**Slide 1 — Title**

**Monsoon & Macro Signals for the NIFTY Midcap 100  
*AI/ML Capstone***

* **Student: Abhilash Panicker**
* **Mentor: Abhijit Chakraborty**
* **M.S. Data Analytics – Capstone 2025**

**Slide 2 — Formatting Guide**

***(keep the template’s reminder, or shorten to one line)***

**All tables/figures labelled; abbreviations defined on first use. No spelling or grammar errors.**

**Slide 3 — Introduction**

* **Problem Statement  
  India’s monsoon drives GDP, inflation and investor sentiment, yet most equity models ignore weather completely. Can rainfall & macro surprises improve next-quarter forecasts for mid-cap returns?**
* **Gap in Literature  
  Existing Indian studies focus on index re-balancing or daily weather quirks—none combine rainfall anomaly + GDP/CPI/PMI + repo-rate in a predictive model.**
* **Goal  
  Build a lightweight, reproducible pipeline—collect public data, test a baseline vs. enriched model, and stress-check results over COVID and rate-hike regimes… all in 7 days.**

**Slide 4 — Scope & Objectives (Research Questions)**

| **Research Question** | **Research Question (plain English)** |
| --- | --- |
| **1. Predictive Boost** | **Does adding rainfall-anomaly + macro surprises help a simple ML model beat a lag-return baseline for 1-quarter-ahead excess returns?** |
| **2. Monsoon Effect** | **Do index returns differ significantly between “good-rain” and “poor-rain” seasons?** |
| **3. Robustness** | **Does the enriched model hold up during COVID crash (Q1–Q2 2020) and rate-hike cycle (Q2 2022–Q4 2023)?** |
| **4. Lead–Lag Channel** | **Does monsoon rainfall anomaly foreshadow GDP growth 1–2 quarters ahead, and does that link further boost equity-return forecasts?** |

**Slide 5 — Sample-Size Calculation**

* **Observations: 2010 Q1 → 2025 Q2 ⇒ 62 quarters.**
* **Regression power check (α = 0.05, desired power = 0.80, expected f² ≈ 0.15): needs 55 points → *we have 62 ⇒ sufficient.***
* **Good-vs-poor monsoon t-test: rule-of-thumb ≥ 30 observations per group; we have 34 “good” & 28 “poor”.  
  *Time-series methods rely more on rolling splits than sheer N, but we comfortably exceed minimums.***

**Slide 6 — Data Description**

| **Layer** | **Source (free)** | **Key Fields** |
| --- | --- | --- |
| **Index price** | **NSE bhavcopy (daily) → quarterly TR** | **Close, splits** |
| **Rainfall anomaly** | **IMD Southwest-Monsoon report (annual XLS)** | **% vs 50-yr mean** |
| **GDP YoY** | **MOSPI / RBI Handbook CSV** | **q/q annualised %** |
| **CPI YoY** | **RBI time-series** | **monthly % (rolled to Q)** |
| **PMI** | **IHS-Markit press releases** | **index (≥ 50 = expansion)** |
| **Repo-rate Δ** | **RBI DB** | **change in bps during Q** |

**All series merged to quarter-end, stored in a single CSV (≈ 62 × 7).**

**Slide 7 — Analytic Approach (overview)**

1. **ETL – read CSV/XLS, resample to Q, forward-fill rainfall.**
2. **Feature set**
   * **Lagged index return (baseline)**
   * **Lagged GDP, CPI, PMI, repo-rate, rainfall anomaly (enriched)**
3. **Models**
   * **Baseline: ElasticNet on lag return only**
   * **Enriched: LightGBM on all features**
4. **Validation: 5-split walk-forward (train : test windows ≈ 10 : 2 Q).**
5. **Tests**
   * **Question 1: R², MAE comparison + SHAP bar chart**
   * **Question 2: two-sample *t* & KS on good vs poor monsoon returns**
   * **Question 3: run fixed model through crisis quarters; check Δ R² / Δ Sharpe < 15 %.**

**Slide 8 — Analytic Approach • Question 1**

* **Target: next-quarter excess return (NIFTY Midcap 100 TR – NIFTY 50 TR).**
* **Baseline vs Enriched: ElasticNet vs LightGBM.**
* **Success Rule: Enriched model raises out-of-sample R² by ≥ 0.10 *and* cuts MAE ≥ 10 %.**

**Slide 9 — Analytic Approach • Question 2**

* **Define seasons**
  + ***Good rain:* rainfall anomaly ≥ +4 % (IMD definition “Above Normal”)**
  + ***Poor rain:* ≤ –4 % (“Below Normal”)**
* **Test: two-sample *t* and KS on the distribution of same-quarter excess returns.**
* **Significance: |Δ mean| > 0.5 σ and p < 0.05 → rainfall matters.**

**Slide 10 — Analytic Approach • Question 3**

* **Shock windows**
  + **COVID crash: 2020 Q1–Q2**
  + **Rate hikes: 2022 Q2–2023 Q4**
* **Metrics: ΔR², ΔSharpe, max drawdown vs full sample.**
* **Pass mark: degradation < 15 %.**

**Slide 11— Analytic Approach • Question 4**

* **Step 1: Econometrics**
  + **OLS: GDP\_{t+1} = α + β · Rain\_t and same for t+2.**
* **Significance: β p < 0.05, R² ≥ 0.08.**
* **Step 2 Pipeline Boost**
  + **Add gdp\_pred\_from\_rain to feature list.**
  + **Re-run walk-forward; compare Δ R² & Δ MAE.**
* **Pass mark – Additional R² ≥ 0.03 or MAE ↓ ≥ 5 % over RQ-A enriched model.**

**Slide 11 — Evaluation Metrics**

| **Metric** | **Formula (text)** | **Used in** |
| --- | --- | --- |
| **R²** | **1 – Σ(ŷ–y)² / Σ(y–ȳ)²** | **A, C** |
| **MAE** | **Σ** | **ŷ–y** |
| **Sharpe** | **(μ\_p – r\_f) / σ\_p** | **C** |
| **KS Statistic** | **max** | **F\_good(x)–F\_poor(x)** |
| **t-statistic** | **(μ\_good–μ\_poor)/SE** | **B** |

**Slide 12 — Recommendations & Applications**

1. **Fund-house mid-cap desks – use rainfall & macro watch-list as an overlay on existing factor models.**
2. **Sell-side research – quarterly “Monsoon Tracker” note becomes a value-add service.**
3. **Policy analysts – quantify how repo-rate actions mediate weather shocks on equity risk premium.**
4. **Open notebook – fast, transparent code for students & quants to extend.**

**Slide 13 — References & Bibliography**

1. **IMD. *End-of-Season Southwest Monsoon Report* (2001-2024).**
2. **RBI. *Handbook of Statistics on Indian Economy*, 2024.**
3. **MOSPI. *National Accounts, Quarterly GDP* press releases, 2010-2025.**
4. **IHS-Markit. *India Manufacturing PMI Methodology*, 2024.**
5. **Fama & French. “Common Risk Factors,” *J. Finance*, 1993.**
6. **Tsay, R. *Analysis of Financial Time Series*, 2023.**
7. **Wright Research Blog. “Does Monsoon Impact Nifty?” 2024.**
8. **StockEdge Blog. “Monsoon & Market Returns,” 2025.**
9. **Bansal et al. “Predicting Nifty Upgrades,” IIM W.P., 2022.**
10. **Lalwani & Meshram. “Machine-Learning Indian Cross-Section,” *SSRN* 4511996, 2023.**

**Slide 14 — Formatting checklist**

***(Use template’s slide if required; otherwise remove.)***

**All figures labelled, abbreviations defined, spell-checked, logical flow confirmed.**

**Slide 15 — Appendix (optional)**

* **SHAP bar chart of enriched model (top 6 drivers).**
* **Table: baseline vs enriched metrics.**