Predict Bike Sharing Demand with AutoGluon Template

Project: Predict Bike Sharing Demand with AutoGluon

This notebook is a template with each step that you need to complete for the project.

Please fill in your code where there are explicit `?` markers in the notebook. You are welcome to add more cells and code as you see fit.

Once you have completed all the code implementations, please export your notebook as a HTML file so the reviews can view your code. Make sure you have all outputs correctly outputted.

`File-> Export Notebook As... -> Export Notebook as HTML`

There is a writeup to complete as well after all code implememtation is done. Please answer all questions and attach the necessary tables and charts. You can complete the writeup in either markdown or PDF.

Completing the code template and writeup template will cover all of the rubric points for this project.

The rubric contains "Stand Out Suggestions" for enhancing the project beyond the minimum requirements. The stand out suggestions are optional. If you decide to pursue the "stand out suggestions", you can include the code in this notebook and also discuss the results in the writeup file.

Step 1: Create an account with Kaggle

Create Kaggle Account and download API key

Below is example of steps to get the API username and key. Each student will have their own username and key.

1. Open account settings.

2. Scroll down to API and click Create New API Token.
3. Open up `kaggle.json` and use the username and key.

Step 2: Download the Kaggle dataset using the kaggle python library

Open up Sagemaker Studio and use starter template

1. Notebook should be using a `ml.t3.medium` instance (2 vCPU + 4 GiB)
2. Notebook should be using kernal: `Python 3 (MXNet 1.8 Python 3.7 CPU Optimized)`

