ExploringTorontoNeighbor

November 14, 2024

[1]:

```python
import pandas as pd

# Step 1: Retrieve the table from Wikipedia
url = 'https://en.wikipedia.org/wiki/List_of_postal_codes_of_Canada:_M'
tables = pd.read_html(url)
df = tables[0]  # Select the first table

# Step 1: Split each cell into Postal Code, Borough, and Neighborhood information
data = []

# Iterate over each row in the DataFrame
for row in df.itertuples(index=False):
    for cell in row:
        # Skip empty cells
        if isinstance(cell, str) and "Not assigned" not in cell:
            # Split the string, e.g., "M1B Scarborough (Malvern / Rouge)"
            parts = cell.split(' ', 1)  # Split on the first space
            postal_code = parts[0]      # Extract the postal code
            # Split into Borough and Neighborhood
            borough_neigh = parts[1].split('(', 1)
            borough = borough_neigh[0].strip()
            neighborhood = borough_neigh[1].strip(')') if len(borough_neigh) > 1 else borough
            # Append the data to the list
            data.append([postal_code, borough, neighborhood])

# Step 2: Convert the cleaned data into a DataFrame
df_cleaned = pd.DataFrame(data, columns=['Postalcode', 'Borough', 'Neighborhood'])

# Filter out rows where Borough is 'Not assigned'
df_cleaned = df_cleaned[df_cleaned['Borough'] != 'Not assigned']

# Combine neighborhoods with the same postal code
df_cleaned = df_cleaned.groupby(['Postalcode', 'Borough'])['Neighborhood'].apply(lambda x: ', '.join(x)).reset_index()
```

```python
# Display the cleaned DataFrame
print(df_cleaned.head())
```

```
  Postalcode      Borough                              Neighborhood
0        M1B  Scarborough                           Malvern / Rouge
1        M1C  Scarborough  Rouge Hill / Port Union / Highland Creek
2        M1E  Scarborough       Guildwood / Morningside / West Hill
3        M1G  Scarborough                                    Woburn
4        M1H  Scarborough                                 Cedarbrae
```

[2]:

```python
import pandas as pd
import requests
import io

# Assume df_cleaned is the cleaned Toronto neighborhoods data from Part 1

# Step 1: Retrieve the latitude and longitude data
url = "https://cocl.us/Geospatial_data"
lat_long_data = requests.get(url).content
lat_long_df = pd.read_csv(io.StringIO(lat_long_data.decode('utf-8')))

# Step 2: Rename columns to match the existing DataFrame
lat_long_df = lat_long_df.rename(columns={'Postal Code': 'Postalcode'})

# Step 3: Merge the two DataFrames on 'Postalcode'
toronto_df = pd.merge(df_cleaned, lat_long_df, on='Postalcode')

# Step 4: Save as a CSV file
toronto_df.to_csv("Toronto_Neighborhoods_with_Coordinates.csv", index=False)

# Display the merged DataFrame
print(toronto_df.head())
```

```
  Postalcode      Borough                              Neighborhood  \
0        M1B  Scarborough                           Malvern / Rouge
1        M1C  Scarborough  Rouge Hill / Port Union / Highland Creek
2        M1E  Scarborough       Guildwood / Morningside / West Hill
3        M1G  Scarborough                                    Woburn
4        M1H  Scarborough                                 Cedarbrae

    Latitude  Longitude
0  43.806686 -79.194353
1  43.784535 -79.160497
2  43.763573 -79.188711
3  43.770992 -79.216917
4  43.773136 -79.239476
```