PowerBI Project: Pakistan Large Scale Manufacturing (LSM) Dashboard

FY 2024-2025, 2025-2026 (Jul-Nov)

Hannan Baig

Contents

1. Executive Summary & Strategic Intent

This project was designed to simulate the real reporting environment of Pakistan's Ministry of Industries and Production. Rather than relying on sanitized or generic practice datasets, the analysis uses official Large Scale Manufacturing (LSM) indices prepared for high-level government briefings.

The objective extended beyond simple visualization to demonstrate applied economic monitoring capability, with emphasis on:

- **Domain Specificity:** Navigating Pakistan Standard Industrial Classification (PSIC) codes and adhering to ministry-defined sector weightages.
- **Policy Relevance:** Differentiating between *Headline Growth* (which can be misleading) and *Weighted Contribution* (which reflects actual economic volume).
- **Conflict Detection:** Identifying structural divergences such as a K-shaped recovery, where sectors like automobiles expand while industrial inputs contract.

2. Data Strategy & Modeling

Government data is frequently distributed in wide, report-ready formats that are unsuitable for scalable analysis. To overcome this, a flat-file approach was rejected in favor of a **Star Schema** architecture.

2.1 Data Transformation (ETL)

Excel and Power Query were used to restructure raw administrative data into a normalized analytical model. Static "July–November" columns were converted into a transactional structure suitable for time-intelligence calculations.

The model was decomposed into three entities:

- **Fact_Production (Transactional Table):**
  - Raw data was unpivoted into a tall table containing production values indexed by Fiscal Year and Period.
  - This design enables scalability; new months can be appended as rows without restructuring columns.

No	PSIC_Code	Item_Name	Fiscal_Year	Period	Value
1	10721	SUGAR	2024-2025	Jul-Nov	191366
2	12000	CIGARETTES	2024-2025	Jul-Nov	15398
3	18110	PRINTING	2024-2025	Jul-Nov	29597
4	18110	WRITING	2024-2025	Jul-Nov	46790
5	17020	PACKING	2024-2025	Jul-Nov	76325
6	17020	PAPER BOARD	2024-2025	Jul-Nov	182003
7	16210	CHIP BOARD	2024-2025	Jul-Nov	0
8	20110	SODA ASH CAUSTIC SODA 1	2024-2025	Jul-Nov	329689
9	20110	SODA ASH CAUSTIC SODA 2	2024-2025	Jul-Nov	196964
10	20120	UREA	2024-2025	Jul-Nov	2887955
11	20120	AMM.NITRATE	2024-2025	Jul-Nov	343536
12	20120	AMM.SULPHATE	2024-2025	Jul-Nov	0
13	20120	NITRO PHOSPHATE	2024-2025	Jul-Nov	338614
14	20120	S.PHOSPHATE	2024-2025	Jul-Nov	33500

Facts_Production | Dim_Items | Dim_Calendar

Figure 1: Facts Production

- **Dim_Items (Context Table):**
  - Metadata such as PSIC codes, item names, units of measurement, and sector weights were isolated into a dedicated dimension.
  - **Strategic Decision:** Items were manually categorized into broader sectors (e.g., "Jeeps/SUVs" under Automotive, "Cement" under Construction Materials) to enable hierarchical drill-downs unavailable in the raw data.

No	PSIC_Code	Item_Name	UOM	Weightage	Source	Category
1	10721	Sugar	Tonnes	3.4270	Cane Commissioner	Food & Beverages
2	12000	CIGARETTES	Mln. Nos	2.0722	Direct Units	Tobacco
3	18110	PRINTING	Tonnes	0.4065	PBS	Paper & Packaging
4	18110	WRITING	Tonnes	0.4065	PBS	Paper & Packaging
5	17020	PACKING	Tonnes	0.4065	PBS	Paper & Packaging
6	17020	PAPER BOARD	Tonnes	0.4065	PBS	Paper & Packaging
7	16210	CHIP BOARD	Tonnes	0.0894	Direct Units	Paper & Packaging
8	20110	SODA ASH CAUSTIC SODA 1	Tonnes	0.0776	Direct Units	Chemicals
9	20110	SODA ASH CAUSTIC SODA 2	Tonnes	0.2304	Direct Units	Chemicals
10	20120	UREA	Nutrient Tonnes	1.143	Direct Units	Fertilizers
11	20120	AMM.NITRATE	Nutrient Tonnes	1.143	Direct Units	Fertilizers
12	20120	AMM.SULPHATE	Nutrient Tonnes	1.143	Direct Units	Fertilizers
13	20120	NITRO PHOSPHATE	Nutrient Tonnes	0.1252	Direct Units	Fertilizers
14	20120	S.PHOSPHATE	Nutrient Tonnes	0.1252	Direct Units	Fertilizers
15	20120	DI.AMM.PHOSPHATE	Nutrient Tonnes	0.1252	Direct Units	Fertilizers
16	20120	NPK	Nutrient Tonnes	0.1252	Direct Units	Fertilizers
17	23110	SHEET/FLOAT GLASS	Th.M2	0.3602	Direct Units	Non-Metallic Minerals
18	23940	CEMENT	Th.Tons	4.650	Direct Units	Construction Material
19	19100	COKE	Tonnes	0.0010	PAK STEEL MILLS	Iron & Steel

Facts_Production Dim_Items Dim_Calendar

Figure 2: Dim Items

- **Dim_Calendar (Time Intelligence):**
  - A dedicated date table was created to manage fiscal year ordering (FY 2024–25 vs. FY 2025–26).
  - This ensures mathematically correct Month-over-Month and Year-over-Year comparisons.

2.2 Why This Matters

The Star Schema ensures that all Power BI measures are computed through relational logic rather than hard-coded spreadsheet formulas, increasing robustness, transparency, and analytical flexibility.

Figure 3: Dim Calendar

3. Data Quality & Modeling Challenges

Real-world administrative data is rarely analysis-ready. Initial inspection of the ministry dataset revealed three structural fractures that would have undermined automated reporting integrity.

3.1 The "Double-Counting" Trap (Granularity)

Issue: The dataset mixed atomic items (e.g., Jeeps, Urea) with aggregated sector totals (e.g., Automobiles Total, Fertilizers Total).

Risk: Direct aggregation would inflate values due to double-counting sector totals alongside their components.

Fix: Strict granularity was enforced by removing all pre-calculated totals and header rows from the fact table.

- **Strategic Insight:** Dashboards should never rely on Excel-based totals. Retaining only atomic data allows Power BI to compute totals dynamically and enables flexible slicing across sectors.

3.2 The "Null vs. Zero" Decision (Statistical Integrity)

Issue: The Value column contained 14 null entries, ambiguous between missing data and zero production.

Decision: All nulls were replaced with zeros.

Reasoning: In policy analysis, a production halt implies zero output, not missing data. Zero values allow detection of severe contractions (e.g., -100% growth), whereas nulls would suppress such signals during aggregation.

3.3 Primary Key Failure (Model Integrity)

Issue: PSIC codes were initially selected as primary keys but were found to be non-unique, with single codes reused across multiple chemical products.

Fix: A cleaned `Item_Name` field was used as the joining key to enforce a one-to-many relationship.

- **Outcome:** This eliminated many-to-many relationship conflicts in Power BI and ensured correct propagation of weights and sector classifications.

4. The Data Model: Moving Beyond Flat Files

Once the data hygiene protocols were executed in Excel, the three cleaned tables—`Facts_Production`, `Dim_Items`, and `Dim_Calendar`—were imported into Power BI.

The primary design objective at this stage was **scalability**. A government dashboard is not a static snapshot; it is a living monitoring system that receives new data every month. Had the data been retained in a single flat table, each new reporting cycle would require manual column restructuring. By normalizing the data into a **Star Schema**, future updates are handled simply by appending rows.

