DMS 692
Advanced Statistical Methods For Business Analytics
Examining Changes in Educational Outcomes in Indian Schools (2017 & 2021)

Group 8
Bhavishya Gupta(220295)
Rishav Raj(220889)
Rohan Nimesh(220907)

April 2025

Presentation Outline

Introduction

- Education is a critical determinant of a nation's development and prosperity
- India's diverse education system requires regular assessment to identify:
  - Performance trends
  - Regional disparities
  - Areas for improvement
- Data is taken from Source.
- Study period (2017 & 2021) includes significant disruption from COVID-19 pandemic
- Analysis focuses on Class 8 performance across multiple subjects

Research Hypotheses

Primary Hypothesis

There is a significant difference in academic performance metrics between 2017 and 2021 across Indian schools.

Secondary Hypothesis

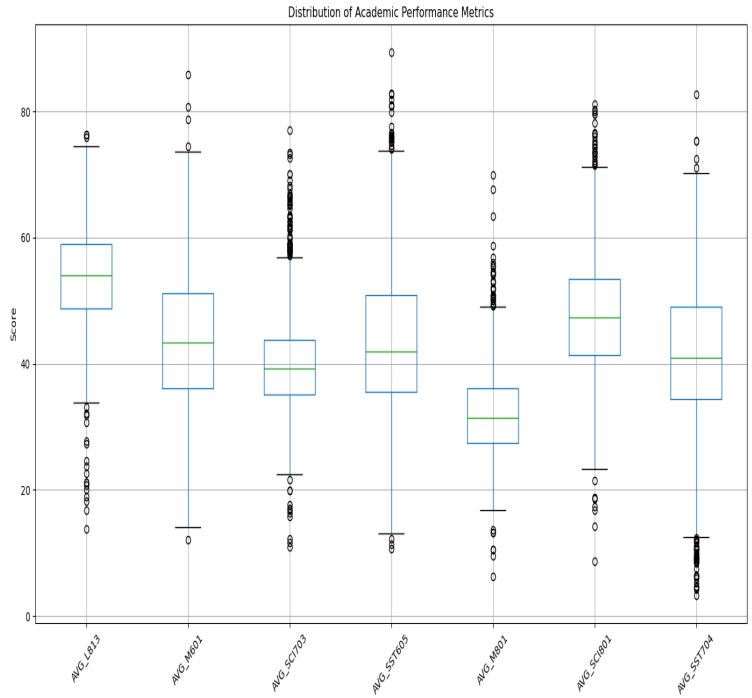
The trends in academic performance vary significantly across different states in India, suggesting regional disparities in educational outcomes.

Tertiary Hypothesis

Performance metrics across different subjects are correlated and can be reduced to fewer underlying factors representing broader academic competencies.

Dataset Overview

- National educational assessment database
- 1450 observations across schools in Indian states & districts
- All observations for Class 8 students
- No missing values
- Performance metrics for 7 subjects:
  - Language (AVG_L813)
  - Mathematics I & II (AVG_M601, AVG_M801)
  - Science I & II (AVG_SCI703, AVG_SCI801)
  - Social Studies I & II (AVG_SST605, AVG_SST704)

