

analysis.

```
#!python3 -m pip install python-dotenv
#!python3 -m pip install pandas
#!python3 -m pip install seaborn matplotlib plotly scikit-learn pyarrow
```

Collecting python-dotenv
 Downloading python_dotenv-1.2.1-py3-none-any.whl.metadata (25 kB)
 Downloading python_dotenv-1.2.1-py3-none-any.whl (21 kB)
Installing collected packages: python-dotenv
Successfully installed python-dotenv-1.2.1

```
Requirement already satisfied: seaborn in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: matplotlib in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: plotly in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: pyarrow in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: numpy in /usr/local/lib/python3.6/dist-packages  
Requirement already satisfied: pandas in /usr/local/lib/python3.6/dist-packages
```

```
Requirement already satisfied: contourpy>=1.0.1 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (1.3.3)
Requirement already satisfied: cycler>=0.10 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (4.61.0)
Requirement already satisfied: kiwisolver>=1.3.1 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (1.4.9)
Requirement already satisfied: packaging>=20.0 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (25.0)
Requirement already satisfied: pillow>=8 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (12.0.0)
Requirement already satisfied: pyparsing>=3 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (3.2.5)
Requirement already satisfied: python-dateutil>=2.7 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from matplotlib) (2.9.0.post0)
Requirement already satisfied: narwhals>=1.15.1 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from plotly) (2.13.0)
Requirement already satisfied: scipy>=1.8.0 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from scikit-learn) (1.16.3)
Requirement already satisfied: joblib>=1.2.0 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from scikit-learn) (1.5.2)
```

Requirement already satisfied: pytz>=2020.1 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from pandas>=1.2>seaborn) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from pandas>=1.2>seaborn) (2025.2)
Requirement already satisfied: six>=1.5 in /Users/sagorika/Library/Python/3.14/lib/python/site-packages (from python-dateutil>=2.7>matplotlib) (1.17)

Fetching and saving raw data

The next cell loads the environment variables, imports the data fetching function, and retrieves all 2024 NYC 311 records from the Socrata API. It handles pagination automatically and prints summary information about the fetched data. Finally, it saves the raw dataset as a parquet file for later cleaning steps.

```
In [ ]: # importing fetch function for NYC 311
from pathlib import Path
import os
import sys

# loading environment variables
try:
    from dotenv import load_dotenv
    load_dotenv(Path('..') / '.env')
except ImportError:
    print("Warning: python-dotenv not installed. Install with: pip install python-dotenv")
    print("Continuing without .env file support...")

PROJECT_ROOT = Path('..').resolve()
if str(PROJECT_ROOT) not in sys.path:
    sys.path.append(str(PROJECT_ROOT))

from src.nyc311_api import fetch_nyc311_data

# reading Socrata app token from environment variable
APP_TOKEN = os.getenv("SOCRATA_APP_TOKEN")
if not APP_TOKEN:
    print("Note: SOCRAITA_APP_TOKEN not found in environment. API calls may be slower.")

# fetching 2024 data
df_raw = fetch_nyc311_data(
    app_token=APP_TOKEN,
    limit=25000,
    year=2024,
)

print(f"DataFrame shape: {df_raw.shape}")
print(f"Columns: {list(df_raw.columns)}")
print(f"\nFirst few rows:")
print(df_raw.head().to_string())

raw_dir = PROJECT_ROOT / 'data' / 'raw'
raw_dir.mkdir(parents=True, exist_ok=True)

raw_path = raw_dir / 'nyc311_2024_raw.parquet'
df_raw.to_parquet(raw_path, index=False)

print(f"\nRaw data saved to: {raw_path}")
raw_path
```

Starting data fetch for year 2024...
Using limit of 25,000 rows per request.
App token detected. Using authenticated requests.

