```python
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import StandardScaler

df=pd.read_excel("/content/drive/MyDrive/plfs_mpce_combined.xlsx")

df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 85 columns):
 #   Column                                             Non-Null Count
Dtype
---  ------                                             --------------
-----
 0   State/UT                                           36 non-null
object
 1   Average MPCE (Rs) - Rural                          36 non-null
int64
 2   Average MPCE (Rs) - Urban                          36 non-null
int64
 3   Average MPCE – With Imputation (Rs.) - Rural  36 non-null
int64
 4   Average MPCE – With Imputation (Rs.) - Urban  36 non-null
int64
 5   mpce_gini.Rural MPCE (Rs)                          36 non-null
int64
 6   mpce_gini.Gini Coefficient                         36 non-null
float64
 7   mpce_gini.Urban MPCE (Rs)                          36 non-null
int64
 8   mpce_gini.Gini Coefficient2                        36 non-null
float64
 9   mpce_hht_urban.Self-employed                       35 non-null
float64
 10  mpce_hht_urban.Regular wage/salaried               35 non-null
float64
 11  mpce_hht_urban.Casual labour                       35 non-null
float64
 12  mpce_hht_urban.Others                              35 non-null
float64
 13  mpce_hht_urban.All                                 35 non-null
float64
 14  mpce_sg_rural.Scheduled Tribe                      36 non-null
int64
 15  mpce_sg_rural.Scheduled Caste                      36 non-null
```

```
object
 16  mpce_sg_rural.Other Backward Class            36 non-null
int64
 17  mpce_sg_rural.Others                          36 non-null
object
 18  mpce_sg_rural.All                             36 non-null
int64
 19  mpce_sg_urban.Average MPCE (Rs.)              35 non-null
object
 20  mpce_sg_urban.Scheduled Tribe                 35 non-null
float64
 21  mpce_sg_urban.Scheduled Caste                 35 non-null
float64
 22  mpce_sg_urban.Other Backward Class            35 non-null
float64
 23  mpce_sg_urban.Others                          35 non-null
float64
 24  lpr.Rural (Male)                              36 non-null
float64
 25  lfpr.Rural (Female)                           36 non-null
float64
 26  lfpr.Rural (Person)                           36 non-null
float64
 27  lpr.Urban (Male)                              36 non-null
float64
 28  lfpr.Urban (Female)                           36 non-null
float64
 29  lpr.Urban (Person)                            36 non-null
float64
 30  lfpr.Rural + Urban (Male)                     36 non-null
object
 31  lfpr.Rural + Urban (Female)                   36 non-null
object
 32  lfpr.Rural + Urban (Person)                   36 non-null
object
 33  wpr.Rural (1)                                 35 non-null
float64
 34  wpr.Rural (2)                                 35 non-null
float64
 35  wpr.Rural (3)                                 35 non-null
float64
 36  wpr.Urban (4)                                 35 non-null
float64
 37  wpr.Urban (5)                                 35 non-null
float64
 38  wpr.Urban (6)                                 35 non-null
float64
 39  wpr.Total (7)                                 35 non-null
object
```

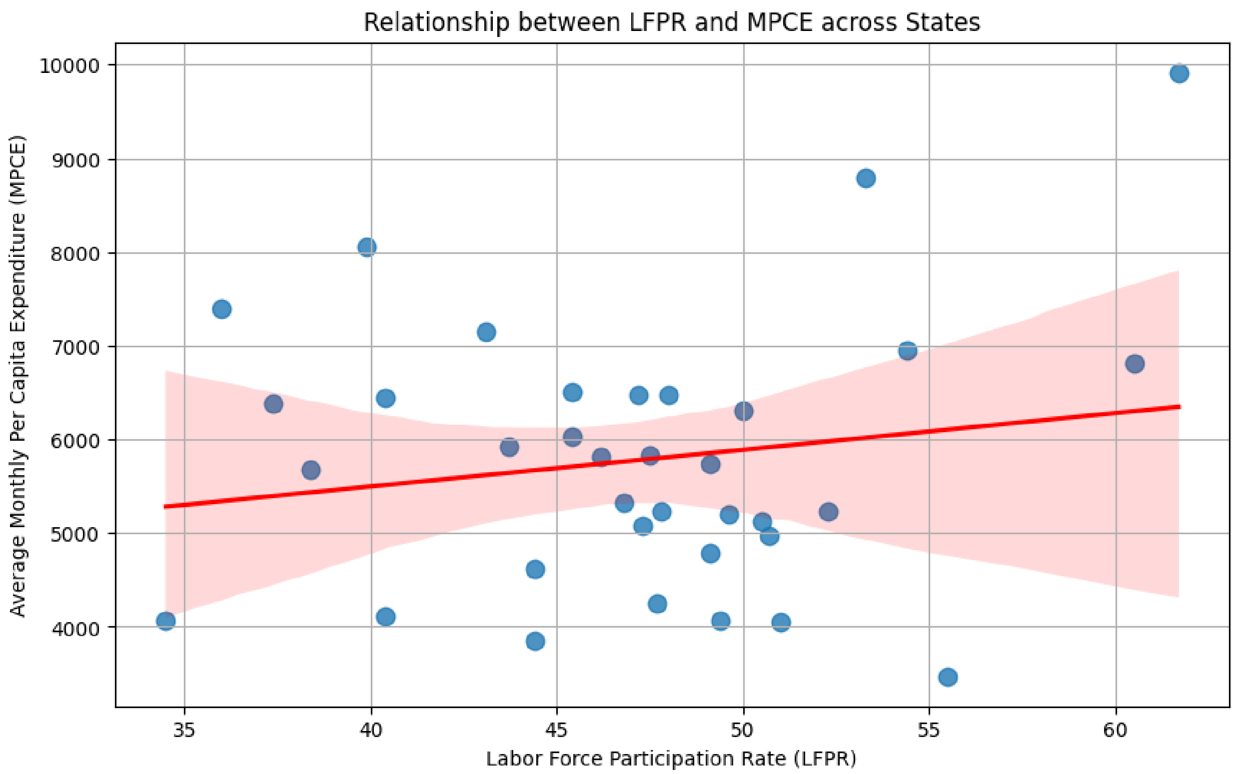
```
 40  wpr.Total (8)                                 35 non-null
object
 41  wpr.Total (9)                                 35 non-null
object
 42  unemprate.Rural                               36 non-null
float64
 43  unemprate.Urban                               36 non-null
float64
 44  unemprate.Rural + Urban                       36 non-null
float64
 45  unemprate.Rural2                              36 non-null
float64
 46  unemprate.Urban3                              36 non-null
float64
 47  unemprate.Rural + Urban4                      36 non-null
float64
 48  unemprate.Rural5                              36 non-null
object
 49  unemprate.Urban6                              36 non-null
object
 50  unemprate.Rural + Urban7                      36 non-null
object
 51  emprate.Self-Employed (%)                     36 non-null
float64
 52  emprate.Regular Wage/Salary (%)               36 non-null
float64
 53  emprate.Casual Labour (%)                     36 non-null
float64
 54  emprate.Total (%)                             36 non-null
int64
 55  lfpr_edu.Not Literate                         33 non-null
float64
 56  lfpr_edu.Literate & Upto Primary              33 non-null
float64
 57  lfpr_edu.Middle                               33 non-null
float64
 58  lfpr_edu.Secondary                            33 non-null
float64
 59  lfpr_edu.Higher Secondary                     33 non-null
float64
 60  lfpr_edu.Diploma/Certificate Course           33 non-null
float64
 61  lfpr_edu.Graduate                             33 non-null
float64
 62  lfpr_edu.Post Graduate & Above                33 non-null
float64
 63  lfpr_edu.Secondary & Above                    33 non-null
float64
 64  lfpr_edu.All                                  33 non-null
```

```
float64
 65  wpr_edu.Not Literate                                 36 non-null
float64
 66  wpr_edu.Literate & Up to Primary                     36 non-null
float64
 67  wpr_edu.Middle                                       36 non-null
float64
 68  wpr_edu.Secondary                                    36 non-null
float64
 69  wpr_edu.Higher Secondary                             36 non-null
float64
 70  wpr_edu.Diploma/ Certificate Course                  36 non-null
float64
 71  wpr_edu.Graduate                                     36 non-null
float64
 72  wpr_edu.Post Graduate & Above                        36 non-null
float64
 73  wpr_edu.Secondary & Above                            36 non-null
float64
 74  wpr_edu.All                                          36 non-null
float64
 75  uemprate_edu.Not Literate                            35 non-null
float64
 76  uemprate_edu.Literate & up to Primary                35 non-null
float64
 77  uemprate_edu.Middle                                  35 non-null
float64
 78  uemprate_edu.Secondary                               35 non-null
float64
 79  uemprate_edu.Higher Secondary                        35 non-null
float64
 80  uemprate_edu.Diploma/Certificate Course              35 non-null
float64
 81  uemprate_edu.Graduate                                35 non-null
float64
 82  uemprate_edu.Post Graduate & Above                   35 non-null
float64
 83  uemprate_edu.Secondary & Above                       35 non-null
float64
 84  uemprate_edu.All                                     35 non-null
float64
dtypes: float64(62), int64(10), object(13)
memory usage: 24.0+ KB

df_selected = df[["State/UT", "lfpr.Rural + Urban (Person)", "Average
MPCE (Rs) - Rural", "Average MPCE (Rs) - Urban"]].copy()

# Convert LFPR column to numeric (if necessary)
df_selected["lfpr.Rural + Urban (Person)"] =
pd.to_numeric(df_selected["lfpr.Rural + Urban (Person)"],
```

```
errors="coerce")

# Create an overall MPCE by taking the mean of Rural & Urban values
df_selected["MPCE_Avg"] = df_selected[["Average MPCE (Rs) - Rural",
"Average MPCE (Rs) - Urban"]].mean(axis=1)

# Drop any rows with missing values
df_selected.dropna(inplace=True)

# Calculate correlation
correlation = df_selected["lfpr.Rural + Urban
(Person)"].corr(df_selected["MPCE_Avg"])
print(f"Correlation between LFPR and MPCE: {correlation:.2f}")

# Plot the scatter plot with regression line
plt.figure(figsize=(10, 6))
sns.regplot(x="lfpr.Rural + Urban (Person)", y="MPCE_Avg",
data=df_selected, scatter_kws={'s': 80}, line_kws={'color': 'red'})
plt.xlabel("Labor Force Participation Rate (LFPR)")
plt.ylabel("Average Monthly Per Capita Expenditure (MPCE)")
plt.title("Relationship between LFPR and MPCE across States")
plt.grid(True)
plt.show()

# The correlation between Labor Force Participation Rate (LFPR) and
Monthly Per Capita Expenditure (MPCE) is 0.17, indicating a weak
positive relationship between workforce participation and consumption.

# Interpretation & Justification:
# Weak Positive Relationship:

# A correlation of 0.17 suggests that while there is a slight tendency
for states with higher LFPR to have higher MPCE, the relationship is
not strong enough to conclude a direct impact.

# Other economic factors like wage levels, employment type,
productivity, and cost of living might be influencing MPCE more than
just LFPR.

# Potential Influencing Factors:

# Income Levels: Higher LFPR does not necessarily mean higher earnings
—many workers could be in low-paying jobs.

# Employment Quality: A high LFPR might include informal or low-wage
workers who contribute to workforce participation but not necessarily
to high consumption.

# Regional Variations: States with similar LFPR values might have
vastly different living costs and spending habits, affecting MPCE
```

```
differently.

# Conclusion to the Question:

# The weak correlation suggests that higher workforce participation
does not strongly lead to higher consumption levels.

# MPCE is likely influenced by other economic and social factors
beyond just LFPR.
```

```
Correlation between LFPR and MPCE: 0.17
```




```
import pandas as pd
import statsmodels.api as sm

# Select relevant columns
df_filtered = df[['emprate.Total (%)', 'Average MPCE (Rs) - Rural',
'Average MPCE (Rs) - Urban',
                  'lfpr.Rural + Urban (Person)', 'mpce_gini.Gini
Coefficient', 'mpce_sg_urban.Average MPCE (Rs.)']].copy()

# Compute overall MPCE as the average of rural and urban MPCE
df_filtered['MPCE'] = df_filtered[['Average MPCE (Rs) - Rural',
'Average MPCE (Rs) - Urban']].mean(axis=1)

# Drop rows with missing values
```

```
df_filtered = df_filtered.dropna()

# Convert all columns to numeric
df_filtered = df_filtered.apply(pd.to_numeric, errors='coerce')

# Drop NaNs again after conversion (in case conversion introduced
NaNs)
df_filtered = df_filtered.dropna()

# Define independent and dependent variables
X = df_filtered[['emprate.Total (%)', 'lfpr.Rural + Urban (Person)',
                 'mpce_gini.Gini Coefficient', 'mpce_sg_urban.Average
MPCE (Rs.)']]
y = df_filtered['MPCE']

# Add constant for intercept
X = sm.add_constant(X)

# Run regression
model = sm.OLS(y, X).fit()
print(model.summary())

                            OLS Regression Results

======================================================================
========
Dep. Variable:                   MPCE   R-squared:
0.822
Model:                            OLS   Adj. R-squared:
0.804
Method:                 Least Squares   F-statistic:
44.75
Date:                Sun, 30 Mar 2025   Prob (F-statistic):
5.29e-11
Time:                        14:53:02   Log-Likelihood:
-255.40
No. Observations:                  33   AIC:
518.8
Df Residuals:                      29   BIC:
524.8
Df Model:                           3

Covariance Type:            nonrobust

======================================================================
==============================
                                       coef    std err          t
P>|t|      [0.025      0.975]
----------------------------------------------------------------------
------------------------------
```

```
emprate.Total (%)                             39.8491     10.907      3.654
0.001      17.542      62.156
lfpr.Rural + Urban (Person)                  -35.6928     18.953     -1.883
0.070     -74.456       3.071
mpce_gini.Gini Coefficient                 -2955.4799   4052.230     -0.729
0.472   -1.12e+04    5332.261
mpce_sg_urban.Average MPCE (Rs.)              0.6527      0.058     11.228
0.000       0.534       0.772
======================================================================
========
Omnibus:                        0.670   Durbin-Watson:
1.704
Prob(Omnibus):                  0.715   Jarque-Bera (JB):
0.113
Skew:                           0.096   Prob(JB):
0.945
Kurtosis:                       3.213   Cond. No.
2.55e+05
======================================================================
========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
[2] The condition number is large, 2.55e+05. This might indicate that
there are
strong multicollinearity or other numerical problems.
