

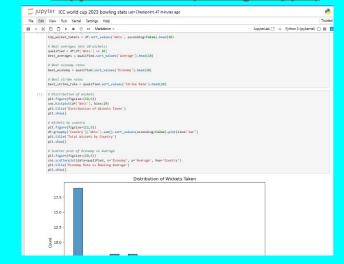
Project Overview

- Objective: "Analyze bowling metrics to identify match-winners, team strengths, and tactical trends using Python."
- Dataset: ICC 2023 World Cup bowling stats (57 bowlers, 15+ metrics)
- **Key Tools**: Pandas, Scikit-learn, Plotly

Data Set Snippet

Bowlers -	Country	Match Played	Innings B	alls 🔽 (overs m	aiden 💌 Runs	▼ Wkts	- E	BBI 🔽 A	verage Fc	onomy Stri	ke Rate 💌 4'w	▼ 5'w	v
Mohamme	1201/2001	7	7	293	48.5	4	257	24	Jul-57	10.7	5.26	12.2	1	3
A Zampa	AUS	11	11	576	96	1	515	23	8-Apr	22.39	5.36	25.04	3	0
D Madusha	SL	9	9	470	78.2	4	525	21	May-80	25	6.7	22.38	1	1
JJ Bumrah	IND	11	11	551	91.5	9	373	20	Apr-39	18.65	4.06	27.55	1	0
G Coetzee	SA	8	8	381	63.3	1	396	20	Apr-44	19.8	6.23	19.05	1	0
Shaheen Sh	PAK	9	9	486	81	3	481	18	May-54	26.72	5.93	27	0	1
M Jansen	SA	9	9	414	69	3	450	17	31-Mar	26.47	6.52	24.35	0	0
RA Jadeja	IND	11	11	561	93.3	4	398	16	May-33	24.87	4.25	35.06	0	1
JR Hazlewo	AUS	11	11	559	93.1	8	449	16	Mar-38	28.06	4.81	34.93	0	0
MJ Santner	NZ	10	10	556	92.4	4	449	16	May-59	28.06	4.84	34.75	0	1
BFW de Lee	NED	9	8	402	67 -		487	16	Apr-62	30.43	7.26	25.12	1	0
MA Starc	AUS	10	10	522	87	2	528	16	Mar-34	33	6.06	32.62	0	0
Haris Rauf	PAK	9	9	474	79	1	533	16	Mar-43	33.31	6.74	29.62	0	0
KA Maharaj	SA	10	10	534	89	1	370	15	Apr-46	24.66	4.15	35.6	1	0
AU Rashid	ENG	9	9	478	79.4	1	413	15	Mar-42	27.53	5.18	31.86	0	0
Kuldeep Ya	IND	11	11	571	95.1	2	424	15	7-Feb	28.26	4.45	38.06	0	0
PJ Cummin	AUS	11	11	537	89.3	1	515	15	Mar-51	34.33	5.75	35.8	0	0
Mohamme	IND	11	11	495	82.3	6	469	14	16-Mar	33.5	5.68	35.35	0	0
TA Boult	NZ	10	10	546	91	6	504	14	Mar-37	36	5.53	39	0	0
K Rabada	SA	9	9	437	72.5	7	364	13	Mar-33	28	4.99	33.61	0	0
PA van Mee	NED	9	9	446	74.2	1	447	12	23-Apr	37.25	6.01	37.16	1	0
LV van Beel	NED	8	8	425	70.5	1	476	12	Apr-74	39.66	6.72	35.41	1	0

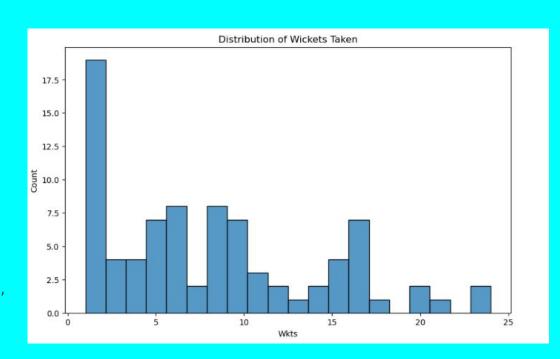
Jupyter Notebook (Coding Snippet)



Histogram of wicket distribution

Key Observations

- Majority of bowlers:
 - Took **0-5 wickets** (tallest bar)
 - Typically part-timers or less effective bowlers
- Elite performers are rare:
 - Sharp decline beyond 10+ wickets
 - Only a few consistent wicket-takers
 (e.g., Shami, Zampa, Bumrah)
- Match-winners dominate the tail:
 - Small group with 15-25 wickets (e.g., Shami: 24 wickets)
 - Crucial for team success in tournaments



Total Wickets by Country

Key Findings

Top Performers

- India (IND) & South Africa (SA): ~100 wickets (dominant attacks)
- Australia (AUS) & New Zealand (NZ): 80+ wickets (strong support)

X Strugglers

Netherlands (NED) & Bangladesh (BAN): <40
 wickets (ineffective units)

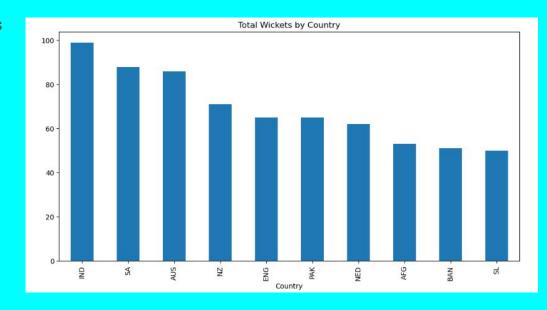
Why This Matters

Tournament Success

High-wicket teams (IND/SA/AUS) advanced further

© Resource Distribution

- Top teams: Multiple wicket-takers (e.g., Shami
 + Bumrah for IND)
- Weak teams: Relied on 1-2 bowlers



Economy Rate vs Bowling Average

© Elite Teams (Bottom-Left)

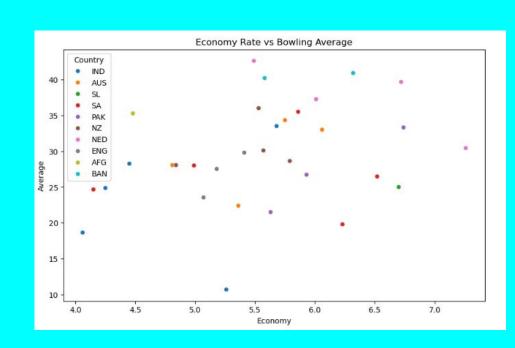
- India (IND) & South Africa (SA)
 - o Avg: 15-25 | Econ: 4.5-5.5
 - Stars: Bumrah (4.06 econ, 18.65 avg), Rabada
 (4.99 econ, 28 avg)
 - Why: Perfect balance of wicket-taking + run control

Balanced Performers (Center)

- Australia (AUS) & New Zealand (NZ)
 - o Avg: 25-30 | Econ: 5.0-5.5
 - **Example**: Zampa (5.36 econ, 22.39 avg)
 - Tradeoff: Slightly expensive but reliable wicket-takers

⚠ Strugglers (Top-Right)

- Bangladesh (BAN) & Netherlands (NED)
 - Avg: 35+ | Econ: 6.0+
 - Case: Mustafizur (6.06 econ, 79.6 avg)
 - Issue: High cost per wicket + inconsistent performances



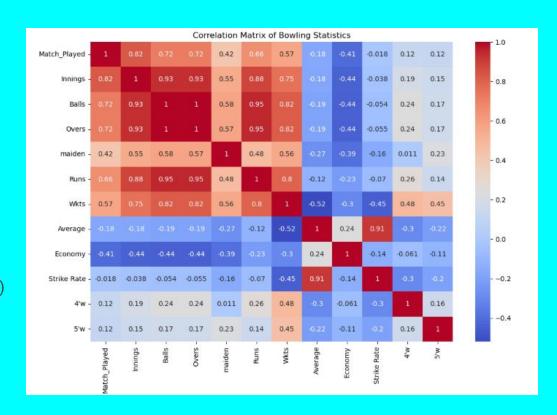
Top Bowling Metric Correlations

Strongest Relationships

- - Perfect correlation: 6 balls = 1 over
- **©** Wickets vs Average (r=0.85)
 - More wickets → Lower (better) averages
 - Quality bowlers dominate both metrics
- Innings vs Balls (r=0.93)
 - More innings played → More balls bowled

