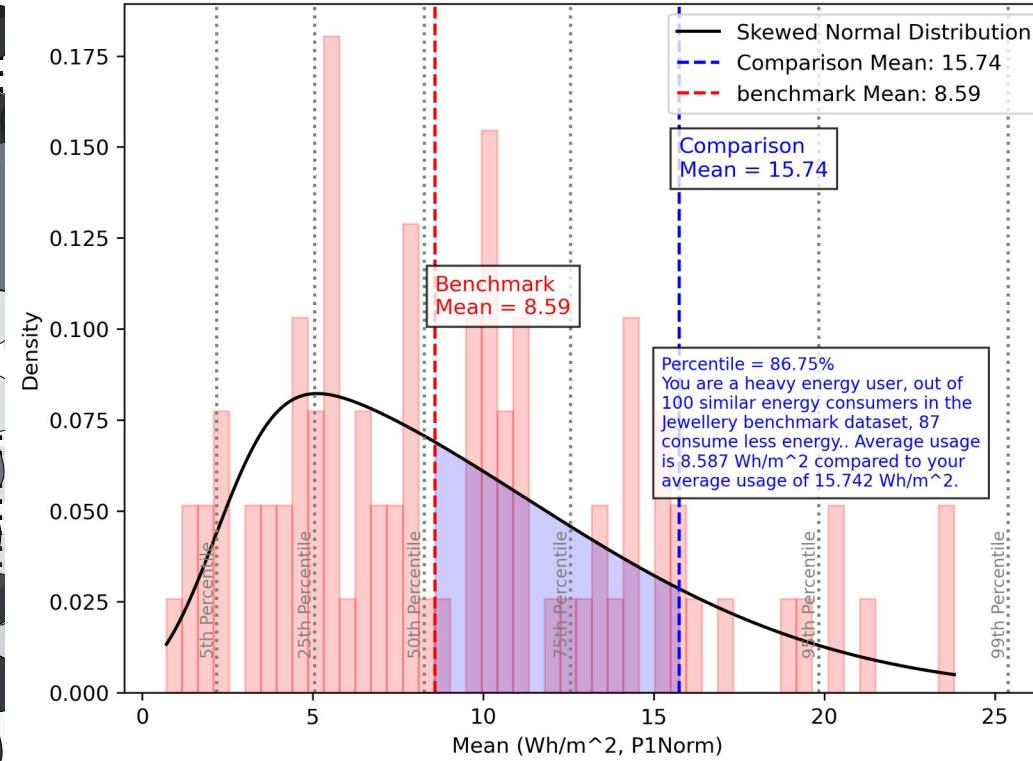


Test-Case #2

Histogram with skewed Probability density

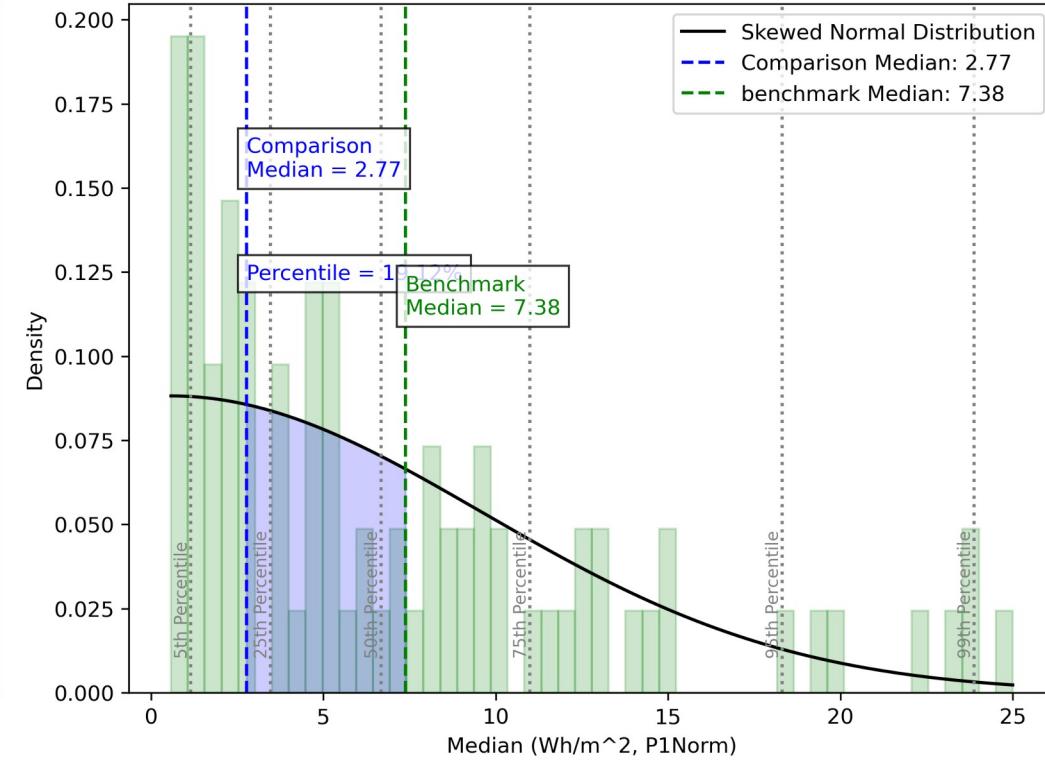
P1 Mean histogram from 91 sampled meters of sector:
Jewellery

