

Chapter 4

April 21, 2025

0.1 Hands-On Data Preprocessing in Python

Learn how to effectively prepare data for successful data analytics

AUTHOR: Dr. Roy Jafari

0.1.1 Chapter 4: Databases

```
[1]: import sqlite3
import pandas as pd
```

```
[2]: Connection = sqlite3.connect('chinook.db')
query_txt = "SELECT * FROM customers;"

df_customers = pd.read_sql_query(query_txt,Connection)
```

```
[3]: import requests

stk_ticker = 'AMZN'
data_resolution = 'W'
timestamp_from = 1577865600
timestamp_to = 1609315200
API_Key = 'bsiqli7rh5rc8orbnkqg'
Address_template = 'https://finnhub.io/api/v1/stock/candle?
↳symbol={}&resolution={}&from={}&to={}&token={}';

API_address = Address_template.format(stk_ticker,data_resolution,
                                     timestamp_from,timestamp_to,API_Key)
r = requests.get(API_address)
print(r)
```

<Response [200]>

```
[4]: print(API_address)
```

```
https://finnhub.io/api/v1/stock/candle?symbol=AMZN&resolution=W&from=1577865600&
to=1609315200&token=bsiqli7rh5rc8orbnkqg
```

```
[5]: print(r.json())
```

```
{'c': [1891.97, 1862.02, 1887.46, 1858, 2039.87, 2160, 2170.22, 1979.59,
1975.83, 1820.86, 1830, 1885.84, 1907.7, 2043, 2307.68, 2363.49, 2372.71,
2351.26, 2367.92, 2497.94, 2410.39, 2478.4, 2647.45, 2640.98, 2734.4, 2878.7,
3081.11, 3008.87, 3099.91, 3033.53, 3205.03, 3162.24, 3260.48, 3441.85, 3531.45,
3268.61, 3078.1, 2999.86, 3148.73, 3195.69, 3363.71, 3184.94, 3162.78, 3241.16,
3137.39, 3105.46, 3185.07, 3104.2, 3240.96, 3185.27, 3285.85], 'h': [1913.89,
1917.82, 1902.5, 1894.99, 2071.02, 2185.95, 2185.1, 2176.79, 1996.33, 1960.72,
1857.775, 1957, 1993.02, 2044, 2333.369, 2460.9981, 2444.88, 2475, 2419.665,
2500.01, 2525.45, 2488, 2722.35, 2671.38, 2796, 2895, 3083.97, 3344.29,
3240.5758, 3098.2713, 3246.82, 3247.47, 3320, 3451.7387, 3552.25, 3488.41,
3349.89, 3133.99, 3212.88, 3224, 3496.24, 3399.6606, 3291.66, 3257.25, 3366.8,
3189.25, 3198, 3228.64, 3247, 3263.51, 3350.65], 'l': [1860, 1855.09, 1857.25,
1815.34, 1850.61, 2024.8001, 2124.105, 1958.42, 1811.129, 1761.2901, 1626.0318,
1812, 1889.29, 1889.15, 2017.66, 2279.66, 2306, 2256.38, 2337.805, 2353.21,
2330, 2378.23, 2437.13, 2503.35, 2636.11, 2630.08, 2871.1021, 2950, 2918.23,
2888, 3005, 3073, 3120, 3238, 3378, 3111.13, 3074.15, 2871, 2965, 3090, 3174.99,
3160, 3121.94, 2950.12, 3019.4828, 3072.685, 3065.46, 3088, 3072.82, 3166,
3169], 'o': [1875, 1909.89, 1882.99, 1885.11, 1858, 2041.02, 2144.99, 2173.07,
1934.38, 1933, 1721.98, 1860, 1902, 1901.64, 2044.3, 2346, 2399.98, 2419.84,
2374.78, 2361.01, 2500, 2384.33, 2477.43, 2603.5, 2647.01, 2739.55, 2912.01,
3115.99, 2971.06, 3098.2713, 3014, 3194.36, 3182.99, 3252, 3450.05, 3485,
3307.2188, 3009.25, 2977.79, 3208, 3224.99, 3292.01, 3189.87, 3201.27, 3319.97,
3159.95, 3105.31, 3205.46, 3088.99, 3250, 3193.9], 's': 'ok', 't': [1577923200,
1578528000, 1579132800, 1579737600, 1580342400, 1580947200, 1581552000,
1582156800, 1582761600, 1583366400, 1583971200, 1584576000, 1585180800,
1585785600, 1586390400, 1586995200, 1587600000, 1588204800, 1588809600,
1589414400, 1590019200, 1590624000, 1591228800, 1591833600, 1592438400,
1593043200, 1593648000, 1594252800, 1594857600, 1595462400, 1596067200,
1596672000, 1597276800, 1597881600, 1598486400, 1599091200, 1599696000,
1600300800, 1600905600, 1601510400, 1602115200, 1602720000, 1603324800,
1603929600, 1604534400, 1605139200, 1605744000, 1606348800, 1606953600, 1607558400,
1608163200, 1608768000], 'v': [19514188, 15160738, 13580875, 14688733, 37459327,
22414811, 11150195, 25788575, 36706851, 30615163, 49635551, 41652435, 26994603,
22810332, 26326086, 37433879, 24405033, 30533452, 17724705, 20568292, 16606580,
14861424, 20347790, 21664534, 20231829, 23825988, 23769178, 32615868, 28992931,
21560522, 27910616, 18289289, 18123114, 22076183, 18755541, 28225747, 23507043,
34061036, 22768258, 23756648, 28019796, 26023841, 23459220, 33986881, 28584900,
18304042, 18491800, 15944105, 18001977, 17782474, 15220947]}}
```

```
[6]: AMZN_df = pd.DataFrame(r.json())
      AMZN_df
```

```
[6]:
```

| | c | h | l | o | s | t | v |
|---|---------|-----------|-----------|-----------|----|------------|----------|
| 0 | 1891.97 | 1913.8900 | 1860.0000 | 1875.0000 | ok | 1577923200 | 19514188 |
| 1 | 1862.02 | 1917.8200 | 1855.0900 | 1909.8900 | ok | 1578528000 | 15160738 |
| 2 | 1887.46 | 1902.5000 | 1857.2500 | 1882.9900 | ok | 1579132800 | 13580875 |
| 3 | 1858.00 | 1894.9900 | 1815.3400 | 1885.1100 | ok | 1579737600 | 14688733 |

| | | | | | | | |
|----|---------|-----------|-----------|-----------|----|------------|----------|
| 4 | 2039.87 | 2071.0200 | 1850.6100 | 1858.0000 | ok | 1580342400 | 37459327 |
| 5 | 2160.00 | 2185.9500 | 2024.8001 | 2041.0200 | ok | 1580947200 | 22414811 |
| 6 | 2170.22 | 2185.1000 | 2124.1050 | 2144.9900 | ok | 1581552000 | 11150195 |
| 7 | 1979.59 | 2176.7900 | 1958.4200 | 2173.0700 | ok | 1582156800 | 25788575 |
| 8 | 1975.83 | 1996.3300 | 1811.1290 | 1934.3800 | ok | 1582761600 | 36706851 |
| 9 | 1820.86 | 1960.7200 | 1761.2901 | 1933.0000 | ok | 1583366400 | 30615163 |
| 10 | 1830.00 | 1857.7750 | 1626.0318 | 1721.9800 | ok | 1583971200 | 49635551 |
| 11 | 1885.84 | 1957.0000 | 1812.0000 | 1860.0000 | ok | 1584576000 | 41652435 |
| 12 | 1907.70 | 1993.0200 | 1889.2900 | 1902.0000 | ok | 1585180800 | 26994603 |
| 13 | 2043.00 | 2044.0000 | 1889.1500 | 1901.6400 | ok | 1585785600 | 22810332 |
| 14 | 2307.68 | 2333.3690 | 2017.6600 | 2044.3000 | ok | 1586390400 | 26326086 |
| 15 | 2363.49 | 2460.9981 | 2279.6600 | 2346.0000 | ok | 1586995200 | 37433879 |
| 16 | 2372.71 | 2444.8800 | 2306.0000 | 2399.9800 | ok | 1587600000 | 24405033 |
| 17 | 2351.26 | 2475.0000 | 2256.3800 | 2419.8400 | ok | 1588204800 | 30533452 |
| 18 | 2367.92 | 2419.6650 | 2337.8050 | 2374.7800 | ok | 1588809600 | 17724705 |
| 19 | 2497.94 | 2500.0100 | 2353.2100 | 2361.0100 | ok | 1589414400 | 20568292 |
| 20 | 2410.39 | 2525.4500 | 2330.0000 | 2500.0000 | ok | 1590019200 | 16606580 |
| 21 | 2478.40 | 2488.0000 | 2378.2300 | 2384.3300 | ok | 1590624000 | 14861424 |
| 22 | 2647.45 | 2722.3500 | 2437.1300 | 2477.4300 | ok | 1591228800 | 20347790 |
| 23 | 2640.98 | 2671.3800 | 2503.3500 | 2603.5000 | ok | 1591833600 | 21664534 |
| 24 | 2734.40 | 2796.0000 | 2636.1100 | 2647.0100 | ok | 1592438400 | 20231829 |
| 25 | 2878.70 | 2895.0000 | 2630.0800 | 2739.5500 | ok | 1593043200 | 23825988 |
| 26 | 3081.11 | 3083.9700 | 2871.1021 | 2912.0100 | ok | 1593648000 | 23769178 |
| 27 | 3008.87 | 3344.2900 | 2950.0000 | 3115.9900 | ok | 1594252800 | 32615868 |
| 28 | 3099.91 | 3240.5758 | 2918.2300 | 2971.0600 | ok | 1594857600 | 28992931 |
| 29 | 3033.53 | 3098.2713 | 2888.0000 | 3098.2713 | ok | 1595462400 | 21560522 |
| 30 | 3205.03 | 3246.8200 | 3005.0000 | 3014.0000 | ok | 1596067200 | 27910616 |
| 31 | 3162.24 | 3247.4700 | 3073.0000 | 3194.3600 | ok | 1596672000 | 18289289 |
| 32 | 3260.48 | 3320.0000 | 3120.0000 | 3182.9900 | ok | 1597276800 | 18123114 |
| 33 | 3441.85 | 3451.7387 | 3238.0000 | 3252.0000 | ok | 1597881600 | 22076183 |
| 34 | 3531.45 | 3552.2500 | 3378.0000 | 3450.0500 | ok | 1598486400 | 18755541 |
| 35 | 3268.61 | 3488.4100 | 3111.1300 | 3485.0000 | ok | 1599091200 | 28225747 |
| 36 | 3078.10 | 3349.8900 | 3074.1500 | 3307.2188 | ok | 1599696000 | 23507043 |
| 37 | 2999.86 | 3133.9900 | 2871.0000 | 3009.2500 | ok | 1600300800 | 34061036 |
| 38 | 3148.73 | 3212.8800 | 2965.0000 | 2977.7900 | ok | 1600905600 | 22768258 |
| 39 | 3195.69 | 3224.0000 | 3090.0000 | 3208.0000 | ok | 1601510400 | 23756648 |
| 40 | 3363.71 | 3496.2400 | 3174.9900 | 3224.9900 | ok | 1602115200 | 28019796 |
| 41 | 3184.94 | 3399.6606 | 3160.0000 | 3292.0100 | ok | 1602720000 | 26023841 |
| 42 | 3162.78 | 3291.6600 | 3121.9400 | 3189.8700 | ok | 1603324800 | 23459220 |
| 43 | 3241.16 | 3257.2500 | 2950.1200 | 3201.2700 | ok | 1603929600 | 33986881 |
| 44 | 3137.39 | 3366.8000 | 3019.4828 | 3319.9700 | ok | 1604534400 | 28584900 |
| 45 | 3105.46 | 3189.2500 | 3072.6850 | 3159.9500 | ok | 1605139200 | 18304042 |
| 46 | 3185.07 | 3198.0000 | 3065.4600 | 3105.3100 | ok | 1605744000 | 18491800 |
| 47 | 3104.20 | 3228.6400 | 3088.0000 | 3205.4600 | ok | 1606953600 | 15944105 |
| 48 | 3240.96 | 3247.0000 | 3072.8200 | 3088.9900 | ok | 1607558400 | 18001977 |
| 49 | 3185.27 | 3263.5100 | 3166.0000 | 3250.0000 | ok | 1608163200 | 17782474 |
| 50 | 3285.85 | 3350.6500 | 3169.0000 | 3193.9000 | ok | 1608768000 | 15220947 |

```
[7]: from datetime import datetime
import requests
import pandas

stk_ticker = 'AMZN'
data_resolution = 'W'
timestamp_from = 1577865600
timestamp_to = 1609315200
API_Key = 'bsiqli7rh5rc8orbnkqg'
Address_template = 'https://finnhub.io/api/v1/stock/candle?
↳symbol={}&resolution={}&from={}&to={}&token={}'

API_address = Address_template.format(stk_ticker,data_resolution,
                                     timestamp_from,timestamp_to,API_Key)

AMZN_df = pd.DataFrame(r.json())
AMZN_df.drop(columns=['s'],inplace=True)
AMZN_df.t = AMZN_df.t.apply(datetime.fromtimestamp)
AMZN_df.t = AMZN_df.t.apply(lambda v:v.date())
AMZN_df.set_index('t',drop=True,inplace=True)
AMZN_df.columns = ['Closing','High','Low','Opening','Volume']
AMZN_df.head()
```

```
[7]:
```

| | Closing | High | Low | Opening | Volume |
|------------|---------|---------|---------|---------|----------|
| t | | | | | |
| 2020-01-01 | 1891.97 | 1913.89 | 1860.00 | 1875.00 | 19514188 |
| 2020-01-08 | 1862.02 | 1917.82 | 1855.09 | 1909.89 | 15160738 |
| 2020-01-15 | 1887.46 | 1902.50 | 1857.25 | 1882.99 | 13580875 |
| 2020-01-22 | 1858.00 | 1894.99 | 1815.34 | 1885.11 | 14688733 |
| 2020-01-29 | 2039.87 | 2071.02 | 1850.61 | 1858.00 | 37459327 |

```
[ ]:
```