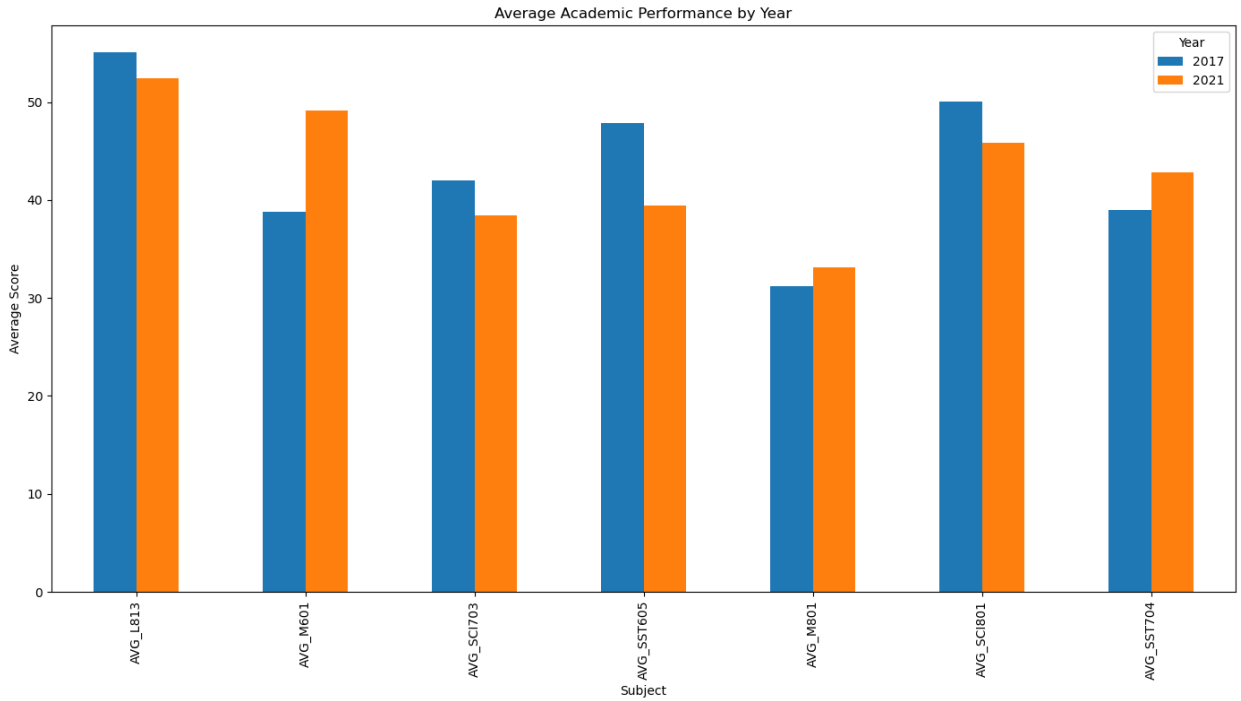


Distribution of Academic performance metrics

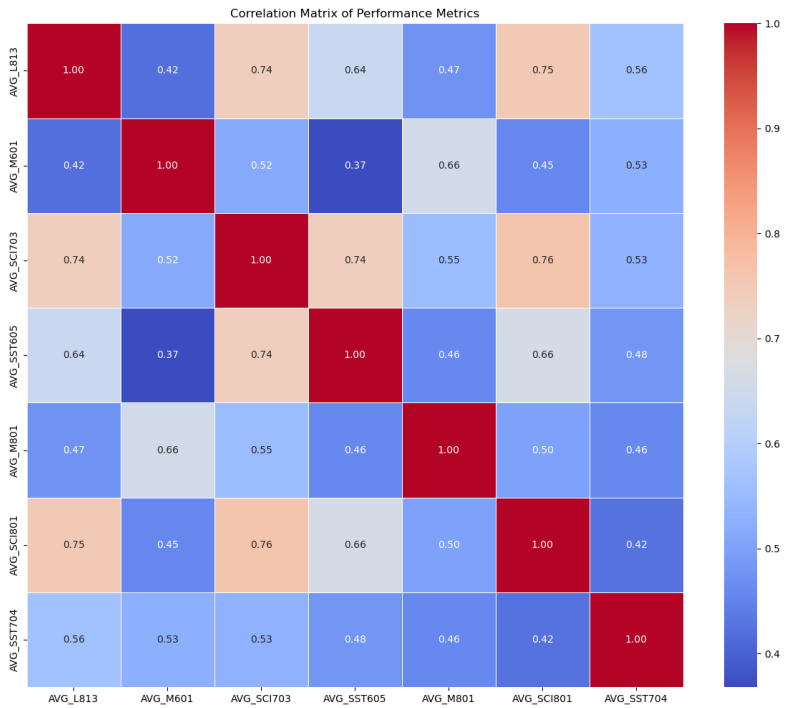
Descriptive Statistics

Table: Summary statistics of performance metrics

Metric	Mean	Std Dev	Min	Max
AVG_L813	53.75	8.15	13.82	76.36
AVG_M601	43.91	10.80	12.08	85.82
AVG_SCI703	40.25	7.91	10.98	77.03
AVG_SST605	43.69	11.47	10.70	89.44
AVG_M801	32.17	7.16	6.25	69.93
AVG_SCI801	47.96	9.51	8.75	81.11
AVG_SST704	40.89	12.55	3.24	82.77



Correlation Analysis



Correlation matrix heatmap

- Strong correlations (>0.7) between:
  - Language (AVG_L813) and Science metrics (AVG_SCI703, AVG_SCI801)
  - Science I (AVG_SCI703) and Social Studies I (AVG_SST605)
- Moderate correlations between Mathematics and other subjects
- All metrics positively correlated, suggesting common underlying factor

Methodology

MANOVA

$$Y = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$$

- Tests differences in means across years and states

ANOVA

$$Y_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

- Follow-up for individual metrics

Principal Component Analysis

$$X = WP + \epsilon$$

- Reduces dimensionality of 7 performance metrics
- Identifies underlying patterns

Factor Analysis

$$X_i = \lambda_{i1}F_1 + \lambda_{i2}F_2 + \ldots + \lambda_{im}F_m + \epsilon_i$$

- Identifies latent constructs
- Explains correlations among metrics

MANOVA & ANOVA Assumptions

Method/Assumption	Test Used	Status	Evidence
MANOVA			
Multivariate Normality	Shapiro-Wilk	✓*	Mixed results, large sample size (N=1450) provides robustness
Homogeneity of Covariance	Levene's test	✓*	Mixed results: Some variables $p>0.05$, others $p<0.05$
Independence	Study design	✓	District-level measures are independent
No Multicollinearity	Correlation matrix	✓	Present but below 0.9 threshold
ANOVA			
Normality	Shapiro-Wilk	✓*	Some non-normality in 2017 data, large sample size provides robustness
Homogeneity of Variance	Levene's test	✓*	Mixed results across metrics
Independence	Study design	✓	Random sampling at district level
Absence of Outliers	Boxplots	✓	Limited influence of outliers

✓= Met, ✓* = Partially met but analysis robust to violations

PCA & Factor Analysis Assumptions

Method/Assumption	Test Used	Status	Evidence
PCA			
Linearity	Correlation matrix	✓	Strong correlations between variables (r >0.7)
Sampling Adequacy	KMO test	✓	KMO = 0.862 (excellent)
Suitability	Bartlett's test	✓	$\chi^2 = 6276.29, p < 0.001$
Variance Explained	PC analysis	✓	First PC: 62.31%, First 3 PCs: 83.77%
Factor Analysis			
Sample Size	N >150	✓	N = 1450, far exceeding requirements
Correlation Structure	Correlation matrix	✓	Values range 0.37 - 0.76, ideal for factor analysis
Factorability	KMO & Bartlett's	✓	KMO = 0.862, Bartlett's: p <0.001
Variable Distribution	Visual inspection	✓*	Some normality violations exist but not critical

✓= Met, ✓* = Partially met but analysis robust to violations

MANOVA Results

Effect of Year on Academic Performance

Wilks' lambda = 0.3548, F(7, 1442) = 374.53, p<0.001

- Significant difference in academic performance between 2017 and 2021

Effect of State on Academic Performance

Wilks' lambda = 0.2703, F(28, 1642) = 25.65, p<0.001

- Academic performance varies significantly across different states

Interaction Effect of Year and State

Wilks' lambda = 0.2036, F(28, 1624) = 32.18, p<0.001

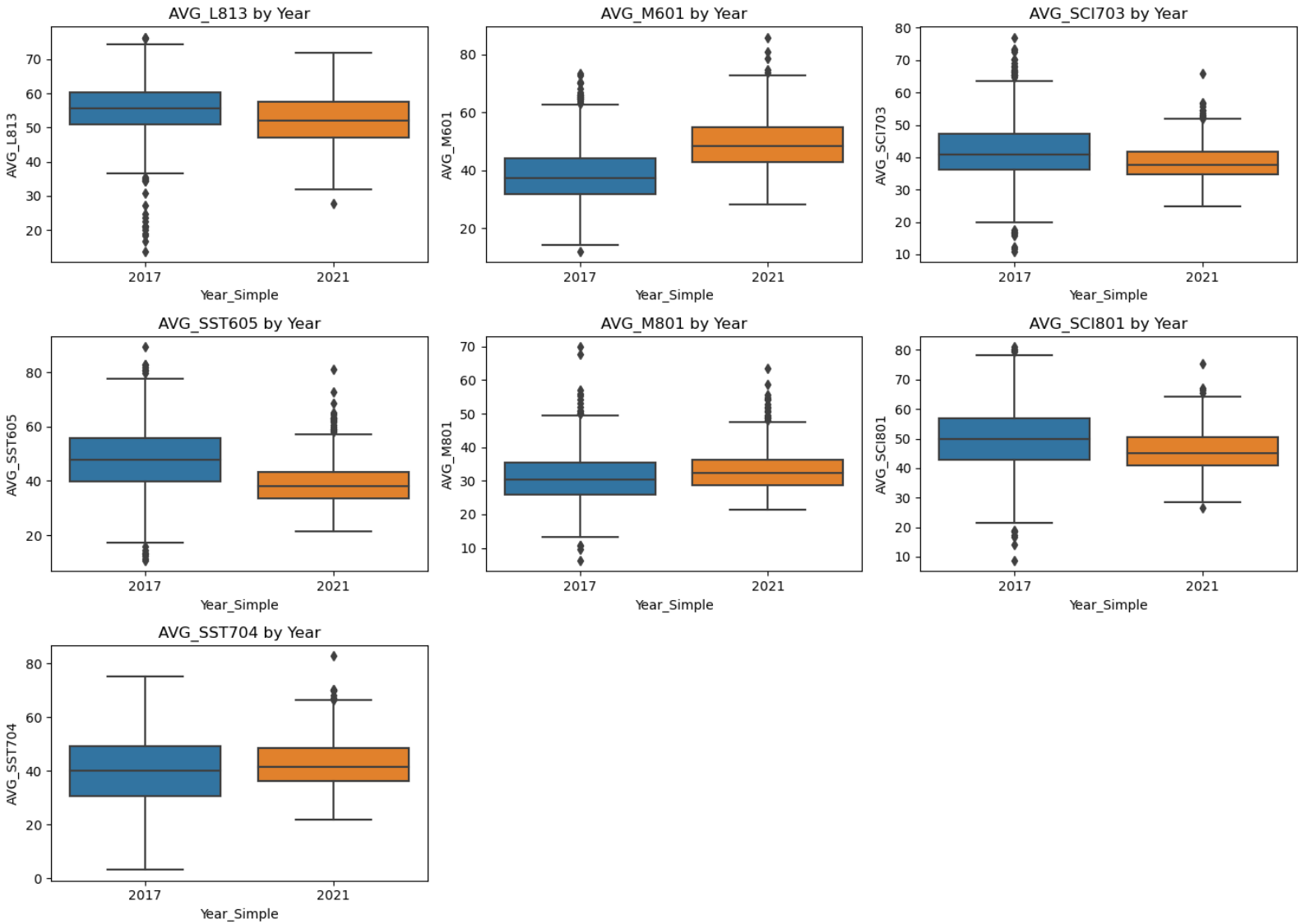
- The change in performance from 2017 to 2021 varied across states

ANOVA Results for Individual Metrics

Table: Change in performance metrics from 2017 to 2021

Subject	2017 Mean	2021 Mean	% Change	p-value
Language (AVG_L813)	55.04	52.42	-4.76%	<0.001
Mathematics I (AVG_M601)	38.80	49.15	+26.67%	<0.001
Science I (AVG_SCI703)	42.04	38.41	-8.63%	<0.001
Social Studies I (AVG_SST605)	47.89	39.39	-17.73%	<0.001
Mathematics II (AVG_M801)	31.22	33.15	+6.19%	<0.001
Science II (AVG_SCI801)	50.03	45.85	-8.37%	<0.001
Social Studies II (AVG_SST704)	39.02	42.82	+9.73%	<0.001

Boxplot of ANOVA Results



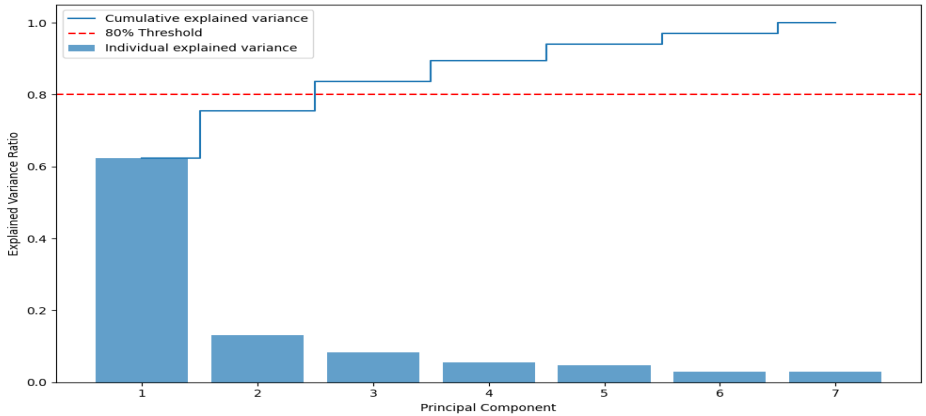
Performance Differences By Year

PCA Results

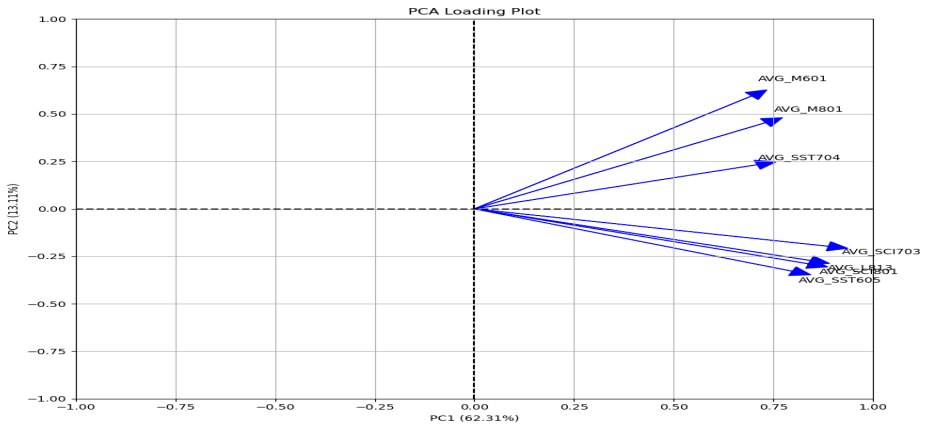
- First PC explains 62.31% of variance
- First three PCs explain 83.77% of variance
- PC1: Overall academic performance
- PC2: Contrasts verbal vs. quantitative abilities

Table: Key component loadings

Metric	PC1	PC2	PC3
AVG_L813	0.84	-0.27	-0.10
AVG_M601	0.70	0.59	0.10
AVG_SCI703	0.89	-0.20	0.08
AVG_SST605	0.80	-0.33	0.01
AVG_M801	0.73	0.45	0.31
AVG_SCI801	0.84	-0.29	0.21
AVG_SST704	0.71	0.23	-0.65

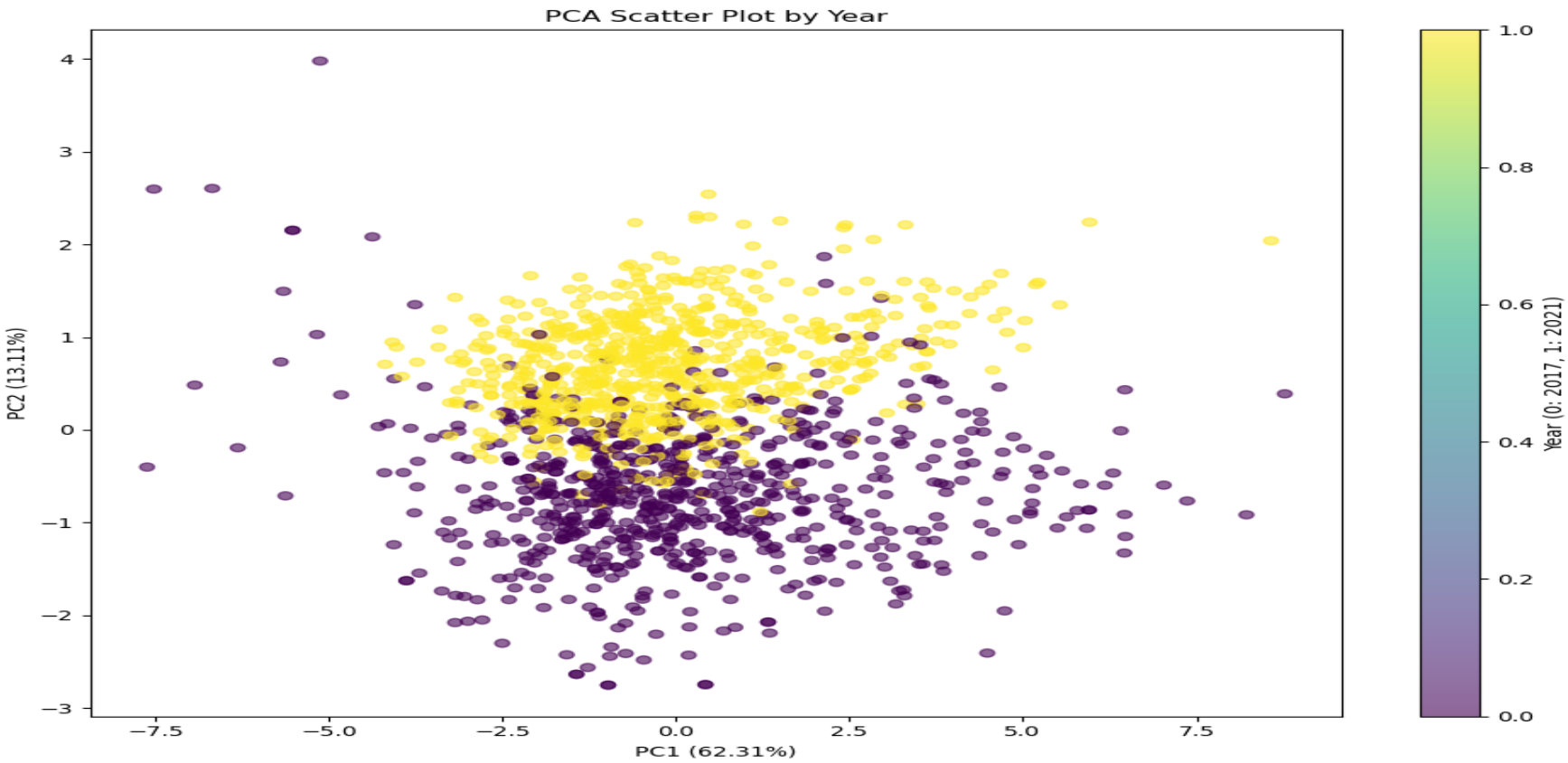


PCA explained variance



PCA loading plot

PCA Visualization

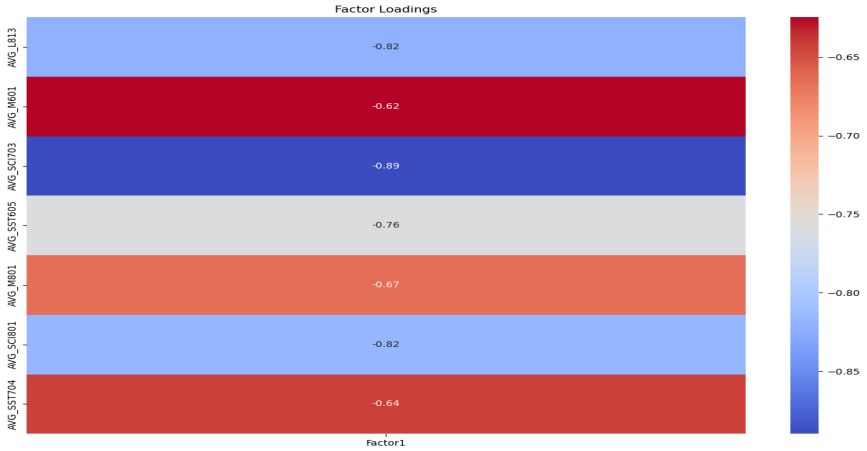


PCA scatter plot by year

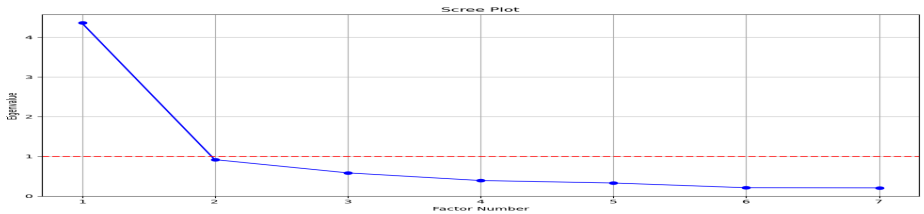
- Clear clustering by year is visible in principal component space
- 2017 and 2021 data points show distinct positioning
- Suggests systematic differences in performance patterns between years

Factor Analysis Results

- Bartlett's test of sphericity:
  - $\chi^2 = 6276.29$, p<0.001
  - Variables are sufficiently correlated
- KMO = 0.862 (excellent sampling adequacy)
- Kaiser criterion suggests 1 factor
- This factor explains $\approx$ 62% of variance



Factor Loadings



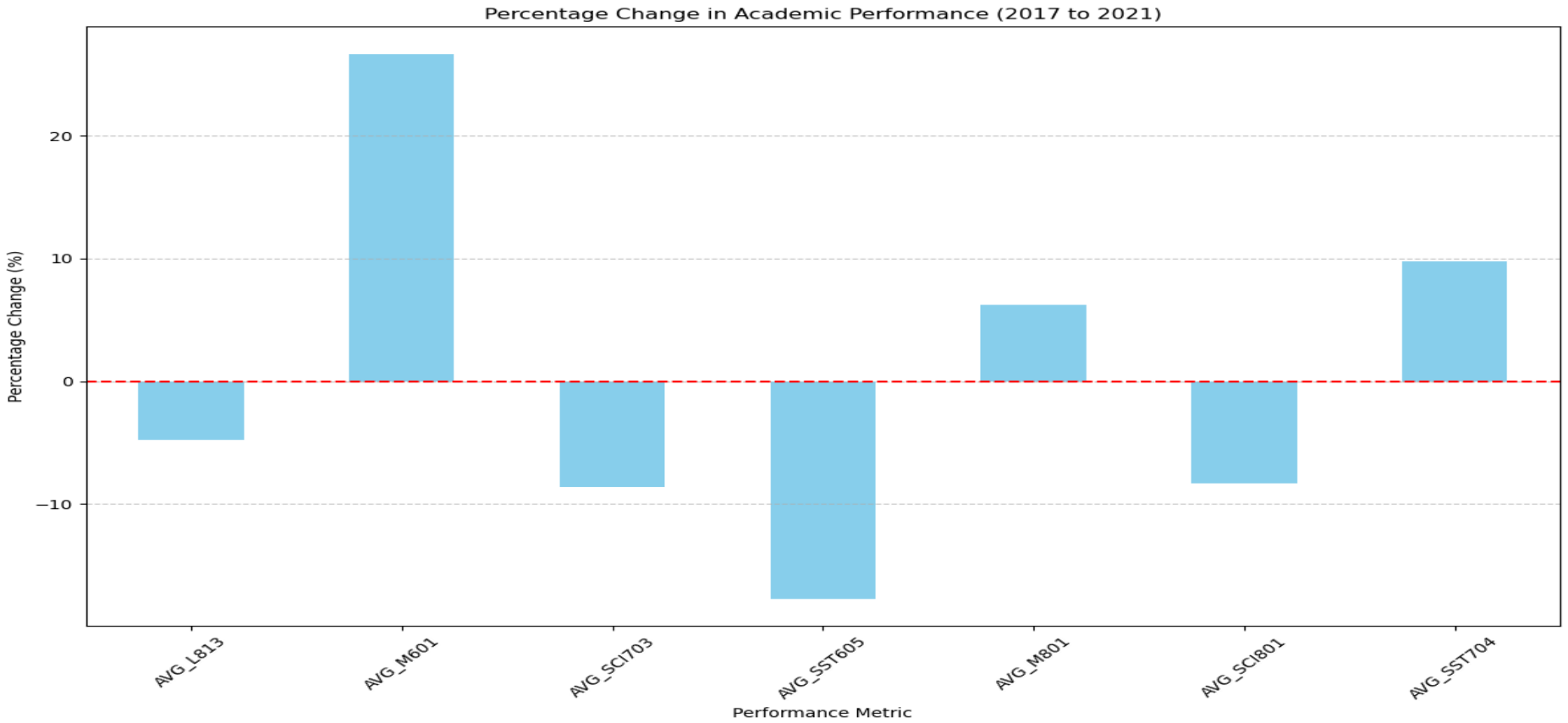
Scree Plot

Interpretation

A single factor representing general academic ability underlies performance across all subjects

Performance Changes (2017 & 2021)

Percentage change in academic performance (2017 & 2021)



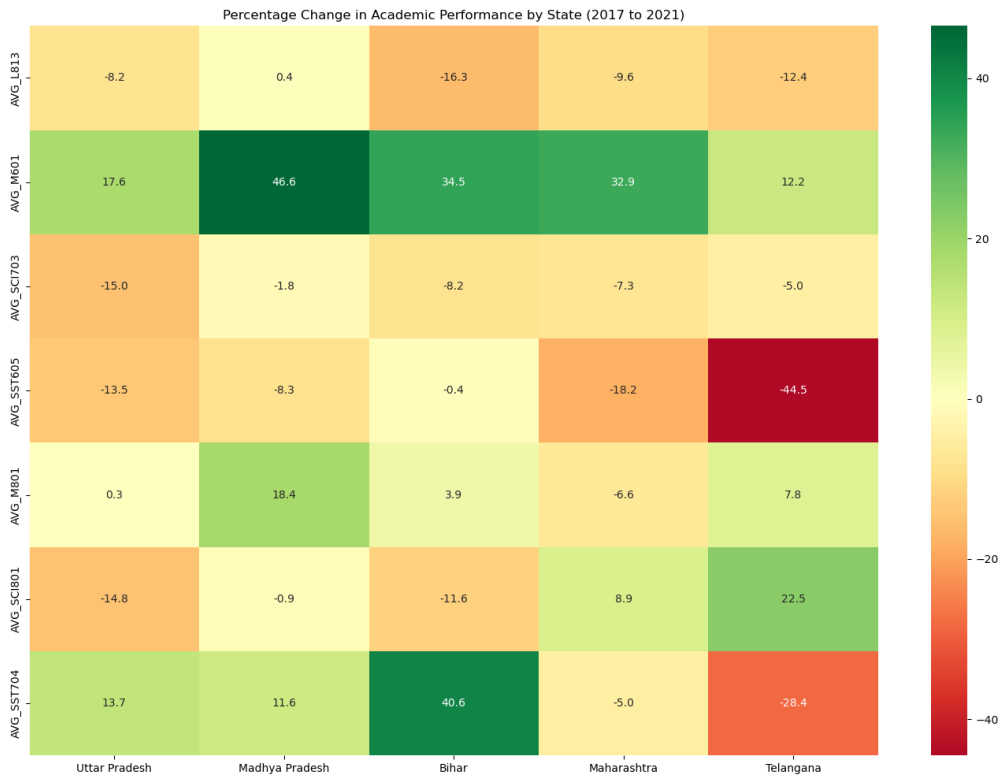
Subjects with Improvement

- Mathematics I (+26.67%)
- Social Studies II (+9.73%)
- Mathematics II (+6.19%)

Subjects with Decline

- Language (-4.76%)
- Science I (-8.63%)
- Science II (-8.37%)
- Social Studies I (-17.73%)

State-wise Performance Changes



Percentage change by state (2017 & 2021)

- Significant regional variations in performance changes
- Madhya Pradesh: Most consistent improvement across subjects
- Telangana: Most extreme variations (both positive and negative)
- Mathematics improved across most states
- Language and Science declined in most states

Key Findings

1. **Significant performance changes between 2017 &2021**
   - All subjects showed statistically significant changes
   - Direction of change varied by subject
2. **Mathematics resilience**
   - Substantial improvement in Mathematics I (+26.67%)
   - Mathematics II also improved (+6.19%)
   - Suggests resilience to pandemic disruptions
3. **Regional disparities**
   - State-wise variations in performance changes
   - Suggests uneven impact of educational policies
4. **Underlying factors**
   - One dominant factor explains most variance
   - Suggests general academic ability as key determinant

Implications & Recommendations

Targeted Interventions

Special attention needed for Language, Science, and Social Studies I where performance declined

Learning from Mathematics Success

Identify and apply successful teaching strategies from mathematics to other subjects

Addressing Regional Disparities

Additional support for states showing significant declines; sharing best practices from successful states

Building Educational Resilience

Develop subject-specific approaches to maintain performance during disruptions

Limitations & Future Research

Limitations:

- Causality not established
- Contextual factors not fully accounted for:
  - Socioeconomic changes
  - Educational policy shifts
  - Pandemic-specific disruptions
- Some statistical assumption violations

Future Research:

- Longitudinal analysis beyond 2021
- Qualitative investigation of regional differences
- Policy evaluation studies
- Integration of socioeconomic data
- Pandemic impact assessment

Conclusion

- Academic performance in Indian schools showed significant changes between 2017 and 2021
- Mathematics demonstrated remarkable resilience and improvement
- Language, Science, and Social Studies I showed concerning declines
- Regional variations highlight educational disparities
- Findings provide valuable insights for educational planning and policy development
- Results emphasize the need for subject-specific and region-specific approaches to educational improvement

References

- Link to the code
- Source of the dataset link
- Johnson, R. and Wichern, D., Applied Multivariate Statistical Analysis, Pearson, 2018, ISBN-13: 978-0134995397.

Thank You!

Questions?