Machine Learning Product Performance Analysis - REPORT

Introduction

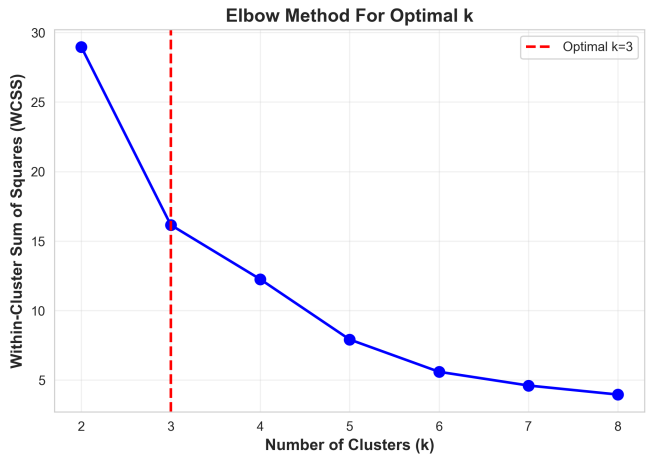
This report summarizes data preprocessing, K-means clustering, and regression analysis performed on the provided product

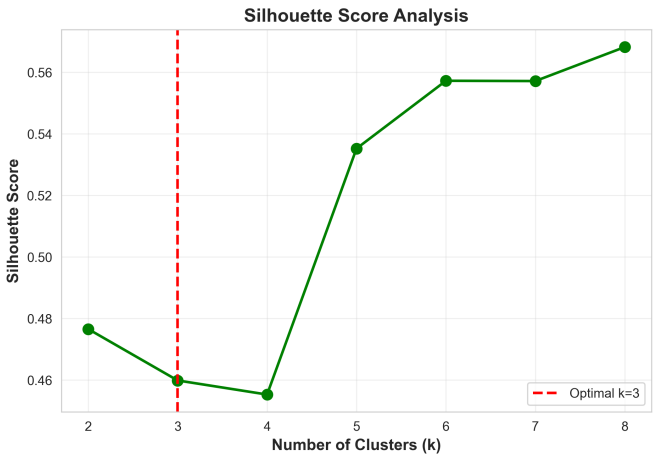
Data Preprocessing

Missing values: handled via median imputation for numeric columns; missing product_name rows dropped.
Outliers: detected via IQR and capped to lower/upper bounds.
Normalization: Min-Max scaling applied to numeric features (0-1).
Final dataset shape: (196, 9)

K-means Clustering

Elbow analysis recommended k = 3. Silhouette score at optimal k: 0.460.
WCSS values: 29.0, 16.2, 12.3, 7.9, 5.6, 4.6, 4.0



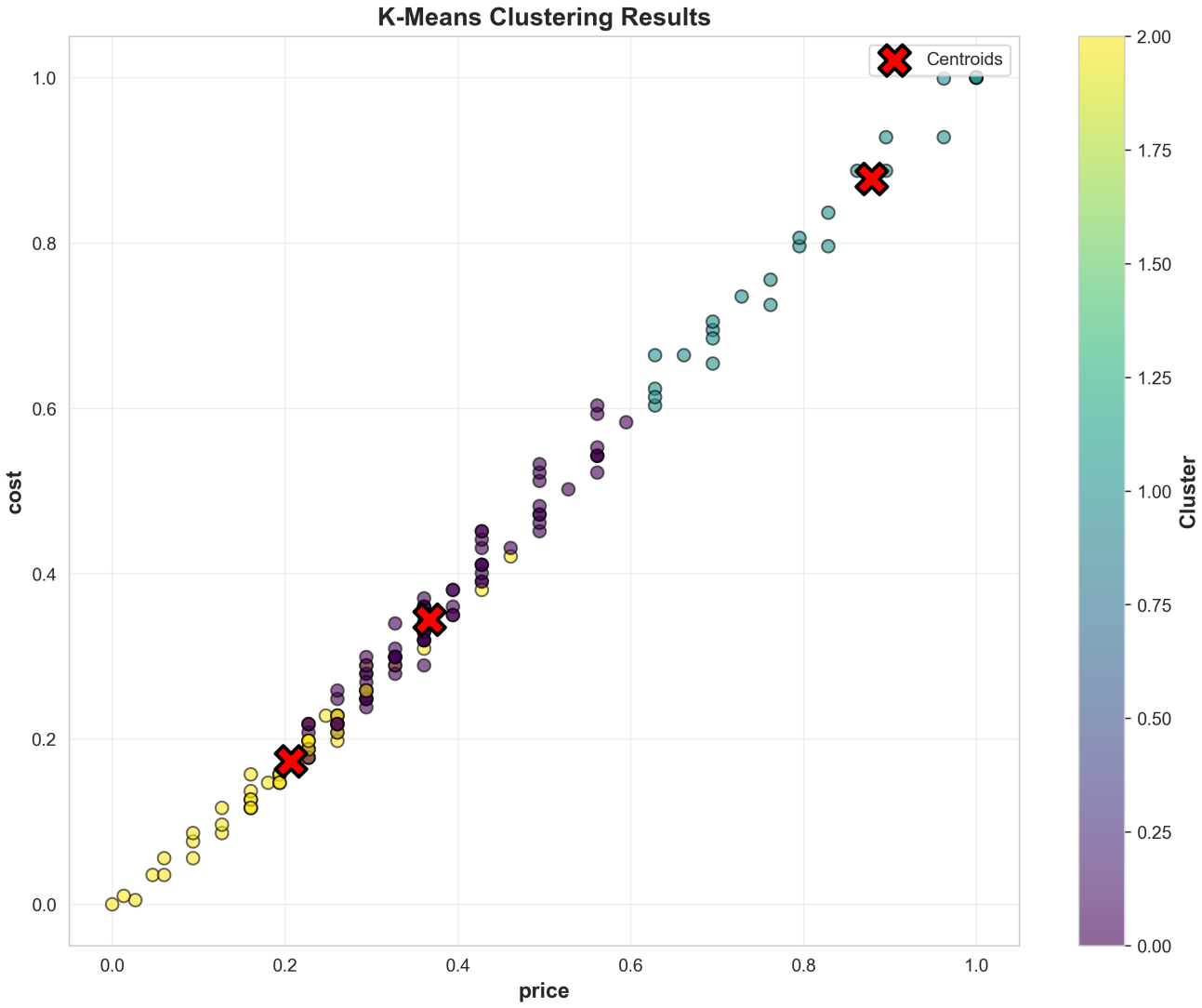


Cluster Summaries & Suggested Names

Cluster 0 (size=92): Suggested name: Mid-Range Steady
  Avg Price: $6.08 Avg Units Sold: 226.7 Avg Profit: $518.24
  Insight: Stable mid-range products; monitor for seasonal shifts.

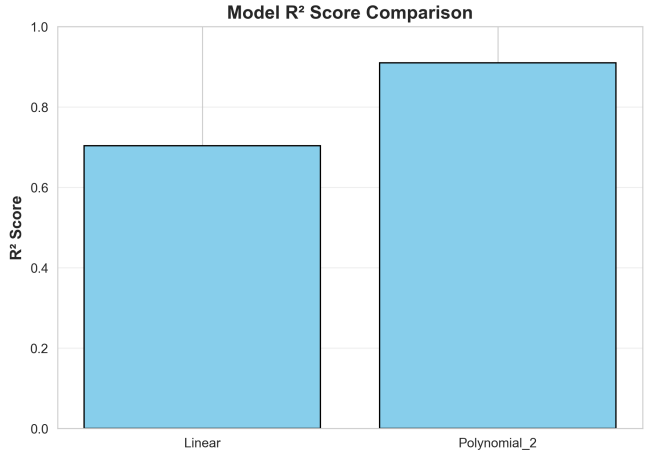
Cluster 1 (size=41): Suggested name: Premium Low-Volume
  Avg Price: $13.74 Avg Units Sold: 78.4 Avg Profit: $397.52
  Insight: High price with low volume; target with promotions to increase trial.

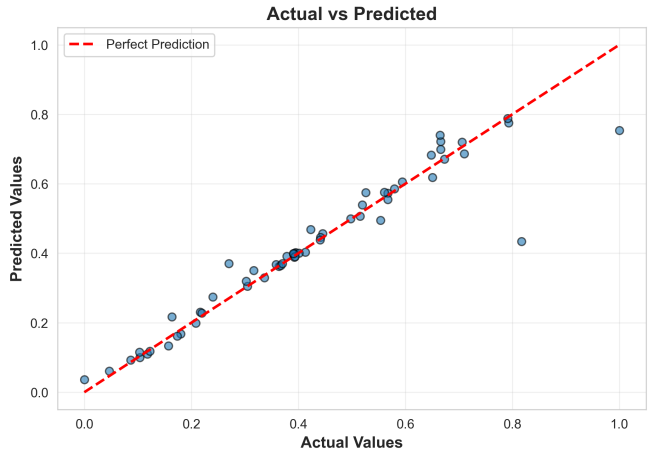
Cluster 2 (size=63): Suggested name: Budget Best-Sellers
  Avg Price: $3.69 Avg Units Sold: 462.0 Avg Profit: $708.26
  Insight: Low price, very high units sold; keep stock and consider small margin optimizations.

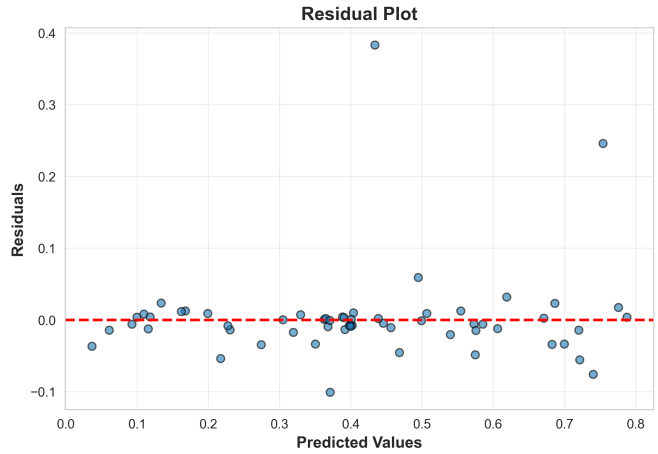


Regression Analysis

Models trained: Linear Regression and Polynomial Regression (degree=2).
Best model by test R^2: Polynomial_2 (R^2=0.9100).
Train/Test split: 137 / 59 (test_size=0.3).







Conclusion & Suggested Fixes

Missing formal REPORT.pdf was generated (this document).
Suggested fixes to meet assignment fully:
- Add a formal REPORT.pdf (done).
- Enhance notebook to automatically name clusters and include interpretation (we added suggested names in this report).
- Add a results/ folder with saved plots (done).
- Include explicit justification for preprocessing choices in the notebook or REPORT (this PDF includes a summary; expand w
- Add a short README section describing how to reproduce the report generation (optional).