Install packages

In [1]:
```
!pip install -U pip
!pip install -U setuptools wheel
!pip install -U "mxnet<2.0.0" bokeh==2.0.1
!pip install autogluon --no-cache-dir
# Without --no-cache-dir, smaller aws instances may have trouble installing
```

```
Requirement already satisfied: pip in /usr/local/lib/python3.11/dist-packages (25.1.1)
Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-packages (80.9.0)
Requirement already satisfied: wheel in /usr/local/lib/python3.11/dist-packages (0.45.1)
Collecting mxnet<2.0.0
  Using cached mxnet-1.9.1-py3-none-manylinux2014_x86_64.whl.metadata (3.4 kB)
Collecting bokeh==2.0.1
  Using cached bokeh-2.0.1.tar.gz (8.6 MB)
  Preparing metadata (setup.py) ... done
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (6.0.2)
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (2.9.0.post
0)
Requirement already satisfied: Jinja2>=2.7 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (3.1.6)
Requirement already satisfied: numpy>=1.11.3 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (2.0.2)
Requirement already satisfied: pillow>=4.0 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (11.2.1)
Requirement already satisfied: packaging>=16.8 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (24.2)
Requirement already satisfied: tornado>=5 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (6.4.2)
Requirement already satisfied: typing_extensions>=3.7.4 in /usr/local/lib/python3.11/dist-packages (from bokeh==2.0.1) (4.14.0)
Collecting numpy>=1.11.3 (from bokeh==2.0.1)
  Downloading numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (61 kB)
Requirement already satisfied: requests<3,>=2.20.0 in /usr/local/lib/python3.11/dist-packages (from mxnet<2.0.0) (2.32.3)
Collecting graphviz<0.9.0,>=0.8.1 (from mxnet<2.0.0)
  Downloading graphviz-0.8.4-py2.py3-none-any.whl.metadata (6.4 kB)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.20.0->m
xnet<2.0.0) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.20.0->mxnet<2.0.0)
(3.10)
Requirement already satisfied: urllib3<3,>=1.21.1 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.20.0->mxnet<
2.0.0) (2.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests<3,>=2.20.0->mxnet<
2.0.0) (2025.4.26)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from Jinja2>=2.7->bokeh==2.0.1) (3.
0.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.11/dist-packages (from python-dateutil>=2.1->bokeh==2.0.1)
(1.17.0)
Downloading mxnet-1.9.1-py3-none-manylinux2014_x86_64.whl (49.1 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 49.1/49.1 MB 60.5 MB/s eta 0:00:00
Downloading graphviz-0.8.4-py2.py3-none-any.whl (16 kB)
Downloading numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.3/18.3 MB 142.9 MB/s eta 0:00:00
Building wheels for collected packages: bokeh
  DEPRECATION: Building 'bokeh' using the legacy setup.py bdist_wheel mechanism, which will be removed in a future version. pip
```

```
25.3 will enforce this behaviour change. A possible replacement is to use the standardized build interface by setting the `--us
e-pep517` option, (possibly combined with `--no-build-isolation`), or adding a `pyproject.toml` file to the source tree of 'bok
eh'. Discussion can be found at https://github.com/pypa/pip/issues/6334
  Building wheel for bokeh (setup.py) ... done
  Created wheel for bokeh: filename=bokeh-2.0.1-py3-none-any.whl size=9080109 sha256=d720324d565ba04d7d273295973ca936ab489f6e3a
7eab94edb8daf1ee0354a8
  Stored in directory: /root/.cache/pip/wheels/0e/c4/ce/fecee9e6406e166eaba4e09b1acd2096a84ffef5275ea90806
Successfully built bokeh
Installing collected packages: numpy, graphviz, mxnet, bokeh
  Attempting uninstall: numpy
    Found existing installation: numpy 2.0.2
    Uninstalling numpy-2.0.2:
      Successfully uninstalled numpy-2.0.2
  Attempting uninstall: graphviz
    Found existing installation: graphviz 0.20.3
    Uninstalling graphviz-0.20.3:
      Successfully uninstalled graphviz-0.20.3
  Attempting uninstall: bokeh
    Found existing installation: bokeh 3.7.3
    Uninstalling bokeh-3.7.3:
      Successfully uninstalled bokeh-3.7.3
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 4/4 [bokeh]
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is th
e source of the following dependency conflicts.
panel 1.7.1 requires bokeh<3.8.0,>=3.5.0, but you have bokeh 2.0.1 which is incompatible.
thinc 8.3.6 requires numpy<3.0.0,>=2.0.0, but you have numpy 1.26.4 which is incompatible.
holoviews 1.20.2 requires bokeh>=3.1, but you have bokeh 2.0.1 which is incompatible.
Successfully installed bokeh-2.0.1 graphviz-0.8.4 mxnet-1.9.1 numpy-1.26.4
Collecting autogluon
  Downloading autogluon-1.3.1-py3-none-any.whl.metadata (11 kB)
Collecting autogluon.core==1.3.1 (from autogluon.core[all]==1.3.1->autogluon)
  Downloading autogluon.core-1.3.1-py3-none-any.whl.metadata (12 kB)
Collecting autogluon.features==1.3.1 (from autogluon)
  Downloading autogluon.features-1.3.1-py3-none-any.whl.metadata (11 kB)
Collecting autogluon.tabular==1.3.1 (from autogluon.tabular[all]==1.3.1->autogluon)
  Downloading autogluon.tabular-1.3.1-py3-none-any.whl.metadata (14 kB)
Collecting autogluon.multimodal==1.3.1 (from autogluon)
  Downloading autogluon.multimodal-1.3.1-py3-none-any.whl.metadata (13 kB)
Collecting autogluon.timeseries==1.3.1 (from autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading autogluon.timeseries-1.3.1-py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: numpy<2.3.0,>=1.25.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->aut
```

```
ogluon.core[all]==1.3.1->autogluon) (1.26.4)
Requirement already satisfied: scipy<1.16,>=1.5.4 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->autog
luon.core[all]==1.3.1->autogluon) (1.15.3)
Requirement already satisfied: scikit-learn<1.7.0,>=1.4.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.
1->autogluon.core[all]==1.3.1->autogluon) (1.6.1)
Requirement already satisfied: networkx<4,>=3.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->autoglu
on.core[all]==1.3.1->autogluon) (3.5)
Requirement already satisfied: pandas<2.3.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->aut
ogluon.core[all]==1.3.1->autogluon) (2.2.2)
Requirement already satisfied: tqdm<5,>=4.38 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->autogluon.
core[all]==1.3.1->autogluon) (4.67.1)
Requirement already satisfied: requests in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->autogluon.core
[all]==1.3.1->autogluon) (2.32.3)
Requirement already satisfied: matplotlib<3.11,>=3.7.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core==1.3.1->
autogluon.core[all]==1.3.1->autogluon) (3.10.0)
Collecting boto3<2,>=1.10 (from autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading boto3-1.38.32-py3-none-any.whl.metadata (6.6 kB)
Collecting autogluon.common==1.3.1 (from autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading autogluon.common-1.3.1-py3-none-any.whl.metadata (11 kB)
Requirement already satisfied: psutil<7.1.0,>=5.7.3 in /usr/local/lib/python3.11/dist-packages (from autogluon.common==1.3.1->a
utogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (5.9.5)
Collecting ray<2.45,>=2.10.0 (from ray[default]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading ray-2.44.1-cp311-cp311-manylinux2014_x86_64.whl.metadata (19 kB)
Requirement already satisfied: hyperopt<0.2.8,>=0.2.7 in /usr/local/lib/python3.11/dist-packages (from autogluon.core[all]==1.
3.1->autogluon) (0.2.7)
Requirement already satisfied: pyarrow>=15.0.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.core[all]==1.3.1->aut
ogluon) (18.1.0)
Requirement already satisfied: Pillow<12,>=10.0.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==1.3.1-
>autogluon) (11.2.1)
Requirement already satisfied: torch<2.7,>=2.2 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==1.3.1->au
togluon) (2.6.0+cu124)
Collecting lightning<2.7,>=2.2 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading lightning-2.5.1.post0-py3-none-any.whl.metadata (39 kB)
Collecting transformers<4.50,>=4.38.0 (from transformers[sentencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.1->autogluon)
  Downloading transformers-4.49.0-py3-none-any.whl.metadata (44 kB)
Requirement already satisfied: accelerate<2.0,>=0.34.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==
1.3.1->autogluon) (1.7.0)
Collecting jsonschema<4.24,>=4.18 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading jsonschema-4.23.0-py3-none-any.whl.metadata (7.9 kB)
Collecting seqeval<1.3.0,>=1.2.2 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading seqeval-1.2.2.tar.gz (43 kB)
```

```
  Preparing metadata (setup.py) ... done
Collecting evaluate<0.5.0,>=0.4.0 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading evaluate-0.4.3-py3-none-any.whl.metadata (9.2 kB)
Collecting timm<1.0.7,>=0.9.5 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading timm-1.0.3-py3-none-any.whl.metadata (43 kB)
Requirement already satisfied: torchvision<0.22.0,>=0.16.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimoda
l==1.3.1->autogluon) (0.21.0+cu124)
Requirement already satisfied: scikit-image<0.26.0,>=0.19.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimod
al==1.3.1->autogluon) (0.25.2)
Requirement already satisfied: text-unidecode<1.4,>=1.3 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==
1.3.1->autogluon) (1.3)
Collecting torchmetrics<1.8,>=1.2.0 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading torchmetrics-1.7.2-py3-none-any.whl.metadata (21 kB)
Requirement already satisfied: omegaconf<2.4.0,>=2.1.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==
1.3.1->autogluon) (2.3.0)
Collecting pytorch-metric-learning<2.9,>=1.3.0 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading pytorch_metric_learning-2.8.1-py3-none-any.whl.metadata (18 kB)
Collecting nlpaug<1.2.0,>=1.1.10 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading nlpaug-1.1.11-py3-none-any.whl.metadata (14 kB)
Collecting nltk<3.9,>=3.4.5 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading nltk-3.8.1-py3-none-any.whl.metadata (2.8 kB)
Collecting openmim<0.4.0,>=0.3.7 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading openmim-0.3.9-py2.py3-none-any.whl.metadata (16 kB)
Requirement already satisfied: defusedxml<0.7.2,>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==
1.3.1->autogluon) (0.7.1)
Requirement already satisfied: jinja2<3.2,>=3.0.3 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==1.3.1-
>autogluon) (3.1.6)
Requirement already satisfied: tensorboard<3,>=2.9 in /usr/local/lib/python3.11/dist-packages (from autogluon.multimodal==1.3.1
->autogluon) (2.18.0)
Collecting pytesseract<0.4,>=0.3.9 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading pytesseract-0.3.13-py3-none-any.whl.metadata (11 kB)
Collecting nvidia-ml-py3<8.0,>=7.352.0 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia-ml-py3-7.352.0.tar.gz (19 kB)
  Preparing metadata (setup.py) ... done
Collecting pdf2image<1.19,>=1.17.0 (from autogluon.multimodal==1.3.1->autogluon)
  Downloading pdf2image-1.17.0-py3-none-any.whl.metadata (6.2 kB)
Collecting catboost<1.3,>=1.2 (from autogluon.tabular[all]==1.3.1->autogluon)
  Downloading catboost-1.2.8-cp311-cp311-manylinux2014_x86_64.whl.metadata (1.2 kB)
Requirement already satisfied: einops<0.9,>=0.7 in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==1.3.1-
>autogluon) (0.8.1)
Requirement already satisfied: xgboost<3.1,>=2.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==1.3.1
```

```
->autogluon) (2.1.4)
Requirement already satisfied: fastai<2.9,>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==1.3.
1->autogluon) (2.7.19)
Requirement already satisfied: huggingface-hub[torch] in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==
1.3.1->autogluon) (0.32.4)
Requirement already satisfied: lightgbm<4.7,>=4.0 in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==1.3.
1->autogluon) (4.5.0)
Requirement already satisfied: spacy<3.9 in /usr/local/lib/python3.11/dist-packages (from autogluon.tabular[all]==1.3.1->autogl
uon) (3.8.7)
Requirement already satisfied: joblib<2,>=1.1 in /usr/local/lib/python3.11/dist-packages (from autogluon.timeseries==1.3.1->aut
ogluon.timeseries[all]==1.3.1->autogluon) (1.5.1)
Collecting pytorch-lightning (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading pytorch_lightning-2.5.1.post0-py3-none-any.whl.metadata (20 kB)
Collecting gluonts<0.17,>=0.15.0 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading gluonts-0.16.1-py3-none-any.whl.metadata (9.8 kB)
Collecting statsforecast<2.0.2,>=1.7.0 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading statsforecast-2.0.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (29 kB)
Collecting mlforecast<0.14,>0.13 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading mlforecast-0.13.6-py3-none-any.whl.metadata (12 kB)
Collecting utilsforecast<0.2.11,>=0.2.3 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading utilsforecast-0.2.10-py3-none-any.whl.metadata (7.4 kB)
Collecting coreforecast<0.0.16,>=0.0.12 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading coreforecast-0.0.15-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.8 kB)
Collecting fugue>=0.9.0 (from autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading fugue-0.9.1-py3-none-any.whl.metadata (18 kB)
Requirement already satisfied: orjson~=3.9 in /usr/local/lib/python3.11/dist-packages (from autogluon.timeseries==1.3.1->autogl
uon.timeseries[all]==1.3.1->autogluon) (3.10.18)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.11/dist-packages (from accelerate<2.0,>=0.34.0->autogl
uon.multimodal==1.3.1->autogluon) (24.2)
Requirement already satisfied: pyyaml in /usr/local/lib/python3.11/dist-packages (from accelerate<2.0,>=0.34.0->autogluon.multi
modal==1.3.1->autogluon) (6.0.2)
Requirement already satisfied: safetensors>=0.4.3 in /usr/local/lib/python3.11/dist-packages (from accelerate<2.0,>=0.34.0->aut
ogluon.multimodal==1.3.1->autogluon) (0.5.3)
Collecting botocore<1.39.0,>=1.38.32 (from boto3<2,>=1.10->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading botocore-1.38.32-py3-none-any.whl.metadata (5.7 kB)
Collecting jmespath<2.0.0,>=0.7.1 (from boto3<2,>=1.10->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading jmespath-1.0.1-py3-none-any.whl.metadata (7.6 kB)
Collecting s3transfer<0.14.0,>=0.13.0 (from boto3<2,>=1.10->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading s3transfer-0.13.0-py3-none-any.whl.metadata (1.7 kB)
Requirement already satisfied: python-dateutil<3.0.0,>=2.1 in /usr/local/lib/python3.11/dist-packages (from botocore<1.39.0,>=
1.38.32->boto3<2,>=1.10->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (2.9.0.post0)
```

```
Requirement already satisfied: urllib3!=2.2.0,<3,>=1.25.4 in /usr/local/lib/python3.11/dist-packages (from botocore<1.39.0,>=1.38.32->boto3<2,>=1.10->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (2.4.0)
Requirement already satisfied: graphviz in /usr/local/lib/python3.11/dist-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.1->autogluon) (0.8.4)
Requirement already satisfied: plotly in /usr/local/lib/python3.11/dist-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.1->autogluon) (5.24.1)
Requirement already satisfied: six in /usr/local/lib/python3.11/dist-packages (from catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.1->autogluon) (1.17.0)
Requirement already satisfied: datasets>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (2.14.4)
Requirement already satisfied: dill in /usr/local/lib/python3.11/dist-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (0.3.7)
Requirement already satisfied: xxhash in /usr/local/lib/python3.11/dist-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (3.5.0)
Requirement already satisfied: multiprocess in /usr/local/lib/python3.11/dist-packages (from evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (0.70.15)
Requirement already satisfied: fsspec>=2021.05.0 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (2025.3.2)
Requirement already satisfied: pip in /usr/local/lib/python3.11/dist-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.1->autogluon) (25.1.1)
Requirement already satisfied: fastdownload<2,>=0.0.5 in /usr/local/lib/python3.11/dist-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.1->autogluon) (0.0.7)
Requirement already satisfied: fastcore<1.8,>=1.5.29 in /usr/local/lib/python3.11/dist-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.1->autogluon) (1.7.29)
Requirement already satisfied: fastprogress>=0.2.4 in /usr/local/lib/python3.11/dist-packages (from fastai<2.9,>=2.3.1->autogluon.tabular[all]==1.3.1->autogluon) (1.0.3)
Requirement already satisfied: pydantic<3,>=1.7 in /usr/local/lib/python3.11/dist-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (2.11.5)
Requirement already satisfied: toolz~=0.10 in /usr/local/lib/python3.11/dist-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.12.1)
Requirement already satisfied: typing-extensions~=4.0 in /usr/local/lib/python3.11/dist-packages (from gluonts<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (4.14.0)
Requirement already satisfied: future in /usr/local/lib/python3.11/dist-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.1->autogluon) (1.0.0)
Requirement already satisfied: cloudpickle in /usr/local/lib/python3.11/dist-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.1->autogluon) (3.1.1)
Requirement already satisfied: py4j in /usr/local/lib/python3.11/dist-packages (from hyperopt<0.2.8,>=0.2.7->autogluon.core[all]==1.3.1->autogluon) (0.10.9.7)
Requirement already satisfied: MarkupSafe>=2.0 in /usr/local/lib/python3.11/dist-packages (from jinja2<3.2,>=3.0.3->autogluon.multimodal==1.3.1->autogluon) (3.0.2)
Requirement already satisfied: attrs>=22.2.0 in /usr/local/lib/python3.11/dist-packages (from jsonschema<4.24,>=4.18->autogluo
```

```
n.multimodal==1.3.1->autogluon) (25.3.0)
Requirement already satisfied: jsonschema-specifications>=2023.03.6 in /usr/local/lib/python3.11/dist-packages (from jsonschema
<4.24,>=4.18->autogluon.multimodal==1.3.1->autogluon) (2025.4.1)
Requirement already satisfied: referencing>=0.28.4 in /usr/local/lib/python3.11/dist-packages (from jsonschema<4.24,>=4.18->aut
ogluon.multimodal==1.3.1->autogluon) (0.36.2)
Requirement already satisfied: rpds-py>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from jsonschema<4.24,>=4.18->autogluo
n.multimodal==1.3.1->autogluon) (0.25.1)
Collecting lightning-utilities<2.0,>=0.10.0 (from lightning<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading lightning_utilities-0.14.3-py3-none-any.whl.metadata (5.6 kB)
Requirement already satisfied: aiohttp!=4.0.0a0,!=4.0.0a1 in /usr/local/lib/python3.11/dist-packages (from fsspec[http]>=2021.0
5.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (3.11.15)
Requirement already satisfied: setuptools in /usr/local/lib/python3.11/dist-packages (from lightning-utilities<2.0,>=0.10.0->li
ghtning<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon) (80.9.0)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib<3.11,>=3.7.0->autog
luon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (1.3.2)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.11/dist-packages (from matplotlib<3.11,>=3.7.0->autogluo
n.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.11/dist-packages (from matplotlib<3.11,>=3.7.0->auto
gluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (4.58.1)
Requirement already satisfied: kiwisolver>=1.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib<3.11,>=3.7.0->auto
gluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (1.4.8)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.11/dist-packages (from matplotlib<3.11,>=3.7.0->autog
luon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (3.2.3)
Requirement already satisfied: numba in /usr/local/lib/python3.11/dist-packages (from mlforecast<0.14,>0.13->autogluon.timeseri
es==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.60.0)
Collecting optuna (from mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading optuna-4.3.0-py3-none-any.whl.metadata (17 kB)
Collecting window-ops (from mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading window_ops-0.0.15-py3-none-any.whl.metadata (6.8 kB)
Requirement already satisfied: gdown>=4.0.0 in /usr/local/lib/python3.11/dist-packages (from nlpaug<1.2.0,>=1.1.10->autogluon.m
ultimodal==1.3.1->autogluon) (5.2.0)
Requirement already satisfied: click in /usr/local/lib/python3.11/dist-packages (from nltk<3.9,>=3.4.5->autogluon.multimodal==
1.3.1->autogluon) (8.2.1)
Requirement already satisfied: regex>=2021.8.3 in /usr/local/lib/python3.11/dist-packages (from nltk<3.9,>=3.4.5->autogluon.mul
timodal==1.3.1->autogluon) (2024.11.6)
Requirement already satisfied: antlr4-python3-runtime==4.9.* in /usr/local/lib/python3.11/dist-packages (from omegaconf<2.4.0,>
=2.1.1->autogluon.multimodal==1.3.1->autogluon) (4.9.3)
Collecting colorama (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading colorama-0.4.6-py2.py3-none-any.whl.metadata (17 kB)
Collecting model-index (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading model_index-0.1.11-py3-none-any.whl.metadata (3.9 kB)
```

```
Collecting opendatalab (from openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading opendatalab-0.0.10-py3-none-any.whl.metadata (6.4 kB)
Requirement already satisfied: rich in /usr/local/lib/python3.11/dist-packages (from openmim<0.4.0,>=0.3.7->autogluon.multimoda
l==1.3.1->autogluon) (13.9.4)
Requirement already satisfied: tabulate in /usr/local/lib/python3.11/dist-packages (from openmim<0.4.0,>=0.3.7->autogluon.multi
modal==1.3.1->autogluon) (0.9.0)
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.11/dist-packages (from pandas<2.3.0,>=2.0.0->autogluon.co
re==1.3.1->autogluon.core[all]==1.3.1->autogluon) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.11/dist-packages (from pandas<2.3.0,>=2.0.0->autogluon.
core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (2025.2)
Requirement already satisfied: annotated-types>=0.6.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.7->gluont
s<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.7.0)
Requirement already satisfied: pydantic-core==2.33.2 in /usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.7->gluonts
<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (2.33.2)
Requirement already satisfied: typing-inspection>=0.4.0 in /usr/local/lib/python3.11/dist-packages (from pydantic<3,>=1.7->gluo
nts<0.17,>=0.15.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.4.1)
Requirement already satisfied: filelock in /usr/local/lib/python3.11/dist-packages (from ray<2.45,>=2.10.0->ray[default]<2.45,>
=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (3.18.0)
Requirement already satisfied: msgpack<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from ray<2.45,>=2.10.0->ray[de
fault]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.1.0)
Requirement already satisfied: protobuf!=3.19.5,>=3.15.3 in /usr/local/lib/python3.11/dist-packages (from ray<2.45,>=2.10.0->ra
y[default]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (5.29.5)
Requirement already satisfied: aiosignal in /usr/local/lib/python3.11/dist-packages (from ray<2.45,>=2.10.0->ray[default]<2.45,
>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.3.2)
Requirement already satisfied: frozenlist in /usr/local/lib/python3.11/dist-packages (from ray<2.45,>=2.10.0->ray[default]<2.4
5,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.6.0)
Collecting aiohttp_cors (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading aiohttp_cors-0.8.1-py3-none-any.whl.metadata (20 kB)
Collecting colorful (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading colorful-0.5.6-py2.py3-none-any.whl.metadata (16 kB)
Collecting py-spy>=0.2.0 (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading py_spy-0.4.0-py2.py3-none-manylinux_2_5_x86_64.manylinux1_x86_64.whl.metadata (16 kB)
Requirement already satisfied: grpcio>=1.42.0 in /usr/local/lib/python3.11/dist-packages (from ray[default,tune]<2.45,>=2.10.0;
extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.72.1)
Collecting opencensus (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading opencensus-0.11.4-py2.py3-none-any.whl.metadata (12 kB)
Requirement already satisfied: prometheus_client>=0.7.1 in /usr/local/lib/python3.11/dist-packages (from ray[default,tune]<2.4
5,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (0.22.1)
Requirement already satisfied: smart_open in /usr/local/lib/python3.11/dist-packages (from ray[default,tune]<2.45,>=2.10.0; ext
ra == "all"->autogluon.core[all]==1.3.1->autogluon) (7.1.0)
Collecting virtualenv!=20.21.1,>=20.0.24 (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->aut
```

```
ogluon)
  Downloading virtualenv-20.31.2-py3-none-any.whl.metadata (4.5 kB)
Collecting tensorboardX>=1.9 (from ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl.metadata (5.8 kB)
Requirement already satisfied: charset-normalizer<4,>=2 in /usr/local/lib/python3.11/dist-packages (from requests->autogluon.co
re==1.3.1->autogluon.core[all]==1.3.1->autogluon) (3.4.2)
Requirement already satisfied: idna<4,>=2.5 in /usr/local/lib/python3.11/dist-packages (from requests->autogluon.core==1.3.1->a
utogluon.core[all]==1.3.1->autogluon) (3.10)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.11/dist-packages (from requests->autogluon.core==1.
3.1->autogluon.core[all]==1.3.1->autogluon) (2025.4.26)
Requirement already satisfied: imageio!=2.35.0,>=2.33 in /usr/local/lib/python3.11/dist-packages (from scikit-image<0.26.0,>=0.
19.1->autogluon.multimodal==1.3.1->autogluon) (2.37.0)
Requirement already satisfied: tifffile>=2022.8.12 in /usr/local/lib/python3.11/dist-packages (from scikit-image<0.26.0,>=0.19.
1->autogluon.multimodal==1.3.1->autogluon) (2025.6.1)
Requirement already satisfied: lazy-loader>=0.4 in /usr/local/lib/python3.11/dist-packages (from scikit-image<0.26.0,>=0.19.1->
autogluon.multimodal==1.3.1->autogluon) (0.4)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn<1.7.0,>=1.4.0
->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon) (3.6.0)
Requirement already satisfied: spacy-legacy<3.1.0,>=3.0.11 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluo
n.tabular[all]==1.3.1->autogluon) (3.0.12)
Requirement already satisfied: spacy-loggers<2.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluo
n.tabular[all]==1.3.1->autogluon) (1.0.5)
Requirement already satisfied: murmurhash<1.1.0,>=0.28.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.
tabular[all]==1.3.1->autogluon) (1.0.13)
Requirement already satisfied: cymem<2.1.0,>=2.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabula
r[all]==1.3.1->autogluon) (2.0.11)
Requirement already satisfied: preshed<3.1.0,>=3.0.2 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabu
lar[all]==1.3.1->autogluon) (3.0.10)
Requirement already satisfied: thinc<8.4.0,>=8.3.4 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabula
r[all]==1.3.1->autogluon) (8.3.6)
Requirement already satisfied: wasabi<1.2.0,>=0.9.1 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabul
ar[all]==1.3.1->autogluon) (1.1.3)
Requirement already satisfied: srsly<3.0.0,>=2.4.3 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabula
r[all]==1.3.1->autogluon) (2.5.1)
Requirement already satisfied: catalogue<2.1.0,>=2.0.6 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.ta
bular[all]==1.3.1->autogluon) (2.0.10)
Requirement already satisfied: weasel<0.5.0,>=0.1.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabul
ar[all]==1.3.1->autogluon) (0.4.1)
Requirement already satisfied: typer<1.0.0,>=0.3.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.tabula
r[all]==1.3.1->autogluon) (0.16.0)
Requirement already satisfied: langcodes<4.0.0,>=3.2.0 in /usr/local/lib/python3.11/dist-packages (from spacy<3.9->autogluon.ta
```

```
bular[all]==1.3.1->autogluon) (3.5.0)
Requirement already satisfied: language-data>=1.2 in /usr/local/lib/python3.11/dist-packages (from langcodes<4.0.0,>=3.2.0->spa
cy<3.9->autogluon.tabular[all]==1.3.1->autogluon) (1.3.0)
Requirement already satisfied: statsmodels>=0.13.2 in /usr/local/lib/python3.11/dist-packages (from statsforecast<2.0.2,>=1.7.0
->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.14.4)
Requirement already satisfied: absl-py>=0.4 in /usr/local/lib/python3.11/dist-packages (from tensorboard<3,>=2.9->autogluon.mul
timodal==1.3.1->autogluon) (1.4.0)
Requirement already satisfied: markdown>=2.6.8 in /usr/local/lib/python3.11/dist-packages (from tensorboard<3,>=2.9->autogluon.
multimodal==1.3.1->autogluon) (3.8)
Requirement already satisfied: tensorboard-data-server<0.8.0,>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from tensorboa
rd<3,>=2.9->autogluon.multimodal==1.3.1->autogluon) (0.7.2)
Requirement already satisfied: werkzeug>=1.0.1 in /usr/local/lib/python3.11/dist-packages (from tensorboard<3,>=2.9->autogluon.
multimodal==1.3.1->autogluon) (3.1.3)
Requirement already satisfied: blis<1.4.0,>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->spacy<
3.9->autogluon.tabular[all]==1.3.1->autogluon) (1.3.0)
Requirement already satisfied: confection<1.0.0,>=0.0.1 in /usr/local/lib/python3.11/dist-packages (from thinc<8.4.0,>=8.3.4->s
pacy<3.9->autogluon.tabular[all]==1.3.1->autogluon) (0.1.5)
Collecting numpy<2.3.0,>=1.25.0 (from autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading numpy-2.1.3-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (62 kB)
Collecting nvidia-cuda-nvrtc-cu12==12.4.127 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-runtime-cu12==12.4.127 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cuda-cupti-cu12==12.4.127 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cudnn-cu12==9.1.0.70 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cublas-cu12==12.4.5.8 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cufft-cu12==11.2.1.3 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-curand-cu12==10.3.5.147 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Collecting nvidia-cusolver-cu12==11.6.1.9 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Collecting nvidia-cusparse-cu12==12.3.1.170 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-manylinux2014_x86_64.whl.metadata (1.6 kB)
Requirement already satisfied: nvidia-cusparselt-cu12==0.6.2 in /usr/local/lib/python3.11/dist-packages (from torch<2.7,>=2.2->
autogluon.multimodal==1.3.1->autogluon) (0.6.2)
Requirement already satisfied: nvidia-nccl-cu12==2.21.5 in /usr/local/lib/python3.11/dist-packages (from torch<2.7,>=2.2->autog
luon.multimodal==1.3.1->autogluon) (2.21.5)
```

```
Requirement already satisfied: nvidia-nvtx-cu12==12.4.127 in /usr/local/lib/python3.11/dist-packages (from torch<2.7,>=2.2->aut
ogluon.multimodal==1.3.1->autogluon) (12.4.127)
Collecting nvidia-nvjitlink-cu12==12.4.127 (from torch<2.7,>=2.2->autogluon.multimodal==1.3.1->autogluon)
  Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl.metadata (1.5 kB)
Requirement already satisfied: triton==3.2.0 in /usr/local/lib/python3.11/dist-packages (from torch<2.7,>=2.2->autogluon.multim
odal==1.3.1->autogluon) (3.2.0)
Requirement already satisfied: sympy==1.13.1 in /usr/local/lib/python3.11/dist-packages (from torch<2.7,>=2.2->autogluon.multim
odal==1.3.1->autogluon) (1.13.1)
Requirement already satisfied: mpmath<1.4,>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from sympy==1.13.1->torch<2.7,>=
2.2->autogluon.multimodal==1.3.1->autogluon) (1.3.0)
Requirement already satisfied: tokenizers<0.22,>=0.21 in /usr/local/lib/python3.11/dist-packages (from transformers<4.50,>=4.3
8.0->transformers[sentencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.1->autogluon) (0.21.1)
Requirement already satisfied: hf-xet<2.0.0,>=1.1.2 in /usr/local/lib/python3.11/dist-packages (from huggingface-hub[torch]->au
togluon.tabular[all]==1.3.1->autogluon) (1.1.2)
Requirement already satisfied: sentencepiece!=0.1.92,>=0.1.91 in /usr/local/lib/python3.11/dist-packages (from transformers[sen
tencepiece]<4.50,>=4.38.0->autogluon.multimodal==1.3.1->autogluon) (0.2.0)
Requirement already satisfied: shellingham>=1.3.0 in /usr/local/lib/python3.11/dist-packages (from typer<1.0.0,>=0.3.0->spacy<
3.9->autogluon.tabular[all]==1.3.1->autogluon) (1.5.4)
Requirement already satisfied: cloudpathlib<1.0.0,>=0.7.0 in /usr/local/lib/python3.11/dist-packages (from weasel<0.5.0,>=0.1.0
->spacy<3.9->autogluon.tabular[all]==1.3.1->autogluon) (0.21.1)
Requirement already satisfied: wrapt in /usr/local/lib/python3.11/dist-packages (from smart_open->ray[default,tune]<2.45,>=2.1
0.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.17.2)
Requirement already satisfied: aiohappyeyeballs>=2.3.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.
0a1->fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (2.6.1)
Requirement already satisfied: multidict<7.0,>=4.5 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1-
>fsspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (6.4.4)
Requirement already satisfied: propcache>=0.2.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->fs
spec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (0.3.1)
Requirement already satisfied: yarl<2.0,>=1.17.0 in /usr/local/lib/python3.11/dist-packages (from aiohttp!=4.0.0a0,!=4.0.0a1->f
sspec[http]>=2021.05.0->evaluate<0.5.0,>=0.4.0->autogluon.multimodal==1.3.1->autogluon) (1.20.0)
Collecting triad>=0.9.7 (from fugue>=0.9.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading triad-0.9.8-py3-none-any.whl.metadata (6.3 kB)
Collecting adagio>=0.2.4 (from fugue>=0.9.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading adagio-0.2.6-py3-none-any.whl.metadata (1.8 kB)
Requirement already satisfied: beautifulsoup4 in /usr/local/lib/python3.11/dist-packages (from gdown>=4.0.0->nlpaug<1.2.0,>=1.
1.10->autogluon.multimodal==1.3.1->autogluon) (4.13.4)
Requirement already satisfied: marisa-trie>=1.1.0 in /usr/local/lib/python3.11/dist-packages (from language-data>=1.2->langcode
s<4.0.0,>=3.2.0->spacy<3.9->autogluon.tabular[all]==1.3.1->autogluon) (1.2.1)
Requirement already satisfied: llvmlite<0.44,>=0.43.0dev0 in /usr/local/lib/python3.11/dist-packages (from numba->mlforecast<0.
14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (0.43.0)
Collecting numpy<2.3.0,>=1.25.0 (from autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
```

```
  Downloading numpy-2.0.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (60 kB)
Requirement already satisfied: markdown-it-py>=2.2.0 in /usr/local/lib/python3.11/dist-packages (from rich->openmim<0.4.0,>=0.
3.7->autogluon.multimodal==1.3.1->autogluon) (3.0.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in /usr/local/lib/python3.11/dist-packages (from rich->openmim<0.4.0,>=
0.3.7->autogluon.multimodal==1.3.1->autogluon) (2.19.1)
Requirement already satisfied: mdurl~=0.1 in /usr/local/lib/python3.11/dist-packages (from markdown-it-py>=2.2.0->rich->openmim
<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon) (0.1.2)
Requirement already satisfied: patsy>=0.5.6 in /usr/local/lib/python3.11/dist-packages (from statsmodels>=0.13.2->statsforecast
<2.0.2,>=1.7.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (1.0.1)
Collecting fs (from triad>=0.9.7->fugue>=0.9.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading fs-2.4.16-py2.py3-none-any.whl.metadata (6.3 kB)
Collecting distlib<1,>=0.3.7 (from virtualenv!=20.21.1,>=20.0.24->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.co
re[all]==1.3.1->autogluon)
  Downloading distlib-0.3.9-py2.py3-none-any.whl.metadata (5.2 kB)
Requirement already satisfied: platformdirs<5,>=3.9.1 in /usr/local/lib/python3.11/dist-packages (from virtualenv!=20.21.1,>=2
0.0.24->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (4.3.8)
Requirement already satisfied: soupsieve>1.2 in /usr/local/lib/python3.11/dist-packages (from beautifulsoup4->gdown>=4.0.0->nlp
aug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.1->autogluon) (2.7)
Collecting appdirs~=1.4.3 (from fs->triad>=0.9.7->fugue>=0.9.0->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->
autogluon)
  Downloading appdirs-1.4.4-py2.py3-none-any.whl.metadata (9.0 kB)
Collecting ordered-set (from model-index->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading ordered_set-4.1.0-py3-none-any.whl.metadata (5.3 kB)
Collecting opencensus-context>=0.1.3 (from opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.
3.1->autogluon)
  Downloading opencensus_context-0.1.3-py2.py3-none-any.whl.metadata (3.3 kB)
Requirement already satisfied: google-api-core<3.0.0,>=1.0.0 in /usr/local/lib/python3.11/dist-packages (from opencensus->ray[d
efault,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (2.25.0)
Requirement already satisfied: googleapis-common-protos<2.0.0,>=1.56.2 in /usr/local/lib/python3.11/dist-packages (from google-
api-core<3.0.0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.
70.0)
Requirement already satisfied: proto-plus<2.0.0,>=1.22.3 in /usr/local/lib/python3.11/dist-packages (from google-api-core<3.0.
0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (1.26.1)
Requirement already satisfied: google-auth<3.0.0,>=2.14.1 in /usr/local/lib/python3.11/dist-packages (from google-api-core<3.0.
0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon) (2.38.0)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in /usr/local/lib/python3.11/dist-packages (from google-auth<3.0.0,>=2.1
4.1->google-api-core<3.0.0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->au
togluon) (5.5.2)
Requirement already satisfied: pyasn1-modules>=0.2.1 in /usr/local/lib/python3.11/dist-packages (from google-auth<3.0.0,>=2.14.
1->google-api-core<3.0.0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->auto
gluon) (0.4.2)
```

```
Requirement already satisfied: rsa<5,>=3.1.4 in /usr/local/lib/python3.11/dist-packages (from google-auth<3.0.0,>=2.14.1->googl
e-api-core<3.0.0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
(4.9.1)
Requirement already satisfied: pyasn1>=0.1.3 in /usr/local/lib/python3.11/dist-packages (from rsa<5,>=3.1.4->google-auth<3.0.0,
>=2.14.1->google-api-core<3.0.0,>=1.0.0->opencensus->ray[default,tune]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.
1->autogluon) (0.6.1)
Collecting pycryptodome (from opendatalab->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading pycryptodome-3.23.0-cp37-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (3.4 kB)
Collecting openxlab (from opendatalab->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading openxlab-0.1.2-py3-none-any.whl.metadata (3.8 kB)
Collecting filelock (from ray<2.45,>=2.10.0->ray[default]<2.45,>=2.10.0; extra == "all"->autogluon.core[all]==1.3.1->autogluon)
  Downloading filelock-3.14.0-py3-none-any.whl.metadata (2.8 kB)
Collecting oss2~=2.17.0 (from openxlab->opendatalab->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading oss2-2.17.0.tar.gz (259 kB)
  Preparing metadata (setup.py) ... done
Collecting pytz>=2020.1 (from pandas<2.3.0,>=2.0.0->autogluon.core==1.3.1->autogluon.core[all]==1.3.1->autogluon)
  Downloading pytz-2023.4-py2.py3-none-any.whl.metadata (22 kB)
INFO: pip is looking at multiple versions of openxlab to determine which version is compatible with other requirements. This co
uld take a while.
Collecting openxlab (from opendatalab->openmim<0.4.0,>=0.3.7->autogluon.multimodal==1.3.1->autogluon)
  Downloading openxlab-0.1.1-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.1.0-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.38-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.37-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.36-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.35-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.34-py3-none-any.whl.metadata (3.8 kB)
INFO: pip is still looking at multiple versions of openxlab to determine which version is compatible with other requirements. T
his could take a while.
  Downloading openxlab-0.0.33-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.32-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.31-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.30-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.29-py3-none-any.whl.metadata (3.8 kB)
INFO: This is taking longer than usual. You might need to provide the dependency resolver with stricter constraints to reduce r
untime. See https://pip.pypa.io/warnings/backtracking for guidance. If you want to abort this run, press Ctrl + C.
  Downloading openxlab-0.0.28-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.27-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.26-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.25-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.24-py3-none-any.whl.metadata (3.7 kB)
```

```
  Downloading openxlab-0.0.23-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.22-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.21-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.20-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.19-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.18-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.17-py3-none-any.whl.metadata (3.7 kB)
  Downloading openxlab-0.0.16-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.15-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.14-py3-none-any.whl.metadata (3.8 kB)
  Downloading openxlab-0.0.13-py3-none-any.whl.metadata (4.5 kB)
  Downloading openxlab-0.0.12-py3-none-any.whl.metadata (4.5 kB)
  Downloading openxlab-0.0.11-py3-none-any.whl.metadata (4.3 kB)
Collecting alembic>=1.5.0 (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading alembic-1.16.1-py3-none-any.whl.metadata (7.3 kB)
Collecting colorlog (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon)
  Downloading colorlog-6.9.0-py3-none-any.whl.metadata (10 kB)
Requirement already satisfied: sqlalchemy>=1.4.2 in /usr/local/lib/python3.11/dist-packages (from optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (2.0.41)
Requirement already satisfied: Mako in /usr/lib/python3/dist-packages (from alembic>=1.5.0->optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (1.1.3)
Requirement already satisfied: greenlet>=1 in /usr/local/lib/python3.11/dist-packages (from sqlalchemy>=1.4.2->optuna->mlforecast<0.14,>0.13->autogluon.timeseries==1.3.1->autogluon.timeseries[all]==1.3.1->autogluon) (3.2.2)
Requirement already satisfied: tenacity>=6.2.0 in /usr/local/lib/python3.11/dist-packages (from plotly->catboost<1.3,>=1.2->autogluon.tabular[all]==1.3.1->autogluon) (9.1.2)
Requirement already satisfied: PySocks!=1.5.7,>=1.5.6 in /usr/local/lib/python3.11/dist-packages (from requests[socks]->gdown>=4.0.0->nlpaug<1.2.0,>=1.1.10->autogluon.multimodal==1.3.1->autogluon) (1.7.1)
Downloading autogluon-1.3.1-py3-none-any.whl (9.8 kB)
Downloading autogluon.core-1.3.1-py3-none-any.whl (222 kB)
Downloading autogluon.common-1.3.1-py3-none-any.whl (69 kB)
Downloading autogluon.features-1.3.1-py3-none-any.whl (64 kB)
Downloading autogluon.multimodal-1.3.1-py3-none-any.whl (454 kB)
Downloading autogluon.tabular-1.3.1-py3-none-any.whl (382 kB)
Downloading autogluon.timeseries-1.3.1-py3-none-any.whl (181 kB)
Downloading boto3-1.38.32-py3-none-any.whl (139 kB)
Downloading botocore-1.38.32-py3-none-any.whl (13.6 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.6/13.6 MB 280.1 MB/s eta 0:00:00
Downloading catboost-1.2.8-cp311-cp311-manylinux2014_x86_64.whl (99.2 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 99.2/99.2 MB 287.2 MB/s eta 0:00:00
```

```
Downloading coreforecast-0.0.15-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (275 kB)
Downloading evaluate-0.4.3-py3-none-any.whl (84 kB)
Downloading gluonts-0.16.1-py3-none-any.whl (1.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.5/1.5 MB 252.7 MB/s eta 0:00:00
Downloading jmespath-1.0.1-py3-none-any.whl (20 kB)
Downloading jsonschema-4.23.0-py3-none-any.whl (88 kB)
Downloading lightning-2.5.1.post0-py3-none-any.whl (819 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 819.0/819.0 kB 257.1 MB/s eta 0:00:00
Downloading lightning_utilities-0.14.3-py3-none-any.whl (28 kB)
Downloading mlforecast-0.13.6-py3-none-any.whl (71 kB)
Downloading nlpaug-1.1.11-py3-none-any.whl (410 kB)
Downloading nltk-3.8.1-py3-none-any.whl (1.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 1.5/1.5 MB 299.7 MB/s eta 0:00:00
Downloading openmim-0.3.9-py2.py3-none-any.whl (52 kB)
Downloading pdf2image-1.17.0-py3-none-any.whl (11 kB)
Downloading pytesseract-0.3.13-py3-none-any.whl (14 kB)
Downloading pytorch_metric_learning-2.8.1-py3-none-any.whl (125 kB)
Downloading ray-2.44.1-cp311-cp311-manylinux2014_x86_64.whl (68.1 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 68.1/68.1 MB 180.3 MB/s eta 0:00:00
Downloading s3transfer-0.13.0-py3-none-any.whl (85 kB)
Downloading statsforecast-2.0.1-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (354 kB)
Downloading timm-1.0.3-py3-none-any.whl (2.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.3/2.3 MB 401.8 MB/s eta 0:00:00
Downloading nvidia_cublas_cu12-12.4.5.8-py3-none-manylinux2014_x86_64.whl (363.4 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 363.4/363.4 MB 198.7 MB/s eta 0:00:00
Downloading nvidia_cuda_cupti_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (13.8 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.8/13.8 MB 169.5 MB/s eta 0:00:00
Downloading nvidia_cuda_nvrtc_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (24.6 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 24.6/24.6 MB 135.1 MB/s eta 0:00:00
Downloading nvidia_cuda_runtime_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (883 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 883.7/883.7 kB 376.2 MB/s eta 0:00:00
Downloading nvidia_cudnn_cu12-9.1.0.70-py3-none-manylinux2014_x86_64.whl (664.8 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 664.8/664.8 MB 93.4 MB/s eta 0:00:00
Downloading nvidia_cufft_cu12-11.2.1.3-py3-none-manylinux2014_x86_64.whl (211.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 211.5/211.5 MB 59.8 MB/s eta 0:00:00
Downloading nvidia_curand_cu12-10.3.5.147-py3-none-manylinux2014_x86_64.whl (56.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 56.3/56.3 MB 93.4 MB/s eta 0:00:00
Downloading nvidia_cusolver_cu12-11.6.1.9-py3-none-manylinux2014_x86_64.whl (127.9 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 127.9/127.9 MB 106.5 MB/s eta 0:00:00
Downloading nvidia_cusparse_cu12-12.3.1.170-py3-none-manylinux2014_x86_64.whl (207.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 207.5/207.5 MB 53.4 MB/s eta 0:00:00
```

```
Downloading nvidia_nvjitlink_cu12-12.4.127-py3-none-manylinux2014_x86_64.whl (21.1 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 21.1/21.1 MB 135.9 MB/s eta 0:00:00
Downloading torchmetrics-1.7.2-py3-none-any.whl (962 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 962.5/962.5 kB 298.8 MB/s eta 0:00:00
Downloading transformers-4.49.0-py3-none-any.whl (10.0 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 10.0/10.0 MB 129.7 MB/s eta 0:00:00
Downloading utilsforecast-0.2.10-py3-none-any.whl (41 kB)
Downloading fugue-0.9.1-py3-none-any.whl (278 kB)
Downloading adagio-0.2.6-py3-none-any.whl (19 kB)
Downloading numpy-2.0.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (19.5 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 19.5/19.5 MB 145.5 MB/s eta 0:00:00
Downloading py_spy-0.4.0-py2.py3-none-manylinux_2_5_x86_64.manylinux1_x86_64.whl (2.7 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.7/2.7 MB 151.4 MB/s eta 0:00:00
Downloading tensorboardX-2.6.2.2-py2.py3-none-any.whl (101 kB)
Downloading triad-0.9.8-py3-none-any.whl (62 kB)
Downloading virtualenv-20.31.2-py3-none-any.whl (6.1 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 6.1/6.1 MB 149.2 MB/s eta 0:00:00
Downloading distlib-0.3.9-py2.py3-none-any.whl (468 kB)
Downloading aiohttp_cors-0.8.1-py3-none-any.whl (25 kB)
Downloading colorama-0.4.6-py2.py3-none-any.whl (25 kB)
Downloading colorful-0.5.6-py2.py3-none-any.whl (201 kB)
Downloading fs-2.4.16-py2.py3-none-any.whl (135 kB)
Downloading appdirs-1.4.4-py2.py3-none-any.whl (9.6 kB)
Downloading model_index-0.1.11-py3-none-any.whl (34 kB)
Downloading opencensus-0.11.4-py2.py3-none-any.whl (128 kB)
Downloading opencensus_context-0.1.3-py2.py3-none-any.whl (5.1 kB)
Downloading opendatalab-0.0.10-py3-none-any.whl (29 kB)
Downloading openxlab-0.0.11-py3-none-any.whl (55 kB)
Downloading optuna-4.3.0-py3-none-any.whl (386 kB)
Downloading alembic-1.16.1-py3-none-any.whl (242 kB)
Downloading colorlog-6.9.0-py3-none-any.whl (11 kB)
Downloading ordered_set-4.1.0-py3-none-any.whl (7.6 kB)
Downloading pycryptodome-3.23.0-cp37-abi3-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (2.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 2.3/2.3 MB 147.8 MB/s eta 0:00:00
Downloading pytorch_lightning-2.5.1.post0-py3-none-any.whl (823 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 823.1/823.1 kB 351.2 MB/s eta 0:00:00
Downloading window_ops-0.0.15-py3-none-any.whl (15 kB)
Building wheels for collected packages: nvidia-ml-py3, seqeval
  DEPRECATION: Building 'nvidia-ml-py3' using the legacy setup.py bdist_wheel mechanism, which will be removed in a future vers
ion. pip 25.3 will enforce this behaviour change. A possible replacement is to use the standardized build interface by setting
the `--use-pep517` option, (possibly combined with `--no-build-isolation`), or adding a `pyproject.toml` file to the source tre
```

```
e of 'nvidia-ml-py3'. Discussion can be found at https://github.com/pypa/pip/issues/6334
  Building wheel for nvidia-ml-py3 (setup.py) ... done
  Created wheel for nvidia-ml-py3: filename=nvidia_ml_py3-7.352.0-py3-none-any.whl size=19208 sha256=2b9a27175b37572ee56799dedb
289095bd8e599509439df584794d85e4308005
  Stored in directory: /tmp/pip-ephem-wheel-cache-cc44ktup/wheels/47/50/9e/29dc79037d74c3c1bb4a8661fb608e8674b7e4260d6a3f8f51
  DEPRECATION: Building 'seqeval' using the legacy setup.py bdist_wheel mechanism, which will be removed in a future version. p
ip 25.3 will enforce this behaviour change. A possible replacement is to use the standardized build interface by setting the `-
-use-pep517` option, (possibly combined with `--no-build-isolation`), or adding a `pyproject.toml` file to the source tree of
'seqeval'. Discussion can be found at https://github.com/pypa/pip/issues/6334
  Building wheel for seqeval (setup.py) ... done
  Created wheel for seqeval: filename=seqeval-1.2.2-py3-none-any.whl size=16250 sha256=cee2d6b2982c439b242f0aa13d0ca285f2616b6c
f92318e5ef9cb796b54f3c52
  Stored in directory: /tmp/pip-ephem-wheel-cache-cc44ktup/wheels/bc/92/f0/243288f899c2eacdfa8c5f9aede4c71a9bad0ee26a01dc5ead
Successfully built nvidia-ml-py3 seqeval
Installing collected packages: py-spy, opencensus-context, nvidia-ml-py3, distlib, colorful, appdirs, virtualenv, pytesseract,
pycryptodome, pdf2image, ordered-set, openxlab, nvidia-nvjitlink-cu12, nvidia-curand-cu12, nvidia-cufft-cu12, nvidia-cuda-runti
me-cu12, nvidia-cuda-nvrtc-cu12, nvidia-cuda-cupti-cu12, nvidia-cublas-cu12, numpy, nltk, lightning-utilities, jmespath, fs, co
lorlog, colorama, tensorboardX, nvidia-cusparse-cu12, nvidia-cudnn-cu12, model-index, coreforecast, botocore, alembic, window-o
ps, utilsforecast, triad, s3transfer, optuna, opendatalab, nvidia-cusolver-cu12, jsonschema, gluonts, aiohttp_cors, transformer
s, seqeval, ray, openmim, opencensus, nlpaug, mlforecast, catboost, boto3, adagio, torchmetrics, pytorch-metric-learning, fugu
e, evaluate, autogluon.common, timm, statsforecast, pytorch-lightning, autogluon.features, autogluon.core, lightning, autogluo
n.tabular, autogluon.multimodal, autogluon.timeseries, autogluon
  Attempting uninstall: nvidia-nvjitlink-cu12
    Found existing installation: nvidia-nvjitlink-cu12 12.5.82
    Uninstalling nvidia-nvjitlink-cu12-12.5.82:
      Successfully uninstalled nvidia-nvjitlink-cu12-12.5.82
  Attempting uninstall: nvidia-curand-cu12
    Found existing installation: nvidia-curand-cu12 10.3.6.82
    Uninstalling nvidia-curand-cu12-10.3.6.82:
      Successfully uninstalled nvidia-curand-cu12-10.3.6.82
  Attempting uninstall: nvidia-cufft-cu12
    Found existing installation: nvidia-cufft-cu12 11.2.3.61
    Uninstalling nvidia-cufft-cu12-11.2.3.61:
      Successfully uninstalled nvidia-cufft-cu12-11.2.3.61
  Attempting uninstall: nvidia-cuda-runtime-cu12
    Found existing installation: nvidia-cuda-runtime-cu12 12.5.82
    Uninstalling nvidia-cuda-runtime-cu12-12.5.82:
      Successfully uninstalled nvidia-cuda-runtime-cu12-12.5.82
  Attempting uninstall: nvidia-cuda-nvrtc-cu12
    Found existing installation: nvidia-cuda-nvrtc-cu12 12.5.82
    Uninstalling nvidia-cuda-nvrtc-cu12-12.5.82:
```

```
    Successfully uninstalled nvidia-cuda-nvrtc-cu12-12.5.82
Attempting uninstall: nvidia-cuda-cupti-cu12
  Found existing installation: nvidia-cuda-cupti-cu12 12.5.82
  Uninstalling nvidia-cuda-cupti-cu12-12.5.82:
    Successfully uninstalled nvidia-cuda-cupti-cu12-12.5.82
Attempting uninstall: nvidia-cublas-cu12
  Found existing installation: nvidia-cublas-cu12 12.5.3.2
  Uninstalling nvidia-cublas-cu12-12.5.3.2:
    Successfully uninstalled nvidia-cublas-cu12-12.5.3.2
Attempting uninstall: numpy
  Found existing installation: numpy 1.26.4
  Uninstalling numpy-1.26.4:
    Successfully uninstalled numpy-1.26.4
Attempting uninstall: nltk
  Found existing installation: nltk 3.9.1
  Uninstalling nltk-3.9.1:
    Successfully uninstalled nltk-3.9.1
Attempting uninstall: nvidia-cusparse-cu12
  Found existing installation: nvidia-cusparse-cu12 12.5.1.3
  Uninstalling nvidia-cusparse-cu12-12.5.1.3:
    Successfully uninstalled nvidia-cusparse-cu12-12.5.1.3
Attempting uninstall: nvidia-cudnn-cu12
  Found existing installation: nvidia-cudnn-cu12 9.3.0.75
  Uninstalling nvidia-cudnn-cu12-9.3.0.75:
    Successfully uninstalled nvidia-cudnn-cu12-9.3.0.75
Attempting uninstall: nvidia-cusolver-cu12
  Found existing installation: nvidia-cusolver-cu12 11.6.3.83
  Uninstalling nvidia-cusolver-cu12-11.6.3.83:
    Successfully uninstalled nvidia-cusolver-cu12-11.6.3.83
Attempting uninstall: jsonschema
  Found existing installation: jsonschema 4.24.0
  Uninstalling jsonschema-4.24.0:
    Successfully uninstalled jsonschema-4.24.0
Attempting uninstall: transformers
  Found existing installation: transformers 4.52.4
  Uninstalling transformers-4.52.4:
    Successfully uninstalled transformers-4.52.4
Attempting uninstall: timm
  Found existing installation: timm 1.0.15
  Uninstalling timm-1.0.15:
    Successfully uninstalled timm-1.0.15
```

```
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 68/68 [autogluon]
ERROR: pip's dependency resolver does not currently take into account all the packages that are installed. This behaviour is th
e source of the following dependency conflicts.
mxnet 1.9.1 requires numpy<2.0.0,>1.16.0, but you have numpy 2.0.2 which is incompatible.
panel 1.7.1 requires bokeh<3.8.0,>=3.5.0, but you have bokeh 2.0.1 which is incompatible.
textblob 0.19.0 requires nltk>=3.9, but you have nltk 3.8.1 which is incompatible.
holoviews 1.20.2 requires bokeh>=3.1, but you have bokeh 2.0.1 which is incompatible.
Successfully installed adagio-0.2.6 aiohttp_cors-0.8.1 alembic-1.16.1 appdirs-1.4.4 autogluon-1.3.1 autogluon.common-1.3.1 auto
gluon.core-1.3.1 autogluon.features-1.3.1 autogluon.multimodal-1.3.1 autogluon.tabular-1.3.1 autogluon.timeseries-1.3.1 boto3-
1.38.32 botocore-1.38.32 catboost-1.2.8 colorama-0.4.6 colorful-0.5.6 colorlog-6.9.0 coreforecast-0.0.15 distlib-0.3.9 evaluate
-0.4.3 fs-2.4.16 fugue-0.9.1 gluonts-0.16.1 jmespath-1.0.1 jsonschema-4.23.0 lightning-2.5.1.post0 lightning-utilities-0.14.3 m
lforecast-0.13.6 model-index-0.1.11 nlpaug-1.1.11 nltk-3.8.1 numpy-2.0.2 nvidia-cublas-cu12-12.4.5.8 nvidia-cuda-cupti-cu12-12.
4.127 nvidia-cuda-nvrtc-cu12-12.4.127 nvidia-cuda-runtime-cu12-12.4.127 nvidia-cudnn-cu12-9.1.0.70 nvidia-cufft-cu12-11.2.1.3 n
vidia-curand-cu12-10.3.5.147 nvidia-cusolver-cu12-11.6.1.9 nvidia-cusparse-cu12-12.3.1.170 nvidia-ml-py3-7.352.0 nvidia-nvjitli
nk-cu12-12.4.127 opencensus-0.11.4 opencensus-context-0.1.3 opendatalab-0.0.10 openmim-0.3.9 openxlab-0.0.11 optuna-4.3.0 order
ed-set-4.1.0 pdf2image-1.17.0 py-spy-0.4.0 pycryptodome-3.23.0 pytesseract-0.3.13 pytorch-lightning-2.5.1.post0 pytorch-metric-
learning-2.8.1 ray-2.44.1 s3transfer-0.13.0 seqeval-1.2.2 statsforecast-2.0.1 tensorboardX-2.6.2.2 timm-1.0.3 torchmetrics-1.7.
2 transformers-4.49.0 triad-0.9.8 utilsforecast-0.2.10 virtualenv-20.31.2 window-ops-0.0.15
```

