Smartphone Procurement Analytics Workflow

Author: ISTIAK ALAM

Affiliation: DataCamp Real World Project

Executive Summary

This project demonstrates an end-to-end data science workflow supporting procurement decision-making for a global university. Built for a real business scenario, this analysis prepares and visualizes a large, varied dataset of smartphones, transforming raw data into actionable, manager-ready insights. Key contributions include robust data cleaning, feature engineering, unit-tested pipeline design, and interactive visualization functions—all essential skills for an analytics or data science role in industry.

Table of Contents

Project Context

To optimize mobile procurement for 1,000+ employees, a global university's procurement team provided a CSV containing extensive smartphone model specs, pricing, ratings, and technical features. As the lead data scientist, I quality-checked and improved a junior colleague's data workflow, ensuring it was robust, reusable, and ready for production.

Approach & Methodology

- **Data Preparation & Cleaning:**
  Developed a function to systematically load and clean raw CSV data, select high-value features (`brand`, `os`, `price`, `processor_speed`, `screen_size`, etc.), and remove missing or corrupt values. Ensured price normalization for clear $ comparison.

- **Reusable Labeling & Visualization:**
  Refactored label generation (via the `column_to_label` function) for DRY, readable plots. Designed a single, flexible plotting function allowing procurement or business users to instantly explore cost-performance tradeoffs (e.g., price versus processor speed or battery).

- **Testing & Validation:**
  Upgraded and passed all unit tests for each transformation stage, reflecting enterprise-level pipeline reliability and test-driven development practices.

Key Results & Visualizations

Cleaned, Analysis-Ready Dataset

After pipeline processing:

- **Data coverage:** Major global brands, multiple OS types, technical specs, and $ prices—all standardized for fair comparison.

- **Missing data removed:** Ensured all visual and numeric analyses are reliable.

Example: Processor Speed vs. Price

(Sample Visualization Output — as seen in notebook.ipynb)

- **Plot:** Scatterplot (color-coded by OS) showing how price relates to mobile processor speed.

- **Value:** Quickly exposes best-value models, outlier pricing, and operating system clusters for direct procurement decision support.

(Imagine image: "Processor Speed vs. Price.png" here)

Impact & Recommendations

- **Business Impact:**
  The workflow enables the university to select smartphones that maximize performance for budget, ensure device diversity, and align with procurement best practices.

- **Technical Excellence:**
  Demonstrates Python (pandas, seaborn, matplotlib), robust feature engineering, code refactoring (DRY), functional and tested code, and clear business communication.

- **Scalable Foundation:**
  Functions are modular—ready to be extended for additional features (e.g., RAM, camera megapixels) or integrated into web dashboards for ongoing procurement analysis.

Files & Portfolio Links

- **Project Notebook:** notebook.ipynb

- **Example Visualization:** *(insert e.g. "Processor Speed vs. Price.png" if included)*

- **Raw Data:** smartphones.csv

- **Live DataCamp Project:** View in DataCamp

- **Portfolio:** istiak-data-analyst

Conclusion

This project proves hands-on ability to design, refactor, test, and communicate data science workflows that drive short-term business value (procurement optimization) and long-term organizational learning.
Recruiters and managers: Review the notebook and code to see modern Python practices, pragmatic business reasoning, and production-ready analytics in action.

For recruiters:
I am ready to deliver the same impactful, reliable, production-level analytics for your business or client team.