Fetching page 1 (offset 0)... Received 25,000 rows (total so far: 25,000)
Fetching page 2 (offset 25,000)... Received 25,000 rows (total so far: 50,000)
Fetching page 3 (offset 50,000)... Received 25,000 rows (total so far: 75,000)
Fetching page 4 (offset 75,000)... Received 25,000 rows (total so far: 100,000)
Fetching page 5 (offset 100,000)... Received 25,000 rows (total so far: 125,000)
Fetching page 6 (offset 125,000)... Received 25,000 rows (total so far: 150,000)
Fetching page 7 (offset 150,000)... Received 25,000 rows (total so far: 175,000)
Fetching page 8 (offset 175,000)... Received 25,000 rows (total so far: 200,000)
Fetching page 9 (offset 200,000)... Received 25,000 rows (total so far: 225,000)
Fetching page 10 (offset 225,000)... Received 25,000 rows (total so far: 250,000)
Fetching page 11 (offset 250,000)... Received 25,000 rows (total so far: 275,000)
Fetching page 12 (offset 275,000)... Received 25,000 rows (total so far: 300,000)
Fetching page 13 (offset 300,000)... Received 25,000 rows (total so far: 325,000)
Fetching page 14 (offset 325,000)... Received 25,000 rows (total so far: 350,000)
Fetching page 15 (offset 350,000)... Received 25,000 rows (total so far: 375,000)
Fetching page 16 (offset 375,000)... Received 25,000 rows (total so far: 400,000)
Fetching page 17 (offset 400,000)... Received 25,000 rows (total so far: 425,000)
Fetching page 18 (offset 425,000)... Received 25,000 rows (total so far: 450,000)
Fetching page 19 (offset 450,000)... Received 25,000 rows (total so far: 475,000)
Fetching page 20 (offset 475,000)... Received 25,000 rows (total so far: 500,000)
Fetching page 21 (offset 500,000)... Received 25,000 rows (total so far: 525,000)
Fetching page 22 (offset 525,000)... Received 25,000 rows (total so far: 550,000)
Fetching page 23 (offset 550,000)... Received 25,000 rows (total so far: 575,000)
Fetching page 24 (offset 575,000)... Received 25,000 rows (total so far: 600,000)
Fetching page 25 (offset 600,000)... Received 25,000 rows (total so far: 625,000)
Fetching page 26 (offset 625,000)... Received 25,000 rows (total so far: 650,000)
Fetching page 27 (offset 650,000)... Received 25,000 rows (total so far: 675,000)
Fetching page 28 (offset 675,000)... Received 25,000 rows (total so far: 700,000)
Fetching page 29 (offset 700,000)... Received 25,000 rows (total so far: 725,000)
Fetching page 30 (offset 725,000)... Received 25,000 rows (total so far: 750,000)
Fetching page 31 (offset 750,000)... Received 25,000 rows (total so far: 775,000)
Fetching page 32 (offset 775,000)... Received 25,000 rows (total so far: 800,000)
Fetching page 33 (offset 800,000)... Received 25,000 rows (total so far: 825,000)
Fetching page 34 (offset 825,000)... Received 25,000 rows (total so far: 850,000)
Fetching page 35 (offset 850,000)... Received 25,000 rows (total so far: 875,000)
Fetching page 36 (offset 875,000)... Received 25,000 rows (total so far: 900,000)
Fetching page 37 (offset 900,000)... Received 25,000 rows (total so far: 925,000)
Fetching page 38 (offset 925,000)... Received 25,000 rows (total so far: 950,000)
Fetching page 39 (offset 950,000)... Received 25,000 rows (total so far: 975,000)
Fetching page 40 (offset 975,000)... Received 25,000 rows (total so far: 1,000,000)
Fetching page 41 (offset 1,000,000)... Received 25,000 rows (total so far: 1,025,000)
Fetching page 42 (offset 1,025,000)... Received 25,000 rows (total so far: 1,050,000)
Fetching page 43 (offset 1,050,000)... Received 25,000 rows (total so far: 1,075,000)
Fetching page 44 (offset 1,075,000)... Received 25,000 rows (total so far: 1,100,000)
Fetching page 45 (offset 1,100,000)... Received 25,000 rows (total so far: 1,125,000)
Fetching page 46 (offset 1,125,000)... Received 25,000 rows (total so far: 1,150,000)
Fetching page 47 (offset 1,150,000)... Received 25,000 rows (total so far: 1,175,000)
Fetching page 48 (offset 1,175,000)... Received 25,000 rows (total so far: 1,200,000)
Fetching page 49 (offset 1,200,000)... Received 25,000 rows (total so far: 1,225,000)
Fetching page 50 (offset 1,225,000)... Received 25,000 rows (total so far: 1,250,000)
Fetching page 51 (offset 1,250,000)... Received 25,000 rows (total so far: 1,275,000)
Fetching page 52 (offset 1,275,000)... Received 25,000 rows (total so far: 1,300,000)
Fetching page 53 (offset 1,300,000)... Received 25,000 rows (total so far: 1,325,000)
Fetching page 54 (offset 1,325,000)... Received 25,000 rows (total so far: 1,350,000)
Fetching page 55 (offset 1,350,000)... Received 25,000 rows (total so far: 1,375,000)
Fetching page 56 (offset 1,375,000)... Received 25,000 rows (total so far: 1,400,000)

```
Fetching page 58 (offset 1,425,000)...
Fetching page 59 (offset 1,450,000)...
Fetching page 60 (offset 1,475,000)...
Fetching page 61 (offset 1,500,000)...
Fetching page 62 (offset 1,525,000)...
Fetching page 63 (offset 1,550,000)...
Fetching page 64 (offset 1,575,000)...
```

```
Fetching page 66 (offset 1,625,000)... Received 25,000 rows (total so far: 1,650,000)
Fetching page 67 (offset 1,650,000)... Received 25,000 rows (total so far: 1,675,000)
Fetching page 68 (offset 1,675,000)... Received 25,000 rows (total so far: 1,700,000)
Fetching page 69 (offset 1,700,000)... Received 25,000 rows (total so far: 1,725,000)
Fetching page 70 (offset 1,725,000)... Received 25,000 rows (total so far: 1,750,000)
Fetching page 71 (offset 1,750,000)... Received 25,000 rows (total so far: 1,775,000)
Fetching page 72 (offset 1,775,000)... Received 25,000 rows (total so far: 1,800,000)
Fetching page 73 (offset 1,800,000)... Received 25,000 rows (total so far: 1,825,000)
Fetching page 74 (offset 1,825,000)... Received 25,000 rows (total so far: 1,850,000)
Fetching page 75 (offset 1,850,000)... Received 25,000 rows (total so far: 1,875,000)
Fetching page 76 (offset 1,875,000)... Received 25,000 rows (total so far: 1,900,000)
```

```

Fetching page 77 (offset 1,925,000)... Received 25,000 rows (total so far: 1,925,000)
Fetching page 78 (offset 1,925,000)... Received 25,000 rows (total so far: 1,950,000)
Fetching page 79 (offset 1,950,000)... Received 25,000 rows (total so far: 1,975,000)
Fetching page 80 (offset 1,975,000)... Received 25,000 rows (total so far: 2,000,000)
Fetching page 81 (offset 2,000,000)... Received 25,000 rows (total so far: 2,025,000)
Fetching page 82 (offset 2,025,000)... Received 25,000 rows (total so far: 2,050,000)
Fetching page 83 (offset 2,050,000)... Received 25,000 rows (total so far: 2,075,000)
Fetching page 84 (offset 2,075,000)... Received 25,000 rows (total so far: 2,100,000)
Fetching page 85 (offset 2,100,000)... Received 25,000 rows (total so far: 2,125,000)
Fetching page 86 (offset 2,125,000)... Request timed out. Retrying in 2 seconds... (attempt 1/3)
Received 25,000 rows (total so far: 2,150,000)
Fetching page 87 (offset 2,150,000)... Received 25,000 rows (total so far: 2,175,000)
Fetching page 88 (offset 2,175,000)... Received 25,000 rows (total so far: 2,200,000)
Fetching page 89 (offset 2,200,000)... Received 25,000 rows (total so far: 2,225,000)
Fetching page 90 (offset 2,225,000)... Received 25,000 rows (total so far: 2,250,000)
Fetching page 91 (offset 2,250,000)... Received 25,000 rows (total so far: 2,275,000)
Fetching page 92 (offset 2,275,000)... Received 25,000 rows (total so far: 2,300,000)
Fetching page 93 (offset 2,300,000)... Received 25,000 rows (total so far: 2,325,000)
Fetching page 94 (offset 2,325,000)... Received 25,000 rows (total so far: 2,350,000)
Fetching page 95 (offset 2,350,000)... Received 25,000 rows (total so far: 2,375,000)
Fetching page 96 (offset 2,375,000)... Received 25,000 rows (total so far: 2,400,000)
Fetching page 97 (offset 2,400,000)... Received 25,000 rows (total so far: 2,425,000)
Fetching page 98 (offset 2,425,000)... Received 25,000 rows (total so far: 2,450,000)
Fetching page 99 (offset 2,450,000)... Received 25,000 rows (total so far: 2,475,000)
Fetching page 100 (offset 2,475,000)... Received 25,000 rows (total so far: 2,500,000)
Fetching page 101 (offset 2,500,000)... Received 25,000 rows (total so far: 2,525,000)
Fetching page 102 (offset 2,525,000)... Received 25,000 rows (total so far: 2,550,000)
Fetching page 103 (offset 2,550,000)... Received 25,000 rows (total so far: 2,575,000)
Fetching page 104 (offset 2,575,000)... Received 25,000 rows (total so far: 2,600,000)
Fetching page 105 (offset 2,600,000)... Received 25,000 rows (total so far: 2,625,000)
Fetching page 106 (offset 2,625,000)... Received 25,000 rows (total so far: 2,650,000)
Fetching page 107 (offset 2,650,000)... Received 25,000 rows (total so far: 2,675,000)
Fetching page 108 (offset 2,675,000)... Received 25,000 rows (total so far: 2,700,000)
Fetching page 109 (offset 2,700,000)... Received 25,000 rows (total so far: 2,725,000)
Fetching page 110 (offset 2,725,000)... Received 25,000 rows (total so far: 2,750,000)
Fetching page 111 (offset 2,750,000)... Received 25,000 rows (total so far: 2,775,000)
Fetching page 112 (offset 2,775,000)... Received 25,000 rows (total so far: 2,800,000)
Fetching page 113 (offset 2,800,000)... Received 25,000 rows (total so far: 2,825,000)
Fetching page 114 (offset 2,825,000)... Received 25,000 rows (total so far: 2,850,000)
Fetching page 115 (offset 2,850,000)... Received 25,000 rows (total so far: 2,875,000)
Fetching page 116 (offset 2,875,000)... Received 25,000 rows (total so far: 2,900,000)
Fetching page 117 (offset 2,900,000)... Received 25,000 rows (total so far: 2,925,000)
Fetching page 118 (offset 2,925,000)... Received 25,000 rows (total so far: 2,950,000)
Fetching page 119 (offset 2,950,000)... Received 25,000 rows (total so far: 2,975,000)
Fetching page 120 (offset 2,975,000)... Received 25,000 rows (total so far: 3,000,000)
Fetching page 121 (offset 3,000,000)... Received 25,000 rows (total so far: 3,025,000)
Fetching page 122 (offset 3,025,000)... Received 25,000 rows (total so far: 3,050,000)
Fetching page 123 (offset 3,050,000)... Received 25,000 rows (total so far: 3,075,000)
Fetching page 124 (offset 3,075,000)... Received 25,000 rows (total so far: 3,100,000)
Fetching page 125 (offset 3,100,000)... Received 25,000 rows (total so far: 3,125,000)
Fetching page 126 (offset 3,125,000)... Received 25,000 rows (total so far: 3,150,000)
Fetching page 127 (offset 3,150,000)... Received 25,000 rows (total so far: 3,175,000)
Fetching page 128 (offset 3,175,000)... Received 25,000 rows (total so far: 3,200,000)
Fetching page 129 (offset 3,200,000)... Received 25,000 rows (total so far: 3,225,000)
Fetching page 130 (offset 3,225,000)... Received 25,000 rows (total so far: 3,250,000)
Fetching page 131 (offset 3,250,000)... Received 25,000 rows (total so far: 3,275,000)
Fetching page 132 (offset 3,275,000)... Received 25,000 rows (total so far: 3,300,000)
Fetching page 133 (offset 3,300,000)... Received 25,000 rows (total so far: 3,325,000)
Fetching page 134 (offset 3,325,000)... Received 25,000 rows (total so far: 3,350,000)
Fetching page 135 (offset 3,350,000)... Received 25,000 rows (total so far: 3,375,000)
Fetching page 136 (offset 3,375,000)... Received 25,000 rows (total so far: 3,400,000)
Fetching page 137 (offset 3,400,000)... Received 25,000 rows (total so far: 3,425,000)
Fetching page 138 (offset 3,425,000)... Received 25,000 rows (total so far: 3,450,000)
Fetching page 139 (offset 3,450,000)... Received 8,320 rows (total so far: 3,458,320)
Last page reached.
Snapshot date (UTC): 2025-11-17 05:35:13 UTC
Total rows fetched: 3458320
DataFrame shape: (3458320, 41)
Columns: ['unique_key', 'created_date', 'closed_date', 'agency', 'agency_name', 'complaint_type', 'descriptor', 'location_type', 'incident_zip', 'incident_address', 'street_name', 'cross_street_1', 'cross_street_2', 'intersection_street_1', 'intersection_street_2', 'address_type', 'city', 'landmark', 'status', 'resolution_description', 'resolution_action', 'updated_date', 'community_board', 'bbi', 'borough', 'x_coordinate_state_plane', 'y_coordinate_state_plane', 'open_data_channel_type', 'park_facility_name', 'park_borough', 'latitude', 'longitude', 'location', 'facility_type', 'bridge_highway_name', 'bridge_highway_segment', 'taxi_pick_up_location', 'vehicle_type', 'bridge_highway_direction', 'road_ramp', 'taxi_company_borough', 'due_date', 'dtype: object']
First few rows:
Raw data saved to: /Users/sagorika/Documents/HCDS/Final Project/A7/data/raw/nyc311_2024_raw.parquet
Out[ ]: PosixPath('/Users/sagorika/Documents/HCDS/Final Project/A7/data/raw/nyc311_2024_raw.parquet')

The next cell just outputs first few rows of the dataframe
In [ ]: print(df_raw.head().to_string())
unique_key      created_date      closed_date agency      agency_name      complaint_type descriptor
location_type  incident_zip  incident_address street_name  cross_street_1  cross_street_2 intersection_street_1 intersection_street_2
t_2_address_type  city      landmark      status      bbl      borough x_coordinate_state_plane y_coordinate_state_plane
resolution_description resolution_action updated_date  community_board      bbl      borough x_coordinate_state_plane y_coordinate_state_plane
data_channel_type  park_facility_name  park_borough      latitude      longitude
location_facility_type  bridge_highway_name  bridge_highway_segment  taxi_pick_up_location  vehicle_type  bridge_highway_direction  road_ramp  taxi_company_borough
gh_due_date
0  59898042  2024-01-01T00:00:00.000  2024-01-02T09:24:45.000  DOHMH  Department of Health and Mental Hygiene  Food Poisoning  1 or 2 Restaurants
ar/Deli/Bakery  10462  711 MORRIS PARK AVENUE  MORRIS PARK AVENUE  WHITE PLAINS ROAD  HUNT AVENUE  WHITE PLAINS ROAD
ENUE ADDRESS  BRONX  MORRIS PARK AVENUE Closed  The Department of Health and Mental Hygiene has determined that the reported incident is isolated. The establishment will be inspected according to its normal schedule for sanitary inspections.  2024-01-02T09:24:45.000  11 BRONX
2040500001  BRONX  1021437  247286  ONLINE  Unspecified  BRONX  40.845339757487484
-73.86559379953668  {'latitude': '40.84533975748463', 'longitude': '-73.86559379953668', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
1  59896272  2024-01-01T00:00:00.000  2024-01-02T00:01:00.000  DOHMH  Department of Health and Mental Hygiene  Food Poisoning  1 or 2 Restaurants
ar/Deli/Bakery  11211  145 GRAND STREET EXTENSION  GRAND STREET EXTENSION  SOUTH 2 STREET  KEAP STREET  SOUTH 2 STREET
REET ADDRESS  BROOKLYN  GRAND STREET EXTENSION Closed  The Department of Health and Mental Hygiene has determined that the reported incident is isolated. The establishment will be inspected according to its normal schedule for sanitary inspections.  2024-01-11T10:04:21.000  01 BROOKLYN
3024110017  BROOKLYN  996989  198198  ONLINE  Unspecified  BROOKLYN  40.7106753063896
-73.954050047883  {'latitude': '40.710675306389646', 'longitude': '-73.954050047883', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
2  59894610  2024-01-01T00:00:00.000  2024-01-02T09:46:32.000  DOHMH  Department of Health and Mental Hygiene  Food Poisoning  1 or 2 Restaurants
ar/Deli/Bakery  10304  7 NAVY PIER COURT  NAVY PIER COURT  FRONT STREET  DEAD END  FRONT STREET
END ADDRESS  STATEN ISLAND  NAVY PIER COURT Closed  The NYC Health Department (DOHMH) found a service request ady exists for the same location. This request has been closed. DOHMH is addressing the original request.  2024-12-27T12:00:00.000  01 STATEN ISLAND
5004877501  STATEN ISLAND  963602  168384  ONLINE  Unspecified  STATEN ISLAND  40.6288276510719
-74.0743866176229  {'latitude': '40.628827651071944', 'longitude': '-74.0743866176229', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
3  59899838  2024-01-01T00:00:00.000  2024-01-02T09:42:25.000  DOHMH  Department of Health and Mental Hygiene  Food Poisoning  1 or 2 Restaurants
ar/Deli/Bakery  10022  1069 1 AVENUE  1 AVENUE  EAST 58 STREET  EAST 59 STREET  EAST 58 STREET  EAST 5
REET ADDRESS  NEW YORK  1 AVENUE Closed  The Department of Health and Mental Hygiene has determined that the reported incident is isolated. The establishment will be inspected according to its normal schedule for sanitary inspections.  2024-01-11T14:23:12.000  06 MANHATTAN
1013510025  MANHATTAN  994697  215873  ONLINE  Unspecified  MANHATTAN  40.759191820164
-73.96228991313669  {'latitude': '40.75919182016476', 'longitude': '-73.96228991313669', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
4  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
5  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
6  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
7  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
8  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
9  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
10  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
11  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.01449651007135  {'latitude': '40.62709413549678', 'longitude': '-74.01449651007135', 'human_address': {'address': '', 'city': '', 'state': '', 'zip': ''}}  NaN  NaN  NaN  NaN  NaN  NaN  NaN
NaN  NaN
12  59887043  2024-01-01T00:00:20.000  2024-01-01T00:41:01.000  NYPD  New York City Police Department  Illegal Fireworks  N/A
street/Sidewalk  11228  7223 FORT HAMILTON PARKWAY  FORT HAMILTON PARKWAY  72 STREET  73 STREET  72 STREET  7
REET ADDRESS  BROOKLYN  FORT HAMILTON PARKWAY Closed  The Police Department responded to the complaint and with the information available observed no evidence of the violation at that time.  2024-01-01T00:41:04.000  10 BROOKLYN
N  3019310001  BROOKLYN  980226  167744  ONLINE  Unspecified  BROOKLYN  40.627094135
-74.0144965100
```

```
import sys, pandas as pd
from pathlib import Path

PROJECT_ROOT = Path('..').resolve()
sys.path.insert(0, str(PROJECT_ROOT))
from src.nyc311_cleaning import
```

```
# keeping top 10 complaint types by default for reliable comparisons
df_clean = clean_nyc311_data(
    df_raw,
    winsorize_percentile=99.0,
    top_complaint_types=10,
)

print("Cleaned data summary:")
print(f"Shape: {df_clean.shape}")
print(f"\nResponse time statistics (days):")
```

```
print(df_clean['response_time_days'].describe())
print(f"\nChannel distribution:")
print(df_clean['open_data_channel_type'].value_counts())
print(f"\nBorough distribution:")
print(df_clean['borough'].value_counts())
```

Starting data cleaning process...
Initial row count: 3,458,320

Step 1: Converting dates to datetime...
Step 2: Computing response time...
Step 3: Filtering invalid rows...
Removed 128,060 rows with missing required fields or invalid response times.
Remaining rows: 3,330,260
Step 4: Standardizing channel types...

Channel type distribution after standardization:
open_data_channel_type
Web 1586605
Phone 1032985
App 710670
Name: count, dtype: int64
Step 5: Removing duplicates...

Removed 2 duplicate rows based on unique_key.
Remaining rows: 3,330,258
Step 6: Winsorizing response times at 99.0th percentile...

Winsorized 33,303 values at 99.0th percentile (273.21 days).
Step 7: Selecting top complaint types...

Filtered to top complaint types:
Kept 10 complaint types
Removed 1,460,942 rows from other complaint types
Remaining rows: 1,869,316

Top complaint types:
- Illegal Parking: 505,732 records
- Noise - Residential: 379,252 records
- HEAT/HOT WATER: 261,748 records
- Blocked Driveway: 170,190 records
- Noise - Street/Sidewalk: 162,997 records
- UNSANITARY CONDITION: 120,699 records
- Abandoned Vehicle: 70,326 records
- Noise - Commercial: 68,334 records
- PLUMBING: 65,767 records
- PAINT/PLASTER: 64,271 records

Cleaning complete!
Final row count: 1,869,316
Cleaned data summary:
Shape: (1869316, 42)

Response time statistics (days):
count 1.869316e+06
mean 2.896917e+00
std 1.316524e+01
min 3.472222e-05
25% 2.879630e-02
50% 9.376157e-02
75% 8.103385e-01
max 2.732056e+02
Name: response_time_days, dtype: float64

Channel distribution:
open_data_channel_type
Web 758634
App 561853
Phone 548829
Name: count, dtype: int64

Borough distribution:
borough
BROOKLYN 580861
BRONX 486025
QUEENS 428697
MANHATTAN 329076
STATEN ISLAND 14524