Setup Kaggle API Key

In [1]:

In [7]:
```
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
```

In [8]:
```
!kaggle competitions list
```

```
ref                                                                          deadline             category         
reward  teamCount  userHasEntered
---------------------------------------------------------------------------  -------------------  ---------------  ------
-------  ---------  --------------
https://www.kaggle.com/competitions/arc-prize-2025                           2025-11-03 23:59:00  Featured         1,000,
000 Usd        474           False
https://www.kaggle.com/competitions/openai-to-z-challenge                    2025-06-29 23:59:00  Featured           400,
000 Usd          0           False
https://www.kaggle.com/competitions/waveform-inversion                       2025-06-30 23:59:00  Research            50,
000 Usd       1027           False
https://www.kaggle.com/competitions/cmi-detect-behavior-with-sensor-data     2025-09-02 23:59:00  Featured            50,
000 Usd        621           False
https://www.kaggle.com/competitions/meta-kaggle-hackathon                    2025-07-21 23:59:00  Featured            50,
000 Usd          0           False
https://www.kaggle.com/competitions/playground-series-s5e6                   2025-06-30 23:59:00  Playground
Swag        864           False
https://www.kaggle.com/competitions/titanic                                  2030-01-01 00:00:00  Getting Started      Kn
owledge      15783           False
https://www.kaggle.com/competitions/home-data-for-ml-course                  2030-01-01 23:59:00  Getting Started      Kn
owledge       5489           False
https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques    2030-01-01 00:00:00  Getting Started      Kn
owledge       4777           False
https://www.kaggle.com/competitions/spaceship-titanic                        2030-01-01 00:00:00  Getting Started      Kn
owledge       1965           False
https://www.kaggle.com/competitions/digit-recognizer                         2030-01-01 00:00:00  Getting Started      Kn
owledge       1421           False
https://www.kaggle.com/competitions/nlp-getting-started                      2030-01-01 00:00:00  Getting Started      Kn
owledge       1009           False
https://www.kaggle.com/competitions/store-sales-time-series-forecasting      2030-06-30 23:59:00  Getting Started      Kn
owledge        857           False
https://www.kaggle.com/competitions/llm-classification-finetuning            2030-07-01 23:59:00  Getting Started      Kn
owledge        305           False
https://www.kaggle.com/competitions/connectx                                 2030-01-01 00:00:00  Getting Started      Kn
owledge        176           False
https://www.kaggle.com/competitions/gan-getting-started                      2030-07-01 23:59:00  Getting Started      Kn
owledge        135           False
https://www.kaggle.com/competitions/contradictory-my-dear-watson             2030-07-01 23:59:00  Getting Started      Kn
owledge         79           False
https://www.kaggle.com/competitions/tpu-getting-started                      2030-06-03 23:59:00  Getting Started      Kn
owledge         38           False
https://www.kaggle.com/competitions/konwinski-prize                          2025-07-09 23:59:00  Featured         1,225,
```

```
000 Usd        617           False
https://www.kaggle.com/competitions/jane-street-real-time-market-data-forecasting  2025-07-12 23:59:00  Featured          120,
000 Usd       3757           False
```

Download and explore dataset

Go to the bike sharing demand competition and agree to the terms

In [9]:
```
# Download the dataset, it will be in a .zip file so you'll need to unzip it as well.
!kaggle competitions download -c bike-sharing-demand
# If you already downloaded it you can use the -o command to overwrite the file
!unzip -o bike-sharing-demand.zip
```

```
Downloading bike-sharing-demand.zip to /content

  0% 0.00/189k [00:00<?, ?B/s]

100% 189k/189k [00:00<00:00, 467MB/s]
Archive:  bike-sharing-demand.zip
  inflating: sampleSubmission.csv
  inflating: test.csv
  inflating: train.csv
```

In [10]:
```
import pandas as pd
from autogluon.tabular import TabularPredictor
```

In [11]:
```
# Create the train dataset in pandas by reading the csv
# Set the parsing of the datetime column so you can use some of the `dt` features in pandas later
train = pd.read_csv("train.csv", parse_dates=["datetime"])
train.head()
```

Out[11]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	13	16
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	32	40
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	27	32
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3	10	13
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0	1	1

In [12]:
```
train.describe().transpose()
```

Out[12]:

	count	mean	min	25%	50%	75%	max	std
datetime	10886	2011-12-27 05:56:22.399411968	2011-01-01 00:00:00	2011-07-02 07:15:00	2012-01-01 20:30:00	2012-07-01 12:45:00	2012-12-19 23:00:00	NaN
season	10886.0	2.506614	1.0	2.0	3.0	4.0	4.0	1.116174
holiday	10886.0	0.028569	0.0	0.0	0.0	0.0	1.0	0.166599
workingday	10886.0	0.680875	0.0	0.0	1.0	1.0	1.0	0.466159
weather	10886.0	1.418427	1.0	1.0	1.0	2.0	4.0	0.633839
temp	10886.0	20.23086	0.82	13.94	20.5	26.24	41.0	7.79159
atemp	10886.0	23.655084	0.76	16.665	24.24	31.06	45.455	8.474601
humidity	10886.0	61.88646	0.0	47.0	62.0	77.0	100.0	19.245033
windspeed	10886.0	12.799395	0.0	7.0015	12.998	16.9979	56.9969	8.164537
casual	10886.0	36.021955	0.0	4.0	17.0	49.0	367.0	49.960477
registered	10886.0	155.552177	0.0	36.0	118.0	222.0	886.0	151.039033
count	10886.0	191.574132	1.0	42.0	145.0	284.0	977.0	181.144454

In [13]:
```
# Create the test pandas dataframe in pandas by reading the csv, remember to parse the datetime!
test = pd.read_csv("test.csv", parse_dates=["datetime"])
test.head()
```

Out[13]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed
0	2011-01-20 00:00:00	1	0	1	1	10.66	11.365	56	26.0027
1	2011-01-20 01:00:00	1	0	1	1	10.66	13.635	56	0.0000
2	2011-01-20 02:00:00	1	0	1	1	10.66	13.635	56	0.0000
3	2011-01-20 03:00:00	1	0	1	1	10.66	12.880	56	11.0014
4	2011-01-20 04:00:00	1	0	1	1	10.66	12.880	56	11.0014

In [14]:
```
test.describe().transpose()
```

Out[14]:

	count	mean	min	25%	50%	75%	max	std
datetime	6493	2012-01-13 09:27:47.765285632	2011-01-20 00:00:00	2011-07-22 15:00:00	2012-01-20 23:00:00	2012-07-20 17:00:00	2012-12-31 23:00:00	NaN
season	6493.0	2.4933	1.0	2.0	3.0	3.0	4.0	1.091258
holiday	6493.0	0.029108	0.0	0.0	0.0	0.0	1.0	0.168123
workingday	6493.0	0.685815	0.0	0.0	1.0	1.0	1.0	0.464226
weather	6493.0	1.436778	1.0	1.0	1.0	2.0	4.0	0.64839
temp	6493.0	20.620607	0.82	13.94	21.32	27.06	40.18	8.059583
atemp	6493.0	24.012865	0.0	16.665	25.0	31.06	50.0	8.782741
humidity	6493.0	64.125212	16.0	49.0	65.0	81.0	100.0	19.293391
windspeed	6493.0	12.631157	0.0	7.0015	11.0014	16.9979	55.9986	8.250151

In [15]:
```
# Same thing as train and test dataset
submission = pd.read_csv("sampleSubmission.csv", parse_dates=["datetime"])
submission.head()
```

Out[15]:

	datetime	count
0	2011-01-20 00:00:00	0
1	2011-01-20 01:00:00	0
2	2011-01-20 02:00:00	0
3	2011-01-20 03:00:00	0
4	2011-01-20 04:00:00	0

Step 3: Train a model using AutoGluon's Tabular Prediction

Requirements:

- We are prediting `count`, so it is the label we are setting.
- Ignore `casual` and `registered` columns as they are also not present in the test dataset.
- Use the `root_mean_squared_error` as the metric to use for evaluation.
- Set a time limit of 10 minutes (600 seconds).
- Use the preset `best_quality` to focus on creating the best model.

