Gishnavi Kolli

Analysing a Crime dataset using Folium

OVERVIEW

A project that is intended in providing a better understanding of creating maps and visualize. Usage of libraries Pandas.
In the course of completing the project, you use folium and create a Choropleth map.

Problem Statement

Creation of maps with markers to explore crimer rate in San Francisco, California.
Eventually creation of a Chloropleth map to visualize the crime rate in San Francisco.

Software Requirements

1. Programming Language : Python

2. Environemnt: Jupyter Notebooks / Google Collab

3. Database: CSV(export type)

4. Operation System: Windows XP or above

5. Librarires Used: Pandas,Folium

6.Datasets used:

San Francisco:https://cocl.us/sanfran_crime_dataset(https://cocl.us/sanfran_crime_dataset)

Geojson file: https://cocl.us/sanfran_geojson (https://cocl.us/sanfran_geojson)

1. **Open a New Notebook and import the required libraires and read the csv file**

```
import pandas as pd
df_sfcrime = pd.read_csv("../content/data.csv")
```

pandas is a Python package providing fast, flexible, and expressive data structures designed to make working with relational or labeled data both easy and intuitive. pandas is well suited for many different kinds of data: Tabular data with heterogeneously-typed columns, as in an SQL table or Excel spreadsheet

df_sfcrime reads the data from the dataset sfcrime

2. **Set the dataset accordingly**

```
df_tmp = df_sfcrime.groupby(['PdDistrict']).count().reset_index()
df_tmp.drop(['Category','Descript','DayOfWeek','Date','Time', 'Resolution','Address','X','Y','Location','PdId'], axis=1, inplace=True)
df_tmp.rename(columns={'PdDistrict':'Neighborhood', 'IncidntNum':'Count'}, inplace=True)
df_tmp
```

	Neighborhood	Count
0	BAYVIEW	14303
1	CENTRAL	17666
2	INGLESIDE	11594
3	MISSION	19503
4	NORTHERN	20100
5	PARK	8699
6	RICHMOND	8922
7	SOUTHERN	28445
8	TARAVAL	11325
9	TENDERLOIN	9942

- A groupby operation involves some combination of splitting the object, applying a function, and combining the results. This can be used to group large amounts of data and compute operations on these groups.

- df_tmp.drop drops the column(which is given as input)

- df_tmp.rename renames the column name

3. Importing Folium

- Here, we Import folium package in the environment

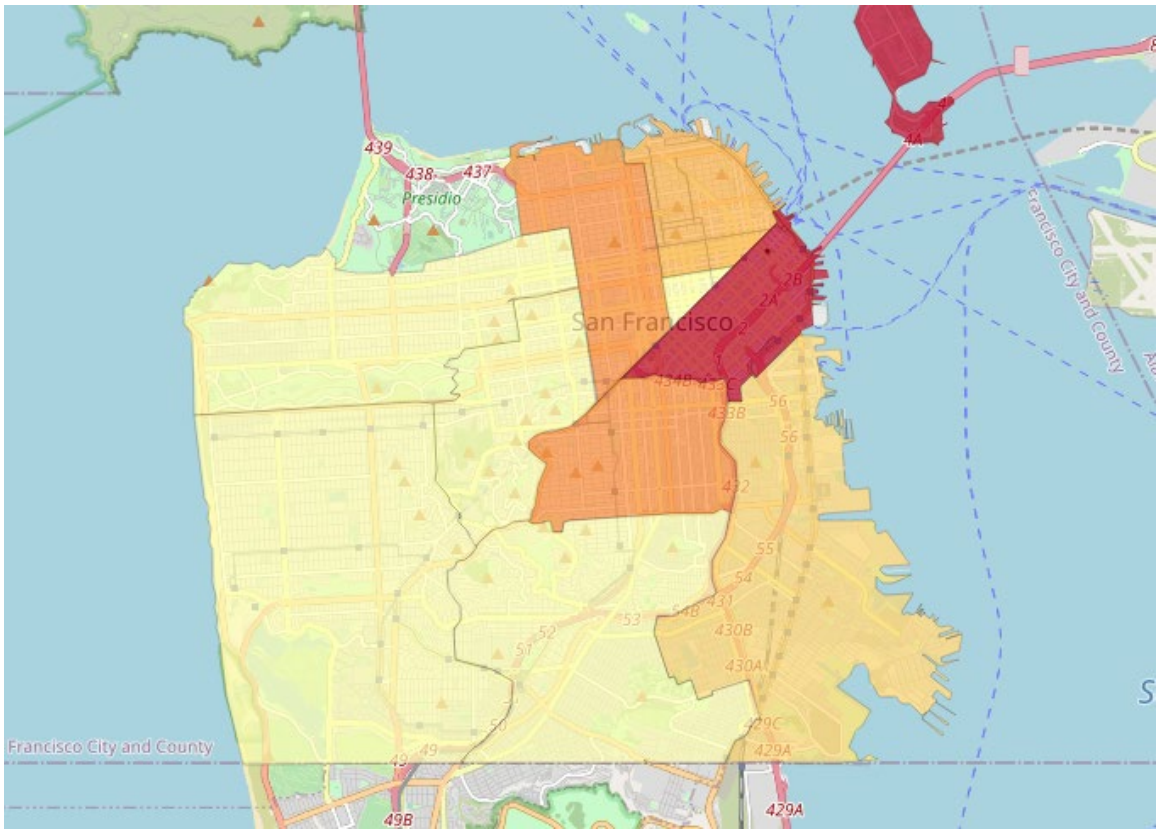
```
import folium
sf_geo = r'..//content/geo.geojson'
print('Folium installed and imported!')
```

- Folium is installed and imported

4. Creating a plain map(Choropleth) of Sanfrancisco:

```
sf_map = folium.Map(location=[37.773972, -122.431297], zoom_start=12) #, tiles='Mapbox Bright')
sf_map.choropleth(
    geo_data=sf_geo,
    data=df_tmp,
    columns=['Neighborhood','Count'],
    key_on='feature.properties.DISTRICT',
    fill_color='YlOrRd',
    fill_opacity=0.7,
    line_opacity=0.2,
    legend_name='San Francisco Crimes'
)


sf_map
```



- A Choropleth Map is a map composed of colored polygons. It is used to represent spatial variations of a quantity. This page documents how to build outline choropleth maps, but you can also build choropleth tilemaps using our Mapbox trace types

- to plot the crimes in San Fransisco, it should be centered around San Fransisco

- If you want to plot a map of a certain area. Here, you need to mention the appropriate starting zoom level so that the map focuses only on that region. Also, if you don't want them to zoom in or out much and lose focus on the map, then you can restrict it. For doing this, Folium gives us three parameters – zoom_start, min_zoom, and max_zoom.

- we use zoom_start of 12

- fill_color → Color scheme used in the visualization.

- we use fill_color = 'YlOrRd',

- fill_opacity → Area fill opacity, range 0–1 (default 0.6).

- we define fill_opacity = 0.7,
- line_opacity → GeoJSON geopath line opacity, range 0-1 (default 1).
- we define line_opacity=0.2
- legend_name → Title for the legend (default empty string).
- we define a legend whose default threshhold scale is used and Import folium package in the environment
- by using the GeoJSON correctly we'll be able to generate A Choropleth Map of crimes in San Francisco as shown above

5. Conclusion

In this Project, We created a Choropleth with markers while exploring and visualising the crime rate in San Franciso.

we restructured the data so that it is in the right format for generating the choloropleth map and create a dataframe that lists each neighborhood in SanFrancisco along with the corresponding total number of crimes.

Based on the San Francisco crime dataset, we found that San Francisco consists of 10 main neighborhoods as mentioned above

And finally ,we created a choloropleth map to visualize crimes in San Francisco