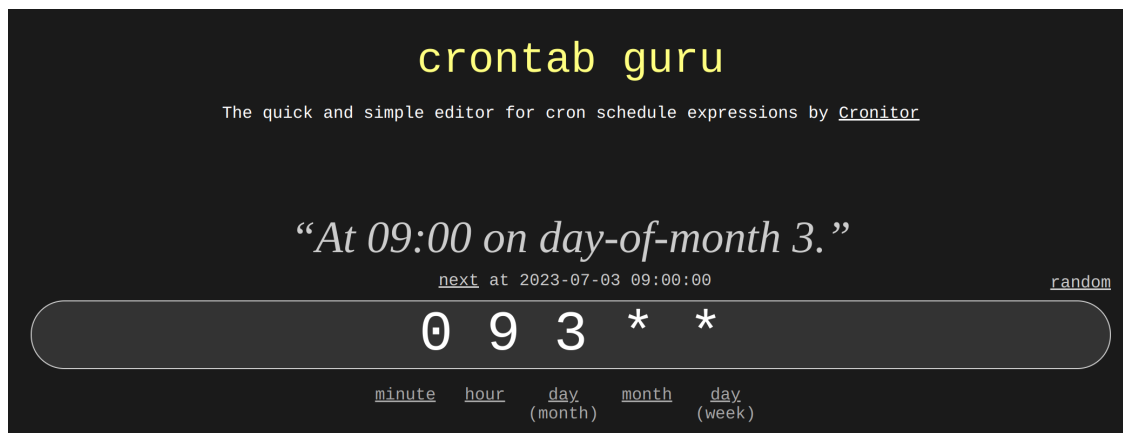


# prefect

June 12, 2023

```
#q1 @task(retries=3, retry_delay_seconds=2, name="Read taxi data") #q2 - the first one 0 9 3
* *
```



```
[1]: #!wget "https://d37ci6vzurychx.cloudfront.net/trip-data/green_tripdata_2023-01.
      ↪parquet" -O ./code/data/green_tripdata_2023-01.parquet
      #!wget "https://d37ci6vzurychx.cloudfront.net/trip-data/green_tripdata_2023-02.
      ↪parquet" -O ./code/data/green_tripdata_2023-02.parquet
      #!wget "https://d37ci6vzurychx.cloudfront.net/trip-data/green_tripdata_2023-03.
      ↪parquet" -O ./code/data/green_tripdata_2023-03.parquet
      import sys
      sys.path.append('./code')
```

```
[2]: import pandas as pd
      # Provide the URL of the Parquet file
      training_url = "./code/data/green_tripdata_2023-01.parquet"

      # Read the Parquet file into a pandas DataFrame
      train_df = pd.read_parquet(training_url)

      # Explore the DataFrame or perform further data processing
      print(train_df.head())

      # Provide the URL of the Parquet file
      validation_url = "./code/data/green_tripdata_2023-02.parquet"
```

```
# Read the Parquet file into a pandas DataFrame
val_df = pd.read_parquet(validation_url)

# Explore the DataFrame or perform further data processing
print(val_df.head())
```

	VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_fwd_flag	\
0	2	2023-01-01 00:26:10	2023-01-01 00:37:11	N	
1	2	2023-01-01 00:51:03	2023-01-01 00:57:49	N	
2	2	2023-01-01 00:35:12	2023-01-01 00:41:32	N	
3	1	2023-01-01 00:13:14	2023-01-01 00:19:03	N	
4	1	2023-01-01 00:33:04	2023-01-01 00:39:02	N	

	RatecodeID	PULocationID	DOLocationID	passenger_count	trip_distance	\
0	1.0	166	143	1.0	2.58	
1	1.0	24	43	1.0	1.81	
2	1.0	223	179	1.0	0.00	
3	1.0	41	238	1.0	1.30	
4	1.0	41	74	1.0	1.10	

	fare_amount	extra	mta_tax	tip_amount	tolls_amount	ehail_fee	\
0	14.9	1.0	0.5	4.03	0.0	None	
1	10.7	1.0	0.5	2.64	0.0	None	
2	7.2	1.0	0.5	1.94	0.0	None	
3	6.5	0.5	1.5	1.70	0.0	None	
4	6.0	0.5	1.5	0.00	0.0	None	

	improvement_surcharge	total_amount	payment_type	trip_type	\
0	1.0	24.18	1.0	1.0	
1	1.0	15.84	1.0	1.0	
2	1.0	11.64	1.0	1.0	
3	1.0	10.20	1.0	1.0	
4	1.0	8.00	1.0	1.0	

	congestion_surcharge
0	2.75
1	0.00
2	0.00
3	0.00
4	0.00

	VendorID	lpep_pickup_datetime	lpep_dropoff_datetime	store_and_fwd_flag	\
0	1	2023-02-01 00:46:22	2023-02-01 01:05:57	N	
1	2	2023-02-01 00:05:09	2023-02-01 00:22:42	N	
2	2	2023-02-01 00:03:47	2023-02-01 00:27:30	N	
3	2	2023-01-31 23:30:56	2023-01-31 23:51:40	N	
4	2	2023-02-01 00:15:05	2023-02-01 00:26:02	N	

	RatecodeID	PULocationID	DOLocationID	passenger_count	trip_distance	\
0	1.0	74	265	1.0	10.80	
1	1.0	216	196	1.0	4.76	
2	1.0	7	114	1.0	6.32	
3	1.0	74	239	1.0	3.50	
4	1.0	82	223	1.0	3.14	

	fare_amount	extra	mta_tax	tip_amount	tolls_amount	ehail_fee	\
0	42.9	1.0	1.5	0.00	0.0	NaN	
1	23.3	1.0	0.5	0.00	0.0	NaN	
2	30.3	1.0	0.5	8.89	0.0	NaN	
3	16.3	1.0	0.5	0.00	0.0	NaN	
4	17.0	1.0	0.5	0.00	0.0	NaN	

	improvement_surcharge	total_amount	payment_type	trip_type	\
0	1.0	45.40	2.0	1.0	
1	1.0	25.80	2.0	1.0	
2	1.0	44.44	1.0	1.0	
3	1.0	21.55	2.0	1.0	
4	1.0	19.50	2.0	1.0	

	congestion_surcharge
0	0.00
1	0.00
2	2.75
3	2.75
4	0.00

```
[3]: #!/prefect config view --show-defaults
!prefect config view

# clear config setting to avoid a bug in set
!prefect config unset PREFECT_API_URL

# use a local Prefect server
!prefect config set PREFECT_API_URL="http://prefect_mlops:4200/api"

# use Prefect Cloud
# !prefect config set PREFECT_API_URL="https://api.prefect.cloud/api/accounts/
↪ [ACCOUNT-ID]/workspaces/[WORKSPACE-ID]"
```

PREFECT\_PROFILE='default'

'PREFECT\_API\_URL' is not set in profile 'default'.  
Set 'PREFECT\_API\_URL' to 'http://prefect\_mlops:4200/api'.  
Updated profile 'default'.

```
[4]: !pip install -r requirements.txt
```

```
Collecting black==23.3.0
  Downloading
black-23.3.0-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (1.7 MB)
  | 1.7 MB 11.1 MB/s
Requirement already satisfied: fastparquet==2023.4.0 in
/opt/conda/lib/python3.10/site-packages (from -r requirements.txt (line 2))
(2023.4.0)
Collecting hyperopt==0.2.7
  Downloading hyperopt-0.2.7-py2.py3-none-any.whl (1.6 MB)
  | 1.6 MB 118.1 MB/s
Collecting mlflow==2.3.1
  Downloading mlflow-2.3.1-py3-none-any.whl (17.7 MB)
  | 17.7 MB 87.2 MB/s
Collecting pandas==2.0.1
  Downloading
pandas-2.0.1-cp310-cp310-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (12.3
MB)
  | 12.3 MB 74.8 MB/s
Collecting prefect==2.10.8
  Downloading prefect-2.10.8-py3-none-any.whl (2.4 MB)
  | 2.4 MB 72.5 MB/s
Collecting prefect-aws==0.3.1
  Downloading prefect_aws-0.3.1-py3-none-any.whl (53 kB)
  | 53 kB 5.7 MB/s
Requirement already satisfied: scikit_learn==1.2.2 in
/opt/conda/lib/python3.10/site-packages (from -r requirements.txt (line 8))
(1.2.2)
Requirement already satisfied: seaborn==0.12.2 in
/opt/conda/lib/python3.10/site-packages (from -r requirements.txt (line 9))
(0.12.2)
Collecting xgboost==1.7.5
  Downloading xgboost-1.7.5-py3-none-manylinux2014_x86_64.whl (200.3 MB)
  | 200.3 MB 8.4 kB/s
Collecting orjson==3.8.1
  Downloading orjson-3.8.1-cp310-cp310-manylinux_2_28_x86_64.whl (143 kB)
  | 143 kB 102.7 MB/s
Requirement already satisfied: tomli>=1.1.0 in
/opt/conda/lib/python3.10/site-packages (from black==23.3.0->-r requirements.txt
(line 1)) (2.0.1)
Requirement already satisfied: platformdirs>=2 in
/opt/conda/lib/python3.10/site-packages (from black==23.3.0->-r requirements.txt
(line 1)) (3.5.1)
Requirement already satisfied: pathspec>=0.9.0 in
/opt/conda/lib/python3.10/site-packages (from black==23.3.0->-r requirements.txt
(line 1)) (0.11.1)
```