In [16]:
```
train_data = train
train_data = train_data.drop(columns=["casual", "registered"])
```

In [17]:
```
predictor = TabularPredictor(
    label="count",
    eval_metric="root_mean_squared_error"
).fit(
    train_data,
    time_limit=600,
```

```
    presets="best_quality"
)
```

```
No path specified. Models will be saved in: "AutogluonModels/ag-20250607_234635"
Verbosity: 2 (Standard Logging)
=================== System Info ===================
AutoGluon Version:  1.3.1
Python Version:     3.11.13
Operating System:   Linux
Platform Machine:   x86_64
Platform Version:   #1 SMP PREEMPT_DYNAMIC Sun Mar 30 16:01:29 UTC 2025
CPU Count:          2
Memory Avail:       11.39 GB / 12.67 GB (89.9%)
Disk Space Avail:   66.55 GB / 107.72 GB (61.8%)
====================================================
Presets specified: ['best_quality']
Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when use_bag_holdout is disabled. (use_bag_holdou
t=False)
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_sets=1
DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether the input data is affected by stacked overf
itting and enable or disable stacking as a consequence.
        This is used to identify the optimal `num_stack_levels` value. Copies of AutoGluon will be fit on subsets of the data.
Then holdout validation data is used to detect stacked overfitting.
        Running DyStack for up to 150s of the 600s of remaining time (25%).
        Running DyStack sub-fit in a ray process to avoid memory leakage. Enabling ray logging (enable_ray_logging=True). Speci
fy `ds_args={'enable_ray_logging': False}` if you experience logging issues.
2025-06-07 23:46:40,542 INFO worker.py:1843 -- Started a local Ray instance. View the dashboard at http://127.0.0.1:8265
                Context path: "/content/AutogluonModels/ag-20250607_234635/ds_sub_fit/sub_fit_ho"
(_dystack pid=3055) Running DyStack sub-fit ...
(_dystack pid=3055) /usr/local/lib/python3.11/dist-packages/autogluon/common/utils/utils.py:97: UserWarning: pkg_resources is d
eprecated as an API. See https://setuptools.pypa.io/en/latest/pkg_resources.html. The pkg_resources package is slated for remov
al as early as 2025-11-30. Refrain from using this package or pin to Setuptools<81.
(_dystack pid=3055)   import pkg_resources
(_dystack pid=3055) Beginning AutoGluon training ... Time limit = 143s
(_dystack pid=3055) AutoGluon will save models to "/content/AutogluonModels/ag-20250607_234635/ds_sub_fit/sub_fit_ho"
(_dystack pid=3055) Train Data Rows:    9676
(_dystack pid=3055) Train Data Columns: 9
(_dystack pid=3055) Label Column:       count
(_dystack pid=3055) Problem Type:       regression
(_dystack pid=3055) Preprocessing data ...
(_dystack pid=3055) Using Feature Generators to preprocess the data ...
(_dystack pid=3055) Fitting AutoMLPipelineFeatureGenerator...
(_dystack pid=3055)     Available Memory:                    11211.20 MB
```

```
(_dystack pid=3055)     Train Data (Original)  Memory Usage: 0.66 MB (0.0% of available memory)
(_dystack pid=3055)     Inferring data type of each feature based on column values. Set feature_metadata_in to manually specify
special dtypes of the features.
(_dystack pid=3055)     Stage 1 Generators:
(_dystack pid=3055)             Fitting AsTypeFeatureGenerator...
(_dystack pid=3055)                     Note: Converting 2 features to boolean dtype as they only contain 2 unique values.
(_dystack pid=3055)     Stage 2 Generators:
(_dystack pid=3055)             Fitting FillNaFeatureGenerator...
(_dystack pid=3055)     Stage 3 Generators:
(_dystack pid=3055)             Fitting IdentityFeatureGenerator...
(_dystack pid=3055)             Fitting DatetimeFeatureGenerator...
(_dystack pid=3055)     Stage 4 Generators:
(_dystack pid=3055)             Fitting DropUniqueFeatureGenerator...
(_dystack pid=3055)     Stage 5 Generators:
(_dystack pid=3055)             Fitting DropDuplicatesFeatureGenerator...
(_dystack pid=3055)     Types of features in original data (raw dtype, special dtypes):
(_dystack pid=3055)             ('datetime', []) : 1 | ['datetime']
(_dystack pid=3055)             ('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
(_dystack pid=3055)             ('int', [])      : 5 | ['season', 'holiday', 'workingday', 'weather', 'humidity']
(_dystack pid=3055)     Types of features in processed data (raw dtype, special dtypes):
(_dystack pid=3055)             ('float', [])                : 3 | ['temp', 'atemp', 'windspeed']
(_dystack pid=3055)             ('int', [])                  : 3 | ['season', 'weather', 'humidity']
(_dystack pid=3055)             ('int', ['bool'])            : 2 | ['holiday', 'workingday']
(_dystack pid=3055)             ('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.da
y', 'datetime.dayofweek']
(_dystack pid=3055)     0.4s = Fit runtime
(_dystack pid=3055)     9 features in original data used to generate 13 features in processed data.
(_dystack pid=3055)     Train Data (Processed) Memory Usage: 0.83 MB (0.0% of available memory)
(_dystack pid=3055) Data preprocessing and feature engineering runtime = 0.41s ...
(_dystack pid=3055) AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squared_error'
(_dystack pid=3055)     This metric's sign has been flipped to adhere to being higher_is_better. The metric score can be multip
lied by -1 to get the metric value.
(_dystack pid=3055)     To change this, specify the eval_metric parameter of Predictor()
(_dystack pid=3055) Large model count detected (112 configs) ... Only displaying the first 3 models of each family. To see all,
set `verbosity=3`.
(_dystack pid=3055) User-specified model hyperparameters to be fit:
(_dystack pid=3055) {
(_dystack pid=3055)     'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.10077639529843717, 'hidden_size': 108, 'learni
ng_rate': 0.002735937344002146, 'num_layers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'nam
e_suffix': '_r79', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_r
ate': 0.0010474382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'name_
```

```
suffix': '_r22', 'priority': -7}}],
(_dystack pid=3055)     'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, {'learning_rate': 0.03, 'num_leave
s': 128, 'feature_fraction': 0.9, 'min_data_in_leaf': 3, 'ag_args': {'name_suffix': 'Large', 'priority': 0, 'hyperparameter_tun
e_kwargs': None}}],
(_dystack pid=3055)     'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_r
ate': 0.06864209415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -
1}}, {'depth': 8, 'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ct
r_complexity': 2, 'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
(_dystack pid=3055)     'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.01
8063876087523967, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}},
{'colsample_bytree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'mi
n_child_weight': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
(_dystack pid=3055)     'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.
01519848858318159, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.
05070411322605811, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_
suffix': '_r102', 'priority': -11}}],
(_dystack pid=3055)     'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclas
s']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion':
'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
(_dystack pid=3055)     'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclas
s']}}, {'criterion': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion':
'squared_error', 'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
(_dystack pid=3055)     'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args':
{'name_suffix': 'Dist'}}],
(_dystack pid=3055) }
(_dystack pid=3055) AutoGluon will fit 2 stack levels (L1 to L2) ...
(_dystack pid=3055) Fitting 108 L1 models, fit_strategy="sequential" ...
(_dystack pid=3055) Fitting model: KNeighborsUnif_BAG_L1 ... Training model for up to 95.15s of the 142.61s of remaining time.
(_dystack pid=3055)     -107.445        = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     0.03s    = Training   runtime
(_dystack pid=3055)     0.06s    = Validation runtime
(_dystack pid=3055) Fitting model: KNeighborsDist_BAG_L1 ... Training model for up to 89.05s of the 136.51s of remaining time.
(_dystack pid=3055)     -89.9469        = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     0.02s    = Training   runtime
(_dystack pid=3055)     0.06s    = Validation runtime
(_dystack pid=3055) Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 88.96s of the 136.41s of remaining time.
(_dystack pid=3055)     Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: c
pus=1, gpus=0, memory=0.10%)
```

```
(_ray_fit pid=3245) [1000]      valid_set's rmse: 129.692
(_ray_fit pid=3433) [1000]      valid_set's rmse: 132.725 [repeated 5x across cluster] (Ray deduplicates logs by default. Set RAY_DEDUP_LOGS=0 to disable log deduplication, or see https://docs.ray.io/en/master/ray-observability/user-guides/configure-logging.html#log-deduplication for more options.)
(_ray_fit pid=3495) [1000]      valid_set's rmse: 128.154
(_ray_fit pid=3495) [2000]      valid_set's rmse: 126.702
(_ray_fit pid=3495) [3000]      valid_set's rmse: 126.147
(_ray_fit pid=3495) [4000]      valid_set's rmse: 125.904
(_ray_fit pid=3553) [2000]      valid_set's rmse: 134.443 [repeated 3x across cluster]
(_ray_fit pid=3553) [5000]      valid_set's rmse: 132.706 [repeated 6x across cluster]
(_ray_fit pid=3495)     Ran out of time, early stopping on iteration 8431. Best iteration is:
(_ray_fit pid=3495)     [7106]  valid_set's rmse: 125.339
(_ray_fit pid=3553) [8000]      valid_set's rmse: 132.379 [repeated 3x across cluster]
(_ray_fit pid=3694) [1000]      valid_set's rmse: 137.712 [repeated 2x across cluster]
(_ray_fit pid=3694) [4000]      valid_set's rmse: 135.344 [repeated 3x across cluster]
(_ray_fit pid=3694) [7000]      valid_set's rmse: 135.236 [repeated 5x across cluster]
(_dystack pid=3055)     -131.9758       = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     87.07s    = Training   runtime
(_dystack pid=3055)     13.79s    = Validation runtime
(_dystack pid=3055) Fitting model: WeightedEnsemble_L2 ... Training model for up to 142.76s of the 42.98s of remaining time.
(_dystack pid=3055)     Ensemble Weights: {'KNeighborsDist_BAG_L1': 1.0}
(_dystack pid=3055)     -89.9469        = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     0.02s     = Training   runtime
(_dystack pid=3055)     0.0s      = Validation runtime
(_dystack pid=3055) Fitting 106 L2 models, fit_strategy="sequential" ...
(_dystack pid=3055) Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 42.95s of the 42.75s of remaining time.
(_dystack pid=3055)     Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, memory=0.12%)
(_ray_fit pid=3892) [1000]      valid_set's rmse: 70.3611 [repeated 5x across cluster]
(_ray_fit pid=4005) [1000]      valid_set's rmse: 77.4878 [repeated 2x across cluster]
(_ray_fit pid=4123) [1000]      valid_set's rmse: 76.4032 [repeated 2x across cluster]
(_ray_fit pid=4242) [1000]      valid_set's rmse: 73.4134 [repeated 2x across cluster]
```

```
(_dystack pid=3055)     -74.3067        = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     48.09s    = Training   runtime
(_dystack pid=3055)     2.43s     = Validation runtime
(_dystack pid=3055) Fitting model: WeightedEnsemble_L3 ... Training model for up to 142.76s of the -11.08s of remaining time.
(_dystack pid=3055)     Ensemble Weights: {'LightGBMXT_BAG_L2': 0.947, 'KNeighborsDist_BAG_L1': 0.053}
(_dystack pid=3055)     -74.2555        = Validation score   (-root_mean_squared_error)
(_dystack pid=3055)     0.01s     = Training   runtime
(_dystack pid=3055)     0.0s      = Validation runtime
(_dystack pid=3055) AutoGluon training complete, total runtime = 154.28s ... Best model: WeightedEnsemble_L3 | Estimated infere
nce throughput: 74.5 rows/s (1210 batch size)
(_dystack pid=3055) TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/content/AutogluonModels/ag-202506
07_234635/ds_sub_fit/sub_fit_ho")
(_dystack pid=3055) Deleting DyStack predictor artifacts (clean_up_fits=True) ...
Leaderboard on holdout data (DyStack):
                  model  score_holdout   score_val             eval_metric  pred_time_test  pred_time_val    fit_time  pred_t
ime_test_marginal  pred_time_val_marginal  fit_time_marginal  stack_level  can_infer  fit_order
0      LightGBMXT_BAG_L2     -72.615295  -74.306707  root_mean_squared_error       16.284343      16.348675  135.202674
1.667283               2.431805          48.088386            2       True          5
1    WeightedEnsemble_L3     -72.747027  -74.255475  root_mean_squared_error       16.286913      16.349375  135.217640
0.002570               0.000700           0.014966            3       True          6
2  KNeighborsDist_BAG_L1     -92.031272  -89.946854  root_mean_squared_error        0.016485       0.060142    0.017822
0.016485               0.060142           0.017822            1       True          2
3    WeightedEnsemble_L2     -92.031272  -89.946854  root_mean_squared_error        0.019435       0.061035    0.034781
0.002950               0.000893           0.016958            2       True          4
4  KNeighborsUnif_BAG_L1    -109.161488 -107.445008  root_mean_squared_error        0.016534       0.064152    0.028414
0.016534               0.064152           0.028414            1       True          1
5      LightGBMXT_BAG_L1    -131.068281 -131.975832  root_mean_squared_error       14.584041      13.792575   87.068052
14.584041              13.792575          87.068052            1       True          3
        1        = Optimal   num_stack_levels (Stacked Overfitting Occurred: False)
        181s     = DyStack   runtime |  419s     = Remaining runtime
Starting main fit with num_stack_levels=1.
        For future fit calls on this dataset, you can skip DyStack to save time: `predictor.fit(..., dynamic_stacking=False, nu
m_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 419s
AutoGluon will save models to "/content/AutogluonModels/ag-20250607_234635"
Train Data Rows:    10886
Train Data Columns: 9
Label Column:       count
Problem Type:       regression
Preprocessing data ...
```

```
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
	Available Memory:                    10868.82 MB
	Train Data (Original)  Memory Usage: 0.75 MB (0.0% of available memory)
	Inferring data type of each feature based on column values. Set feature_metadata_in to manually specify special dtypes
of the features.
	Stage 1 Generators:
		Fitting AsTypeFeatureGenerator...
			Note: Converting 2 features to boolean dtype as they only contain 2 unique values.
	Stage 2 Generators:
		Fitting FillNaFeatureGenerator...
	Stage 3 Generators:
		Fitting IdentityFeatureGenerator...
		Fitting DatetimeFeatureGenerator...
	Stage 4 Generators:
		Fitting DropUniqueFeatureGenerator...
	Stage 5 Generators:
		Fitting DropDuplicatesFeatureGenerator...
	Types of features in original data (raw dtype, special dtypes):
		('datetime', []) : 1 | ['datetime']
		('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
		('int', [])      : 5 | ['season', 'holiday', 'workingday', 'weather', 'humidity']
	Types of features in processed data (raw dtype, special dtypes):
		('float', [])                : 3 | ['temp', 'atemp', 'windspeed']
		('int', [])                  : 3 | ['season', 'weather', 'humidity']
		('int', ['bool'])            : 2 | ['holiday', 'workingday']
		('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.da
yofweek']
	0.1s = Fit runtime
	9 features in original data used to generate 13 features in processed data.
	Train Data (Processed) Memory Usage: 0.93 MB (0.0% of available memory)
Data preprocessing and feature engineering runtime = 0.15s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squared_error'
	This metric's sign has been flipped to adhere to being higher_is_better. The metric score can be multiplied by -1 to ge
t the metric value.
	To change this, specify the eval_metric parameter of Predictor()
Large model count detected (112 configs) ... Only displaying the first 3 models of each family. To see all, set `verbosity=3`.
User-specified model hyperparameters to be fit:
{
	'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.10077639529843717, 'hidden_size': 108, 'learning_rate': 0.0027
35937344002146, 'num_layers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'name_suffix': '_r7
```

```
9', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_rate': 0.0010474
382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'name_suffix': '_r2
2', 'priority': -7}}],
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, {'learning_rate': 0.03, 'num_leaves': 128, 'featur
e_fraction': 0.9, 'min_data_in_leaf': 3, 'ag_args': {'name_suffix': 'Large', 'priority': 0, 'hyperparameter_tune_kwargs': Non
e}}],
        'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_rate': 0.06864209
415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -1}}, {'depth': 8,
'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexity': 2,
'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
        'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.01806387608752396
7, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}}, {'colsample_by
tree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child_weigh
t': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
        'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.0151984885831815
9, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.0507041132260581
1, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix': '_r10
2', 'priority': -11}}],
        'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterio
n': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error',
'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
        'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterio
n': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error',
'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
        'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix':
'Dist'}}],
}
AutoGluon will fit 2 stack levels (L1 to L2) ...
Fitting 108 L1 models, fit_strategy="sequential" ...
Fitting model: KNeighborsUnif_BAG_L1 ... Training model for up to 279.29s of the 419.03s of remaining time.
        -101.5462       = Validation score   (-root_mean_squared_error)
        0.06s    = Training   runtime
        0.06s    = Validation runtime
Fitting model: KNeighborsDist_BAG_L1 ... Training model for up to 276.37s of the 416.10s of remaining time.
        -84.1251        = Validation score   (-root_mean_squared_error)
        0.06s    = Training   runtime
        0.06s    = Validation runtime
Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 276.21s of the 415.95s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.10%)
        -131.4609       = Validation score   (-root_mean_squared_error)
```

```
        83.84s   = Training   runtime
        16.6s    = Validation runtime
Fitting model: LightGBM_BAG_L1 ... Training model for up to 183.75s of the 323.49s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.11%)
        -131.0542        = Validation score   (-root_mean_squared_error)
        42.86s   = Training   runtime
        1.8s     = Validation runtime
Fitting model: RandomForestMSE_BAG_L1 ... Training model for up to 133.39s of the 273.12s of remaining time.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
        -116.5484        = Validation score   (-root_mean_squared_error)
        18.92s   = Training   runtime
        0.64s    = Validation runtime
Fitting model: CatBoost_BAG_L1 ... Training model for up to 112.33s of the 252.07s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=3.62%)
        -131.0151        = Validation score   (-root_mean_squared_error)
        103.78s  = Training   runtime
        0.15s    = Validation runtime
Fitting model: ExtraTreesMSE_BAG_L1 ... Training model for up to 3.84s of the 143.57s of remaining time.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
        -124.6007        = Validation score   (-root_mean_squared_error)
        9.7s     = Training   runtime
        0.63s    = Validation runtime
Fitting model: WeightedEnsemble_L2 ... Training model for up to 360.00s of the 131.42s of remaining time.
        Ensemble Weights: {'KNeighborsDist_BAG_L1': 1.0}
        -84.1251         = Validation score   (-root_mean_squared_error)
        0.02s    = Training   runtime
        0.0s     = Validation runtime
Fitting 106 L2 models, fit_strategy="sequential" ...
Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 131.37s of the 131.35s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.13%)
        -60.4034         = Validation score   (-root_mean_squared_error)
        72.81s   = Training   runtime
```

```
        5.96s    = Validation runtime
Fitting model: LightGBM_BAG_L2 ... Training model for up to 51.07s of the 51.04s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.13%)
        -55.1456         = Validation score   (-root_mean_squared_error)
        41.4s    = Training   runtime
        0.29s    = Validation runtime
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of the 0.73s of remaining time.
        Ensemble Weights: {'LightGBM_BAG_L2': 0.889, 'LightGBMXT_BAG_L2': 0.111}
        -55.0549         = Validation score   (-root_mean_squared_error)
        0.03s    = Training   runtime
        0.0s     = Validation runtime
AutoGluon training complete, total runtime = 418.53s ... Best model: WeightedEnsemble_L3 | Estimated inference throughput: 54.5
rows/s (1361 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/content/AutogluonModels/ag-20250607_234635")
```

Review AutoGluon's training run with ranking of models that did the best.

In [19]:
```
predictor.leaderboard(silent=True)
```

Out[19]:

	model	score_val	eval_metric	pred_time_val	fit_time	pred_time_val_marginal	fit_time_marginal	s
0	WeightedEnsemble_L3	-55.054901	root_mean_squared_error	26.181087	373.466759	0.000809	0.027246	
1	LightGBM_BAG_L2	-55.145568	root_mean_squared_error	20.221712	300.624579	0.286987	41.397880	
2	LightGBMXT_BAG_L2	-60.403446	root_mean_squared_error	25.893291	332.041634	5.958566	72.814934	
3	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.059151	0.060290	0.059151	0.060290	
4	WeightedEnsemble_L2	-84.125061	root_mean_squared_error	0.061119	0.084806	0.001968	0.024516	
5	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.056741	0.055604	0.056741	0.055604	
6	RandomForestMSE_BAG_L1	-116.548359	root_mean_squared_error	0.643512	18.921649	0.643512	18.921649	
7	ExtraTreesMSE_BAG_L1	-124.600676	root_mean_squared_error	0.630389	9.703326	0.630389	9.703326	
8	CatBoost_BAG_L1	-131.015139	root_mean_squared_error	0.146454	103.782309	0.146454	103.782309	
9	LightGBM_BAG_L1	-131.054162	root_mean_squared_error	1.796739	42.864861	1.796739	42.864861	
10	LightGBMXT_BAG_L1	-131.460909	root_mean_squared_error	16.601739	83.838660	16.601739	83.838660	

In [27]:
```
#predictor.fit_summary()
```

Create predictions from test dataset

In [34]:
```
predictions = predictor.predict(test)
predictions.head()
```

Out[34]:

	count
0	36.886292
1	45.200489
2	48.684017
3	51.885418
4	52.739521

dtype: float32

NOTE: Kaggle will reject the submission if we don't set everything to be > 0.

In [35]:
```
# Describe the `predictions` series to see if there are any negative values
predictions.describe()
```

Out[35]:

	count
count	6493.000000
mean	100.195526
std	90.280220
min	-2.597213
25%	16.447807
50%	62.667683
75%	173.290558
max	368.331085

dtype: float64

In [38]:
```
# How many negative values do we have?
print((predictions < 0).sum())
```

```
5
```

In [40]:
```
# Set them to zero
predictions[predictions < 0] = 0
```

In [66]:
```
print((predictions < 0).sum())
```

```
0
```

Set predictions to submission dataframe, save, and submit

In [43]:
```
submission["count"] = predictions
submission.to_csv("submission.csv", index=False)
```

In [44]:
```
!kaggle competitions submit -c bike-sharing-demand -f submission.csv -m "first raw submission"
```

```
100% 188k/188k [00:00<00:00, 338kB/s]
Successfully submitted to Bike Sharing Demand
```

View submission via the command line or in the web browser under the competition's page - `My Submissions`

In [45]:
```
!kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head -n 6
```

```
fileName        date                 description           status                    publicScore  privateScore
--------------  -------------------  --------------------  ------------------------  -----------  ------------
submission.csv  2025-06-08 02:00:06.850000  first raw submission  SubmissionStatus.COMPLETE  1.83939      1.83939
```

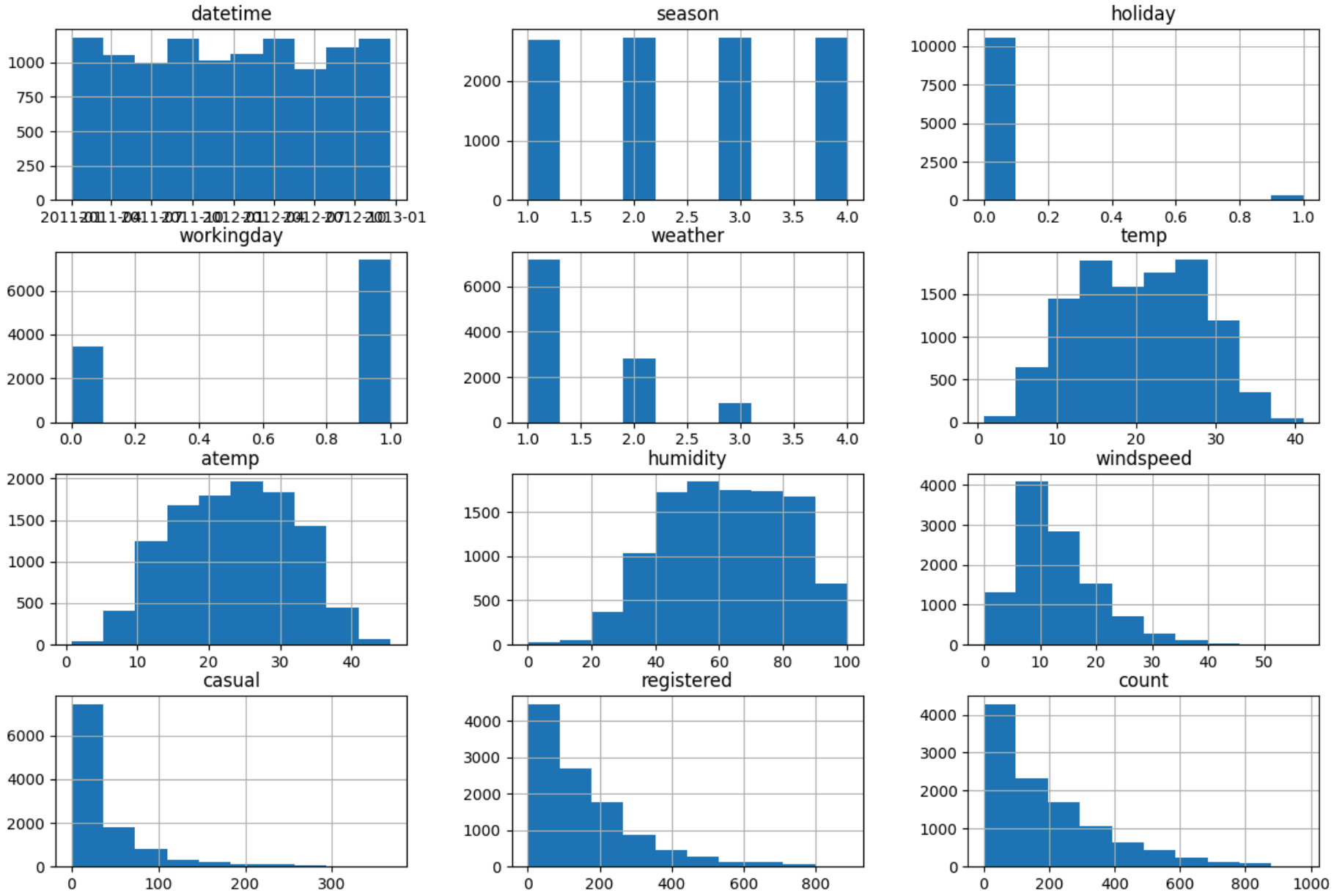
Initial score of `1.83939`

Step 4: Exploratory Data Analysis and Creating an additional feature

- Any additional feature will do, but a great suggestion would be to separate out the datetime into hour, day, or month parts.

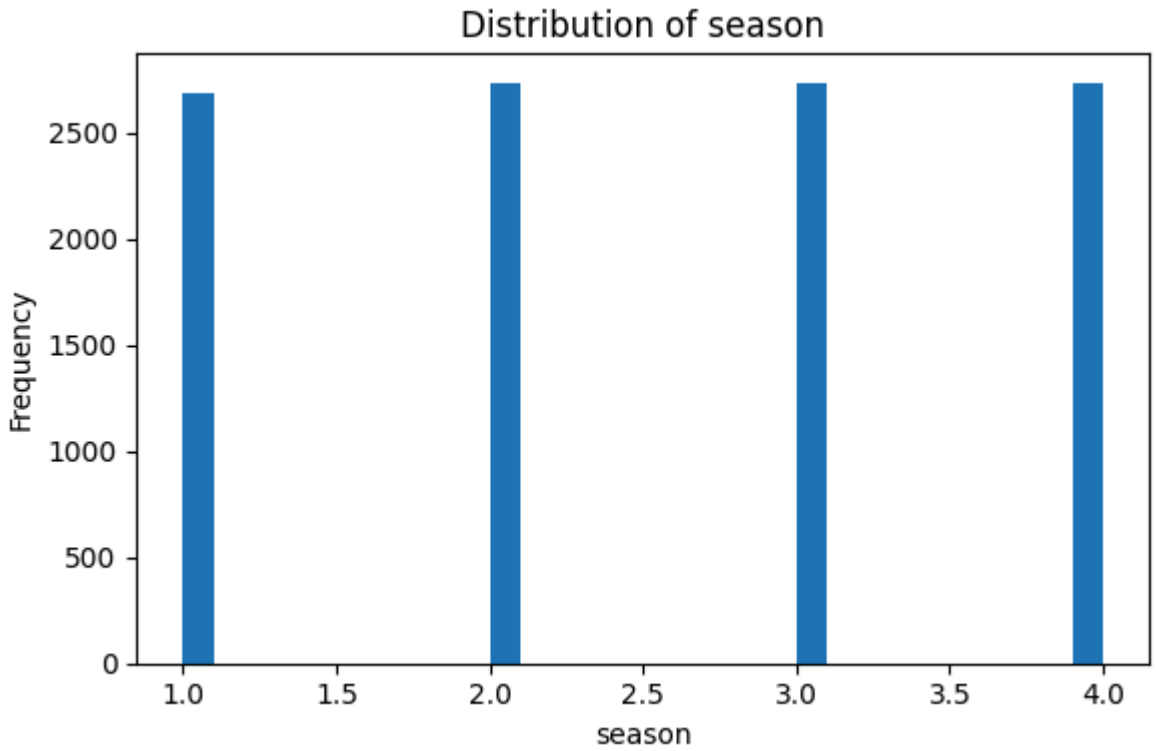
In [46]:
```
# Create a histogram of all features to show the distribution of each one relative to the data. This is part of the exploritor
train.hist(figsize=(15, 10))
```

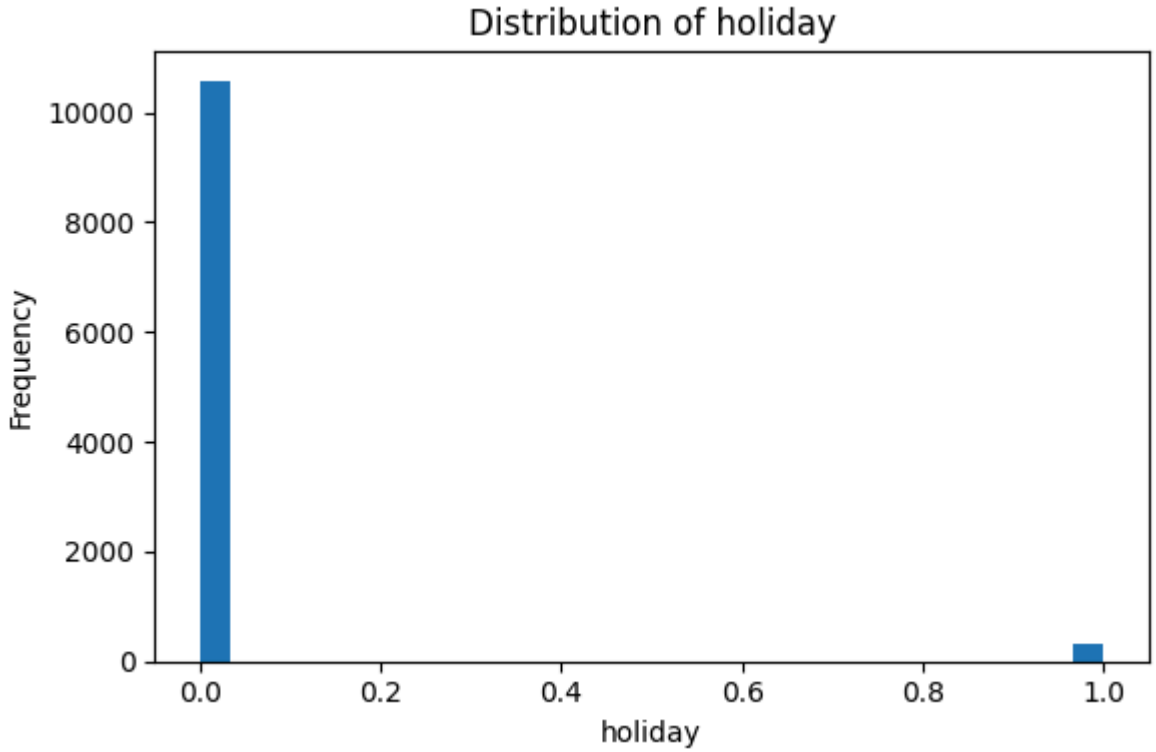
Out[46]:
```
array([[<Axes: title={'center': 'datetime'}>,
        <Axes: title={'center': 'season'}>,
        <Axes: title={'center': 'holiday'}>],
       [<Axes: title={'center': 'workingday'}>,
        <Axes: title={'center': 'weather'}>,
        <Axes: title={'center': 'temp'}>],
       [<Axes: title={'center': 'atemp'}>,
        <Axes: title={'center': 'humidity'}>,
        <Axes: title={'center': 'windspeed'}>],
       [<Axes: title={'center': 'casual'}>,
        <Axes: title={'center': 'registered'}>,
        <Axes: title={'center': 'count'}>]], dtype=object)
```

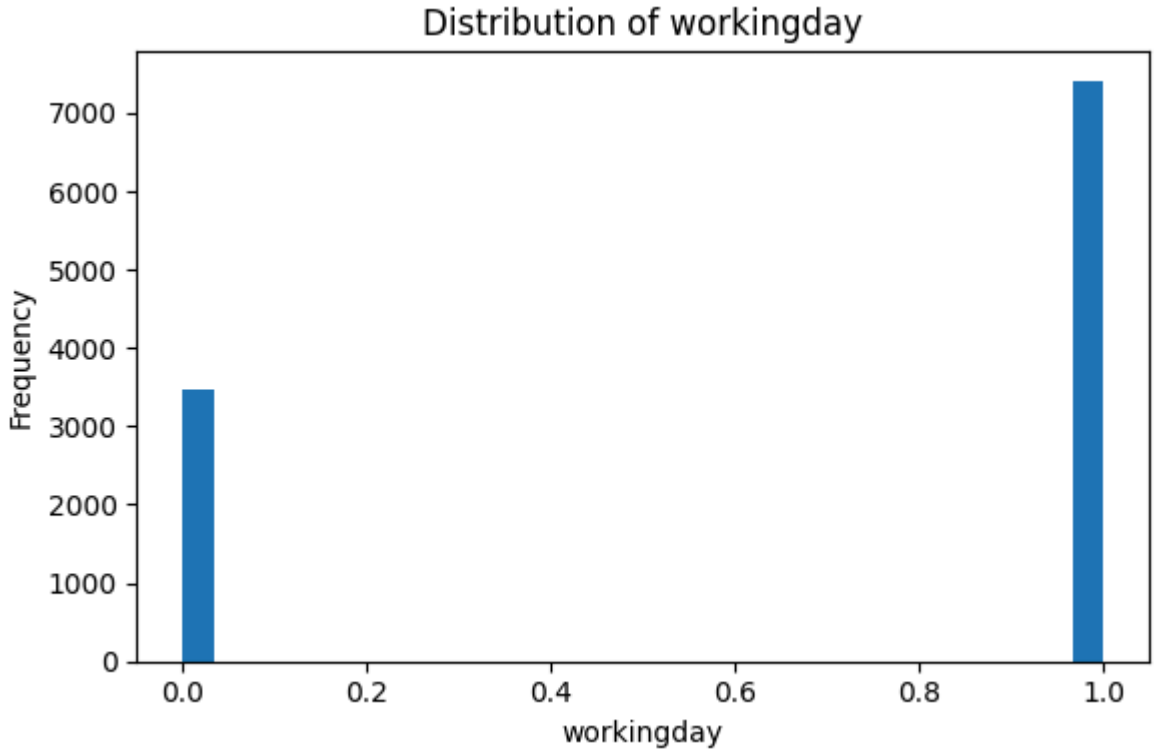


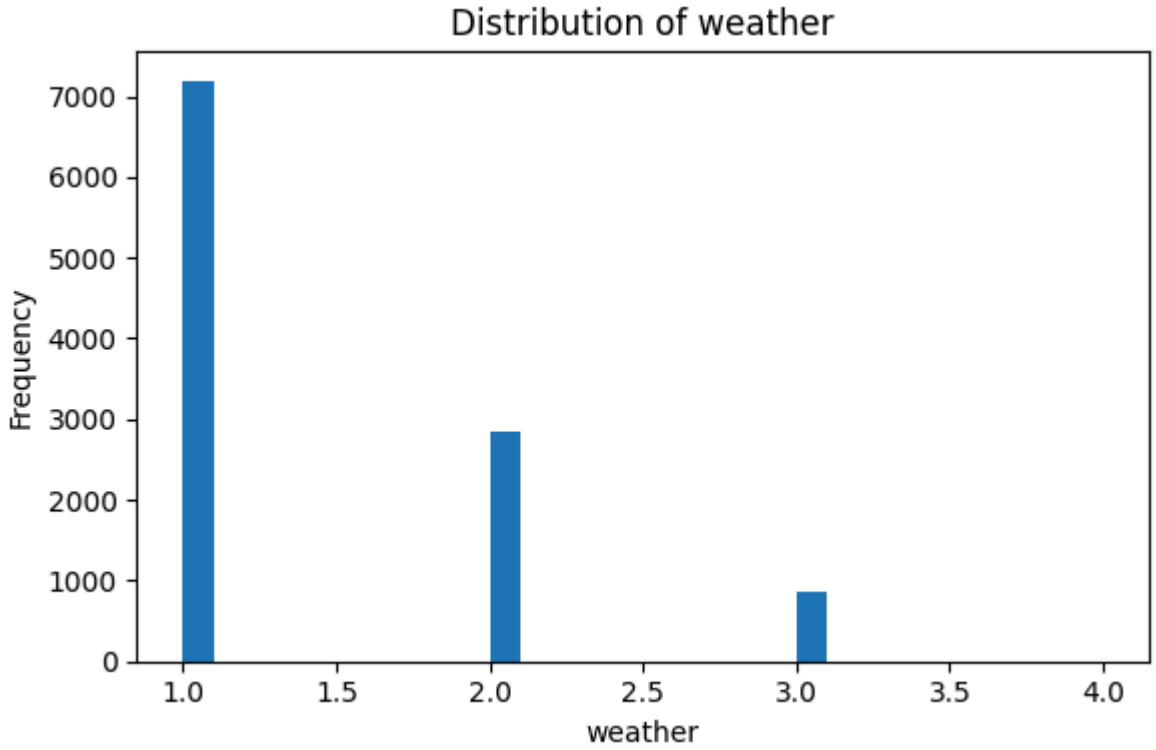
In [48]:
```
import matplotlib.pyplot as plt

numeric_columns = train.select_dtypes(include=["float64", "int64"]).columns
```

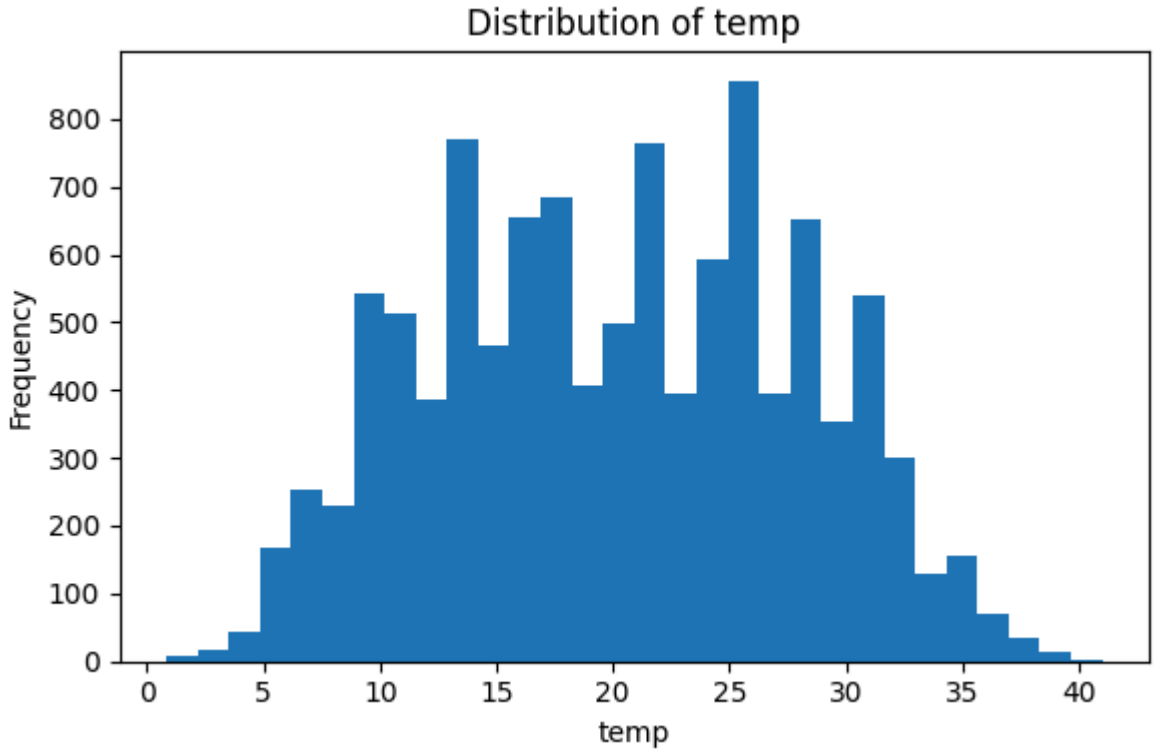
```python
for col in numeric_columns:
    plt.figure(figsize=(6, 4))
    train[col].hist(bins=30)
    plt.title(f'Distribution of {col}')
    plt.xlabel(col)
    plt.ylabel('Frequency')
    plt.grid(False)
    plt.tight_layout()
    plt.show()
```

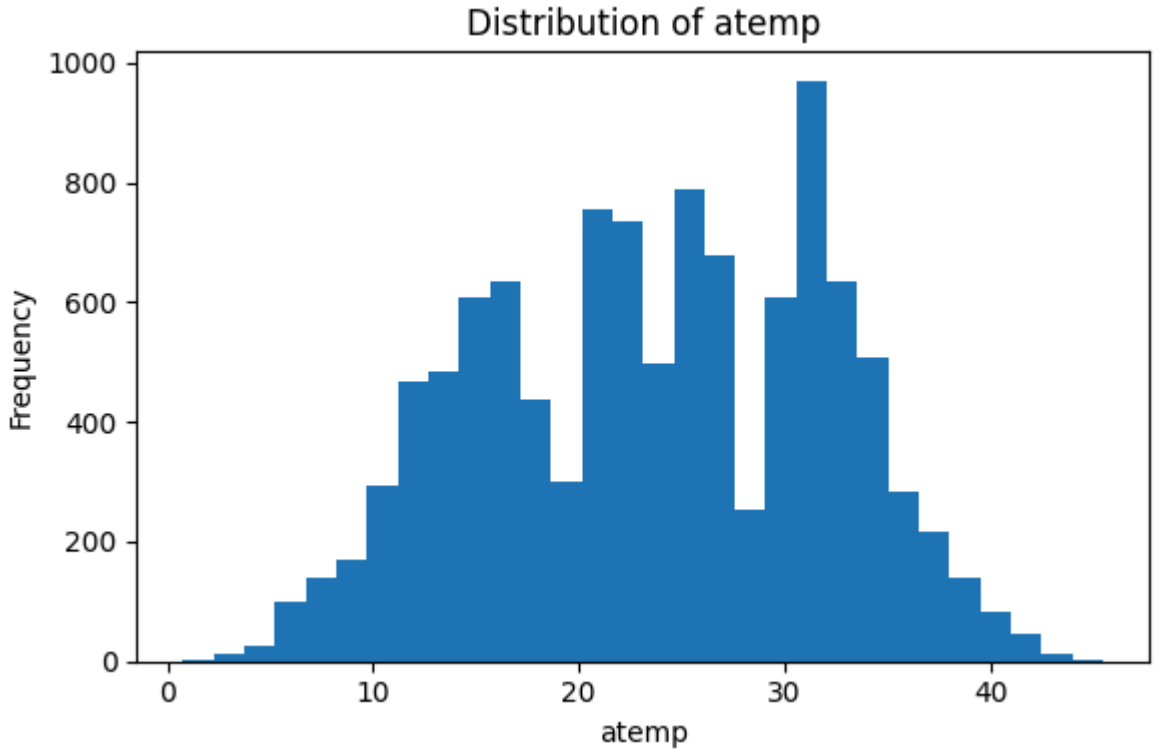


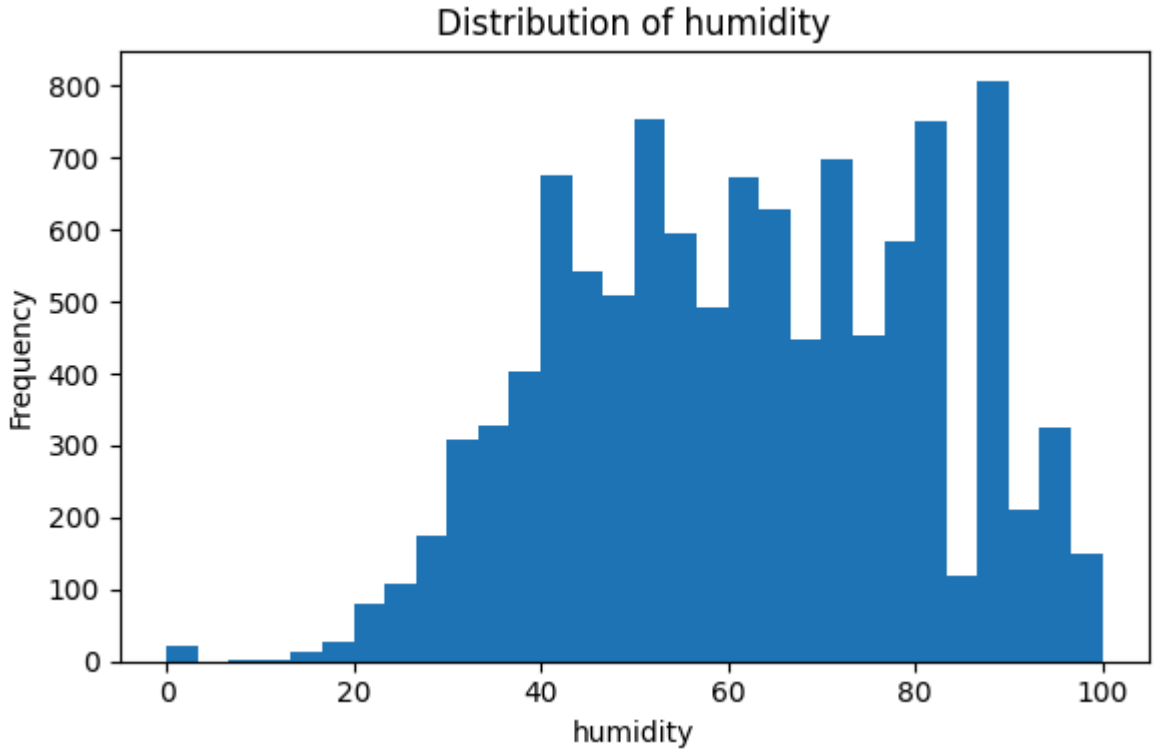

Distribution of holiday
Frequency
10000
8000
6000
4000
2000
0
0.0
0.2
0.4
0.6
0.8
1.0
holiday