4.1 Relationship Logic (The One-to-Many Principle)

Relationships were defined to ensure that contextual information from the dimension tables flows correctly into the transactional fact table.

4.1.1 Item Context (`Dim_Items` → `Facts_Production`)

- **Cardinality:** One-to-Many (1:*).
- **Join Key:** `Item_Name`.
- **Reasoning:** After cleaning item names and removing ambiguous PSIC codes, this relationship became the analytical backbone of the model. It allows sector weightages and categories from the dimension table to correctly filter and calculate raw production values stored in the fact table.

4.1.2 Time Intelligence (`Dim_Calendar` → `Facts_Production`)

- **Cardinality:** One-to-Many (1:*).
- **Join Key:** `Fiscal_Year`.
- **Reasoning:** This relationship enables accurate "Current vs. Last Year" comparisons. Selecting FY 2025–26 through a slicer automatically maps the correct historical baseline (FY 2024–25) without manual intervention.

4.2 The Schema Advantage

This architecture cleanly separates definitions from data.

- If the Ministry revises the weight of "Sugar" in a future year, only a single row in `Dim_Items` needs updating; the historical production records remain untouched.
- The structure mirrors enterprise-grade data warehouse design, demonstrating that the solution is built for long-term operational use rather than a one-off visualization.

5. Analytical Logic & DAX Strategy

To prevent analytical distortion caused by default aggregations, a dedicated measure table (`_Key Measures`) was created. All business logic was explicitly scripted using DAX (Data Analysis Expressions).

5.1 Fiscal Year Logic (Time Intelligence)

Standard time-intelligence functions such as `SAMEPERIODLASTYEAR` often fail under custom fiscal calendars. To ensure robustness, calculations were engineered using the `Sort_Order` column from the calendar dimension.

5.1.1 Production (Current Fiscal Year)

```
Production CY =
CALCULATE(
    [Total Production],
    Dim_Calendar[Sort_Order] = 2
)
```

Logic: This measure isolates the active monitoring period (FY 2025–26), ensuring that no historical data contaminates the current status KPIs.

5.1.2 Production (Previous Fiscal Year)

```
Production PY =
CALCULATE(
    [Total Production],
    Dim_Calendar[Sort_Order] = 1
)
```

Logic: This creates a fixed baseline (FY 2024–25) against which all growth calculations are measured.

5.2 Weighted Impact (The Critical Measure)

Headline growth percentages are economically misleading when viewed in isolation. A small industry doubling output is far less consequential than marginal growth in a dominant sector.

To address this, a **Weighted Contribution** measure was implemented using an iterator function.

5.2.1 Formula

```
Weighted Impact =
SUMX(
    Dim_Items,
    [Growth %] * Dim_Items[Weightage]
)
```

5.2.2 Why `SUMX` Was Required

A simple aggregation would fail in this context. The engine must:

1. Iterate over each individual industry.
2. Compute item-specific growth.
3. Multiply that growth by the Ministry-assigned weight.
4. Aggregate the weighted results to compute total index impact.

This logic enables a policy-relevant narrative: a sector may grow rapidly, but another may contribute more materially to economic output.

6. Technical Challenges: The Many-to-Many Trap

During relationship modeling, Power BI attempted to default the link between `Dim_Items` and `Facts_Production` to a Many-to-Many relationship.

6.1 Diagnosis

Power BI detected non-unique values in the `Item_Name` column of the dimension table, caused by invisible whitespace and capitalization inconsistencies (e.g., "Sugar" vs. "Sugar ").

6.2 Strategic Decision

The process was halted immediately. Accepting a Many-to-Many relationship in a Star Schema introduces ambiguity in filter propagation and creates a high risk of silent double-counting during slicing.