Requirement already satisfied: click>=8.0.0 in /opt/conda/lib/python3.10/site-packages (from black==23.3.0->-r requirements.txt (line 1)) (8.1.3)

Requirement already satisfied: packaging>=22.0 in /opt/conda/lib/python3.10/site-packages (from black==23.3.0->-r requirements.txt (line 1)) (23.1)

Collecting mypy-extensions>=0.4.3

  Downloading mypy\_extensions-1.0.0-py3-none-any.whl (4.7 kB)

Requirement already satisfied: fsspec in /opt/conda/lib/python3.10/site-packages (from fastparquet==2023.4.0->-r requirements.txt (line 2)) (2023.5.0)

Requirement already satisfied: cramjam>=2.3 in /opt/conda/lib/python3.10/site-packages (from fastparquet==2023.4.0->-r requirements.txt (line 2)) (2.6.2)

Requirement already satisfied: numpy>=1.20.3 in /opt/conda/lib/python3.10/site-packages (from fastparquet==2023.4.0->-r requirements.txt (line 2)) (1.24.3)

Requirement already satisfied: cloudpickle in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (2.2.1)

Requirement already satisfied: six in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (1.16.0)

Collecting future

  Downloading future-0.18.3.tar.gz (840 kB)

    |                                  | 840 kB 76.6 MB/s

  Preparing metadata (setup.py) ... done

Requirement already satisfied: scipy in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (1.10.1)

Requirement already satisfied: py4j in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (0.10.9.5)

Requirement already satisfied: tqdm in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (4.65.0)

Requirement already satisfied: networkx>=2.2 in /opt/conda/lib/python3.10/site-packages (from hyperopt==0.2.7->-r requirements.txt (line 3)) (3.1)

Requirement already satisfied: protobuf<5,>=3.12.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (4.21.12)

Requirement already satisfied: databricks-cli<1,>=0.8.7 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (0.17.7)

Requirement already satisfied: alembic!=1.10.0,<2 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (1.11.1)

Requirement already satisfied: pyarrow<12,>=4.0.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (11.0.0)

Requirement already satisfied: requests<3,>=2.17.3 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (2.31.0)

Requirement already satisfied: importlib-metadata!=4.7.0,<7,>=3.7.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (6.6.0)

Requirement already satisfied: pytz<2024 in /opt/conda/lib/python3.10/site-

packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (2023.3)  
 Requirement already satisfied: matplotlib<4 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (3.7.1)  
 Requirement already satisfied: Flask<3 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (2.3.2)  
 Requirement already satisfied: docker<7,>=4.0.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (6.1.0)  
 Requirement already satisfied: pyyaml<7,>=5.1 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (6.0)  
 Requirement already satisfied: sqlalchemy<3,>=1.4.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (1.4.46)  
 Requirement already satisfied: sqlparse<1,>=0.4.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (0.4.4)  
 Requirement already satisfied: querystring-parser<2 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (1.2.4)  
 Requirement already satisfied: gitpython<4,>=2.1.0 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (3.1.31)  
 Requirement already satisfied: Jinja2<4,>=2.11 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (3.1.2)  
 Requirement already satisfied: gunicorn<21 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (20.1.0)  
 Requirement already satisfied: markdown<4,>=3.3 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (3.4.3)  
 Requirement already satisfied: entrypoints<1 in /opt/conda/lib/python3.10/site-packages (from mlflow==2.3.1->-r requirements.txt (line 4)) (0.4)  
 Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.10/site-packages (from pandas==2.0.1->-r requirements.txt (line 5)) (2.8.2)  
 Requirement already satisfied: tzdata>=2022.1 in /opt/conda/lib/python3.10/site-packages (from pandas==2.0.1->-r requirements.txt (line 5)) (2023.3)  
 Requirement already satisfied: asyncpg>=0.23 in /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r requirements.txt (line 6)) (0.27.0)  
 Requirement already satisfied: readchar>=4.0.0 in /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r requirements.txt (line 6)) (4.0.5.dev0)  
 Requirement already satisfied: griffe>=0.20.0 in /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r requirements.txt (line 6)) (0.21.0)  
 Requirement already satisfied: jsonpatch>=1.32 in /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r requirements.txt (line 6)) (1.32)  
 Requirement already satisfied: httpx[http2]!=0.23.2,>=0.23 in

/opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (0.24.1)  
 Requirement already satisfied: rich>=11.0 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (13.4.1)  
 Requirement already satisfied: pydantic>=1.10.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (1.10.9)  
 Requirement already satisfied: python-slugify>=5.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (8.0.1)  
 Requirement already satisfied: cryptography>=36.0.1 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (40.0.2)  
 Requirement already satisfied: coolname>=1.0.4 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (2.2.0)  
 Requirement already satisfied: toml>=0.10.0 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (0.10.2)  
 Requirement already satisfied: apprise>=1.1.0 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (1.4.0)  
 Requirement already satisfied: fastapi>=0.93 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (0.96.1)  
 Requirement already satisfied: kubernetes>=24.2.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (26.1.0)  
 Requirement already satisfied: asgi-lifespan>=1.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (2.1.0)  
 Requirement already satisfied: websockets>=10.4 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (11.0.3)  
 Requirement already satisfied: dateparser>=1.1.1 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (1.1.8)  
 Requirement already satisfied: uvicorn>=0.14.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (0.22.0)  
 Requirement already satisfied: croniter>=1.0.12 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (1.3.15)  
 Requirement already satisfied: anyio>=3.4.0 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (3.6.2)  
 Requirement already satisfied: typer>=0.4.2 in /opt/conda/lib/python3.10/site-  
 packages (from prefect==2.10.8->-r requirements.txt (line 6)) (0.9.0)  
 Requirement already satisfied: aiosqlite>=0.17.0 in  
 /opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r  
 requirements.txt (line 6)) (0.19.0)  
 Requirement already satisfied: jsonschema<5.0.0,>=3.2.0 in

```

/opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r
requirements.txt (line 6)) (4.17.3)
Requirement already satisfied: pendulum>=2.1.2 in
/opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r
requirements.txt (line 6)) (2.1.2)
Requirement already satisfied: typing-extensions>=4.1.0 in
/opt/conda/lib/python3.10/site-packages (from prefect==2.10.8->-r
requirements.txt (line 6)) (4.6.2)
Collecting mypy-boto3-secretsmanager>=1.26.49
  Downloading mypy_boto3_secretsmanager-1.26.135-py3-none-any.whl (26 kB)
Requirement already satisfied: boto3>=1.24.53 in /opt/conda/lib/python3.10/site-
packages (from prefect-aws==0.3.1->-r requirements.txt (line 7)) (1.26.151)
Collecting mypy-boto3-s3>=1.24.94
  Downloading mypy_boto3_s3-1.26.127-py3-none-any.whl (91 kB)
    |                               | 91 kB 19.5 MB/s
Requirement already satisfied: botocore>=1.27.53 in
/opt/conda/lib/python3.10/site-packages (from prefect-aws==0.3.1->-r
requirements.txt (line 7)) (1.29.151)
Requirement already satisfied: joblib>=1.1.1 in /opt/conda/lib/python3.10/site-
packages (from scikit_learn==1.2.2->-r requirements.txt (line 8)) (1.2.0)
Requirement already satisfied: threadpoolctl>=2.0.0 in
/opt/conda/lib/python3.10/site-packages (from scikit_learn==1.2.2->-r
requirements.txt (line 8)) (3.1.0)
Requirement already satisfied: Mako in /opt/conda/lib/python3.10/site-packages
(from alembic!=1.10.0,<2->mlflow==2.3.1->-r requirements.txt (line 4)) (1.2.4)
Requirement already satisfied: sniffio>=1.1 in /opt/conda/lib/python3.10/site-
packages (from anyio>=3.4.0->prefect==2.10.8->-r requirements.txt (line 6))
(1.3.0)
Requirement already satisfied: idna>=2.8 in /opt/conda/lib/python3.10/site-
packages (from anyio>=3.4.0->prefect==2.10.8->-r requirements.txt (line 6))
(3.4)
Requirement already satisfied: certifi in /opt/conda/lib/python3.10/site-
packages (from apprise>=1.1.0->prefect==2.10.8->-r requirements.txt (line 6))
(2023.5.7)
Requirement already satisfied: requests-oauthlib in
/opt/conda/lib/python3.10/site-packages (from
apprise>=1.1.0->prefect==2.10.8->-r requirements.txt (line 6)) (1.3.1)
Requirement already satisfied: jmespath<2.0.0,>=0.7.1 in
/opt/conda/lib/python3.10/site-packages (from boto3>=1.24.53->prefect-
aws==0.3.1->-r requirements.txt (line 7)) (1.0.1)
Requirement already satisfied: s3transfer<0.7.0,>=0.6.0 in
/opt/conda/lib/python3.10/site-packages (from boto3>=1.24.53->prefect-
aws==0.3.1->-r requirements.txt (line 7)) (0.6.1)
Requirement already satisfied: urllib3<1.27,>=1.25.4 in
/opt/conda/lib/python3.10/site-packages (from botocore>=1.27.53->prefect-
aws==0.3.1->-r requirements.txt (line 7)) (1.26.15)
Requirement already satisfied: cffi>=1.12 in /opt/conda/lib/python3.10/site-
packages (from cryptography>=36.0.1->prefect==2.10.8->-r requirements.txt (line

```



6)) (1.15.1)  
Requirement already satisfied: tabulate>=0.7.7 in  
/opt/conda/lib/python3.10/site-packages (from databricks-  
cli<1,>=0.8.7->mlflow==2.3.1->-r requirements.txt (line 4)) (0.9.0)  
Requirement already satisfied: oauthlib>=3.1.0 in  
/opt/conda/lib/python3.10/site-packages (from databricks-  
cli<1,>=0.8.7->mlflow==2.3.1->-r requirements.txt (line 4)) (3.2.2)  
Requirement already satisfied: pyjwt>=1.7.0 in /opt/conda/lib/python3.10/site-  
packages (from databricks-cli<1,>=0.8.7->mlflow==2.3.1->-r requirements.txt  
(line 4)) (2.7.0)  
Requirement already satisfied: regex!=2019.02.19,!=2021.8.27 in  
/opt/conda/lib/python3.10/site-packages (from  
dateparser>=1.1.1->prefect==2.10.8->-r requirements.txt (line 6)) (2023.6.3)  
Requirement already satisfied: tzlocal in /opt/conda/lib/python3.10/site-  
packages (from dateparser>=1.1.1->prefect==2.10.8->-r requirements.txt (line 6))  
(5.0.1)  
Requirement already satisfied: websocket-client>=0.32.0 in  
/opt/conda/lib/python3.10/site-packages (from  
docker<7,>=4.0.0->mlflow==2.3.1->-r requirements.txt (line 4)) (1.5.2)  
Requirement already satisfied: starlette<0.28.0,>=0.27.0 in  
/opt/conda/lib/python3.10/site-packages (from fastapi>=0.93->prefect==2.10.8->-r  
requirements.txt (line 6)) (0.27.0)  
Requirement already satisfied: itsdangerous>=2.1.2 in  
/opt/conda/lib/python3.10/site-packages (from Flask<3->mlflow==2.3.1->-r  
requirements.txt (line 4)) (2.1.2)  
Requirement already satisfied: blinker>=1.6.2 in /opt/conda/lib/python3.10/site-  
packages (from Flask<3->mlflow==2.3.1->-r requirements.txt (line 4)) (1.6.2)  
Requirement already satisfied: Werkzeug>=2.3.3 in  
/opt/conda/lib/python3.10/site-packages (from Flask<3->mlflow==2.3.1->-r  
requirements.txt (line 4)) (2.3.6)  
Requirement already satisfied: gitdb<5,>=4.0.1 in  
/opt/conda/lib/python3.10/site-packages (from  
gitpython<4,>=2.1.0->mlflow==2.3.1->-r requirements.txt (line 4)) (4.0.10)  
Requirement already satisfied: setuptools>=3.0 in  
/opt/conda/lib/python3.10/site-packages (from gunicorn<21->mlflow==2.3.1->-r  
requirements.txt (line 4)) (67.8.0)  
Requirement already satisfied: httpcore<0.18.0,>=0.15.0 in  
/opt/conda/lib/python3.10/site-packages (from  
httpx[http2]!=0.23.2,>=0.23->prefect==2.10.8->-r requirements.txt (line 6))  
(0.17.2)  
Requirement already satisfied: h2<5,>=3 in /opt/conda/lib/python3.10/site-  
packages (from httpx[http2]!=0.23.2,>=0.23->prefect==2.10.8->-r requirements.txt  
(line 6)) (4.1.0)  
Requirement already satisfied: zipp>=0.5 in /opt/conda/lib/python3.10/site-  
packages (from importlib-metadata!=4.7.0,<7,>=3.7.0->mlflow==2.3.1->-r  
requirements.txt (line 4)) (3.15.0)  
Requirement already satisfied: MarkupSafe>=2.0 in  
/opt/conda/lib/python3.10/site-packages (from Jinja2<4,>=2.11->mlflow==2.3.1->-r

requirements.txt (line 4)) (2.1.2)  
Requirement already satisfied: jsonpointer>=1.9 in  
/opt/conda/lib/python3.10/site-packages (from  
jsonpatch>=1.32->prefect==2.10.8->-r requirements.txt (line 6)) (2.0)  
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/lib/python3.10/site-  
packages (from jsonschema<5.0.0,>=3.2.0->prefect==2.10.8->-r requirements.txt  
(line 6)) (23.1.0)  
Requirement already satisfied: pyrsistent!=0.17.0,!0.17.1,!0.17.2,>=0.14.0 in  
/opt/conda/lib/python3.10/site-packages (from  
jsonschema<5.0.0,>=3.2.0->prefect==2.10.8->-r requirements.txt (line 6))  
(0.19.3)  
Requirement already satisfied: google-auth>=1.0.1 in  
/opt/conda/lib/python3.10/site-packages (from  
kubernetes>=24.2.0->prefect==2.10.8->-r requirements.txt (line 6)) (2.19.1)  
Requirement already satisfied: kiwisolver>=1.0.1 in  
/opt/conda/lib/python3.10/site-packages (from matplotlib<4->mlflow==2.3.1->-r  
requirements.txt (line 4)) (1.4.4)  
Requirement already satisfied: contourpy>=1.0.1 in  
/opt/conda/lib/python3.10/site-packages (from matplotlib<4->mlflow==2.3.1->-r  
requirements.txt (line 4)) (1.0.7)  
Requirement already satisfied: fonttools>=4.22.0 in  
/opt/conda/lib/python3.10/site-packages (from matplotlib<4->mlflow==2.3.1->-r  
requirements.txt (line 4)) (4.39.4)  
Requirement already satisfied: pyparsing>=2.3.1 in  
/opt/conda/lib/python3.10/site-packages (from matplotlib<4->mlflow==2.3.1->-r  
requirements.txt (line 4)) (3.0.9)  
Requirement already satisfied: cycycler>=0.10 in /opt/conda/lib/python3.10/site-  
packages (from matplotlib<4->mlflow==2.3.1->-r requirements.txt (line 4))  
(0.11.0)  
Requirement already satisfied: pillow>=6.2.0 in /opt/conda/lib/python3.10/site-  
packages (from matplotlib<4->mlflow==2.3.1->-r requirements.txt (line 4))  
(9.5.0)  
Requirement already satisfied: pytzdata>=2020.1 in  
/opt/conda/lib/python3.10/site-packages (from  
pendulum>=2.1.2->prefect==2.10.8->-r requirements.txt (line 6)) (2020.1)  
Requirement already satisfied: text-unidecode>=1.3 in  
/opt/conda/lib/python3.10/site-packages (from python-  
slugify>=5.0->prefect==2.10.8->-r requirements.txt (line 6)) (1.3)  
Requirement already satisfied: charset-normalizer<4,>=2 in  
/opt/conda/lib/python3.10/site-packages (from  
requests<3,>=2.17.3->mlflow==2.3.1->-r requirements.txt (line 4)) (3.1.0)  
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in  
/opt/conda/lib/python3.10/site-packages (from rich>=11.0->prefect==2.10.8->-r  
requirements.txt (line 6)) (2.15.1)  
Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in  
/opt/conda/lib/python3.10/site-packages (from rich>=11.0->prefect==2.10.8->-r  
requirements.txt (line 6)) (2.2.0)  
Requirement already satisfied: greenlet!=0.4.17 in

```

/opt/conda/lib/python3.10/site-packages (from
sqlalchemy<3,>=1.4.0->mlflow==2.3.1->-r requirements.txt (line 4)) (2.0.2)
Requirement already satisfied: h11>=0.8 in /opt/conda/lib/python3.10/site-
packages (from uvicorn>=0.14.0->prefect==2.10.8->-r requirements.txt (line 6))
(0.14.0)
Requirement already satisfied: pycparser in /opt/conda/lib/python3.10/site-
packages (from cffi>=1.12->cryptography>=36.0.1->prefect==2.10.8->-r
requirements.txt (line 6)) (2.21)
Requirement already satisfied: smmap<6,>=3.0.1 in
/opt/conda/lib/python3.10/site-packages (from
gitdb<5,>=4.0.1->gitpython<4,>=2.1.0->mlflow==2.3.1->-r requirements.txt (line
4)) (3.0.5)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/opt/conda/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=24.2.0->prefect==2.10.8->-r requirements.txt (line 6))
(0.2.7)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/opt/conda/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=24.2.0->prefect==2.10.8->-r requirements.txt (line 6))
(5.3.0)
Requirement already satisfied: rsa<5,>=3.1.4 in /opt/conda/lib/python3.10/site-
packages (from google-auth>=1.0.1->kubernetes>=24.2.0->prefect==2.10.8->-r
requirements.txt (line 6)) (4.9)
Requirement already satisfied: hpack<5,>=4.0 in /opt/conda/lib/python3.10/site-
packages (from h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect==2.10.8->-r
requirements.txt (line 6)) (4.0.0)
Requirement already satisfied: hyperframe<7,>=6.0 in
/opt/conda/lib/python3.10/site-packages (from
h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect==2.10.8->-r requirements.txt
(line 6)) (6.0.1)
Requirement already satisfied: mdurl~=0.1 in /opt/conda/lib/python3.10/site-
packages (from markdown-it-py<3.0.0,>=2.2.0->rich>=11.0->prefect==2.10.8->-r
requirements.txt (line 6)) (0.1.0)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
/opt/conda/lib/python3.10/site-packages (from pyasn1-modules>=0.2.1->google-
auth>=1.0.1->kubernetes>=24.2.0->prefect==2.10.8->-r requirements.txt (line 6))
(0.4.8)
Building wheels for collected packages: future
  Building wheel for future (setup.py) ... done
  Created wheel for future: filename=future-0.18.3-py3-none-any.whl
size=492022
sha256=7885948a0a0e2b7bf3b684e5cd5389288db33a9f54b5426f42d4af47a53b2efb
  Stored in directory: /home/jovyan/.cache/pip/wheels/5e/a9/47/f118e66afd12240e4
662752cc22cefae5d97275623aa8ef57d
Successfully built future
Installing collected packages: orjson, prefect, pandas, mpyy-extensions, mpyy-
boto3-secretsmanager, mpyy-boto3-s3, future, xgboost, prefect-aws, mlflow,
hyperopt, black