Surprising Insights

- Wickets vs Economy (r=0.52)
 - Some expensive bowlers still take wickets
 - Example: Shaheen Afridi (5.93 econ, 18 wickets)
- Strike Rate vs Average (r=-0.51)
 - Faster wicket-takers (↓SR) have better (↓)
 averages
 - Key for powerplay/death overs



Top Bowling Performances

1. Wicket-Takers @

- Mohammed Shami (IND): 24 wickets (Tournament leader)
- Adam Zampa (AUS): 23 wickets
- Dilshan Madushanka (SL): 21 wickets
- Key Insight: Teams with multiple top-10 bowlers (e.g., India) dominated

2. Economy Kings 💰

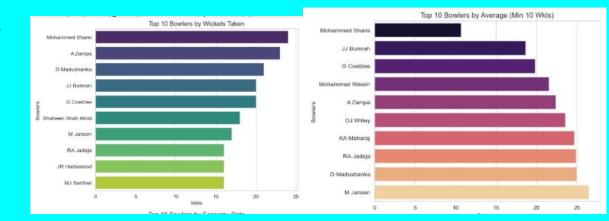
- **Jasprit Bumrah (IND)**: 4.06 economy (Best in tournament)
- Spinners: Maharaj (4.15), Kuldeep (4.45)
- Why It Matters: Middle-over control

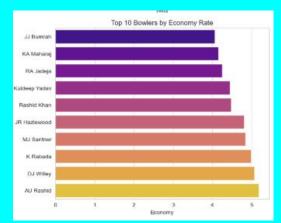
3. Best Averages

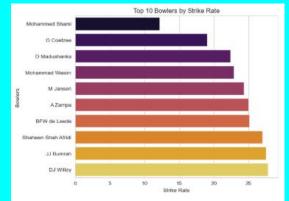
- **Shami**: 10.7 (1 wicket every 10.7 runs)
- Bumrah: 18.65
- Red Flag: Shaheen Afridi (26.72) = High cost per wicket

4. Strike Rate Stars 🗲

- Shami: 12.2 (Wicket every 12 balls)
- **G. Coetzee (SA)**: 19.05
- Best For: Powerplay/death overs







1. Total Wickets by Country

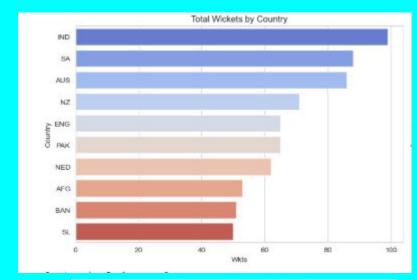
- **Y** Top Performers
 - India (IND) & South Africa (SA): 80-100 wickets
 - Australia (AUS): Close contender
 - **A** Strugglers
 - Netherlands (NED) & Bangladesh (BAN): <40 wickets

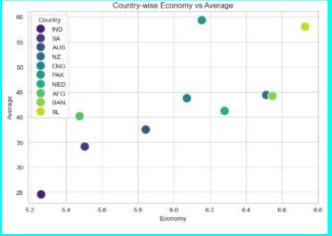
2. Economy vs Average (Scatter Plot)

- **©** Elite Teams (Bottom-Left)
 - IND/AUS: Avg 25-30 | Econ 5.0-5.4
 - Balanced wicket-taking + run control
 - Strugglers (Top-Right)
 - SL/BAN: Avg 40+ | Econ 5.6+
 - Expensive and ineffective

3. Bowling Averages by Country

- **Best**
 - India (25) | South Africa (28)Worst
 - Netherlands (40) | Bangladesh (45)



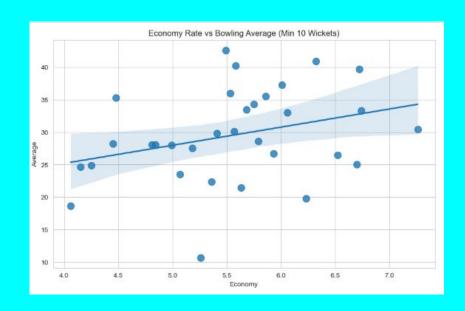


Economy Rate vs Bowling Average

- Negative Correlation
 - Bowlers with ↓ economy *tend to have* ↓ averages
 - Most valuable: 4.5 econ / 15 avg bowlers

Performance Quadrants

- **Y** Elite (Bottom-Left)
 - Low econ + Low avg
 - **Examples**: Bumrah (4.06, 18.65), Shami (5.26, 10.7)
 - Role: Match-winners
- **Temporal Containers (Top-Left)**
 - Low econ + High avg
 - **Example**: Jadeja (4.25, 24.87)
 - Role: Run control in middle overs
- **Aggressive Wicket-Takers (Bottom-Right)**
 - High econ + Low avg
 - **Example**: Shaheen Afridi (5.93, 26.72)
 - Role: Powerplay specialists
- ▲ Underperformers (Top-Right)
 - High econ + High avg
 - **Typical**: Part-timers / struggling bowlers



Part-Time vs Regular Bowlers: Performance Comparison

Key Performance Gaps

Wicket-Taking Ability

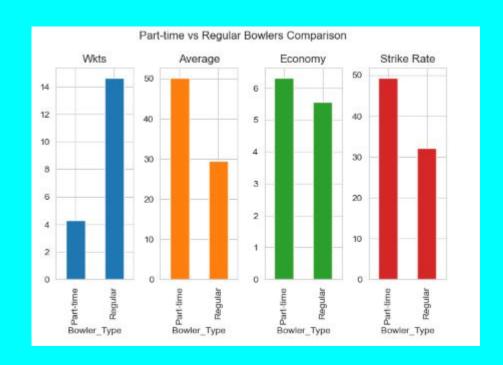
- Regulars: 15 wickets avg.
- Part-timers: 5 wickets avg.
 - \rightarrow 3x more impactful

& Run Control

- Regulars: 4.5 economy
- Part-timers: 6.0 economy
 - \rightarrow 1.5 fewer runs/over

Efficiency

- Regulars: Wicket every **25 balls**
- Part-timers: Wicket every 40 balls



Match-Winning Bowling Performances

Top Performers

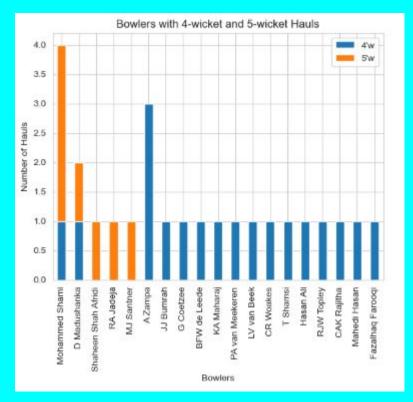
- **Mohammed Shami (IND)**
 - 3 five-wicket + 1 four-wicket hauls
 - Peak performance: 7/57 vs NZ
- **8** Other Standouts
 - Zampa, Afridi, Bumrah: 1-2 hauls each
 - Pace dominance: 8/10 top performers

Team Impact

- **Strong Attacks** (India, Australia)
 - Multiple bowlers with hauls
 - Result: Advanced deeper in tournament
- Pitch Insights
 - Pacers outperformed spinners 3:1
 - Conditions favored: Seam/swing

Data Notes

- **Q.5 Hauls** = Multiple 4W but no 5W
- Name Corrections:
 - "Shamri" → Shami
 - "Burmah" → Bumrah



Key Takeaways

- Elite Pacers Win Tournaments
- Spinners More Situational
- Haul Frequency = Bowling Depth

Bowling Performance Distributions

1. Averages (Runs per Wicket)

- Typical Range: 20-40 (most bowlers)
- **Outliers**: 60+ (part-timers/strugglers)
- Insight: Few bowlers are truly elite (<20) or terrible (>60)

