



Procurement Transformation & Digitalization (Theme 3)

**Creating an Intelligent Cost Database
(Challenge 3.3)**

Team SAU Storms | Ashirwad Sinha, Sumit Pandey, Utkarsh Shukla
(spandey_mca24)

Problem Recap: Unstructured Procurement Data

Inconsistent Descriptions

Same items written in many styles

Slow Cost Estimation

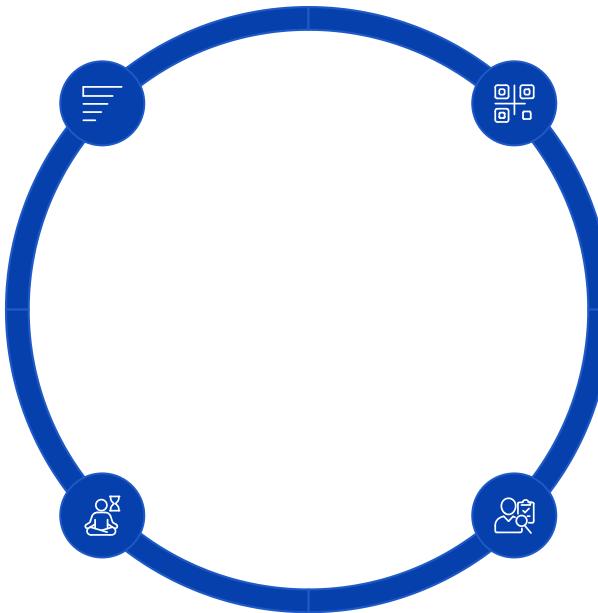
Manual processes hinder efficiency

No Unique Codes

Lack of standardized item identifiers

Elusive Past Prices

Difficulty finding accurate historical costs



A	B	C	D	E	F	G	H	I
po_id	item_description	unit_price	quantity	unit	po_date	region	department	supplier
1	CS Pipe 100mm dia	1200	10	pcs	01-01-2024	North	Maintenance	ABC Metals
2	Carbon Steel Pipe Dia 100 mm	1180	8	pcs	01-01-2024	North	Maintenance	ABC Metals
3	CS PIPE 4 inch diameter	1250	6	pcs	02-01-2024	North	Maintenance	XYZ Traders
4	Carbon steel pipe 100 mm OD	1210	9	pcs	02-01-2024	East	Maintenance	XYZ Traders
5	CS pipe dia 100mm	1195	7	pcs	03-01-2024	East	Maintenance	ABC Metals
6	Carbon Steel Pipe 4 in	49600	5	pcs	03-01-2024	West	Maintenance	SteelHub
7	CS PIPE 100 MM	1220	12	pcs	04-01-2024	South	Maintenance	SteelHub
8	Carbon steel pipeline 100mm	1230	11	pcs	04-01-2024	South	Maintenance	ABC Metals
9	CS Pipe OD 100 mm	1205	10	pcs	05-01-2024	North	Maintenance	XYZ Traders
10	Carbon steel pipe 4 inch dia	1260	6	pcs	05-01-2024	North	Maintenance	XYZ Traders
11	SS Pipe 50mm dia	2100	5	pcs	06-01-2024	South	Production	SteelCorp
12	Stainless Steel Pipe 50 mm	2080	6	pcs	06-01-2024	South	Production	SteelCorp
13	SS PIPE DIA 50MM	2120	4	pcs	07-01-2024	South	Production	MetalWorks
14	Stainless steel pipe 2 inch	2150	5	pcs	07-01-2024	West	Production	MetalWorks
15	SS pipe 50 mm OD	10475	7	pcs	08-01-2024	South	Production	SteelCorp

