Procurement Transformation & Digitalization (Theme 3)

Creating an Intelligent Cost Database (Challenge 3.3)

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

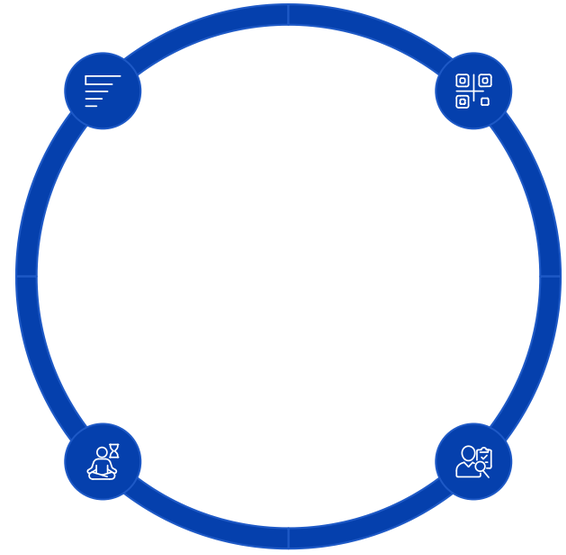
Problem Recap: Unstructured Procurement Data

Inconsistent Descriptions

Same items written in many styles

No Unique Codes

Lack of standardized item identifiers



Slow Cost Estimation

Manual processes hinder efficiency

Elusive Past Prices

Difficulty finding accurate historical costs

	A	B	C	D	E	F	G	H	I
1	po_id	item_description	unit_price	quantity	unit	po_date	region	department	supplier
2	PO001	CS Pipe 100mm dia	1200	10	pcs	01-01-2024	North	Maintenance	ABC Metals
3	PO002	Carbon Steel Pipe Dia 100 mm	1180	8	pcs	01-01-2024	North	Maintenance	ABC Metals
4	PO003	CS PIPE 4 inch diameter	1250	6	pcs	02-01-2024	North	Maintenance	XYZ Traders
5	PO004	Carbon steel pipe 100 mm OD	1210	9	pcs	02-01-2024	East	Maintenance	XYZ Traders
6	PO005	CS pipe dia 100mm	1195	7	pcs	03-01-2024	East	Maintenance	ABC Metals
7	PO006	Carbon Steel Pipe 4 in	49600	5	pcs	03-01-2024	West	Maintenance	SteelHub
8	PO007	CS PIPE 100 MM	1220	12	pcs	04-01-2024	South	Maintenance	SteelHub
9	PO008	Carbon steel pipeline 100mm	1230	11	pcs	04-01-2024	South	Maintenance	ABC Metals
10	PO009	CS Pipe OD 100 mm	1205	10	pcs	05-01-2024	North	Maintenance	XYZ Traders
11	PO010	Carbon steel pipe 4 inch dia	1260	6	pcs	05-01-2024	North	Maintenance	XYZ Traders
12	PO011	SS Pipe 50mm dia	2100	5	pcs	06-01-2024	South	Production	SteelCorp
13	PO012	Stainless Steel Pipe 50 mm	2080	6	pcs	06-01-2024	South	Production	SteelCorp
14	PO013	SS PIPE DIA 50MM	2120	4	pcs	07-01-2024	South	Production	MetalWorks
15	PO014	Stainless steel pipe 2 inch	2150	5	pcs	07-01-2024	West	Production	MetalWorks
16	PO015	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:

ABBREVIATIONS

Deciphering short-forms and acronyms

SPELLING VARIATIONS

Correcting typos and inconsistent spellings

UNORDERED SENTENCES

Structuring free-form text descriptions

MIXED UNITS

Standardizing different units of measure

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

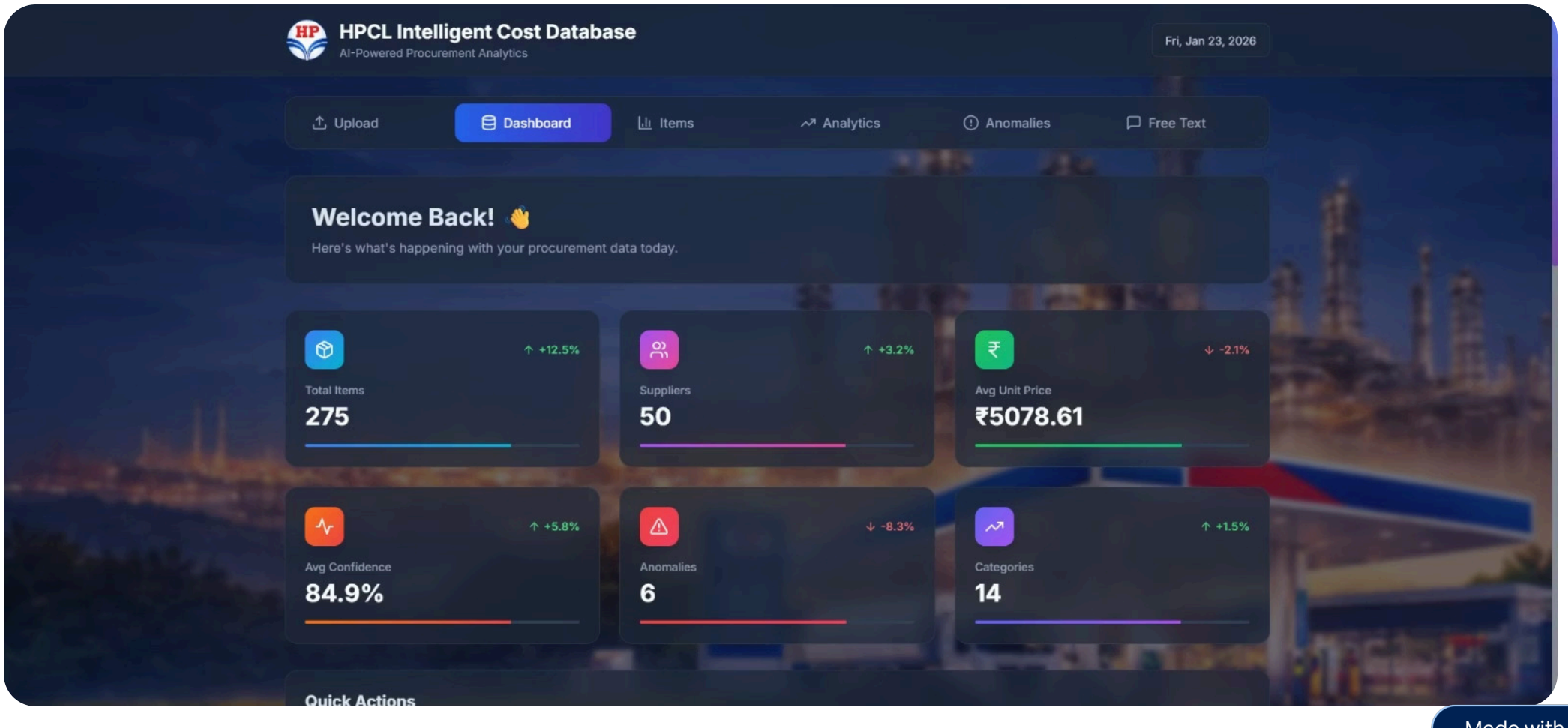
300-400

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.

Raw Procurement Text

Initial unstructured data from procurement sources.