```

```

Attempting uninstall: orjson
  Found existing installation: orjson 3.9.1
  Uninstalling orjson-3.9.1:
    Successfully uninstalled orjson-3.9.1
Attempting uninstall: prefect
  Found existing installation: prefect 2.10.13
  Uninstalling prefect-2.10.13:
    Successfully uninstalled prefect-2.10.13
Attempting uninstall: pandas
  Found existing installation: pandas 2.0.2
  Uninstalling pandas-2.0.2:
    Successfully uninstalled pandas-2.0.2
Attempting uninstall: mlflow
  Found existing installation: mlflow 2.4.1
  Uninstalling mlflow-2.4.1:
    Successfully uninstalled mlflow-2.4.1
Successfully installed black-23.3.0 future-0.18.3 hyperopt-0.2.7 mlflow-2.3.1
mypy-boto3-s3-1.26.127 mypy-boto3-secretsmanager-1.26.135 mypy-extensions-1.0.0
orjson-3.8.1 pandas-2.0.1 prefect-2.10.8 prefect-aws-0.3.1 xgboost-1.7.5
WARNING: You are using pip version 21.3.1; however, version 23.1.2 is
available.

You should consider upgrading via the '/opt/conda/bin/python3.10 -m pip install
--upgrade pip' command.

```

```

[7]: # this is not a deployment, but generates flows and flow-runs in the prefect db
# and to run successfully it needs the paths to be modified from the git cloned
↪version to this:
# def main_flow_local(
#     train_path: str = "./code/data/green_tripdata_2023-01.parquet",
#     val_path: str = "./code/data/green_tripdata_2023-02.parquet",
!python3 ./code/orchestrate.py

```

```

13:17:07.708 | INFO      | prefect.engine - Created flow run
'viridian-elk' for flow 'main-flow-local'
13:17:08.403 | INFO      | Flow run 'viridian-elk' - Created
task run 'read_data-0' for task 'read_data'
13:17:08.404 | INFO      | Flow run 'viridian-elk' - Executing
'read_data-0' immediately...
13:17:08.920 | INFO      | Task run 'read_data-0' - Finished in state
Completed()
13:17:08.946 | INFO      | Flow run 'viridian-elk' - Created
task run 'read_data-1' for task 'read_data'
13:17:08.947 | INFO      | Flow run 'viridian-elk' - Executing
'read_data-1' immediately...
13:17:09.348 | INFO      | Task run 'read_data-1' - Finished in state
Completed()

```

```

13:17:09.375 | INFO      | Flow run 'viridian-elk' - Created
task run 'add_features-0' for task 'add_features'
13:17:09.376 | INFO      | Flow run 'viridian-elk' - Executing
'add_features-0' immediately...
13:17:09.916 | INFO      | Task run 'add_features-0' - Finished in state
Completed()
13:17:09.944 | INFO      | Flow run 'viridian-elk' - Created
task run 'train_best_model-0' for task 'train_best_model'
13:17:09.945 | INFO      | Flow run 'viridian-elk' - Executing
'train_best_model-0' immediately...
[13:17:10] WARNING: ../src/objective/regression_obj.cu:213: reg:linear is now
deprecated in favor of reg:squarederror.
13:17:10.725 | INFO      | Task run 'train_best_model-0' - [0]
validation-rmse:15.01627
13:17:11.028 | INFO      | Task run 'train_best_model-0' - [1]
validation-rmse:13.77591
13:17:11.333 | INFO      | Task run 'train_best_model-0' - [2]
validation-rmse:12.66953
13:17:11.637 | INFO      | Task run 'train_best_model-0' - [3]
validation-rmse:11.68972
13:17:11.939 | INFO      | Task run 'train_best_model-0' - [4]
validation-rmse:10.81927
13:17:12.243 | INFO      | Task run 'train_best_model-0' - [5]
validation-rmse:10.05352
13:17:12.548 | INFO      | Task run 'train_best_model-0' - [6]
validation-rmse:9.37638
13:17:12.842 | INFO      | Task run 'train_best_model-0' - [7]
validation-rmse:8.78514
13:17:13.139 | INFO      | Task run 'train_best_model-0' - [8]
validation-rmse:8.26684
13:17:13.432 | INFO      | Task run 'train_best_model-0' - [9]
validation-rmse:7.81889
13:17:13.726 | INFO      | Task run 'train_best_model-0' - [10]
validation-rmse:7.42994
13:17:13.961 | INFO      | Task run 'train_best_model-0' - [11]
validation-rmse:7.09348
13:17:14.264 | INFO      | Task run 'train_best_model-0' - [12]
validation-rmse:6.80836
13:17:14.622 | INFO      | Task run 'train_best_model-0' - [13]
validation-rmse:6.56211
13:17:14.918 | INFO      | Task run 'train_best_model-0' - [14]
validation-rmse:6.35332
13:17:15.152 | INFO      | Task run 'train_best_model-0' - [15]
validation-rmse:6.17838
13:17:15.434 | INFO      | Task run 'train_best_model-0' - [16]
validation-rmse:6.02763
13:17:16.045 | INFO      | Task run 'train_best_model-0' - [17]
validation-rmse:5.90232

```

```

13:17:16.262 | INFO      | Task run 'train_best_model-0' - [18]
validation-rmse:5.79691
13:17:16.540 | INFO      | Task run 'train_best_model-0' - [19]
validation-rmse:5.70817
13:17:16.818 | INFO      | Task run 'train_best_model-0' - [20]
validation-rmse:5.63182
13:17:17.026 | INFO      | Task run 'train_best_model-0' - [21]
validation-rmse:5.56852
13:17:17.240 | INFO      | Task run 'train_best_model-0' - [22]
validation-rmse:5.51520
13:17:17.448 | INFO      | Task run 'train_best_model-0' - [23]
validation-rmse:5.47115
13:17:17.726 | INFO      | Task run 'train_best_model-0' - [24]
validation-rmse:5.43421
13:17:17.926 | INFO      | Task run 'train_best_model-0' - [25]
validation-rmse:5.40206
13:17:18.132 | INFO      | Task run 'train_best_model-0' - [26]
validation-rmse:5.37653
13:17:18.339 | INFO      | Task run 'train_best_model-0' - [27]
validation-rmse:5.35308
13:17:18.538 | INFO      | Task run 'train_best_model-0' - [28]
validation-rmse:5.33437
13:17:18.748 | INFO      | Task run 'train_best_model-0' - [29]
validation-rmse:5.31842
13:17:18.946 | INFO      | Task run 'train_best_model-0' - [30]
validation-rmse:5.30434
13:17:19.141 | INFO      | Task run 'train_best_model-0' - [31]
validation-rmse:5.29293
13:17:19.338 | INFO      | Task run 'train_best_model-0' - [32]
validation-rmse:5.28270
13:17:19.532 | INFO      | Task run 'train_best_model-0' - [33]
validation-rmse:5.27307
13:17:19.653 | INFO      | Task run 'train_best_model-0' - [34]
validation-rmse:5.26638
13:17:19.840 | INFO      | Task run 'train_best_model-0' - [35]
validation-rmse:5.26043
13:17:19.964 | INFO      | Task run 'train_best_model-0' - [36]
validation-rmse:5.25512
13:17:20.140 | INFO      | Task run 'train_best_model-0' - [37]
validation-rmse:5.25088
13:17:20.260 | INFO      | Task run 'train_best_model-0' - [38]
validation-rmse:5.24664
13:17:20.443 | INFO      | Task run 'train_best_model-0' - [39]
validation-rmse:5.24209
13:17:20.563 | INFO      | Task run 'train_best_model-0' - [40]
validation-rmse:5.23909
13:17:20.744 | INFO      | Task run 'train_best_model-0' - [41]
validation-rmse:5.23607

```

13:17:20.864 | INFO | Task run 'train\_best\_model-0' - [42]  
validation-rmse:5.23344  
13:17:21.044 | INFO | Task run 'train\_best\_model-0' - [43]  
validation-rmse:5.23200  
13:17:21.536 | INFO | Task run 'train\_best\_model-0' - [44]  
validation-rmse:5.23097  
13:17:21.662 | INFO | Task run 'train\_best\_model-0' - [45]  
validation-rmse:5.22865  
13:17:21.827 | INFO | Task run 'train\_best\_model-0' - [46]  
validation-rmse:5.22775  
13:17:21.938 | INFO | Task run 'train\_best\_model-0' - [47]  
validation-rmse:5.22687  
13:17:22.056 | INFO | Task run 'train\_best\_model-0' - [48]  
validation-rmse:5.22582  
13:17:22.221 | INFO | Task run 'train\_best\_model-0' - [49]  
validation-rmse:5.22487  
13:17:22.335 | INFO | Task run 'train\_best\_model-0' - [50]  
validation-rmse:5.22442  
13:17:22.455 | INFO | Task run 'train\_best\_model-0' - [51]  
validation-rmse:5.22332  
13:17:22.622 | INFO | Task run 'train\_best\_model-0' - [52]  
validation-rmse:5.22236  
13:17:22.744 | INFO | Task run 'train\_best\_model-0' - [53]  
validation-rmse:5.22142  
13:17:22.861 | INFO | Task run 'train\_best\_model-0' - [54]  
validation-rmse:5.22022  
13:17:23.029 | INFO | Task run 'train\_best\_model-0' - [55]  
validation-rmse:5.21987  
13:17:23.144 | INFO | Task run 'train\_best\_model-0' - [56]  
validation-rmse:5.21908  
13:17:23.257 | INFO | Task run 'train\_best\_model-0' - [57]  
validation-rmse:5.21873  
13:17:23.427 | INFO | Task run 'train\_best\_model-0' - [58]  
validation-rmse:5.21820  
13:17:23.544 | INFO | Task run 'train\_best\_model-0' - [59]  
validation-rmse:5.21787  
13:17:23.660 | INFO | Task run 'train\_best\_model-0' - [60]  
validation-rmse:5.21702  
13:17:23.827 | INFO | Task run 'train\_best\_model-0' - [61]  
validation-rmse:5.21632  
13:17:23.941 | INFO | Task run 'train\_best\_model-0' - [62]  
validation-rmse:5.21571  
13:17:24.060 | INFO | Task run 'train\_best\_model-0' - [63]  
validation-rmse:5.21404  
13:17:24.228 | INFO | Task run 'train\_best\_model-0' - [64]  
validation-rmse:5.21348  
13:17:24.346 | INFO | Task run 'train\_best\_model-0' - [65]  
validation-rmse:5.21290

```

13:17:24.522 | INFO      | Task run 'train_best_model-0' - [66]
validation-rmse:5.21262
13:17:24.636 | INFO      | Task run 'train_best_model-0' - [67]
validation-rmse:5.21181
13:17:24.823 | INFO      | Task run 'train_best_model-0' - [68]
validation-rmse:5.21179
13:17:24.941 | INFO      | Task run 'train_best_model-0' - [69]
validation-rmse:5.21113
13:17:25.060 | INFO      | Task run 'train_best_model-0' - [70]
validation-rmse:5.21051
13:17:25.225 | INFO      | Task run 'train_best_model-0' - [71]
validation-rmse:5.21000
13:17:25.343 | INFO      | Task run 'train_best_model-0' - [72]
validation-rmse:5.20944
13:17:25.457 | INFO      | Task run 'train_best_model-0' - [73]
validation-rmse:5.20928
13:17:25.622 | INFO      | Task run 'train_best_model-0' - [74]
validation-rmse:5.20890
13:17:25.744 | INFO      | Task run 'train_best_model-0' - [75]
validation-rmse:5.20840
13:17:25.865 | INFO      | Task run 'train_best_model-0' - [76]
validation-rmse:5.20723
13:17:26.026 | INFO      | Task run 'train_best_model-0' - [77]
validation-rmse:5.20677
13:17:26.145 | INFO      | Task run 'train_best_model-0' - [78]
validation-rmse:5.20653
13:17:26.261 | INFO      | Task run 'train_best_model-0' - [79]
validation-rmse:5.20608
13:17:26.431 | INFO      | Task run 'train_best_model-0' - [80]
validation-rmse:5.20599
13:17:26.555 | INFO      | Task run 'train_best_model-0' - [81]
validation-rmse:5.20558
13:17:27.060 | INFO      | Task run 'train_best_model-0' - [82]
validation-rmse:5.20516
13:17:27.232 | INFO      | Task run 'train_best_model-0' - [83]
validation-rmse:5.20428
13:17:27.347 | INFO      | Task run 'train_best_model-0' - [84]
validation-rmse:5.20377
13:17:27.463 | INFO      | Task run 'train_best_model-0' - [85]
validation-rmse:5.20323
13:17:27.633 | INFO      | Task run 'train_best_model-0' - [86]
validation-rmse:5.20281
13:17:27.751 | INFO      | Task run 'train_best_model-0' - [87]
validation-rmse:5.20294
13:17:27.863 | INFO      | Task run 'train_best_model-0' - [88]
validation-rmse:5.20271
13:17:28.025 | INFO      | Task run 'train_best_model-0' - [89]
validation-rmse:5.20245

```



```
[8]: !prefect project init
```

```
[ ]: # create mlflow-prefect-pool via UI (docker logs prefect_mlops -> click on ↵
      ↵link) as this command doesn't work:
# prefect work-pool create -t process -p mlflow-prefect-pool
# under File -> new -> Terminal open new terminal to execute this code (in ↵
      ↵VSCode terminal user and urls to prefect and mlflow containers will be ↵
      ↵different)
# prefect agent start -p mlops-prefect-pool -q default
# under File -> new -> Terminal open another terminal to execute this code (as ↵
      ↵prefect agent start hangs up terminal indefinitely)
# prefect worker start -p mlops-prefect-pool
# mlops-prefect-pool status should have changed to "green" in the UI
```

```
| _/ / _|| _|| _| ( _ | | / _ \ ( | _|| . ^ | | |
|_| | _\ _ _| | _ _\ _ _| | | / _ \ _ _| _ _| _\ | | |
```

Agent started! Looking for work from work pool 'mlops-prefect-pool'...

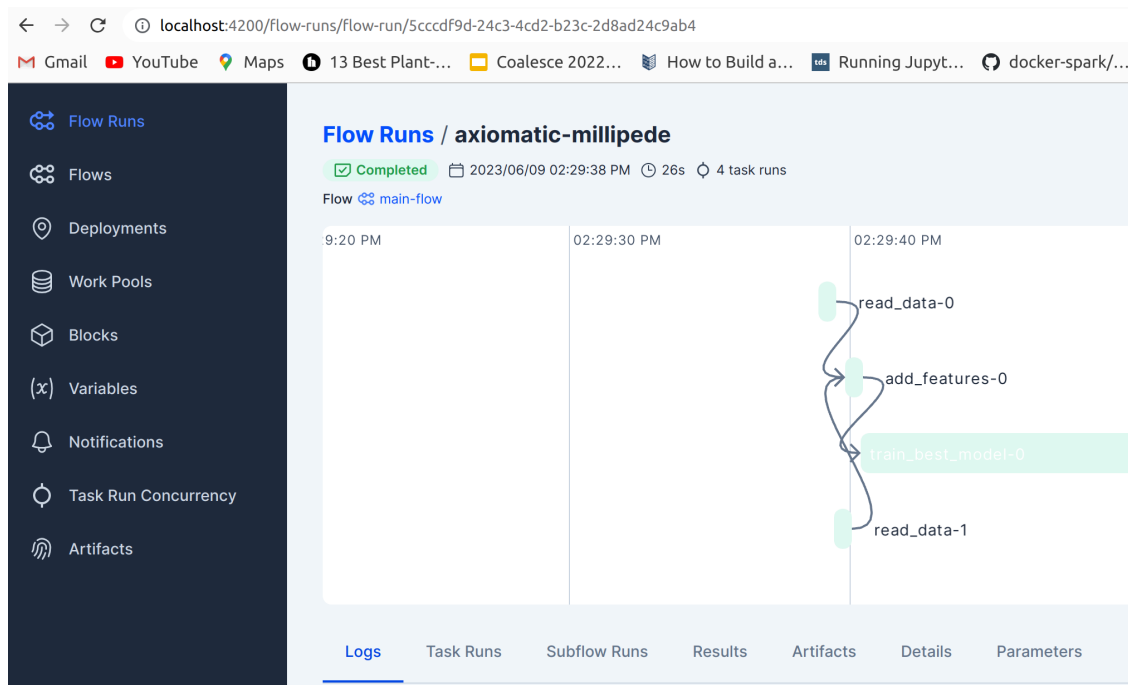
```
[21]: # under File -> new -> Terminal open another terminal to execute this code or
      ↪ try it with ! in jupyter (seems to work)
      !prefect deploy ./code/orchestrate.py:main_flow_local -n taxi_local -p
      ↪ mlops-prefect-pool
      # then start a worker in the new terminal window and run:
      # prefect deployment run main-flow-local/taxi_local
      # NOTICE THAT main_flow_local/taxi_local HAS CHANGED TO main-flow-local/
      ↪ taxi_local
      # USE this command to see what deployments are actually available and their
      ↪ exact names!!!
      # prefect deployment ls
```

Deployment 'main-flow-local/taxi\_local' successfully created with id '67838ea5-574d-463f-8de2-6ec8683deae'.

