Big Data Analytics

 # 8: In-Memory Analytics with Pandas. ABC Analysis

Instructor: Oleh Tymchuk

#10: Agenda

- ABC Analysis
- XYZ Analysis
- Other Data Classification Methods
- Practical cases
- Useful Links

ABC Analysis

What is ABC Analysis?

ABC Analysis is a categorization technique used to classify items, products, or activities based on their importance or value to an organization. It is rooted in the Pareto Principle (80/20 rule), which states that roughly 80% of the effects come from 20% of the causes.

In the context of ABC Analysis:

- A small percentage of items (Category A) contribute the most value
- A larger percentage of items (Categories B and C) contribute progressively less value

Purpose of ABC Analysis

The primary goal of ABC Analysis is to prioritize resources and efforts by focusing on the most critical items.

It helps organizations:

Optimize Inventory Management

Focus on high-value items to reduce costs and improve efficiency.

Improve Decision-Making

Allocate resources (time, money, effort) to the most impactful areas.

Enhance Profitability

Identify and prioritize high-revenue or high-margin products or customers.

Concept

Items are categorized into three classes:

A: High-value items (e.g., 20% of items contributing to 80% of value).

B: Moderate-value items.

C: Low-value items (e.g., 50% of items contributing to 5% of value).

Pareto Principle (80/20 Rule): Often used as the basis for ABC classification

Steps to Perform ABC Analysis

Collect Data: gather data on items (e.g., sales, revenue, inventory levels)

Calculate Contribution: compute the contribution of each item to the total value

Sort Items: sort items in descending order of contribution

Classify Items: assign items to categories A, B, or C based on cumulative contribution thresholds

Data:

Product	Revenue
P1	1000
P2	600
P3	300
P4	100
P5	50

1. **Total Revenue:** 1000 + 600 + 300 + 100 + 50 = **2050**

2. **Calculate Contribution:**
   - P1: 1000 / 2050 = 48.78%
   - P2: 600 / 2050 = 29.27%
   - P3: 300 / 2050 = 14.63%
   - P4: 100 / 2050 = 4.88%
   - P5: 50 / 2050 = 2.44%

3. **Sort Items:**
   - P1 (48.78%), P2 (29.27%), P3 (14.63%), P4 (4.88%), P5 (2.44%)

4. **Calculate Cumulative Percentage:**
   - P1: 48.78%
   - P2: 48.78% + 29.27% = 78.05%
   - P3: 78.05% + 14.63% = 92.68%
   - P4: 92.68% + 4.88% = 97.56%
   - P5: 97.56% + 2.44% = 100%

Classify Items:

- **A:** P1, P2 (Cumulative ≤ 80%)
- **B:** P3 (Cumulative ≤ 95%)
- **C:** P4, P5 (Cumulative ≤ 100%)

Applications of ABC Analysis

Inventory Management:

Focus on high-value items (Category A) for better control.

Sales and Marketing:

Prioritize high-revenue customers or products.

Resource Allocation:

Allocate resources efficiently based on item importance.

XYZ Analysis

XYZ Analysis

Concept:

- Classifies items based on their demand variability or consistency over time.
- Focuses on understanding how predictable or stable the demand for items is.

Purpose:

- Optimize inventory management and supply chain planning.
- Identify items with stable, moderate, or highly variable demand.

Categories:

- X: Stable demand (low variability).
- Y: Moderate variability.
- Z: Highly variable demand.

Applications:

- Inventory management: Stock items with stable demand more efficiently.
- Supply chain: Plan for variability in demand.
- Forecasting: Improve accuracy for items with predictable demand.

Example:

- X: Essential goods with consistent demand (e.g., bread, milk); Y: Seasonal products (e.g., winter coats); Z: Unpredictable items (e.g., luxury goods).

Formula for XYZ Analysis

The most common metric used is the **coefficient of variation (CV)**, which is defined as:

$$CV = \frac{\text{Standard Deviation of Demand}}{\text{Mean Demand}} \times 100$$

- $CV < 10\%$ → X-class

- $10\% \leq CV < 25\%$ → Y-class

- $CV \geq 25\%$ → Z-class

Example of XYZ Analysis

Product	Jan	Feb	Mar	Apr	May	Jun
A	100	102	98	101	99	100
B	80	70	85	90	60	95
C	30	5	80	0	100	10

Product A

- **Mean** = (100 + 102 + 98 + 101 + 99 + 100)/6 = **100**
- **Standard Deviation** ≈ **1.41**
- **CV** = (1.41 / 100) × 100 ≈ **1.41%**

→ **Product A = X-class**

Product B

- Mean ≈ 80
- Std ≈ 12.91
- CV ≈ (12.91 / 80) × 100 ≈ **16.14%**

→ **Product B = Y-class**

Product C

- Mean ≈ 37.5
- Std ≈ 41.56
- CV ≈ (41.56 / 37.5) × 100 ≈ **110.83%**

→ **Product C = Z-class**

Other Data Classification Methods

Pareto Analysis

Concept:

- Focuses on identifying the most significant factors that contribute to a problem or outcome.
- Based on the Pareto Principle (80/20 Rule): 80% of effects come from 20% of causes.

Purpose:

- Prioritize efforts on the most impactful factors.
- Solve problems efficiently by addressing the root causes.

Applications:

- Quality control: Identify key defects in manufacturing.
- Sales: Focus on top-performing products or customers.
- Time management: Prioritize high-impact tasks.

Example:

- In a business, 20% of customers may generate 80% of revenue.

Cluster Analysis

Concept:

- Groups items or data points into clusters based on their similarity.
- Uses algorithms to identify patterns and relationships in data.

Purpose:

- Discover natural groupings within data.
- Simplify complex datasets for analysis.

Applications:

- Customer segmentation: Group customers with similar behaviors.
- Market research: Identify target audiences.
- Image recognition: Group similar images.

Example:

- Segmenting customers into groups like "high spenders," "frequent buyers," and "occasional shoppers."

Comparison of Classification Methods

Method	Basis	Use Case
ABC Analysis	Value/Importance	Inventory management, sales analysis
XYZ Analysis	Demand Variability	Supply chain optimization
Pareto Analysis	Significant Factors	Problem-solving, decision-making
Cluster Analysis	Similarity/Grouping	Customer segmentation, market research

Practical cases

Useful Links

[ABC-XYZ Inventory Classification with Python](#)

[Inventory Management -- ABC Analysis(Python)](#)

Q&A