data-pipelining-with-polygon

August 10, 2025

1 Data Pipelining With Polygon

1.1 Python Imports

[1]:
```
# Standard Library
import datetime
import io
import os
import random
import sys
import warnings

from datetime import datetime, timedelta
from pathlib import Path

# Data Handling
import numpy as np
import pandas as pd

# Data Visualization
import matplotlib.dates as mdates
import matplotlib.pyplot as plt
import matplotlib.ticker as mtick
import seaborn as sns
from matplotlib.ticker import FormatStrFormatter, FuncFormatter, MultipleLocator

# Data Sources
import yfinance as yf

# Statistical Analysis
import statsmodels.api as sm

# Machine Learning
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler

# Suppress warnings
warnings.filterwarnings("ignore")
```

1.2 Add Directories To Path

[2]:
```
# Add the source subdirectory to the system path to allow import config from␣
 ↪settings.py
current_directory = Path(os.getcwd())
website_base_directory = current_directory.parent.parent.parent
src_directory = website_base_directory / "src"
sys.path.append(str(src_directory)) if str(src_directory) not in sys.path else␣
 ↪None

# Import settings.py
from settings import config

# Add configured directories from config to path
SOURCE_DIR = config("SOURCE_DIR")
sys.path.append(str(Path(SOURCE_DIR))) if str(Path(SOURCE_DIR)) not in sys.path␣
 ↪else None

# Add other configured directories
BASE_DIR = config("BASE_DIR")
CONTENT_DIR = config("CONTENT_DIR")
POSTS_DIR = config("POSTS_DIR")
PAGES_DIR = config("PAGES_DIR")
PUBLIC_DIR = config("PUBLIC_DIR")
SOURCE_DIR = config("SOURCE_DIR")
DATA_DIR = config("DATA_DIR")
DATA_MANUAL_DIR = config("DATA_MANUAL_DIR")

# Print system path
for i, path in enumerate(sys.path):
    print(f"{i}: {path}")
```

```
0: /usr/lib/python313.zip
1: /usr/lib/python3.13
2: /usr/lib/python3.13/lib-dynload
3:
4: /home/jared/python-virtual-envs/general_313/lib/python3.13/site-packages
5: /home/jared/Cloud_Storage/Dropbox/Websites/jaredszajkowski.github.io/src
```

1.3 Track Index Dependencies

[3]:
```
# Create file to track markdown dependencies
dep_file = Path("index_dep.txt")
dep_file.write_text("")
```

[3]: 0

1.4 Python Functions

[4]:
```
from export_track_md_deps import export_track_md_deps
from polygon_fetch_full_history import polygon_fetch_full_history
from polygon_pull_data import polygon_pull_data
```