Vendor Description

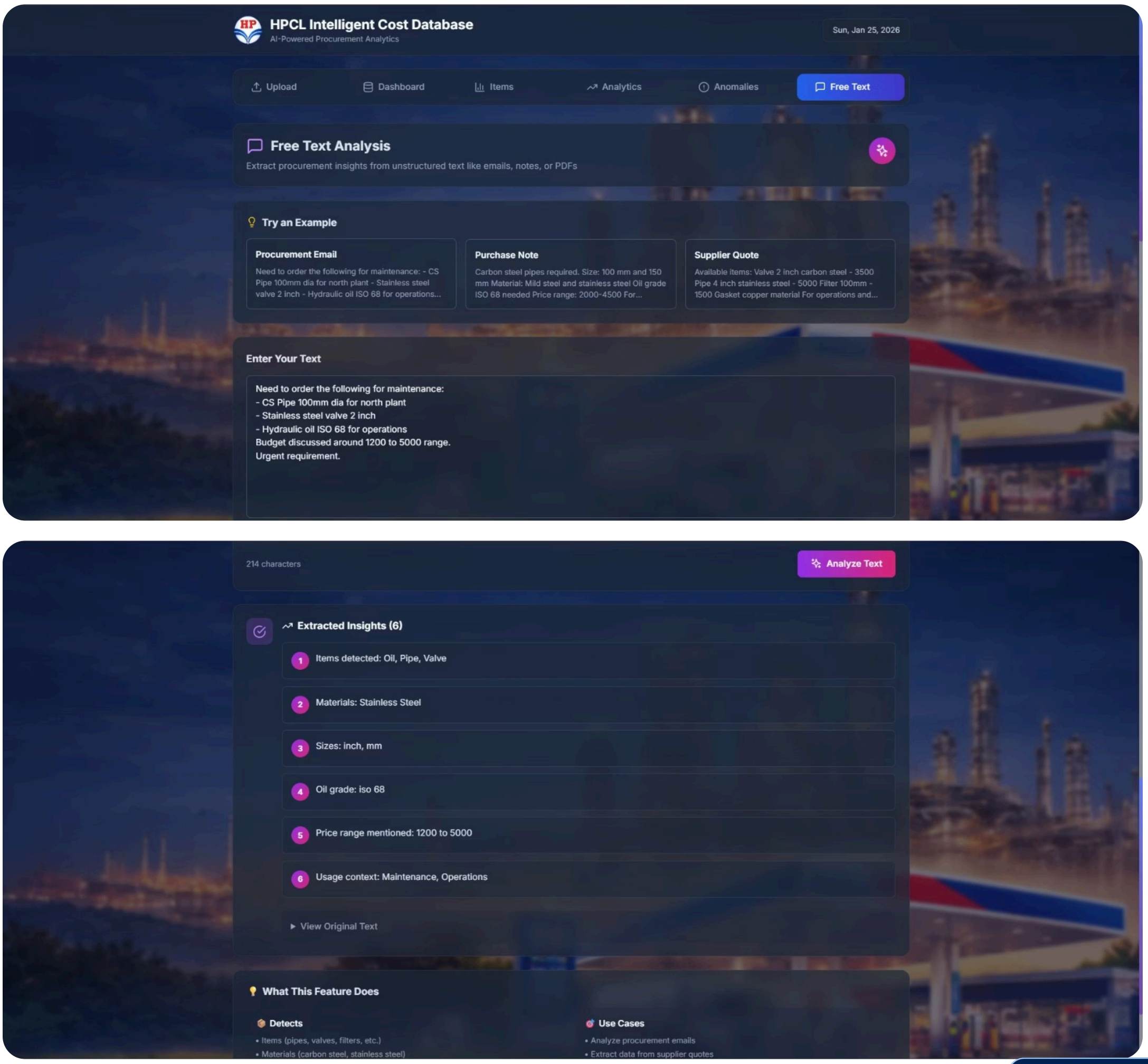
Information and details provided by the vendor.

NLP Engine Normalization

Processing text to standardize terms and structure.

Standardized Item Code

Assigning a uniform code for system-wide use.