6.3 Resolution

The issue was resolved at the ETL layer:

1. **Normalization:** Duplicate values were removed from the `Item_Name` column.
2. **Validation:** The row count reduced to the expected unique total (33 items).
3. **Result:** Power BI correctly detected a One-to-Many (1:*) relationship with single-direction filtering.

This enforcement guarantees that each production record maps to exactly one unique item.

7. Model Auditing: Overriding Automated Defaults

Reliance on automatic relationship detection is a common failure point in Power BI models. A manual audit revealed a critical logic error introduced by the engine.

7.1 The Double-Wiring Conflict

Power BI detected a shared column (`No`, representing row indices) in both the dimension and fact tables and automatically linked them.

7.2 Risk

This relationship was semantically invalid. Row numbers carry no business meaning and would have linked unrelated records based purely on spreadsheet position, silently corrupting all calculations.

7.3 Correction

A manual override was executed:

1. **Deactivation:** The erroneous relationship based on `No` was deleted.
2. **Enforcement:** Relationships were restricted exclusively to:
   - `Item_Name` for product context
   - `Fiscal_Year` for temporal alignment

By removing this "double wiring," the model was rendered fully deterministic—calculations now depend solely on explicit business rules rather than structural coincidences.

8. Advanced Data Modeling: Resolving Cardinality Conflicts

Building a robust analytical model requires more than visual relationship mapping; it requires strict enforcement of **referential integrity**.

During the relationship modeling phase, a persistent Many-to-Many warning emerged between the dimension table (`Dim_Items`) and the fact table (`Facts_Production`). This was treated as a critical stop-work condition. Accepting a Many-to-Many relationship in a Star Schema introduces ambiguity in filter propagation and creates a high risk of incorrect aggregations (double-counting) during interactive slicing.

8.1 Diagnosis: The "Invisible" Duplicate

The issue persisted even after applying a standard `Remove Duplicates` transformation, indicating a deeper conflict between the two engines operating within Power BI:

1. **ETL Layer (Power Query):** Case-sensitive; treats `''Sugar''` and `''SUGAR''` as distinct values.

2. **Modeling Layer (VertiPaq Engine):** Case-insensitive; treats `''Sugar''` and `''SUGAR''` as identical.

This discrepancy allowed text variants to pass as unique during transformation while being flagged as duplicates during model validation. Additionally, invisible whitespace (e.g., trailing spaces) caused strings such as `''Cement ''` and `''Cement''` to be interpreted as separate items in the ETL layer.

8.2 Solution: Text Standardization Protocol

To enforce strict one-to-one uniqueness in the dimension table, a deep-clean normalization procedure was implemented in Power Query. Three transformations were applied to the `Item_Name` primary key in both tables:

1. **TRIM:** Removed leading and trailing whitespace common in manually entered administrative data.

2. **CLEAN:** Eliminated non-printable control characters such as line breaks and tab markers.

3. **UPPERCASE:** Neutralized case-sensitivity conflicts by converting all text to a uniform uppercase representation.

This ensured that `''Sugar''`, `''sugar''`, and `''SUGAR''` were treated as mathematically identical values across all layers.

8.3 Outcome

After reapplying the `Remove Duplicates` step to the cleaned dimension table, the model successfully validated a One-to-Many (1:*) relationship with a single-direction cross-filter.

This rigorous standardization guaranteed that every production record in the fact table maps to exactly one unique item in the dimension table, ensuring 100% accurate filtering and weight propagation throughout the dashboard.

9. Data Validation & Audit (The QA Matrix)

Before designing any visualizations, a raw matrix visual was constructed to audit the mathematical integrity of the model.

- **Sanity Check:** High-growth items such as Cars (81.79%) correctly generated a positive weighted impact (0.74), confirming accurate weight application.
- **Conflict Detection:** The Cast/Rolled Billet sector correctly displayed negative growth (-10.19%), validating the K-shaped recovery hypothesis where industrial inputs contract despite headline growth.

10. Quality Assurance: The "Mixed Unit" Trap

During final measure validation, a critical logical flaw was identified in the `Total Production` aggregation. The initial calculation returned a headline figure of 11.7 million.

10.1 Diagnosis

Further inspection revealed that this value was a sum of heterogeneous units of measurement. The model was aggregating counts (Cars, Jeeps), weights (Fertilizers in tons), and volumes (Cigarettes in cases).

- **Mathematically:** The aggregation was valid.
- **Economically:** The result was meaningless. Units such as "1 Car" and "1 Ton of Urea" cannot be combined into a coherent indicator.

10.2 Correction

The `Total Production` measure was immediately removed from all high-level KPI cards to prevent stakeholder misinterpretation.

- **Decision:** Raw production volumes were restricted to row-level views where units remain consistent.
- **Substitution:** For aggregate reporting, `Total Weighted Impact` (2.67) was enforced as the headline metric. As a dimensionless index contribution, it is the only mathematically valid measure for cross-sector aggregation.

11. Advanced Segmentation: The Policy Narrative

To move the analysis beyond simple growth rates, the dashboard was designed to answer a core policy question: *Who is driving the recovery, and who is dragging it down?*

Traditional industry classifications (e.g., Textile, Automotive) are static and descriptive. For policy analysis, a dynamic segmentation based on *current economic impact* is required.

11.1 The Performance Flag Measure

A dynamic segmentation measure was engineered using the DAX `SWITCH` function to classify each sector into four actionable performance buckets based on its weighted contribution.

11.1.1 Logic

```
Performance Flag =
SWITCH(
    TRUE(),
    [Weighted Impact] > 0.1,   "High Growth Driver",
    [Weighted Impact] > 0,     "Stable / Moderate",
    [Weighted Impact] > -0.05, "Minor Contraction",
    "Major Drag"
)
```

11.2 Business Value

This measure fundamentally transformed the dashboard's analytical capability. Instead of filtering only by static sector names, decision-makers can now filter by *economic behavior.*

- It immediately isolates structurally weak industries such as Printing (0.13 weighted impact) and Soda Ash, enabling targeted root-cause analysis without manually scanning dozens of rows.
- It shifts the analytical narrative from "What happened?" to "Where is the crisis?".

12. Dashboard Architecture: The "Executive Monitor" Framework

Senior government officials do not have the time to explore detailed tables. They require a concise, decision-ready view that communicates risks and opportunities within seconds.

The dashboard canvas was structured into three hierarchical zones, beginning with a Headline Layer that establishes immediate economic context.

12.1 The Headline Layer (KPI Strategy)

The top section of the dashboard was engineered to answer the three most common briefing questions:

1. What is the net performance?
2. What is driving growth?
3. What is the largest drag on the economy?

12.2 The Anchor Metric: Net Weighted Growth

As established during data validation, aggregating raw production volumes across heterogeneous units creates a statistical fallacy. Therefore, the primary KPI card displays **Net Weighted Growth** (2.67).

- **Design Choice:** All secondary labels and visual clutter were removed, positioning this value as the "North Star" metric of the report.
- **Interpretation:** This metric represents the true, weighted pulse of the industrial sector.

12.3 Automated Outlier Detection (Top/Bottom N Logic)

Static narrative text becomes obsolete as soon as new data arrives. To ensure scalability, algorithmic filtering was used to automatically identify top contributors and major drags.

12.3.1 Top Growth Driver

- **Technique:** A Top N (1) visual-level filter was applied to `Item_Name`, sorted by `[Weighted Impact]`.
- **Result:** The dashboard automatically identified **CARS** as the highest positive contributor, reinforcing the "Auto Illusion" narrative in which automotive recovery masks broader industrial weakness.

12.3.2 Major Drag Sector

- **Technique:** The visual was duplicated and a Bottom N (1) filter applied.
- **Result: PRINTING** was instantly flagged as the largest negative contributor.

12.4 Significance of the Architecture

By relying on algorithmic filters rather than static text, the dashboard remains fully dynamic. As new monthly data is ingested, the headline cards automatically update to reflect the evolving industrial landscape—whether growth shifts to Sugar or contraction deepens in Textiles—without requiring manual analyst intervention.

13. Visualization Strategy: The “Traffic Light” Protocol

A recurring failure in government reporting is the creation of a “wall of data”—dense tables that demand excessive cognitive effort to interpret. The objective of the LSM Executive Monitor was to reduce time-to-insight to under five seconds.

To achieve this, standard Power BI legends (which are static) were deliberately avoided. Instead, a dynamic conditional-formatting system was engineered directly into the visuals.

13.1 The Technical Challenge: Dynamic Coloring

A key limitation in Power BI is that calculated measures (such as `[Performance Flag]`) cannot be used as legends in standard charts.

13.1.1 Solution: The Performance Color Measure

A dedicated formatting measure, `[Performance Color]`, was created using explicit hexadecimal color codes to function as a decision-maker–friendly “traffic light” system:

- **#006400 (Dark Green):** High Growth Drivers (Impact > 0.1)
- **#32CD32 (Lime Green):** Stable / Moderate (Impact > 0)
- **#FF8C00 (Dark Orange):** Minor Contraction (Impact > -0.05)
- **#8B0000 (Dark Red):** Major Drag (Crisis Zone)

By binding this measure to the “Field Value” property of bars and markers, visual signals were guaranteed to align with mathematical reality. If a leading sector collapses in a future update, it will automatically shift to red without manual intervention.

13.2 Visual 1: Winners & Losers (Bar Chart)

Goal: To visualize the K-shaped recovery identified in earlier analysis.

Design: Sectors were placed along a horizontal axis and sorted strictly by `[Weighted Impact]`.

Insight: The chart immediately confirms structural divergence. The upper range is dominated by Automobiles (green), while the lower range is anchored by Printing, Paper, and Soda Ash (red).

13.3 Visual 2: Risk vs. Relevance Matrix (Scatter Plot)

Goal: To separate economic signal from statistical noise.

Design: A scatter plot correlating two competing variables:

- **X-axis:** Growth % (Velocity)

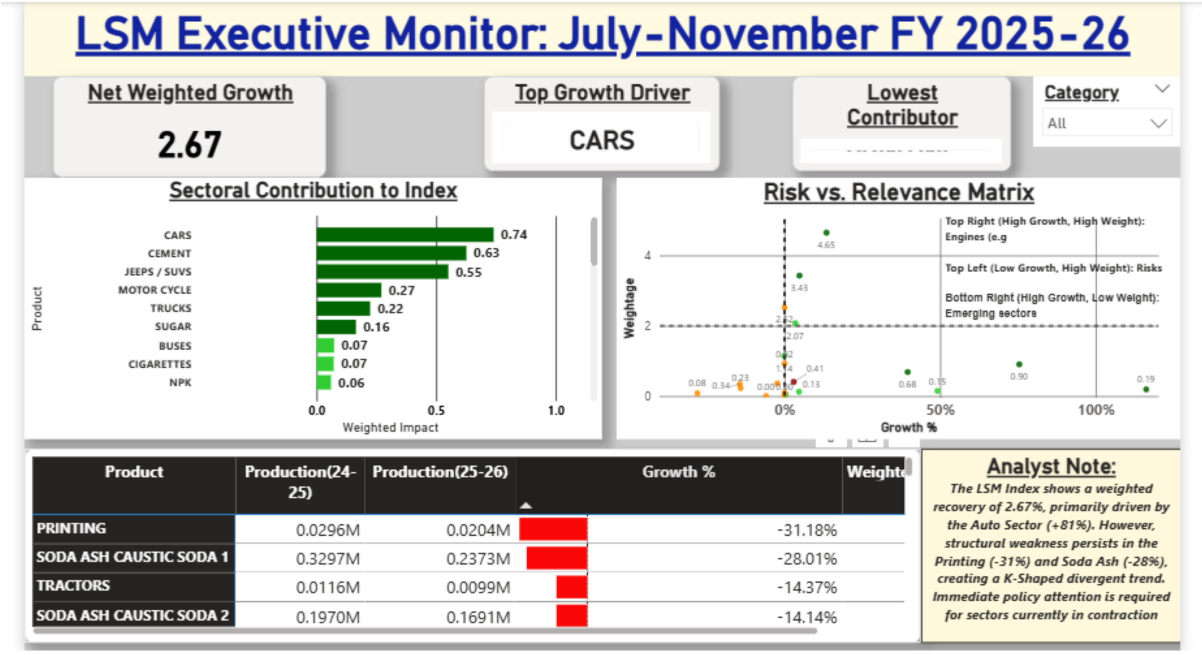

Figure 4: LSM Executive Dashboard Overview

- **Y-axis:** Weightage (Volume)
- **Tooltips:** Sector name

Strategic Value: This visualization serves as the primary policy-prioritization tool.

- **Top-Right Quadrant (High Weight, High Growth):** Engines of the economy (e.g., Cars).
- **Top-Left Quadrant (High Weight, Negative Growth):** "Code Red" risks. Movement of a heavy sector (e.g., Textiles or Food) into this quadrant signals an immediate GDP threat, even if smaller sectors show growth.

14. The Operational Layer: Auditability & Interaction

While visuals identify patterns, government decisions require precise, auditable numbers. To support this, the lower third of the dashboard was reserved for a detailed matrix.

14.1 The "Government-Style" Ledger

A matrix visual was designed to function as the dashboard's audit trail.

- **Columns:** Weightage → Production (PY) → Production (CY) → Growth %.
- **Logic:** This structure allows stakeholders to mentally validate calculations by tracing raw volumes, assigned weights, and resulting growth rates.

14.2 Cognitive Aids: Data Bars

Tables containing dozens of percentages are cognitively demanding and error-prone. To mitigate this, conditional formatting with data bars was applied to the Growth % column.

- **Green Bars (Right):** Magnitude of positive growth.
- **Red Bars (Left):** Magnitude of contraction.

14.3 The Interaction Layer (Slicers)

A category slicer (e.g., Food, Automotive, Textile) was implemented to enable interactive interrogation of the data.

- **Function:** The slicer filters all visuals simultaneously.
- **Use Case:** Selecting "Automobiles" instantly updates KPI cards, bar charts, and the matrix to display only automotive performance.

15. Final Executive Briefing: The "Auto Illusion"

By moving beyond headline figures and implementing a weighted analytical framework, a critical structural divergence in Pakistan's Large Scale Manufacturing sector for July–November FY 2025–26 was uncovered.

15.1 The Verdict

The dashboard reveals a K-shaped recovery that is currently masked by a statistical distortion referred to as the **Auto Illusion**.

1. **Headline Distortion:** Aggregate industrial performance appears positive due to outsized growth in Automobiles (Cars +81%, Trucks +116%). This reflects a base-effect rebound following import restrictions rather than broad-based industrial expansion.

2. **The Real Economy (The Drag):** While automobiles register green indicators, the risk matrix exposes contraction in industrial inputs:
   - Soda Ash (-28%) and Paper/Board (-31%) are in decline.
   - Steel billets (-10%) are contracting despite growth in Cement (+13%), indicating a disrupted construction supply chain—likely driven by energy costs forcing steel melters offline.

15.2 Conclusion

Through rigorous data cleaning, scalable Star Schema modeling, and weighted DAX measures, a fragmented administrative dataset was transformed into a decision-grade executive monitor. This dashboard does not merely report outcomes; it explains why they matter, equipping the Ministry to shift from generalized industrial policy toward targeted, sector-specific intervention.