UNEMPLOYMENT IN INDIA DUE TO COVID-19

Danish Taufic Dharwadkar (2012)

Mackensie Fernandes (2017)

Vaibhavraj Kamlakar Verlekar (2059)

Goa University

Data Analytics (CS 636)

Instructor : S. Basker

Date : 04/05/2022

Abstract

Unemployment is a term referring to individuals who are employable and actively seeking a job but are unable to find a job, unemployment serves as one of the indicators of a country's economic status. The COVID-19 pandemic in India has very severely impacted, rather negatively, the employment figures of India since early 2020. While the pandemic has pushed several industries and small businesses into a tight spot, the current job crisis finds its roots in pre-pandemic times. Now this rising unemployment, fuelled by the pandemic-led lockdowns, threatens to translate into a "vicious cycle" of a severe economic downturn. From this project we hope to see what the effect of the pandemic has had on Indian employment.

Motivation

Our main motive in this project is to find how Covid-19 pandemic affected unemployment in India and how responsible it was for the increase of unemployment rates.

The Covid-19 pandemic has had a detrimental effect on the labor market worldwide causing many individuals to lose their jobs and businesses to close.

Many services such as school, gym, shops and movie theaters were forced to shut down in various states temporarily. This had a significant impact on economic activity and contributed to a rise in the unemployment rate.

However, the Covid-19 pandemic can be only partly blamed, since unemployment is one of the key problems in India.

Objectives

This project has the following objectives:

1. To examine the status of unemployment in India.
2. To study the unemployment rate of states during the pandemic.
3. To study the change in employment after the lockdown.

Data collection

Most of the data used in this study was collected from the Centre for Monitoring Indian Economy (CMIE).

CMIE produces this vitally important indicator for India as a public good. The rate of unemployment in India estimated by CMIE is for free public consumption by all.

Introduction of a reliable measure of the unemployment rate would help the domestic and global financial markets understand India's economic health better than it is understood in the absence of such a measure.