View Deployment in UI:

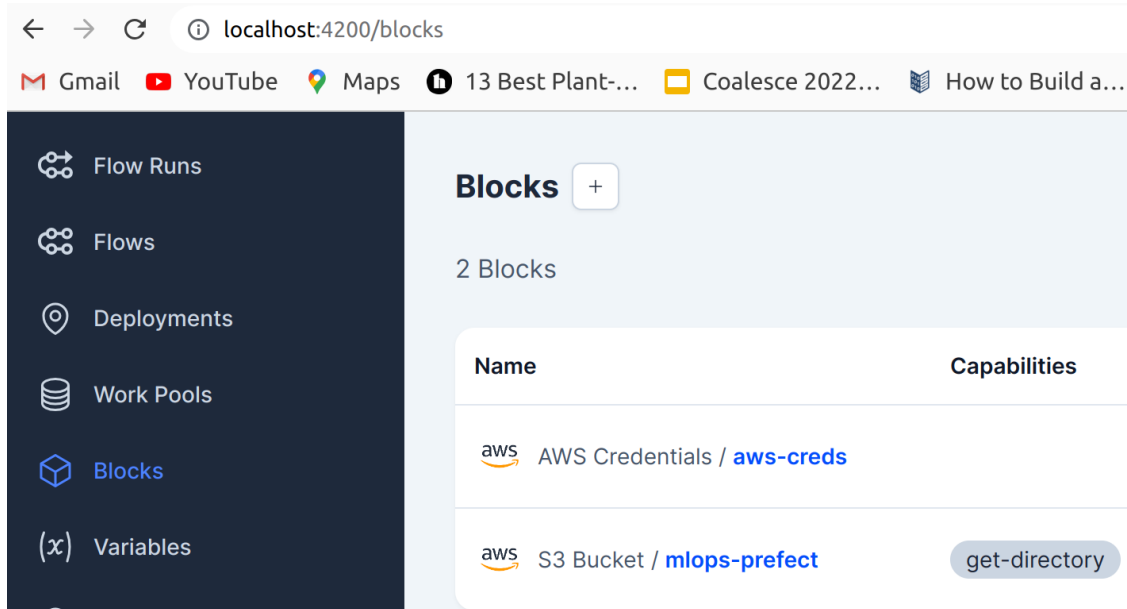
[http://prefect\\_mlops:4200/deployments/deployment/67838ea5-574d-463f-8de2-6ec8683deae](http://prefect_mlops:4200/deployments/deployment/67838ea5-574d-463f-8de2-6ec8683deae)

To execute flow runs from this deployment, start a worker that pulls work from the 'mlops-prefect-pool' work pool



#q3: rmse 5.19931

```
[11]: # run this to add aws creds..  
!python3 ./code/s3_block.py
```



```
[8]: # upload .parquet files to s3 bucket before deleting them locally  
#from utils import upload_to_s3  
#import os  
  
#files = os.listdir("data")  
#for file in files:  
#    if file.endswith(".parquet"):  
#        upload_to_s3(f"data/{file}", "mlops-prefect")
```

```
[22]: # under File -> new -> Terminal open another terminal to execute this code or  
      ↪ try it with ! in jupyter (seems to work)  
!prefect deploy ./code/orchestrate_s3.py:main_flow_s3 -n taxi_s3 -p  
      ↪ mlops-prefect-pool  
# then start a worker in the new terminal window and run:  
# prefect deployment run main-flow-s3/taxi_s3  
# NOTICE THAT main_flow_s3/taxi_s3 HAS CHANGED TO main-flow-s3/taxi_s3  
# USE this command to see what deployments are actually available and their  
      ↪ exact names!!!  
# prefect deployment ls
```

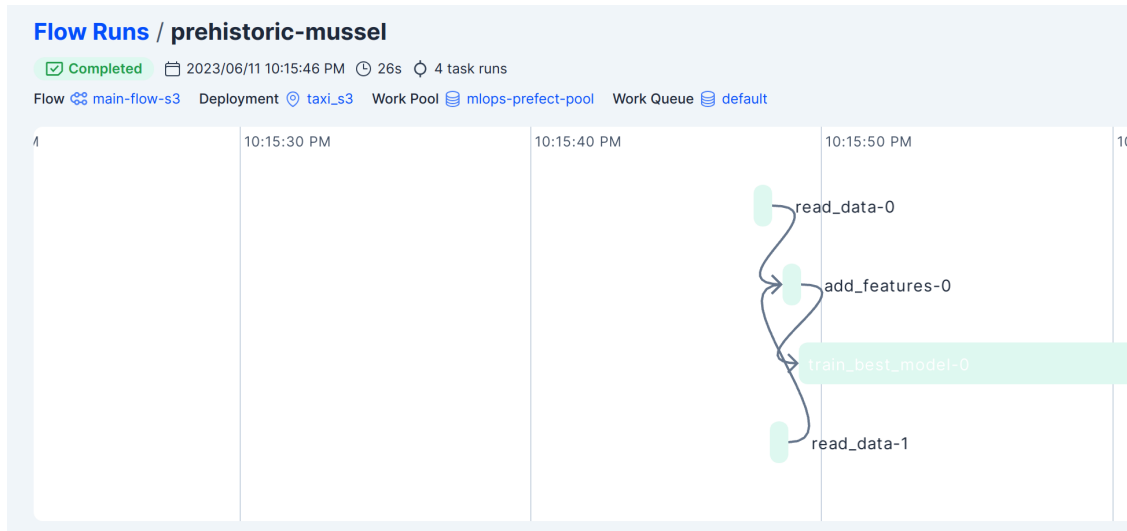
Deployment 'main-flow-s3/taxi\_s3' successfully created with id  
'ab3ab912-2df9-4f20-85e5-99ed991647a0'.

View Deployment in UI:

[http://prefect\\_mlops:4200/deployments/deployment/ab3ab912-2df9-4f20-85e5-99ed991](http://prefect_mlops:4200/deployments/deployment/ab3ab912-2df9-4f20-85e5-99ed991)

647a0

To execute flow runs from this deployment, start a worker that pulls work from the 'mlops-prefect-pool' work pool



```
[23]: # before trying this with git clone - test run with regular prefect.yaml:
!prefect deploy ./code/orchestrate_params.py:main_flow_params -n taxi_params -p
      ↪ mlops-prefect-pool
# then in the UI create new custom run and set paths to data
# "./code/data/green_tripdata_2023-01.parquet"
# "./code/data/green_tripdata_2023-02.parquet"
```

Deployment 'main-flow-params/taxi\_params' successfully created with id '69e0cfe3-3d5d-4d7d-ad1d-7ce6ec03ac70'.

View Deployment in UI:

[http://prefect\\_mlops:4200/deployments/deployment/69e0cfe3-3d5d-4d7d-ad1d-7ce6ec03ac70](http://prefect_mlops:4200/deployments/deployment/69e0cfe3-3d5d-4d7d-ad1d-7ce6ec03ac70)

To execute flow runs from this deployment, start a worker that pulls work from the 'mlops-prefect-pool' work pool

Flow Runs / zeta4-harnen

Completed

2023/06/11 10:24:54 PM

25s

4 task runs

Flow [main-flow-params](#) Deployment [taxi\\_params](#) Work Pool [mlops-prefect-pool](#) Work Queue [default](#)

10:24:40 PM

10:24:50 PM

10:25:00 PM

read\_data-0

add\_features-0

read\_data-1

Logs

Task Runs

Subflow Runs

Results

Artifacts

Details

Parameters

{

"train\_path": "./code/data/green\_tripdata\_2023-01.parquet",

"val\_path": "./code/data/green\_tripdata\_2023-02.parquet"

}

[24]: !prefect deployment ls

Deployments	
Name	ID
<a href="#">main-flow-local/taxi_local</a>	<a href="#">67838ea5-574d-463f-8de2-6ec8683deeeae</a>
<a href="#">main-flow-params/taxi_params</a>	<a href="#">69e0cfe3-3d5d-4d7d-ad1d-7ce6ec03ac70</a>
<a href="#">main-flow-s3/taxi_s3</a>	<a href="#">ab3ab912-2df9-4f20-85e5-99ed991647a0</a>

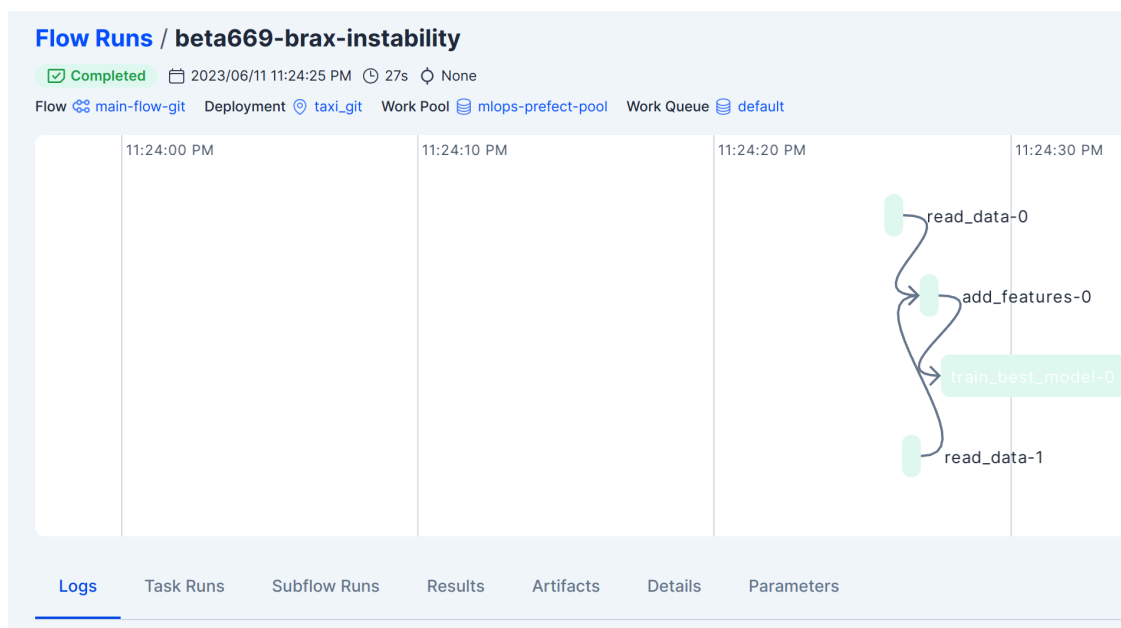
```
[43]: # then rename prefect.yaml to prefect_nogit.yaml and prefect_git.yaml into
      ↪prefect.yaml to run with pull settings
      # repeat the deployment under taxi_git name as deployment.yaml settings are
      ↪different for it - and copy file orchestrate_git.py from ./code into .
      # otherwise this deployment if starts with ./code or code will always double
      ↪the code/code in the path - it seems like prefect's bug:
      !prefect deploy --name taxi_git orchestrate_git.py:main_flow_git
      --params "{\"train_path\": \"./data/green_tripdata_2023-01.parquet\",
      ↪\"val_path\": \"./data/green_tripdata_2023-02.parquet\"}" --pool
      ↪mlops-prefect-pool
      # and after running this command - DELETE the ./code directory (better to
      ↪rename just in case) otherwise git clone will fail!!!
      # then in the UI create new custom run and set paths to data
      # "./data/green_tripdata_2023-01.parquet"
      # "./data/green_tripdata_2023-02.parquet"
```

Deployment 'main-flow-git/taxi\_git' successfully created with id '153566a3-7e04-4f55-974d-acebc55703c0'.

View Deployment in UI:

[http://prefect\\_mlops:4200/deployments/deployment/153566a3-7e04-4f55-974d-acebc55703c0](http://prefect_mlops:4200/deployments/deployment/153566a3-7e04-4f55-974d-acebc55703c0)

To execute flow runs from this deployment, start a worker that pulls work from the 'mlops-prefect-pool' work pool



```
[26]: # in UI create a custom run for taxi_s3 with params:
      #{
      # "train_path": "./data/green_tripdata_2023-02.parquet",
```

```
# "val_path": "./data/green_tripdata_2023-03.parquet"  
#}  
# to get the artifacts for Feb/Mar combo RMSE = 5.37
```

## Artifacts / duration-model-report

📁 Latest

11:32 PM  
Jun 11th, 2023

**95be0f32**

Flow run  gamma3-tolfalas  
Task run  train\_best\_model-0

### RMSE Report

## Summary

Duration Prediction

## RMSE XGBoost Model

Region	RMSE
:-----	-----:
2023-06-11	5.37

10:16 PM  
Jun 11th, 2023

**6a5f1f34**

Flow run  prehistoric-mussel  
Task run  train\_best\_model-0

[44]: !prefect project recipe ls

Available project recipes



Name	Description
<code>docker</code>	Store project within a custom docker image alongside its runtime environment
<code>docker-gcs</code>	Store project within GCS and build a custom docker image for runtime
<code>docker-git</code>	Store project within a git repository and build a custom docker image for runtime
<code>docker-s3</code>	Store project within S3 and build a custom docker image for runtime
<code>gcs</code>	Store project within a GCS bucket
<code>git</code>	Store project within git repository
<code>local</code>	Store project on a local filesystem
<code>s3</code>	Store project within an S3 bucket

Run ``prefect project init --recipe <recipe>`` to initialize a project with a `<recipe>`.

```
[45]: !pip install prefect_email
```

```
Collecting prefect_email
  Downloading prefect_email-0.2.2-py3-none-any.whl (11 kB)
```

Requirement already satisfied: prefect>=2.0.0 in /opt/conda/lib/python3.10/site-packages (from prefect\_email) (2.10.8)

Requirement already satisfied: websockets>=10.4 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (11.0.3)

Requirement already satisfied: fsspec>=2022.5.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (2023.5.0)

Requirement already satisfied: cloudpickle>=2.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (2.2.1)

Requirement already satisfied: jsonpatch>=1.32 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.32)

Requirement already satisfied: pytz>=2021.1 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (2023.3)

Requirement already satisfied: coolname>=1.0.4 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (2.2.0)

Requirement already satisfied: asyncpg>=0.23 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (0.27.0)

Requirement already satisfied: uvicorn>=0.14.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (0.22.0)

Requirement already satisfied: apprise>=1.1.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.4.0)

Requirement already satisfied: croniter>=1.0.12 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.3.15)

Requirement already satisfied: typing-extensions>=4.1.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (4.6.2)

Requirement already satisfied: dateparser>=1.1.1 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.1.8)

Requirement already satisfied: asgi-lifespan>=1.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (2.1.0)

Requirement already satisfied: alembic>=1.7.5 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.11.1)

Requirement already satisfied: aiosqlite>=0.17.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (0.19.0)

Requirement already satisfied: jsonschema<5.0.0,>=3.2.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (4.17.3)

Requirement already satisfied: python-slugify>=5.0 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)

(8.0.1)  
Requirement already satisfied: kubernetes>=24.2.0 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(26.1.0)  
Requirement already satisfied: typer>=0.4.2 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (0.9.0)  
Requirement already satisfied: docker>=4.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (6.1.0)  
Requirement already satisfied: packaging>=21.3 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(23.1)  
Requirement already satisfied: pydantic>=1.10.0 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(1.10.9)  
Requirement already satisfied: toml>=0.10.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (0.10.2)  
Requirement already satisfied: httpx[http2]!=0.23.2,>=0.23 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(0.24.1)  
Requirement already satisfied: click<8.2,>=8.0 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(8.1.3)  
Requirement already satisfied: pyyaml>=5.4.1 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (6.0)  
Requirement already satisfied: readchar>=4.0.0 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(4.0.5.dev0)  
Requirement already satisfied: griffe>=0.20.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (0.21.0)  
Requirement already satisfied: jinja2>=3.0.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (3.1.2)  
Requirement already satisfied: rich>=11.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (13.4.1)  
Requirement already satisfied: pathspec>=0.8.0 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(0.11.1)  
Requirement already satisfied: cryptography>=36.0.1 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(40.0.2)  
Requirement already satisfied: orjson>=3.7 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (3.8.1)  
Requirement already satisfied: fastapi>=0.93 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (0.96.1)  
Requirement already satisfied: pendulum>=2.1.2 in  
/opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email)  
(2.1.2)  
Requirement already satisfied: anyio>=3.4.0 in /opt/conda/lib/python3.10/site-  
packages (from prefect>=2.0.0->prefect\_email) (3.6.2)

Requirement already satisfied: sqlalchemy[asyncio]!=1.4.33,<2.0,>=1.4.22 in /opt/conda/lib/python3.10/site-packages (from prefect>=2.0.0->prefect\_email) (1.4.46)

Requirement already satisfied: Mako in /opt/conda/lib/python3.10/site-packages (from alembic>=1.7.5->prefect>=2.0.0->prefect\_email) (1.2.4)

Requirement already satisfied: idna>=2.8 in /opt/conda/lib/python3.10/site-packages (from anyio>=3.4.0->prefect>=2.0.0->prefect\_email) (3.4)

Requirement already satisfied: sniffio>=1.1 in /opt/conda/lib/python3.10/site-packages (from anyio>=3.4.0->prefect>=2.0.0->prefect\_email) (1.3.0)

Requirement already satisfied: markdown in /opt/conda/lib/python3.10/site-packages (from apprise>=1.1.0->prefect>=2.0.0->prefect\_email) (3.4.3)

Requirement already satisfied: requests in /opt/conda/lib/python3.10/site-packages (from apprise>=1.1.0->prefect>=2.0.0->prefect\_email) (2.31.0)

Requirement already satisfied: requests-oauthlib in /opt/conda/lib/python3.10/site-packages (from apprise>=1.1.0->prefect>=2.0.0->prefect\_email) (1.3.1)

Requirement already satisfied: certifi in /opt/conda/lib/python3.10/site-packages (from apprise>=1.1.0->prefect>=2.0.0->prefect\_email) (2023.5.7)

Requirement already satisfied: python-dateutil in /opt/conda/lib/python3.10/site-packages (from croniter>=1.0.12->prefect>=2.0.0->prefect\_email) (2.8.2)

