

```
In [155... import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
In [156... df_novos_alunos = pd.read_csv('C://Users//HENRIQUE//Documents//processo_seletivo_esc
```

```
In [158... df_novos_alunos.head()
```

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
0	xxx	não	-	871	93,05	-
1	xxx	não	-	498	66,56	2.001
2	xxx	não	-	974	95,50	3.001
3	xxx	não	-	790	88,20	acima de 10.000
4	xxx	não	-	816	93,05	-

```
In [160... df_novos_alunos.dtypes
```

```
Out[160... Possível aluno          object
Deve? (Sim ou Não)      object
Sim? Quanto?           object
Score                  object
Chance de Pagamento (%) object
Valor Mínimo de Salário object
dtype: object
```

REALIZANDO A LIMPEZA E CONVERSÃO DOS DADOS

Coluna 'Deve? (Sim ou Não)'

```
In [164... df_novos_alunos['Deve? (Sim ou Não)'].unique()
```

```
Out[164... array(['não', 'sim', 'não ', 'Não', 'Sim', '2,99'], dtype=object)
```

```
In [165... df_novos_alunos[df_novos_alunos['Deve? (Sim ou Não)'] == '2,99']
```

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
404	xxx	2,99	-	841	93,05	-

```
In [166... df_novos_alunos['Deve? (Sim ou Não)'] = df_novos_alunos['Deve? (Sim ou Não)'].map({'sim': 1, 'não': 0})
```

```
In [167... df_novos_alunos['Deve? (Sim ou Não)'] = df_novos_alunos['Deve? (Sim ou Não)'].fillna(0)
```

```
In [168... df_novos_alunos.describe()
```

Out[168...]

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
count	494	494	491	494	494	493
unique	1	3	39	291	39	19
top	xxx	não	-	974	95,50	-
freq	494	355	449	8	119	135

In [169...]

Precisamos classificar Deve? (Sim ou não) como SIM

In [170...]

df_novos_alunos[(df_novos_alunos['Deve? (Sim ou Não)'] == '-') & (df_novos_alunos['S

Out[170...]

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
295	xxx	-	18448.1	251	0.1787	7001
297	xxx	-	936.97	140	0.0299	7001
305	xxx	-	26889.86	105	0.0299	acima de 10.000
311	xxx	-	6894.79	96	0.0086	1001
320	xxx	-	5018.66	290	0.1787	1001
321	xxx	-	121.5	366	0.4969	3001
356	xxx	-	5545.91	81	0.86	1001
376	xxx	-	4985.26	32	0.86	2001
384	xxx	-	216.66	225	17.87	2001

In [171...]

```
df_novos_alunos['Deve? (Sim ou Não)'].iloc[295] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[297] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[305] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[311] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[320] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[321] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[356] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[376] = 'sim'
df_novos_alunos['Deve? (Sim ou Não)'].iloc[384] = 'sim'
```

Coluna 'Sim? Quanto?'

In [176...]

df_novos_alunos['Sim? Quanto?'].unique()

Out[176...]

```
array(['-', '40.878,87', '51.876,99', nan, '2.081,6', '885,75',
       'R$ 2.145,14', 'R$ 3.057,49', '6.059,83', '22.407,42', '5.298,24',
       'R$ 44.929,79', 'R$ 68,17', 'R$ 691,15', 'R$ 200,78',
       'R$ 5.890,96', 'R$ 310,72', 'R$ 2.839,16', 'R$ 575,95',
       'R$ 5.134,22', 'R$ 2.917,06', '8.236,69', 'R$ 1.029,67',
       'R$ 59.160,21', 'R$ 10.845,69', '18448.1', '936.97', '26889.86',
       '285,10\t', '6894.79', '5018.66', '121.5', '5545.91', '4985.26',
       '216.66', 'R$ 3.342,17', 'R$ 6.414,28', 'R$ 5.803,97', 'R$ 158,08',
       'R$ 927,80'], dtype=object)
```

In [177...]

```
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace('.', '')
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace('\t', '')
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace(' ', '')
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace('R', ''')
```

```
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace('$', '')
df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace(',', '')
df_novos_alunos['Sim? Quanto?'].unique()
```

```
Out[177... array(['-', '40878.87', '51876.99', 'nan', '2081.6', '885.75', '2145.14',
       '3057.49', '6059.83', '22407.42', '5298.24', '44929.79', '68.17',
       '691.15', '200.78', '5890.96', '310.72', '2839.16', '575.95',
       '5134.22', '2917.06', '8236.69', '1029.67', '59160.21', '10845.69',
       '184481', '93697', '2688986', '285.10', '689479', '501866', '1215',
       '554591', '498526', '21666', '3342.17', '6414.28', '5803.97',
       '158.08', '927.80'], dtype=object)
```

```
In [178... df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].fillna('-')
```

```
In [179... df_novos_alunos['Sim? Quanto?'].unique()
```

```
Out[179... array(['-', '40878.87', '51876.99', '2081.6', '885.75', '2145.14',
       '3057.49', '6059.83', '22407.42', '5298.24', '44929.79', '68.17',
       '691.15', '200.78', '5890.96', '310.72', '2839.16', '575.95',
       '5134.22', '2917.06', '8236.69', '1029.67', '59160.21', '10845.69',
       '184481', '93697', '2688986', '285.10', '689479', '501866', '1215',
       '554591', '498526', '21666', '3342.17', '6414.28', '5803.97',
       '158.08', '927.80'], dtype=object)
```

Coluna 'Score'

```
In [181... df_novos_alunos['Score'].unique()
```

```
Out[181... array(['871', '498', '974', '790', '816', '922', '169', '966', '433',
       '314', '831', '51.876,99', '952', '349', '327', '240', '873',
       '395', '948', '898', '969', '858', '574', '683', '829', '540',
       '604', '596', '32', '408', '383', '999', '569', '768', '58', '515',
       '995', '928', '1000', '987', '362', '963', '621', '973', '552',
       '810', '855', '454', '406', '847', '917', '971', '937', '959',
       '524', '976', '895', '196', '945', '460', '691', '252', '997',
       '251', '840', '216', '641', '970', '590', '418', '752', '890',
       '84', '972', '437', '629', '587', '686', '913', '815', '492',
       '788', '513', '813', '522', '16', '817', '939', '602', '865',
       '967', '841', '882', '809', '440', '781', '984', '993', '951',
       '991', '435', '626', '516', '676', '588', '405', '909', '784',
       '571', '793', '455', '576', '494', '371', '468', '335', '328',
       '885', '965', '860', '981', '441', '62', '730', '394', '958',
       '448', '256', '388', '961', '664', '953', '617', '384', '741',
       '354', '986', '777', '489', '929', '964', '502', '954', '906',
       '476', '348', '648', '533', '943', '830', '915', '484', '253',
       '955', '478', '962', '850', '979', '868', '710', '404', '541',
       '763', '926', '941', '554', '806', '531', '935', '438', '960',
       '914', '826', '467', '417', '523', '181', '932', '727', '234',
       '760', '54', '879', '536', '886', '862', '985', '702', '933',
       '488', '834', '983', '757', '228', '545', '150', '762', '409',
       '428', '990', '968', '377', '892', '573', '443', '140', '364',
       '654', '105', '918', '717', '96', '363', '857', '290', '366',
       '612', '877', '848', '439', '712', '820', '714', '902', '908',
       '624', '980', '837', '705', '743', '81', '771', '669', '988',
       '360', '661', '720', '359', '426', '863', '735', '911', '225',
       '800', '920', '910', '947', '876', '905', '166', '403', '430',
       '379', '544', '859', '639', '372', '782', '693', '875', '385',
       '568', '491', '843', '884', '989', '709', '566', '707', '924',
       '946', '656', '754', '260', '429', '636', '69', '326', '378',
       '434', '310', '304', '802', '353', '558', '485', '382', '452',
       '120', '411', '662'], dtype=object)
```

```
In [182... valores_estranhos = df_novos_alunos[df_novos_alunos['Score'] == '51.876,99']
valores_estranhos
```

```
Out[182...
```

Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
12	xxx	não	- 51.876,99	95,50	7.001

```
In [183... df_novos_alunos = df_novos_alunos[df_novos_alunos['Score'] != '51.876,99']
```

Coluna 'Chance de pagamento'

```
In [187... df_novos_alunos['Chance de Pagamento (%)'].unique()
```

```
Out[187... array(['93,05', '66,56', '95,50', '88,20', '2,99', '49,69', '49,49',
       '17,87', '77,51', '83,12', '0,86', '66,65', '95,50', '83,12',
       '88,20', '0.4969', '0.9305', '0.955', '95,50%', '0.7751',
       '0.1787', '0.6656', '0.0299', '0.8312', '17,87%', '95.5', '88.2',
       '0.0086', '77,51%', '0.882', '88,20%', '0.86', '83.12', '77.51',
       '49.69', '66.56', '93.05', '17.87', '46,69'], dtype=object)
```

```
In [188... valores_estranhos = pd.concat([valores_estranhos, df_novos_alunos[df_novos_alunos['Chance de Pagamento (%)'] == '95.5%']])
```

Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
12	xxx	não	- 51.876,99	95,50	7.001
311	xxx	sim	689479	96	0.0086

```
In [189... df_novos_alunos['Chance de Pagamento (%)']
```

```
Out[189... 0    93,05
1    66,56
2    95,50
3    88,20
4    93,05
...
489   2,99
490   66,56
491   49,69
492   83,12
493   95,50
Name: Chance de Pagamento (%), Length: 493, dtype: object
```

```
In [190... df_novos_alunos['Chance de Pagamento (%)'] = df_novos_alunos['Chance de Pagamento (%)'].str.replace(',', '.')
df_novos_alunos['Chance de Pagamento (%)'] = df_novos_alunos['Chance de Pagamento (%)'].str.replace('95.5%', '95,50')
df_novos_alunos['Chance de Pagamento (%)'] = df_novos_alunos['Chance de Pagamento (%)'].str.replace('95.5', '95,50')
df_novos_alunos['Chance de Pagamento (%)'].unique()
```

```
Out[190... array(['93.05', '66.56', '95.50', '88.20', '2.99', '49.69', '49.49',
       '17.87', '77.51', '83.12', '0.86', '66.65', '0.4969', '0.9305',
       '0.955', '0.7751', '0.1787', '0.6656', '0.0299', '0.8312', '95.5',
       '88.2', '0.0086', '0.882', '46.69'], dtype=object)
```

Coluna 'Valor Mínimo de salário'

```
In [196... df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[196... array(['-', '2.001', '3.001', 'acima de 10.000', '5.001', '4.001',
       '1.001', '7.001', 'acima de 10.000', 'acima 10.000', '3001',
       '10000', '4001', '2001', '7001', '1001', '5001', '70001', nan],
       dtype=object)
```

```
In [197... df_novos_alunos['Valor Mínimo de Salário'] = df_novos_alunos['Valor Mínimo de Salário']
df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[197... array(['-', '2.001', '3.001', 'acima de 10.000', '5.001', '4.001',
   '1.001', '7.001', 'acima de 10.000 ', 'acima 10.000', ' - ', '3001',
   '10000', '4001', '2001', '7001', '1001', '5001', '70001'],
  dtype=object)
```

```
In [198... df_novos_alunos['Valor Mínimo de Salário'] = df_novos_alunos['Valor Mínimo de Salário']
df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[198... array(['-', '2.001', '3.001', 'acima de 10.000', '5.001', '4.001',
   '1.001', '7.001', 'acima de 10.000 ', 'acima 10.000', '3001',
   '10000', '4001', '2001', '7001', '1001', '5001', '70001'],
  dtype=object)
```

```
In [199... df_novos_alunos['Valor Mínimo de Salário'] = df_novos_alunos['Valor Mínimo de Salário']
df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[199... array(['-', '2.001', '3.001', 'acimade10.000', '5.001', '4.001', '1.001',
   '7.001', 'acima10.000', '3001', '10000', '4001', '2001', '7001',
   '1001', '5001', '70001'], dtype=object)
```

```
In [200... df_novos_alunos['Valor Mínimo de Salário'] = df_novos_alunos['Valor Mínimo de Salário']
df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[200... array(['-', '2001', '3001', 'acimade10000', '5001', '4001', '1001',
   '7001', 'acima10000', '10000', '70001'], dtype=object)
```

```
In [201... df_novos_alunos[df_novos_alunos['Valor Mínimo de Salário'] == 'acima10000']
```

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
190	xxx	não	-	348	49.69	acima10000

```
In [202... df_novos_alunos.iloc[189, 5]
```

```
Out[202... 'acima10000'
```

```
In [203... df_novos_alunos.iloc[189, 5] = 'acimade10000'
```

```
In [204... df_novos_alunos['Valor Mínimo de Salário'].unique()
```

```
Out[204... array(['-', '2001', '3001', 'acimade10000', '5001', '4001', '1001',
   '7001', '10000', '70001'], dtype=object)
```

Vamos verificarmos a tipagem dos dados

```
In [205... valores_estranhos
```

	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
12	xxx	não	-	51.876,99	95,50	7.001
311	xxx	sim	689479	96	0.0086	1001

```
In [207... df_novos_alunos.dtypes
```

```
Out[207... Possível aluno          object
```

```
Deve? (Sim ou Não)      object
Sim? Quanto?            object
Score                   object
Chance de Pagamento (%) object
Valor Mínimo de Salário object
dtype: object
```

VAMOS PRIMEIRAMENTE ISOLAR OS DADOS QUE NÃO POSSUEM INFORMAÇÕES DE SE DEVE OU NÃO, OU NÃO POSSUI VALOR DEVEDOR

```
In [208... pagamento_desconhecido = df_novos_alunos[df_novos_alunos['Deve? (Sim ou Não)'] == '-']
pagamento_desconhecido
```

Out[208...]	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
285	xxx	-	-	377	0.4969	3001
286	xxx	-	-	862	0.9305	10000
287	xxx	-	-	862	0.9305	10000
288	xxx	-	-	892	0.9305	4001
289	xxx	-	-	966	0.955	10000
...
394	xxx	-	-	876	93.05	10000
395	xxx	-	-	963	95.5	10000
396	xxx	-	-	596	77.51	3001
397	xxx	-	-	971	95.5	5001
404	xxx	-	-	841	93.05	-

96 rows × 6 columns

```
In [209... df_novos_alunos = df_novos_alunos[df_novos_alunos['Deve? (Sim ou Não)'] != '-']
```

```
In [210... df_novos_alunos['Deve? (Sim ou Não)'].unique()
```

```
Out[210... array(['não', 'sim'], dtype=object)
```

```
In [211... pagamento_desconhecido = pd.concat([pagamento_desconhecido, df_novos_alunos[(df_novo
```

```
In [212... df_novos_alunos[(df_novos_alunos['Deve? (Sim ou Não)'] == 'sim') & (df_novos_alunos[
```

Out[212...]	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
17	xxx	sim	-	240	17.87	acimade10000

```
In [213... df_novos_alunos = df_novos_alunos.drop([17])
```

VAMOS AGORA TRANSFORMAR ESSES NÚMEROS PARA VALORES INTEIROS

```
In [214... # Primeiro precisamos transformar esses valores em '-' como valor igual a 0
df_novos_alunos[df_novos_alunos['Sim? Quanto?'] == '-']
```

```
Out[214...]
```

Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
0	xxx	não	-	871	93.05
1	xxx	não	-	498	66.56
2	xxx	não	-	974	95.50
3	xxx	não	-	790	88.20
4	xxx	não	-	816	93.05
...
488	xxx	não	-	452	66.56
490	xxx	não	-	411	66.56
491	xxx	não	-	360	49.69
492	xxx	não	-	662	83.12
493	xxx	não	-	979	95.50

354 rows × 6 columns

```
In [215...]: df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].str.replace('-', '0')
In [216...]: df_novos_alunos['Sim? Quanto?'] = df_novos_alunos['Sim? Quanto?'].astype('float64')
```

VAMOS AGORA TRANSFORMAR O SCORE PARA VALORES INTEIROS

```
In [217...]: df_novos_alunos['Score'] = df_novos_alunos['Score'].astype('float64')
```

VAMOS AGORA TRANSFORMAR O CHANCE DE PAGAMENTO PARA VALORES INTEIROS

```
In [218...]: df_novos_alunos['Chance de Pagamento (%)'] = df_novos_alunos['Chance de Pagamento (%)'].astype('int')
```

FILTRO DOS ALUNOS

```
In [219...]: negados = df_novos_alunos[(df_novos_alunos['Score'] <= 246.86) | (df_novos_alunos['Chance de Pagamento (%)'] <= 17.87)]
negados
```

Out[219...]	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
6	xxx	sim	40878.87	169.0	2.9900	4001
16	xxx	sim	51876.99	240.0	17.8700	acimade10000
18	xxx	sim	51876.99	240.0	17.8700	acimade10000
31	xxx	sim	2081.60	32.0	0.8600	2001
37	xxx	sim	885.75	58.0	0.8600	1001
62	xxx	sim	2145.14	196.0	2.9900	3001
63	xxx	sim	2145.14	196.0	2.9900	3001
69	xxx	sim	3057.49	252.0	17.8700	3001

Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
71	xxx	não	0.00	251.0	17.8700 -
74	xxx	sim	6059.83	216.0	17.8700 acimade10000
82	xxx	sim	22407.42	84.0	0.8600 3001
98	xxx	sim	5298.24	16.0	0.8600 5001
120	xxx	sim	6059.83	216.0	17.8700 acimade10000
121	xxx	sim	44929.79	58.0	0.8600 acimade10000
156	xxx	sim	200.78	62.0	0.8600 1001
167	xxx	sim	310.72	256.0	17.8700 2001
175	xxx	sim	200.78	62.0	0.8600 1001
200	xxx	sim	2839.16	253.0	17.8700 2001
203	xxx	sim	575.95	196.0	2.9900 5001
238	xxx	sim	5134.22	181.0	2.9900 -
244	xxx	sim	2917.06	234.0	17.8700 1001
246	xxx	sim	8236.69	54.0	0.8600 3001
269	xxx	sim	1029.67	228.0	17.8700 2001
273	xxx	sim	59160.21	150.0	2.9900 acimade10000
278	xxx	sim	10845.69	181.0	2.9900 3001
295	xxx	sim	184481.00	251.0	0.1787 7001
297	xxx	sim	93697.00	140.0	0.0299 7001
305	xxx	sim	2688986.00	105.0	0.0299 acimade10000
307	xxx	sim	285.10	228.0	17.8700 2001
311	xxx	sim	689479.00	96.0	0.0086 1001
320	xxx	sim	501866.00	290.0	0.1787 1001
321	xxx	sim	1215.00	366.0	0.4969 3001
356	xxx	sim	554591.00	81.0	0.8600 1001
376	xxx	sim	498526.00	32.0	0.8600 2001
384	xxx	sim	21666.00	225.0	17.8700 2001
399	xxx	não	0.00	166.0	2.9900 -
401	xxx	sim	3342.17	96.0	0.8600 -
423	xxx	sim	6414.28	81.0	0.8600 -
453	xxx	não	0.00	260.0	17.8700 -
458	xxx	sim	5803.97	253.0	17.8700 -
465	xxx	sim	158.08	69.0	0.8600 -
489	xxx	sim	927.80	120.0	2.9900 -

In [221]: df_novos_alunos = df_novos_alunos[(df_novos_alunos['Score'] > 246.86) & (df_novos_al

df_novos_alunos

Out[221...]	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
0	xxx	não	0.0	871.0	93.05	-
1	xxx	não	0.0	498.0	66.56	2001
2	xxx	não	0.0	974.0	95.50	3001
3	xxx	não	0.0	790.0	88.20	acimade10000
4	xxx	não	0.0	816.0	93.05	-
...
488	xxx	não	0.0	452.0	66.56	-
490	xxx	não	0.0	411.0	66.56	-
491	xxx	não	0.0	360.0	49.69	-
492	xxx	não	0.0	662.0	83.12	-
493	xxx	não	0.0	979.0	95.50	-

354 rows × 6 columns

In [222...]: valores_estranhos

Out[222...]	Possível aluno	Deve? (Sim ou Não)	Sim? Quanto?	Score	Chance de Pagamento (%)	Valor Mínimo de Salário
12	xxx	não	-	51.876,99	95,50	7.001
311	xxx	sim	689479	96	0.0086	1001

In [228...]: negados.to_csv('Alunos_recusados.csv')

In [232...]: valores_estranhos.to_csv('Valores_revisados.csv')

In [231...]: df_novos_alunos.to_csv('Alunos_aprovados.csv')

In []: