

SALES FORECAST 2024: PREDICTIVE MODELING FOR UNIEURO'S FULL PRODUCT PORTFOLIO



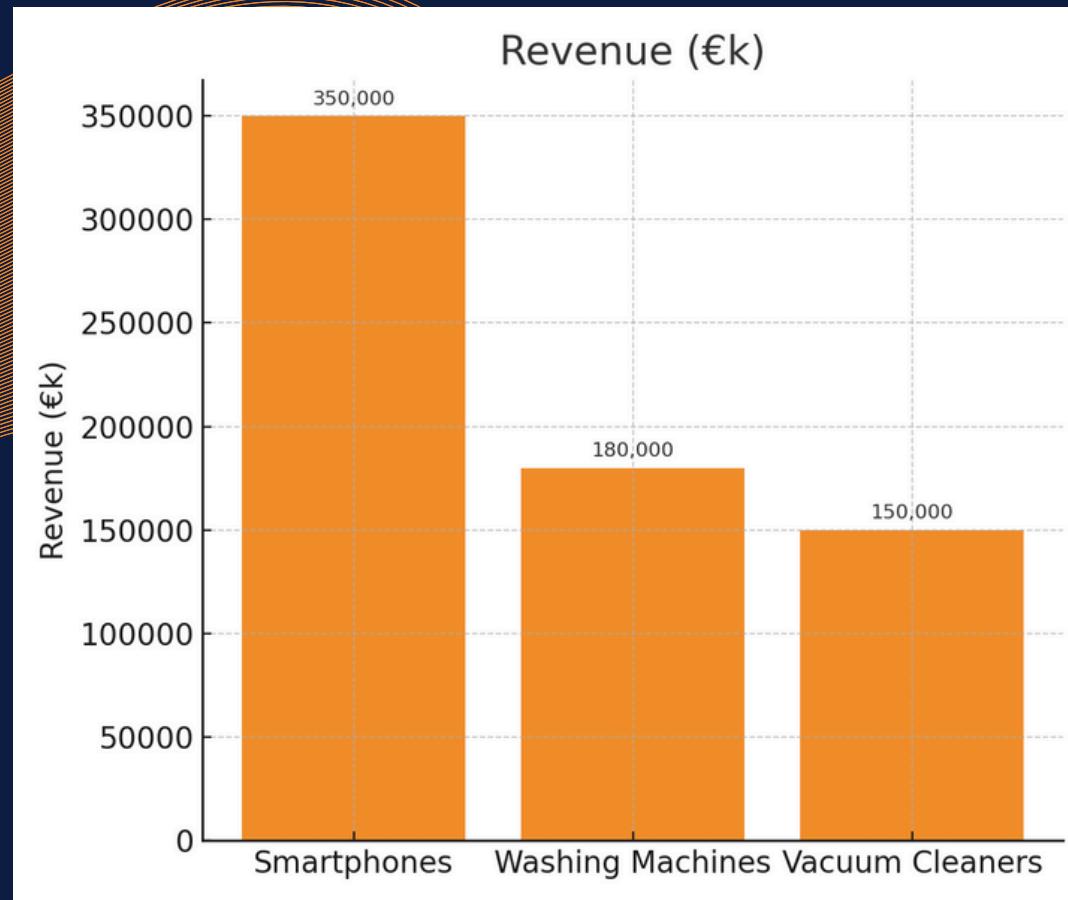
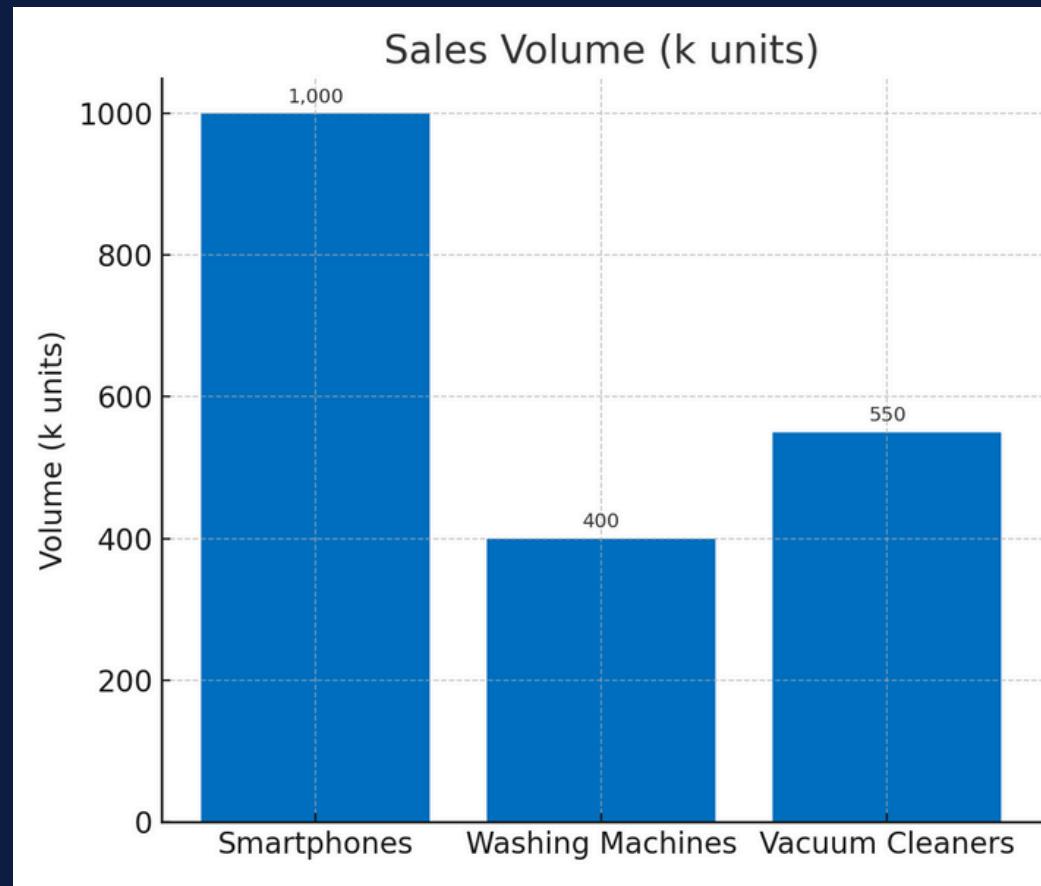
AN INTEGRATED ANALYTICS APPROACH TO SUPPORT
PRICING, PURCHASING, AND PROMOTION DECISIONS

24/04/25

PRESENTED BY:

ANALaura Granata
Daniele Lupico
Gabriele Rizzo

BUSINESS CONTEXT - KEY PRODUCT GROUPS



Unieuro operates in the Italian consumer electronics market, characterized by:

- **Fragmentation** (tech superstores ~50% market share)
- **Customer preference** for proximity shopping, despite online growth (online penetration ~25% vs. mature markets >40%)

Three strategic Product Groups for Unieuro:

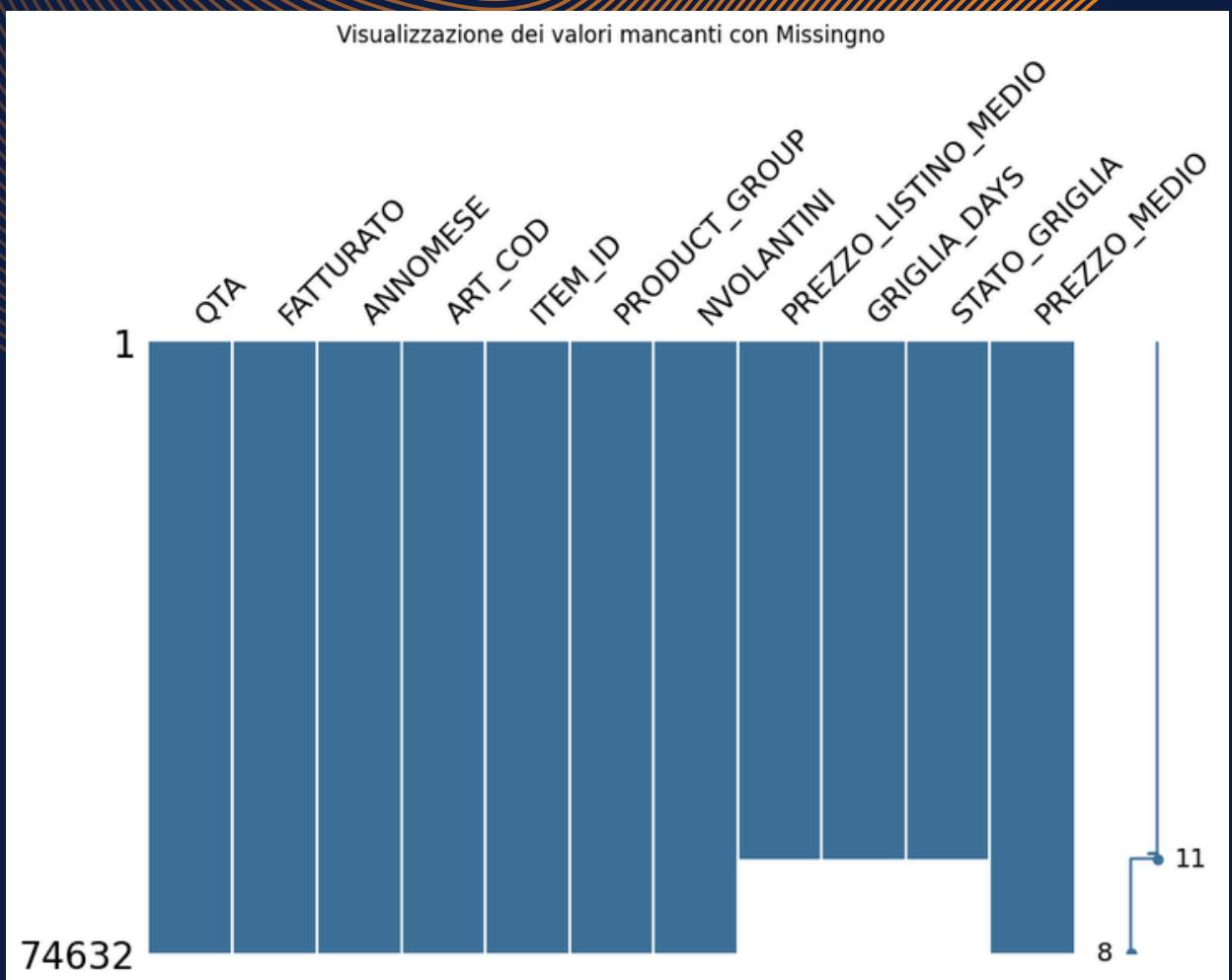
- Smartphones (Telecom)
- Washing Machines (Large Domestic Appliances)
- Vacuum Cleaners (Small Domestic Appliances)

Key figures (2017-2023):

- **Smartphones:**
 - ~1200 unique products
 - Highest sales volume among product groups
- **Washing Machines:**
 - ~450 unique products
 - Stable revenue, with seasonal peaks (Black Friday, Christmas)
- **Vacuum Cleaners:**
 - ~600 unique products
 - Post-pandemic growth, increasing model variety

PROJECT GOAL AND FIRST APPROACH

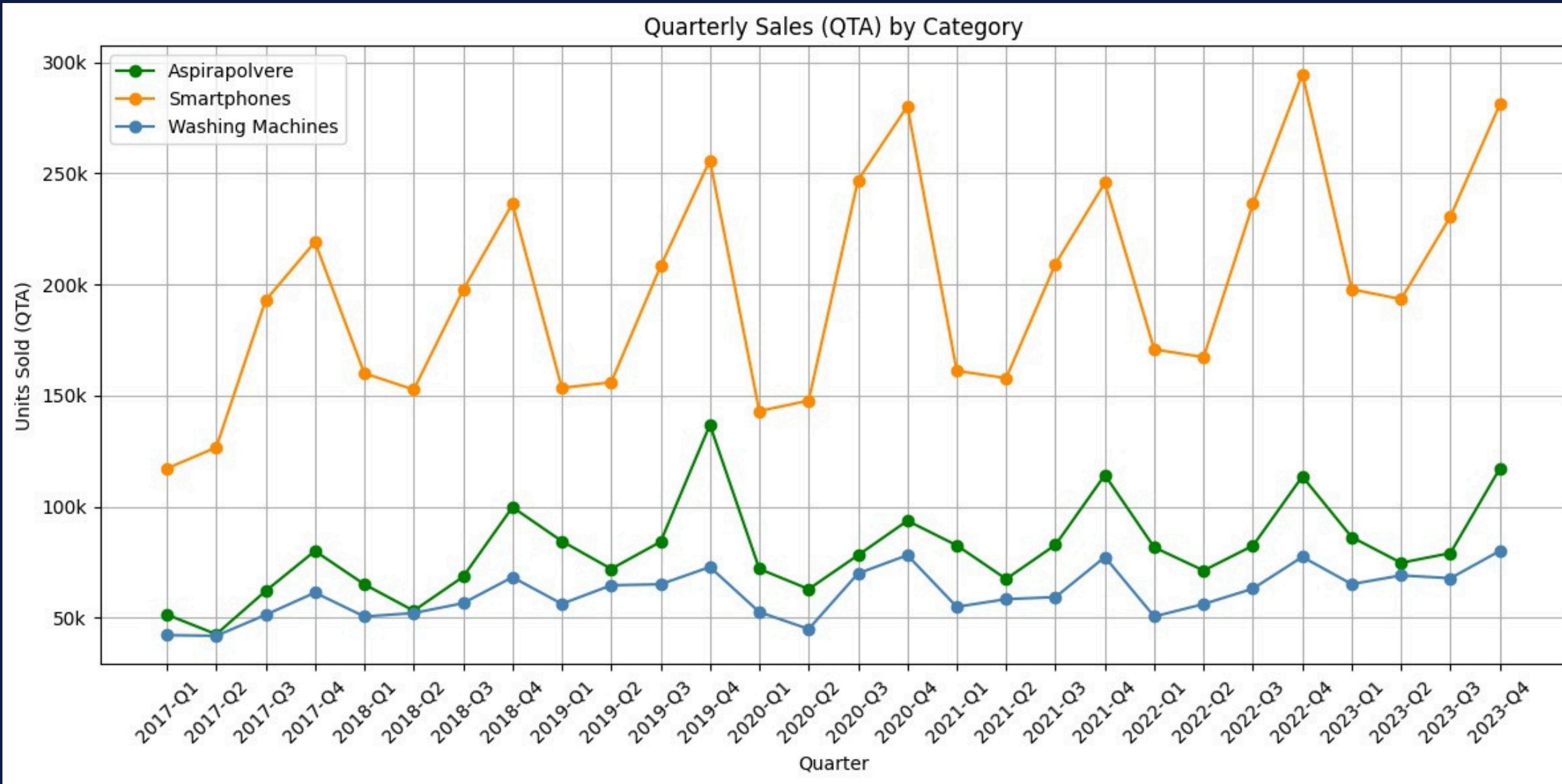
Main objective: Forecast the number of units sold in 2024 for every product-brand-category combination



The approach up to now:

1. PRIMA_VENDITA' to datetime format (YYYY-MM).
2. Dropped rows where ITEM_ID == "ND"
3. Inner joined AN_PROD with RISULTATI_ANNAMESE.
 - a. ~37,000 rows were dropped due to unmatched values, these lacked both ITEM_ID and ART_COD.
4. Merged again with INFO_BUSINESS.
5. Filled missing values in PREZZO_LISTINO_MEDIO using the median per ITEM_ID group.
6. Filling missing values in STATO_GRIGLIA with the mode per ITEM_ID group.
7. Filling GRIGLIA_DAYS missing values using a three-step hierarchical imputation:
 - a. Step 1: Filled NaN using the median per (ITEM_ID, STATO_GRIGLIA) group.
 - b. Step 2: For remaining NaNs, used the median per ITEM_ID group.
 - c. Step 3: Final fallback — filled any still missing values with the overall median of the GRIGLIA_DAYS column.
8. Main dataset split into three separate datasets based on the PRODUCT_GROUP column

HERE IS WHAT WE UNDERSTOOD AFTER THE CLEANING



TEMPORAL TRENDS:

- BLACK FRIDAY
- CHRISTMAS

SAME PICKS
BUT DIFFERENT
MAGNITUDE

PANDEMIC-RELATED
GROWTH

MODELING APPROACH

WE AIM TO TEST A COMBINATION OF SIMPLE AND ADVANCED MODELS TO FORECAST 2024 PRODUCT QUANTITIES. MODELS WERE TRAINED ON 2017–2023 DATA.

Model	Purpose	Notes
Naïve model	Simple benchmark ($2024 = 2023$)	Fast and interpretable
Moving Average	Captures short-term trends	Baseline
ARIMA	Main forecasting model	Robust with seasonality, requires tuning
Prophet (limited)	To be tested on complex or irregular series	Easy to scale, handles holidays well
Random Forest	Understand feature importance (explanatory)	Not for time-series prediction

ECONOMIC INTERPRETATION

- Identify growing or declining products → BCG Matrix
- Optimize purchasing decisions with predicted demand
- Reduce out-of-stock or overstock situations
- Support pricing and promotion strategy with feature insights
- Insights about product lifecycle and demand shifts

QUESTIONS

1. Should we fine tune all the models and show the before and after of all the models with a possible (desirable) performance improvement? Or should we select only one model to fine-tune (e.g. Random Forest) and, with the important variables found, retune the models before the RF?
2. Could it be optimal to find the model tailored for each different product_group dataset? Or is it counterproductive?
3. What do you think about the cleaning part is it sufficient? Do you suggest a different approach?

THANK YOU