



# Dicionário de dados

Dados utilizados para o desenvolvimento do projeto Ebettha

Este é o dicionário de dados do Projeto Ebettha. Ele é um recurso essencial para o entendimento do trabalho desenvolvido pelo grupo, haja vista que traz uma descrição de todos os dados utilizados para construção do modelo preditivo. Além da descrição básica acerca de cada coluna, este documento conta com uma análise matemática que compreende a estatística descritiva de cada *feature*.

## Tabela prep\_df\_job\_opp\_workstyle

### Exemplo de preenchimento

id	job_opportunity_id	score_classic	score_order	score_change	score_tireless	score_explorer	score_specialist	score_generalist
1915	2900	0.0	2.0	4.0	6.0	3.5	2.0	3.0
1804	2727	0.5	5.0	3.5	4.0	1.5	3.0	2.0
1916	2902	2.0	4.5	3.5	4.0	5.0	3.0	1.0
1803	2711	1.0	4.5	4.0	4.5	3.0	1.0	3.0
1865	2806	0.0	4.0	5.0	5.0	4.0	2.0	2.0

**Descrição:** Esta tabela armazena os resultados dos testes do tipo “workstyle” preenchido pelos gestores das vagas postadas dentro da plataforma Bettha.

**Tamanho:** 168 linhas (1 head + 167 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Numérica (float)	Identificador único utilizado para indicar o gestor.
job_opportunity_id	Numérica (float)	Identificador utilizado para indicar a qual vaga aquele gestor pertence.
score_classic	Numérica (float)	Score do gestor ao avaliar a competência “classic”
score_order	Numérica (float)	Score do gestor ao avaliar a competência “order”
score_change	Numérica (float)	Score do gestor ao avaliar a competência “change”
score_tireless	Numérica (float)	Score do gestor ao avaliar a competência “tireless”
score_explorer	Numérica (float)	Score do gestor ao avaliar a competência “explorer”
score_specialist	Numérica (float)	Score do gestor ao avaliar a competência “specialist”
score_generalist	Numérica (float)	Score do gestor ao avaliar a competência “generalist”

## Análise matemática das colunas

Nome	Valor mínimo	Valor máximo	Moda	Média aritmética	Desvio padrão	Amplitude	Percentil 25	Percentil 50	Percentil 75
score_classic	0.0	4.0	1.0	1.607143	0.897673	4.0	1.0	1.5	2.0
score_order	1.0	7.0	4.0	4.238095	1.270287	6.0	3.5	4.0	5.0
score_change	1.5	7.0	5.0	4.330357	1.091379	5.5	3.5	4.5	5.0
score_tireless	1.5	7.5	4.0	4.107143	1.113530	6.0	3.5	4.0	5.0
score_explorer	0.0	7.0	3.0	3.330357	1.426682	7.0	2.0	3.0	4.5
score_specialist	0.0	4.0	3.0	2.136905	1.077398	4.0	1.0	2.0	3.0
score_generalist	0.0	4.0	3.0	2.136905	1.152584	4.0	1.0	3.0	3.0

**Observações:** Colunas “id” e “job\_opportunity\_id” não foram utilizadas nesta análise pois são identificadores. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

## Tabela prep\_df\_jb\_opp\_superfit

### Exemplo de preenchimento

id	job_opportunity_id	score_res	score_eng	score_int	score_cur	score_sin	score_dis
1815	2727	44.0	40.0	38.0	38.0	30.0	42.0
1816	2686	42.0	34.0	44.0	28.0	26.0	38.0
1928	2900	38.0	20.0	36.0	30.0	34.0	32.0
1862	2755	26.0	30.0	34.0	24.5	42.0	26.0
1817	2749	24.0	16.0	42.0	20.0	32.0	24.0

**Descrição:** Esta tabela armazena os resultados dos testes do tipo “*genius*” preenchido pelos gestores das vagas postadas dentro da plataforma Bettha.

**Tamanho:** 169 linhas (1 head + 168 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Numérica (float)	Identificador único utilizado para indicar o gestor.
job_opportunity_id	Numérica (float)	Identificador utilizado para indicar a qual vaga aquele gestor pertence.
score_res	Numérica (float)	Score do gestor ao avaliar a competência “resiliente”
score_eng	Numérica (float)	Score do gestor ao avaliar a competência “engajado”
score_int	Numérica (float)	Score do gestor ao avaliar a competência “colaborativo”
score_cur	Numérica (float)	Score do gestor ao avaliar a competência “original”
score_sin	Numérica (float)	Score do gestor ao avaliar a competência “pragmático”
score_dis	Numérica (float)	Score do gestor ao avaliar a competência “consistente”

## Análise matemática das colunas

Nome	Valor mínimo	Valor máximo	Moda	Média aritmética	Desvio padrão	Amplitude	Percentil 25	Percentil 50	Percentil 75
score_res	16.0	46.0	32.0	32.532544	6.204871	30.0	28.0	32.0	36.0
score_eng	10.0	46.0	30.0	29.502959	6.910245	36.0	26.0	30.0	34.0
score_int	16.0	52.0	36.0	33.988166	7.356555	36.0	30.0	34.0	38.0
score_cur	12.0	46.0	30.0	27.076923	7.030546	34.0	22.0	26.0	32.0
score_sin	16.0	42.0	34.0	33.372781	6.020557	32.0	28.0	34.0	38.0
score_dis	8.0	52.0	36.0	35.846154	7.562439	44.0	30.0	36.0	42.0

**Observações:** Colunas “id” e “job\_opportunity\_id” não foram utilizadas nesta análise pois são identificadores. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

## Tabela prep\_df\_user\_lifestyle

### Exemplo de preenchimento

id	key	lifestyle_classic	lifestyle_order	lifestyle_change	lifestyle_tireless	lifestyle_explorer	lifestyle_specialist	lifestyle_generalist	lifestyle_hybrid
1019062	ls_explorer_generalist	2.50	4.00	4.00	5.00	5.50	1.00	3.00	0.00
1019131	ls_change_generalist	1.00	2.50	6.00	4.50	4.50	0.00	3.00	1.00
1019138	ls_explorer_generalist	1.50	1.50	4.00	6.00	8.50	2.00	3.00	0.00
1017352	ls_explorer_specialist	1.00	2.50	4.50	3.50	6.00	4.00	0.00	0.00
1019110	ls_explorer_generalist	1.50	3.00	3.50	4.00	6.00	0.00	4.00	0.00

**Descrição:** Esta tabela armazena os resultados dos testes do tipo “lifestyle” preenchido pelos candidatos que estão tentando alguma vaga dentro da plataforma Bettha.

**Tamanho:** 67.508 linhas (1 head + 67.507 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Numérica (float)	Identificador único utilizado para indicar o candidato.
key	Categórica (nominal)	Indicador utilizado para identificar qual as duas competências mais presentes do usuário formado por “is + (order, change, tireless, explorer) + (specialist, generalista, hybrid).
lifestyle_classic	Numérica (float)	Score do candidato ao avaliar a competência “classic”
lifestyle_order	Numérica (float)	Score do candidato ao avaliar a competência “order”
lifestyle_change	Numérica (float)	Score do candidato ao avaliar a competência “change”
lifestyle_tireless	Numérica (float)	Score do candidato ao avaliar a competência “tireless”
lifestyle_explorer	Numérica (float)	Score do candidato ao avaliar a competência “explorer”
lifestyle_specialist	Numérica (float)	Score do candidato ao avaliar a competência “specialist”
lifestyle_generalist	Numérica (float)	Score do candidato ao avaliar a competência “generalist”
lifestyle_hybrid	Numérica (float)	Score do candidato ao avaliar a competência “hybrid”

## Análise matemática das colunas

Nome	Valor mínimo	Valor máximo	Moda	Média aritmética	Desvio padrão	Amplitude	Percentil 25	Percentil 50	Percentil 75
lifestyle_classic	0.0	8.5	2.0	2.216078	0.869838	8.5	1.5	2.0	3.0
lifestyle_order	0.0	24.5	4.0	3.464056	1.324757	24.5	2.5	3.5	4.5
lifestyle_change	0.0	18.0	6.0	5.775182	1.552299	18.0	4.5	6.0	7.0
lifestyle_tireless	0.0	22.5	5.0	4.740016	1.334286	22.5	4.0	4.5	5.5
lifestyle_explorer	0.0	23.5	3.0	5.032241	1.381420	23.5	4.0	5.0	6.0
lifestyle_specialist	0.0	5.0	3.0	1.859143	1.165238	5.0	1.0	2.0	3.0
lifestyle_generalist	0.0	5.0	1.0	1.470907	1.112443	5.0	1.0	1.0	2.0
lifestyle_hybrid	0.0	4.0	0.0	0.876444	0.931191	4.0	0.0	1.0	1.0

**Observações:** Colunas “id” não foi utilizada nesta análise pois se caracteriza como um identificador. Além disso, a coluna “key” é categórica. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

**Transformação:** Considerando o teor categórico da coluna “key”, foi necessário transformar esses dados em dados numéricos. A correlação entre o valor original e o novo valor pode ser observada abaixo:

Valor original	Novo valor	Valor original	Novo valor	Valor original	Novo valor
'ls_explorer_generalist'	1	'ls_change_hybrid'	6	'ls_tireless_hybrid'	11
'ls_change_generalist'	2	'ls_tireless_generalist'	7	'ls_order_generalist'	12
'ls_explorer_specialist'	3	'ls_order_specialist'	8	'ls_classic_specialist'	13
'ls_explorer_hybrid'	4	'ls_tireless_specialist'	9	'ls_classic_hybrid'	14
'ls_change_specialist'	5	'ls_order_hybrid'	10	'ls_classic_generalist'	15

## Exemplo de preenchimento após alterações

id	key	lifestyle_classic	lifestyle_order	lifestyle_change	lifestyle_tireless	lifestyle_explorer	lifestyle_specialist	lifestyle_generalist	lifestyle_hybrid
1019062	1	2.50	4.00	4.00	5.00	5.50	1.00	3.00	0.00
1019131	2	1.00	2.50	6.00	4.50	4.50	0.00	3.00	1.00
1019138	1	1.50	1.50	4.00	6.00	8.50	2.00	3.00	0.00
1017352	3	1.00	2.50	4.50	3.50	6.00	4.00	0.00	0.00
1019110	1	1.50	3.00	3.50	4.00	6.00	0.00	4.00	0.00



## Tabela prep\_df\_user\_superfit

### Exemplo de preenchimento

id	key	superfit_dis	superfit_sin	superfit_cur	superfit_int	superfit_eng	superfit_res
1018194	SF_DR	60.00	50.00	48.00	50.00	50.00	66.00
1018228	SF_DS	58.00	54.00	44.00	52.00	50.00	42.00
1018214	SF_DR	68.00	54.00	56.00	52.00	58.00	62.00
1018221	SF_DS	58.00	50.00	14.00	36.00	32.00	46.00
1016674	SF_DS	62.00	56.00	48.00	52.00	54.00	54.00

**Descrição:** Esta tabela armazena os resultados dos testes do tipo “genius” preenchido pelos candidatos que estão tentando alguma vaga dentro da plataforma Bettha.

**Tamanho:** 73.959 linhas (1 head + 73.958 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Numérica (float)	Identificador único utilizado para indicar o gestor.
key	Categórica (nominal)	Indicador utilizado para identificar as duas competências mais presentes no candidato. Formado por “SF” + as duas iniciais de cada competência.
superfit_dis	Numérica (float)	Score do candidato ao avaliar a competência “resiliente”
superfit_sin	Numérica (float)	Score do candidato ao avaliar a competência “engajado”
superfit_cur	Numérica (float)	Score do candidato ao avaliar a competência “colaborativo”
superfit_int	Numérica (float)	Score do candidato ao avaliar a competência “original”
superfit_eng	Numérica (float)	Score do candidato ao avaliar a competência “pragmático”
superfit_res	Numérica (float)	Score do candidato ao avaliar a competência “consistente”

## Análise matemática das colunas

Nome	Valor mínimo	Valor máximo	Moda	Média aritmética	Desvio padrão	Amplitude	Percentil 25	Percentil 50	Percentil 75
superfit_dis	0.0	90.0	58.0	57.326681	9.131010	90.0	52.0	58.0	64.0
superfit_sin	0.0	90.0	50.0	49.855569	8.213067	90.0	44.0	50.0	56.0
superfit_cur	0.0	94.0	44.0	44.041591	10.283265	94.0	38.0	44.0	50.0
superfit_int	0.0	92.0	54.0	52.177369	9.636396	92.0	46.0	52.0	58.0
superfit_eng	2.0	90.0	50.0	48.233643	10.148993	88.0	42.0	50.0	56.0
superfit_res	4.0	88.0	54.0	51.197664	9.477999	84.0	46.0	52.0	58.0

**Observações:** Coluna “id” não foi utilizada nesta análise pois se caracteriza como um identificador. Além disso, a coluna “key” é categórica. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

**Transformação:** Considerando o teor categórico da coluna “key”, foi necessário transformar esses dados em dados numéricos. A correlação entre o valor original e o novo valor pode ser observada abaixo:

Valor original	Novo valor	Valor original	Novo valor	Valor original	Novo valor
'SF_DR'	1	'SF_IR'	6	'SF_ER'	11
'SF_DS'	2	'SF_SI'	7	'SF_CI'	12
'SF_DI'	3	'SF_DC'	8	'SF_SE'	13
'SF_SC'	4	'SF_IE'	9	'SF_CE'	14
'SF_SR'	5	'SF_DE'	10	'SF_CR'	15

## Exemplo de preenchimento após alterações

id	key	superfit_dis	superfit_sin	superfit_cur	superfit_int	superfit_eng	superfit_res
1018194	1	60.00	50.00	48.00	50.00	50.00	66.00
1018228	2	58.00	54.00	44.00	52.00	50.00	42.00
1018214	1	68.00	54.00	56.00	52.00	58.00	62.00
1018221	2	58.00	50.00	14.00	36.00	32.00	46.00
1016674	2	62.00	56.00	48.00	52.00	54.00	54.00

## Tabela prep\_df\_users\_contratados

### Exemplo de preenchimento

user_id
135908
215549
628688
639513
628688

**Descrição:** Esta tabela armazena os IDs dos usuários que foram contratados para alguma vaga oferecida dentro da plataforma Bettha.

**Tamanho:** 256 linhas (1 head + 255 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
user_id	Numérica (float)	Identificador único utilizado para indicar algum candidato.

**Observações:** Coluna “id” não foi utilizada nesta análise pois se caracteriza como um identificador. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

## Tabela prep\_df\_users\_tests

### Exemplo de preenchimento

id	ingles_conceito	ingles_nota	excel_conceito	excel_nota	solu_conceito	solu_nota
1018194	0	0	0	0	Pré-Eficiente	3.20
1018228	C1	9.244	0	0	Eficiente	6.80
1018214	0	0	0	0	Eficiente	5.90
1018221	A1	0.64	Básico	2.0	Iniciante	1.50
1016674	A1	824	Intermediário	9.00	Iniciante	0.70

**Descrição:** Esta tabela armazena os resultados dos testes de inglês, excel e solu (lógica) preenchido pelos candidatos que estão tentando alguma vaga dentro da plataforma Bettha.

**Tamanho:** 72.573 linhas (1 head + 72.572 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Numérica (float)	Identificador único utilizado para indicar o candidato.
ingles_conceito	Categórica (ordinal)	Conceito utilizado para identificar a proficiência do candidato em inglês com base no Quadro Europeu Comum de Referência para Línguas.
ingles_nota	Numérica (float)	Score do candidato ao avaliar a competência “íngles”
excel_conceito	Categórica (ordinal)	Conceito utilizado para indicar o nível de habilidade do candidato no Microsoft Excel.
excel_nota	Numérica (float)	Score do candidato ao avaliar a competência “excel”
solu_conceito	Categórica (ordinal)	Conceito utilizado para indicar o nível de habilidade do candidato em resolver problemas que envolvem raciocínio lógico.
solu_nota	Numérica (float)	Score do candidato ao avaliar a competência “lógica”

## Análise matemática das colunas

Nome	Valor mínimo	Valor máximo	Moda	Média aritmética	Desvio padrão	Amplitude	Percentil 25	Percentil 50	Percentil 75
ingles_nota	0.0	9.56	0.0	0.791519	2.235348	9.56	0.0	0.0	0.0
excel_nota	0.0	14.0	0.0	0.383049	1.551370	14.0	0.0	0.0	0.0
solu_nota	0.0	10.0	0.0	1.331855	2.444841	10.0	0.0	0.0	0.0

**Observações:** Coluna “id” não foi utilizada nesta análise pois se caracteriza como um identificador. Além disso, as colunas “ingles\_conceito”, “excel\_conceito” e “solu\_conceito” são categóricas. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

**Transformação:** Considerando o teor categórico da coluna “key”, foi necessário transformar esses dados em dados numéricos. A correlação entre o valor original e o novo valor pode ser observada abaixo:

Valor original	Novo valor	Valor original	Novo valor	Valor original	Novo valor
'A1'	1	'Básico'	1	Eficiente	3
'A2'	2	'Intermediário'	2	Proficiente	4
'B1'	3	'Avançado'	3		
'B2'	4	'Iniciante'	1		
'C1'	5	Pré-Eficiente	2		

## Exemplo de preenchimento após alterações

id	ingles_conceito	ingles_nota	excel_conceito	excel_nota	solu_conceito	solu_nota
1018194	0	0	0	0	2	3.20
1018228	5	9.244	0	0	3	6.80
1018214	0	0	0	0	3	5.90
1018221	1	0.64	1	2.0	1	1.50
1016674	1	824	2	9.00	1	0.70

## Tabela prep\_df\_user\_objectives

### Exemplo de preenchimento

id	name
1022022	Procurar emprego
379020	Procurar emprego
378020	Melhorar minha carreira
1022022	Autoconhecimento
1022022	Desenvolvimento Pessoal

**Descrição:** Esta tabela armazena os objetivos dos usuários ao ingressar dentro da plataforma Bettha.

**Tamanho:** 48.883 linhas (1 head + 48.882 linhas de dados).

### Descrição das colunas

Nome	Tipo	Descrição
id	Númerica (float)	Identificador único utilizado para indicar algum candidato.
name	Catégorica (nominal)	Indica qual o objetivo do usuário ao ingressar na plataforma Bettha.

**Observações:** Coluna “id” não foi utilizada nesta análise pois se caracteriza como um identificador. Além disso, “name” é uma feature catégorica. Dessa forma, não existe sentido em fazer uma análise descritiva desses dados.

**Transformação:** Considerando o teor catégorico da coluna “name”, foi necessário transformar esses dados em dados numéricos. A correlação entre o valor original e o novo valor pode ser observada abaixo:

Valor original	Novo valor
'Procurar emprego'	1
'Melhorar minha carreira'	2

'Autoconhecimento'	3
'Desenvolvimento Pessoal'	4

### Exemplo de preenchimento após alterações

id	name
1022022	1
379020	1
378020	2
1022022	3
1022022	4