Propreturns Data Science Internship Assignment

Web Scraping

By:- Madhar Jadhav

```
 'Joshi Road',
 'Kailash Colony*',
 'Kala Mahal',
 'Kala Masjid',
 'Kalinidi Colony*',
 'Kalka Ji*',
 'Kalkaji*',
 'Kamla Market',
 'Kapashera*',
 'Karol Bagh',
 'Kaseruwalan*',
 'Kashmere Gate*',
 'Katra Chhotey Lal (Darya Ganj)',
 'Kaushri Wallan',
 'Kautilya Marg*',
 'Khaira*',
 'Khalsa Nagar',
 'Khari Baoli*',
 'Khirki Extension*',
 'Kilokari*',
 'Kinari Bazar*',
 'Kirti Nagar*',
 'Kishan Ganj*',
 'Kotla Mubarak Pur*',
 'Krishna Nagar',
 'Krishna Park*',
 'Kucha Chelan*',
 'Kucha Pandit (Bazar Sita Ram)',
 'Kucha Pati Ram (Pahar Ganj)',
 'Kucha Sohan Lal (Pahar Ganj)',
 'Kuchalal Man (Darya Ganj)',
 'Kunde Walan',
 'Laddu Ghati*',
 'Lajpat Nagar I*',
 'Lajpat Nagar*',
 'Lal Kuan',
 'Loha Mandi Naraina*',
 'Madanpur Khadar*',
 'Madipur*',
 'Mahipal Pur*',
 'Maidan Garhi*',
 'Main Bazar Pahar Ganj*',
 'Malka Ganj*',
 'Malviya Nagar*',
 'Man Singh Road*',
 'Manak Pura',
 'Mandawali*',
 'Mantola*',
 'Masih Garh*',
 'Masjid Moth*',
 'Masood Pur*',
 'Mata Rameshwari Nagar',
 'Mata Sundari Rly Colony (Minto Road)',
 'Mata Sundri Road*',
 'Matia Mahal*',
 'May Fair Garden*',
 'Meena Bazar (Jama Masjid)',
 'Mehrauli*',
 'Militry Road',
 'Minto Road',
 'Mirdard Road*',
 'Mithapur*',
 'Model Basti (Manakpura)',
 'Molarband*',
 'Mori Gate*',
 'Moti Bagh -1*',
 'Moti Nagar*',
 'Motia Bagh*',
 'Motia Khan*',
 'Multani Dhanda',
 'Munirka*',
 'Nabi Karim',
 'Nai Basti*',
 'Nai Sarak*',
 'Nai Wala',
 'Nai Wara*',
 'Naiwala*',
 'Nanak Pura*',
 'Nangal Raya*',
 'Naraina Vihar*',
 'Naraina*',
 'Nawab Ganj*',
 'Naya Bazar*',
 'Nehru Nagar*',
 'Netaji Nagar*',
 'New Friends Colony*',
 'New Moti Nagar*',
 'New Rajinder Nagar',
 'New Ranjit Nagar*',
 'New Rohtak Road*',
 'Niti Bagh*',
 'Nizamuddin West*',
 'North Extn. Area, Pusa Road',
 'Okhla Village*',
 'Old Daryaganj Area( Pataudi House)',
 'Old Rajinder Nagar',
 'Old Subzi Mandi*',
 'Others',
 'Padam Singh Road',
 'Padmini Enclave*',
 'Pahar Ganj',
 'Pahari Bhojla*',
 'Pahari Gajaan',
 'Pahari Imli',
 'Pai Walan (Chandni Chowk)',
 'Pamposh Enclave*',
 'Pandara Road*',
 'Pandav Nagar*',
 'Pant Nagar*',
 'Parsad Nagar*',
 'Partap Nagar*',
 'Paschim Puri*',
 'Paschim Vihar*',
 'Patel Nagar*',
 'Peshwa Road*',
 'Pitampura*',
 'Prashad Nagar',
 'Prem Nagar*',
 'Press Area (Darya Ganj)',
 'Prithvi Raj Road*',
 'Punjabi Bagh Extension*',
 'Punjabi Bagh*',
 'Punjabi Basti*',
 'Pusa Institute*',
 'Pusa Road',
 'Pyare Lal Road',
 'Qamra Bangush',
 'Qasab Pura*',
 'Qutab Road*',
 'Raigar Pura*',
 'Rajender Nagar*',
 'Rajendra Place',
 'Rajinder Nagar',
 'Rajindra Park, Pusa Road',
 'Rajouri Garden*',
 'Rajpur Road*',
 'Rakab Ganj (Darya Ganj)',
 'Rakab Ganj*',
 'Ram Nagar',
 'Rama Krishna Ashram Marg*',
 'Ramesh Nagar*',
 'Ramjas Road*',
 'Rampura*',
 'Rangpuri*',
 'Rani Jhansi Road',
 'Ravi Nagar*',
 'Regar Pura*',
 'Reghar Pura',
 'Rishi Nagar*',
 'Rodgran*',
 'Rohini*',
 'Roshanara Road*',
 'Rouse Avenue (I.T.O.)',
 'Sabzi Mandi*',
 'Sadar Bazar*',
 'Sadar Nala Road*',
 'Sadar Thana Road*',
 'Safdarjung Enclave*',
 'Sainik Farm*',
 'Samalka*',
 'Sanjay Amar Colony (Minto Road)',
 'Sanjay Nagar*',
 'Sant Nagar*',
 'Sarai Kale Khan*',
 'Sardar Patel Marg*',
 'Sarita Vihar*',
 'Sarojini Nagar*',
 'Sarvodaya Enclave*',
 'Sarvpriya Vihar*',
 'Sat Nagar',
 'Savitri Nagar*',
 'Sewa Nagar*',
 'Shadi Khampur*',
 'Shah Ganj (Bazar Sita Ram)',
 'Shahbad Mohammad Pur*',
 'Shahpur Jat*',
 'Shahurpur*',
 'Shakti Nagar*',
 'Shakur Basti*',
 'Shalimar Bagh*',
 'Shastri Park(Beadan Pura)',
 'Sheesh Mahal (Darya Ganj)',
 'Shiv Nagar*',
 'Shivaji Park*',
 'Shivalik*',
 'Shora Kothi (Pahar Ganj)',
 'Shora Kothi*',
 'Shyam Nagar*',
 'Siddi Pura',
 'Sidharth Basti*',
 'Sita Ram Bazar*',
 'South Patel Nagar*',
 'Sriniwaspuri*',
 'Subhash Nagar',
 'Subzi Mandi*',
 'Sui Walan',
 'Suiwalan*',
 'Sukhdev Vihar*',
 'Sunder Nagar*',
 'Sunder Vihar*',
 'Sunlight Colony*',
 'Swami Ramtirth Nagar (manakpura)',
 'Tagore Garden*',
 'Tagore Road*',
 'Taimoor Nagar*',
 'Tajpur*',
 'Thomson Road*',
 'Tibbia College*',
 'Tilak Bridge*',
 'Tilak Nagar*',
 'Timarpur*',
 'Tiraha Behram Khan',
 'Tiraha Behram Khan*',
 'Tis Hazari*',
 'Turkman Gate',
 'Uday Park*',
 'Uttam Nagar*',
 'Varinder Nagar*',
 'Vasant Kunj*',
 'Vijay Nagar*',
 'Vikas Puri*',
 'Vikram Nagar (Minto Road)',
 'Vikram Nagar*',
 'W.E.A. Karol Bagh',
 'Wazir Nagar*',
 'West Patel Nagar*',
 'Yamuna Bazar*',
 'Yusuf Sarai*',
 'Zakhira*',
 'Zakir Nagar*',
 'Zamrud Pur*']
```

Observation: Here we can see the list for locations

```
[4]: len(locations) # Lenghth Of location, we have
```

```
392
```

```
[7]:
from selenium import webdriver
from selenium.webdriver.common.by import By
from time import sleep
import pandas as pd
import numpy as np
from bs4 import BeautifulSoup
import urllib.request

# Initialize the webdriver

# Loop through all locations
for i in locations:
    # Open the website and select SRO, locality
    sleep(1)

    # Wait for user to manually enter captcha
    sleep(1)
    # Click the search button
    try:
```