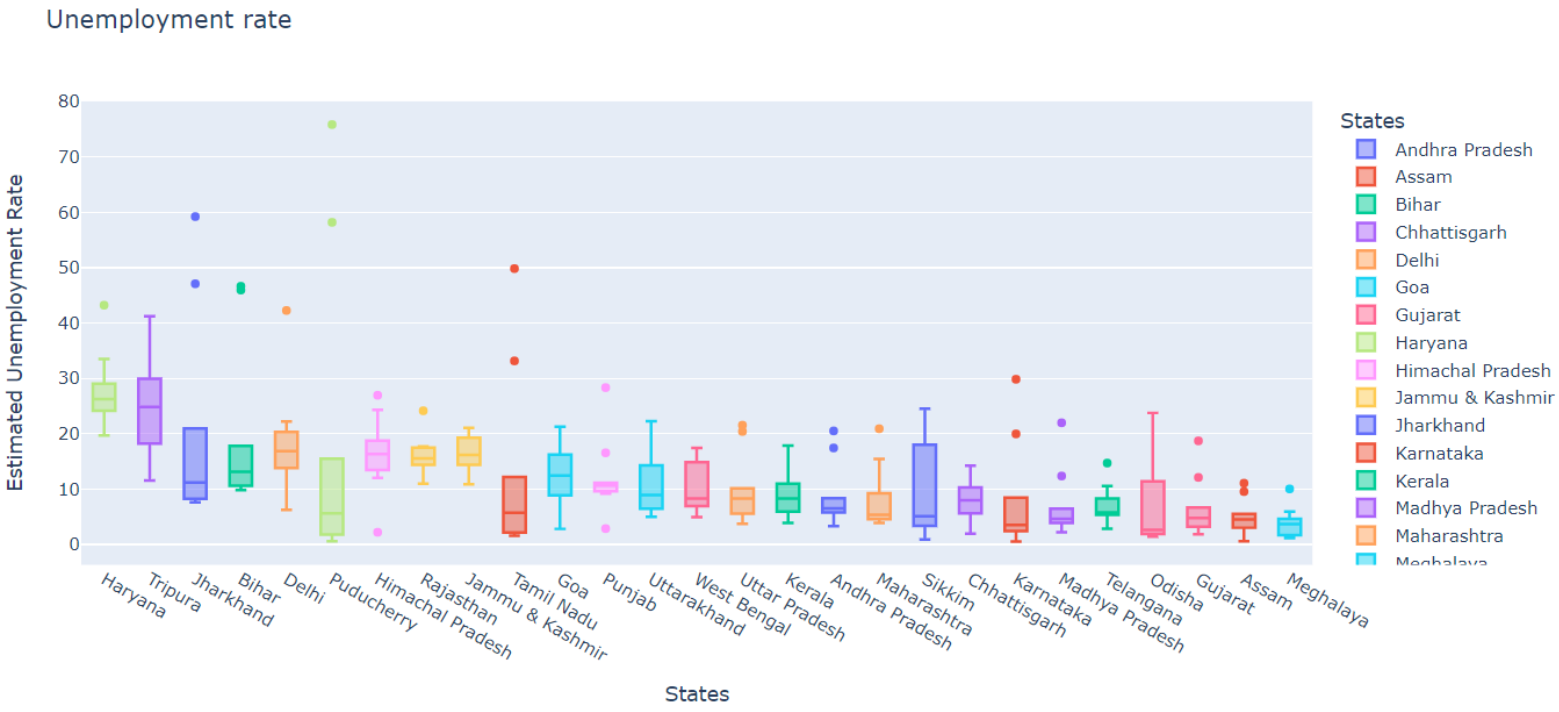
Exploratory Data Analysis

Link to Google Colab Workbook

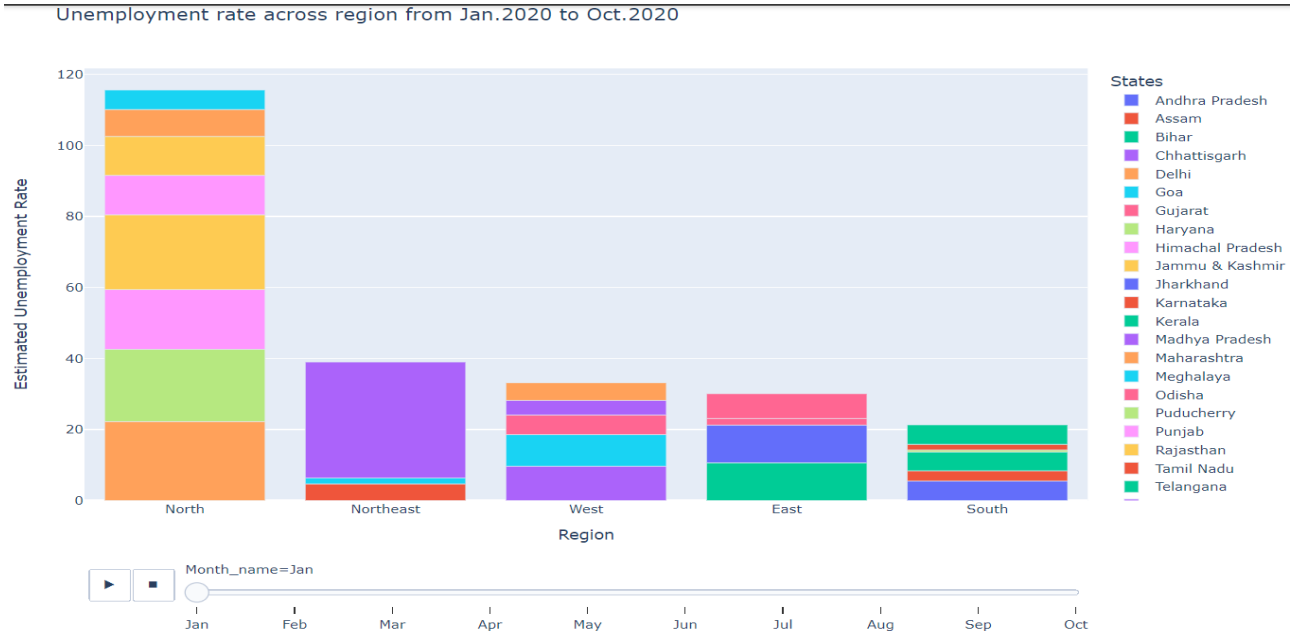
https://colab.research.google.com/drive/1sYI1eiHPUPNRS35xozdzRYry65lX4qLy?usp=sharing

Heatmap

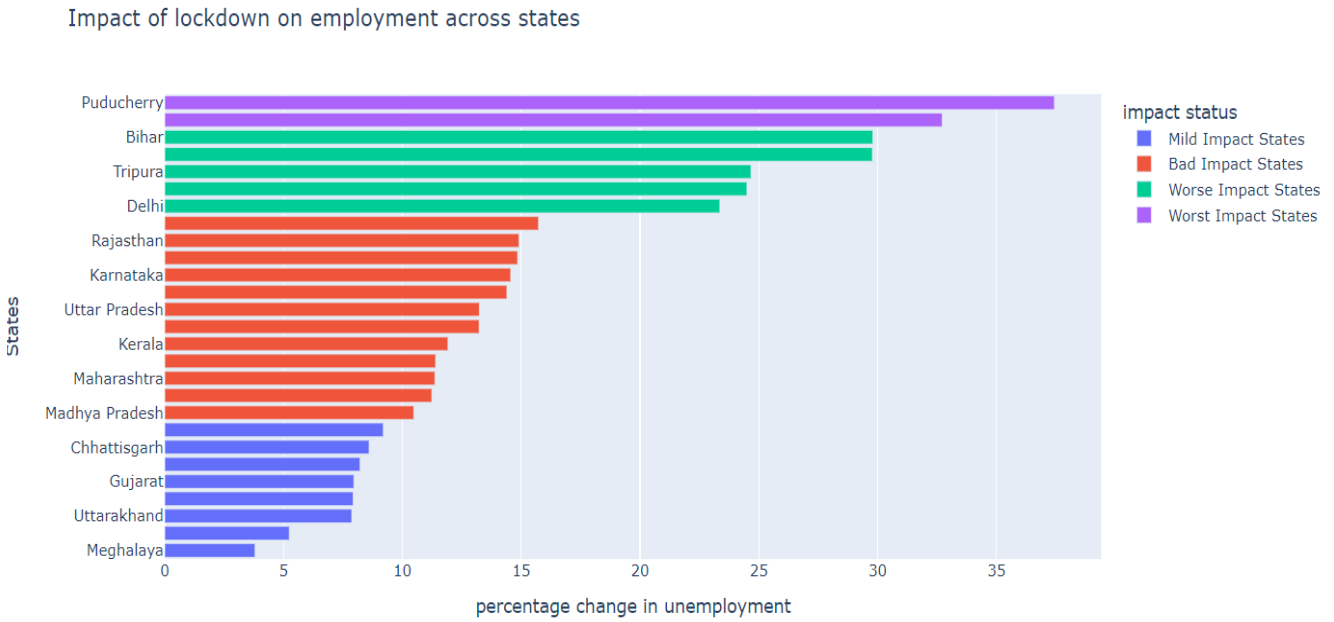
Boxplot

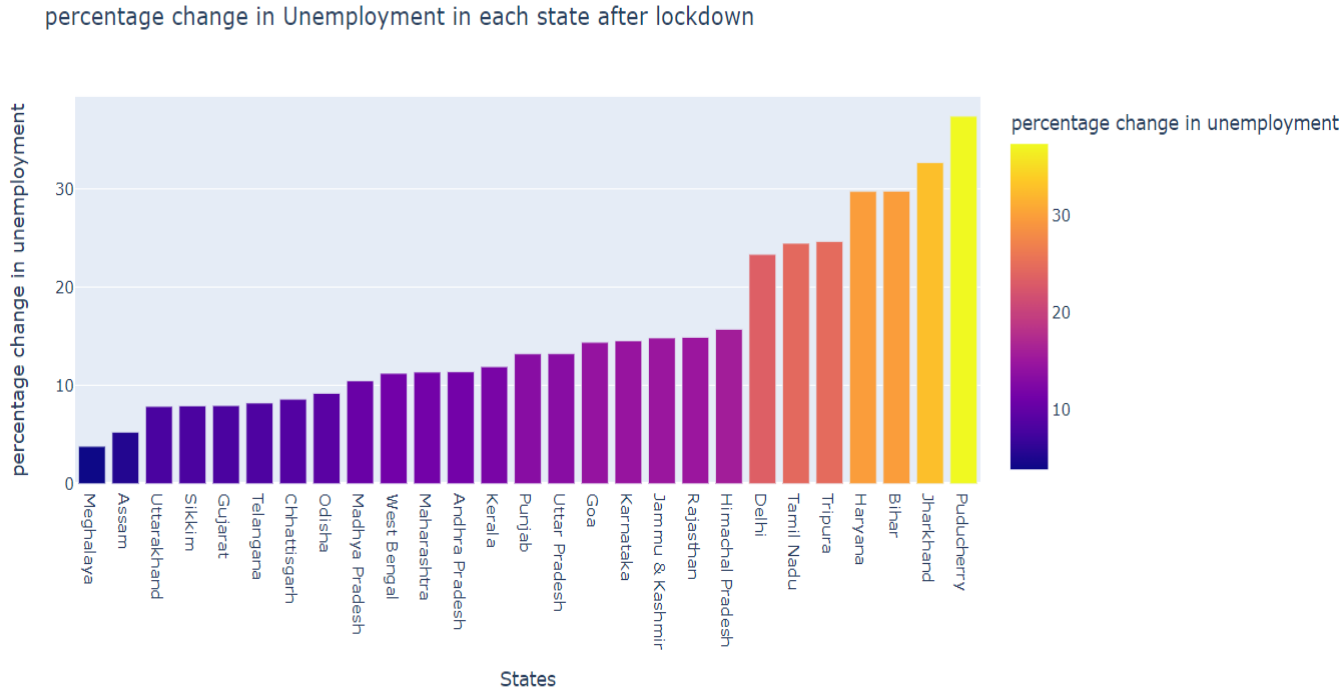


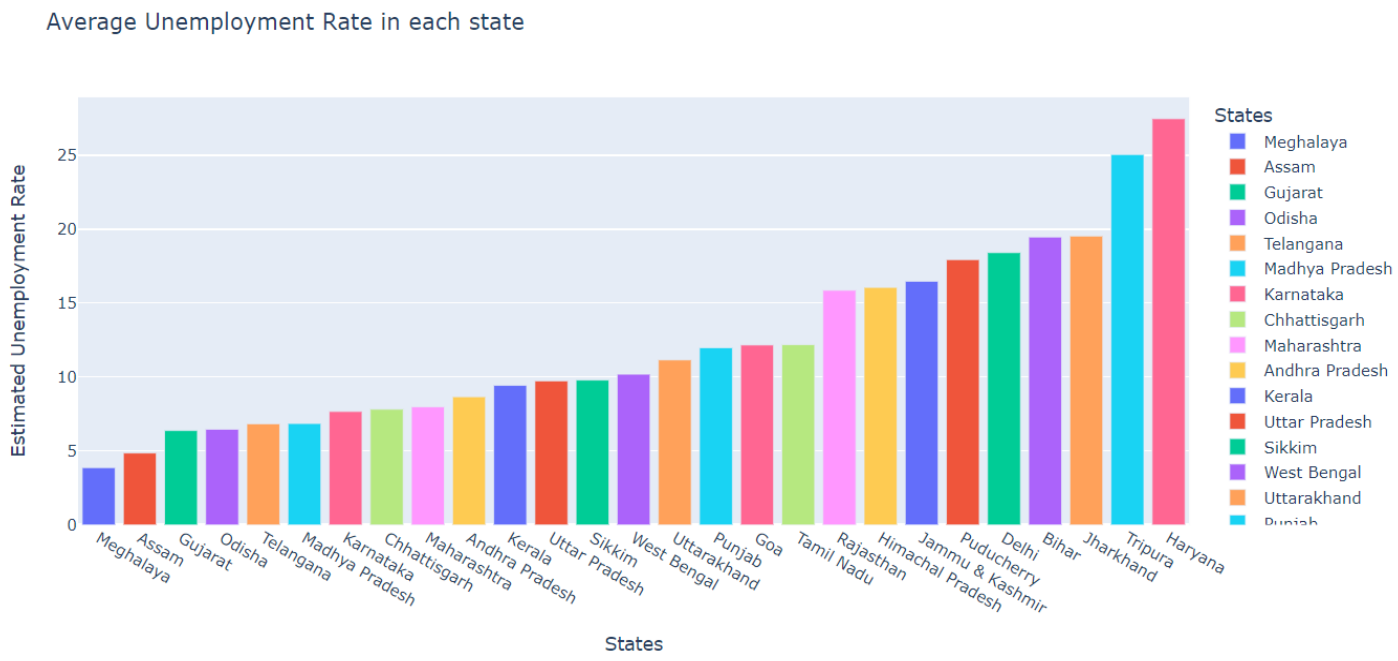
BoxPlot for unemployment rate across region



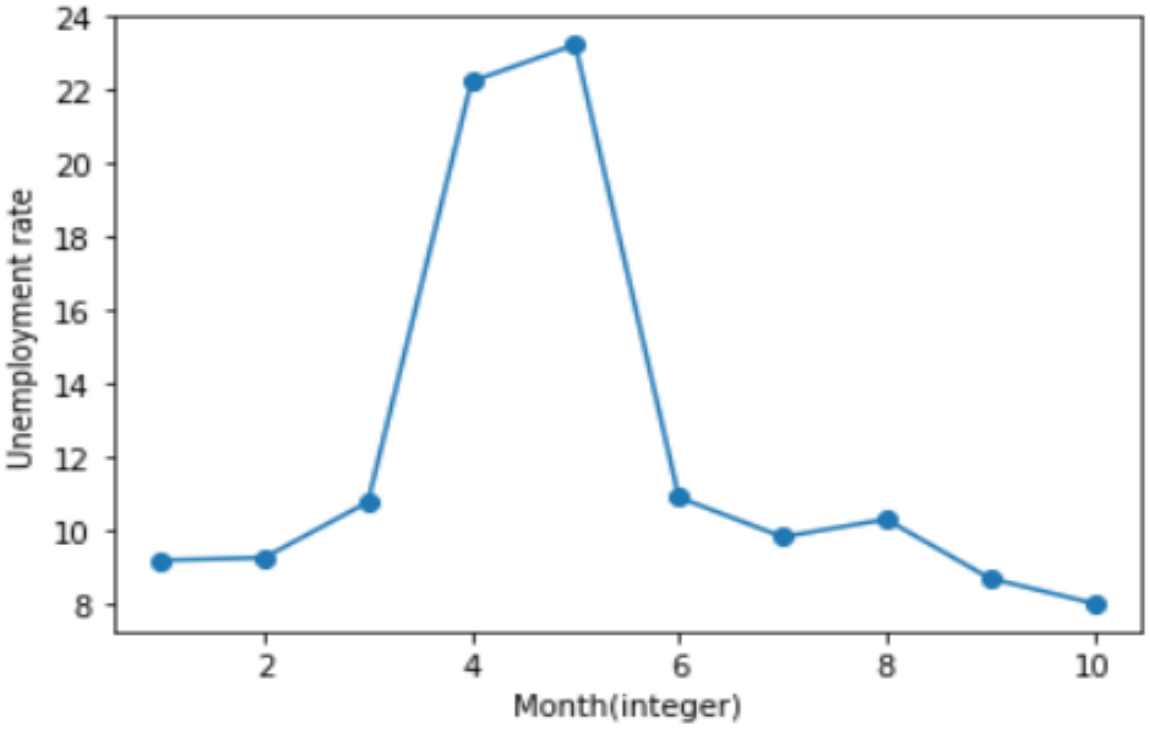
BarGraph


Impact of lockdown on employment across states
Puducherry
Bihar
Tripura
Delhi
Rajasthan
Karnataka
Uttar Pradesh
Kerala
Maharashtra
Madhya Pradesh
Chhattisgarh
Gujarat
Uttarakhand
Meghalaya
States
0
5
10
15
20
25
30
35
percentage change in unemployment
impact status
Mild Impact States
Bad Impact States
Worse Impact States
Worst Impact States


percentage change in Unemployment in each state after lockdown
percentage change in unemployment
0
10
20
30
Meghalaya
Assam
Uttarakhand
Sikkim
Gujarat
Telangana
Chhattisgarh
Odisha
Madhya Pradesh
West Bengal
Maharashtra
Andhra Pradesh
Kerala
Punjab
Uttar Pradesh
Goa
Karnataka
Jammu & Kashmir
Rajasthan
Himachal Pradesh
Delhi
Tamil Nadu
Tripura
Haryana
Bihar
Jharkhand
Puducherry
States
percentage change in unemployment
10
20
30



Graph representing unemployment through the months



Tensorflow Validation

IBM Quality Check

- Class Overlap : dc9d0dec-efd0-44d9-a5b7-c5671ee97543


```
      "total_overlap_fraction":
        "definition": "Total number of points in all overlapping regions / Total
number of points in the dataset",
        "value": 0.4

      "explanation": " Total overlapping fraction present in dataset: 0.4<br> Number
of overlapping rows: 107<br> Number of classes with overlap: 22<br> Class: Odisha has
highest amount of overlap i.e 9.35% of total overlap.",
      "score": 0.6,
      "title": "Class Overlap"
```


- Class Parity : d5e5fa13-2fa6-4c3c-bb91-83aad14c1e9c


```
    "method_details":
      "categorical_encoding": "nominal",
      "definition": "Identifies noise, overlap, size and sample counts in the data
to give a class parity score<br>The score is 1 when the ratio of majority class
samples to minority class samples is less than 70:30 regardless of any other issue
present in the dataset.",
      "hyperparameters": {},
      "name": "Class Parity",
      "type": "Quality"
    "runtime": 0.021294,
    "starting_timestamp": "2022-05-03 18:45:13",
    "user_input_details":
      "label_column": "Region"
  "results":
    "details":
    "explanation": " The dataset is balanced.",
    "score": 1,
    "title": "Class Parity"
```

- Correlation Detection : 8b297524-32f7-46cd-a3f9-eb13eed5b03e


```
    "method_details": {
        "definition": "Identifies correlated numerical columns in the data. A score of
1 indicates no correlated columns found in the data.",
        "hyperparameters":
          "correlation_threshold": 0.8
        "name": "Correlation Detection",
        "type": "Quality"
      "runtime": 0.011886,
      "starting_timestamp": "2022-05-03 18:49:51",
      "user_input_details": {
        "label_column": " Estimated Unemployment Rate (%)"
    "results": {
      "details": {
        "correlated_cols": []
      "explanation": "Correlated column pairs found in the data are 0 giving a quality
score of 1.0.",
      "score": 1,
      "title": "Correlation Detection"
```


- Data Completeness : d9d5ffab-bd35-463d-80a8-350657380575


```
    "method_details":
        "definition": "Identifies missing values in the given data. A score of 1
indicates no missing values found in the data.",
        "hyperparameters": {},
        "name": "Data Completeness",
        "type": "Quality"
      "runtime": 0.003033,
      "starting_timestamp": "2022-05-03 18:50:56",
      "user_input_details": {}
    "results":
      "details":
        "col_wise_score":
          " Date": 1,
          " Estimated Employed": 1,
          " Estimated Labour Participation Rate (%)": 1,
          " Estimated Unemployment Rate (%)": 1,
          " Frequency": 1,
          "Region": 1,
          "State": 1,
          "latitude": 1,
          "longitude": 1
        "missing_samples_index": []
      "explanation": "Missing values detected in 0 / 2403 entries giving a quality
score of 1.0.",
      "score": 1,
      "title": "Data Completeness"
```

- Data Duplicates : eac4e0ef-a239-4fb5-8ce6-af8224f9abd6


```
    "method_details":
        "definition": "Find duplicates in the data using all values in the record.
Quality score equals to 1 indicates no duplicates found.",
        "hyperparameters": {},
        "name": "Data Duplicates",
        "type": "Quality"
      "runtime": 0.003669,
      "starting_timestamp": "2022-05-03 18:51:57",
      "user_input_details": {}
    "results":
      "details":
        "duplicate_samples_index": []
      "explanation": "Duplicates detected in 0 / 267 rows giving a quality score of
1.0.",
      "score": 1,
      "title": "Data Duplicates"
```


- Data Homogeneity : 52c60203-7ed7-425c-b8fe-00bfa9f225af


```
    "method_details":
        "definition": "Identifies homogeneity in each column in the data. A score of 1
indicates that no Inhomogeneity is present in the given data.",
        "hyperparameters": {},
        "name": "Data Homogeneity",
        "type": "Quality"
    "explanation": "Number of columns analysed: 9<br>Number of columns with no
homogeneity issues: 8 ([' Date', ' Frequency', ' Estimated Unemployment Rate (%)', '
Estimated Employed', ' Estimated Labour Participation Rate (%)', 'Region',
'longitude', 'latitude'])<br>Number of columns with homogeneity issues: 1 (State)",
      "score": 0.96,
      "title": "Data Homogeneity"
```

- Data Profiler : b0548fbc-1258-428d-925f-c520cfee5009


```
"method_details":
        "definition": "Find (a) count of numerical/categorical/date columns present in
the dataset, (b) min and max values in numerical columns, (c) min and max string
length in categorical columns, and (d) count of unique columns.",
        "hyperparameters": {},
        "name": "Data Profiler",
        "type": "Profiling"
       "Max Categorical Column String Length": {
          " Frequency": 1,
          "Region": 9,
          "State": 16
        },
        "Max Numerical Column Value": {
          " Estimated Employed": 59433759,
          " Estimated Labour Participation Rate (%)": 69.69,
          " Estimated Unemployment Rate (%)": 75.85,
          "latitude": 92.9376,
          "longitude": 33.7782
        },
        "Min Categorical Column String Length": {
          " Frequency": 1,
          "Region": 4,
          "State": 3

        "Min Numerical Column Value": {
          " Estimated Employed": 117542,
          " Estimated Labour Participation Rate (%)": 16.77,
          " Estimated Unemployment Rate (%)": 0.5,
          "latitude": 71.1924,
          "longitude": 10.8505
        "Unique Columns": {
          " Date": {
            "is_unique": false,
            "num_unique_values": 10
          " Estimated Employed": {
            "is_unique": true,
            "num_unique_values": 267
          " Estimated Labour Participation Rate (%)": {
            "is_unique": false,
            "num_unique_values": 248
          " Estimated Unemployment Rate (%)": {
            "is_unique": false,
            "num_unique_values": 252
          " Frequency": {
            "is_unique": false,
            "num_unique_values": 1
          "Region": {
            "is_unique": false,
            "num_unique_values": 5
          "State": {
            "is_unique": false,
            "num_unique_values": 27
          "latitude": {
            "is_unique": false,
```

```
            "num_unique_values": 24
          "longitude": {
            "is_unique": false,
            "num_unique_values": 27
```


- Feature Relevance : 2f55ff8b-6f20-437b-bd70-a97ee2416780


```
    "method_details":
        "definition": "Indentifies and ranks the feature based on Relevance. A score
of 1 indicates that all features are relevant.",
        "hyperparameters":
        "confidence_threshold": 0.4
        "name": "Feature Relevance",
        "type": "Quality"
      "runtime": 0.474974,
      "starting_timestamp": "2022-05-03 18:55:51",
      "user_input_details": {}
    "results":
      "details":
        "high_relevant_features": "[0, 1, 4]",
        "less_relevant_features": "[]",
        "medium_relevant_features": "[5, 3, 7, 6, 2]",
        "ranked_feature_list": "[0, 1, 4, 5, 3, 7, 6, 2]"
      "explanation": "Less relevant features found in the dataset are 0 / 8 giving a
quality score of 1.0",
      "score": 1,
      "title": "Feature Relevance"
```


- Outlier Detection : 54023967-af48-4fd2-b3d2-6f8093620935


```
    "method_details":
        "categorical_encoding": "nominal",
        "definition": "Identifies outlier samples in the data. A score of 1 indicates
no outliers found in the data.",
        "hyperparameters": {},
        "name": "Outlier Detection",
        "type": "Quality"
      "runtime": 0.023031,
      "starting_timestamp": "2022-05-03 18:59:22",
      "user_input_details":
        "algorithm": "LocalOutlierFactor",
        "label_column": " Estimated Employed"
    "results":
      "details":
        "outlier_samples_index": [
          172
      "explanation": "Outliers detected in 1 / 267 rows giving a quality score of
0.99.",
      "score": 0.99,
      "title": "Outlier Detection"
```

Model Built

```
[ ] test.shape

    (267, 5)

[ ] X = test.drop('Estimated Unemployment Rate',axis=1)
    y = test['Estimated Unemployment Rate']


[ ] from sklearn.model_selection import train_test_split
    train_x, test_x,train_y,test_y = train_test_split(X,y,test_size=0.3,random_state=1)

[ ] from sklearn.linear_model import LinearRegression
    reg = LinearRegression().fit(train_x,train_y)

[ ] reg.score(test_x,test_y)

    0.0826993255899664

[ ] reg.score(train_x, train_y)

    0.1253100242154629

[ ] from sklearn.linear_model import Ridge
    ridge_reg=Ridge(alpha=50,max_iter=100,tol=0.1)
    ridge_reg.fit(train_x,train_y)

    Ridge(alpha=50, max_iter=100, tol=0.1)

[ ] ridge_reg.score(test_x, test_y)

    0.08779962791335105

[ ] ridge_reg.score(train_x,train_y)

    0.12419034644922844
```

Inspiring Questions Asked and insights obtained.

1. In which month of the year unemployment was at its peak?
2. Which state was affected the most by the Pandemic?
3. Unemployment before and after the first lockdown.

Conclusion

India is a developing country, moving on the path of progression. It is necessary, in this process that available resources should be used to the full extent possible.

Unemployment is a grave problem in India, while a single unemployed person has minimal influence on society, high unemployment rates in some states contribute to increased poverty rates and poorer neighborhoods, exacerbating the socioeconomic impact of unemployment.

These regions with high unemployment rates are more likely to have restricted career opportunities, inadequate housing, fewer recreational activities accessible, as well as limited access to public services and underfunded schools.

With more people being unemployed, more individuals earning less income. Thus, they will spend less money, resulting in decreased economic contribution in terms of services and products supplied and produced.

Nonetheless, steps must be taken to increase efficiency and boost living conditions for sustainable economic growth by the government.

References

Dataset from, https://unemploymentinindia.cmie.com/