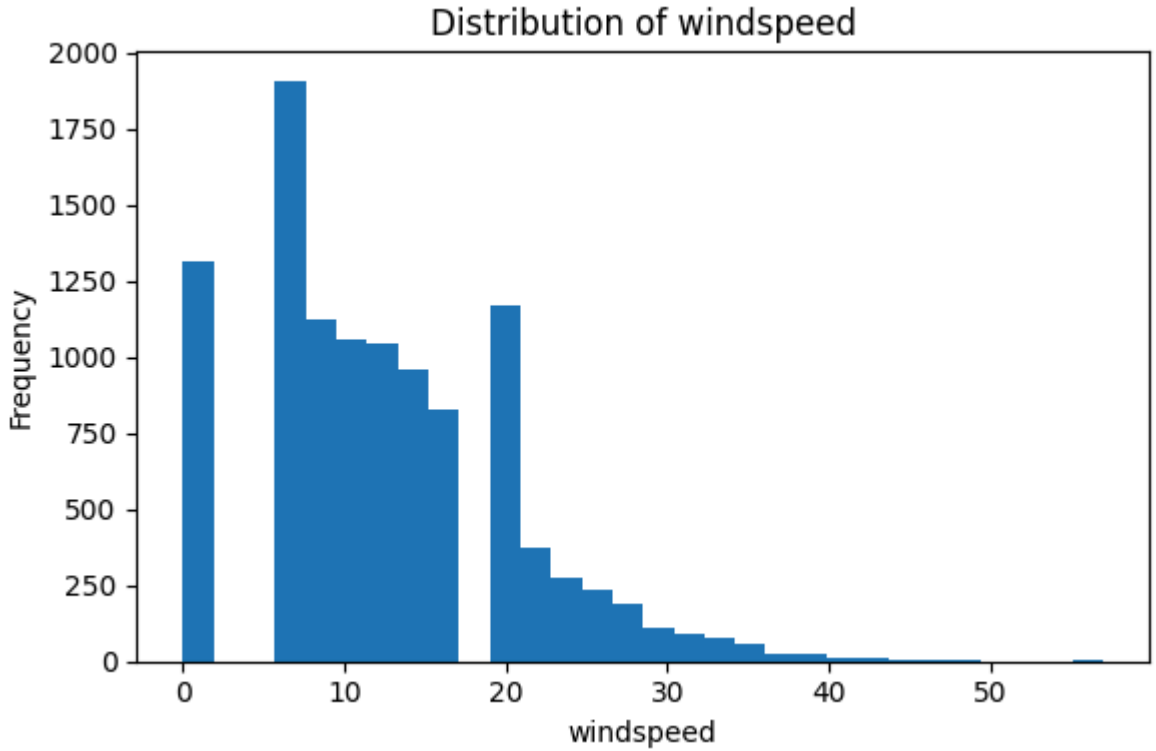

Distribution of workingday
Frequency
7000
6000
5000
4000
3000
2000
1000
0
0.0
0.2
0.4
0.6
0.8
1.0
workingday

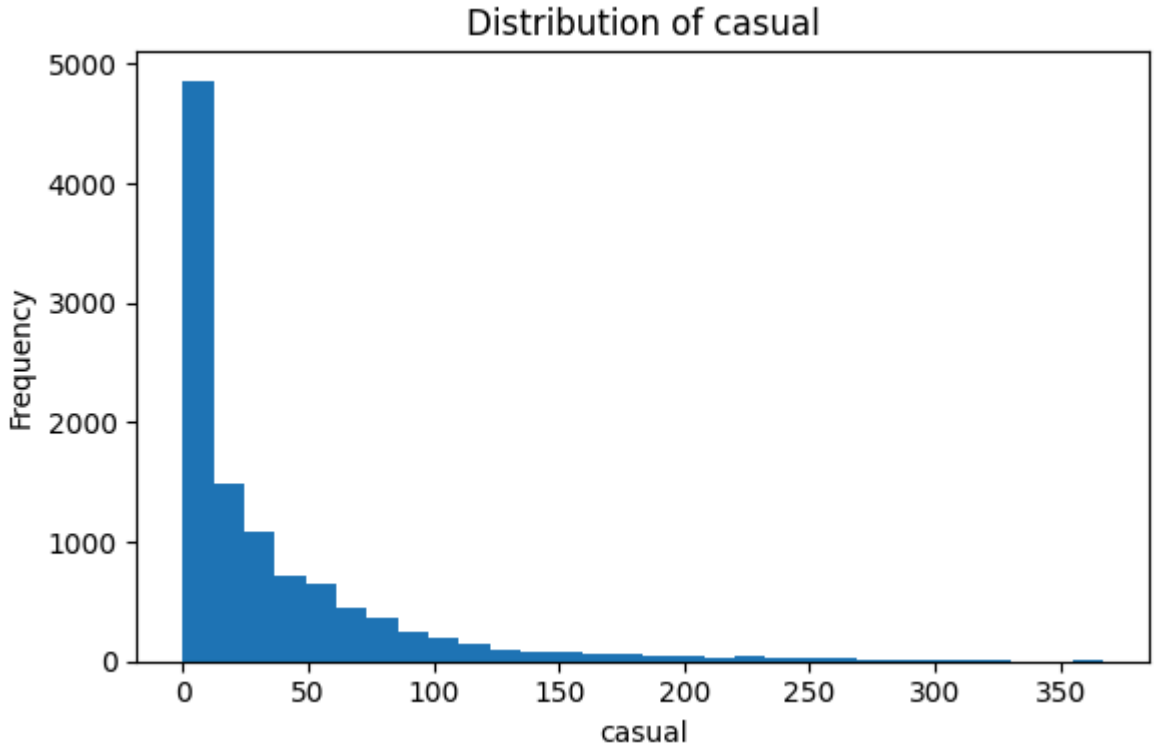

Distribution of weather
Frequency
7000
6000
5000
4000
3000
2000
1000
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
weather

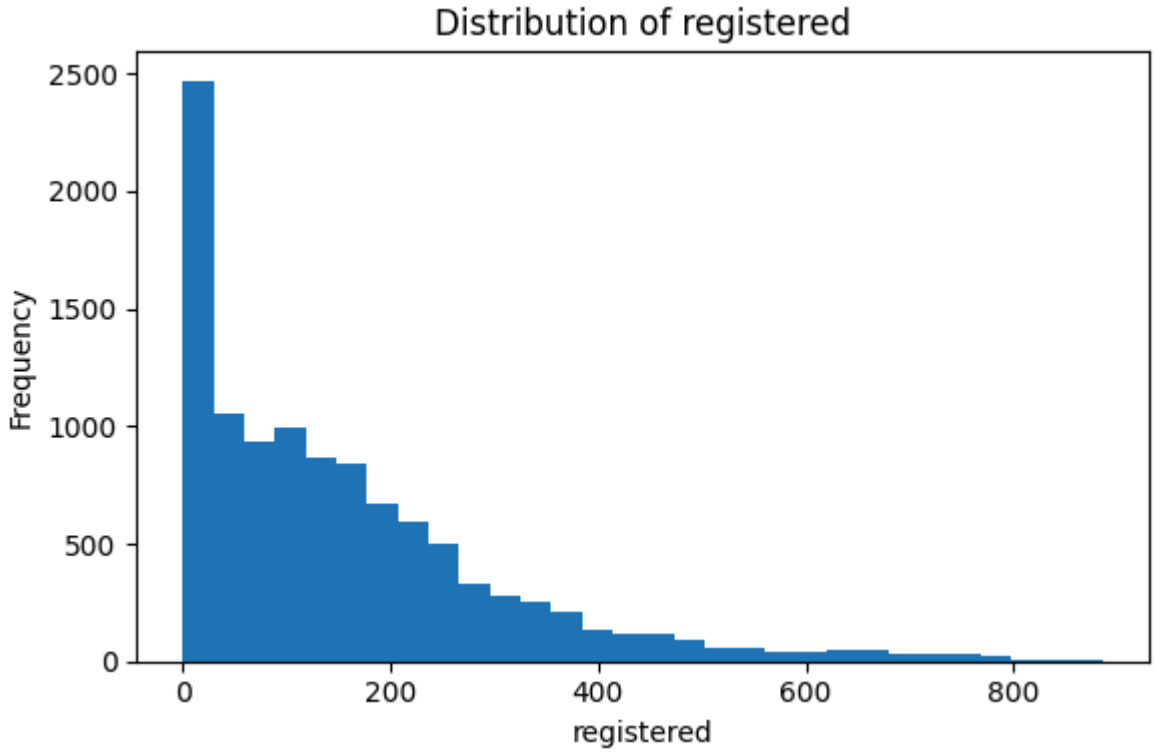

Distribution of temp
Frequency
temp
0
100
200
300
400
500
600
700
800
0
5
10
15
20
25
30
35
40

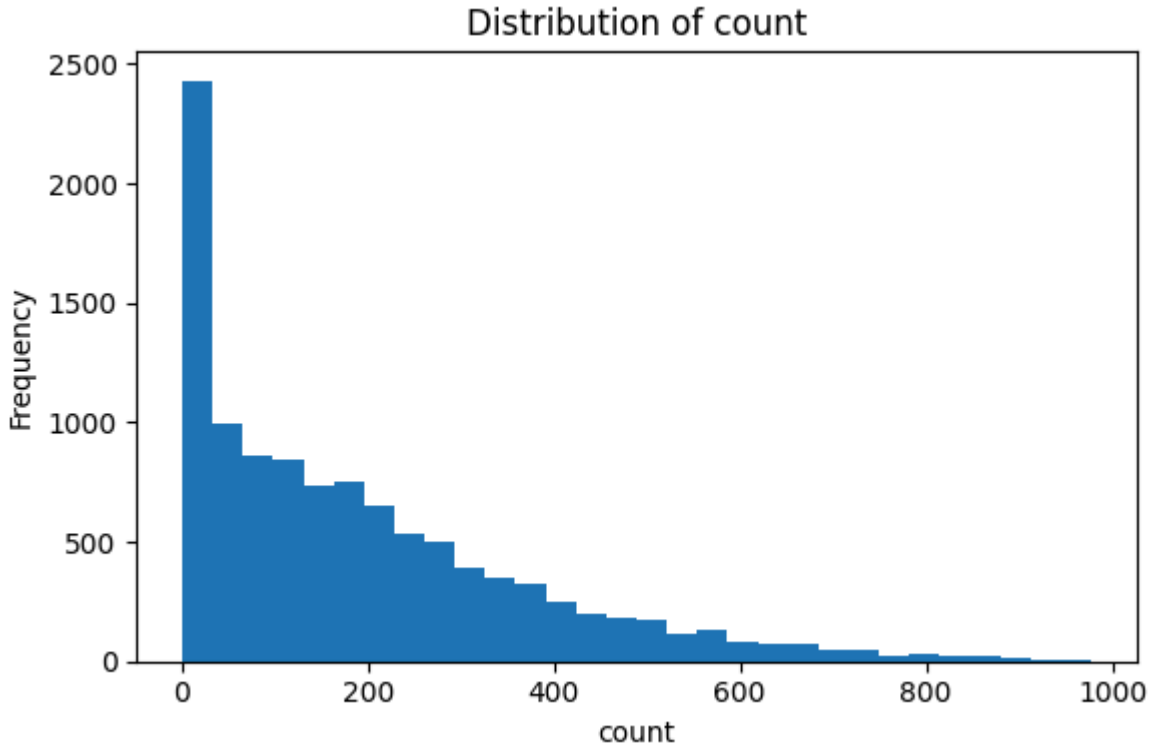

Distribution of atemp
Frequency
1000
800
600
400
200
0
0
10
20
30
40
atemp


Distribution of humidity
Frequency
humidity
0
100
200
300
400
500
600
700
800
0
20
40
60
80
100


Distribution of windspeed
Frequency
2000
1750
1500
1250
1000
750
500
250
0
0
10
20
30
40
50
windspeed


Distribution of casual
Frequency
casual
0
1000
2000
3000
4000
5000
0
50
100
150
200
250
300
350


Distribution of registered
Frequency
2500
2000
1500
1000
500
0
0
200
400
600
800
registered



In [49]:
```
# create a new feature
train["hour"] = train["datetime"].dt.hour
test["hour"] = test["datetime"].dt.hour
```

Make category types for these so models know they are not just numbers

- AutoGluon originally sees these as ints, but in reality they are int representations of a category.
- Setting the dtype to category will classify these as categories in AutoGluon.

In [50]:
```
train["season"] = train["season"].astype("category")
train["weather"] = train["weather"].astype("category")
test["season"] = test["season"].astype("category")
test["weather"] = test["weather"].astype("category")
```

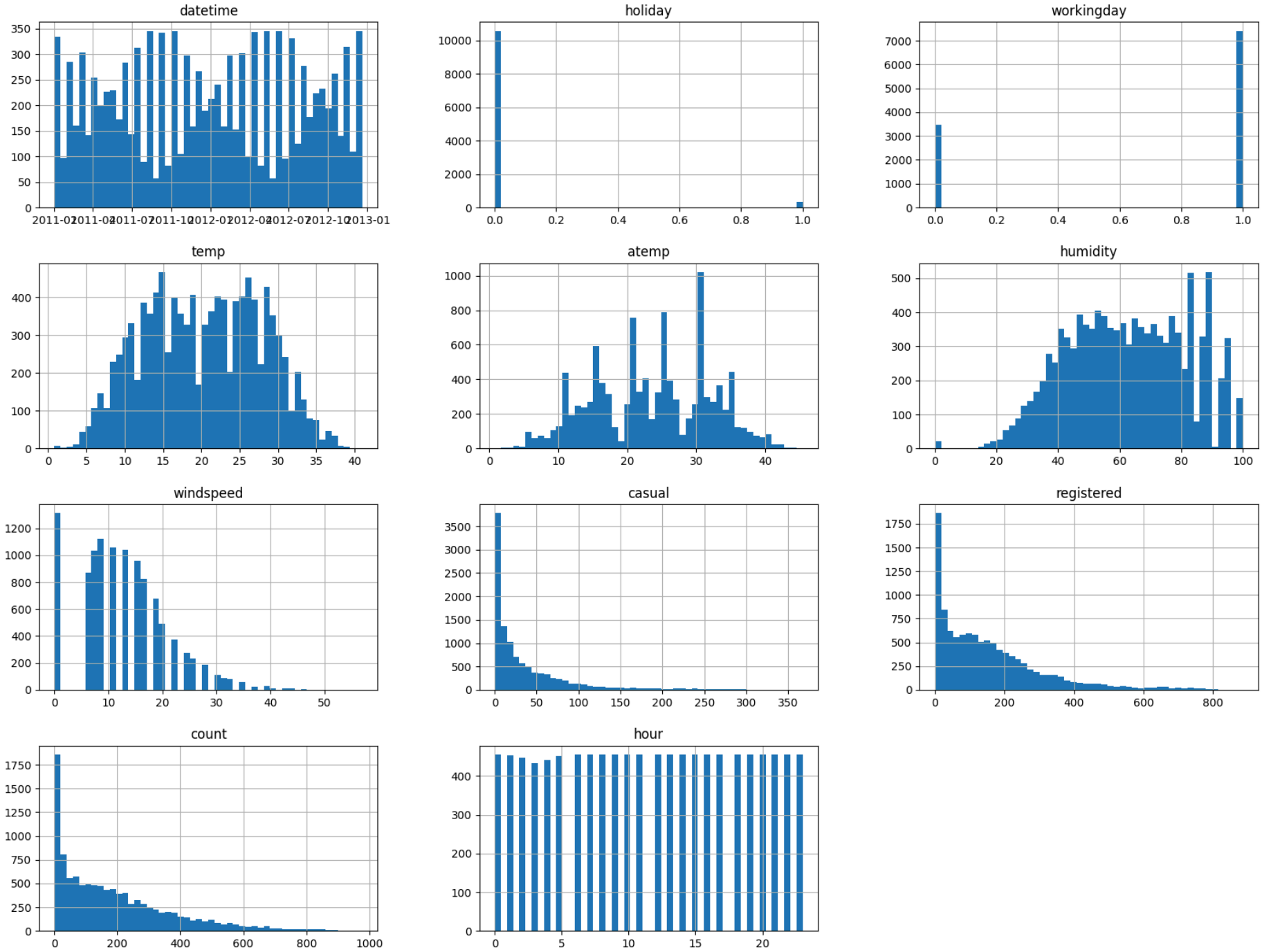
In [54]:
```
# View are new feature
train.head()
```

Out[54]:

	datetime	season	holiday	workingday	weather	temp	atemp	humidity	windspeed	casual	registered	count	hour
0	2011-01-01 00:00:00	1	0	0	1	9.84	14.395	81	0.0	3	13	16	0
1	2011-01-01 01:00:00	1	0	0	1	9.02	13.635	80	0.0	8	32	40	1
2	2011-01-01 02:00:00	1	0	0	1	9.02	13.635	80	0.0	5	27	32	2
3	2011-01-01 03:00:00	1	0	0	1	9.84	14.395	75	0.0	3	10	13	3
4	2011-01-01 04:00:00	1	0	0	1	9.84	14.395	75	0.0	0	1	1	4

In [58]:
```
# View histogram of all features again now with the hour feature
train.hist(bins=50, figsize=(20,15))
```

Out[58]:
```
array([[<Axes: title={'center': 'datetime'}>,
        <Axes: title={'center': 'holiday'}>,
        <Axes: title={'center': 'workingday'}>],
       [<Axes: title={'center': 'temp'}>,
        <Axes: title={'center': 'atemp'}>,
        <Axes: title={'center': 'humidity'}>],
       [<Axes: title={'center': 'windspeed'}>,
        <Axes: title={'center': 'casual'}>,
        <Axes: title={'center': 'registered'}>],
       [<Axes: title={'center': 'count'}>,
        <Axes: title={'center': 'hour'}>, <Axes: >]], dtype=object)
```


datetime
holiday
workingday
temp
atemp
humidity
windspeed
casual
registered
count
hour

Step 5: Rerun the model with the same settings as before, just with more features

Requirements:

- We are prediting `count`, so it is the label we are setting.
- Ignore `casual` and `registered` columns as they are also not present in the test dataset.
- Use the `root_mean_squared_error` as the metric to use for evaluation.
- Set a time limit of 10 minutes (600 seconds).
- Use the preset `best_quality` to focus on creating the best model.