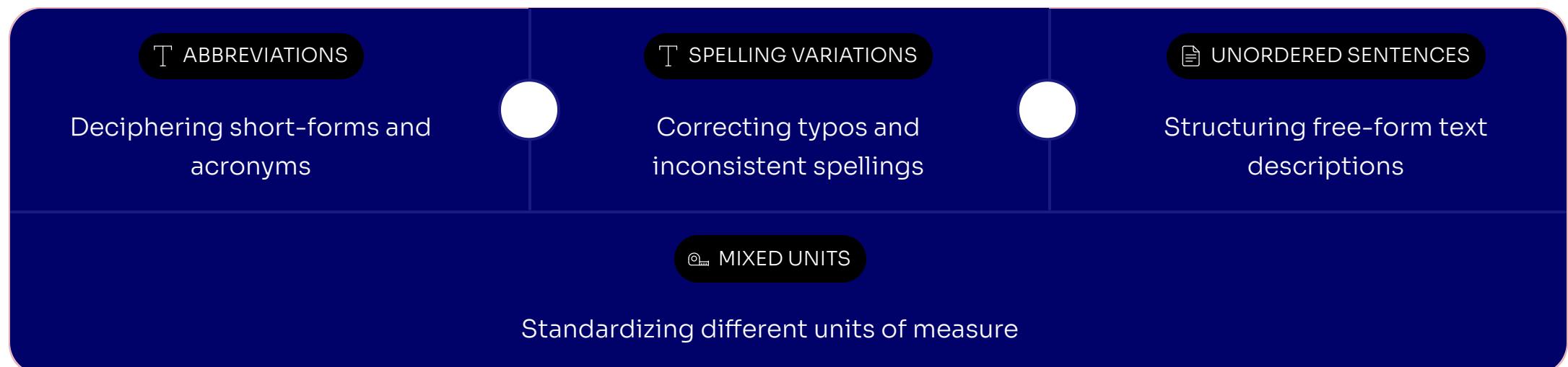
raw data: poorly managed, no outlier detection, no visualization and statistics.

Dataset and Real Inputs

To tackle the challenges of inconsistent procurement data, we began by analyzing raw Purchase Order (PO) CSV files, which served as our primary dataset. These files provided a rich source of real-world procurement scenarios.

Addressing Data Inconsistencies

A crucial step involved meticulously handling various data issues present in the raw files to ensure accuracy and standardization:



Key Dataset Statistics

The scale of the data transformation effort is highlighted by these metrics:

5K+

PO Rows

Total purchase order entries

1.2K+

Raw Descriptions

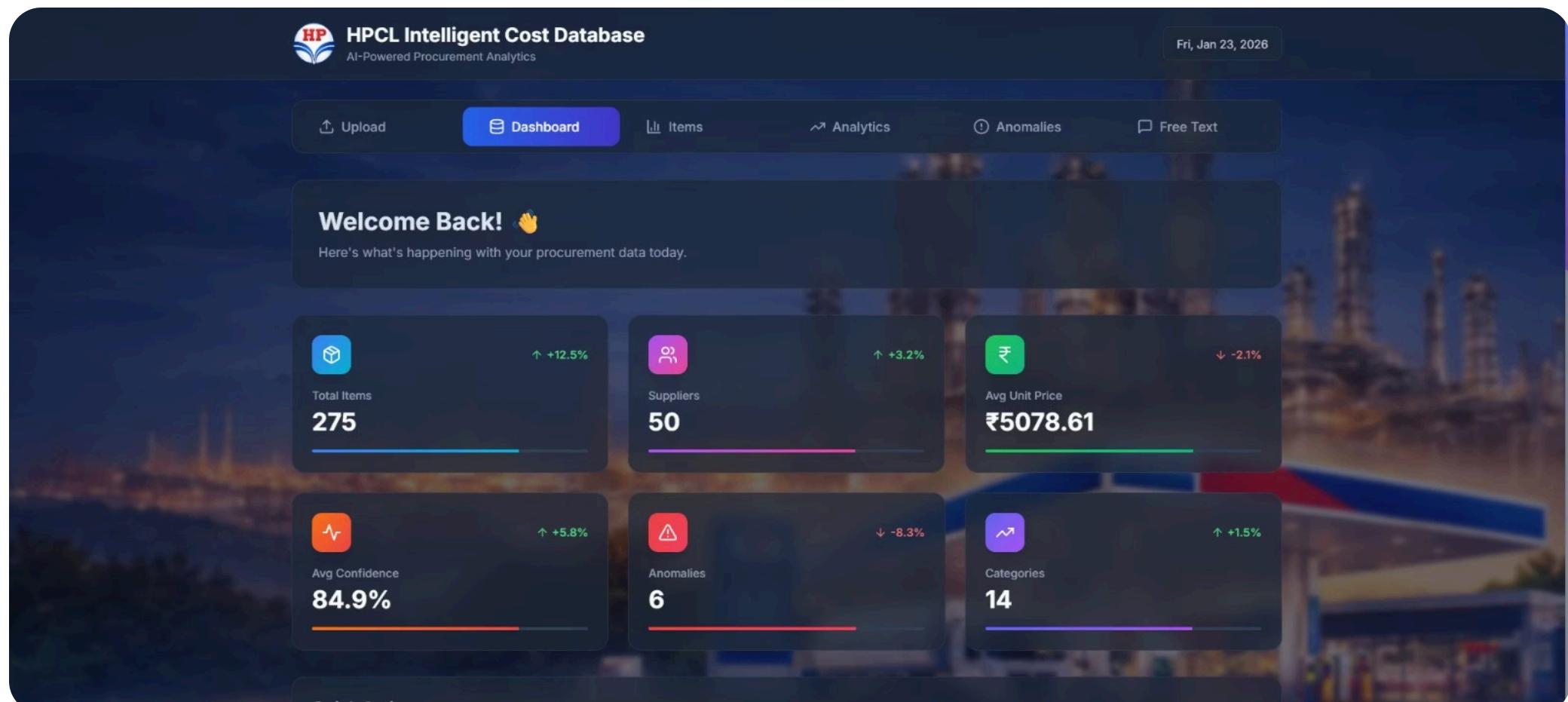
Unique item descriptions before standardization