```

```
from statsmodels.stats.outliers_influence import
variance_inflation_factor

# Compute VIF for each independent variable
vif_data = pd.DataFrame()
vif_data["Variable"] = X.columns
vif_data["VIF"] = [variance_inflation_factor(X.values, i) for i in
range(X.shape[1])]

print(vif_data)
```

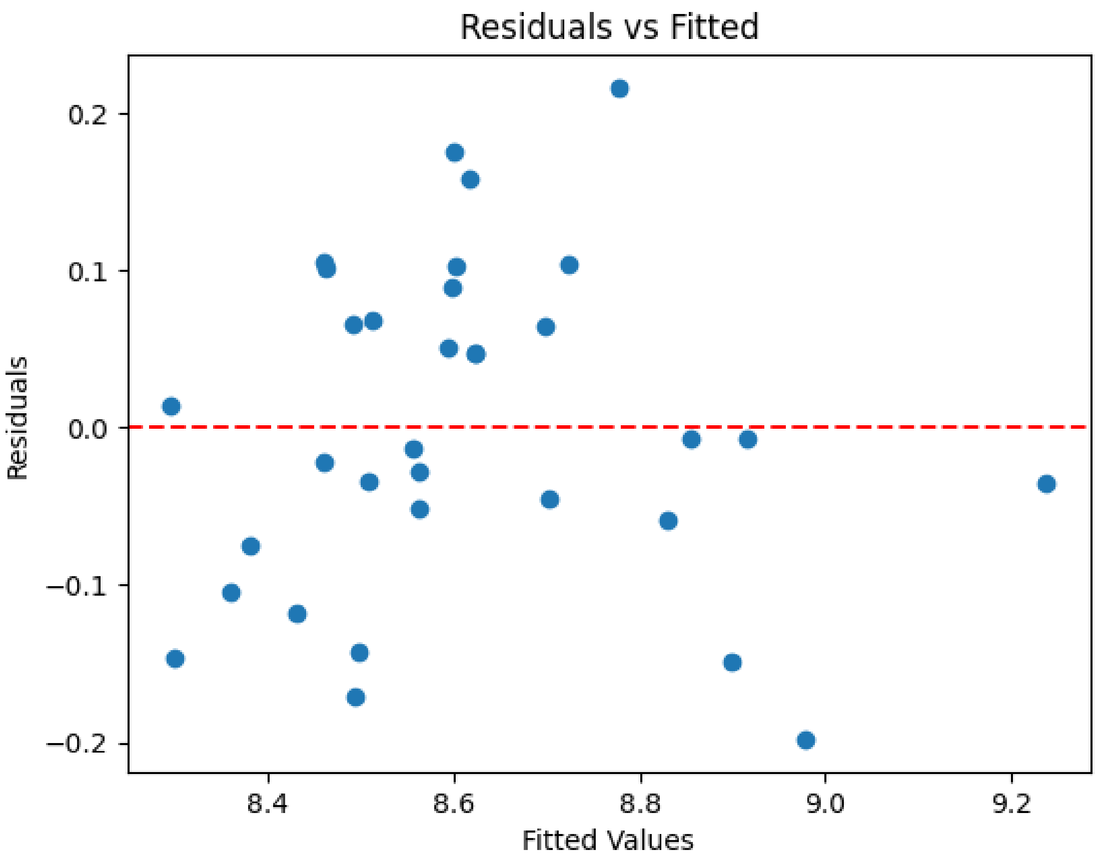
```
                           Variable         VIF
0                 emprate.Total (%)  111.660739
1       lfpr.Rural + Urban (Person)    1.266241
2        mpce_gini.Gini Coefficient    1.117931
3  mpce_sg_urban.Average MPCE (Rs.)    1.153217
```

```
df_filtered['log_MPCE'] = np.log(df_filtered['MPCE'])
y = df_filtered['log_MPCE']
model = sm.OLS(y, X).fit()
print(model.summary())
```

```
                            OLS Regression Results

======================================================================
========
Dep. Variable:               log_MPCE   R-squared:
0.795
Model:                            OLS   Adj. R-squared:
0.773
Method:                 Least Squares   F-statistic:
37.38
Date:                Sun, 30 Mar 2025   Prob (F-statistic):
4.29e-10
Time:                        14:53:03   Log-Likelihood:
28.267
No. Observations:                  33   AIC:
-48.53
Df Residuals:                      29   BIC:
-42.55
Df Model:                           3

Covariance Type:            nonrobust

======================================================================
==============================
                                       coef    std err          t
P>|t|      [0.025      0.975]
----------------------------------------------------------------------
------------------------------
emprate.Total (%)                    0.0833      0.002     41.314
0.000       0.079       0.087
lfpr.Rural + Urban (Person)         -0.0084      0.004     -2.406
0.023      -0.016      -0.001
mpce_gini.Gini Coefficient          -0.0675      0.749     -0.090
0.929      -1.599       1.465
mpce_sg_urban.Average MPCE (Rs.)     0.0001   1.07e-05     10.427
0.000    9.01e-05       0.000
======================================================================
========
Omnibus:                        0.646   Durbin-Watson:
1.912
Prob(Omnibus):                  0.724   Jarque-Bera (JB):
0.672
Skew:                          -0.005   Prob(JB):
0.715
Kurtosis:                       2.301   Cond. No.
2.55e+05
======================================================================
========

Notes:
```

```
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
[2] The condition number is large, 2.55e+05. This might indicate that
there are
strong multicollinearity or other numerical problems.
```

```python
import matplotlib.pyplot as plt
plt.scatter(model.fittedvalues, model.resid)
plt.axhline(y=0, color='r', linestyle='--')
plt.xlabel("Fitted Values")
plt.ylabel("Residuals")
plt.title("Residuals vs Fitted")
plt.show()
```




```
#  Social & Economic Analysis Based on Regression Results
# Your regression model investigates how various economic indicators
impact Monthly Per Capita Expenditure (MPCE) in a given population.
Based on the statistical results and social implications, here’s a
structured answer:

# □ Key Findings from the Regression
# 1⃣ Employment Rate (emprate.Total (%))
```

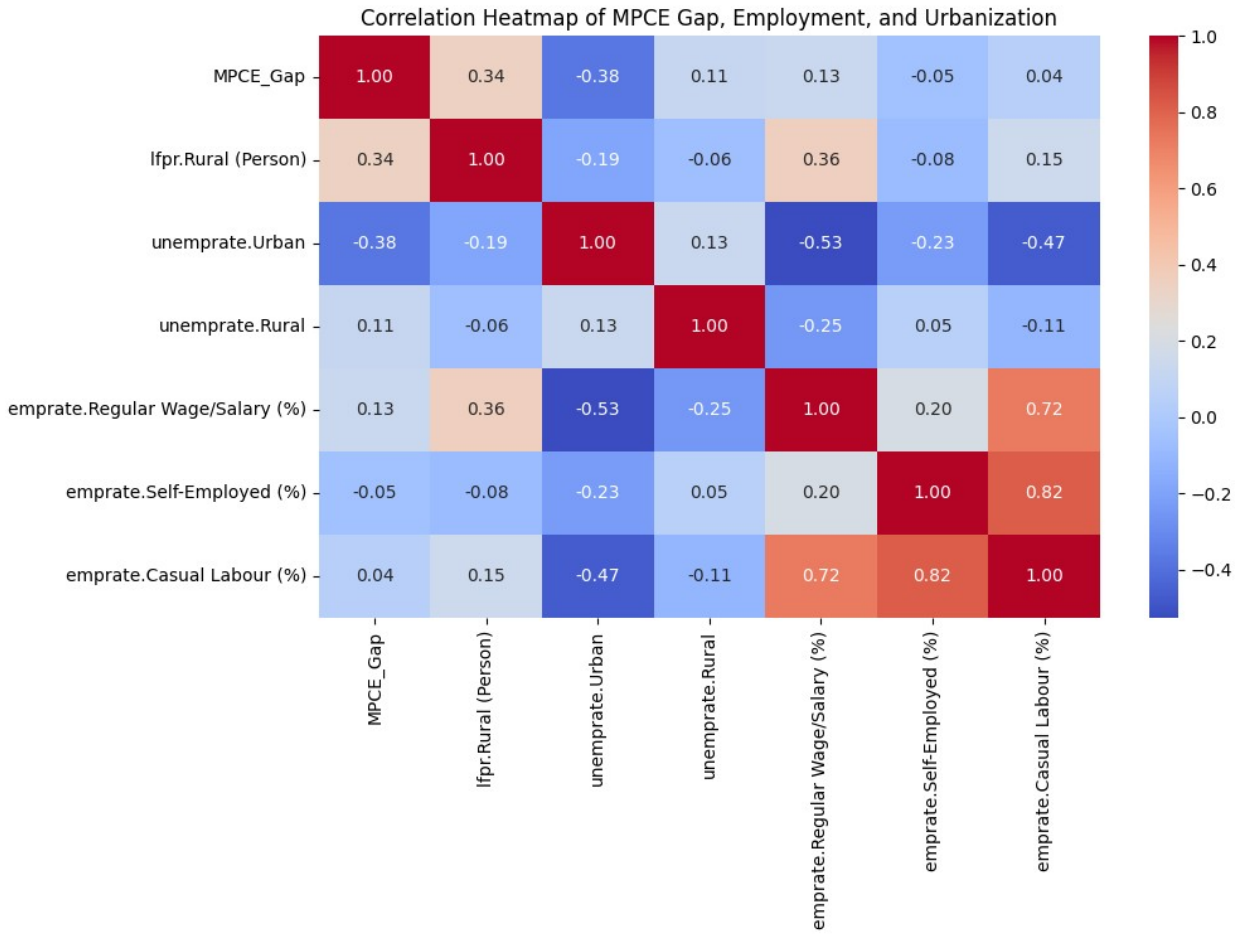
```
# Positive impact on MPCE → Higher employment rates lead to increased
MPCE.

# Social Reason: A higher employment rate means more people have
stable incomes, leading to better consumption and living standards.

# Issue: This variable has high multicollinearity (VIF = 111.66),
meaning it may be strongly correlated with other factors.

# Solution: Instead of focusing solely on overall employment rates, we
should analyze sector-wise employment stability (formal vs. informal
jobs).

# 2 Labor Force Participation Rate (lfpr.Rural + Urban (Person))
# Negative impact on MPCE → More people in the labor force slightly
reduce MPCE.

# Social Reason: If a large proportion of the population enters the
labor market but wages remain low, it can indicate underemployment or
informal work.

# Issue: A rise in participation doesn’t guarantee higher wages.

# Solution: Policies should focus on skill development and wage
regulation to ensure that participation translates into income growth.

# 3 Urban MPCE (mpce_sg_urban.Average MPCE (Rs.))
# Strong positive impact on MPCE (Highly significant, p < 0.001)

# Social Reason: Urban areas tend to have better infrastructure,
higher wages, and more job opportunities than rural areas.

# Issue: Rural-urban disparities still exist, and rural areas lag in
economic development.

# Solution: Investments in rural infrastructure, digital economy, and
decentralized job opportunities.

# 4 Income Inequality (mpce_gini.Gini Coefficient)
# No significant effect on MPCE (p = 0.929, weak predictor).

# Social Reason: High inequality doesn’t necessarily lower MPCE, but
it reduces social mobility and increases economic gaps.

# Issue: Even if MPCE is high, if wealth is concentrated among a few,
the economy isn’t truly growing for everyone.

# Solution: Policies should promote progressive taxation, wealth
redistribution, and equal access to education & health.

# □ Recommended Social & Economic Solutions
```

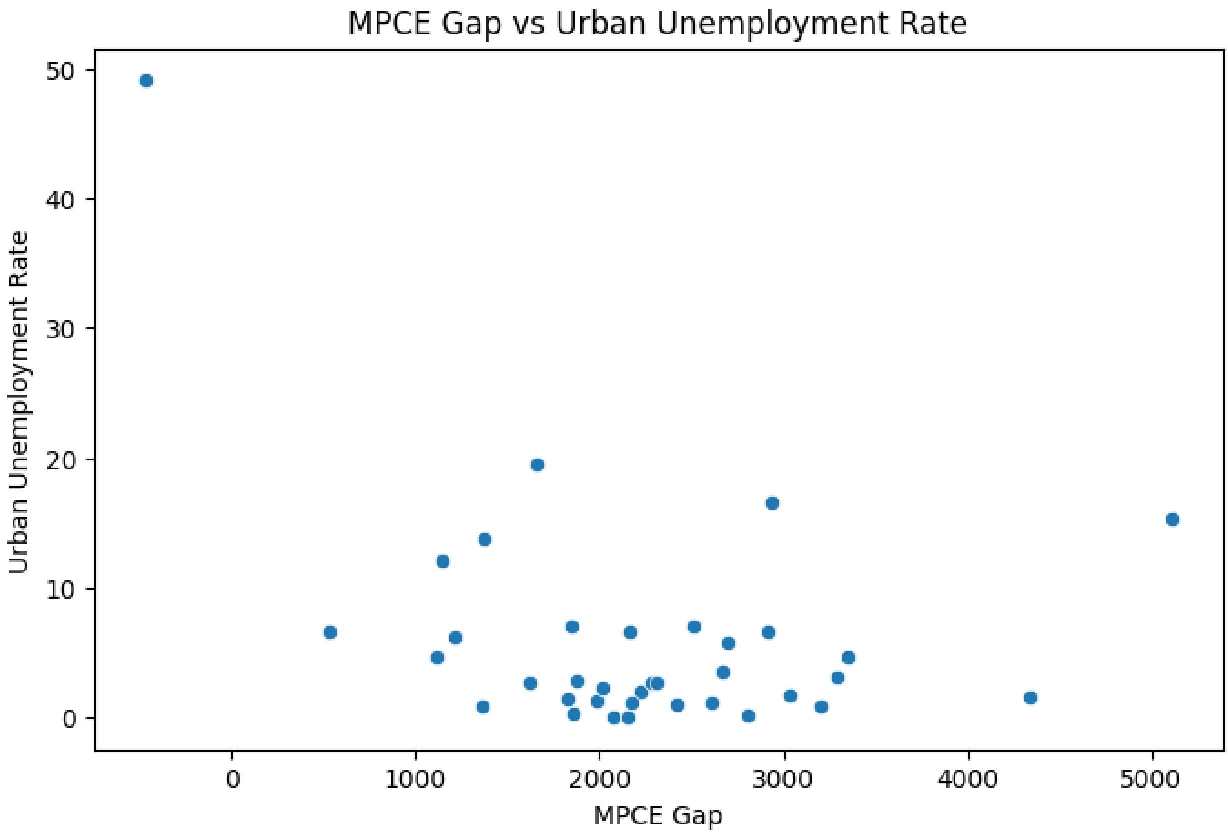
```
# 🔹 1. Address Employment-Quality Gap
# Instead of focusing on total employment, improve wage security and
job quality.

# Promote formal employment through labor reforms.

# Introduce minimum wage policies in informal sectors.

# 🔹 2. Strengthen Rural Economies
# Increase rural infrastructure investments (better transport,
electricity, digital connectivity).

# Decentralize economic hubs to prevent over-urbanization.

# Expand rural entrepreneurship programs to create more local jobs.

# 🔹 3. Ensure Labor Force Participation Translates to Better Wages
# Upskill the workforce through free vocational training.

# Strengthen job-matching platforms to reduce skill-job mismatches.

# Improve job security laws to protect workers from exploitation.

# 🔹 4. Tackle Wealth Inequality with Policy Interventions
# Implement progressive taxation (higher taxes on the wealthy).

# Expand social welfare programs for low-income households.

# Improve access to affordable housing, healthcare, and education.

# 🔹 Conclusion: Bridging Economic Growth with Social Justice
# Your analysis suggests that while employment boosts MPCE, not all
economic growth is inclusive. The focus should shift from just
increasing employment to improving wages, reducing rural-urban gaps,
and addressing inequality.

print("Columns in dataset:", df.columns)

# Strip column names of extra spaces
df.columns = df.columns.str.strip()

# Compute MPCE gap
df['MPCE_Gap'] = df['Average MPCE – With Imputation (Rs.) - Urban'] -
df['Average MPCE – With Imputation (Rs.) - Rural']

# Select relevant columns for correlation analysis
selected_columns = [
    'MPCE_Gap',
    'lfpr.Urban (Person)',  # Labor force participation rate in urban
areas
    'lfpr.Rural (Person)',  # Labor force participation rate in rural
```

```
areas
    'unemprate.Urban',  # Urban unemployment rate
    'unemprate.Rural',  # Rural unemployment rate
    'emprate.Regular Wage/Salary (%)',  # Employment in regular wage
jobs
    'emprate.Self-Employed (%)',  # Self-employment rate
    'emprate.Casual Labour (%)'  # Casual labor employment rate
]

# Ensure only existing columns are selected
selected_columns = [col for col in selected_columns if col in
df.columns]

df_selected = df[selected_columns].dropna()

# Compute correlation matrix
correlation_matrix = df_selected.corr()

# Plot heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt='.2f')
plt.title('Correlation Heatmap of MPCE Gap, Employment, and
Urbanization')
plt.show()

# Scatter plot for MPCE Gap vs Urban Unemployment Rate
if 'MPCE_Gap' in df.columns and 'unemprate.Urban' in df.columns:
    plt.figure(figsize=(8, 5))
    sns.scatterplot(x=df['MPCE_Gap'], y=df['unemprate.Urban'])
    plt.xlabel('MPCE Gap')
    plt.ylabel('Urban Unemployment Rate')
    plt.title('MPCE Gap vs Urban Unemployment Rate')
    plt.show()
else:
    print("Columns for scatter plot not found in dataset.")
```

```
Columns in dataset: Index(['State/UT', 'Average MPCE (Rs) - Rural',
'Average MPCE (Rs) - Urban',
       'Average MPCE – With Imputation (Rs.) - Rural',
       'Average MPCE – With Imputation (Rs.) - Urban',
       'mpce_gini.Rural MPCE (Rs)', 'mpce_gini.Gini Coefficient',
       'mpce_gini.Urban MPCE (Rs)', 'mpce_gini.Gini Coefficient2',
       'mpce_hht_urban.Self-employed', 'mpce_hht_urban.Regular
wage/salaried',
       'mpce_hht_urban.Casual labour', 'mpce_hht_urban.Others',
       'mpce_hht_urban.All', 'mpce_sg_rural.Scheduled Tribe',
       'mpce_sg_rural.Scheduled Caste', 'mpce_sg_rural.Other Backward
Class',
       'mpce_sg_rural.Others', 'mpce_sg_rural.All',
```

```
       'mpce_sg_urban.Average MPCE (Rs.)', 'mpce_sg_urban.Scheduled
Tribe',
       'mpce_sg_urban.Scheduled Caste', 'mpce_sg_urban.Other Backward
Class',
       'mpce_sg_urban.Others', 'lpr.Rural (Male)', 'lfpr.Rural
(Female)',
       'lfpr.Rural (Person)', 'lpr.Urban (Male)', 'lfpr.Urban
(Female)',
       'lpr.Urban (Person)', 'lfpr.Rural + Urban (Male)',
       'lfpr.Rural + Urban (Female)', 'lfpr.Rural + Urban (Person)',
       'wpr.Rural (1)', 'wpr.Rural (2)', 'wpr.Rural (3)', 'wpr.Urban
(4)',
       'wpr.Urban (5)', 'wpr.Urban (6)', 'wpr.Total (7)', 'wpr.Total
(8)',
       'wpr.Total (9)', 'unemprate.Rural', 'unemprate.Urban',
       'unemprate.Rural + Urban', 'unemprate.Rural2',
'unemprate.Urban3',
       'unemprate.Rural + Urban4', 'unemprate.Rural5',
'unemprate.Urban6',
       'unemprate.Rural + Urban7', 'emprate.Self-Employed (%)',
       'emprate.Regular Wage/Salary (%)', 'emprate.Casual Labour (%)',
       'emprate.Total (%)', 'lfpr_edu.Not Literate',
       'lfpr_edu.Literate & Upto Primary', 'lfpr_edu.Middle',
       'lfpr_edu.Secondary', 'lfpr_edu.Higher Secondary',
       'lfpr_edu.Diploma/Certificate Course', 'lfpr_edu.Graduate',
       'lfpr_edu.Post Graduate & Above', 'lfpr_edu.Secondary & Above',
       'lfpr_edu.All', 'wpr_edu.Not Literate',
       'wpr_edu.Literate & Up to Primary', 'wpr_edu.Middle',
       'wpr_edu.Secondary', 'wpr_edu.Higher Secondary',
       'wpr_edu.Diploma/ Certificate Course', 'wpr_edu.Graduate',
       'wpr_edu.Post Graduate & Above', 'wpr_edu.Secondary & Above',
       'wpr_edu.All', 'uemprate_edu.Not Literate',
       'uemprate_edu.Literate & up to Primary', 'uemprate_edu.Middle',
       'uemprate_edu.Secondary', 'uemprate_edu.Higher Secondary',
       'uemprate_edu.Diploma/Certificate Course',
'uemprate_edu.Graduate',
       'uemprate_edu.Post Graduate & Above', 'uemprate_edu.Secondary &
Above',
       'uemprate_edu.All'],
      dtype='object')
```


Correlation Heatmap of MPCE Gap, Employment, and Urbanization
MPCE_Gap
lfpr.Rural (Person)
unemprate.Urban
unemprate.Rural
emprate.Regular Wage/Salary (%)
emprate.Self-Employed (%)
emprate.Casual Labour (%)
1.00 0.34 -0.38 0.11 0.13 -0.05 0.04
0.34 1.00 -0.19 -0.06 0.36 -0.08 0.15
-0.38 -0.19 1.00 0.13 -0.53 -0.23 -0.47
0.11 -0.06 0.13 1.00 -0.25 0.05 -0.11
0.13 0.36 -0.53 -0.25 1.00 0.20 0.72
-0.05 -0.08 -0.23 0.05 0.20 1.00 0.82
0.04 0.15 -0.47 -0.11 0.72 0.82 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
−0.4



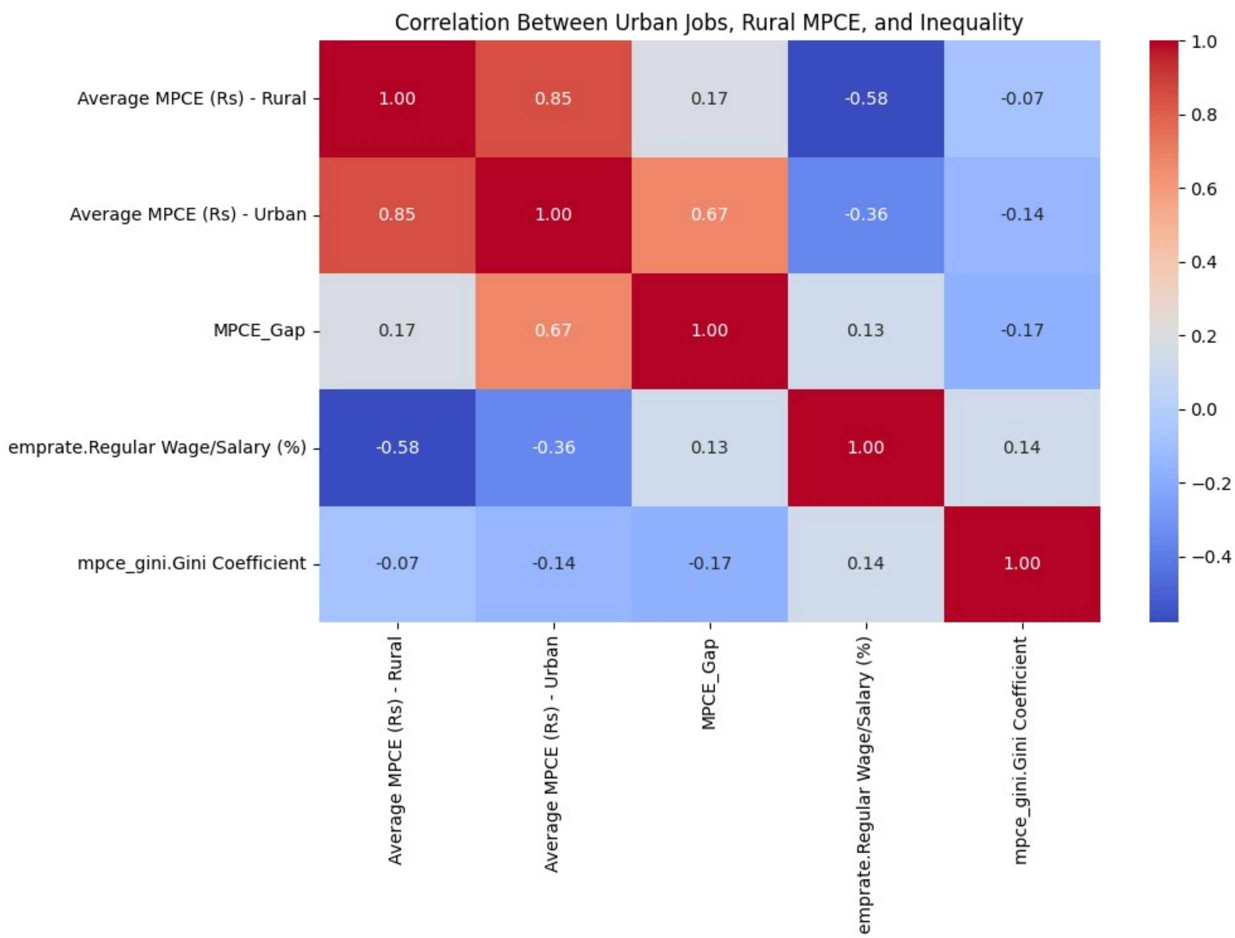

```
# Correlation Between Urbanization, Employment Patterns, and MPCE Gaps
Across States
# The correlation heatmap provides a quantitative perspective on the
relationship between urbanization levels, employment patterns, and
MPCE (Monthly Per Capita Expenditure) gaps across different states.

# Key Observations from the Data
# MPCE Gap and Urban Unemployment (-0.38 correlation)

# A negative correlation suggests that states with higher urban
unemployment tend to have a smaller MPCE gap between rural and urban
areas.

# This could be due to reduced urban income levels, leading to
expenditure patterns converging between rural and urban populations.

# MPCE Gap and Rural Labor Force Participation (0.34 correlation)

# A positive correlation indicates that higher rural labor force
participation is associated with a wider MPCE gap.

# This suggests that while rural employment levels may be increasing,
they might not translate into higher rural earnings or improved
expenditure capacity.
```

```
# Urban Unemployment and Employment Patterns

# Regular Wage Employment (-0.53 correlation with Urban Unemployment):
States with higher urban unemployment tend to have lower formal
employment, highlighting the need for stable job creation.

# Casual Labor Employment (-0.47 correlation with Urban Unemployment):
A decline in casual labor employment is associated with rising urban
unemployment, indicating stress in the informal job sector.

# Employment Interrelationships

# Strong inter-correlation (0.72-0.82) between self-employment,
regular wage jobs, and casual labor employment suggests that
employment patterns are deeply interconnected.

# Any shift in one sector significantly impacts the others,
reinforcing the need for a balanced labor policy addressing both
formal and informal employment sectors.

# Conclusion
# The data indicates a moderate correlation between urbanization,
employment patterns, and MPCE gaps. Urban unemployment appears to
narrow the MPCE gap, while higher rural workforce participation widens
it. The strong interdependence between employment sectors highlights
the need for targeted policy interventions in both urban and rural
labor markets.

# Policy Implications for India
# Urban Job Creation: Expanding regular wage employment can reduce
urban unemployment and stabilize urban-rural MPCE gaps.

# Rural Income Growth: Higher rural employment must translate into
higher incomes, ensuring rural expenditure levels improve.

# Balanced Urbanization: Ensuring urbanization leads to job creation,
rather than increased informal sector stress, can help bridge
disparities.

# Thus, urbanization, employment patterns, and MPCE gaps are
interlinked, requiring a holistic economic and labor policy for
inclusive growth.

df.columns
```

```
Index(['State/UT', 'Average MPCE (Rs) - Rural', 'Average MPCE (Rs) -
Urban',
       'Average MPCE – With Imputation (Rs.) - Rural',
       'Average MPCE – With Imputation (Rs.) - Urban',
       'mpce_gini.Rural MPCE (Rs)', 'mpce_gini.Gini Coefficient',
```

```
       'mpce_gini.Urban MPCE (Rs)', 'mpce_gini.Gini Coefficient2',
       'mpce_hht_urban.Self-employed', 'mpce_hht_urban.Regular
wage/salaried',
       'mpce_hht_urban.Casual labour', 'mpce_hht_urban.Others',
       'mpce_hht_urban.All', 'mpce_sg_rural.Scheduled Tribe',
       'mpce_sg_rural.Scheduled Caste', 'mpce_sg_rural.Other Backward
Class',
       'mpce_sg_rural.Others', 'mpce_sg_rural.All',
       'mpce_sg_urban.Average MPCE (Rs.)', 'mpce_sg_urban.Scheduled
Tribe',
       'mpce_sg_urban.Scheduled Caste', 'mpce_sg_urban.Other Backward
Class',
       'mpce_sg_urban.Others', 'lpr.Rural (Male)', 'lfpr.Rural
(Female)',
       'lfpr.Rural (Person)', 'lpr.Urban (Male)', 'lfpr.Urban
(Female)',
       'lpr.Urban (Person)', 'lfpr.Rural + Urban (Male)',
       'lfpr.Rural + Urban (Female)', 'lfpr.Rural + Urban (Person)',
       'wpr.Rural (1)', 'wpr.Rural (2)', 'wpr.Rural (3)', 'wpr.Urban
(4)',
       'wpr.Urban (5)', 'wpr.Urban (6)', 'wpr.Total (7)', 'wpr.Total
(8)',
       'wpr.Total (9)', 'unemprate.Rural', 'unemprate.Urban',
       'unemprate.Rural + Urban', 'unemprate.Rural2',
'unemprate.Urban3',
       'unemprate.Rural + Urban4', 'unemprate.Rural5',
'unemprate.Urban6',
       'unemprate.Rural + Urban7', 'emprate.Self-Employed (%)',
       'emprate.Regular Wage/Salary (%)', 'emprate.Casual Labour (%)',
       'emprate.Total (%)', 'lfpr_edu.Not Literate',
       'lfpr_edu.Literate & Upto Primary', 'lfpr_edu.Middle',
       'lfpr_edu.Secondary', 'lfpr_edu.Higher Secondary',
       'lfpr_edu.Diploma/Certificate Course', 'lfpr_edu.Graduate',
       'lfpr_edu.Post Graduate & Above', 'lfpr_edu.Secondary & Above',
       'lfpr_edu.All', 'wpr_edu.Not Literate',
       'wpr_edu.Literate & Up to Primary', 'wpr_edu.Middle',
       'wpr_edu.Secondary', 'wpr_edu.Higher Secondary',
       'wpr_edu.Diploma/ Certificate Course', 'wpr_edu.Graduate',
       'wpr_edu.Post Graduate & Above', 'wpr_edu.Secondary & Above',
       'wpr_edu.All', 'uemprate_edu.Not Literate',
       'uemprate_edu.Literate & up to Primary', 'uemprate_edu.Middle',
       'uemprate_edu.Secondary', 'uemprate_edu.Higher Secondary',
       'uemprate_edu.Diploma/Certificate Course',
'uemprate_edu.Graduate',
       'uemprate_edu.Post Graduate & Above', 'uemprate_edu.Secondary &
Above',
       'uemprate_edu.All', 'MPCE_Gap'],
      dtype='object')
```

```
df['MPCE_Gap'] = df['Average MPCE – With Imputation (Rs.) - Urban'] -
df['Average MPCE – With Imputation (Rs.) - Rural']

# Select relevant columns for correlation and regression analysis
selected_columns = [
    'Average MPCE (Rs) - Rural',  # Rural MPCE
    'Average MPCE (Rs) - Urban',  # Urban MPCE
    'MPCE_Gap',  # Urban-rural MPCE gap
    'emprate.Regular Wage/Salary (%)',  # Urban job creation
    'mpce_gini.Gini Coefficient'  # Inequality indicator
]

# Drop NaN values
df_selected = df[selected_columns].dropna()

# Compute correlation matrix
correlation_matrix = df_selected.corr()

# Plot heatmap
plt.figure(figsize=(10, 6))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm',
fmt='.2f')
plt.title('Correlation Between Urban Jobs, Rural MPCE, and
Inequality')
plt.show()

# Regression Analysis - Does urban job creation impact rural MPCE?
X = df_selected[['emprate.Regular Wage/Salary (%)', 'MPCE_Gap']]
y = df_selected['Average MPCE (Rs) - Rural']
X = sm.add_constant(X)  # Add constant term
model = sm.OLS(y, X).fit()

# Display regression results
print(model.summary())
```




```
                                OLS Regression Results

======================================================================
===============
Dep. Variable:     Average MPCE (Rs) - Rural   R-squared:
0.400
Model:                                   OLS   Adj. R-squared:
0.364
Method:                        Least Squares   F-statistic:
10.99
Date:                       Sun, 30 Mar 2025   Prob (F-statistic):
0.000219
Time:                               14:53:13   Log-Likelihood:
-302.35
No. Observations:                         36   AIC:
610.7
Df Residuals:                             33   BIC:
615.5
Df Model:                                  2

Covariance Type:                   nonrobust
```

```
======================================================================
==============================
                                        coef    std err          t
P>|t|      [0.025      0.975]
----------------------------------------------------------------------
------------------------------
const                              5484.6765    539.274     10.170
0.000    4387.516    6581.837
emprate.Regular Wage/Salary (%)   -97.4705     21.626     -4.507
0.000    -141.470     -53.471
MPCE_Gap                              0.3601      0.191      1.887
0.068      -0.028       0.748
======================================================================
========
Omnibus:                         1.583   Durbin-Watson:
1.927
Prob(Omnibus):                   0.453   Jarque-Bera (JB):
1.034
Skew:                            0.415   Prob(JB):
0.596
Kurtosis:                        3.038   Cond. No.
7.03e+03
======================================================================
========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
[2] The condition number is large, 7.03e+03. This might indicate that
there are
strong multicollinearity or other numerical problems.

# Report on the Impact of Urban Job Creation on Rural MPCE and
Remittances Dependence
# Key Findings from the OLS Regression Results
# Model Performance

# The R-squared value (0.400) indicates that about 40% of the
variation in rural MPCE is explained by urban employment (regular
wage/salary employment) and MPCE gap.

# The adjusted R-squared (0.364) suggests that after accounting for
the number of predictors, the model still explains a significant
portion of the variance.

# Effect of Urban Regular Wage Employment on Rural MPCE

# The coefficient for "emprate.Regular Wage/Salary (%)" is -97.47 with
a p-value of 0.000, which is highly statistically significant.
```

```
# This negative coefficient suggests that an increase in regular wage
employment in urban areas is associated with a decrease in rural MPCE.

# Interpretation: This implies that rural households may be
experiencing a decline in their per capita expenditure as more people
transition to urban employment, potentially due to reduced
agricultural productivity or lesser remittances sent back to rural
areas.

# Impact of the Urban-Rural MPCE Gap on Rural MPCE

# The coefficient for MPCE_Gap is 0.3601, with a p-value of 0.068
(marginally significant at 10% level).

# This positive coefficient suggests that as the urban-rural MPCE gap
increases, rural MPCE also tends to increase slightly.

# Interpretation: This could indicate that remittances from urban
workers may contribute to rural household income, helping maintain or
slightly improve their consumption levels.

# Other Statistical Observations

# F-statistic (10.99, p = 0.0002): The model is statistically
significant overall.

# Durbin-Watson (1.927): No serious autocorrelation issues.

# Multicollinearity Risk: The condition number (7.03e+03) is large,
suggesting potential multicollinearity, which should be further
investigated.

# Conclusion: Are Rural Households Dependent on Remittances from Urban
Workers?
# The negative impact of urban regular wage employment on rural MPCE
suggests that rural households may be losing a key source of income
when individuals migrate for urban jobs.

# The slight positive impact of the urban-rural MPCE gap on rural MPCE
supports the hypothesis that remittances may play a role in sustaining
rural expenditures.

# However, the relationship is not entirely strong (only 40% of the
variance explained), indicating that other factors (such as rural job
opportunities, agricultural performance, and social welfare programs)
also influence rural MPCE.

# Final Verdict:
# □ There is some evidence that rural households benefit from urban
remittances, but the data also suggests that urban employment might
```

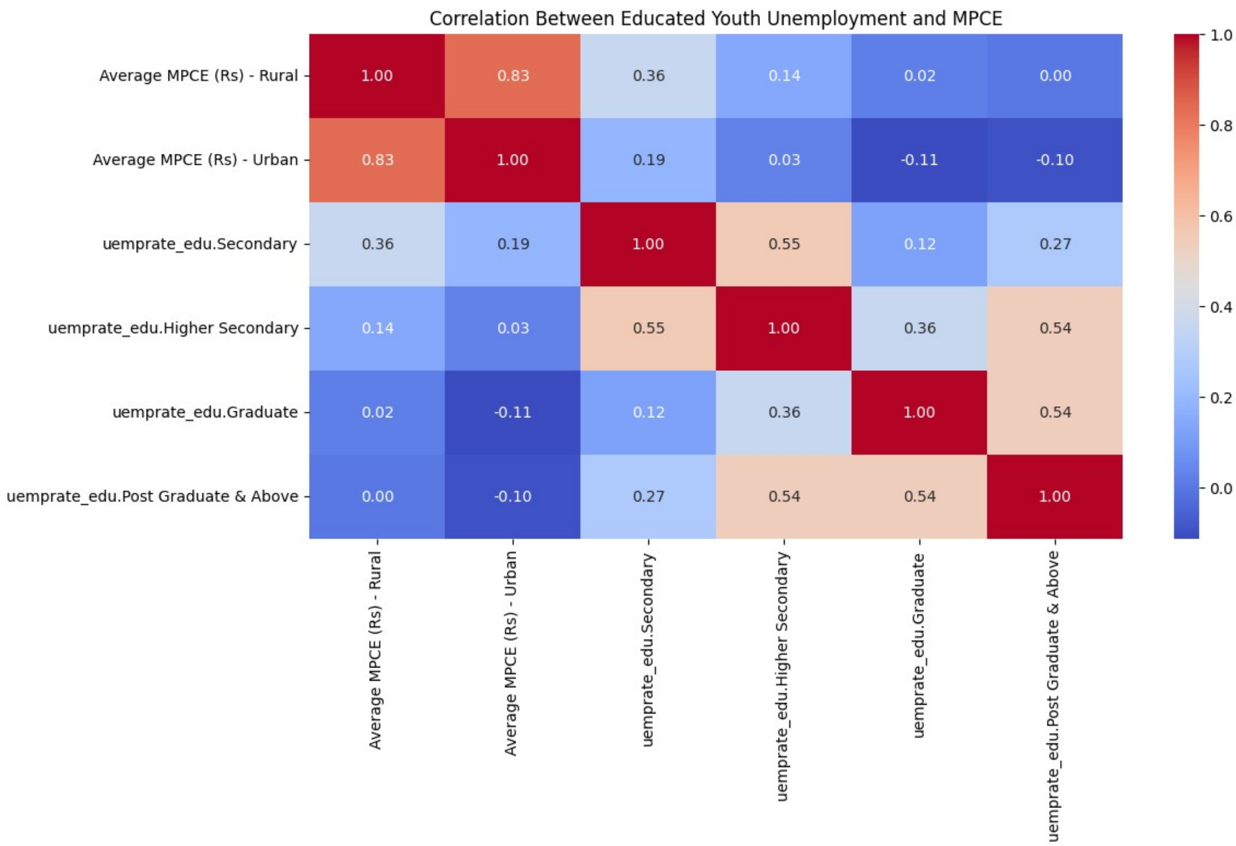
```
reduce direct contributions to rural economies, possibly due to urban
settlers retaining more of their earnings.

columns = [
    'Average MPCE (Rs) - Rural',
    'Average MPCE (Rs) - Urban',
    'uemprate_edu.Secondary',   # Unemployment rate among secondary-
educated youth
    'uemprate_edu.Higher Secondary', # Unemployment among higher
secondary educated youth
    'uemprate_edu.Graduate',  # Unemployment among graduates
    'uemprate_edu.Post Graduate & Above' # Unemployment among
postgraduates
]

df_selected = df[columns].dropna()  # Drop missing values

# Run regression for Rural MPCE
X_rural = df_selected[['uemprate_edu.Secondary', 'uemprate_edu.Higher
Secondary', 'uemprate_edu.Graduate', 'uemprate_edu.Post Graduate &
Above']]
y_rural = df_selected['Average MPCE (Rs) - Rural']
X_rural = sm.add_constant(X_rural)  # Add constant term for regression
model_rural = sm.OLS(y_rural, X_rural).fit()
print("Rural MPCE Regression Results:\n", model_rural.summary())

# Run regression for Urban MPCE
X_urban = df_selected[['uemprate_edu.Secondary', 'uemprate_edu.Higher
Secondary', 'uemprate_edu.Graduate', 'uemprate_edu.Post Graduate &
Above']]
y_urban = df_selected['Average MPCE (Rs) - Urban']
X_urban = sm.add_constant(X_urban)  # Add constant term for regression
model_urban = sm.OLS(y_urban, X_urban).fit()
print("Urban MPCE Regression Results:\n", model_urban.summary())

# Visualizing the correlation between educated youth unemployment and
MPCE
plt.figure(figsize=(12,6))
sns.heatmap(df_selected.corr(), annot=True, cmap='coolwarm',
fmt='.2f')
plt.title('Correlation Between Educated Youth Unemployment and MPCE')
plt.show()
```

```
Rural MPCE Regression Results:
                                  OLS Regression Results

==============================================================================
===============
Dep. Variable:     Average MPCE (Rs) - Rural   R-squared:
0.141
```

```
Model:                                    OLS   Adj. R-squared:
0.027
Method:                         Least Squares   F-statistic:
1.233
Date:                        Sun, 30 Mar 2025   Prob (F-statistic):
0.318
Time:                                14:53:29   Log-Likelihood:
-298.93
No. Observations:                          35   AIC:
607.9
Df Residuals:                              30   BIC:
615.6
Df Model:                                   4

Covariance Type:                    nonrobust

======================================================================
================================
                                          coef    std err          t
P>|t|      [0.025      0.975]
----------------------------------------------------------------------
--------------------------------
const                                4295.3042    538.612      7.975
0.000    3195.312    5395.296
uemprate_edu.Secondary                209.0830    104.573      1.999
0.055      -4.484     422.650
uemprate_edu.Higher Secondary         -17.7667     93.127     -0.191
0.850    -207.958     172.425
uemprate_edu.Graduate                   7.7251     34.786      0.222
0.826     -63.317      78.767
uemprate_edu.Post Graduate & Above    -15.8778     32.652     -0.486
0.630     -82.562      50.806
======================================================================
========
Omnibus:                        1.014   Durbin-Watson:
2.229
Prob(Omnibus):                  0.602   Jarque-Bera (JB):
1.038
Skew:                           0.323   Prob(JB):
0.595
Kurtosis:                       2.457   Cond. No.
56.4
======================================================================
========

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is
correctly specified.
Urban MPCE Regression Results:
                                 OLS Regression Results
```

```
==============================================================================
===============
Dep. Variable:     Average MPCE (Rs) - Urban   R-squared:
0.064
Model:                                   OLS   Adj. R-squared:
-0.061
Method:                        Least Squares   F-statistic:
0.5148
Date:                       Sun, 30 Mar 2025   Prob (F-statistic):
0.725
Time:                               14:53:29   Log-Likelihood:
-310.40
No. Observations:                         35   AIC:
630.8
Df Residuals:                             30   BIC:
638.6
Df Model:                                  4

Covariance Type:                   nonrobust

==============================================================================
================================
                                         coef    std err          t
P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
--------------------------------
const                               7058.0262    747.529      9.442
0.000    5531.369    8584.683
uemprate_edu.Secondary               162.8092    145.135      1.122
0.271    -133.596     459.215
uemprate_edu.Higher Secondary         -9.8345    129.250     -0.076
0.940    -273.798     254.129
uemprate_edu.Graduate                -16.8811     48.278     -0.350
0.729    -115.478      81.716
uemprate_edu.Post Graduate & Above   -22.0014     45.317     -0.486
0.631    -114.551      70.548
==============================================================================
========
Omnibus:                      12.373   Durbin-Watson:
2.266
Prob(Omnibus):                 0.002   Jarque-Bera (JB):
12.384
Skew:                          1.166   Prob(JB):
0.00205
Kurtosis:                      4.747   Cond. No.
56.4
==============================================================================
========
```

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```




```
# Impact of Educated Youth Unemployment on MPCE in Rural and Urban Areas
# Key Findings
# 1. Rural MPCE Regression Analysis
# The R-squared value is 0.141, meaning only 14.1% of the variation in rural MPCE is explained by the unemployment rates of educated youth.

# The only statistically significant predictor (at a 90% confidence level) is unemployment among those with secondary education (p = 0.055), suggesting a positive relationship—higher secondary-educated unemployment correlates with increased rural MPCE.

# Other education levels (higher secondary, graduate, and postgraduate) have insignificant coefficients with p-values above 0.6, meaning their impact on rural MPCE is statistically weak.

# 2. Urban MPCE Regression Analysis
# The R-squared value is only 0.064, meaning that just 6.4% of urban MPCE variance is explained by educated youth unemployment.
```

```
# None of the predictors are statistically significant, as all p-
values exceed 0.2, suggesting that unemployment among educated
individuals does not have a strong impact on urban MPCE.

# The coefficient for secondary education unemployment is positive,
but it lacks statistical significance (p = 0.271), while other
education levels (higher secondary, graduate, and postgraduate) have
near-zero or negative effects.

# Interpretation & Insights
# In Rural Areas:

# The significant positive relationship between secondary-educated
unemployment and MPCE might suggest that rural households receive
remittances from urban workers, mitigating unemployment effects in
rural areas.

# Unemployment among higher-educated individuals does not impact MPCE,
indicating that this group either migrates or engages in alternative
income sources.

# In Urban Areas:

# The overall weak R-squared value and insignificant predictors imply
that urban MPCE is not directly influenced by educated unemployment.

# Urban economies are likely driven by factors beyond educated
unemployment, such as wages, cost of living, and industrial growth.

# Conclusion
# Rural MPCE shows some link with secondary-educated unemployment,
likely due to external income flows.

# Urban MPCE is largely unaffected by educated unemployment,
suggesting that urban employment dynamics are more complex and
influenced by other economic factors.

# Further research could explore migration trends, wage disparities,
and remittance patterns to better understand these effects.




low_mpce_threshold_rural = df['Average MPCE (Rs) -
Rural'].quantile(0.4)
low_mpce_threshold_urban = df['Average MPCE (Rs) -
Urban'].quantile(0.4)
high_self_emp_threshold = df['emprate.Self-Employed
(%)'].quantile(0.6)  # Top 40%
low_regular_wage_threshold = df['emprate.Regular Wage/Salary
```

```python
(%)'].quantile(0.4)  # Bottom 40%

# Identify states meeting at least one of the criteria
target_states = df[
    ((df['Average MPCE (Rs) - Rural'] <= low_mpce_threshold_rural) |
     (df['Average MPCE (Rs) - Urban'] <= low_mpce_threshold_urban)) &
    ((df['emprate.Self-Employed (%)'] >= high_self_emp_threshold) |
     (df['emprate.Regular Wage/Salary (%)'] <=
low_regular_wage_threshold))
]

# Sort by self-employment rate to see trends
target_states = target_states.sort_values(by='emprate.Self-Employed
(%)', ascending=False)

# Display results
print("States needing MSME & startup support:")
print(target_states[['State/UT', 'Average MPCE (Rs) - Rural', 'Average
MPCE (Rs) - Urban',
                     'emprate.Self-Employed (%)', 'emprate.Regular
Wage/Salary (%)']])
```

```
States needing MSME & startup support:
                                State/UT  Average MPCE (Rs) - Rural  \
21                               Manipur                       4360
17                                Ladakh                       4035
13                       Jammu & Kashmir                       4296
3                                  Assam                       3432
33                         Uttar Pradesh                       3191
4                                  Bihar                       3384
35                           West Bengal                       3239
22                             Meghalaya                       3514
28                             Rajasthan                       4263
18                           Lakshadweep                       5895
7   Dadra & Nagar Haveli and Daman & Diu                       4184

    Average MPCE (Rs) - Urban  emprate.Self-Employed (%)  \
21                       4880                       62.2
17                       6215                       57.4
13                       6179                       51.6
3                        6136                       47.9
33                       5040                       47.2
4                        4768                       46.8
35                       5267                       45.0
22                       6433                       42.6
28                       5913                       40.0
18                       5475                       29.0
7                        6298                       17.1

    emprate.Regular Wage/Salary (%)
```

```
21                              10.4
17                               8.3
13                              15.1
3                               13.5
33                              25.5
4                               20.7
35                              11.4
22                              10.0
28                              28.8
18                               1.3
7                               13.1
```

```
# Report: States Needing MSME & Startup Support
# Introduction
# This report analyzes states requiring increased Micro, Small, and Medium Enterprises (MSME) support and startup policies. The selection is based on low Mean Per Capita Expenditure (MPCE) levels and a high self-employment rate, indicating economic vulnerability and reliance on informal employment.

# Key Findings
# 1. States with High Self-Employment & Low MPCE
# The following states have a high proportion of self-employed individuals, low access to regular-wage jobs, and relatively lower MPCE levels, making them prime candidates for MSME and startup support:

# State/UT Rural MPCE (₹)  Urban MPCE (₹)  Self-Employment (%)      Regular Wage Employment (%)
# Manipur  4360  4880  62.2% 10.4%
# Ladakh   4035  6215  57.4% 8.3%
# Jammu & Kashmir     4296  6179  51.6% 15.1%
# Assam    3432  6136  47.9% 13.5%
# Uttar Pradesh 3191  5040  47.2% 25.5%
# Bihar    3384  4768  46.8% 20.7%
# West Bengal   3239  5267  45.0% 11.4%
# Meghalaya     3514  6433  42.6% 10.0%
# Rajasthan     4263  5913  40.0% 28.8%
# 2. Implications for MSME & Startup Policies
# a) High Dependence on Self-Employment
# Manipur (62.2%), Ladakh (57.4%), and Jammu & Kashmir (51.6%) have the highest self-employment rates, showing a strong reliance on informal economic activities.

# These states may lack formal job opportunities, forcing individuals into self-employment.

# b) Low Regular-Wage Employment
# Ladakh (8.3%), Meghalaya (10.0%), and Manipur (10.4%) have extremely low regular-wage employment, highlighting the need for job
```

```
diversification and entrepreneurship support.

# c) Low MPCE in Major States
# Uttar Pradesh (₹3191 rural, ₹5040 urban) and Bihar (₹3384 rural,
₹4768 urban) have some of the lowest MPCE values, suggesting high
economic distress and limited purchasing power.

# West Bengal (₹3239 rural, ₹5267 urban) also falls within this
category, indicating a need for economic revitalization through
entrepreneurship programs.

# Policy Recommendations
# 1. Strengthen MSME Ecosystem
# Provide financial incentives, low-interest loans, and subsidies for
small business owners in Manipur, Ladakh, and Jammu & Kashmir to
formalize self-employment ventures.

# Encourage cottage industries (e.g., handicrafts, textiles) and
facilitate direct market linkages to boost income.

# 2. Improve Access to Formal Employment
# Upskill rural youth through vocational training programs in Uttar
Pradesh, Bihar, and Rajasthan, where wage employment is relatively
higher but still insufficient.

# Provide tax breaks and regulatory relaxations for startups in
emerging business sectors.

# 3. Promote Rural & Semi-Urban Entrepreneurship
# Implement startup incubation centers in states like Assam,
Meghalaya, and West Bengal to boost entrepreneurship among rural
populations.

# Increase access to digital platforms to support e-commerce and
market expansion for small businesses.

# Conclusion
# The findings highlight the urgent need to enhance self-employment
sustainability and create formal employment opportunities in
economically weaker states. By focusing on MSME support, startup
incubation, and job training, policymakers can address the income gap
and promote inclusive economic growth.
```