Statistical & Inferential Analysis on Agriculture Dataset

Project Aim

To perform Statistical & Inferential Analysis to:

- **Analysing Farmer’s pay**
- **Effect of climate change on Farming**
- **Minimising crop loss in storage management**

Essential Tools

Libraries:

- **dplyr package** *(data manipulation)*
- **ggplot2 for plotting & visualisation**
- **tidyr package** *(cleaning the data)*
- **Stats package**

R-Studio

Datasets Chosen

- Agriculture crop production in India
  https://www.kaggle.com/srinivas1/agricuture-crops-production-in-india?select=datafile+%282%29.csv
- Rainfall in India
  https://data.world/rajanand/rainfall-in-india/workspace/file?filename=rainfall+in+india+1901-2015.csv
- Environment indicators in India
  https://data.world/hdx/892f72f8-29b5-451e-8d27-7b89d1c45f63/workspace/file?filename=environment-indicators-for-india-1.csv

data.world

kaggle

Dataset Description

1) Agriculture crop production in India

Size: 5 files
60 columns

About: Production costs
Cultivation costs
Yield costs

data.gov.in
Open Government Data (OGD) Platform India

kaggle

Dataset Description

2) Rainfall in India

Size: 9 columns
4,116 rows

About: Annual Rainfall
Monthly & statewise readings

Type: ***.csv*** format

data.gov.in
Open Government Data (OGD) Platform India

kaggle

Dataset Description

3) Environment indicators in India

Size: 6 columns
3,675 rows

About: % usage of agricultural land
CO2 emissions
greenhouse gas emissions
year-wise % values

data.world

kaggle

Describing Dataset

```
1
2  file_csv = read.csv("/Users/akash/Downloads/rainfall in india 1901-2015.csv")
3
4  df <- data.frame(file_csv)
5  head(df,6)
6
7
8
9
10
```

11:1 (Top Level)

Console Terminal Jobs

~/

```
49    85.9   487.0  2419.4   456.7
50    85.5   742.2  1648.3   873.6
51   126.4   656.0  1604.4   693.3
52    47.3   166.0  1833.6   460.3
 [ reached 'max' / getOption("max.print") -- omitted 4064 rows ]
> head(df,6)
                    SUBDIVISION YEAR  JAN   FEB  MAR   APR   MAY   JUN   JUL   AUG   SEP   OCT   NOV   DEC ANNUAL
1 ANDAMAN & NICOBAR ISLANDS 1901 49.2  87.1 29.2   2.3 528.8 517.5 365.1 481.1 332.6 388.5 558.2  33.6 3373.2
2 ANDAMAN & NICOBAR ISLANDS 1902  0.0 159.8 12.2   0.0 446.1 537.1 228.9 753.7 666.2 197.2 359.0 160.5 3520.7
3 ANDAMAN & NICOBAR ISLANDS 1903 12.7 144.0  0.0   1.0 235.1 479.9 728.4 326.7 339.0 181.2 284.4 225.0 2957.4
4 ANDAMAN & NICOBAR ISLANDS 1904  9.4  14.7  0.0 202.4 304.5 495.1 502.0 160.1 820.4 222.2 308.7  40.1 3079.6
5 ANDAMAN & NICOBAR ISLANDS 1905  1.3   0.0  3.3  26.9 279.5 628.7 368.7 330.5 297.0 260.7  25.4 344.7 2566.7
6 ANDAMAN & NICOBAR ISLANDS 1906 36.6   0.0  0.0   0.0 556.1 733.3 247.7 320.5 164.3 267.8 128.9  79.2 2534.4
  Jan.Feb Mar.May Jun.Sep Oct.Dec
1   136.3   560.3  1696.3   980.3
2   159.8   458.3  2185.9   716.7
3   156.7   236.1  1874.0   690.6
4    24.1   506.9  1977.6   571.0
5     1.3   309.7  1624.9   630.8
6    36.6   556.1  1465.8   475.9
```

Machine Learning Algorithm

- K-Means Clustering
- K-Nearest Neighbor
- Random Forest
- Decision Tree