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

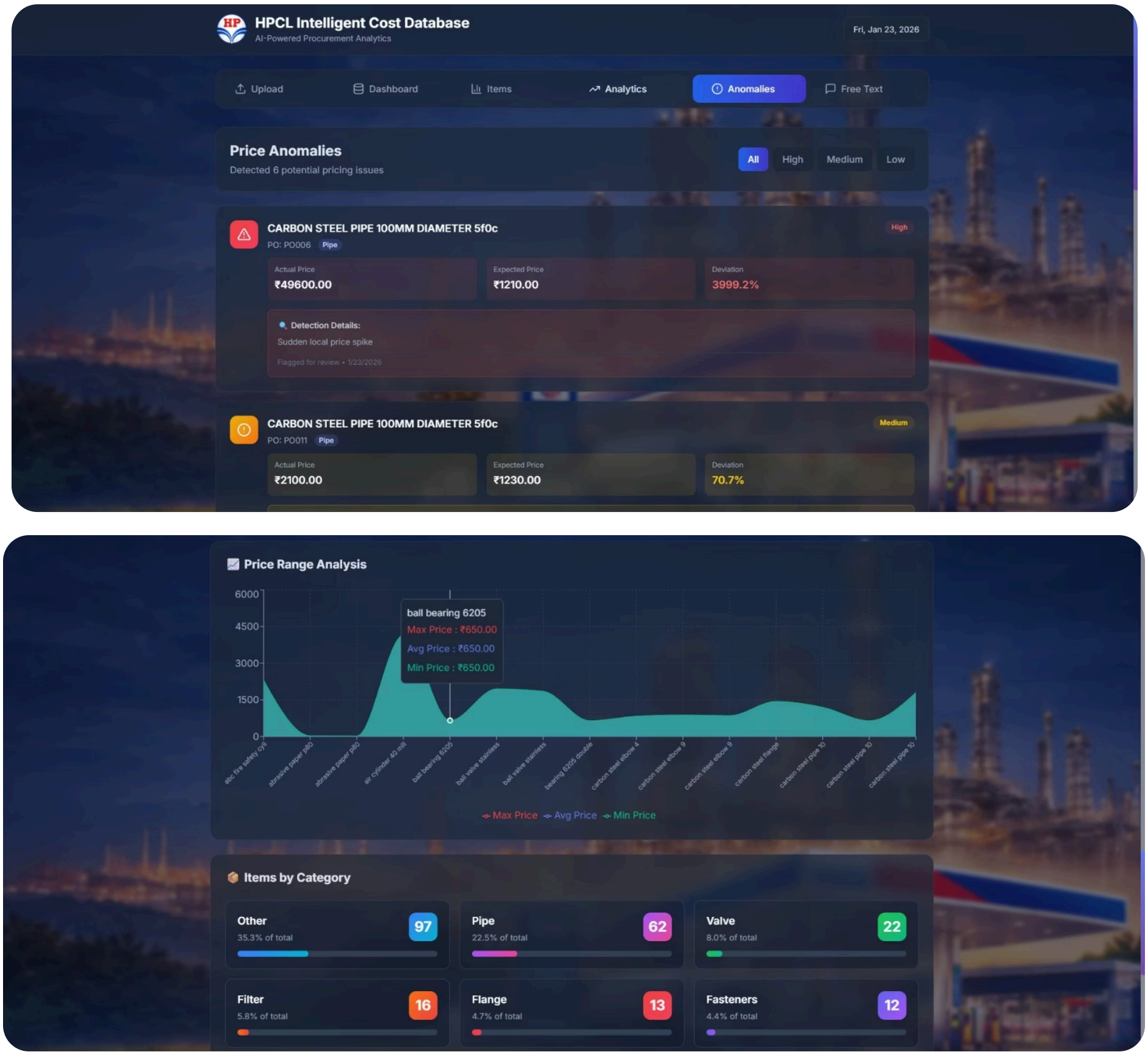
Pricing from various vendors

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.



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.

View Standardized Output

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

Free Text Search using NLP

Effortlessly find items using natural language queries.

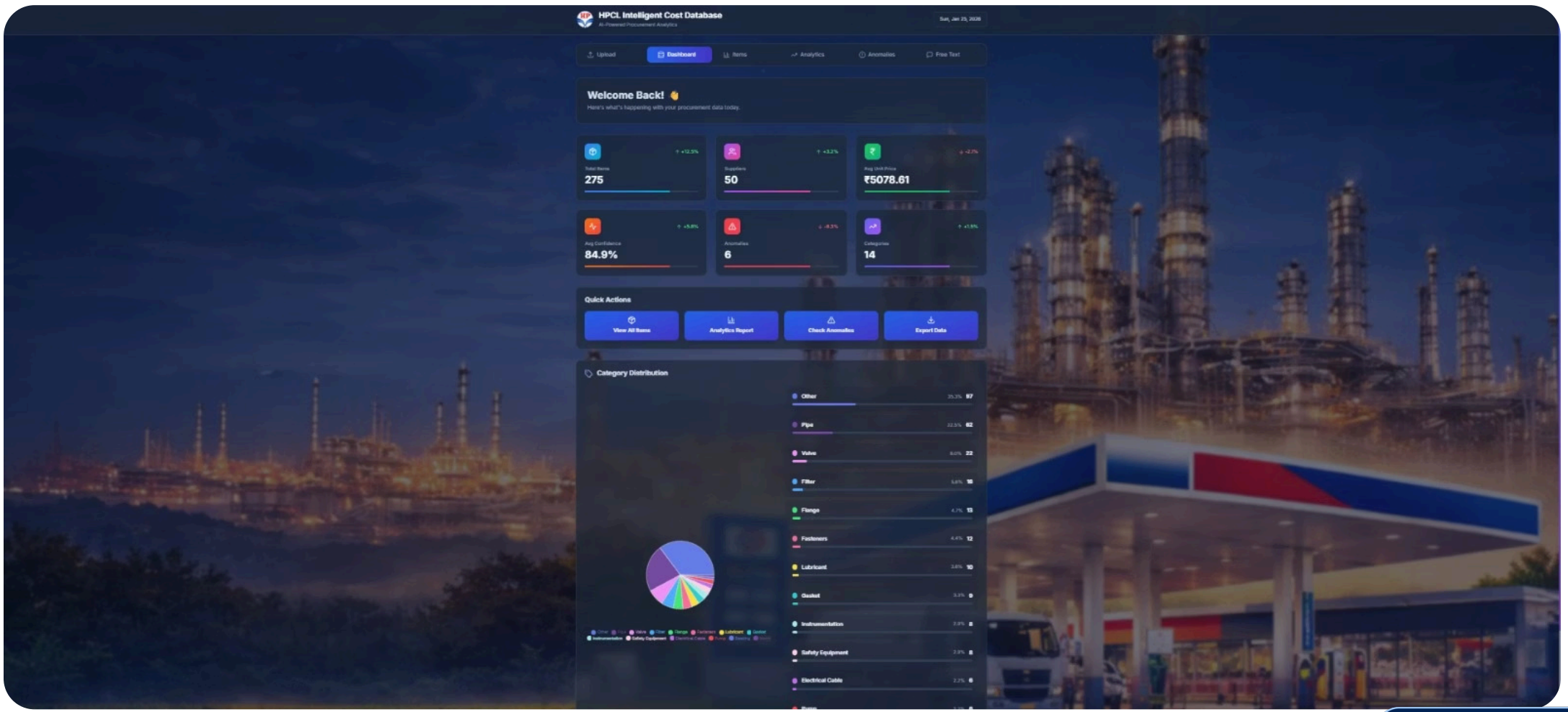
Explore Charts and Insights

Visualize cost trends, supplier comparisons, and regional data.

Identify Pricing Anomalies

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

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

Works across all procurement categories

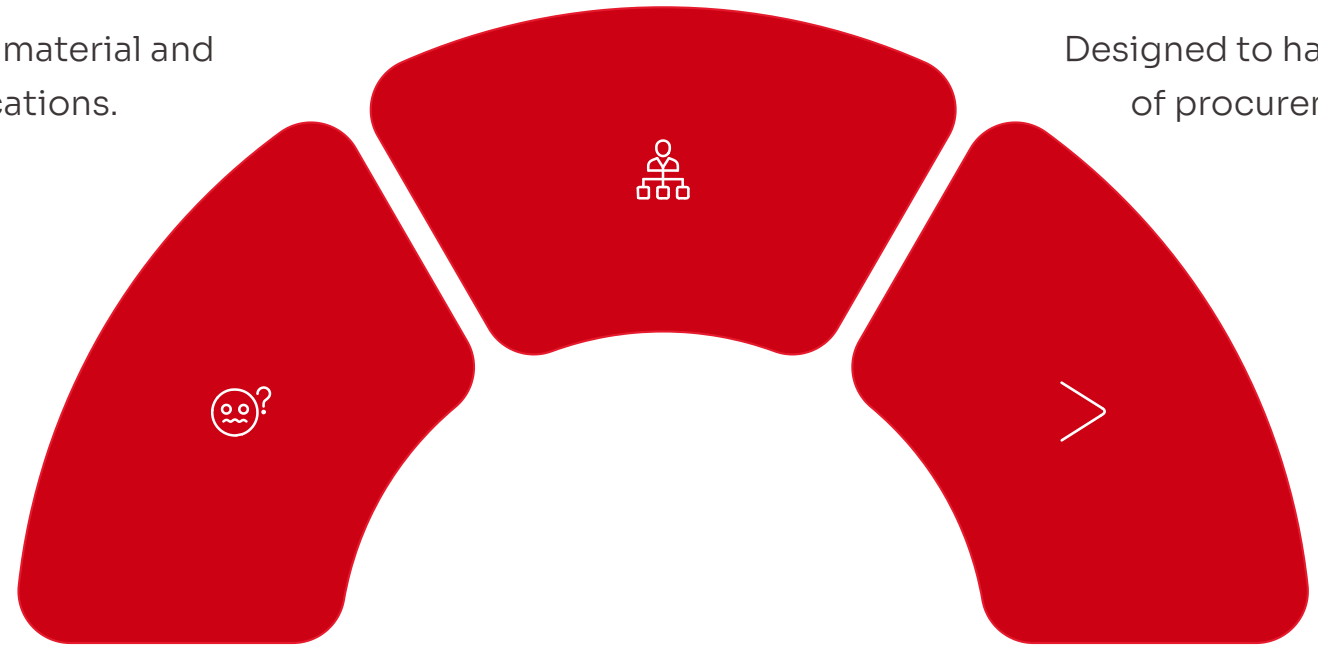
Adaptable to diverse material and service classifications.

Can integrate with ERP systems

Seamless data flow with existing enterprise resource planning solutions.

Ready for larger datasets

Designed to handle increasing volumes of procurement data efficiently.



From raw purchase orders to actionable cost intelligence — live.