



**ANÁLISE DE CLUSTERIZAÇÃO PARA DIVISÃO DE TURMAS DE CAPACITAÇÃO:  
DIAGNÓSTICO DE MATURIDADE EM DATA & ANALYTICS  
[ UNIDADE DE PROJETOS, PROCESSOS E INTELIGÊNCIA DE DADOS ]**

**I. OBJETIVOS:**

Esta ação faz parte do Projeto Impulsionador de Inteligência de Dados do Sebrae/ES (vinculado ao Programa Nacional de Inteligência de Dados), e tem como foco trabalhar o eixo de Cultura Analítica no Sebrae/ES.

O presente relatório é fruto de um esforço de qualificação da tomada de decisão baseada em dados, e serve como um complemento ao relatório analítico dos resultados da pesquisa de Diagnóstico, no qual foram identificados os colaboradores que atuarão como Pontos Focais de Dados em cada unidade, e onde foram listados os colaboradores aderentes aos perfis indicados para capacitações em ferramentas de análise de dados.

A partir dessa primeira análise, foi identificada a necessidade de fornecer subsídio para priorização de seleção de colaboradores para seu agrupamento em turmas de capacitação nas ferramentas de análise de dados. Com isso, os objetivos desta análise são:

- 1) Agrupar colaboradores em grupos menores (turmas) de acordo com seus perfis, para cada grupo de pessoas que compartilham da mesma necessidade de capacitação (Excel Básico, Excel Intermediário/Avançado, PowerBI Básico, PowerBI Intermediário/Avançado);
- 2) Subsidiar a UGP com informações para seleção e priorização de turmas para capacitação;
- 3) Otimizar o nivelamento das turmas e impulsionar o aproveitamento dos conteúdos dos cursos.

Para as análises foram incluídos os dados do diagnóstico com as respostas recebidas até o dia 24/08/2023, totalizando 98 respondentes. As respostas duplicadas foram tratadas com a exclusão das respostas mais antigas.



## **II. METODOLOGIA:**

Para a realização da divisão de turmas foi utilizado o algoritmo K-means, um método de aprendizado não supervisionado que particiona os dados em um número K de agrupamentos não sobrepostos, chamados clusters. O algoritmo foi rodado em linguagem R, utilizando variáveis de interesse para o agrupamento ótimo dos indivíduos, para cada ferramenta e nível de capacitação, de acordo com suas características comuns.

Os resultados obtidos da clusterização dos indivíduos foram validados com uso de métodos estatísticos e analisados com suporte de visualização gráfica em PowerBI. As análises resultaram nas sugestões de divisões de turmas listadas abaixo, para cada ferramenta e nível de capacitação.



### III. RESULTADOS DA CLUSTERIZAÇÃO - DIVISÃO DE TURMAS:

#### ➤ CAPACITAÇÃO NÍVEL BÁSICO EM EXCEL:

##### Critérios para seleção:

1. Utiliza o Excel;
2. Possui conhecimentos básicos ou menos em Excel;
3. Possui dificuldade para realizar algumas tarefas no Excel;
4. Possui interesse em aprofundar conhecimentos em análise de dados.

##### Filtros aplicados:

1. Utiliza Excel - Incluídos (1): Sim
2. Conhecimento: Microsoft Excel - Incluídos (2): Básico; Nenhum conhecimento.
3. Interesse em aprofundar conhecimentos em análise de dados - Incluídos (1) Sim

##### Variáveis utilizadas para clusterização:

1. Variáveis de nível de conhecimentos em análise de dados, estatística e raciocínio lógico;
2. Variáveis de nível de conhecimento em ferramentas de análise de dados;
3. Variáveis de nível de dificuldade na interpretação de gráficos;
4. Variáveis de nível de dificuldade na execução de tarefas em Excel.

##### Número ideal de clusters: 2 clusters

- **Cluster 1:** 17 colaboradores (menores médias)
- **Cluster 2:** 14 colaboradores (maiores médias)



➤ **CAPACITAÇÃO NÍVEL INTERMEDIÁRIO/AVANÇADO EM EXCEL:**

**Critérios para seleção:**

1. Utiliza o Excel;
2. Possui conhecimentos básicos ou mais em Excel;
3. Possui interesse em aprofundar conhecimentos em análise de dados.

**Filtros aplicados:**

1. Utiliza Excel - Incluídos (1): Sim
2. Conhecimento: Microsoft Excel - Incluídos (2): Intermediário; Avançado.
3. Interesse em aprofundar conhecimentos em análise de dados - Incluídos (1) Sim

**Variáveis utilizadas para clusterização:**

1. Variáveis de nível de conhecimentos em análise de dados, estatística e raciocínio lógico;
2. Variáveis de nível de conhecimento em ferramentas de análise de dados;
3. Variáveis de nível de dificuldade na interpretação de gráficos;
4. Variáveis de nível de dificuldade na execução de tarefas em Excel.

**Número ideal de clusters:** 4 clusters

- **Cluster 1:** 9 colaboradores (menores médias - não devem ser mesclados)
- **Cluster 2:** 17 colaboradores (médias intermediárias, se aproximam mais do cluster 3 do que 1)
- **Cluster 3:** 30 colaboradores (médias próximas ao cluster 4, podem ser mesclados ou divididos)
- **Cluster 4:** 7 colaboradores (maiores médias)



➤ **CAPACITAÇÃO NÍVEL BÁSICO EM POWER BI:**

**Critérios para seleção:**

1. Utiliza o Excel;
2. Utiliza o PowerBI (ainda que apenas para consultas);
3. Possui conhecimentos intermediários ou mais em Excel;
4. Consulta dados em painéis com frequência mensal ou maior;
5. Nível de conhecimento básico ou menor em PowerBI;
6. Possui interesse em aprofundar conhecimentos em análise de dados

**Filtros aplicados:**

1. Utiliza Excel - Incluídos (1): Sim
2. Utiliza Power BI - Incluídos (1): Sim
3. Frequência de consulta a painéis - Incluídos (4): Mensalmente; Algumas vezes por mês; Semanalmente; Diariamente.
4. Conhecimento: Microsoft Excel - Incluídos (2): Intermediário; Avançado.
5. Conhecimento: PowerBI - Incluídos (2): Básico; Nenhum conhecimento.
6. Interesse em aprofundar conhecimentos em análise de dados - Incluídos (1) Sim

**Variáveis utilizadas para clusterização:**

1. Variáveis de nível de conhecimentos em análise de dados, estatística e raciocínio lógico;
2. Variáveis de nível de conhecimento em ferramentas de análise de dados;
3. Variáveis de nível de dificuldade na interpretação de gráficos;
4. Variáveis de nível de dificuldade na execução de tarefas em Excel;
5. Variáveis de nível de dificuldade na execução de tarefas em PowerBI;
6. Frequência de consulta a painéis PowerBI.

**Número ideal de clusters: 3 clusters**

- **Cluster 1:** 9 colaboradores (menores médias - não devem ser mesclados com 2 e 3)
- **Cluster 2:** 13 colaboradores (médias próximas ao cluster 3 - podem ser mesclados)
- **Cluster 3:** 19 colaboradores (maiores médias)



➤ **CAPACITAÇÃO NÍVEL INTERMEDIÁRIO EM POWER BI:**

**Critérios para seleção:**

1. Utiliza o Excel
2. Utiliza o PowerBI (ainda que apenas para consultas)
3. Consulta dados em painéis com frequência mensal ou maior
4. Possui conhecimentos intermediários ou mais em Excel
5. Nível de conhecimento básico ou maior em PowerBI
6. Possui interesse em aprofundar conhecimentos em análise de dados

**Filtros aplicados:**

1. Utiliza Excel - Incluídos (1): Sim
2. Utiliza Power BI - Incluídos (1): Sim
3. Frequência de consulta a painéis - Incluídos (4): Mensalmente; Algumas vezes por mês; Semanalmente; Diariamente.
4. Conhecimento: Microsoft Excel - Incluídos (2): Intermediário; Avançado.
5. Conhecimento: PowerBI - Incluídos (2): Intermediário; Avançado.
6. Interesse em aprofundar conhecimentos em análise de dados - Incluídos (1) Sim

**Variáveis utilizadas para clusterização:**

1. Variáveis de nível de conhecimentos em análise de dados, estatística e raciocínio lógico;
2. Variáveis de nível de conhecimento em ferramentas de análise de dados;
3. Variáveis de nível de dificuldade na interpretação de gráficos;
4. Variáveis de nível de dificuldade na execução de tarefas em Excel;
5. Variáveis de nível de dificuldade na execução de tarefas em PowerBI;
6. Frequência de consulta a painéis PowerBI.

**Número ideal de clusters: 1 cluster**

- **Cluster 1:** 7 colaboradores