1.5 Function Usage

1.5.1 Polygon Fetch Full History

[5]:
```
from load_api_keys import load_api_keys
from polygon import RESTClient

# Load API keys from the environment
api_keys = load_api_keys()

# Get the environment variable for where data is stored
DATA_DIR = config("DATA_DIR")

# Open client connection
client = RESTClient(api_key=api_keys["POLYGON_KEY"])

# Create an empty DataFrame
df = pd.DataFrame({
    'Date': pd.Series(dtype="datetime64[ns]"),
    'open': pd.Series(dtype="float64"),
    'high': pd.Series(dtype="float64"),
    'low': pd.Series(dtype="float64"),
    'close': pd.Series(dtype="float64"),
    'volume': pd.Series(dtype="float64"),
    'vwap': pd.Series(dtype="float64"),
    'transactions': pd.Series(dtype="int64"),
    'otc': pd.Series(dtype="object")
})

# Example usage - minute
df = polygon_fetch_full_history(
    client=client,
    ticker="AMZN",
    timespan="day",
    multiplier=1,
    adjusted=True,
    full_history_df=df,
    current_start=datetime(2025, 1, 1),
    free_tier=True,
)
```

```
Pulling day data for 2025-01-01 00:00:00 thru 2026-01-01 00:00:00 for AMZN…
```

```
New data:
                   Date     open     high       low   close      volume  \
0   2025-01-02 05:00:00  222.030  225.150  218.1900  220.22  33956579.0
1   2025-01-03 05:00:00  222.505  225.360  221.6200  224.19  27503606.0
2   2025-01-06 05:00:00  226.780  228.835  224.8400  227.61  31849831.0
3   2025-01-07 05:00:00  227.900  228.381  221.4600  222.11  28084164.0
4   2025-01-08 05:00:00  223.185  223.520  220.2000  222.13  25033292.0
..                  …      …        …         …     …         …
145 2025-08-04 04:00:00  217.400  217.440  211.4200  211.65  77890146.0
146 2025-08-05 04:00:00  213.050  216.300  212.8700  213.75  51505121.0
147 2025-08-06 04:00:00  214.695  222.650  213.7409  222.31  54823045.0
148 2025-08-07 04:00:00  221.000  226.220  220.8200  223.13  40603513.0
149 2025-08-08 04:00:00  223.140  223.800  221.8836  222.69  32970477.0

         vwap  transactions   otc
0    221.2745        449631  None
1    223.7046        346975  None
2    227.0921        410686  None
3    223.4033        379570  None
4    222.0414        325539  None
..        …             …   …
145  213.1312       1046525  None
146  214.5142        639055  None
147  219.4299        654274  None
148  223.1357        553279  None
149  222.6698        397504  None

[150 rows x 9 columns]
Combined data:
                   Date     open     high       low   close      volume  \
0   2025-01-02 05:00:00  222.030  225.150  218.1900  220.22  33956579.0
1   2025-01-03 05:00:00  222.505  225.360  221.6200  224.19  27503606.0
2   2025-01-06 05:00:00  226.780  228.835  224.8400  227.61  31849831.0
3   2025-01-07 05:00:00  227.900  228.381  221.4600  222.11  28084164.0
4   2025-01-08 05:00:00  223.185  223.520  220.2000  222.13  25033292.0
..                  …      …        …         …     …         …
145 2025-08-04 04:00:00  217.400  217.440  211.4200  211.65  77890146.0
146 2025-08-05 04:00:00  213.050  216.300  212.8700  213.75  51505121.0
147 2025-08-06 04:00:00  214.695  222.650  213.7409  222.31  54823045.0
148 2025-08-07 04:00:00  221.000  226.220  220.8200  223.13  40603513.0
149 2025-08-08 04:00:00  223.140  223.800  221.8836  222.69  32970477.0

         vwap  transactions   otc
0    221.2745        449631  None
1    223.7046        346975  None
2    227.0921        410686  None
3    223.4033        379570  None
4    222.0414        325539  None
```

```
..        …            …   …
145  213.1312       1046525  None
146  214.5142        639055  None
147  219.4299        654274  None
148  223.1357        553279  None
149  222.6698        397504  None

[150 rows x 9 columns]
Sleeping for 12 seconds to avoid hitting API rate limits…
```

[6]:
```
# Copy this <!-- INSERT_polygon_fetch_full_history_HERE --> to index_temp.md
export_track_md_deps(dep_file=dep_file, md_filename="polygon_fetch_full_history.
↪md", content=df.to_markdown(floatfmt=".5f"))
```

```
Exported and tracked: polygon_fetch_full_history.md
```

1.5.2 Polygon Pull Data

[7]:
```
current_year = datetime.now().year
current_month = datetime.now().month
current_day = datetime.now().day

# Example usage - daily
df = polygon_pull_data(
    base_directory=DATA_DIR,
    ticker="AMZN",
    source="Polygon",
    asset_class="Equities",
    start_date=datetime(current_year - 2, current_month, current_day),
    timespan="day",
    multiplier=1,
    adjusted=True,
    excel_export=True,
    pickle_export=True,
    output_confirmation=True,
)
```

```
File found…updating the AMZN day data.
Existing data:
                  Date     open    high       low   close      volume  \
0  2023-07-28 04:00:00  129.690  133.01  129.3300  132.21  46269781.0
1  2023-07-31 04:00:00  133.200  133.87  132.3800  133.68  41901516.0
0  2023-08-01 04:00:00  133.550  133.69  131.6199  131.69  42250989.0
1  2023-08-02 04:00:00  130.154  130.23  126.8200  128.21  50988614.0
2  2023-08-03 04:00:00  127.480  129.84  126.4100  128.91  90855736.0
..                 …      …      …         …      …         …
3  2025-08-04 04:00:00  217.400  217.44  211.4200  211.65  77890146.0
1  2025-08-05 04:00:00  213.050  216.30  212.8700  213.75  51505121.0
```

```
2  2025-08-06 04:00:00  214.695  222.65  213.7409  222.31  54823045.0
2  2025-08-07 04:00:00  221.000  226.22  220.8200  223.13  40603513.0
3  2025-08-08 04:00:00  223.140  223.80  221.8836  222.69  32970477.0

        vwap  transactions   otc
0   131.8837        413438  None
1   133.3410        406644  None
0   132.2470        385743  None
1   128.3973        532942  None
2   131.4941        746639  None
..       …            …   …
3   213.1312       1046525  None
1   214.5142        639055  None
2   219.4299        654274  None
2   223.1357        553279  None
3   222.6698        397504  None

[510 rows x 9 columns]
Last date in existing data: 2025-08-08 04:00:00
Pulling day data for 2025-08-07 04:00:00 thru 2026-08-07 04:00:00 for AMZN…

New data:
                 Date    open    high       low   close      volume      vwap  \
0 2025-08-07 04:00:00  221.00  226.22  220.8200  223.13  40603513.0  223.1357
1 2025-08-08 04:00:00  223.14  223.80  221.8836  222.69  32970477.0  222.6698

   transactions   otc
0        553279  None
1        397504  None
Combined data:
                  Date     open    high       low   close      volume  \
0  2023-07-28 04:00:00  129.690  133.01  129.3300  132.21  46269781.0
1  2023-07-31 04:00:00  133.200  133.87  132.3800  133.68  41901516.0
0  2023-08-01 04:00:00  133.550  133.69  131.6199  131.69  42250989.0
1  2023-08-02 04:00:00  130.154  130.23  126.8200  128.21  50988614.0
2  2023-08-03 04:00:00  127.480  129.84  126.4100  128.91  90855736.0
..                 …      …     …       …     …         …
3  2025-08-04 04:00:00  217.400  217.44  211.4200  211.65  77890146.0
1  2025-08-05 04:00:00  213.050  216.30  212.8700  213.75  51505121.0
2  2025-08-06 04:00:00  214.695  222.65  213.7409  222.31  54823045.0
2  2025-08-07 04:00:00  221.000  226.22  220.8200  223.13  40603513.0
3  2025-08-08 04:00:00  223.140  223.80  221.8836  222.69  32970477.0

        vwap  transactions   otc
0   131.8837        413438  None
1   133.3410        406644  None
0   132.2470        385743  None
1   128.3973        532942  None
```

```
2   131.4941        746639  None
..       …           …   …
3   213.1312       1046525  None
1   214.5142        639055  None
2   219.4299        654274  None
2   223.1357        553279  None
3   222.6698        397504  None

[510 rows x 9 columns]
Sleeping for 12 seconds to avoid hitting API rate limits…


Exporting AMZN day data to Excel…

Exporting AMZN day data to pickle…
The first and last date of data for AMZN is:

                 Date    open    high     low   close      volume      vwap  \
0 2023-07-28 04:00:00  129.69  133.01  129.33  132.21  46269781.0  131.8837

   transactions   otc
0        413438  None

                 Date    open   high       low   close      volume      vwap  \
3 2025-08-08 04:00:00  223.14  223.8  221.8836  222.69  32970477.0  222.6698

   transactions   otc
3        397504  None

Polygon data complete for AMZN
--------------------
```

[8]:
```
# Copy this <!-- INSERT_polygon_pull_data_HERE --> to index_temp.md
export_track_md_deps(dep_file=dep_file, md_filename="polygon_pull_data.md",␣
 ↪content=df.to_markdown(floatfmt=".5f"))
```

```
Exported and tracked: polygon_pull_data.md
```