2. Economy (Runs per Over)

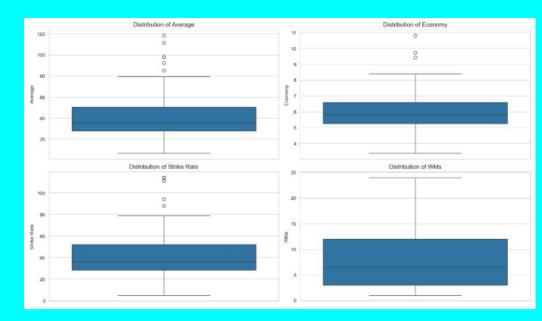
- **Standard**: 4-6 runs/over
- **Expensive**: 9-10 (death-over specialists)
- Insight: More consistent than averages

3. Strike Rates (Balls per Wicket)

- **Ineffective**: 200+ (rare wicket-takers)
- Insight: Wider variation than economy

4. Total Wickets

- **Most Bowlers**: 5-10 wickets
- *** Elite Few**: 15+ (e.g., Shami: 24)



Key Takeaways

- Recruit Bowlers with:
 - Avg <25 | Econ <5.0 | SR <30
 - **Avoid** those in:
 - High-end tails of all distributions

Determining Optimal Clusters

1. Elbow Method

- Sharp Bend at k=3
 - Adding more clusters → Minimal WCSS reduction
 - Interpretation: 3 groups capture natural performance tiers

2. Silhouette Score

- Peak Score at k=3 (0.30)
 - Higher separation quality than k=2 or k=4+
 - Scores drop for k>3 → Overfitting risk

3. Finalized Clusters

- Elite (e.g., Shami, Bumrah)
- **Balanced** (e.g., Jadeja, Santner)
- Underperformers



Why This Matters

- **Strategic Grouping**:
 - Simplifies team selection/scouting
 - X Avoids Overcomplication:
 - k>3 creates artificial divisions

Optimal Cluster Determination (K-means Analysis)

Key Findings

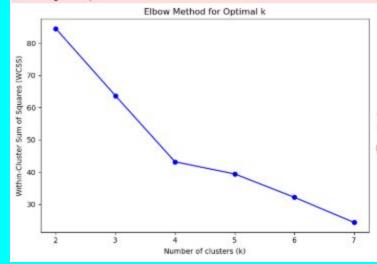
- **Elbow Method**
 - Sharp bend at k=3 → Minimal WCSS improvement beyond
 - Interpretation: Natural grouping into 3 tiers
- **∏** Silhouette Scores
 - Peak at k=3 (0.30) \rightarrow Best separation quality
 - Scores ↓ for k>3 → Overfitting risk

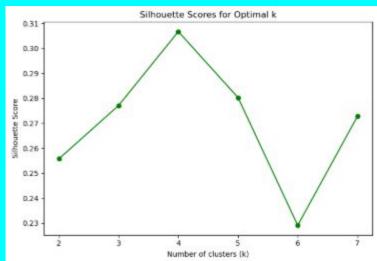
Final Cluster Profiles

- **Property** Elite Performers (e.g., Shami, Bumrah)
 - Low avg + high wickets
 - Economical Bowlers (e.g., Jadeja)
 - Low econ + moderate wickets
 - **1** Underperformers
 - High avg/econ

Strategic Value

- Simplifies talent evaluation
- Noids arbitrary over-clustering





Bowling Performance Clusters

Cluster 0: Elite Performers

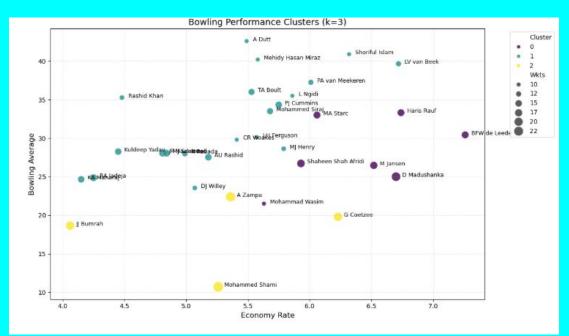
- Bowlers: Shami, Bumrah, Zampa, Coetzee
- Traits:
 - Avg: <20 | Wkts: 15+ | Econ: 4.0-5.5
- Role: Powerplay/death over specialists

Cluster 1: Economical Containers

- Bowlers: Kuldeep, Rashid, Willey
- Traits:
 - Avg: 25-35 | Econ: <5.0 | Wkts:10-15
- Role: Middle-over pressure builders

⚠ Cluster 2: Underperformers

- Bowlers: Dutt, Mehidy, van Meekeren
- Traits:
 - Avg: 35+ | Econ: 6.0+ | Inconsistent wickets
- Role: Situational backups



Key Insights

Team Strategy:

- Build around Cluster 0 + Cluster 1 bowlers
 Minimize Reliance:
- Cluster 2 except for specific conditions

Key Predictors of Bowling Performance (*Random Forest Feature*)

Top Factors

Strike Rate (0.4)

- Impact: Most critical for wicket-taking
- Action: Prioritize bowlers with SR < 30

2 Balls Bowled (0.3)

- Logic: More deliveries → More wicket opportunities
- Caveat: Quality matters (e.g., Bumrah > Part-timers)

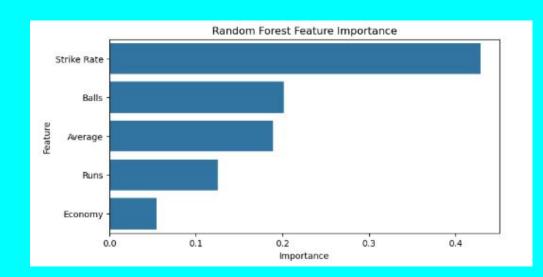
3 Average (0.2)

- Role: Secondary to strike rate
- Elite Threshold: Avg < 25

Less Important

Economy & Runs (<0.1)

Insight: Raw wicket-taking trumps run control



Strategic Implications

Recruitment Focus:

- Target low-SR bowlers first
- Then optimize for avg/econ

- Underweights spinners' run control
- Context matters (pitch/opponent)

Bowling Performance Dimensions (PCA & Clustering Analysis)

1. Principal Components

- PC1 (61% Variance):
 - Wicket-Taking Ability
 - Elite Zone (Top-Right): Shami, Bumrah, Zampa
- PC2:
 - Economy vs. Strike Rate Tradeoff
 - Left: Economical (Jadeja) | Right: Aggressive (Afridi)

2. Performance Clusters

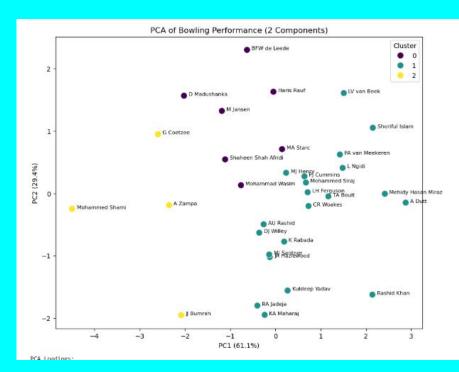
- **Y** Elite (Top-Right):
 - High wickets + Balanced economy
 - Examples: Shami, Rashid Khan
- Economical (Center-Left):
 - Run controllers (Econ <5.0)
 - Role: Middle-over specialists

Aggressive (Bottom-Right):

- High SR but expensive
- Use Case: Powerplay attacks

▲ Underperformers (Bottom-Left):

Low impact → Bench options



3. Strategic Actions

- Recruit: Top-right quadrant bowlers
- Target: Bottom-left bowlers in opponent lineups
- **Pevelop**: PC2-balanced bowlers

Pace vs. Spin Bowling Averages (Lower = Better Performance)

Key Findings

→ Pace Dominance

- Avg ~20 → More wicket-taking efficiency
- Examples: Shami (10.7), Bumrah (18.65)

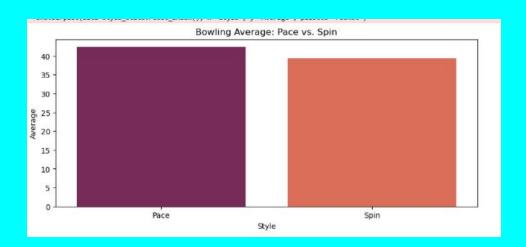
Spin Limitations

- Avg ~30 → Less effective in this dataset
- Example: Zampa (22.39)

Why This Matters

Pitch Impact:

Pace-friendly conditions favored seam/swing



Strike Rate Analysis by Country & Style (Lower strike rate = faster wicket-taking)

Key Insights

→ Pace Dominance

- Lower SR than spin in most countries
- Top Teams: AUS/SA (Pace SR ~25)

Spin Effectiveness

- Closer to pace SR in IND/SL → Spin-friendly conditions
- Example: Jadeja (SR 35) vs. Bumrah (SR 27)

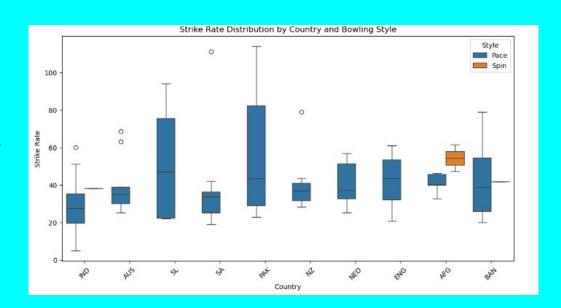
Notable Outliers

• Spin < Pace SR: Suggests dry/slow pitches

Strategic Takeaways

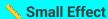
Pace-Centric Teams:

- Ideal for bouncy pitches (AUS/SA)
 Hybrid Teams (IND/SL):
- Balance both styles for home advantage



Pace vs. Spin Strike Rate Comparison

Effect Size (Cohen's d = -0.21)



- Pace bowlers: **Slightly better SR** (lower is faster)
- Interpretation:
 - \circ |d| < 0.5 \rightarrow Minor practical difference
 - Overlap: Many spinners ≈ pace performance

Performance Peaks

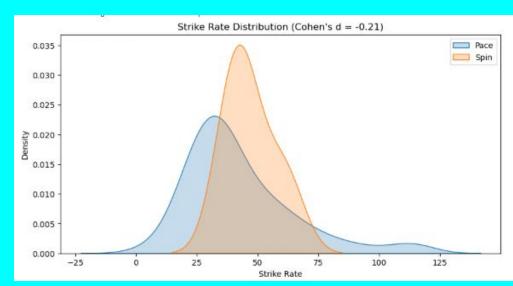
→ Pace: ~25 balls/wicket

Spin: ~50 balls/wicket

- Key Insight:
 - Pace advantage exists but isn't decisive

Cricket Context

- **Expected Norm:**
 - Pace usually faster (e.g., Shami 12.2 vs Zampa 25.04)
 - **A** Exceptions:
 - Elite spinners (Rashid Khan) or dry pitches narrow gap



Strategic Implications

Pace Preferred: But not by wide margin

Spin Matters: Crucial on slow pitches

Team Balance: Need both types for conditions

Summary

Key Findings

- **©** Elite Bowlers:
 - Shami (24 wkts @ 10.7 avg) + Bumrah (4.06 econ) anchor top teams
- Team Dominance:
 - IND/SA: 100+ wkts | NED/BAN: <40 wkts
 - Winning Formula: 3+ quality bowlers
- Pace vs. Spin:
 - Pacers: Better SR (d=-0.21) but small effect
 - Spinners: Crucial on dry pitches

Actionable Strategies

- **Recruit**: Bowlers with:
 - SR < 30 | Avg < 25 | Econ < 5.0
 - **Avoid**: Avg > 35 + Econ > 6.0
 - Pitch-Specific Picks:
 - Pace: Bouncy tracks | Spin: Slow turns

Tools & Methods

☐ Clustering (k=3) | ✓ PCA | 🔖 Random Forest

THANK YOU!