

Government Dataset

**"Analyzing Long-Term Rainfall Trends: A District-Wise Study
(1970-2021)"**

Introduction :

- Rainfall plays a crucial role in agriculture, water resource management, and climate studies. This dataset provides annual average rainfall data for various districts from 1970 to 2021. By analyzing this data, we can identify long-term trends, seasonal variations, and potential climate changes affecting different regions. Understanding these trends helps policymakers, researchers, and farmers make informed decisions about water conservation and disaster preparedness. This study aims to provide insights into the fluctuations in rainfall over the decades and its impact on different districts.

Overview:

	District/Year	2021	2020	2019	2018	2017	2016	2015	2014	2013	...	1982	1981	1980	1979	1978	1977	1976	1975	1974	1970
0	Gurdaspur	714.8	1282.8	1200.6	1306.9	1273.3	944.8	1166.7	1047.7	499.1	...	946.0	944.0	1155.0	911.6	903.1	1210.9	1370.6	944.3	599.5	926.3
1	Pathankot	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	Amritsar	518.4	581.9	480.7	511.5	579.8	552.1	562.9	323.8	685.8	...	544.0	542.7	869.8	511.9	542.7	702.3	1233.2	542.9	378.9	594.6
3	Tarn Taran	401.6	393.2	446.0	560.5	270.5	341.0	451.3	352.1	398.3	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	Kapurthala	1108.3	637.3	959.9	566.2	731.7	764.8	453.2	605.5	696.7	...	0.0	0.0	683.0	582.0	180.0	609.5	592.5	619.6	338.9	554.9
5	Jalandhar	828.3	561.0	542.4	575.7	459.9	326.6	378.5	311.3	593.9	...	616.0	636.3	873.9	666.3	645.6	774.7	627.0	648.5	367.9	171.4
6	SBS Nagar	769.2	745.6	971.6	1109.4	1048.6	650.0	824.7	567.9	835.7	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
7	Hoshiarpur	470.3	444.3	681.6	904.0	509.3	444.1	593.8	508.0	586.0	...	788.0	769.0	906.1	709.8	712.7	950.0	991.5	787.8	510.9	999.3
8	Rupnagar	758.3	1010.5	1018.6	1378.9	1022.7	845.0	821.7	622.7	957.8	...	776.0	753.9	759.1	609.9	815.7	732.0	628.7	866.2	557.2	983.4
9	SAS Nagar	492.1	668.7	705.5	924.8	670.4	586.0	832.0	611.0	646.3	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
10	Ludhiana	725.5	531.5	696.1	742.5	575.2	359.1	622.4	360.4	554.5	...	647.0	637.2	38.0	430.1	651.9	878.2	692.3	632.9	365.0	756.7
11	Ferozepur	437.9	399.7	245.1	107.2	93.7	131.5	171.0	204.7	300.0	...	665.0	361.5	956.2	386.5	359.8	464.8	763.8	349.7	172.0	232.3
12	Fazilka	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
13	Faridkot	510.6	667.1	416.8	411.8	342.2	412.6	490.0	396.0	796.9	...	429.0	423.1	511.4	305.3	425.9	513.6	608.3	410.6	273.3	NaN
14	Sri Muktsar Sahib	426.3	516.0	513.7	287.1	329.7	351.9	372.8	412.4	584.3	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
15	Moga	390.0	405.5	326.3	340.3	534.3	458.9	434.8	396.3	606.4	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
16	Bathinda	492.6	455.6	483.7	271.5	356.8	357.5	391.1	335.8	592.5	...	316.0	292.5	355.9	240.6	210.3	353.7	421.6	602.0	240.0	499.2
17	Mansa	211.8	224.4	296.4	159.5	264.5	203.0	391.1	183.5	260.0	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
18	Sangrur	337.3	604.6	373.8	492.1	300.6	377.5	492.7	270.2	325.4	...	0.0	0.0	521.4	997.4	853.0	680.3	625.5	485.2	275.1	521.9
19	Barnala	315.2	566.4	404.4	568.5	338.8	330.0	326.4	122.5	275.6	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
20	Patiala	678.4	678.2	928.2	842.9	542.2	325.6	514.1	395.0	503.8	...	681.0	665.0	835.7	488.8	640.7	943.7	881.7	660.2	418.9	555.6
21	Fatehgarh Sahib	530.3	677.5	579.9	712.5	496.7	304.6	685.9	263.8	695.3	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
22	Punjab	556.9	602.6	578.6	598.3	493.0	426.7	546.9	384.9	619.7	...	610.8	602.8	739.1	520.0	501.7	NaN	NaN	625.0	NaN	672.3

Column Description :

Column Description for the Rainfall Dataset

1. **District/Year** – Name of the district for which the annual average rainfall data is recorded.
2. **2021 to 1970** – Annual average rainfall (in millimeters) for each respective year.

Each column from **2021 to 1970** represents the recorded rainfall data for that particular year in different districts. Some years may have missing values due to unavailable records. This dataset helps in analyzing rainfall trends over the past five decades.

Data Cleaning:

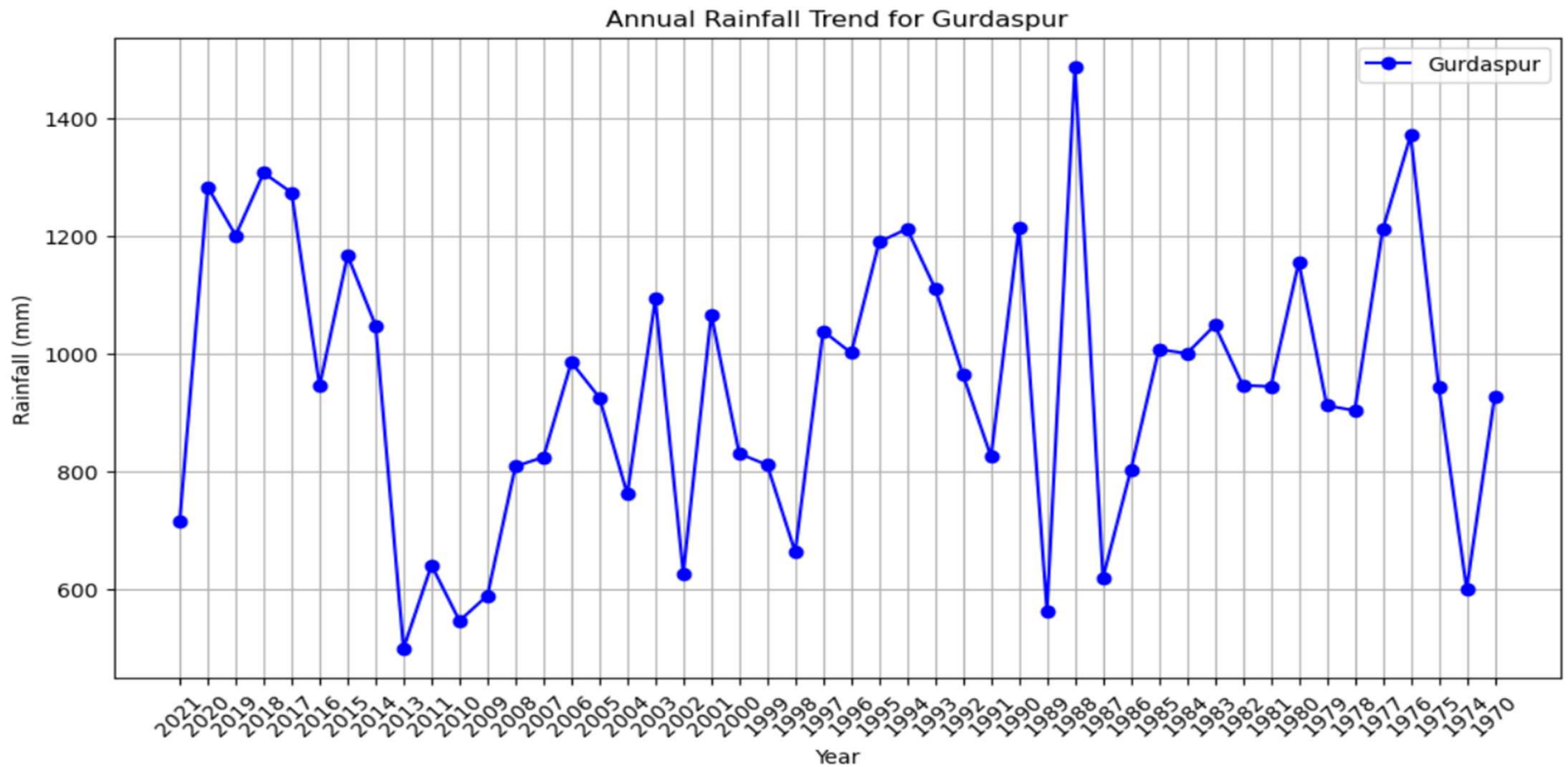
```
df.rename(columns={df.columns[0]: "District"}, inplace=True)
df.iloc[:, 1:] = df.iloc[:, 1:].apply(pd.to_numeric, errors='coerce')
df.fillna(df.mean(numeric_only=True), inplace=True)
df.dropna(thresh=len(df.columns) * 0.7, inplace=True)
df.drop_duplicates(subset=["District"], keep="first", inplace=True)
df.to_csv("Cleaned_AnnualAverageRainFall.csv", index=False)
print("Data cleaning completed. Cleaned file saved.")
```

Data cleaning completed. Cleaned file saved.

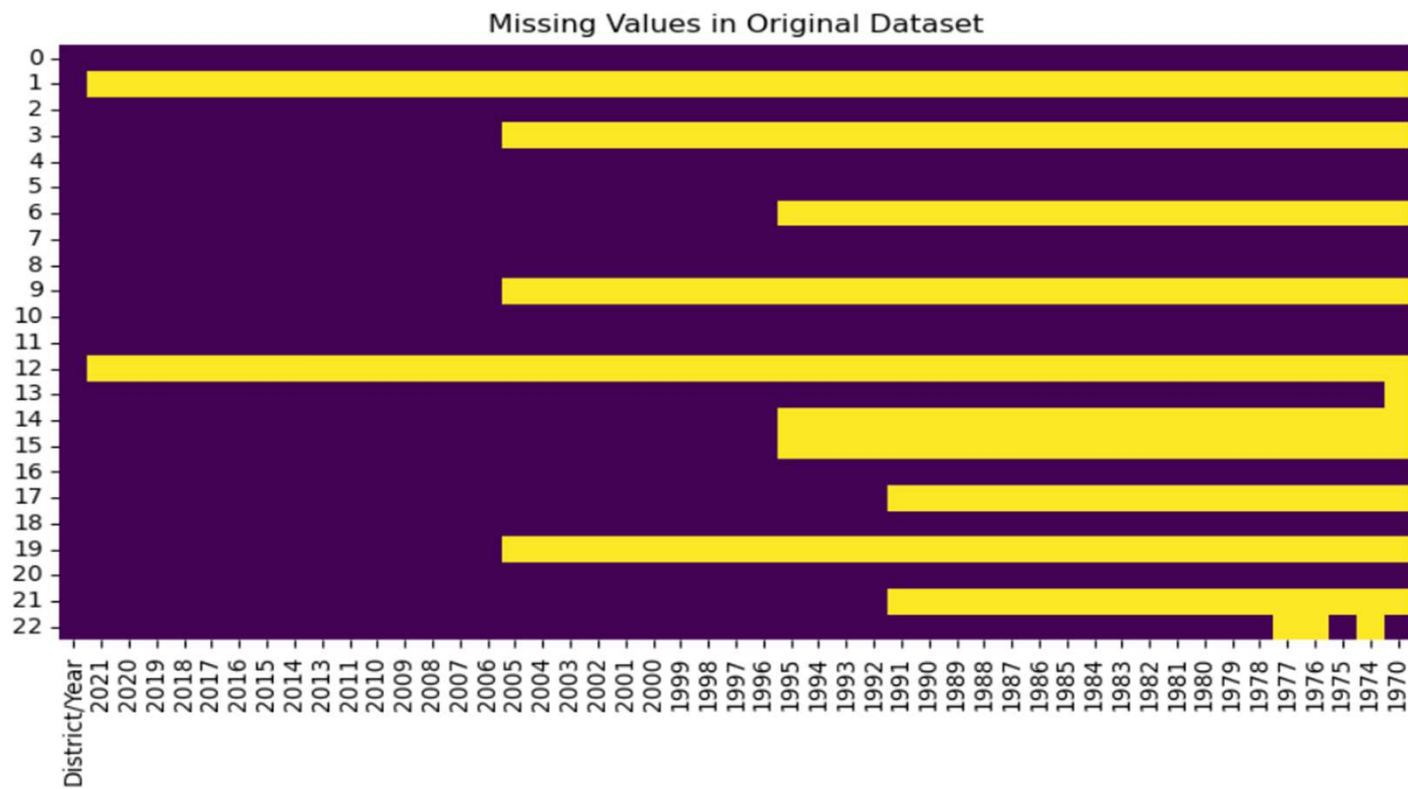
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21 entries, 0 to 20
Data columns (total 49 columns):
#   Column      Non-Null Count  Dtype
---  -
0   District    21 non-null    object
1   2021         21 non-null    float64
2   2020         21 non-null    float64
3   2019         21 non-null    float64
4   2018         21 non-null    float64
5   2017         21 non-null    float64
6   2016         21 non-null    float64
7   2015         21 non-null    float64
8   2014         21 non-null    float64
9   2013         21 non-null    float64
10  2011         21 non-null    float64
11  2010         21 non-null    float64
12  2009         21 non-null    float64
13  2008         21 non-null    float64
14  2007         21 non-null    float64
15  2006         21 non-null    float64
16  2005         21 non-null    float64
17  2004         21 non-null    float64
18  2003         21 non-null    float64
19  2002         21 non-null    float64
20  2001         21 non-null    float64
21  2000         21 non-null    float64
22  1999         21 non-null    float64
23  1998         21 non-null    float64
24  1997         21 non-null    float64
25  1996         21 non-null    float64
26  1995         21 non-null    float64
27  1994         21 non-null    float64
28  1993         21 non-null    float64
29  1992         21 non-null    float64
30  1991         21 non-null    float64
31  1990         21 non-null    float64
32  1989         21 non-null    float64
33  1988         21 non-null    float64
34  1987         21 non-null    float64
35  1986         21 non-null    float64
36  1985         21 non-null    float64
37  1984         21 non-null    float64
38  1983         21 non-null    float64
39  1982         21 non-null    float64
```

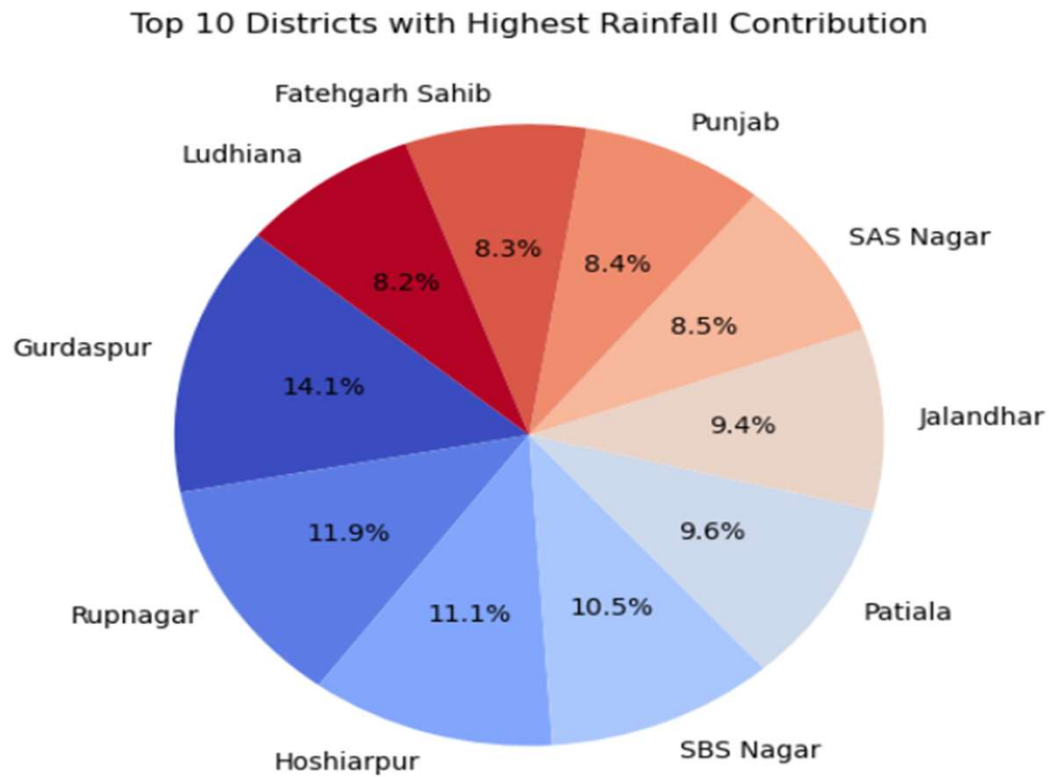
Rainfall in Gurdaspur:



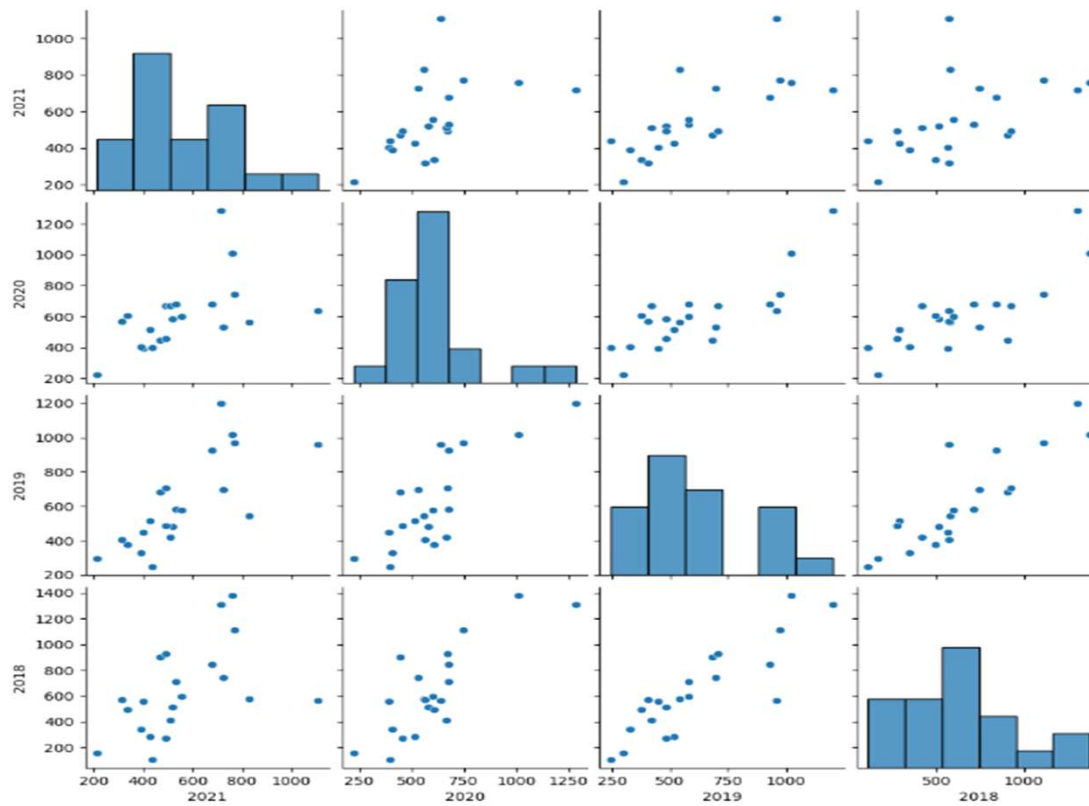
Missing Values in Dataset :



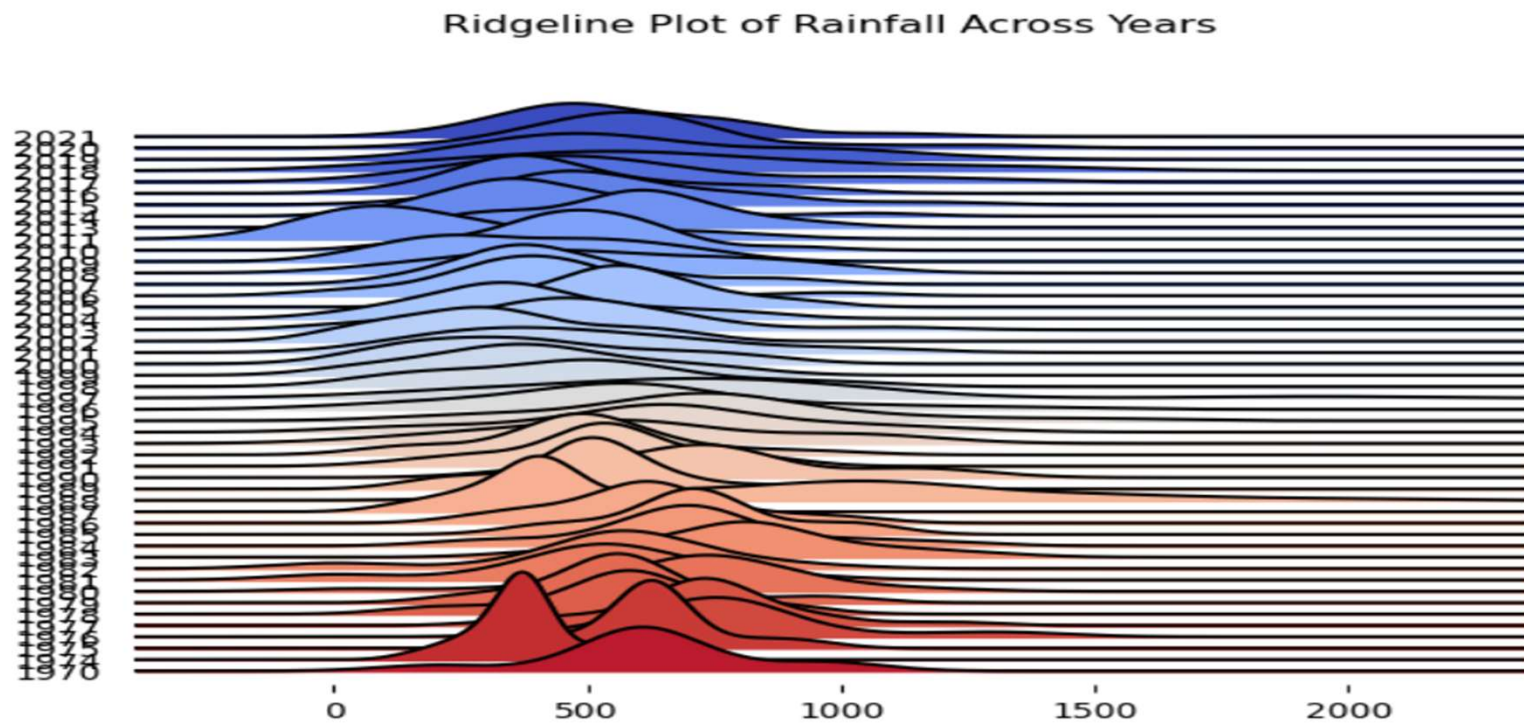
Highest Rainfall in 10 districts:



columns for visualization:

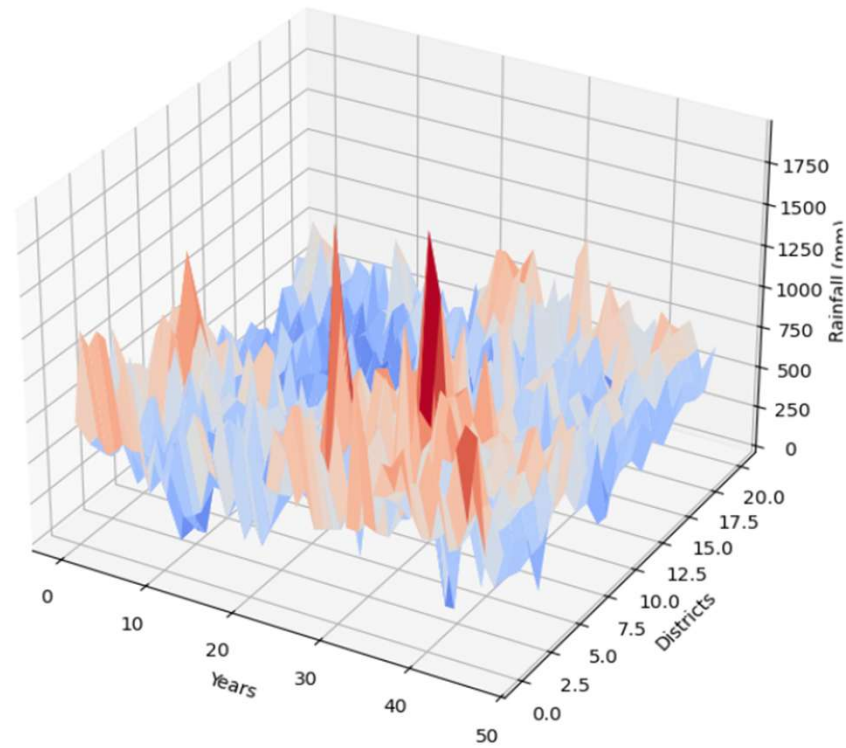


Rainfall Across Years:



3D surface Plot:

3D Surface Plot of Rainfall Variation



Conclusion :

The analysis of annual average rainfall data from **1970 to 2021** provides valuable insights into long-term precipitation trends across various districts. The dataset highlights **fluctuations in rainfall patterns**, with some regions experiencing steady rainfall while others show significant variations over the decades.

By cleaning and processing the data, we identified missing values and ensured data consistency for accurate trend analysis. The findings from this dataset can be crucial for **climate studies, water resource management, agricultural planning, and disaster preparedness**. Understanding these rainfall trends helps policymakers and researchers make informed decisions for sustainable environmental management.