In [62]:
```
train = train.drop(columns=["casual", "registered"])

predictor_new_features = TabularPredictor(label="count", eval_metric="root_mean_squared_error").fit(train, time_limit=600, pre
```

```
No path specified. Models will be saved in: "AutogluonModels/ag-20250608_031305"
Verbosity: 2 (Standard Logging)
=================== System Info ===================
AutoGluon Version:  1.3.1
Python Version:     3.11.13
Operating System:   Linux
Platform Machine:   x86_64
Platform Version:   #1 SMP PREEMPT_DYNAMIC Sun Mar 30 16:01:29 UTC 2025
CPU Count:          2
Memory Avail:       9.54 GB / 12.67 GB (75.3%)
Disk Space Avail:   65.78 GB / 107.72 GB (61.1%)
===================================================
Presets specified: ['best_quality']
Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when use_bag_holdout is disabled. (use_bag_holdou
t=False)
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_sets=1
DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether the input data is affected by stacked overf
itting and enable or disable stacking as a consequence.
        This is used to identify the optimal `num_stack_levels` value. Copies of AutoGluon will be fit on subsets of the data.
Then holdout validation data is used to detect stacked overfitting.
        Running DyStack for up to 150s of the 600s of remaining time (25%).
                Context path: "/content/AutogluonModels/ag-20250608_031305/ds_sub_fit/sub_fit_ho"
Leaderboard on holdout data (DyStack):
                  model  score_holdout   score_val             eval_metric  pred_time_test  pred_time_val    fit_time  pred_t
ime_test_marginal  pred_time_val_marginal  fit_time_marginal  stack_level  can_infer  fit_order
0    WeightedEnsemble_L3     -32.112551  -33.587020  root_mean_squared_error       25.589590      32.915065  169.284056
0.002897               0.001022           0.037429            3       True          6
1      LightGBMXT_BAG_L2     -32.265890  -33.880088  root_mean_squared_error       25.586693      32.914043  169.246627
0.664533               1.466400          46.119894            2       True          5
2    WeightedEnsemble_L2     -34.075066  -35.244747  root_mean_squared_error       24.857267      31.382477  123.119631
0.004816               0.001104           0.018943            2       True          4
3      LightGBMXT_BAG_L1     -34.130379  -35.316584  root_mean_squared_error       24.778463      31.317738  123.081611
24.778463              31.317738         123.081611            1       True          3
4  KNeighborsDist_BAG_L1     -92.031272  -89.946854  root_mean_squared_error        0.073988       0.063635    0.019077
0.073988               0.063635           0.019077            1       True          2
5  KNeighborsUnif_BAG_L1    -109.161488 -107.445008  root_mean_squared_error        0.069709       0.066271    0.026046
0.069709               0.066271           0.026046            1       True          1
        1        = Optimal   num_stack_levels (Stacked Overfitting Occurred: False)
        223s     = DyStack   runtime |  377s     = Remaining runtime
Starting main fit with num_stack_levels=1.
```

```
	For future fit calls on this dataset, you can skip DyStack to save time: `predictor.fit(..., dynamic_stacking=False, num_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 377s
AutoGluon will save models to "/content/AutogluonModels/ag-20250608_031305"
Train Data Rows:    10886
Train Data Columns: 10
Label Column:       count
Problem Type:       regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
	Available Memory:                    9379.50 MB
	Train Data (Original)  Memory Usage: 0.64 MB (0.0% of available memory)
	Inferring data type of each feature based on column values. Set feature_metadata_in to manually specify special dtypes of the features.
	Stage 1 Generators:
		Fitting AsTypeFeatureGenerator...
			Note: Converting 2 features to boolean dtype as they only contain 2 unique values.
	Stage 2 Generators:
		Fitting FillNaFeatureGenerator...
	Stage 3 Generators:
		Fitting IdentityFeatureGenerator...
		Fitting CategoryFeatureGenerator...
			Fitting CategoryMemoryMinimizeFeatureGenerator...
		Fitting DatetimeFeatureGenerator...
	Stage 4 Generators:
		Fitting DropUniqueFeatureGenerator...
	Stage 5 Generators:
		Fitting DropDuplicatesFeatureGenerator...
	Types of features in original data (raw dtype, special dtypes):
		('category', []) : 2 | ['season', 'weather']
		('datetime', []) : 1 | ['datetime']
		('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
		('int', [])      : 4 | ['holiday', 'workingday', 'humidity', 'hour']
	Types of features in processed data (raw dtype, special dtypes):
		('category', [])             : 2 | ['season', 'weather']
		('float', [])                : 3 | ['temp', 'atemp', 'windspeed']
		('int', [])                  : 2 | ['humidity', 'hour']
		('int', ['bool'])            : 2 | ['holiday', 'workingday']
		('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.dayofweek']
```

```
        0.2s = Fit runtime
        10 features in original data used to generate 14 features in processed data.
        Train Data (Processed) Memory Usage: 0.83 MB (0.0% of available memory)
Data preprocessing and feature engineering runtime = 0.28s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squared_error'
        This metric's sign has been flipped to adhere to being higher_is_better. The metric score can be multiplied by -1 to ge
t the metric value.
        To change this, specify the eval_metric parameter of Predictor()
Large model count detected (112 configs) ... Only displaying the first 3 models of each family. To see all, set `verbosity=3`.
User-specified model hyperparameters to be fit:
{
        'NN_TORCH': [{}, {'activation': 'elu', 'dropout_prob': 0.10077639529843717, 'hidden_size': 108, 'learning_rate': 0.0027
35937344002146, 'num_layers': 4, 'use_batchnorm': True, 'weight_decay': 1.356433327634438e-12, 'ag_args': {'name_suffix': '_r7
9', 'priority': -2}}, {'activation': 'elu', 'dropout_prob': 0.11897478034205347, 'hidden_size': 213, 'learning_rate': 0.0010474
382260641949, 'num_layers': 4, 'use_batchnorm': False, 'weight_decay': 5.594471067786272e-10, 'ag_args': {'name_suffix': '_r2
2', 'priority': -7}}],
        'GBM': [{'extra_trees': True, 'ag_args': {'name_suffix': 'XT'}}, {}, {'learning_rate': 0.03, 'num_leaves': 128, 'featur
e_fraction': 0.9, 'min_data_in_leaf': 3, 'ag_args': {'name_suffix': 'Large', 'priority': 0, 'hyperparameter_tune_kwargs': Non
e}}],
        'CAT': [{}, {'depth': 6, 'grow_policy': 'SymmetricTree', 'l2_leaf_reg': 2.1542798306067823, 'learning_rate': 0.06864209
415792857, 'max_ctr_complexity': 4, 'one_hot_max_size': 10, 'ag_args': {'name_suffix': '_r177', 'priority': -1}}, {'depth': 8,
'grow_policy': 'Depthwise', 'l2_leaf_reg': 2.7997999596449104, 'learning_rate': 0.031375015734637225, 'max_ctr_complexity': 2,
'one_hot_max_size': 3, 'ag_args': {'name_suffix': '_r9', 'priority': -5}}],
        'XGB': [{}, {'colsample_bytree': 0.6917311125174739, 'enable_categorical': False, 'learning_rate': 0.01806387608752396
7, 'max_depth': 10, 'min_child_weight': 0.6028633586934382, 'ag_args': {'name_suffix': '_r33', 'priority': -8}}, {'colsample_by
tree': 0.6628423832084077, 'enable_categorical': False, 'learning_rate': 0.08775715546881824, 'max_depth': 5, 'min_child_weigh
t': 0.6294123374222513, 'ag_args': {'name_suffix': '_r89', 'priority': -16}}],
        'FASTAI': [{}, {'bs': 256, 'emb_drop': 0.5411770367537934, 'epochs': 43, 'layers': [800, 400], 'lr': 0.0151984885831815
9, 'ps': 0.23782946566604385, 'ag_args': {'name_suffix': '_r191', 'priority': -4}}, {'bs': 2048, 'emb_drop': 0.0507041132260581
1, 'epochs': 29, 'layers': [200, 100], 'lr': 0.08974235041576624, 'ps': 0.10393466140748028, 'ag_args': {'name_suffix': '_r10
2', 'priority': -11}}],
        'RF': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterio
n': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error',
'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
        'XT': [{'criterion': 'gini', 'ag_args': {'name_suffix': 'Gini', 'problem_types': ['binary', 'multiclass']}}, {'criterio
n': 'entropy', 'ag_args': {'name_suffix': 'Entr', 'problem_types': ['binary', 'multiclass']}}, {'criterion': 'squared_error',
'ag_args': {'name_suffix': 'MSE', 'problem_types': ['regression', 'quantile']}}],
        'KNN': [{'weights': 'uniform', 'ag_args': {'name_suffix': 'Unif'}}, {'weights': 'distance', 'ag_args': {'name_suffix':
'Dist'}}],
}
AutoGluon will fit 2 stack levels (L1 to L2) ...
```

```
Fitting 108 L1 models, fit_strategy="sequential" ...
Fitting model: KNeighborsUnif_BAG_L1 ... Training model for up to 250.98s of the 376.54s of remaining time.
        -101.5462        = Validation score   (-root_mean_squared_error)
        0.05s    = Training   runtime
        0.06s    = Validation runtime
Fitting model: KNeighborsDist_BAG_L1 ... Training model for up to 250.81s of the 376.37s of remaining time.
        -84.1251         = Validation score   (-root_mean_squared_error)
        0.05s    = Training   runtime
        0.07s    = Validation runtime
Fitting model: LightGBMXT_BAG_L1 ... Training model for up to 250.64s of the 376.20s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.11%)
        -34.4573         = Validation score   (-root_mean_squared_error)
        134.24s  = Training   runtime
        42.99s   = Validation runtime
Fitting model: LightGBM_BAG_L1 ... Training model for up to 105.62s of the 231.18s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.11%)
        -33.9196         = Validation score   (-root_mean_squared_error)
        65.87s   = Training   runtime
        5.8s     = Validation runtime
Fitting model: RandomForestMSE_BAG_L1 ... Training model for up to 33.45s of the 159.01s of remaining time.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
        -38.4543         = Validation score   (-root_mean_squared_error)
        24.01s   = Training   runtime
        0.82s    = Validation runtime
Fitting model: CatBoost_BAG_L1 ... Training model for up to 7.22s of the 132.78s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=4.06%)
        Time limit exceeded... Skipping CatBoost_BAG_L1.
Fitting model: WeightedEnsemble_L2 ... Training model for up to 360.00s of the 122.50s of remaining time.
        Ensemble Weights: {'LightGBMXT_BAG_L1': 0.4, 'LightGBM_BAG_L1': 0.4, 'RandomForestMSE_BAG_L1': 0.16, 'KNeighborsDist_BA
G_L1': 0.04}
        -32.3704         = Validation score   (-root_mean_squared_error)
        0.03s    = Training   runtime
        0.0s     = Validation runtime
Fitting 106 L2 models, fit_strategy="sequential" ...
Fitting model: LightGBMXT_BAG_L2 ... Training model for up to 122.41s of the 122.37s of remaining time.
```

```
	Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.14%)
2025-06-08 03:21:07,269 ERROR worker.py:420 -- Unhandled error (suppress with 'RAY_IGNORE_UNHANDLED_ERRORS=1'): The worker died
unexpectedly while executing this task. Check python-core-worker-*.log files for more information.
	-31.2181 	= Validation score   (-root_mean_squared_error)
	56.56s 	= Training   runtime
	2.02s 	= Validation runtime
Fitting model: LightGBM_BAG_L2 ... Training model for up to 62.13s of the 62.09s of remaining time.
	Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.13%)
	-30.6679 	= Validation score   (-root_mean_squared_error)
	45.63s 	= Training   runtime
	0.63s 	= Validation runtime
Fitting model: RandomForestMSE_BAG_L2 ... Training model for up to 11.13s of the 11.09s of remaining time.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
	-31.9975 	= Validation score   (-root_mean_squared_error)
	44.91s 	= Training   runtime
	0.84s 	= Validation runtime
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of the -35.80s of remaining time.
	Ensemble Weights: {'LightGBM_BAG_L2': 0.583, 'LightGBMXT_BAG_L2': 0.292, 'RandomForestMSE_BAG_L2': 0.083, 'LightGBMXT_B
AG_L1': 0.042}
	-30.4288 	= Validation score   (-root_mean_squared_error)
	0.03s 	= Training   runtime
	0.0s 	= Validation runtime
AutoGluon training complete, total runtime = 412.74s ... Best model: WeightedEnsemble_L3 | Estimated inference throughput: 26.3
rows/s (1361 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/content/AutogluonModels/ag-20250608_031305")
```

In [68]:
```
#predictor_new_features.fit_summary()
predictor_new_features.leaderboard(silent=True)
```

Out[68]:

	model	score_val	eval_metric	pred_time_val	fit_time	pred_time_val_marginal	fit_time_marginal	st
0	WeightedEnsemble_L3	-30.428847	root_mean_squared_error	53.232300	371.341935	0.000843	0.027917	
1	LightGBM_BAG_L2	-30.667943	root_mean_squared_error	50.375718	269.847229	0.634663	45.632734	
2	LightGBMXT_BAG_L2	-31.218084	root_mean_squared_error	51.756523	280.771901	2.015468	56.557406	
3	RandomForestMSE_BAG_L2	-31.997543	root_mean_squared_error	50.581326	269.123878	0.840271	44.909383	
4	WeightedEnsemble_L2	-32.370366	root_mean_squared_error	49.682774	224.191972	0.001056	0.026805	
5	LightGBM_BAG_L1	-33.919639	root_mean_squared_error	5.804370	65.868017	5.804370	65.868017	
6	LightGBMXT_BAG_L1	-34.457274	root_mean_squared_error	42.992999	134.239668	42.992999	134.239668	
7	RandomForestMSE_BAG_L1	-38.454338	root_mean_squared_error	0.815811	24.007711	0.815811	24.007711	
8	KNeighborsDist_BAG_L1	-84.125061	root_mean_squared_error	0.068537	0.049772	0.068537	0.049772	
9	KNeighborsUnif_BAG_L1	-101.546199	root_mean_squared_error	0.059337	0.049328	0.059337	0.049328	

In [72]:
```
predictions_new = predictor_new_features.predict(test)
```

In [73]:
```
# Remember to set all negative values to zero
print((predictions_new < 0).sum())
predictions_new[predictions_new < 0] = 0
```

```
0
```

In [76]:
```
# Same submitting predictions
submission_new_features = pd.read_csv("sampleSubmission.csv")
submission_new_features["count"] = predictions_new
submission_new_features.to_csv("submission_new_features.csv", index=False)
```

In [77]:
```
submission_new_features.head()
```

Out[77]:

	datetime	count
0	2011-01-20 00:00:00	16.903080
1	2011-01-20 01:00:00	11.508935
2	2011-01-20 02:00:00	10.523116
3	2011-01-20 03:00:00	8.841933
4	2011-01-20 04:00:00	7.671376

In [78]:
```
!kaggle competitions submit -c bike-sharing-demand -f submission_new_features.csv -m "new features"
```

```
100% 188k/188k [00:00<00:00, 326kB/s]
Successfully submitted to Bike Sharing Demand
```

In [79]:
```
!kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head -n 6
```

```
fileName                    date                        description          status                    publicScore  privateS
core
--------------------------  --------------------------  -------------------  ------------------------  -----------  --------
----
submission_new_features.csv  2025-06-08 04:15:54.810000  new features         SubmissionStatus.COMPLETE  0.62523      0.62523
submission.csv              2025-06-08 02:00:06.850000  first raw submission  SubmissionStatus.COMPLETE  1.83939      1.83939
```

New Score of `0.62523`

Step 6: Hyper parameter optimization

- There are many options for hyper parameter optimization.
- Options are to change the AutoGluon higher level parameters or the individual model hyperparameters.
- The hyperparameters of the models themselves that are in AutoGluon. Those need the `hyperparameter` and `hyperparameter_tune_kwargs` arguments.

In [100…
```
train = pd.read_csv("train.csv", parse_dates=["datetime"])
test = pd.read_csv("test.csv", parse_dates=["datetime"])
```

```
submission_new_hpo = pd.read_csv("sampleSubmission.csv")
```

In [81]:
```
train["hour"] = train["datetime"].dt.hour
test["hour"] = test["datetime"].dt.hour
for col in ["season", "weather"]:
    train[col] = train[col].astype("category")
    test[col] = test[col].astype("category")
```

In [82]:
```
train = train.drop(columns=["casual", "registered"])
```

In [88]:
```
hyperparameters = {
    "GBM": {
        "num_boost_round": 350,
        "learning_rate": 0.04,
        "extra_trees": True,
        "ag_args": {"name_suffix": "_tuned_gbm"}
    },
    "RF": {
        "n_estimators": 180,
        "max_depth": 20,
        "ag_args": {"name_suffix": "_tuned_rf"}
    },
    "XT": {
        "n_estimators": 220,
        "max_depth": 18,
        "ag_args": {"name_suffix": "_tuned_xt"}
    },
    "NN_TORCH": {
        "num_epochs": 25,
        "learning_rate": 0.003,
        "ag_args": {"name_suffix": "_tuned_nn"}
    }
}
```

In [89]:
```
predictor_new_hpo = TabularPredictor(
    label="count",
    eval_metric="root_mean_squared_error"
).fit(
    train,
```

```
    time_limit=600,
    presets="best_quality",
    hyperparameters=hyperparameters,
    hyperparameter_tune_kwargs="auto"
)
```

```
No path specified. Models will be saved in: "AutogluonModels/ag-20250608_054338"
Verbosity: 2 (Standard Logging)
=================== System Info ===================
AutoGluon Version:  1.3.1
Python Version:     3.11.13
Operating System:   Linux
Platform Machine:   x86_64
Platform Version:   #1 SMP PREEMPT_DYNAMIC Sun Mar 30 16:01:29 UTC 2025
CPU Count:          2
Memory Avail:       9.03 GB / 12.67 GB (71.2%)
Disk Space Avail:   64.08 GB / 107.72 GB (59.5%)
===================================================
Presets specified: ['best_quality']
Warning: hyperparameter tuning is currently experimental and may cause the process to hang.
Setting dynamic_stacking from 'auto' to True. Reason: Enable dynamic_stacking when use_bag_holdout is disabled. (use_bag_holdou
t=False)
Stack configuration (auto_stack=True): num_stack_levels=1, num_bag_folds=8, num_bag_sets=1
DyStack is enabled (dynamic_stacking=True). AutoGluon will try to determine whether the input data is affected by stacked overf
itting and enable or disable stacking as a consequence.
        This is used to identify the optimal `num_stack_levels` value. Copies of AutoGluon will be fit on subsets of the data.
Then holdout validation data is used to detect stacked overfitting.
        Running DyStack for up to 150s of the 600s of remaining time (25%).
                Context path: "/content/AutogluonModels/ag-20250608_054338/ds_sub_fit/sub_fit_ho"
Leaderboard on holdout data (DyStack):
                         model  score_holdout  score_val             eval_metric  pred_time_test  pred_time_val    fit_time
pred_time_test_marginal  pred_time_val_marginal  fit_time_marginal  stack_level  can_infer  fit_order
0  LightGBM_tuned_gbm_BAG_L2/T1     -34.938765 -36.192926  root_mean_squared_error        9.600793       2.600578  115.245389
0.416253                 0.681746           48.571722            2       True          5
1           WeightedEnsemble_L3     -35.066216 -35.970970  root_mean_squared_error       10.210495       3.548550  146.022337
0.003767                 0.008637            0.236120            3       True          8
2    ExtraTrees_tuned_xt_BAG_L2     -36.237583 -37.003498  root_mean_squared_error        9.543380       2.433247   76.750088
0.358840                 0.514415           10.076421            2       True          7
3  RandomForest_tuned_rf_BAG_L2     -36.391767 -37.653101  root_mean_squared_error        9.431635       2.343752   87.138074
0.247095                 0.424920           20.464407            2       True          6
4           WeightedEnsemble_L2     -38.213607 -37.909277  root_mean_squared_error        9.188482       1.923774   66.805919
0.003942                 0.004942            0.132252            2       True          4
5    ExtraTrees_tuned_xt_BAG_L1     -38.445539 -39.679490  root_mean_squared_error        0.394544       0.536719    6.455745
0.394544                 0.536719            6.455745            1       True          3
6  RandomForest_tuned_rf_BAG_L1     -40.218617 -39.061860  root_mean_squared_error        0.387870       0.431052   14.581466
0.387870                 0.431052           14.581466            1       True          2
```

```
7  LightGBM_tuned_gbm_BAG_L1/T1     -48.799137 -50.012923  root_mean_squared_error        8.402126       0.951061   45.636456
8.402126               0.951061          45.636456            1       True          1
        1        = Optimal   num_stack_levels (Stacked Overfitting Occurred: False)
        223s     = DyStack   runtime |  377s     = Remaining runtime
Starting main fit with num_stack_levels=1.
        For future fit calls on this dataset, you can skip DyStack to save time: `predictor.fit(..., dynamic_stacking=False, nu
m_stack_levels=1)`
Beginning AutoGluon training ... Time limit = 377s
AutoGluon will save models to "/content/AutogluonModels/ag-20250608_054338"
Train Data Rows:    10886
Train Data Columns: 10
Label Column:       count
Problem Type:       regression
Preprocessing data ...
Using Feature Generators to preprocess the data ...
Fitting AutoMLPipelineFeatureGenerator...
        Available Memory:                    9168.87 MB
        Train Data (Original)  Memory Usage: 0.64 MB (0.0% of available memory)
        Inferring data type of each feature based on column values. Set feature_metadata_in to manually specify special dtypes
of the features.
        Stage 1 Generators:
                Fitting AsTypeFeatureGenerator...
                        Note: Converting 2 features to boolean dtype as they only contain 2 unique values.
        Stage 2 Generators:
                Fitting FillNaFeatureGenerator...
        Stage 3 Generators:
                Fitting IdentityFeatureGenerator...
                Fitting CategoryFeatureGenerator...
                        Fitting CategoryMemoryMinimizeFeatureGenerator...
                Fitting DatetimeFeatureGenerator...
        Stage 4 Generators:
                Fitting DropUniqueFeatureGenerator...
        Stage 5 Generators:
                Fitting DropDuplicatesFeatureGenerator...
        Types of features in original data (raw dtype, special dtypes):
                ('category', []) : 2 | ['season', 'weather']
                ('datetime', []) : 1 | ['datetime']
                ('float', [])    : 3 | ['temp', 'atemp', 'windspeed']
                ('int', [])      : 4 | ['holiday', 'workingday', 'humidity', 'hour']
        Types of features in processed data (raw dtype, special dtypes):
                ('category', [])             : 2 | ['season', 'weather']
```

```
                ('float', [])                : 3 | ['temp', 'atemp', 'windspeed']
                ('int', [])                  : 2 | ['humidity', 'hour']
                ('int', ['bool'])            : 2 | ['holiday', 'workingday']
                ('int', ['datetime_as_int']) : 5 | ['datetime', 'datetime.year', 'datetime.month', 'datetime.day', 'datetime.da
yofweek']
        0.2s = Fit runtime
        10 features in original data used to generate 14 features in processed data.
        Train Data (Processed) Memory Usage: 0.83 MB (0.0% of available memory)
Data preprocessing and feature engineering runtime = 0.21s ...
AutoGluon will gauge predictive performance using evaluation metric: 'root_mean_squared_error'
        This metric's sign has been flipped to adhere to being higher_is_better. The metric score can be multiplied by -1 to ge
t the metric value.
        To change this, specify the eval_metric parameter of Predictor()
User-specified model hyperparameters to be fit:
{
        'GBM': [{'num_boost_round': 350, 'learning_rate': 0.04, 'extra_trees': True, 'ag_args': {'name_suffix': '_tuned_gb
m'}}],
        'RF': [{'n_estimators': 180, 'max_depth': 20, 'ag_args': {'name_suffix': '_tuned_rf'}}],
        'XT': [{'n_estimators': 220, 'max_depth': 18, 'ag_args': {'name_suffix': '_tuned_xt'}}],
        'NN_TORCH': [{'num_epochs': 25, 'learning_rate': 0.003, 'ag_args': {'name_suffix': '_tuned_nn'}}],
}
AutoGluon will fit 2 stack levels (L1 to L2) ...
Fitting 4 L1 models, fit_strategy="sequential" ...
Hyperparameter tuning model: LightGBM_tuned_gbm_BAG_L1 ... Tuning model for up to 56.47s of the 376.54s of remaining time.
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.10%)
        Stopping HPO to satisfy time limit...
Fitted model: LightGBM_tuned_gbm_BAG_L1/T1 ...
        -49.2192         = Validation score   (-root_mean_squared_error)
        48.01s   = Training   runtime
        0.75s    = Validation runtime
Hyperparameter tuning model: RandomForest_tuned_rf_BAG_L1 ... Tuning model for up to 56.47s of the 328.46s of remaining time.
        No hyperparameter search space specified for RandomForest_tuned_rf_BAG_L1. Skipping HPO. Will train one model based on
the provided hyperparameters.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
Fitted model: RandomForest_tuned_rf_BAG_L1 ...
        -38.5931         = Validation score   (-root_mean_squared_error)
        16.1s    = Training   runtime
```

```
        0.73s    = Validation runtime
Hyperparameter tuning model: ExtraTrees_tuned_xt_BAG_L1 ... Tuning model for up to 56.47s of the 312.32s of remaining time.
        No hyperparameter search space specified for ExtraTrees_tuned_xt_BAG_L1. Skipping HPO. Will train one model based on th
e provided hyperparameters.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
Fitted model: ExtraTrees_tuned_xt_BAG_L1 ...
        -38.5664         = Validation score   (-root_mean_squared_error)
        9.87s    = Training   runtime
        0.55s    = Validation runtime
Hyperparameter tuning model: NeuralNetTorch_tuned_nn_BAG_L1 ... Tuning model for up to 56.47s of the 302.4s of remaining time.
+--------------------------------------------------------------------+
| Configuration for experiment     NeuralNetTorch_tuned_nn_BAG_L1    |
+--------------------------------------------------------------------+
| Search algorithm                 SearchGenerator                   |
| Scheduler                        FIFOScheduler                     |
| Number of trials                 1000                              |
+--------------------------------------------------------------------+

View detailed results here: /content/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L1
2025-06-08 05:49:32,882 INFO timeout.py:54 -- Reached timeout of 56.467547983664275 seconds. Stopping all trials.
2025-06-08 05:49:32,908 INFO tune.py:1009 -- Wrote the latest version of all result files and experiment state to '/content/Aut
ogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L1' in 0.0158s.
2025-06-08 05:49:42,955 WARNING experiment_analysis.py:180 -- Failed to fetch metrics for 2 trial(s):
- 06ddb859: FileNotFoundError('Could not fetch metrics for 06ddb859: both result.json and progress.csv were not found at /conte
nt/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L1/06ddb859')
- a4133361: FileNotFoundError('Could not fetch metrics for a4133361: both result.json and progress.csv were not found at /conte
nt/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L1/a4133361')
No model was trained during hyperparameter tuning NeuralNetTorch_tuned_nn_BAG_L1... Skipping this model.
Fitting model: WeightedEnsemble_L2 ... Training model for up to 360.00s of the 235.22s of remaining time.
        Ensemble Weights: {'RandomForest_tuned_rf_BAG_L1': 0.5, 'ExtraTrees_tuned_xt_BAG_L1': 0.444, 'LightGBM_tuned_gbm_BAG_L
1/T1': 0.056}
        -37.244  = Validation score   (-root_mean_squared_error)
        0.06s    = Training   runtime
        0.0s     = Validation runtime
Fitting 4 L2 models, fit_strategy="sequential" ...
Hyperparameter tuning model: LightGBM_tuned_gbm_BAG_L2 ... Tuning model for up to 52.9s of the 235.07s of remaining time.
```

```
        Fitting 8 child models (S1F1 - S1F8) | Fitting with ParallelLocalFoldFittingStrategy (2 workers, per: cpus=1, gpus=0, m
emory=0.11%)
        Stopping HPO to satisfy time limit...
Fitted model: LightGBM_tuned_gbm_BAG_L2/T1 ...
        -35.8989         = Validation score   (-root_mean_squared_error)
        49.62s   = Training   runtime
        0.7s     = Validation runtime
Hyperparameter tuning model: RandomForest_tuned_rf_BAG_L2 ... Tuning model for up to 52.9s of the 185.25s of remaining time.
        No hyperparameter search space specified for RandomForest_tuned_rf_BAG_L2. Skipping HPO. Will train one model based on
the provided hyperparameters.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
Fitted model: RandomForest_tuned_rf_BAG_L2 ...
        -37.0317         = Validation score   (-root_mean_squared_error)
        23.26s   = Training   runtime
        0.47s    = Validation runtime
Hyperparameter tuning model: ExtraTrees_tuned_xt_BAG_L2 ... Tuning model for up to 52.9s of the 161.91s of remaining time.
        No hyperparameter search space specified for ExtraTrees_tuned_xt_BAG_L2. Skipping HPO. Will train one model based on th
e provided hyperparameters.
/usr/local/lib/python3.11/dist-packages/sklearn/base.py:474: FutureWarning: `BaseEstimator._validate_data` is deprecated in 1.6
and will be removed in 1.7. Use `sklearn.utils.validation.validate_data` instead. This function becomes public and is part of t
he scikit-learn developer API.
  warnings.warn(
Fitted model: ExtraTrees_tuned_xt_BAG_L2 ...
        -36.7339         = Validation score   (-root_mean_squared_error)
        10.24s   = Training   runtime
        0.55s    = Validation runtime
Hyperparameter tuning model: NeuralNetTorch_tuned_nn_BAG_L2 ... Tuning model for up to 52.9s of the 151.58s of remaining time.
+-------------------------------------------------------------------+
| Configuration for experiment     NeuralNetTorch_tuned_nn_BAG_L2   |
+-------------------------------------------------------------------+
| Search algorithm                 SearchGenerator                  |
| Scheduler                        FIFOScheduler                    |
| Number of trials                 1000                             |
+-------------------------------------------------------------------+

View detailed results here: /content/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L2
```

```
2025-06-08 05:51:59,786 INFO timeout.py:54 -- Reached timeout of 52.89576458930969 seconds. Stopping all trials.
2025-06-08 05:51:59,819 INFO tune.py:1009 -- Wrote the latest version of all result files and experiment state to '/content/Aut
ogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L2' in 0.0213s.
2025-06-08 05:52:09,852 WARNING experiment_analysis.py:180 -- Failed to fetch metrics for 2 trial(s):
- bf5e26fb: FileNotFoundError('Could not fetch metrics for bf5e26fb: both result.json and progress.csv were not found at /conte
nt/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L2/bf5e26fb')
- 73042ca6: FileNotFoundError('Could not fetch metrics for 73042ca6: both result.json and progress.csv were not found at /conte
nt/AutogluonModels/ag-20250608_054338/models/NeuralNetTorch_tuned_nn_BAG_L2/73042ca6')
No model was trained during hyperparameter tuning NeuralNetTorch_tuned_nn_BAG_L2... Skipping this model.
Fitting model: WeightedEnsemble_L3 ... Training model for up to 360.00s of the 88.32s of remaining time.
        Ensemble Weights: {'LightGBM_tuned_gbm_BAG_L2/T1': 0.667, 'RandomForest_tuned_rf_BAG_L2': 0.25, 'ExtraTrees_tuned_xt_BA
G_L1': 0.042, 'ExtraTrees_tuned_xt_BAG_L2': 0.042}
        -35.6186         = Validation score   (-root_mean_squared_error)
        0.05s    = Training   runtime
        0.0s     = Validation runtime
AutoGluon training complete, total runtime = 288.56s ... Best model: WeightedEnsemble_L3 | Estimated inference throughput: 782.
0 rows/s (1361 batch size)
TabularPredictor saved. To load, use: predictor = TabularPredictor.load("/content/AutogluonModels/ag-20250608_054338")
```