Requirement already satisfied: cffi>=1.12 in /opt/conda/lib/python3.10/site-packages (from cryptography>=36.0.1->prefect>=2.0.0->prefect\_email) (1.15.1)

Requirement already satisfied: regex!=2019.02.19,!=2021.8.27 in /opt/conda/lib/python3.10/site-packages (from dateparser>=1.1.1->prefect>=2.0.0->prefect\_email) (2023.6.3)

Requirement already satisfied: tzlocal in /opt/conda/lib/python3.10/site-packages (from dateparser>=1.1.1->prefect>=2.0.0->prefect\_email) (5.0.1)

Requirement already satisfied: websocket-client>=0.32.0 in /opt/conda/lib/python3.10/site-packages (from docker>=4.0->prefect>=2.0.0->prefect\_email) (1.5.2)

Requirement already satisfied: urllib3>=1.26.0 in /opt/conda/lib/python3.10/site-packages (from docker>=4.0->prefect>=2.0.0->prefect\_email) (1.26.15)

Requirement already satisfied: starlette<0.28.0,>=0.27.0 in /opt/conda/lib/python3.10/site-packages (from fastapi>=0.93->prefect>=2.0.0->prefect\_email) (0.27.0)

Requirement already satisfied: httpcore<0.18.0,>=0.15.0 in /opt/conda/lib/python3.10/site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect>=2.0.0->prefect\_email) (0.17.2)

Requirement already satisfied: h2<5,>=3 in /opt/conda/lib/python3.10/site-packages (from httpx[http2]!=0.23.2,>=0.23->prefect>=2.0.0->prefect\_email) (4.1.0)

Requirement already satisfied: MarkupSafe>=2.0 in /opt/conda/lib/python3.10/site-packages (from jinja2>=3.0.0->prefect>=2.0.0->prefect\_email) (2.1.2)

Requirement already satisfied: jsonpointer>=1.9 in /opt/conda/lib/python3.10/site-packages (from

```

jsonpach>=1.32->prefect>=2.0.0->prefect_email) (2.0)
Requirement already satisfied: pyrsistent!=0.17.0,!=0.17.1,!=0.17.2,>=0.14.0 in
/opt/conda/lib/python3.10/site-packages (from
jsonschema<5.0.0,>=3.2.0->prefect>=2.0.0->prefect_email) (0.19.3)
Requirement already satisfied: attrs>=17.4.0 in /opt/conda/lib/python3.10/site-
packages (from jsonschema<5.0.0,>=3.2.0->prefect>=2.0.0->prefect_email) (23.1.0)
Requirement already satisfied: google-auth>=1.0.1 in
/opt/conda/lib/python3.10/site-packages (from
kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (2.19.1)
Requirement already satisfied: setuptools>=21.0.0 in
/opt/conda/lib/python3.10/site-packages (from
kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (67.8.0)
Requirement already satisfied: six>=1.9.0 in /opt/conda/lib/python3.10/site-
packages (from kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (1.16.0)
Requirement already satisfied: pytzdata>=2020.1 in
/opt/conda/lib/python3.10/site-packages (from
pendulum>=2.1.2->prefect>=2.0.0->prefect_email) (2020.1)
Requirement already satisfied: text-unidecode>=1.3 in
/opt/conda/lib/python3.10/site-packages (from python-
slugify>=5.0->prefect>=2.0.0->prefect_email) (1.3)
Requirement already satisfied: markdown-it-py<3.0.0,>=2.2.0 in
/opt/conda/lib/python3.10/site-packages (from
rich>=11.0->prefect>=2.0.0->prefect_email) (2.2.0)
Requirement already satisfied: pygments<3.0.0,>=2.13.0 in
/opt/conda/lib/python3.10/site-packages (from
rich>=11.0->prefect>=2.0.0->prefect_email) (2.15.1)
Requirement already satisfied: greenlet!=0.4.17 in
/opt/conda/lib/python3.10/site-packages (from
sqlalchemy[asyncio]!=1.4.33,<2.0,>=1.4.22->prefect>=2.0.0->prefect_email)
(2.0.2)
Requirement already satisfied: h11>=0.8 in /opt/conda/lib/python3.10/site-
packages (from uvicorn>=0.14.0->prefect>=2.0.0->prefect_email) (0.14.0)
Requirement already satisfied: pycparser in /opt/conda/lib/python3.10/site-
packages (from cffi>=1.12->cryptography>=36.0.1->prefect>=2.0.0->prefect_email)
(2.21)
Requirement already satisfied: pyasn1-modules>=0.2.1 in
/opt/conda/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (0.2.7)
Requirement already satisfied: rsa<5,>=3.1.4 in /opt/conda/lib/python3.10/site-
packages (from google-
auth>=1.0.1->kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (4.9)
Requirement already satisfied: cachetools<6.0,>=2.0.0 in
/opt/conda/lib/python3.10/site-packages (from google-
auth>=1.0.1->kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (5.3.0)
Requirement already satisfied: hpach<5,>=4.0 in /opt/conda/lib/python3.10/site-
packages (from
h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect>=2.0.0->prefect_email) (4.0.0)
Requirement already satisfied: hyperframe<7,>=6.0 in

```

```

/opt/conda/lib/python3.10/site-packages (from
h2<5,>=3->httpx[http2]!=0.23.2,>=0.23->prefect>=2.0.0->prefect_email) (6.0.1)
Requirement already satisfied: mdurl~=0.1 in /opt/conda/lib/python3.10/site-
packages (from markdown-it-
py<3.0.0,>=2.2.0->rich>=11.0->prefect>=2.0.0->prefect_email) (0.1.0)
Requirement already satisfied: charset-normalizer<4,>=2 in
/opt/conda/lib/python3.10/site-packages (from
requests->apprise>=1.1.0->prefect>=2.0.0->prefect_email) (3.1.0)
Requirement already satisfied: oauthlib>=3.0.0 in
/opt/conda/lib/python3.10/site-packages (from requests-
oauthlib->apprise>=1.1.0->prefect>=2.0.0->prefect_email) (3.2.2)
Requirement already satisfied: pyasn1<0.5.0,>=0.4.6 in
/opt/conda/lib/python3.10/site-packages (from pyasn1-modules>=0.2.1->google-
auth>=1.0.1->kubernetes>=24.2.0->prefect>=2.0.0->prefect_email) (0.4.8)
Installing collected packages: prefect-email
Successfully installed prefect-email-0.2.2
WARNING: You are using pip version 21.3.1; however, version 23.1.2 is
available.

You should consider upgrading via the '/opt/conda/bin/python3.10 -m pip install
--upgrade pip' command.

```

```
[46]: !prefect block register -m prefect_email
```

Successfully registered 1 block

### Registered Blocks

#### Email Server Credentials

To configure the newly registered blocks, go to the Blocks page in the Prefect UI: [http://prefect\\_mlops:4200/blocks/catalog](http://prefect_mlops:4200/blocks/catalog)

```
[11]: # switch to cloud to use email notifications and creds
# clear config setting to avoid a bug in set
!prefect config unset PREFECT_API_URL

# use Prefect Cloud
!prefect config set PREFECT_API_URL=$PREFECT_API_URL
!prefect config set PREFECT_API_KEY=$PREFECT_API_KEY
```

```
[10]: #q5 answer: email_send_message
# we switched configs to the server, now need to modify prefect.yaml to
  ↪prefect_git.yaml and prefect_nogit.yaml back to prefect.yaml (very annoying)
```

```

# mlops-prefect-pool should be created in the cloud, then start agent and
  ↳ worker in new terminal windows:
# prefect agent start -p mlops-prefect-pool -q default
# prefect worker start -p mlops-prefect-pool
# under File -> new -> Terminal open another terminal to execute this code or
  ↳ try it with ! in jupyter (seems to work)
!prefect deploy ./code/orchestrate_email.py:main_flow_email -n taxi_email -p
  ↳ mlops-prefect-pool
# then start a worker in the new terminal window and run or use custom run from
  ↳ the cloud UI:
# prefect deployment run main-flow-email/taxi_email
# NOTICE THAT main_flow_email/taxi_email HAS CHANGED TO main-flow-email/
  ↳ taxi_email
# USE this command to see what deployments are actually available and their
  ↳ exact names!!!
# prefect deployment ls

```

This deployment references a work pool that does not exist. This means no worker will be able to pick up its runs. You can create a work pool in the Prefect UI.

Deployment 'main-flow-email/taxi\_email' successfully created with id '36d645ef-b157-4f88-8023-f51b1e60d570'.

View Deployment in UI:

<https://app.prefect.cloud/account/e9796d3f-7997-45df-9f2a-e9d402122ef8/workspace/a9b7dc58-eede-4f3d-99f7-c18595aa28cc/deployments/deployment/36d645ef-b157-4f88-8023-f51b1e60d570>

This deployment does not specify a work pool or queue, which means no worker will be able to pick up its runs. To add a work pool, edit the deployment spec and re-run this command, or visit the deployment in the UI.

[1]: #q6 answer is: Actions

## Automations / Create

01

Trigger

02

Actions

Trigger Type

Select trigger

## Flows / main-flow-email


Runs


Deployments


1 Flow run

**main-flow-email** > gamma845-lebbenin

 Late

 Scheduled for 2023/06/12 11:20:32 AM (4m 41s late)

 None

Deployment  taxi\_email

[ ]: