

TOPIC - Descriptive statistics

Question 1 . Create a python function that takes two input years, year1 and year2, and returns a data frame containing statistical measures such as minimum, maximum, mean, variance, standard deviation, skewness, kurtosis, and interquartile range (IQR) for both temperature and precipitation datasets ?

Question 2 . Create a python function that accepts four parameters: Latitude, Longitude, year1, and year2. This function generates scatter plots, line plots, box plots, histogram plots, and violin plots based on the given input for both data sets. The goal is to derive meaningful inferences from the generated plots. Additionally, if any irregularities(e.g maybe min temp having a year) are detected in a particular year, further exploration will be conducted at the monthly level ?

Question 3 . Generate spatial plots (e.g., fig-01) illustrating minimum, maximum, mean, variance, standard deviation, skewness, kurtosis, and interquartile range (IQR) for both temperature and precipitation datasets. Additionally, create a correlation plot to explore the relationship between the two datasets. Analyze the correlation plot to gain insights and identify any patterns or trends ?

Question 4 . Generate spatial plots illustrating clustering for both the dataset .Use K Means clustering and for choosing optimal number of clusters use elbow method . Analyze how the same cluster points are related by statistical measures such as minimum, maximum, mean, variance, standard deviation, skewness, kurtosis, and interquartile range (IQR) ?

NOTE : In the zip file two datasets are attached containing NWH (northern western himalayas) region temperature and precipitation(rainfall) .

NWH Region - In the dataset, NWH region includes Jammu and Kashmir,Ladakh, Himachal Pradesh and Uttarakhand .

DATA - Temperature.csv contains temperature of NWH region in degree celsius .
Precipitation.csv contains rainfall data in mm .

To generate spatial plots 4 files are given, You have to upload all four files on your system but have to call only dot shp file .

For spatial plots you can use GEOPANDAS .

To install geopandas :

```
pip install geopandas
```

To read NWH region shape file :

```
import geopandas as gpd
```

```
shapefile=gpd.read_file("/content/4-17-2018-899072.shp")
```

To plot the shape file :

```
shapefile.plot(ax=ax,color='red')
```

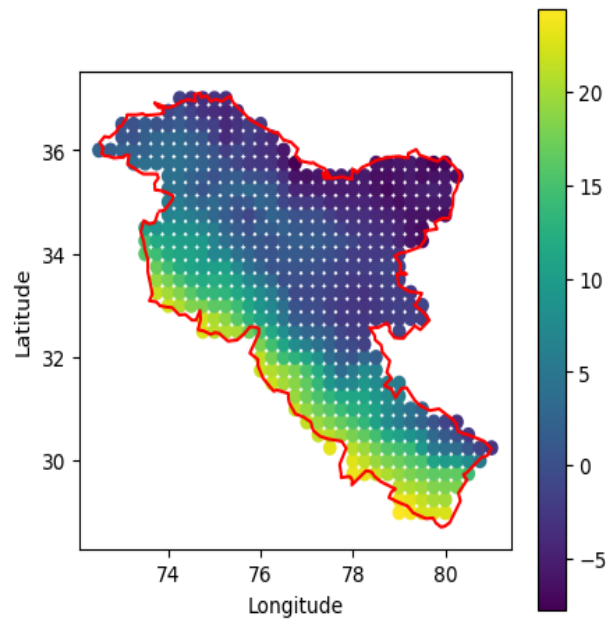


FIG :01 (For mean of temperature(yearwise))

K means clustering -

```
from sklearn.cluster import KMeans
X = data
kmeans = KMeans(n_clusters=2).fit(X)
kmeans.labels_
kmeans.predict()
kmeans.cluster_centers_
```

Elbow Method to detect optimal numbers of clusters -

```
sse = {}
For k in range(1, 10):
    kmeans = KMeans(n_clusters=k, max_iter=1000).fit(data)
    data["clusters"] = kmeans.labels_
#print(data["clusters"])
# Inertia: Sum of distances of samples to their closest cluster center
```

```
sse[k] = kmeans.inertia_
```

```
plt.figure()  
plt.plot(list(sse.keys()), list(sse.values()))  
plt.xlabel("Number of cluster")  
plt.ylabel("SSE")  
plt.show()
```