Walmart Retail Sales Analysis Dashboard

Project Team: Adithyan R (RA2211027010040),Mohammed Zeeshan Samseer (RA2211027010049),Ishan Dey (RA2211027010069),Azam Tanzeem (RA2211027010081),Shreeyansh Verma (RA2211027010085)**Institution:** SRM Institute of Science & Technology**Course:** 21CSE421T – Business Intelligence and Analytics**Guide:** Ms. Gayatri M

Project Overview

Objective

Transform raw Walmart sales data into actionable business intelligence using Python and Power BI

Scope

Comprehensive analysis of 420,000+ sales records across multiple stores and departments

Data Source

Kaggle "Walmart Store Sales Forecasting" dataset with temporal and economic indicators

This project demonstrates the integration of Python-based data engineering with Power BI visualization to deliver strategic retail insights.

Motivation & Strategic Objectives

1 Real-World Challenge

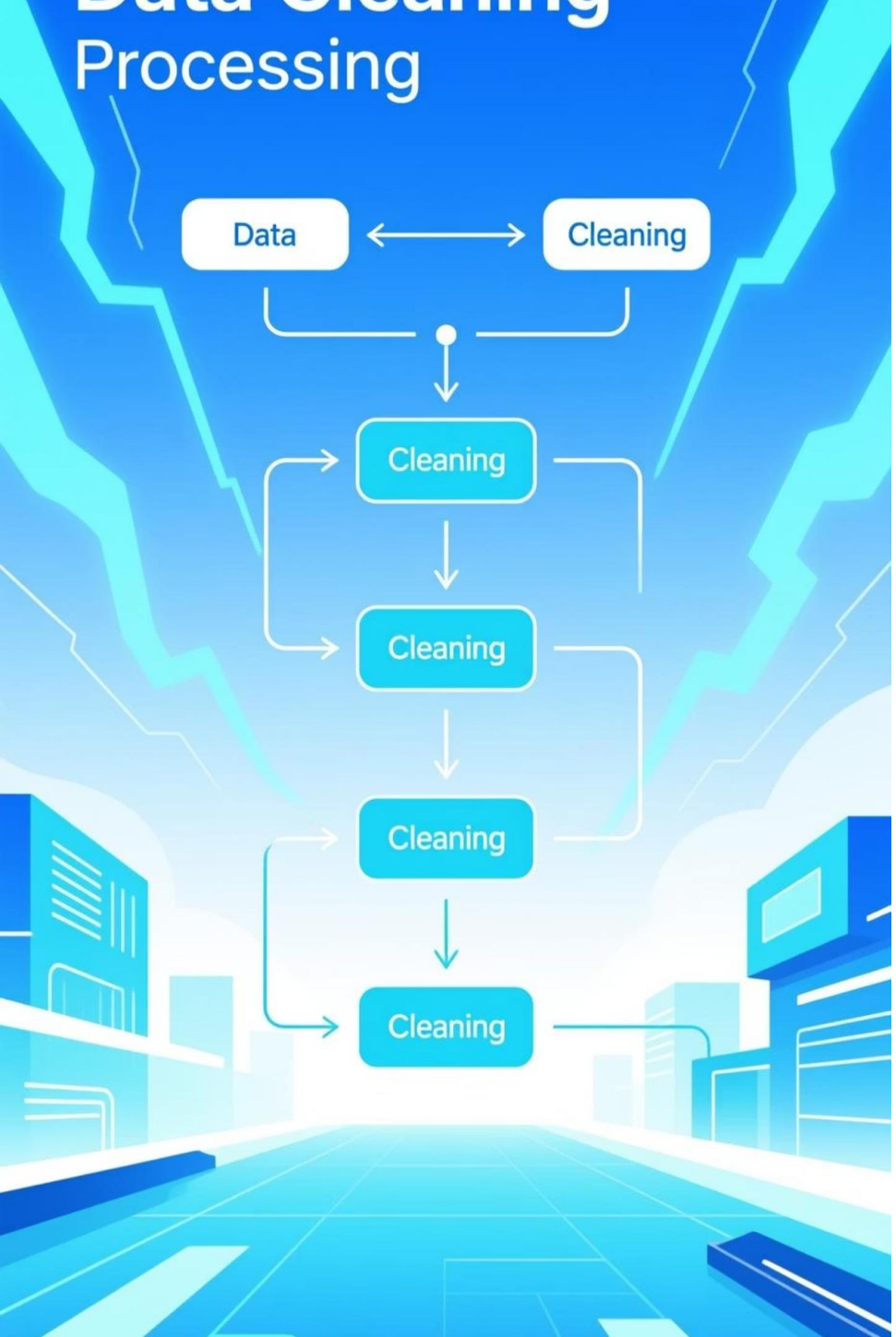
Modern retail demands rapid data-driven decisions to optimize inventory, staffing, and promotional strategies across geographically dispersed store networks.

2 Analytical Goals

Identify seasonal patterns, store performance variances, economic correlations, and actionable trends to enhance operational efficiency and revenue forecasting.

3 Technical Objectives

Master ETL processes, feature engineering, exploratory analysis, and interactive dashboard development—critical competencies for business intelligence professionals.



Problem Statement & Key Challenges

Core Problem

Raw Walmart sales data lacks consistency and structure. Identifying performance drivers—seasonal effects, store disparities, and economic correlations—requires rigorous data cleaning and statistical analysis.

Business Impact: Without clarity, retailers cannot optimize inventory allocation or forecast accurately.

Technical Challenges

- Data quality inconsistencies and missing values
- Seasonal and holiday effects requiring temporal feature engineering
- Multicollinearity among economic indicators (CPI, unemployment, fuel price)
- Integrating disparate data sources and platforms seamlessly

Dataset Architecture

The Kaggle Walmart Store Sales Forecasting dataset encompasses comprehensive retail metrics across multiple dimensions:

Dimension	Attribute	Type	Description
Store	Store ID	Categorical	Store identifier (1–45)
Temporal	Date, Week, Month, Year	DateTime	Weekly sales observation period
Sales	Weekly_Sales, Dept	Numeric	Department-level weekly revenue
Economic	CPI, Unemployment, Fuel_Price	Numeric	Macroeconomic indicators
Event	IsHoliday, Temperature	Boolean/Numeric	Holiday flags and seasonal factors

Dataset Size: ~420,000 weekly sales records | **Time Period:** 2010–2014 | **Stores:** 45 locations | **Departments:** 81 unique product categories

Data Preparation & Python Engineering

01

Data Import & Validation

Load CSV files from Kaggle, validate schema, inspect data types, and assess missing value patterns using pandas and NumPy.

02

Cleaning & Imputation

Handle missing values through statistical imputation; standardize date formats; remove erroneous records and outliers using Z-score and IQR methods.

03

Feature Engineering

Extract temporal features (Day of Week, Month, Quarter, Year); create binary indicators (HolidayWeek, IsWeekend); derive rolling averages and lag features.

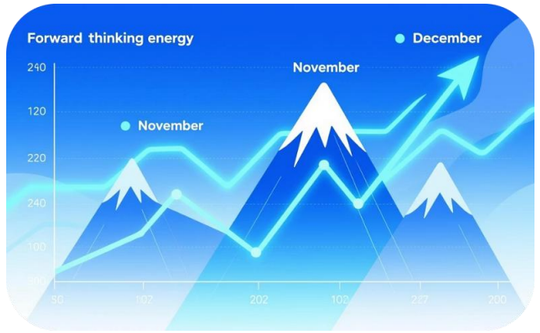
04

Aggregation & Export

Aggregate sales by store, department, and week; normalize economic indicators; export cleaned datasets as CSV and Excel for Power BI consumption.

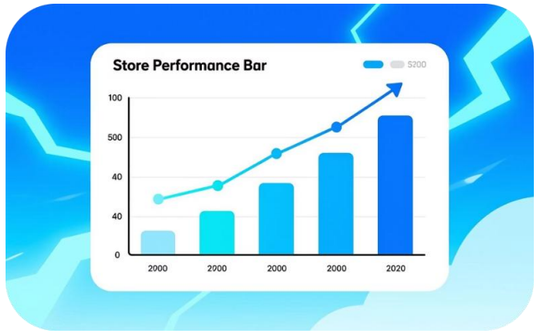
Exploratory Data Analysis: Key Findings

Analysis of cleaned data reveals critical retail dynamics and performance patterns:



Seasonal Patterns

Holiday periods (Thanksgiving, Christmas, Black Friday) drive 35–45% sales uplift. Clear seasonal cycles align with consumer spending behavior.

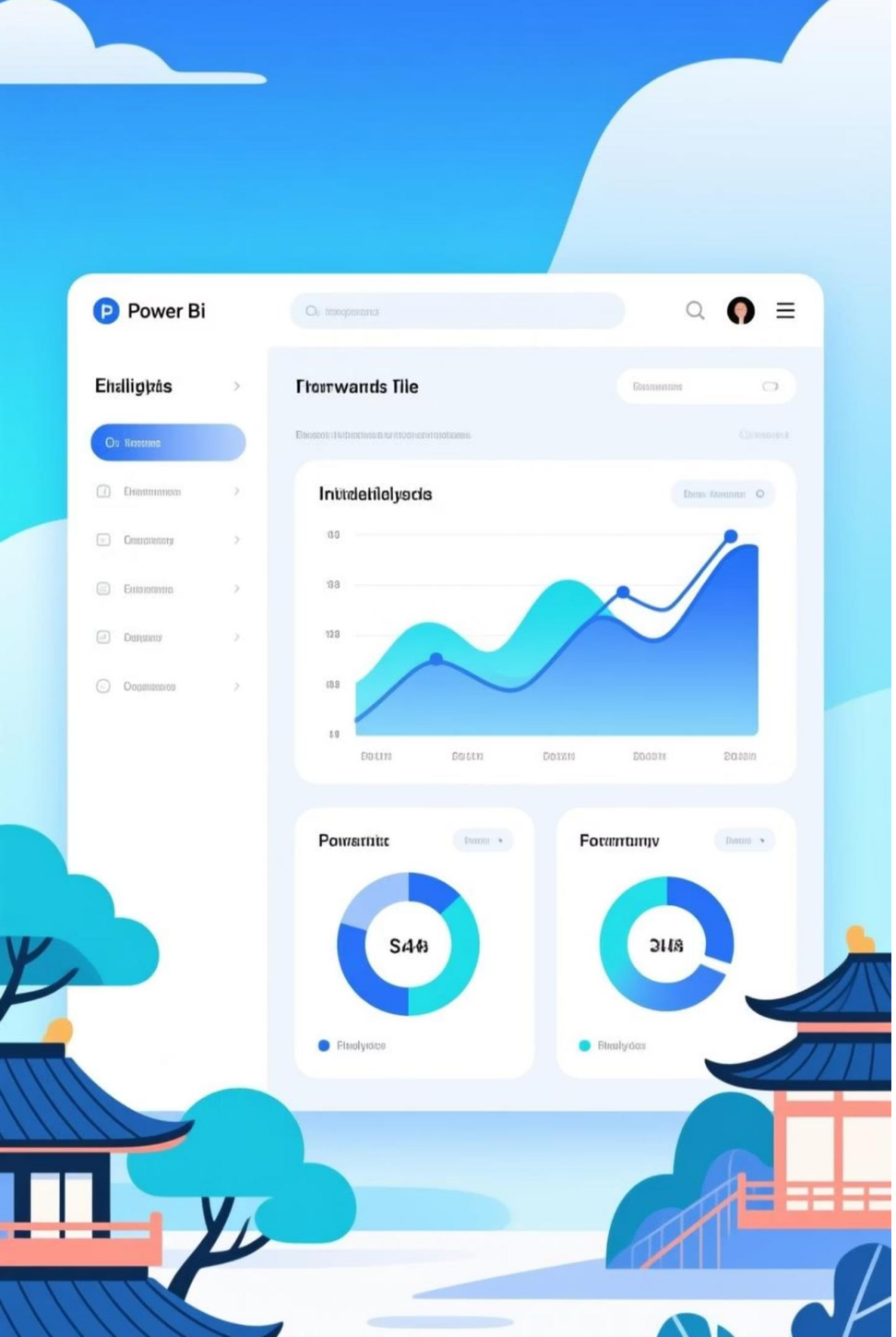


Store Disparities

Store 20 ranks highest in sales volume, while geographic and demographic factors create 3–4x revenue variance across the 45-store network.

Economic Correlations

Sales exhibit strong negative correlation with CPI (–0.31) and unemployment rates (–0.21); fuel prices show weak positive influence on consumer activity.



Results & Business Intelligence Insights

Holiday Impact

35–45% average sales increase during holiday weeks, with strongest performance in Q4. Strategic promotions yield exponential returns during peak periods.

Store Optimization

High-performing stores (Top 10) require distinct inventory and staffing strategies. Geographic clustering reveals regional demand patterns enabling targeted resource allocation.

Economic Sensitivity

Sales inversely track unemployment and inflation metrics, enabling predictive modeling using macroeconomic forecasts. Recession periods correlate with 15–20% sales dips.

Dashboard Intelligence

Interactive Power BI dashboards enable real-time KPI monitoring, drill-down analysis by store/department/period, and automated alerts for anomalies and opportunities.

Conclusion & Future Enhancements

Key Takeaways

- Python + Power BI integration streamlines retail analytics
- Seasonal insights drive targeted promotions and inventory planning
- Economic indicators enable predictive forecasting
- Interactive dashboards democratize data-driven decision-making

Future Roadmap

- **ML Forecasting:** ARIMA and LSTM models for demand prediction
- **Real-Time Data:** Streaming pipelines for live sales ingestion
- **Geospatial:** Store location mapping and neighborhood analytics
- **Cloud Deployment:** Azure/AWS for scalable BI infrastructure

References & Acknowledgments

Data & Tools

Kaggle: Walmart Store Sales Forecasting dataset**Python Libraries:** Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn**Visualization:** Microsoft Power BI Pro

Documentation

Power BI Docs: Official Microsoft documentation**OpenAI:** Data analysis and interpretation support**GitHub:** Project repository and code

Institutional Support

SRM Institute of Science & TechnologyGuide: Ms. Gayatri M**Course:** 21CSE421T – Business Intelligence and Analytics

Thank you for exploring Walmart's retail analytics landscape. May these insights drive strategic excellence in data-driven retail management.