300-400

Unique Items

Actual unique items after standardization

From Chaos to Clarity: Data Transformation Clean Data Example



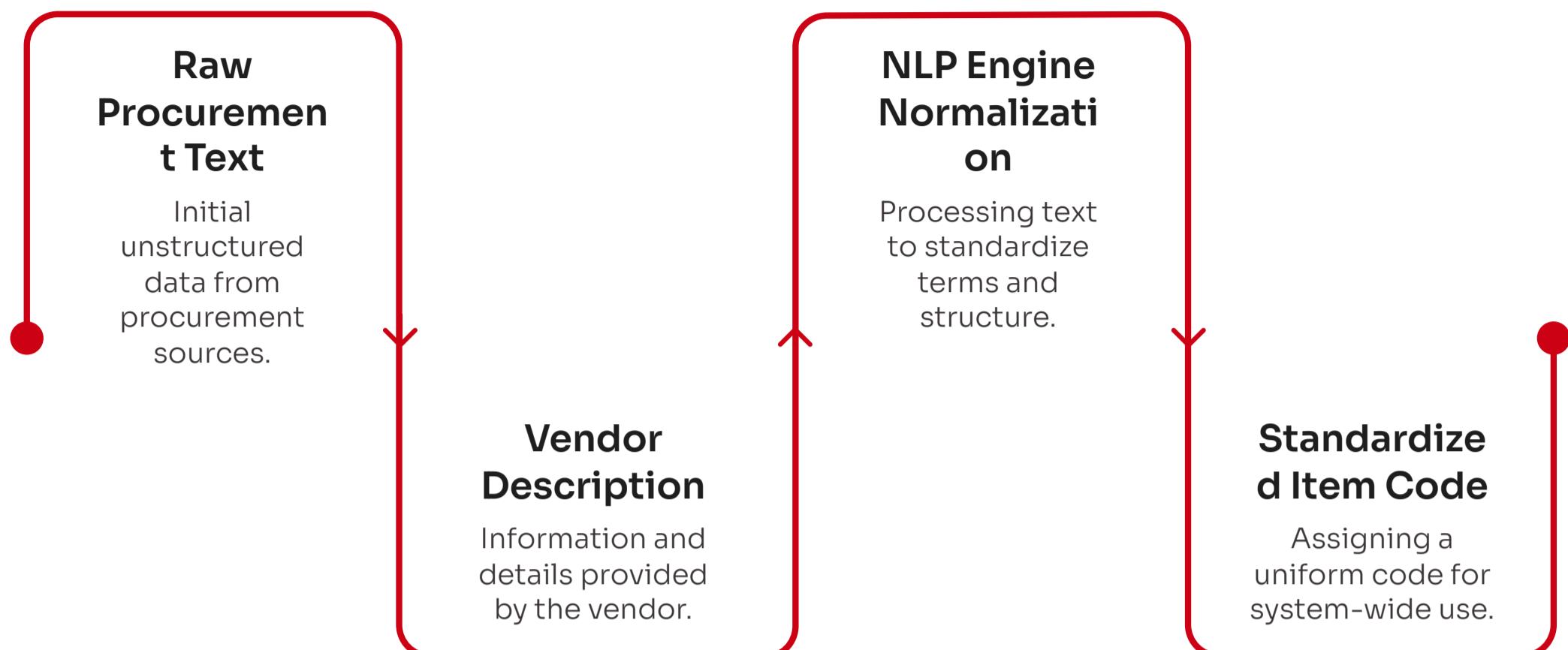
What We Built: System Pipeline



NLP Item Standardization Engine

Turning Text into Structure:

This transformation process is crucial for creating an intelligent cost database.



The screenshot shows the HPCL Intelligent Cost Database interface. At the top, there's a navigation bar with 'Upload', 'Dashboard', 'Items', 'Analytics', 'Anomalies', and a highlighted 'Free Text' button. Below the navigation is a section titled 'Free Text Analysis' with a sub-section 'Try an Example'. It displays three examples: 'Procurement Email', 'Purchase Note', and 'Supplier Quote', each with a snippet of text and a 'View Details' button. Below this is a large input field labeled 'Enter Your Text' containing a sample procurement email. To the right of the input field is a pink 'Analyze Text' button.

The screenshot shows the results of the 'Free Text Analysis' for the entered procurement email. The 'Extracted Insights' section lists six items detected: 1. Items detected: Oil, Pipe, Valve; 2. Materials: Stainless Steel; 3. Sizes: inch, mm; 4. Oil grade: iso 68; 5. Price range mentioned: 1200 to 5000; 6. Usage context: Maintenance, Operations. Below this is a 'What This Feature Does' section with 'Detects' (Items (pipes, valves, filters, etc.) and Materials (carbon steel, stainless steel)) and 'Use Cases' (Analyze procurement emails and Extract data from supplier quotes). The background of the interface features a blurred image of an industrial facility at night.

