Desafio 05 - Equipamento Defeituoso

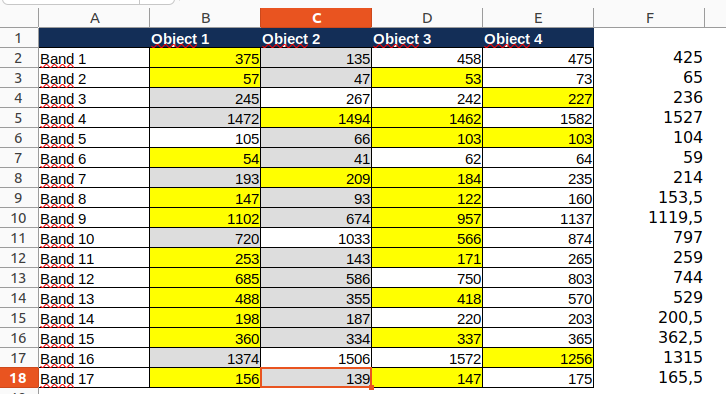
## **Descrição do Problema:**

A planilha apresenta 4 equipamentos (denominados “Object” de 1 a 4) e valores de sensores (denominados “Band” de 1 a 17), o objetivo do desafio é saber qual dos equipamentos é o defeituoso com base nos resultados apresentados. A grande dificuldade é a vastidão de informações que não permitem tomar uma decisão concreta sobre qual é o equipamento defeituoso.

## **Abordagem adotada:**

### **Análise Exploratória de dados (EDA):**

Inicialmente, chequei os dados para identificar algum padrão, ou até mesmo algum padrão semelhante, tentei calcular a média de cada um dos sensores e ver os equipamentos que estavam abaixo da média, mas dois equipamentos tiveram mesmo relatório, então comparei os dois e vi qual dos dois apresentava o menor valor dos sensores.



Pela minha análise o Objeto 2 seria o equipamento defeituoso, no entanto esse método era muito superficial, então fui em busca de mais informações.

### **Tecnologia Utilizada: Detecção de Anomalias:**

Em minhas pesquisas, descobri a ideia da detecção de anomalias, baseada na suposição de que equipamentos saudáveis devem ter comportamentos semelhantes (valores próximos entre si) e um equipamento com defeito deve apresentar valores diferentes.

### **Técnica Aplicada: Análise de Distância e PCA:**

Para representar os dados em um espaço onde os padrões anômalos se destacam, fiz uso da redução de dimensionalidade (PCA - Análise de Padrões Comportamentais). Após, apliquei a “distância Mahalanobis”- Detector de padrões estranhos.

## **Como a técnica foi aplicada?**

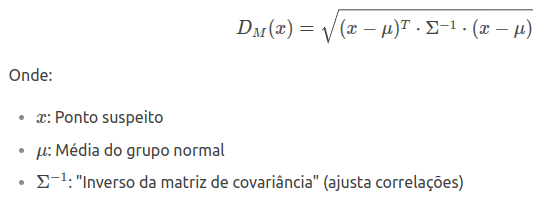
1. Fiz uma matriz transposta dos dados para obter um valor de leitura por equipamento.
2. Padronizei os dados (z-score).
3. Apliquei o PCA para reduzir a dimensão dos dados para 2 componentes principais.
4. Calculei a distância mahalanobis entre os equipamentos no espaço de PCA.
5. Análise do resultado - o equipamento com maior distância média em relação aos outros foi identificado como defeituoso.

## **Justificativa das técnicas:**

* **PCA - Análise de padrões comportamentais:** sua ideia consiste em “compactar informações”, ou seja, juntar medidas parecidas, além disso é possível descartar informações que não ajudam a diferenciar os equipamentos e cria um “mapa simplificado” do contexto. Matematicamente falando, o PCA encontra autovetores da matriz de covariância dos dados, que representam as direções máximas de variância, isso permite a visualização em 2D ou 3D mantendo a estrutura essencial dos dados.
* **Outliers:** são pontos que se desviam significativamente da distribuição normal dos dados, no espaço PCA, os outliers aparecem distantes do aglomerado principal.
* **Distância de mahalanobis:** é a medida estatística que considera a estrutura da covariância dos dados para identificar outliers. Ela considera os seguintes parâmetros: Onde a maioria está? (média do grupo), o que é comum? (dispersão dos dados), Correlação entre as variáveis.

**OBS:** Importante destacar que a distância mahalanobis difere da euclidiana pelos seguintes aspectos: A distância mahalanobis mede o desvio padrão, enquanto a euclidiana mede linha reta, além disso a distância mahalanobis é sensível a escala e considera a correlação, enquanto a euclidiana não, por fim, a mahalanobis é melhor para dados com variância diferentes, ao contrário da euclidiana é ideal para dados normalizados.

**Fórmula da distância de mahalanobis:**



**Curiosidade:** A distância mahalanobis foi criada por Prasanta Chandra Mahalanobis em 1936 para estudar os crânios humanos.

## **Conclusão:**

Com base na análise PCA combinada com a métrica de distância de Mahalanobis, o equipamento 2 (Object 2) foi identificado como o que apresenta comportamento anômalo em relação aos demais. Isso sugere que este equipamento está potencialmente com defeito, apresentando padrões de leitura de sensores significativamente diferentes dos outros equipamentos operando normalmente.