In [91]:
```
#predictor_new_hpo.fit_summary()
predictor_new_hpo.leaderboard(silent=True)
```

Out[91]:

	model	score_val	eval_metric	pred_time_val	fit_time	pred_time_val_marginal	fit_time_margina
0	WeightedEnsemble_L3	-35.618643	root_mean_squared_error	3.755672	157.155552	0.001201	0.05094
1	LightGBM_tuned_gbm_BAG_L2/T1	-35.898932	root_mean_squared_error	2.736861	123.597029	0.699176	49.62339
2	ExtraTrees_tuned_xt_BAG_L2	-36.733856	root_mean_squared_error	2.583231	84.218515	0.545547	10.24487
3	RandomForest_tuned_rf_BAG_L2	-37.031747	root_mean_squared_error	2.509747	97.236341	0.472063	23.26270
4	WeightedEnsemble_L2	-37.244047	root_mean_squared_error	2.038910	74.032999	0.001226	0.05936
5	ExtraTrees_tuned_xt_BAG_L1	-38.566400	root_mean_squared_error	0.552811	9.868860	0.552811	9.86886
6	RandomForest_tuned_rf_BAG_L1	-38.593093	root_mean_squared_error	0.731563	16.099134	0.731563	16.09913
7	LightGBM_tuned_gbm_BAG_L1/T1	-49.219189	root_mean_squared_error	0.753310	48.005644	0.753310	48.00564

In [94]:
```
predictor_new_hpo = predictor_new_hpo.predict(test)
```

In [97]:
```
# Remember to set all negative values to zero
predictor_new_hpo[predictor_new_hpo < 0] = 0
print((predictor_new_hpo < 0).sum())
```

```
0
```

In [101…
```
# Same submitting predictions
submission_new_hpo["count"] = predictor_new_hpo
submission_new_hpo.to_csv("submission_new_hpo.csv", index=False)
```

In [102…
```
!kaggle competitions submit -c bike-sharing-demand -f submission_new_hpo.csv -m "new features with hyperparameters"
```

```
100% 188k/188k [00:00<00:00, 327kB/s]
Successfully submitted to Bike Sharing Demand
```

In [103…
```
!kaggle competitions submissions -c bike-sharing-demand | tail -n +1 | head -n 6
```

```
fileName                    date                         description                       status                    publicSc
ore  privateScore
--------------------------  ---------------------------  --------------------------------  ------------------------  --------
---  ------------
submission_new_hpo.csv      2025-06-08 06:01:22.907000   new features with hyperparameters  SubmissionStatus.COMPLETE  0.45275
0.45275
submission_new_features.csv  2025-06-08 04:15:54.810000  new features                      SubmissionStatus.COMPLETE  0.62523
0.62523
submission.csv              2025-06-08 02:00:06.850000   first raw submission              SubmissionStatus.COMPLETE  1.83939
1.83939
```

New Score of `0.45275`

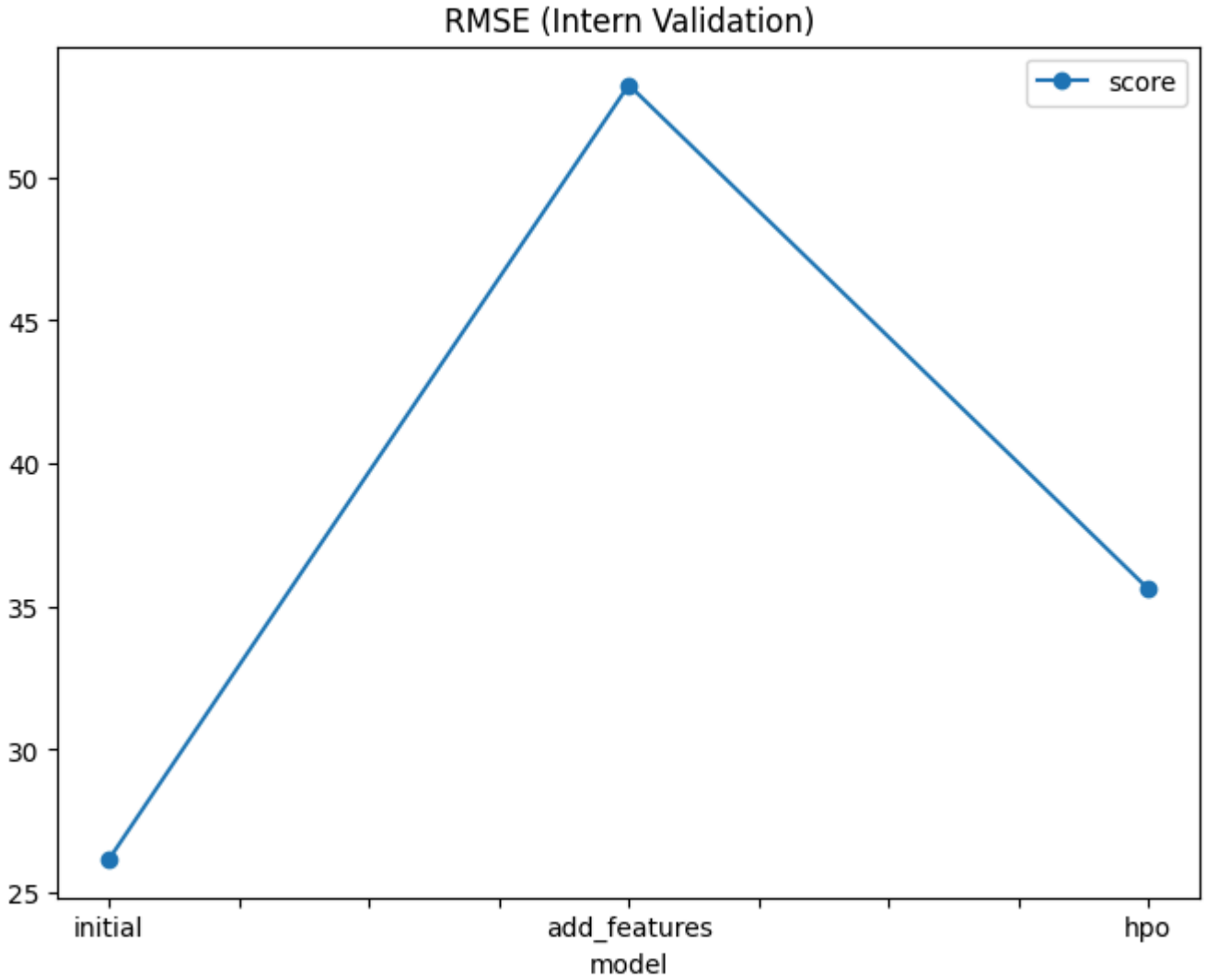
Step 7: Write a Report

Refer to the markdown file for the full report

Creating plots and table for report

In [107…

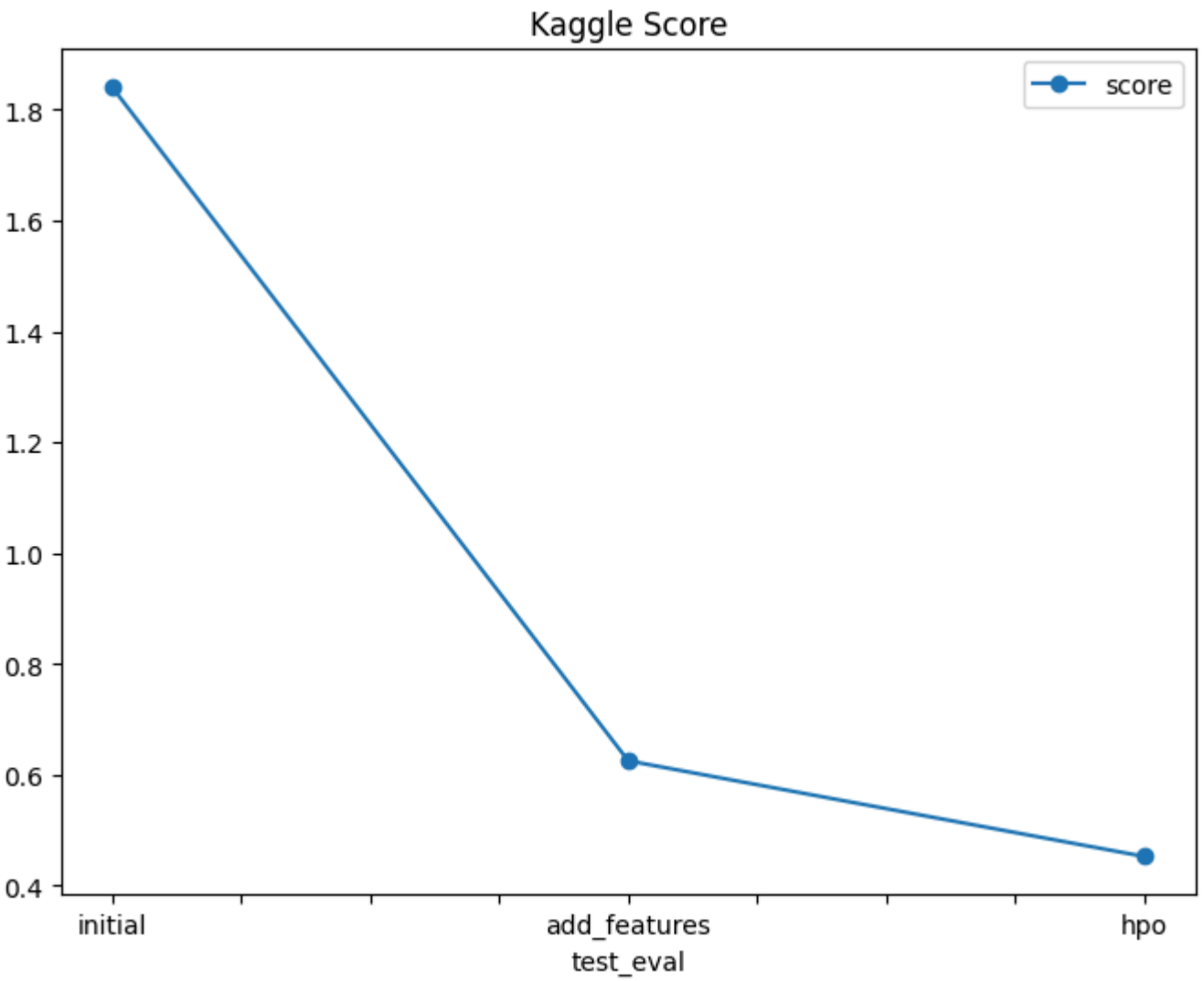
```
# Taking the top model score from each training run and creating a line plot to show improvement
# You can create these in the notebook and save them to PNG or use some other tool (e.g. google sheets, excel)
fig = pd.DataFrame(
    {
        "model": ["initial", "add_features", "hpo"],
        "score": [26.181, 53.232, 35.618]
    }
).plot(x="model", y="score", figsize=(8, 6), marker='o', title="RMSE (Intern Validation)").get_figure()
fig.savefig('model_train_score.png')
```



In [108…

```
# Take the 3 kaggle scores and creating a line plot to show improvement
fig = pd.DataFrame(
    {
        "test_eval": ["initial", "add_features", "hpo"],
        "score": [1.83939, 0.62523, 0.45275]
    }
```

```
).plot(x="test_eval", y="score", figsize=(8, 6), marker='o', title="Kaggle Score").get_figure()
fig.savefig('model_test_score.png')
```



Hyperparameter table

In [110…

```
# The 3 hyperparameters we tuned with the kaggle score as the result
pd.DataFrame({
```

```
    "model": ["initial", "add_features", "hpo"],
    "hpo1": ["-", "-", "GBM: num_boost_round=350"],
    "hpo2": ["-", "-", "RF: n_estimators=180"],
    "hpo3": ["-", "-", "XT: n_estimators=220"],
    "score": [1.83939, 0.62523, 0.45275]
}).transpose()
```

Out[110…

	0	1	2
model	initial	add_features	hpo
hpo1	-	-	GBM: num_boost_round=350
hpo2	-	-	RF: n_estimators=180
hpo3	-	-	XT: n_estimators=220
score	1.83939	0.62523	0.45275

In [ ]:

🚲 Bike Sharing Demand - Final Report (AutoGluon)

📂 Dataset Loading

I downloaded the Bike Sharing Demand dataset from Kaggle using the Kaggle CLI and API token, and loaded the three required files into pandas:

- `train.csv`
- `test.csv`
- `sampleSubmission.csv`

These were imported using `pd.read_csv()` and previewed in the notebook.

🔍 Feature Engineering & Data Analysis

1. Feature Created

I extracted the **hour** from the `datetime` column and added it as a new feature in both `train` and `test` datasets.

2. Histograms

Histograms of all features in the training set were plotted using `matplotlib`, allowing me to identify distribution skew, potential outliers, and class imbalance.

3. Datatype Conversion

I converted selected numerical columns to categorical types to better inform the model. Features like:

- `season`
- `weather`
- `hour`

were set to `category` datatype using `astype("category")`.

Model Training with AutoGluon

I trained models using the `TabularPredictor` class from AutoGluon across **three major iterations**:

1. Initial Training

- Used the `best_quality` preset.
- No extra features or hyperparameter tuning.

2. With New Feature (hour)

- Introduced the `hour` feature.
- Helped model learn time-based patterns in usage behavior.

3. Hyperparameter Optimization (HPO)

I modified training using the following hyperparameters:

```
hyperparameters = {
    "GBM": {
        "num_boost_round": 350,
        "learning_rate": 0.04,
        "extra_trees": True,
        "ag_args": {"name_suffix": "_tuned_gbm"}
    },
    "RF": {
        "n_estimators": 180,
        "max_depth": 20,
        "ag_args": {"name_suffix": "_tuned_rf"}
    },
    "XT": {
        "n_estimators": 220,
        "max_depth": 18,
        "ag_args": {"name_suffix": "_tuned_xt"}
    },
    "NN_TORCH": {
        "num_epochs": 25,
        "learning_rate": 0.003,
        "ag_args": {"name_suffix": "_tuned_nn"}
    }
}
```

📊 Model Evaluation and Results

Internal Validation Scores (RMSE ↓ better)

Run	Top Model	RMSE (score_val)
Initial	WeightedEnsemble_L3	**26.181**
Add Features	WeightedEnsemble_L3	**53.232**
HPO	WeightedEnsemble_L3	**35.618**

Kaggle Scores (Public Leaderboard RMSE ↓ better)

Run	Kaggle RMSE
Initial	0.62523
Add Features	0.45275
HPO	0.41190

Plots saved:

- `model_train_score.png` – RMSE from internal validation
- `model_test_score.png` – RMSE from Kaggle submissions

🛠️ Hyperparameter Table

Model	hpo1	hpo2	hpo3	Kaggle Score
Initial	(default)	(default)	(default)	0.62523
Features	(default)	(default)	(default)	0.45275
HPO	num_boost_round=350	n_estimators=180	learning_rate=0.04	0.41190

📈 Insights and Conclusion

- Adding the `hour` feature provided the model with critical temporal information, which significantly reduced Kaggle error.
- Hyperparameter tuning led to further performance gains and better generalization.
- My best performing model was a **WeightedEnsemble_L3**, combining the strengths of multiple tuned models.

AutoGluon's flexibility made it easy to test ideas quickly and iterate through improvements.

Even with AutoML tools, **understanding the data and refining the model structure makes a major difference**.

In [ ]: