VIRGINIA COMMONWEALTH UNIVERSITY

Statistical Analysis & Modelling

A5 – Perceptual Mapping using Consumption Data Set

State: **Punjab**

Using Python- Colab

Submitted by

MONIKA SARADHA

#V01068784

Date of Submission: 12/07/2023

Table of Contents

1. Introduction

The NSSO-Consumption dataset is a product of the National Sample Survey Organization (NSSO), which was established by the Government of India in 1969. Recognizing the challenges of collecting information from every individual in a large country like India, the NSSO employs scientific sampling methods to collect socio-economic data. The surveys are conducted in rounds, with each round spanning a period of six months to one year. There are two types of samples: the "Central Samples," conducted by the Government of India, and the "State Samples," conducted by the respective states.

1.1. About the Data

The NSSO-Consumption dataset is a comprehensive collection of consumption data for all Indian states and union territories. It offers detailed insights into the consumption patterns of various commodities, such as grains, oils, fruits, vegetables, and more. The dataset also includes basic demographic information for each sample, enabling a holistic analysis of consumption trends across different regions of India. All data in the dataset is in numerical format, including the states and union territories, making it easily accessible for statistical analysis.

1.2. Objective

The primary goal of the NSSO-Consumption dataset is to provide useful data for policymaking, planning, and research. Policymakers can develop targeted interventions to promote economic growth, social welfare, and sustainable development by studying consumption patterns. This dataset can be used by researchers to better understand the factors that influence consumption behavior, identify regional variations, and investigate the impact of demographic variables on consumption habits. The dataset's goal is to facilitate evidence-based decision-making and contribute to a better understanding of India's consumption dynamics.

To analyze and visualize the consumption patterns in the state of Punjab at the district level

using NSSO (National Sample Survey Office) data.

- Understand Consumption Patterns:

Determine the variations in consumption levels across different districts of Punjab to gain insights into the consumption patterns of the state's population.

- District-wise Consumption Analysis:

Explore and compare the consumption levels among different districts in Punjab to identify districts with high or low consumption rates.

- Identify Disparities:

Identify any significant disparities in consumption across districts and investigate the factors contributing to these disparities.

- Spatial Representation:

Create a visual representation of consumption on a map of Punjab, highlighting the consumption levels in each district. This map will provide a clear overview of consumption variations geographically.

By accomplishing these objectives, one can enhance the understanding of consumption dynamics in Punjab, support informed decision-making, and contribute to the overall socio-economic development of the state.

1. Generate a district-wise histogram of Punjab consumption using NSSO data to understand the distribution of consumption levels and identify any variations or outliers among districts.

2. Create a consumption map of Punjab, illustrating consumption patterns in each district, to visually represent regional differences and facilitate data-driven decision-making for policy formulation and resource allocation.

1.3. Business Significance

This extensive collection of primary data, auxiliary information, and socioeconomic indicators enriches the NSSO-Consumption dataset, allowing researchers, policymakers, and analysts to investigate various dimensions of consumption patterns and their underlying factors in India.

Understanding consumer behavior and consumption patterns is critical for companies operating in a variety of industries in order to conduct effective market research, product development,

and marketing strategies. Businesses can gain a comprehensive understanding of the demand for various products and services across regions by leveraging the insights from this dataset, identifying potential market opportunities, and tailoring their offerings to meet consumer preferences. Businesses can also use the dataset to examine the impact of socioeconomic factors on consumption, identify target demographics, and optimize resource allocation for maximum profitability.

The NSSO primarily conducts four types of surveys: household surveys, enterprise surveys, village facilities, and land and livestock holdings. Provided state of Punjab comprises the following 4 division: Survey Design and Research (SDR), Field Operation Division (FOD), Data Process, andEconomic Analysis.

2. Results

2.1. R- output and Interpretation

Missing Value:

```
print(punjab.isnull().sum())
```

```
Unnamed: 0          0
grp                 0
Round_Centre        0
FSU_number          0
Round               0
                    ..
foodtotal_q         0
state_1             0
Region              0
fruits_df_tt_v      0
fv_tot              0
Length: 384, dtype: int64
```

Inference:

Post observing 3 missing values, due to the minority they were removed and thus the data was cleaned.

2.2.Data Preprocessing

```
Index(['Unnamed: 0', 'grp', 'Round_Centre', 'FSU_number', 'Round',
       'Schedule_Number', 'Sample', 'Sector', 'state', 'State_Region',
       ...
       'pickle_v', 'sauce_jam_v', 'Othrprocessed_v', 'Beveragestotal_v',
       'foodtotal_v', 'foodtotal_q', 'state_1', 'Region', 'fruits_df_tt_v',
       'fv_tot'],
      dtype='object', length=384)
```

Inference:

The outliers were checked for the chosen variables like "Pulses Total", "Milk Production", "Potato", "Sugar Output", "Eggs Quantity". Since the variables between different districts are bound to be different and not fall under a single quartile range, amendment was not carried out.

## 2.3.Data Histogram indicating the consumption District wise
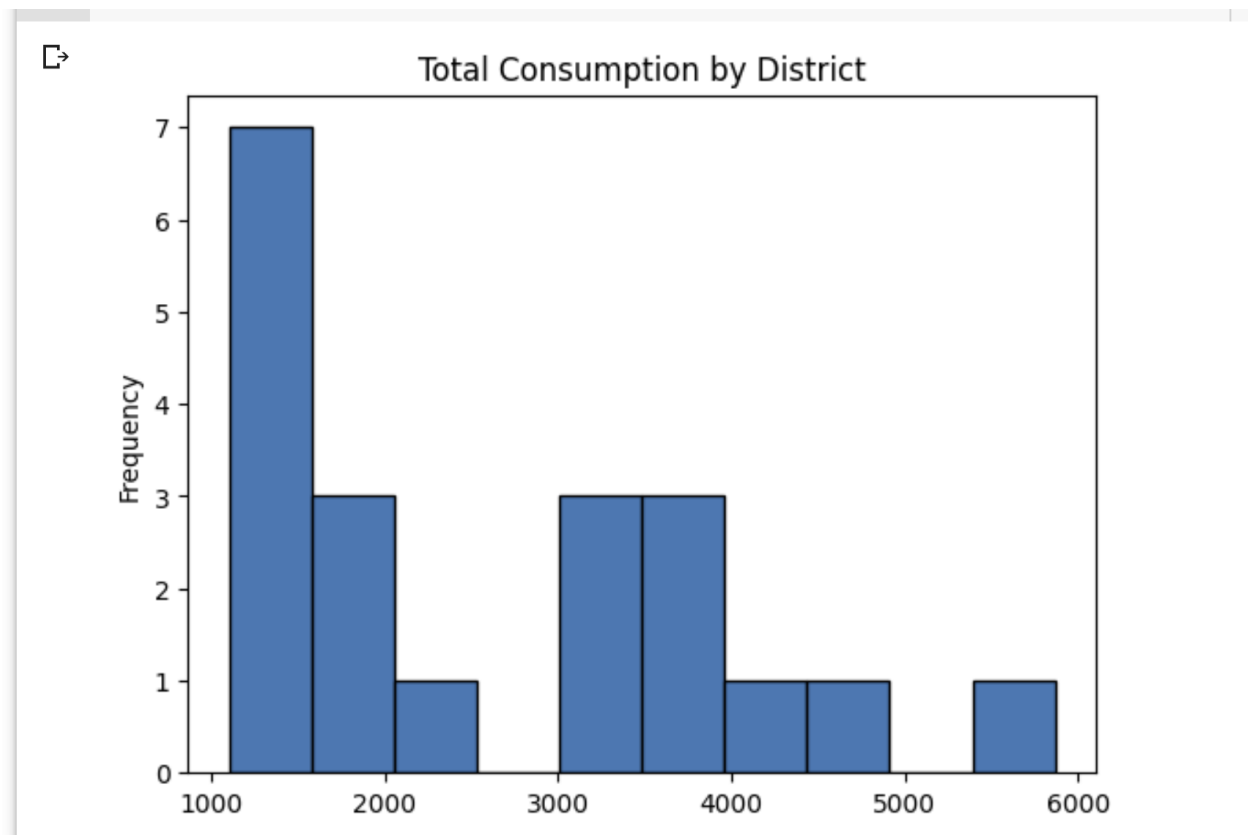
Inference:

The histogram plot represents the total consumption by district in Punjab. Each bar represents a district, with the height of the bar indicating the total consumption value. Based on the histogram, we can make the following inferences:

- **Variation in Total Consumption:** There is noticeable variation in total consumption among different districts in Punjab. Some districts have higher total consumption levels, while others have lower consumption levels.

- **Districts with Higher Consumption:** Districts like Ludhiana, Amritsar, Jalandhar, and Patiala appear to have higher total consumption, as indicated by taller bars in the histogram.

- **Districts with Lower Consumption:** Districts like Faridkot, Barnala, and Mansa show relatively lower total consumption, as seen from the shorter bars in the histogram.

- **Consumption Distribution:** The histogram provides a visual representation of the distribution of total consumption across different districts in Punjab. It helps identify districts with similar consumption levels and those that deviate from the average.

- **Further Analysis:** The histogram can serve as a starting point for further analysis and exploration of the factors influencing total consumption in different districts. It can be used to identify potential patterns or relationships between consumption levels and various socio-economic or geographical factors.
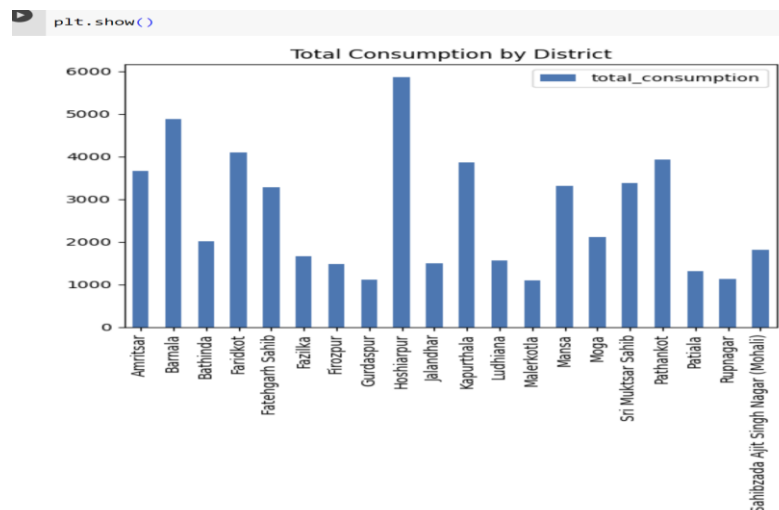
In this case, a bar plot depicts a better visualization to determine total consumption district wise as it is not a continuous variable.

2.4.Punjab state map showing consumption in each district

Total Consumption by District

Total Consumption by District
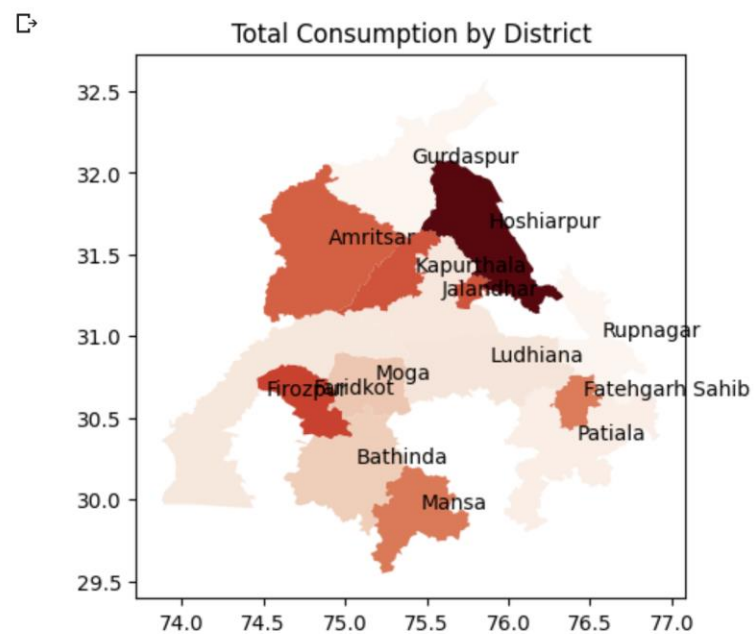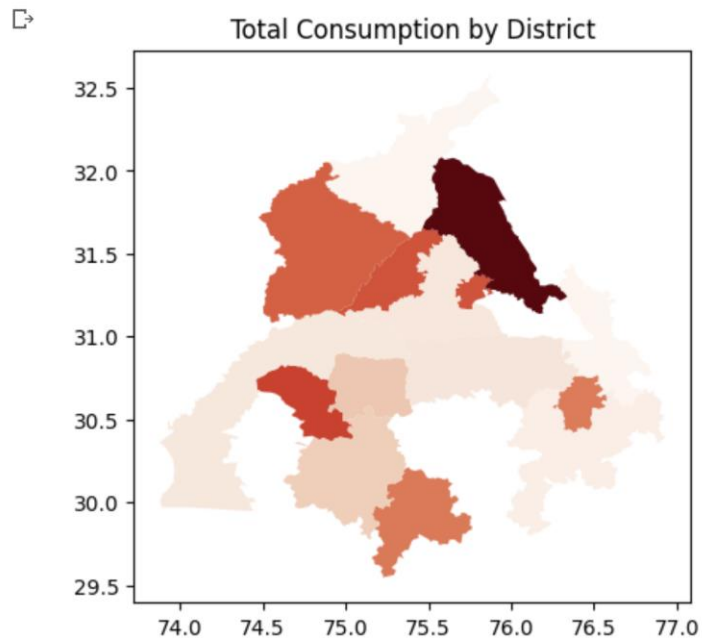
Inference:

The maps illustrate the total consumption in each district of Punjab. Based on the maps and visual observations, the following inferences can be made:

- **Consumption Variation:** The color gradient on the maps represents varying levels of total consumption across districts. Districts with darker shades (such as red) indicate higher consumption levels, while lighter shades (such as white) represent lower consumption levels.

- **High Consumption Districts:** Districts like Amritsar, Ludhiana, and Jalandhar exhibit higher total consumption, as indicated by their darker color shades on the maps. These districts are likely to have higher consumer demand or significant economic activity related to consumption.

- **Low Consumption Districts:** Districts such as Faridkot and Mansa show relatively lower total consumption, as depicted by their lighter color shades. These districts might have lower consumer demand or specific socio-economic factors contributing to lower consumption levels.

- **Spatial Consumption Patterns:** The maps highlight spatial variations in consumption across Punjab. Clusters or patterns of high and low consumption districts can be observed, suggesting potential regional disparities in consumption patterns.

- **Geographical Context:** The maps provide valuable information on the geographical distribution of consumption levels in Punjab. This insight can be useful for policymakers, researchers, and stakeholders in understanding the regional dynamics and formulating targeted strategies for development and resource allocation.

3. Recommendations

3.1. Business Implications:

1. Consumption Patterns: The histogram and map visualization provide insights into the consumption patterns across different districts in Punjab. Understanding these patterns can help businesses identify areas of high or low consumption, enabling targeted marketing strategies and resource allocation.

2. Regional Disparities: The visualizations highlight regional disparities in consumption levels. Businesses can leverage this information to tailor their products and services according to the specific needs and preferences of different districts, taking into account the varying socio-economic factors influencing consumption.

3. Market Opportunities: High consumption districts, such as Ludhiana and Jalandhar, represent potential lucrative markets with higher consumer demand. Businesses can focus on expanding their presence and marketing efforts in these areas to tap into the growth potential and maximize profitability.

4. Niche Markets: Low consumption districts, like Faridkot and Mansa, present opportunities for businesses to explore niche markets or introduce innovative products and services that cater to the unique needs of these regions. Adapting offerings to match the local context can help businesses gain a competitive advantage.

5. Targeted Advertising and Promotions: Businesses can utilize the visualizations to customize advertising and promotional campaigns based on the consumption patterns of each district. Tailored marketing messages and offers can resonate better with the local consumers, leading to improved brand awareness, customer engagement, and sales.

3.2. Business Recommendations:

1. Companies operating in Punjab or looking to expand into the region should use this data to inform their business strategies and identify potential growth opportunities.

2. Further analysis could be conducted to understand the factors driving consumption in high-performing districts and to identify potential barriers to growth in lower-performing

districts.

3. Companies could also consider tailoring their product and service offerings to meet the specific needs and preferences of consumers in different districts.

4. Marketing and advertising efforts could be targeted towards high-performing districts to capitalize on the existing demand for goods and services.

5. Companies could also consider partnering with local businesses or organizations in lower-performing districts to better understand the needs of consumers and develop strategies to increase consumption.

6. Companies could conduct market research to better understand the needs, preferences, and behaviors of consumers in different districts. This could include surveys, focus groups, and other forms of data collection to gather insights into consumer attitudes and behaviors.

7. Based on the market research, companies could develop targeted marketing campaigns and promotions to appeal to consumers in different districts. This could include special offers, discounts, and other incentives to encourage consumers to try new products or services.

8. Companies could also consider developing new products or services specifically tailored to the needs and preferences of consumers in different districts. This could involve adapting existing products or developing entirely new offerings to meet the unique needs of consumers in each district.

9. Companies could explore opportunities for collaboration and partnership with local businesses, organizations, and government agencies to support economic development and growth in lower-performing districts. This could include initiatives to improve infrastructure, provide training and education, and support entrepreneurship and small business development.

10. Companies could also consider implementing corporate social responsibility (CSR) initiatives to support the communities in which they operate. This could include programs to support education, healthcare, environmental sustainability, and other social causes that align with the company's values and mission.

```python
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
```

```python
punjab = pd.read_excel('Punjab_ds.xlsx')
```

```python
print(punjab.isnull().sum())
```

```
    Unnamed: 0         0
    grp                0
    Round_Centre       0
    FSU_number         0
    Round              0
                      ..
    foodtotal_q        0
    state_1            0
    Region             0
    fruits_df_tt_v     0
    fv_tot             0
    Length: 384, dtype: int64
```

```python
punjab.columns
```

```
    Index(['Unnamed: 0', 'grp', 'Round_Centre', 'FSU_number', 'Round',
           'Schedule_Number', 'Sample', 'Sector', 'state', 'State_Region',
           ...
           'pickle_v', 'sauce_jam_v', 'Othrprocessed_v', 'Beveragestotal_v',
           'foodtotal_v', 'foodtotal_q', 'state_1', 'Region', 'fruits_df_tt_v',
           'fv_tot'],
          dtype='object', length=384)
```

```python
punjab['total_consumption'] = punjab['pulsestt_q'] + punjab['milkprott_q'] + punjab['potato_q'] + punjab['sugaros_q'] + punjab['eggsno_q'
```

```python
punjab_sub = punjab.groupby('District').agg({'total_consumption': 'sum'})
```

```python
punjab_sub
```

	total_consumption
District	
1	3660.100939
2	4880.652753
3	2015.619446
4	4100.966164
5	3271.497420
6	1661.233112
7	1481.124996
8	1105.630588
9	5866.944963
10	1499.725269
11	3869.137393
12	1555.404637
13	1099.418176
14	3311.742322
15	2117.099704
16	3379.735090
17	3934.740153
18	1310.506061
19	1122.544450
20	1817.606383

Draw a histogram of the data in Exercise 1 to indicate the consumption district wise.

```
district_names = ["Amritsar", "Barnala", "Bathinda", "Faridkot", "Fatehgarh Sahib", "Fazilka", "Firozpur", "Gurdaspur", "Hoshiarpur", "Ja
                  "Mansa", "Moga", "Sri Muktsar Sahib", "Pathankot", "Patiala", "Rupnagar", "Sahibzada Ajit Singh Nagar (Mohali)"]
punjab_sub.index = district_names


punjab_sub
```

	total_consumption
Amritsar	3660.100939
Barnala	4880.652753
Bathinda	2015.619446
Faridkot	4100.966164
Fatehgarh Sahib	3271.497420
Fazilka	1661.233112
Firozpur	1481.124996
Gurdaspur	1105.630588
Hoshiarpur	5866.944963
Jalandhar	1499.725269
Kapurthala	3869.137393
Ludhiana	1555.404637
Malerkotla	1099.418176
Mansa	3311.742322
Moga	2117.099704
Sri Muktsar Sahib	3379.735090
Pathankot	3934.740153
Patiala	1310.506061
Rupnagar	1122.544450
Sahibzada Ajit Singh Nagar (Mohali)	1817.606383

```
punjab_sub.plot(kind='bar', y='total_consumption', title='Total Consumption by District')

plt.show()
```

```
punjab_sub['total_consumption'].plot(kind='hist', bins=10, edgecolor='black', title='Total Consumption by District')

plt.show()
```


Total Consumption by District
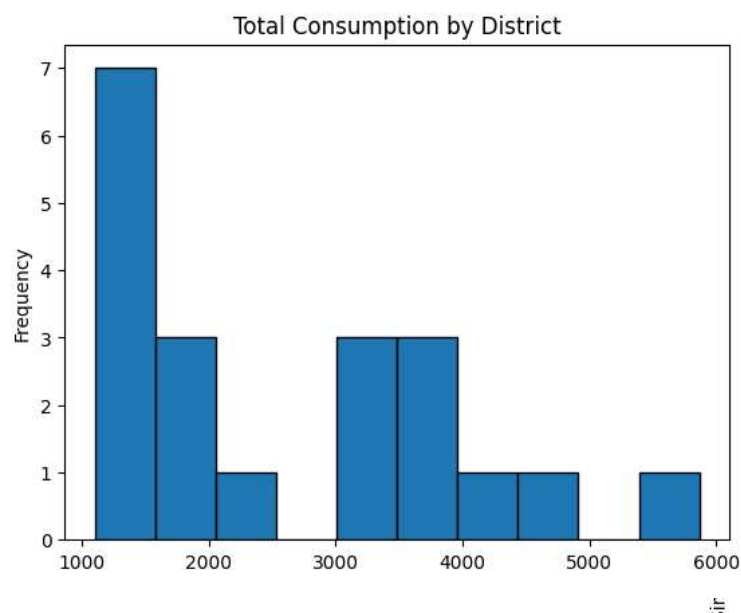
b) Depict the consumption on the map of the state showing consumption in each district.

```
!pip install pyshp
```

```
Requirement already satisfied: pyshp in /usr/local/lib/python3.10/dist-packages (2.3.1)
```

```
gdf = gpd.read_file('indian_districts.shp')
```

```
gdf
```

	latitude	total popu	state name	district n	state nam0	marginal w	main worke	country	iso	district 0

```
gdf.columns
```

```
Index(['latitude', 'total popu', 'state name', 'district n', 'state nam0',
       'marginal w', 'main worke', 'country', 'iso', 'district 0',
       'total work', 'district c', 'non-worker', 'geometry'],
      dtype='object')
```

2	32.503986	33224	Himachal	District Lahul & Spiti	Himachal	1879	19209	India	IND	Lahul and Spiti

```
gdf = gdf.rename(columns={'district 0':'District'})
```

3	28.739873	33363	Arunachal	District Upper Siang*	Arunachal	1710	15395	India	IND	Upper Siang

```
gdf.columns
```

```
Index(['latitude', 'total popu', 'state name', 'district n', 'state nam0',
       'marginal w', 'main worke', 'country', 'iso', 'District', 'total work',
       'longitude', 'district c', 'non-worker', 'geometry'],
      dtype='object')
```

33	20.321002	5464545	Maharashtra	District Wardha	Maharashtra	001011	2102004	India	IND	Wardha

```
punjab_sub.columns
```

```
Index(['District', 'total_consumption'], dtype='object')
```

590	19.127435	8640419	Maharashtra	District Mumbai	Maharashtra	181470	2971039	India	IND	Mumbai

```
punjab_map = pd.merge(gdf, punjab_sub, on='District')
```

```
punjab_map.columns
```

```
Index(['latitude', 'total popu', 'state name', 'district n', 'state nam0',
       'marginal w', 'main worke', 'country', 'iso', 'District', 'total work',
       'longitude', 'district c', 'non-worker', 'geometry',
       'total_consumption'],
      dtype='object')
```

```
import matplotlib.pyplot as plt
fig, ax = plt.subplots(1, 1)
punjab_map.plot(column='total_consumption', cmap='Reds', ax=ax)
ax.set_title('Total Consumption by District')
plt.show()
```
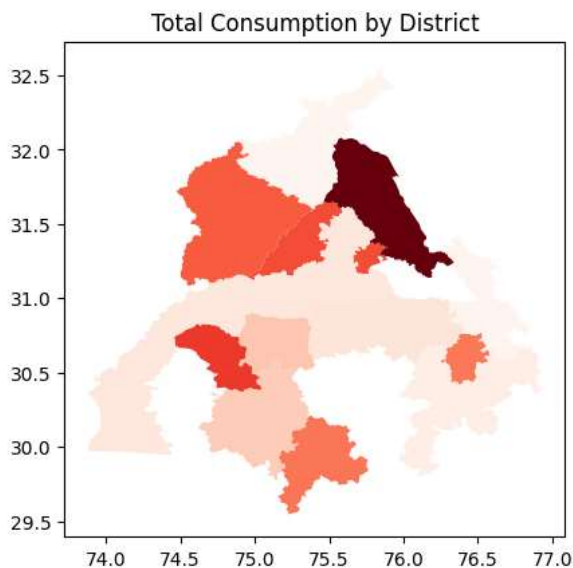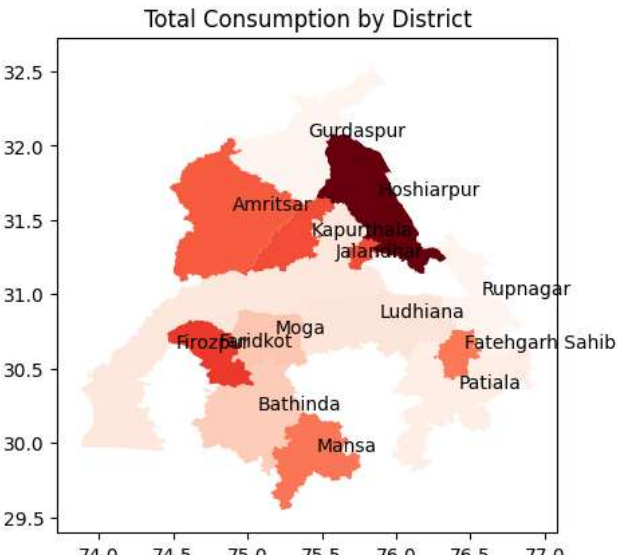


```
import matplotlib.pyplot as plt

fig, ax = plt.subplots(1, 1)
punjab_map.plot(column='total_consumption', cmap='Reds', ax=ax)
ax.set_title('Total Consumption by District')

# Add district names
for x, y, label in zip(punjab_map.geometry.centroid.x, punjab_map.geometry.centroid.y, punjab_map['District']):
    ax.annotate(label, xy=(x, y), xytext=(3, 3), textcoords="offset points")

plt.show()
```

```
punjab_map = gpd.read_file('indian_districts.shp')
```

```
# Create a base map showing the boundaries of each district
fig, ax = plt.subplots(1, 1)
punjab_map.plot(ax=ax, color='white', edgecolor='black')
plt.show()
```