Total area of bars in histogram equals 1 (Density=True)
Constructed with 2,425,392 samples



P1 Median histogram from 91 sampled meters of sector:
Jewellery

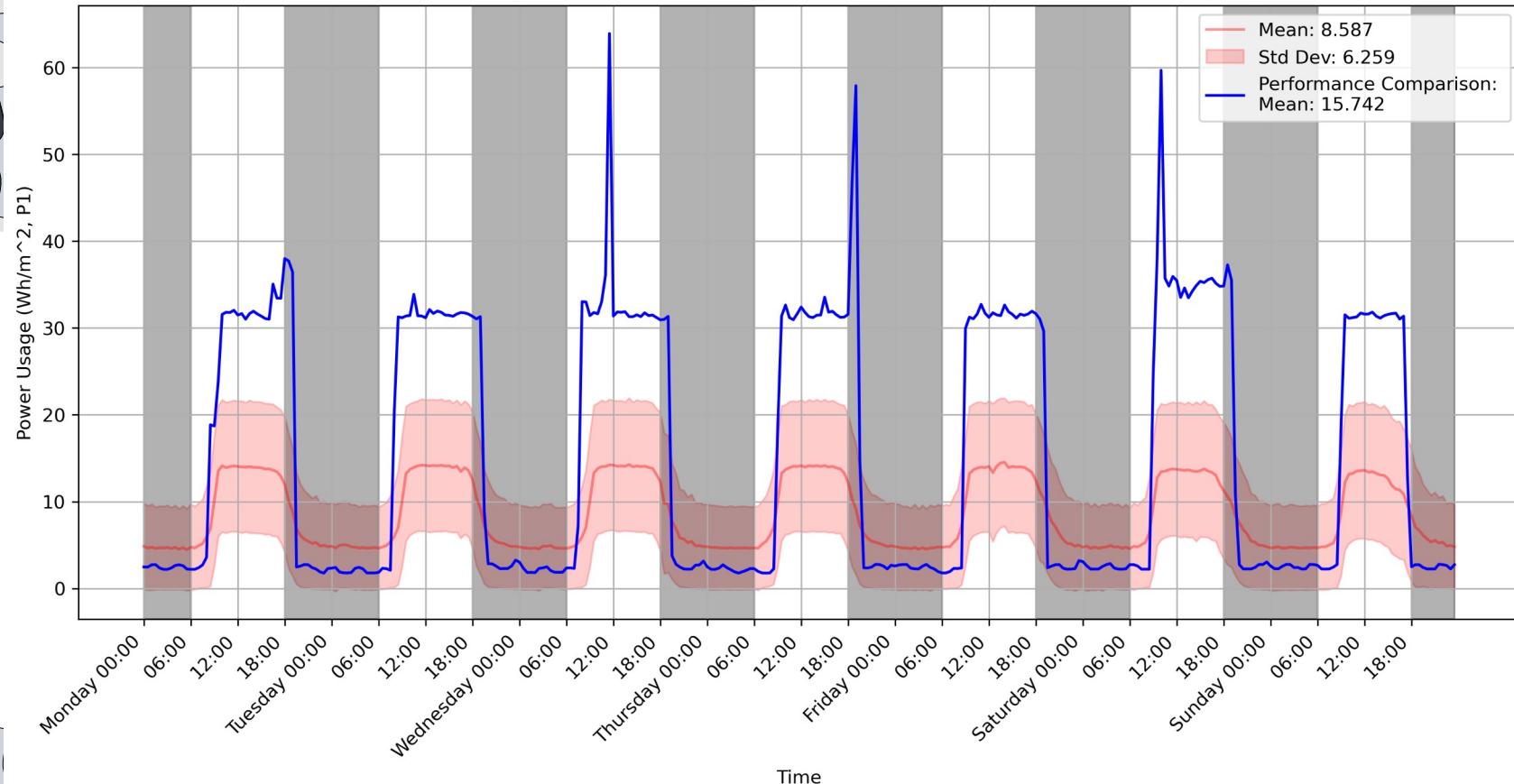
Total area of bars in histogram equals 1 (Density=True)
Constructed with 2,425,392 samples



Test-Case #2

Time-series w mean and standard deviation

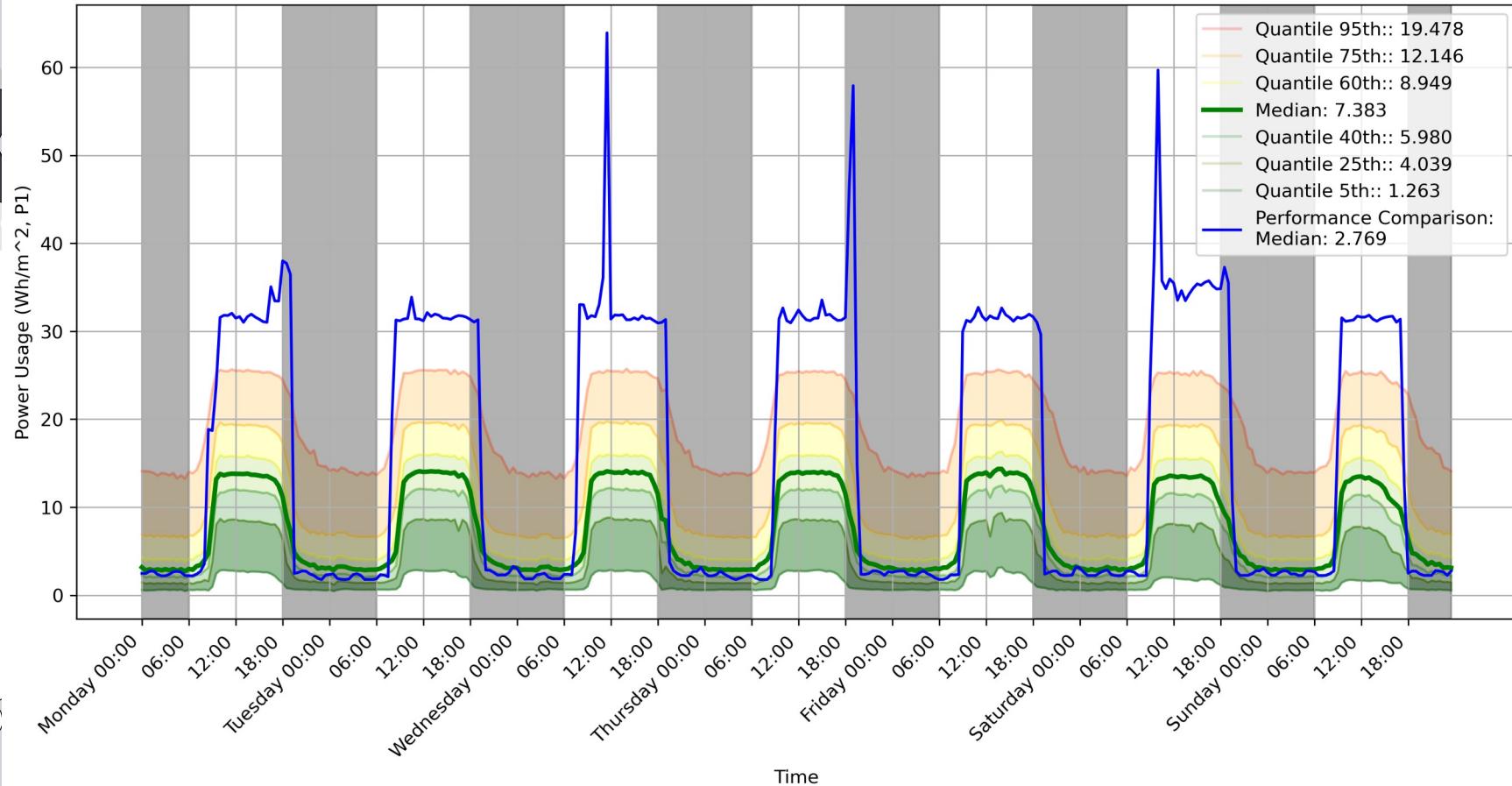
benchmarkdata for "Jewellery" Market sector using 91 sampled meters, constructed with 2,425,392 samples



Test-Case #2

Time-series w median and quantile range

benchmarkdata for "Jewellery" Market sector using 91 sampled meters, constructed with 2,425,392 samples

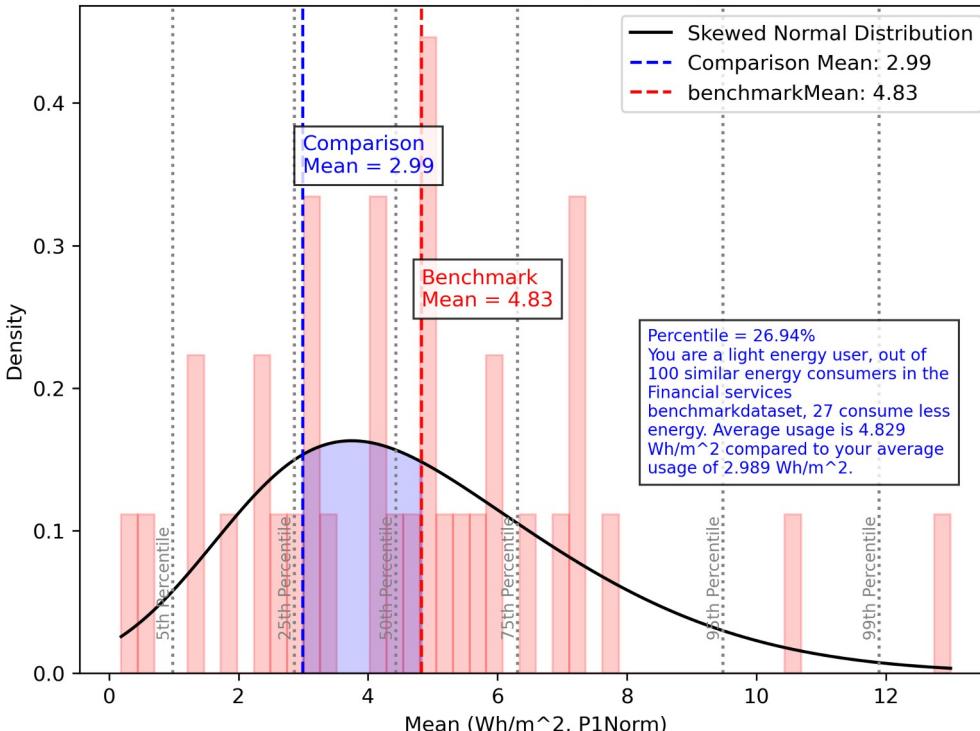


Test-Case #3

Histogram with skewed Probability density

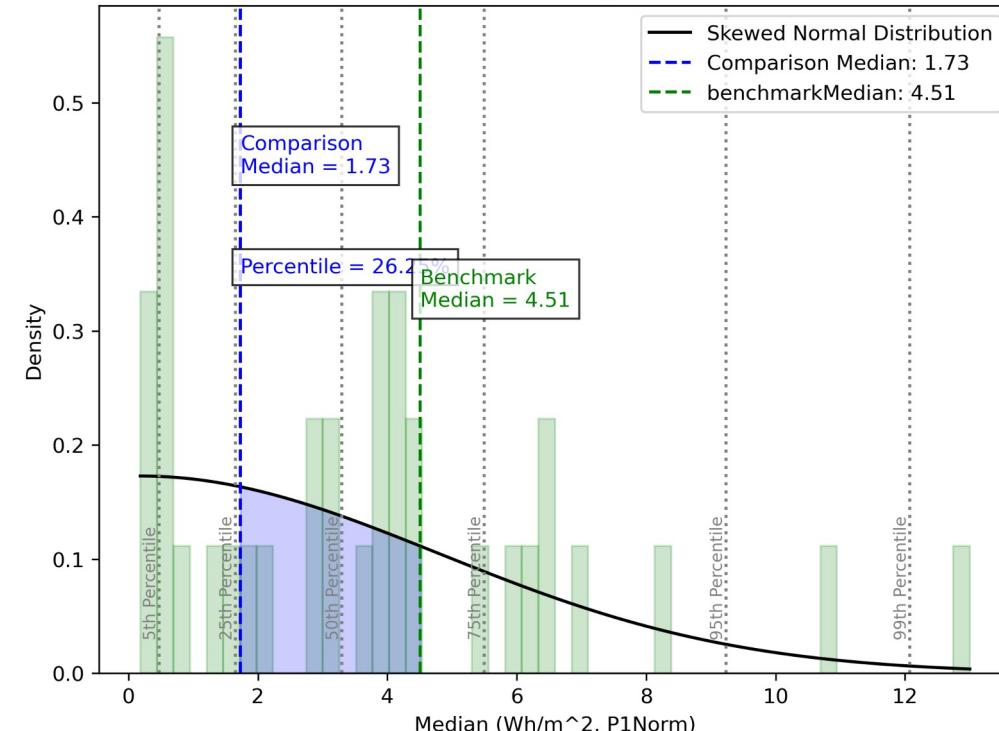
P1 Mean histogram from 41 sampled meters of sector:
Financial services

Total area of bars in histogram equals 1 (Density=True)
Constructed with 994,492 samples



P1 Median histogram from 41 sampled meters of sector:
Financial services

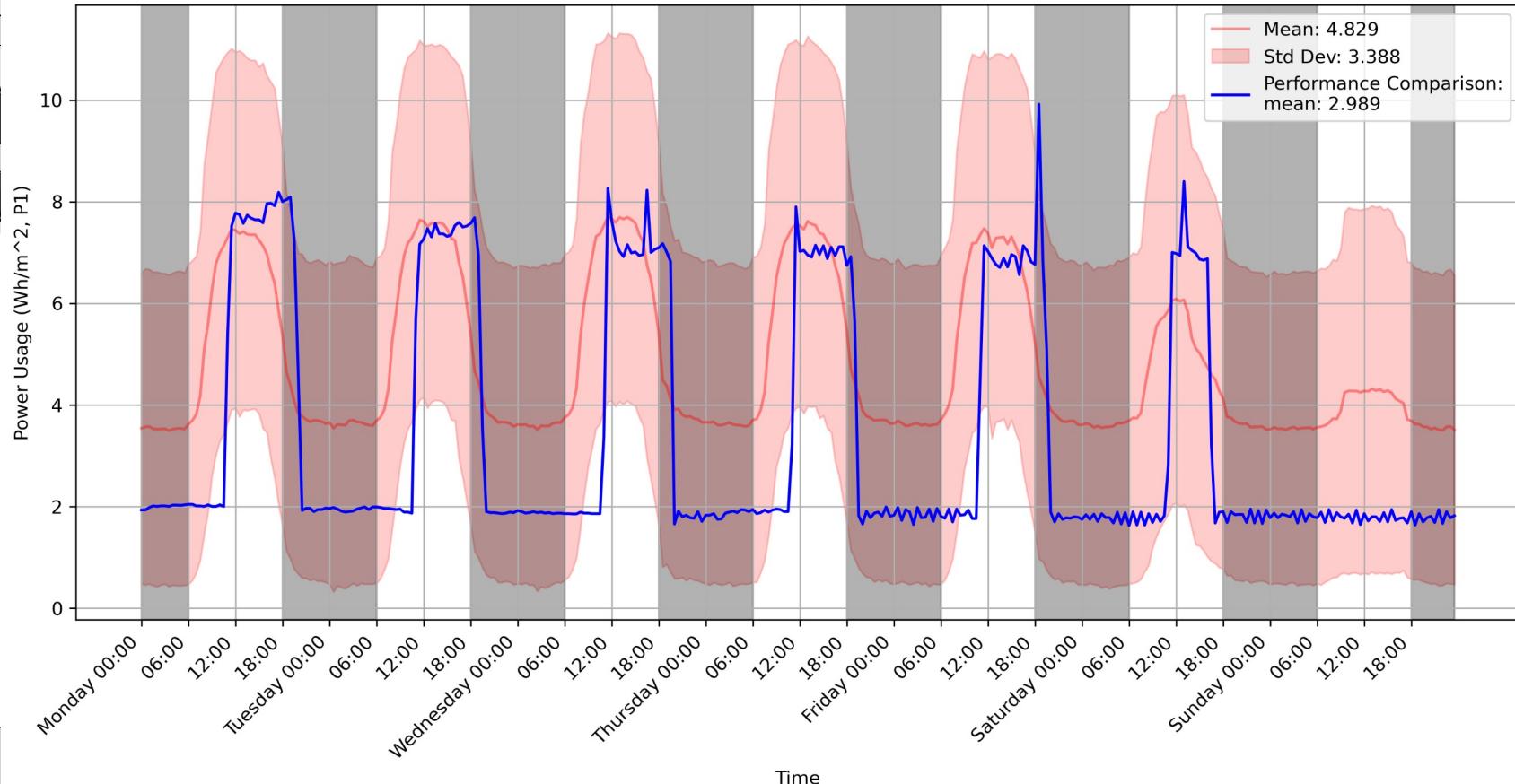
Total area of bars in histogram equals 1 (Density=True)
Constructed with 994,492 samples



Test-Case #3

Time-series w mean and standard deviation

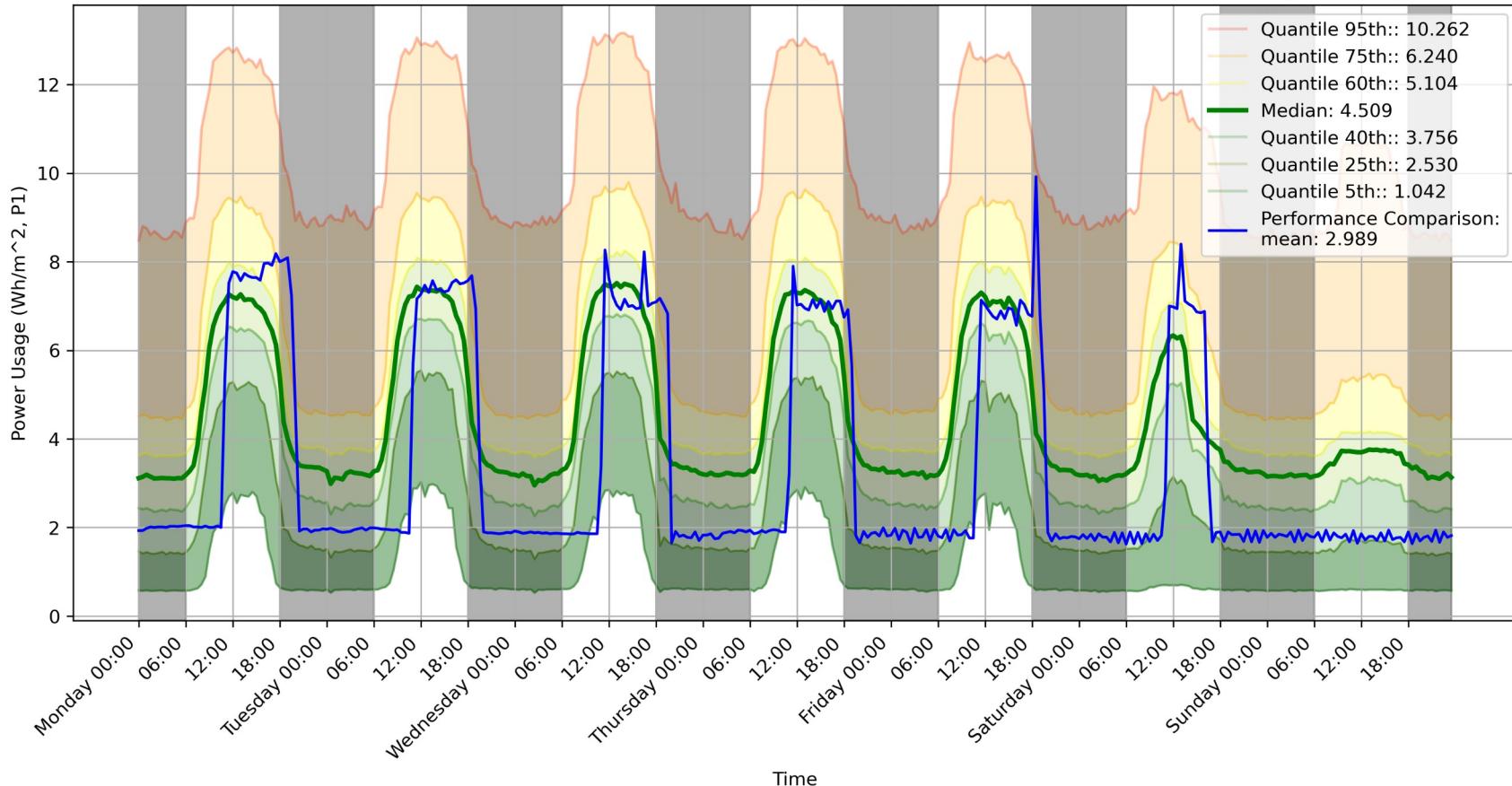
benchmarkdata for "Financial services" Market sector using 41 sampled meters, constructed with 994,492 samples



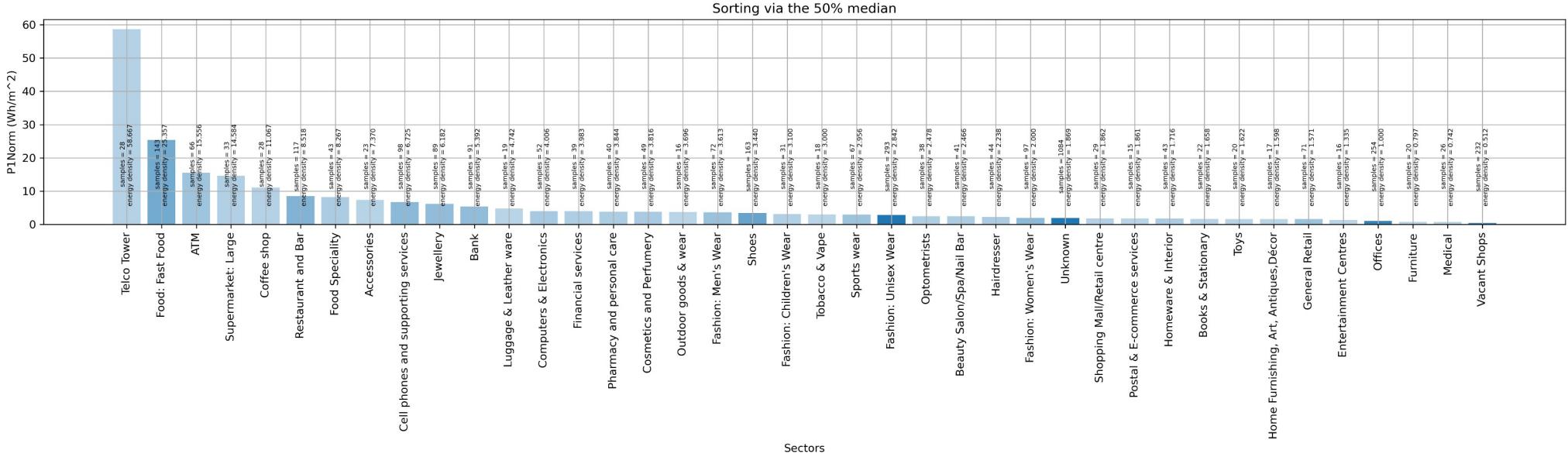
Test-Case #3

Time-series w median and quantile range

benchmarkdata for "Financial services" Market sector using 41 sampled meters, constructed with 994,492 samples



What does the median power consumption look like?



Questions?