Cost Analytics Built on Clean Data

With clean, standardized procurement data in place, our system moves beyond mere record-keeping to deliver intelligent cost analytics. For each standardized item, we calculate comprehensive insights, transforming raw numbers into actionable intelligence.



Average Price

Overall mean cost for an item



Median Price

Middle value in price distribution



Min-Max Range

Highest and lowest recorded prices



Time-Based Trend

Price evolution over time



Region-wise Comparison

Prices across different geographies



Supplier-wise Comparison

Pricing from various vendors



Automatic Detection

System automatically identifies abnormal prices that deviate from expected patterns.



Out-of-Range Alerts

Highlights entries that fall outside predefined or learned price ranges for immediate attention.

The screenshot shows the 'Price Anomalies' section of the dashboard. It displays two items flagged as anomalies:

- CARBON STEEL PIPE 100MM DIAMETER 5f0c**: PO: P0006, Pipe. Actual Price: ₹49600.00, Expected Price: ₹1210.00, Deviation: 3999.2%. Status: High.
- CARBON STEEL PIPE 100MM DIAMETER 5f0c**: PO: P0011, Pipe. Actual Price: ₹2100.00, Expected Price: ₹1230.00, Deviation: 70.7%. Status: Medium.

The interface includes navigation tabs like Upload, Dashboard, Items, Analytics, Anomalies (highlighted in blue), and Free Text, along with a date filter for Friday, Jan 23, 2026.

The screenshot shows the 'Price Range Analysis' and 'Items by Category' sections of the dashboard.

Price Range Analysis: A line chart showing price trends for various items. A callout highlights the 'ball bearing 6205' item with the following details:

- Max Price: ₹650.00
- Avg Price: ₹650.00
- Min Price: ₹650.00

Items by Category: A summary of procurement categories and their counts:

Category	Count
Other	97
Pipe	62
Valve	22
Filter	16
Flange	13
Fasteners	12

Dashboard: From Data to Decisions

The interactive dashboard is the final, user-facing component of our solution, translating complex backend processes into intuitive, actionable insights for procurement teams.

Upload Raw PO File

Seamlessly ingest procurement data directly into the system.

Free Text Search using NLP

Effortlessly find items using natural language queries.

Identify Pricing Anomalies

Receive automatic alerts for unusual and out-of-range prices.

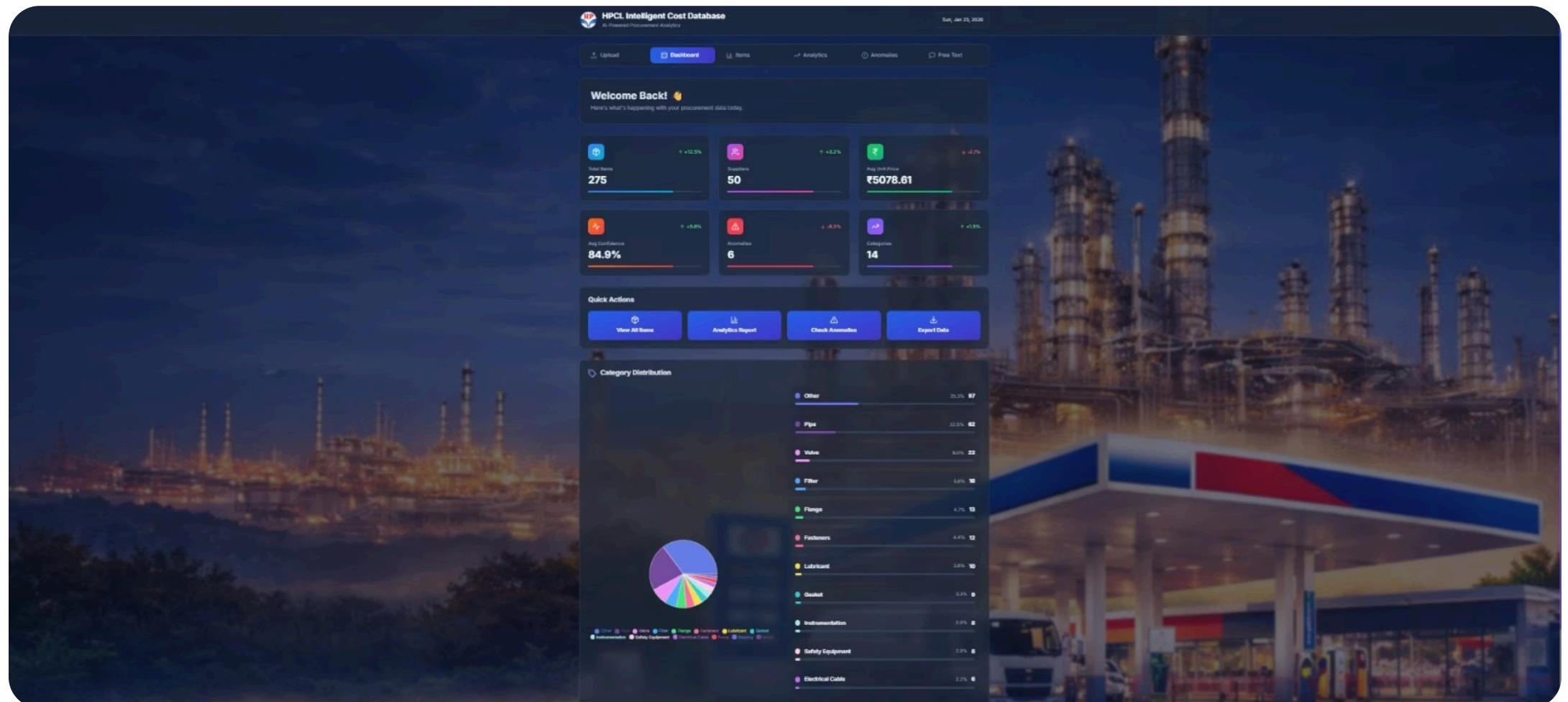
View Standardized Output

Access clean, categorized item data transformed by the NLP engine.

Explore Charts and Insights

Visualize cost trends, supplier comparisons, and regional data.

This dashboard brings all the backend work to life, empowering users with the clarity and insights needed for informed decision-making and optimized procurement strategies.



Impact for HPCL and What Comes Next



Faster cost estimation

Quickly access accurate cost data for better planning.



Reduced manual effort

Automate data cleaning and analysis, freeing up resources.



Clear visibility across regions

Gain insights into pricing trends and anomalies globally.

Scalability & Future Readiness

Can integrate with ERP systems

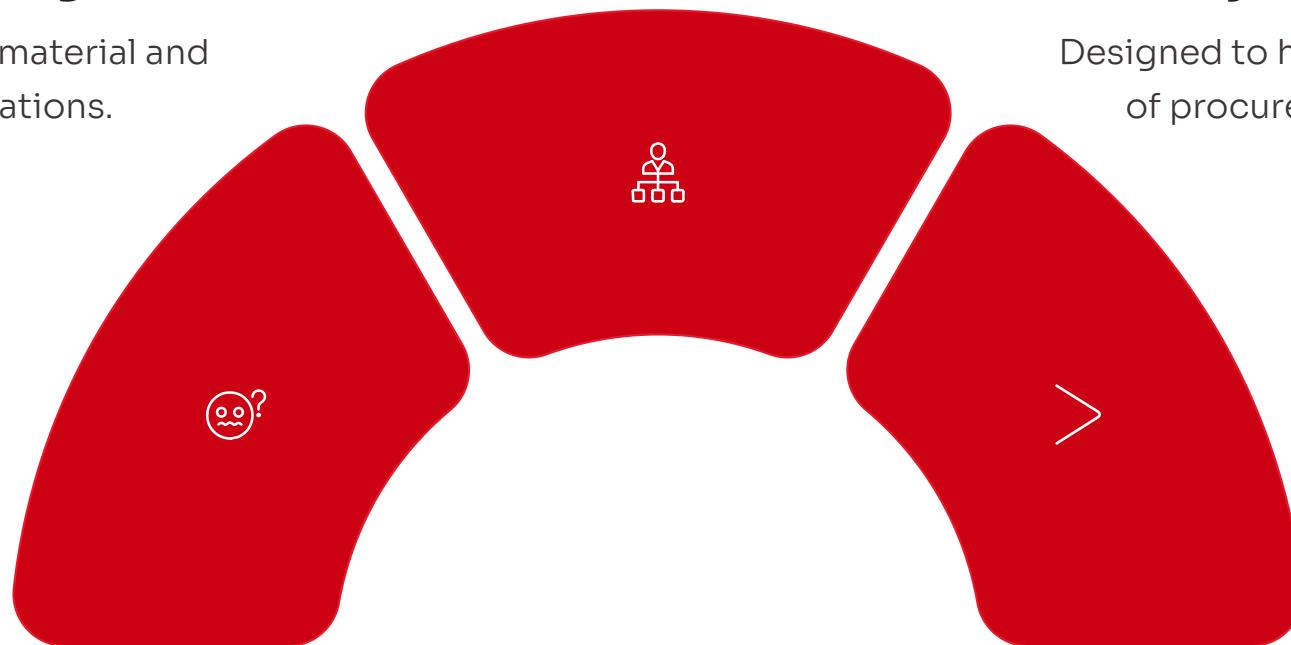
Seamless data flow with existing enterprise resource planning solutions.

Works across all procurement categories

Adaptable to diverse material and service classifications.

Ready for larger datasets

Designed to handle increasing volumes of procurement data efficiently.



From raw purchase orders to actionable cost intelligence — live.