



Ciência de Dados Aplicada | IMPA 2019

Seleção de características para predição de notas no ENEM 2018

Kizzy Terra





— objetivos

os microdados do ENEM

seleção de características

visualização de características

conclusão





objetivos

Identificar as características mais relevantes

Explorar as características identificadas





objetivos

Identificar as características mais relevantes

Explorar as características identificadas





objetivos

— **os microdados do ENEM**

seleção de características

visualização de características

conclusão



os microdados do ENEM 2018

5.513.747 estudantes

137 colunas

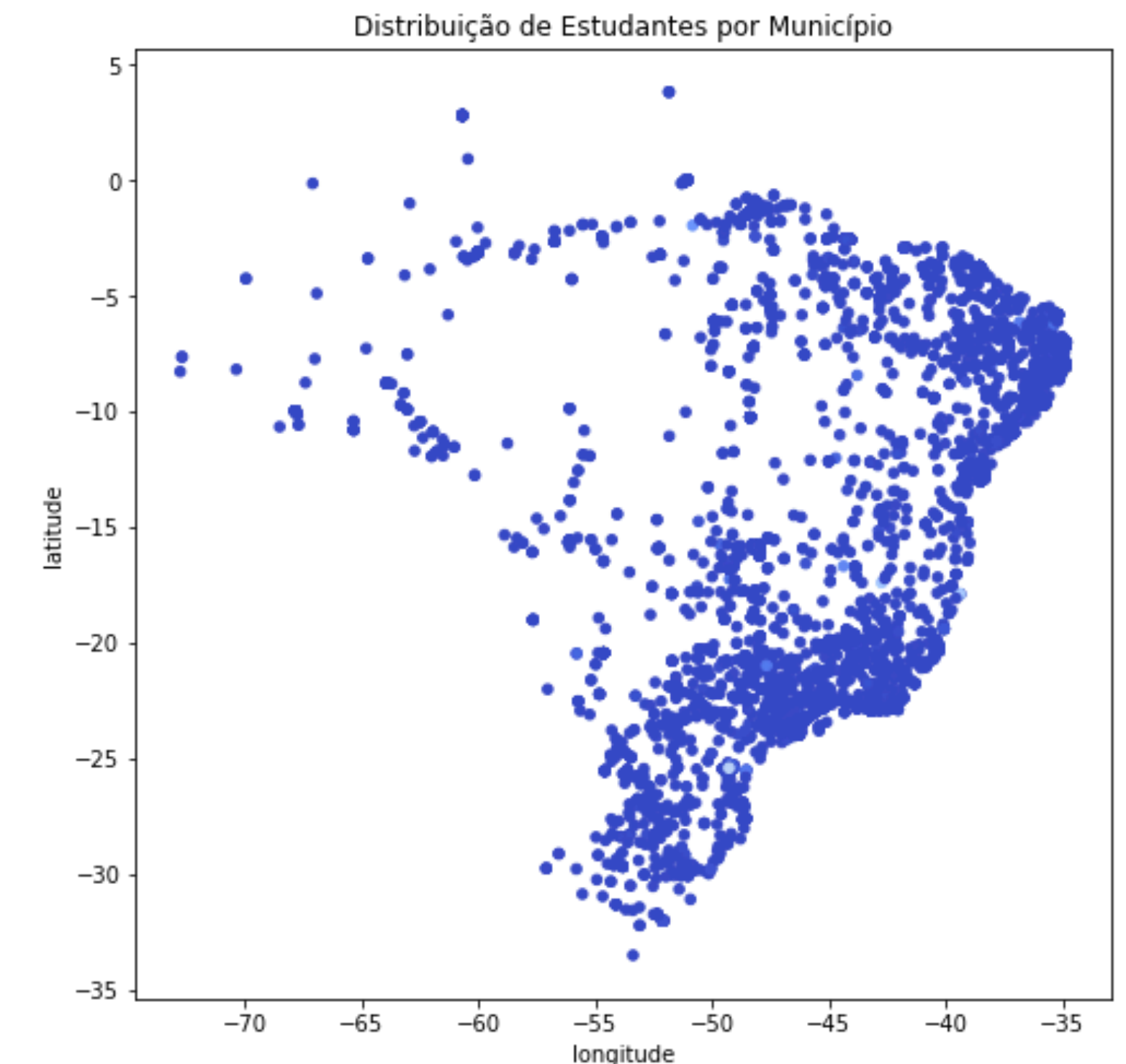
dados do estudante

dados da escola

dados de atendimento especializado

dados do local de prova

dados das provas objetivas e redação



Tamanho do arquivo: 3.4 GB

Fonte dos microdados: INEP

objetivos

os microdados do ENEM

— **seleção de características**

visualização de características

conclusão

seleção de características

```
microdados_enem.columns.values
```

```
array(['NU_INSCRICAO', 'NU_ANO', 'CO_MUNICIPIO_RESIDENCIA',  
      'NO_MUNICIPIO_RESIDENCIA', 'CO_UF_RESIDENCIA', 'SG_UF_RESIDENCIA',  
      'NU_IDADE', 'TP_SEXO', 'TP_ESTADO_CIVIL', 'TP_COR_RACA',  
      'TP_NACIONALIDADE', 'CO_MUNICIPIO_NASCIMENTO',  
      'NO_MUNICIPIO_NASCIMENTO', 'CO_UF_NASCIMENTO', 'SG_UF_NASCIMENTO',  
      'TP_ST_CONCLUSAO', 'TP_ANO_CONCLUIU', 'TP_ESCOLA', 'TP_ENSINO',  
      'IN_TREINEIRO', 'CO_ESCOLA', 'CO_MUNICIPIO_ESC',  
      'NO_MUNICIPIO_ESC', 'CO_UF_ESC', 'SG_UF_ESC',  
      'TP_DEPENDENCIA_ADM_ESC', 'TP_LOCALIZACAO_ESC', 'TP_SIT_FUNC_ESC',  
      'IN_BAIXA_VISAO', 'IN_CEGUEIRA', 'IN_SURDEZ',  
      'IN_DEFICIENCIA_AUDITIVA', 'IN_SURDO_CEGUEIRA',  
      'IN_DEFICIENCIA_FISICA', 'IN_DEFICIENCIA_MENTAL',  
      'IN_DEFICIT_ATENCAO', 'IN_DISLEXIA', 'IN_DISCALCULIA',  
      'IN_AUTISMO', 'IN_VISAO_MONOCULAR', 'IN_OUTRA_DEF', 'IN_GESTANTE',  
      'IN_LACTANTE', 'IN_IDOSO', 'IN_ESTUDA_CLASSE_HOSPITALAR',  
      'IN_SEM_RECURSO', 'IN_BRILLE', 'IN_AMPLIADA_24', 'IN_AMPLIADA_18',  
      'IN_LEDOR', 'IN_ACESSO', 'IN_TRANSCRICAO', 'IN_LIBRAS',  
      'IN_LEITURA_LABIAL', 'IN_MESA_CADEIRA_RODAS',  
      'IN_MESA_CADEIRA_SEPARADA', 'IN_APOIO_PERNA', 'IN_GUIA_INTERPRETE',  
      'IN_COMPUTADOR', 'IN_CADEIRA_ESPECIAL', 'IN_CADEIRA_CANHOTO',  
      'IN_CADEIRA_ACOLCHOADA', 'IN_PROVA_DEITADO', 'IN_MOBILIARIO_OBESO',  
      'IN_LAMINA_OVERLAY', 'IN_PROTETOR_AURICULAR', 'IN_MEDIDOR_GLICOSE',  
      'IN_MAQUINA_BRAILE', 'IN_SOROBAN', 'IN_MARCA_PASSO', 'IN_SONDA',  
      'IN_MEDICAMENTOS', 'IN_SALA_INDIVIDUAL', 'IN_SALA_ESPECIAL',  
      'IN_SALA_ACOMPANHANTE', 'IN_MOBILIARIO_ESPECIFICO',  
      'IN_MATERIAL_ESPECIFICO', 'IN_NOME_SOCIAL', 'CO_MUNICIPIO_PROVA',  
      'NO_MUNICIPIO_PROVA', 'CO_UF_PROVA', 'SG_UF_PROVA',  
      'TP_PRESENCA_CN', 'TP_PRESENCA_CH', 'TP_PRESENCA_LC',  
      'TP_PRESENCA_MT', 'CO_PROVA_CN', 'CO_PROVA_CH', 'CO_PROVA_LC',  
      'CO_PROVA_MT', 'NU_NOTA_CN', 'NU_NOTA_CH', 'NU_NOTA_LC',  
      'NU_NOTA_MT', 'TX_RESPOSTAS_CN', 'TX_RESPOSTAS_CH',  
      'TX_RESPOSTAS_LC', 'TX_RESPOSTAS_MT', 'TP_LINGUA',  
      'TX_GABARITO_CN', 'TX_GABARITO_CH', 'TX_GABARITO_LC',  
      'TX_GABARITO_MT', 'TP_STATUS_REDACAO', 'NU_NOTA_COMP1',  
      'NU_NOTA_COMP2', 'NU_NOTA_COMP3', 'NU_NOTA_COMP4', 'NU_NOTA_COMP5',  
      'NU_NOTA_REDACAO', 'Q001', 'Q002', 'Q003', 'Q004', 'Q005', 'Q006',  
      'Q007', 'Q008', 'Q009', 'Q010', 'Q011', 'Q012', 'Q013', 'Q014',  
      'Q015', 'Q016', 'Q017', 'Q018', 'Q019', 'Q020', 'Q021', 'Q022',  
      'Q023', 'Q024', 'Q025', 'Q026', 'Q027'], dtype=object)
```


|

seleção de características

remoção de características

remoção de dados ausentes

seleção univariada

seleção a partir de modelo

análise de componentes PCA

|



remoção de características

características com baixa variância

características redundantes



remoção baixa variância

```
['IN_BAIXA_VISAO', 'IN_CEGUEIRA', 'IN_SURDEZ',  
'IN_DEFICIENCIA_AUDITIVA', 'IN_SURDO_CEGUEIRA',  
'IN_DEFICIENCIA_FISICA', 'IN_DEFICIENCIA_MENTAL',  
'IN_DEFICIT_ATENCAO', 'IN_DISLEXIA', 'IN_DISCALCULIA',  
'IN_AUTISMO', 'IN_VISAO_MONOCULAR', 'IN_OUTRA_DEF', 'IN_GESTANTE',  
'IN_LACTANTE', 'IN_IDOSO', 'IN_ESTUDA_CLASSE_HOSPITALAR',  
'IN_SEM_RECURSO', 'IN_BRILLE', 'IN_AMPLIADA_24', 'IN_AMPLIADA_18',  
'IN_LEDOR', 'IN_ACESSO', 'IN_TRANSCRICAO', 'IN_LIBRAS',  
'IN_LEITURA_LABIAL', 'IN_MESA_CADEIRA_RODAS',  
'IN_MESA_CADEIRA_SEPARADA', 'IN_APOIO_PERNA', 'IN_GUIA_INTERPRETE',  
'IN_COMPUTADOR', 'IN_CADEIRA_ESPECIAL', 'IN_CADEIRA_CANHOTO',  
'IN_CADEIRA_ACOLCHOADA', 'IN_PROVA_DEITADO', 'IN_MOBILIARIO_OBESO',  
'IN_LAMINA_OVERLAY', 'IN_PROTETOR_AURICULAR', 'IN_MEDIDOR_GLIPOSE',  
'IN_MAQUINA_BRAILE', 'IN_SOROBAN', 'IN_MARCA_PASSO', 'IN_SONDA',  
'IN_MEDICAMENTOS', 'IN_SALA_INDIVIDUAL', 'IN_SALA_ESPECIAL',  
'IN_SALA_ACOMPANHANTE', 'IN_MOBILIARIO_ESPECIFICO',  
'IN_MATERIAL_ESPECIFICO', 'IN_NOME_SOCIAL']
```

min(var) = 0.0000001 | max(var) = 0.001723

IN_BAIXA_VISAO	0.000951
IN_CEGUEIRA	0.000137
IN_SURDEZ	0.000262
IN_DEFICIENCIA_AUDITIVA	0.000438
IN_SURDO_CEGUEIRA	0.000002
IN_DEFICIENCIA_FISICA	0.001617
IN_DEFICIENCIA_MENTAL	0.000354
IN_DEFICIT_ATENCAO	0.001304
IN_DISLEXIA	0.000257
IN_DISCALCULIA	0.000054
IN_AUTISMO	0.000140
IN_VISAO_MONOCULAR	0.000254
IN_OUTRA_DEF	0.000851
IN_GESTANTE	0.000323
IN_LACTANTE	0.000974
IN_IDOSO	0.000046
IN_ESTUDA_CLASSE_HOSPITALAR	0.000006
IN_SEM_RECURSO	0.001723
IN_BRILLE	0.000063
IN_AMPLIADA_24	0.000230
IN_AMPLIADA_18	0.000711
IN_LEDOR	0.001046
IN_ACESSO	0.001700
IN_TRANSCRICAO	0.000979
IN_LIBRAS	0.000200
IN_LEITURA_LABIAL	0.000124
IN_MESA_CADEIRA_RODAS	0.000343
IN_MESA_CADEIRA_SEPARADA	0.000352
IN_APOIO_PERNA	0.000379
IN_GUIA_INTERPRETE	0.000001
IN_COMPUTADOR	0.000001
IN_CADEIRA_ESPECIAL	0.000003
IN_CADEIRA_CANHOTO	0.000007
IN_CADEIRA_ACOLCHOADA	0.000028
IN_PROVA_DEITADO	0.000002
IN_MOBILIARIO_OBESO	0.000012
IN_LAMINA_OVERLAY	0.000003
IN_PROTETOR_AURICULAR	0.000001
IN_MEDIDOR_GLIPOSE	0.000082
IN_MAQUINA_BRAILE	0.000019
IN_SOROBAN	0.000011
IN_MARCA_PASSO	0.000002
IN_SONDA	0.000002
IN_MEDICAMENTOS	0.000036
IN_SALA_INDIVIDUAL	0.000052
IN_SALA_ESPECIAL	0.000039
IN_SALA_ACOMPANHANTE	0.000010
IN_MOBILIARIO_ESPECIFICO	0.000011
IN_MATERIAL_ESPECIFICO	0.000017
IN_NOME_SOCIAL	0.000046

dtype: float64

remoção de redundâncias

```
Index([ 'CO_MUNICIPIO_RESIDENCIA', 'NO_MUNICIPIO_RESIDENCIA',  
       'CO_UF_RESIDENCIA', 'SG_UF_RESIDENCIA', 'CO_MUNICIPIO_NASCIMENTO',  
       'NO_MUNICIPIO_NASCIMENTO', 'CO_UF_NASCIMENTO', 'SG_UF_NASCIMENTO',  
       'CO_MUNICIPIO_ESC', 'NO_MUNICIPIO_ESC', 'CO_UF_ESC', 'SG_UF_ESC',  
       'CO_MUNICIPIO_PROVA', 'NO_MUNICIPIO_PROVA', 'CO_UF_PROVA',  
       'SG_UF_PROVA'],  
      dtype='object')
```

|

seleção de características

remoção de características

remoção de dados ausentes

seleção univariada

seleção a partir de modelo

análise de componentes PCA

|

— remoção de dados ausentes —

```
microdados_enem.isna().sum()
```

NU_INSCRICAO	0	CO_MUNICIPIO_PROVA	0	Q001	0
NU_ANO	0	CO_UF_PROVA	0	Q002	0
CO_MUNICIPIO_RESIDENCIA	0	TP_PRESENCA_CN	155	Q003	0
CO_UF_RESIDENCIA	0	TP_PRESENCA_CH	155	Q004	0
NU_IDADE	97	TP_PRESENCA_LC	155	Q005	0
TP_SEXO	0	TP_PRESENCA_MT	155	Q006	0
TP_ESTADO_CIVIL	217637	CO_PROVA_CN	1608648	Q007	0
TP_COR_RACA	0	CO_PROVA_CH	1365483	Q008	0
TP_NACIONALIDADE	0	CO_PROVA_LC	1365483	Q009	0
CO_MUNICIPIO_NASCIMENTO	185783	CO_PROVA_MT	1608648	Q010	0
CO_UF_NASCIMENTO	185783	NU_NOTA_CN	1608648	Q011	0
TP_ST_CONCLUSAO	0	NU_NOTA_CH	1365483	Q012	0
TP_ANO_CONCLUIU	0	NU_NOTA_LC	1365483	Q013	0
TP_ESCOLA	0	NU_NOTA_MT	1608648	Q014	0
TP_ENSINO	2030666	TX_RESPOSTAS_CN	1608648	Q015	0
IN_TREINEIRO	0	TX_RESPOSTAS_CH	1365483	Q016	0
CO_ESCOLA	4064926	TX_RESPOSTAS_LC	1365483	Q017	0
CO_MUNICIPIO_ESC	4064926	TX_RESPOSTAS_MT	1608648	Q018	0
CO_UF_ESC	4064926	TP_LINGUA	0	Q019	0
TP_DEPENDENCIA_ADM_ESC	4064926	TX_GABARITO_CN	1608648	Q020	0
TP_LOCALIZACAO_ESC	4064926	TX_GABARITO_CH	1365483	Q021	0
TP_SIT_FUNC_ESC	4064926	TX_GABARITO_LC	1365483	Q022	0
dtype: int64		TX_GABARITO_MT	1608648	Q023	0
		TP_STATUS_REDACAO	1365483	Q024	0
		NU_NOTA_COMP1	1365483	Q025	0
		NU_NOTA_COMP2	1365483	Q026	4
		NU_NOTA_COMP3	1365483	Q027	0
		NU_NOTA_COMP4	1365483	dtype: int64	
		NU_NOTA_COMP5	1365483		
		NU_NOTA_REDACAO	1365483		
		dtype: int64			

estratégias de remoção de ausentes

exclusão das entradas

exclusão de características

adição de categoria

substituição por dado correlacionado

exclusão de entradas

- exclusão das linhas que não possuem os dados sobre as notas
- daqueles estudantes que não fizeram alguma das provas

CO_PROVA_CN	1608648
CO_PROVA_CH	1365483
CO_PROVA_LC	1365483
CO_PROVA_MT	1608648
NU_NOTA_CN	1608648
NU_NOTA_CH	1365483
NU_NOTA_LC	1365483
NU_NOTA_MT	1608648
TX_RESPOSTAS_CN	1608648
TX_RESPOSTAS_CH	1365483
TX_RESPOSTAS_LC	1365483
TX_RESPOSTAS_MT	1608648
TP_LINGUA	0
TX_GABARITO_CN	1608648
TX_GABARITO_CH	1365483
TX_GABARITO_LC	1365483
TX_GABARITO_MT	1608648
TP_STATUS_REDACAO	1365483
NU_NOTA_COMP1	1365483
NU_NOTA_COMP2	1365483
NU_NOTA_COMP3	1365483
NU_NOTA_COMP4	1365483
NU_NOTA_COMP5	1365483
NU_NOTA_REDACAO	1365483
dtype:	int64

- dados da escola não são de preenchimento obrigatório
- aproximadamente 80% das linhas não possuem esse dado

Missing values:

CO_ESCOLA	4064926
CO_MUNICIPIO_ESC	4064926
CO_UF_ESC	4064926
TP_DEPENDENCIA_ADM_ESC	4064926
TP_LOCALIZACAO_ESC	4064926
TP_SIT_FUNC_ESC	4064926
TP_ENSINO	2030666

- a coluna que indica o estado civil é uma variável categórica numérica

0	Solteiro(a)
1	Casado(a)/Mora com companheiro(a)
2	Divorciado(a)/Desquitado(a)/Separado(a)
3	Viúvo(a)

- adicionada uma nova categoria as 217.637 dados faltantes

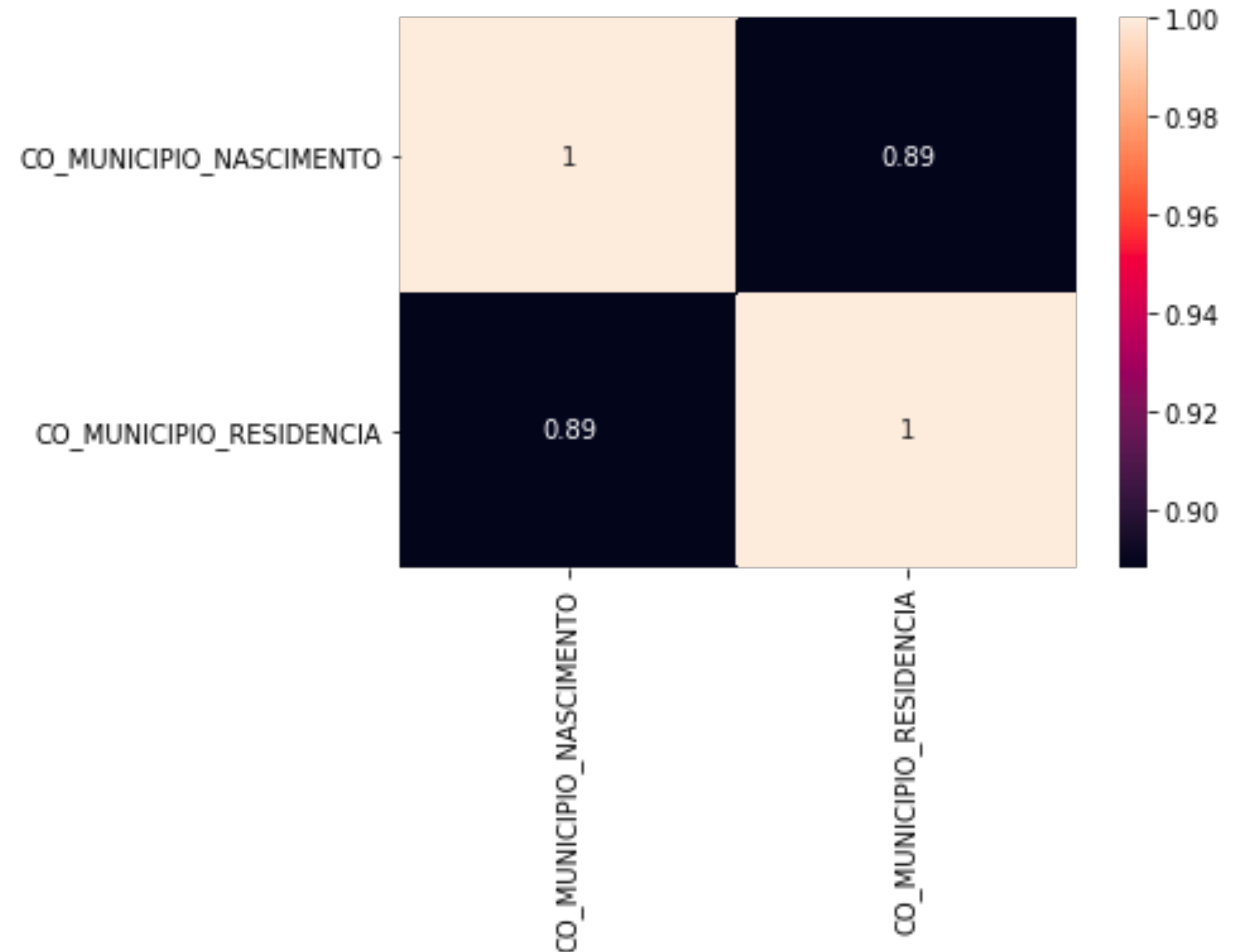
0	Solteiro(a)
1	Casado(a)/Mora com companheiro(a)
2	Divorciado(a)/Desquitado(a)/Separado(a)
3	Viúvo(a)
4	Não-informado

— substituição por dado correlacionado —

- valores ausentes preenchidos a partir das colunas correlacionadas

Missing values:

CO_MUNICIPIO_NASCIMENTO	185783
CO_UF_NASCIMENTO	185783



|

seleção de características

remoção de características

remoção de dados ausentes

seleção univariada

seleção a partir de modelo

análise de componentes PCA

|

Downloaded from <http://ajphaphysocpharm.sagepub.com/> at 11:01 11 November 2014

1000000

target

- criação de novas colunas com médias
+ **MEAN_GROUPS**

```
Index([ 'NU_NOTA_CN', 'NU_NOTA_CH', 'NU_NOTA_LC', 'NU_NOTA_MT',  
       'NU_NOTA_REDACAO', 'MEAN', 'DIR_UFRJ', 'ENG_UFRJ', 'MED_UFRJ',  
       'MED_UFRJ_AC_LABEL', 'MED_UFRJ_EP_LABEL', 'MED_USP', 'MED_USP_AC_LABEL',  
       'MED_USP_EP_LABEL', 'MED_USP_PPI_LABEL', 'MEAN_GROUPS'],  
      dtype='object')
```


target

```
med_weights_usp = {"CN": 2.0, "CH": 2.0, "LC": 4.0, "MT": 3.0, "REDACAO": 2.0, "TOTAL": 13.0}
med_cuts_usp = {"AC": 858.44, "EP": 793.29, "PPI": 767.15}
med_weights_ufrj = {"CN": 2.0, "CH": 1.0, "LC": 2.0, "MT": 1.0, "REDACAO": 3.0, "TOTAL": 9.0}
med_cuts_ufrj = {"AC": 822.46, "EP": 805.53}

eng_weights_ufrj = {"CN": 4.0, "CH": 2.0, "LC": 2.0, "MT": 4.0, "REDACAO": 3.0, "TOTAL": 15.0}
eng_cuts_ufrj = {"AC": 774.29, "EP": 664.61}

dir_weights_ufrj = {"CN": 1.0, "CH": 2.0, "LC": 2.0, "MT": 1.0, "REDACAO": 3.0, "TOTAL": 8.0}
dir_cuts_ufrj = {"AC": 789.17, "EP": 771.61}
```

Fonte: Simulador SISU

|

target

	MEAN	DIR_UFRJ	ENG_UFRJ	MED_UFRJ	MED_USP
MEAN	1.000000	0.982404	0.995595	0.982271	0.995285
DIR_UFRJ	0.982404	1.000000	0.969944	0.996584	0.966688
ENG_UFRJ	0.995595	0.969944	1.000000	0.975361	0.990613
MED_UFRJ	0.982271	0.996584	0.975361	1.000000	0.965819
MED_USP	0.995285	0.966688	0.990613	0.965819	1.000000



seleção univariada

1.13.2. Univariate feature selection

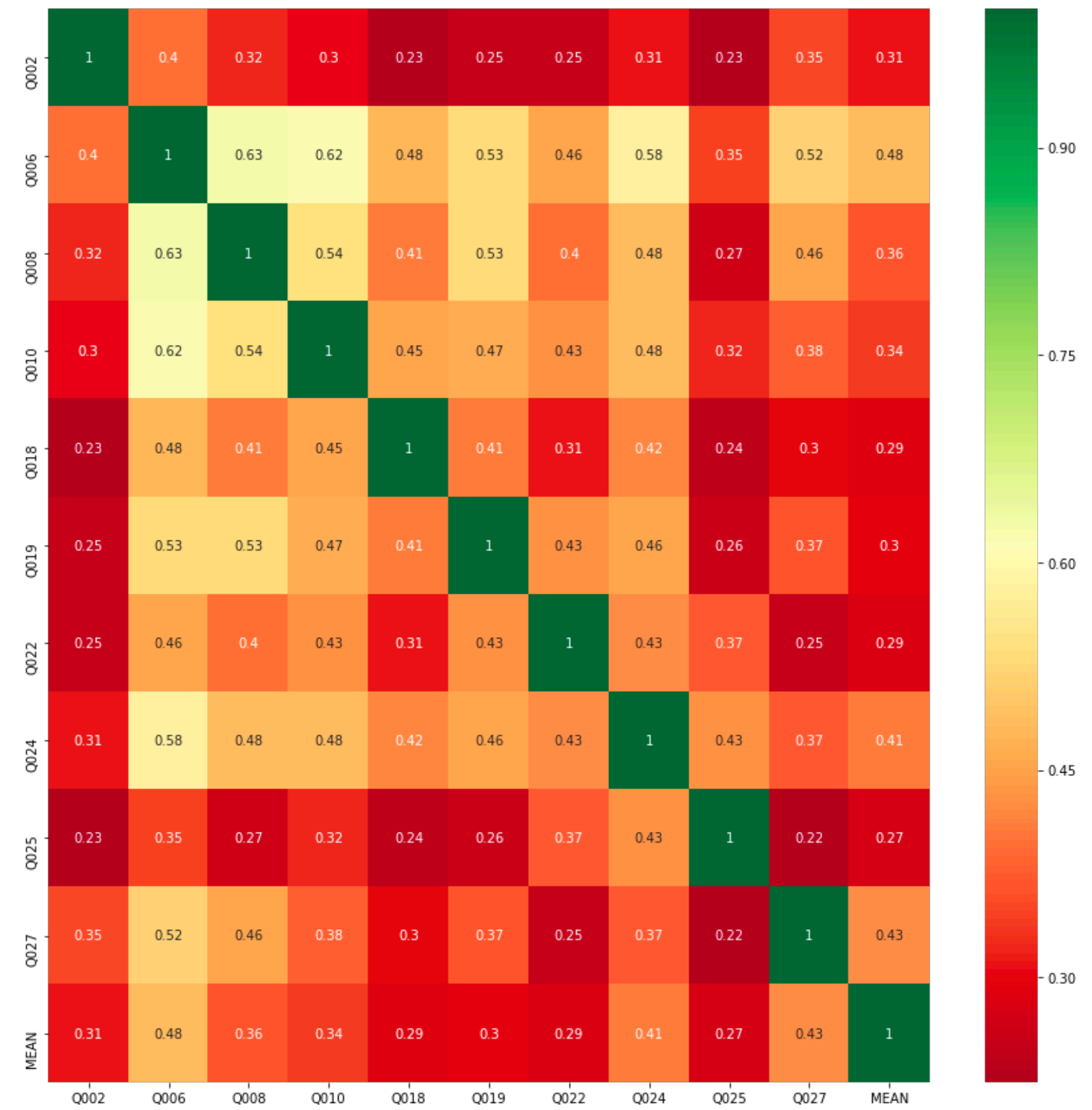
Univariate feature selection works by selecting the best features based on univariate statistical tests. It can be seen as a preprocessing step to an estimator. Scikit-learn exposes feature selection routines as objects that implement the `transform` method:

- **SelectKBest** removes all but the k highest scoring features

```
feature_selector = SelectKBest(score_func=f_regression, k=10)
feature_selector.fit_transform(X, y.MEAN)
```

selectKBest - resultado

0	Q006	Qual é a renda mensal de sua família? (Some a ...	1.183062e+06
1	Q027	Em que tipo de escola você frequentou o Ensino...	8.701589e+05
2	Q024	Na sua residência tem computador?	7.962705e+05
3	Q008	Na sua residência tem banheiro?	5.943384e+05
4	Q010	Na sua residência tem carro?	5.116835e+05
5	Q002	Até que série sua mãe, ou a mulher responsável...	4.098133e+05
6	Q019	Na sua residência tem televisão em cores?	3.737618e+05
7	Q018	Na sua residência tem aspirador de pó?	3.568314e+05
8	Q022	Na sua residência tem telefone celular?	3.459437e+05
9	Q025	Na sua residência tem acesso à Internet?	3.161334e+05



|

seleção de características

remoção de características

remoção de dados ausentes

seleção univariada

seleção a partir de modelo

análise de componentes PCA

|

seleção a partir de regressor - Lasso

```
import matplotlib.pyplot as plt
import numpy as np

from sklearn.feature_selection import SelectFromModel
from sklearn.linear_model import Lasso

# We use the base estimator LassoCV since the L1 norm promotes sparsity of features.
clf_lasso = Lasso(alpha=0.1)

# Set a minimum threshold of 0.25
sfm = SelectFromModel(clf_lasso)
sfm.fit(X_train, y_train)
X_transform = sfm.transform(X_train)
n_features = X_transform.shape[1]

# Reset the threshold till the number of features equals two.
# Note that the attribute can be set directly instead of repeatedly
# fitting the metatransformer.

# Plot the selected two features from X.
plt.title(
    "Features selected from Boston using SelectFromModel")
X_transform

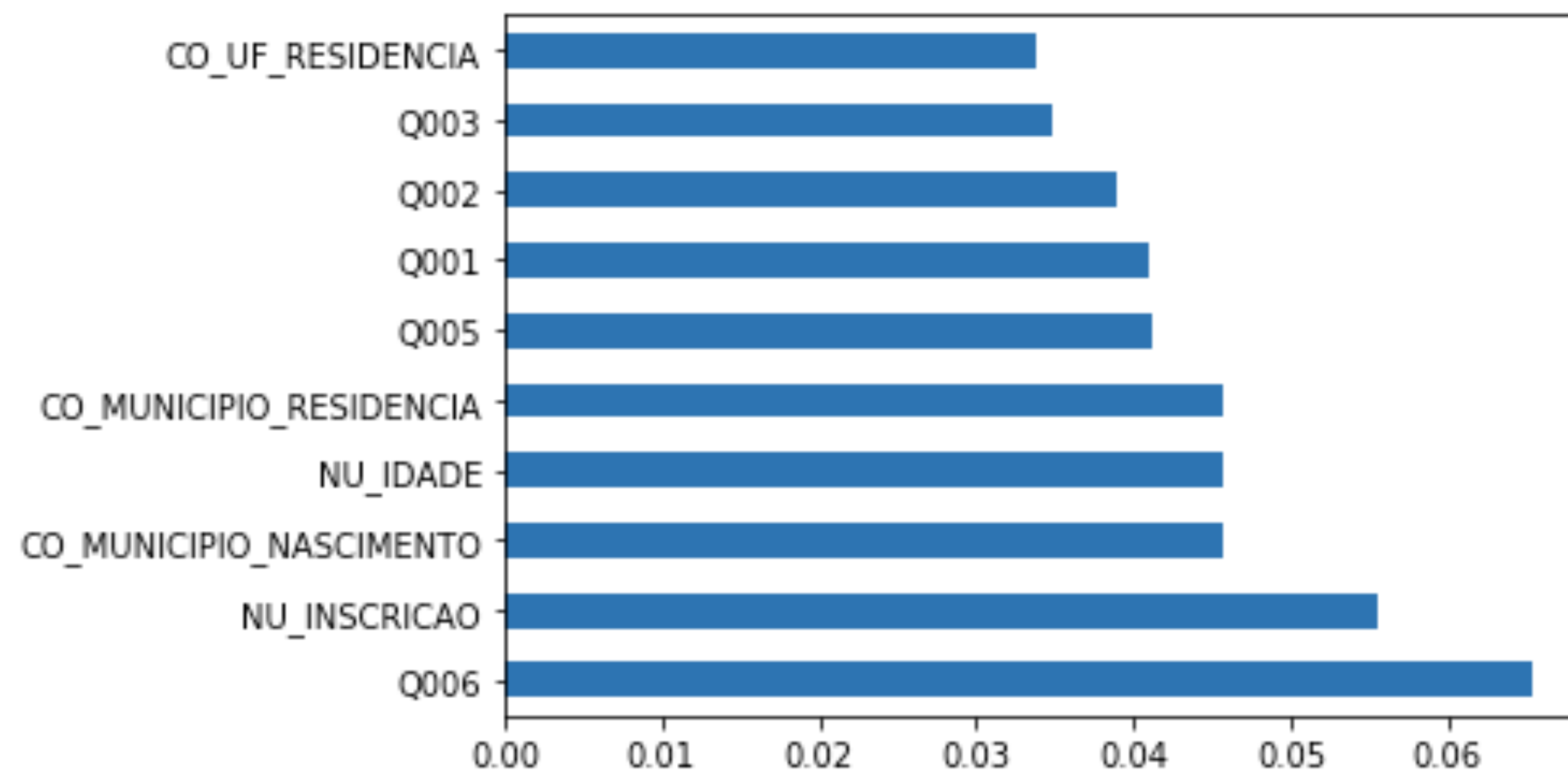
X_train.columns[sfm.get_support()]

Index(['TP_SEXO', 'TP_NACIONALIDADE', 'CO_UF_NASCIMENTO', 'TP_ST_CONCLUSAO',
       'Q006', 'Q011', 'Q012', 'Q013', 'Q015', 'Q024', 'Q025', 'Q026', 'Q027'],
      dtype='object')
```

seleção a partir de classificador

```
from sklearn.ensemble import ExtraTreesClassifier

model = ExtraTreesClassifier()
model.fit(X_train, y_train2)
print(model.feature_importances_) #use inbuilt class feature_importances of tree based classifiers
#plot graph of feature importances for better visualization
feat_importances = pd.Series(model.feature_importances_, index=X_train.columns)
feat_importances.nlargest(10).plot(kind='barh')
plt.show()
```



|

seleção de características

remoção de características

remoção de dados ausentes

seleção univariada

seleção a partir de modelo

análise de componentes PCA

|

análise de componentes PCA

```
pca = decomposition.PCA(n_components=10)
pca.fit(X_train)
```

```
PCA(copy=True, iterated_power='auto', n_components=10, random_state=None,
     svd_solver='auto', tol=0.0, whiten=False)
```

```
pca.explained_variance_ratio_
```

```
array([6.49073456e-01, 3.31328854e-01, 1.95976897e-02, 1.11147051e-11,
       2.75733008e-12, 8.35264578e-13, 7.80059318e-13, 3.95907538e-13,
       3.48153806e-13, 3.16146727e-13])
```

análise de componentes PCA

CO_MUNICIPIO_RESIDENCIA	1.697187e-02
CO_MUNICIPIO_NASCIMENTO	1.617195e-02
NU_IDADE	1.988258e-07
CO_UF_RESIDENCIA	1.678983e-07
CO_UF_NASCIMENTO	1.600319e-07
Q006	1.448310e-07
TP_ANO_CONCLUIU	1.178327e-07
Q027	3.343202e-08
Q010	2.614031e-08
Q004	2.525815e-08
Q002	2.524580e-08
Q008	2.059892e-08
Q009	1.920878e-08
TP_SEXO	1.882424e-08
IN_TREINEIRO	1.661433e-08
Q001	1.654071e-08
Q024	1.552582e-08
Q014	1.523811e-08
Q022	1.507852e-08

Name: PC-1, dtype: float64

Q006	5.745953e-07
Q010	1.457337e-07
Q003	1.315824e-07
Q022	1.289860e-07
Q024	1.259260e-07
Q001	1.179431e-07
Q004	1.158899e-07
Q016	1.134454e-07
Q014	1.106480e-07
Q002	1.026217e-07
Q013	9.289598e-08
Q023	7.918958e-08
Q025	7.695937e-08
Q018	7.427917e-08
Q009	7.382918e-08
Q008	7.345535e-08

Name: PC-2, dtype: float64

objetivos

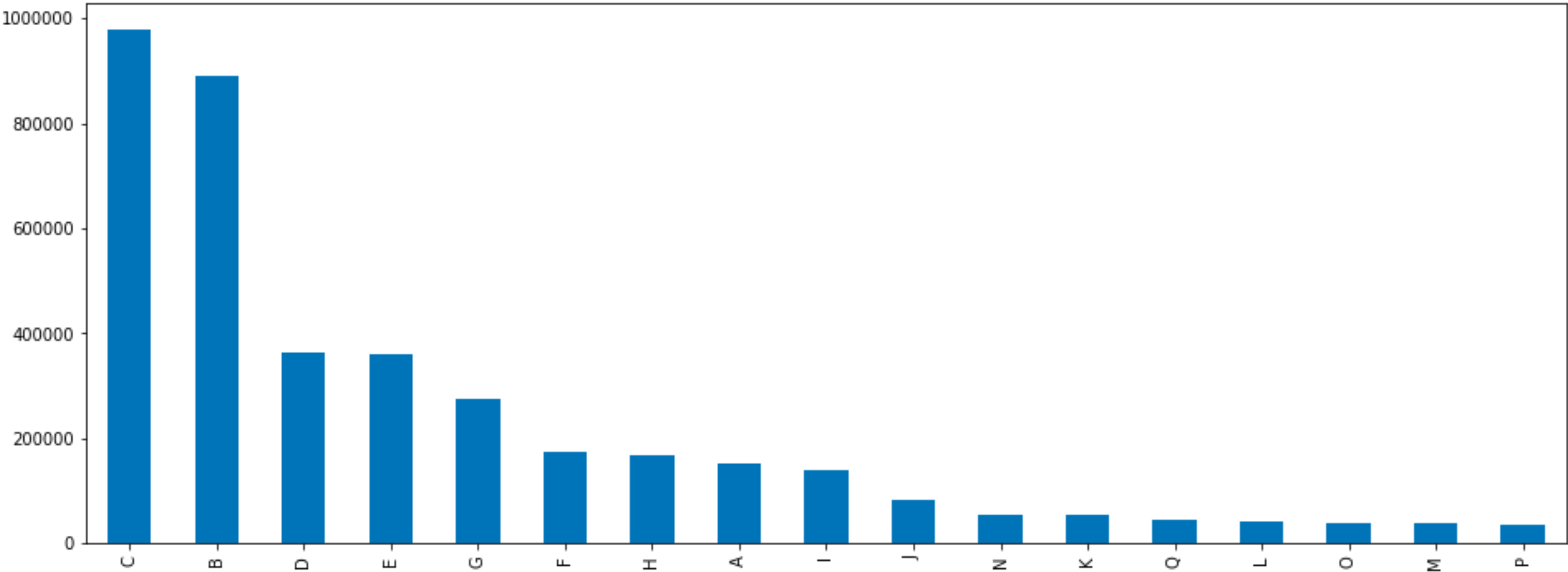
os microdados do ENEM

seleção de características

— **visualização de características**

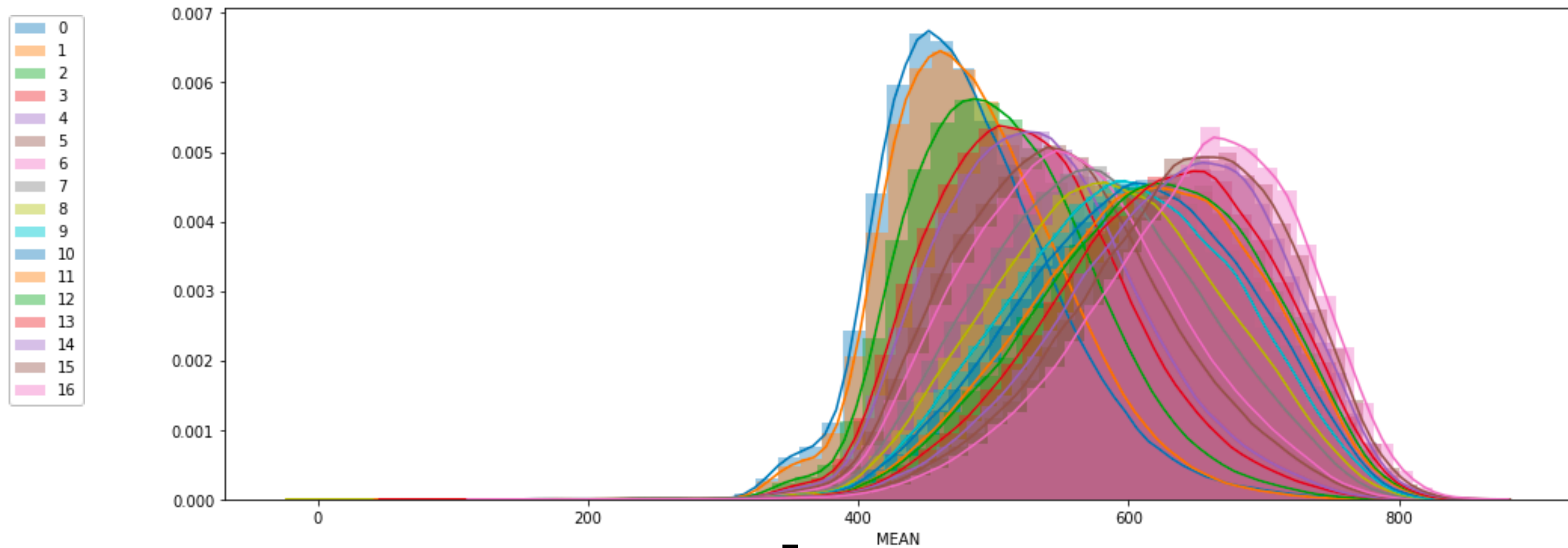
conclusão

Q006- Qual é a renda mensal de sua família?

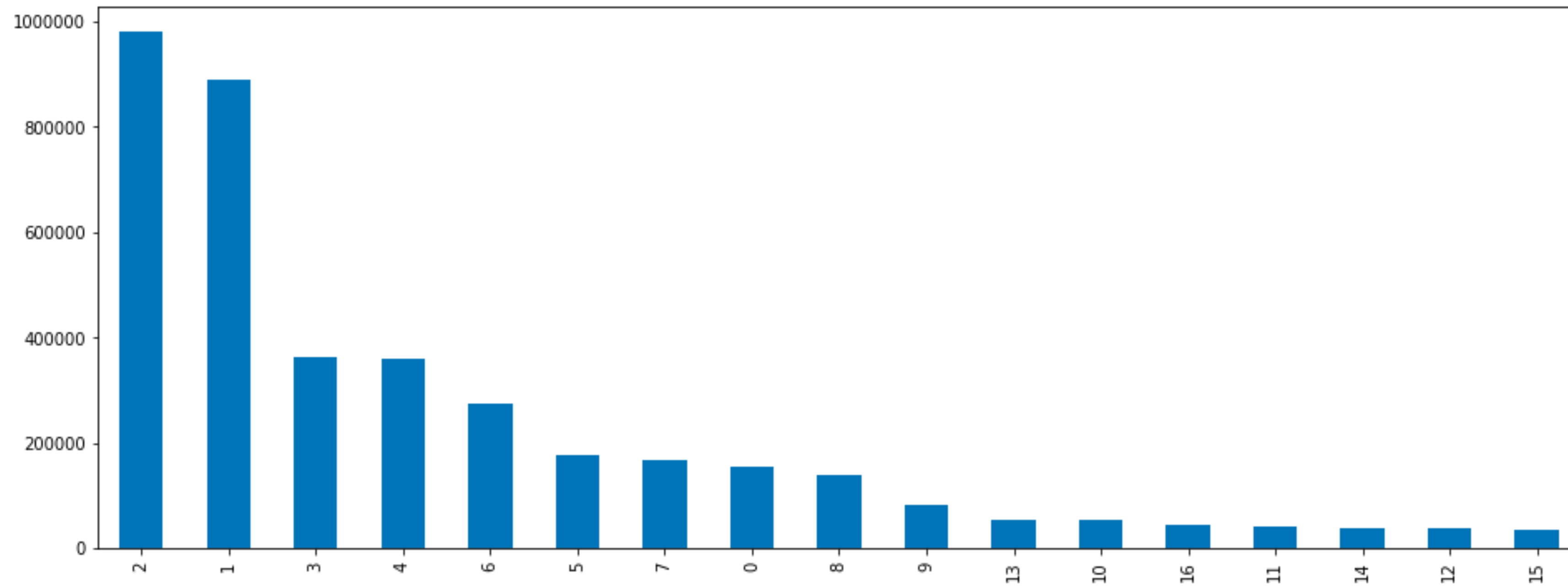


A	Nenhuma renda.
B	Até R\$ 954,00.
C	De R\$ 954,01 até R\$ 1.431,00.
D	De R\$ 1.431,01 até R\$ 1.908,00.
E	De R\$ 1.908,01 até R\$ 2.385,00.
F	De R\$ 2.385,01 até R\$ 2.862,00.
G	De R\$ 2.862,01 até R\$ 3.816,00.
H	De R\$ 3.816,01 até R\$ 4.770,00.
I	De R\$ 4.770,01 até R\$ 5.724,00.
J	De R\$ 5.724,01 até R\$ 6.678,00.
K	De R\$ 6.678,01 até R\$ 7.632,00.
L	De R\$ 7.632,01 até R\$ 8.586,00.
M	De R\$ 8.586,01 até R\$ 9.540,00.
N	De R\$ 9.540,01 até R\$ 11.448,00.
O	De R\$ 11.448,01 até R\$ 14.310,00.
P	De R\$ 14.310,01 até R\$ 19.080,00.
Q	Mais de R\$ 19.080,00.

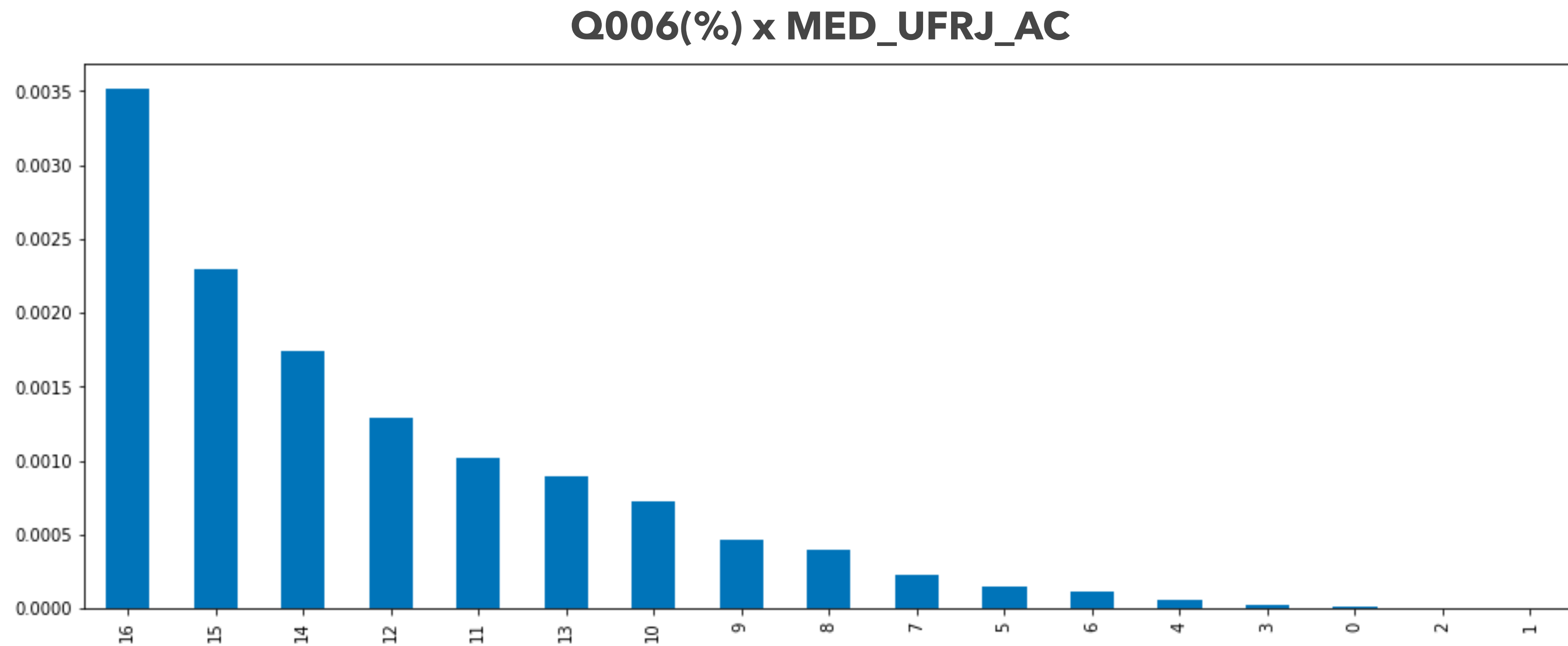
Q006- Qual é a renda mensal de sua família?



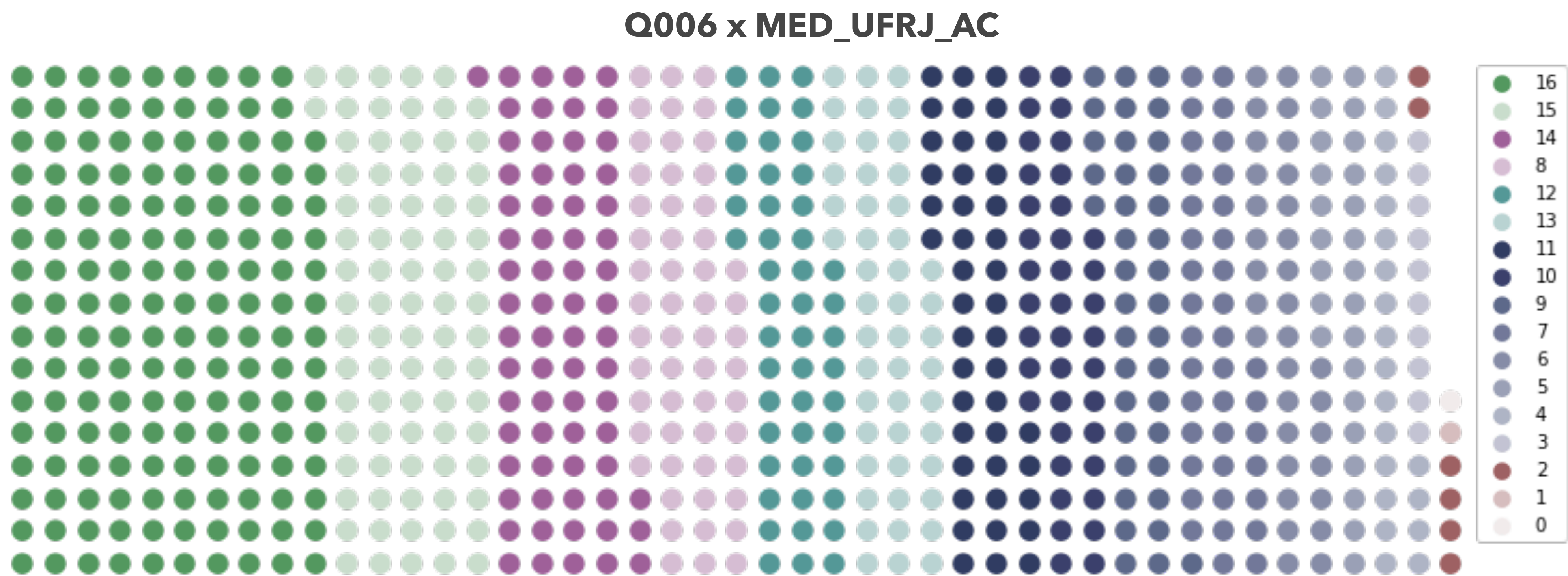
Q006- Qual é a renda mensal de sua família?



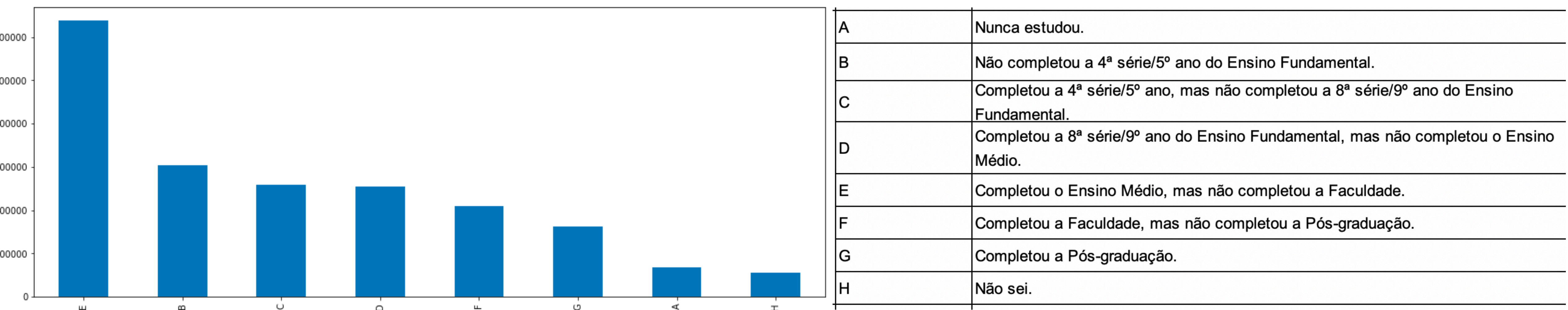
Q006- Qual é a renda mensal de sua família?



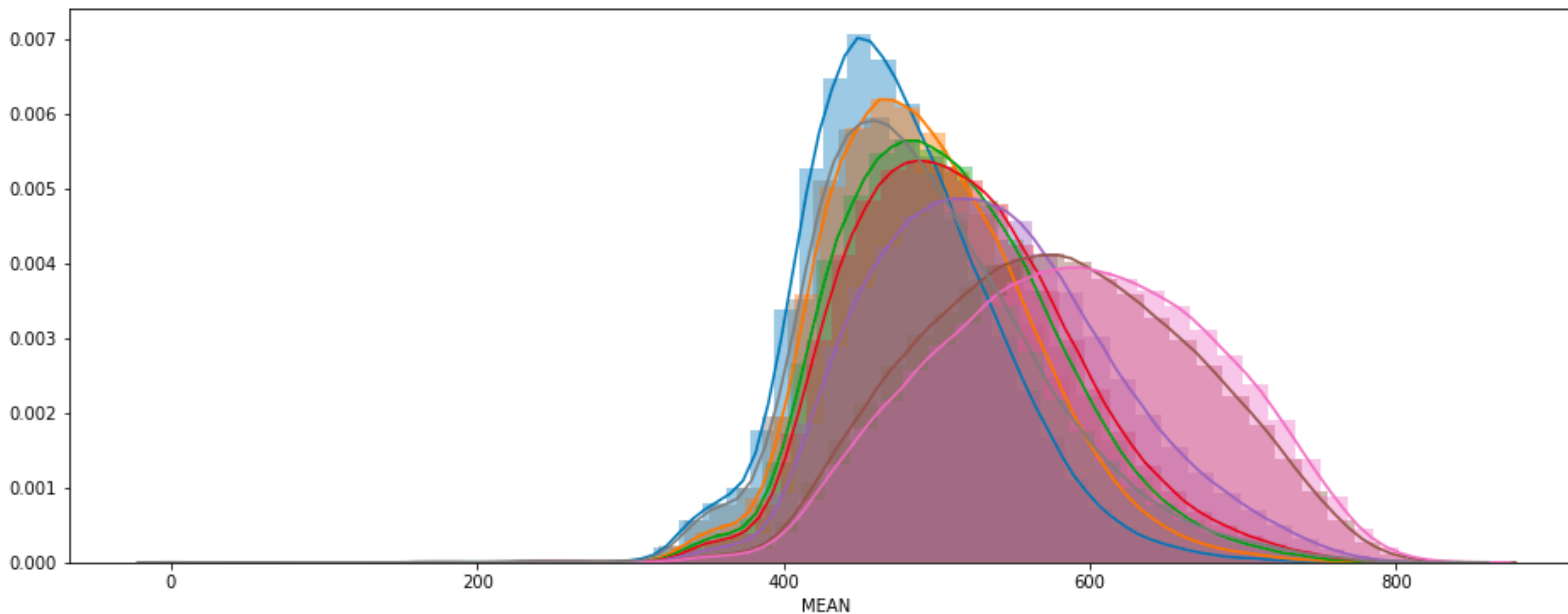
Q006- Qual é a renda mensal de sua família?



Q002 - Até que série sua mãe, ou a mulher responsável por você, estudou?

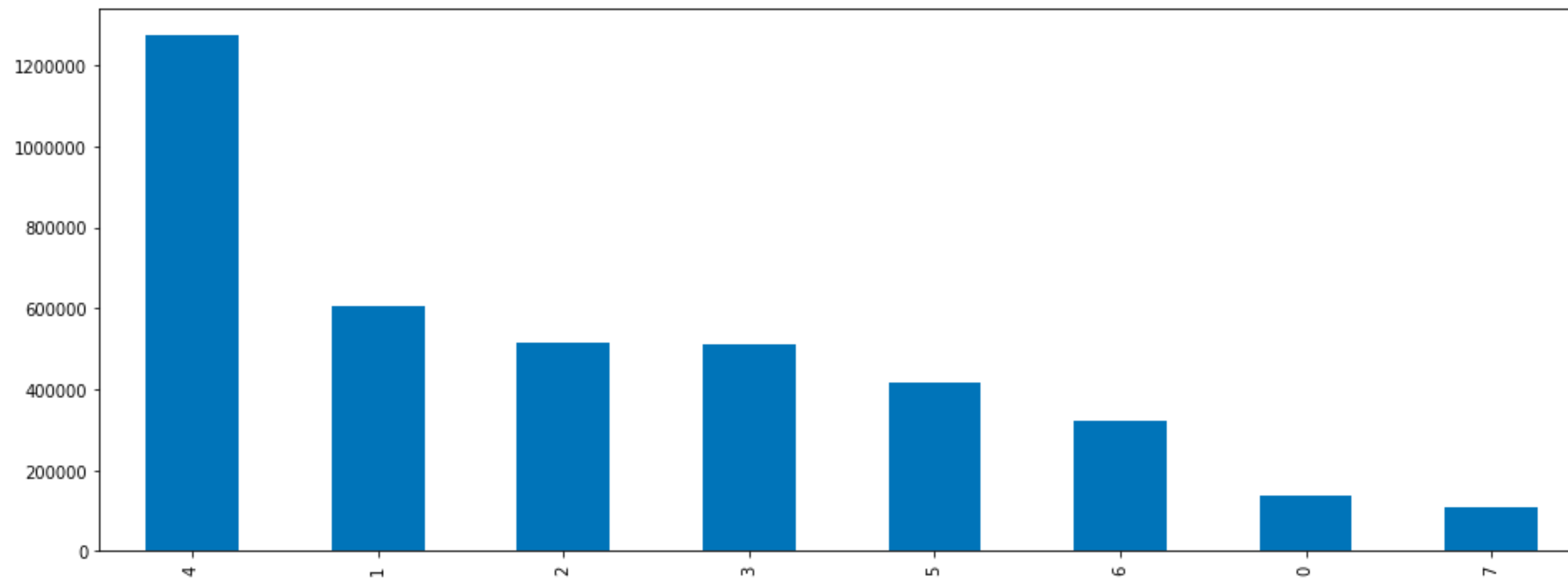


Q002



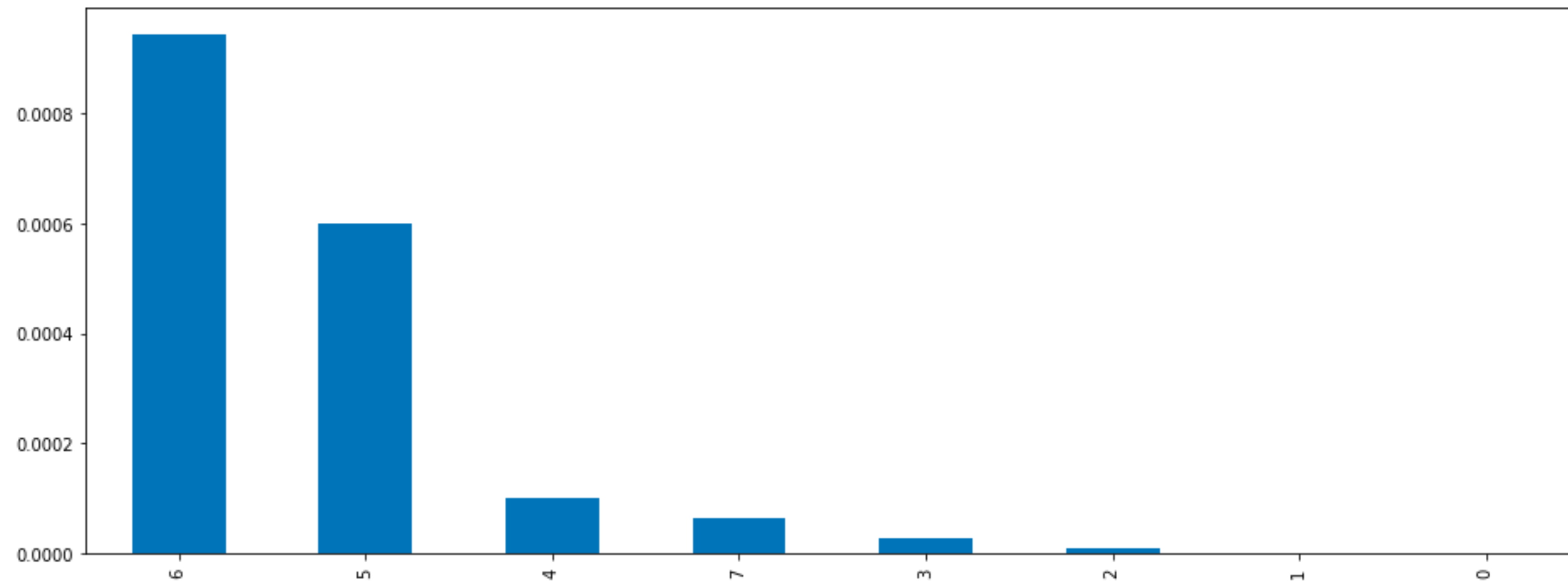
- '0': 'Nunca estudou'
- '1': 'Não completou a 4ª série/5º ano do Ensino Fundamental'
- '2': 'Completo a 4ª série/5º ano, mas não completou a 8ª série/9º ano do Ensino Fundamental'
- '3': 'Completo a 8ª série/9º ano do Ensino Fundamental, mas não completou o Ensino Médio'
- '4': 'Completo o Ensino Médio, mas não completou a Faculdade'
- '5': 'Completo a Faculdade, mas não completou a Pós-graduação'
- '6': 'Completo a Pós-graduação'
- '7': 'Não sei'

Q002 - Até que série sua mãe, ou a mulher responsável por você, estudou?

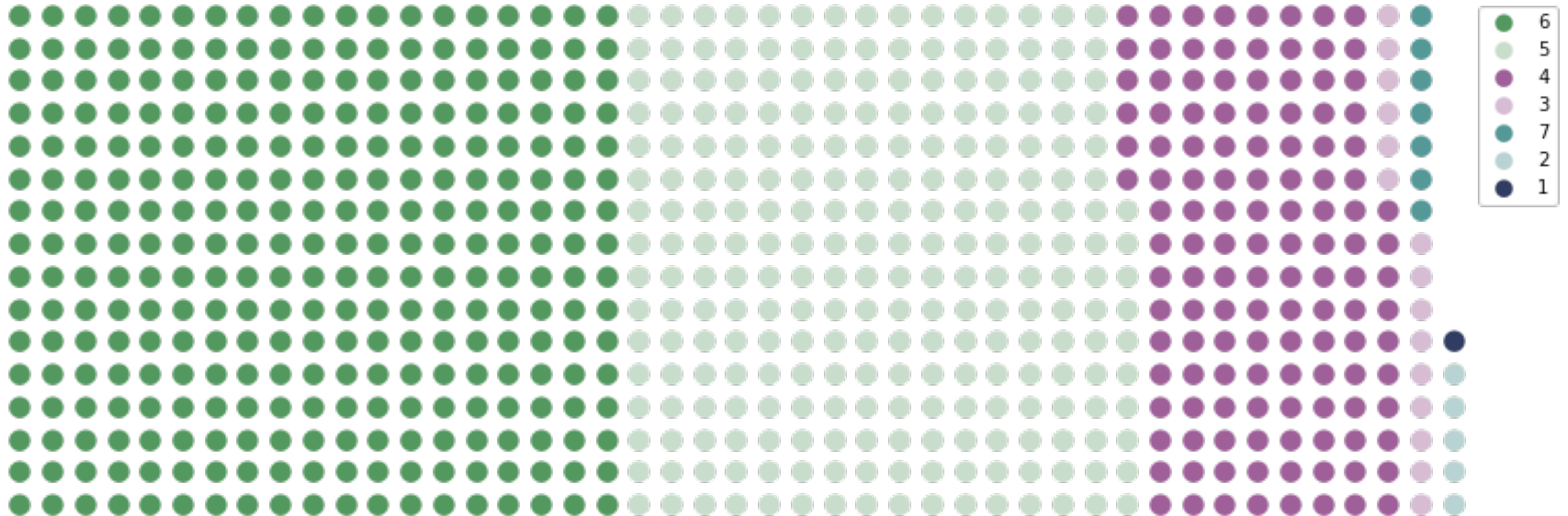


Q002 - Até que série sua mãe, ou a mulher responsável por você, estudou?

Q002(%) x MED_UFRJ_AC



Q002

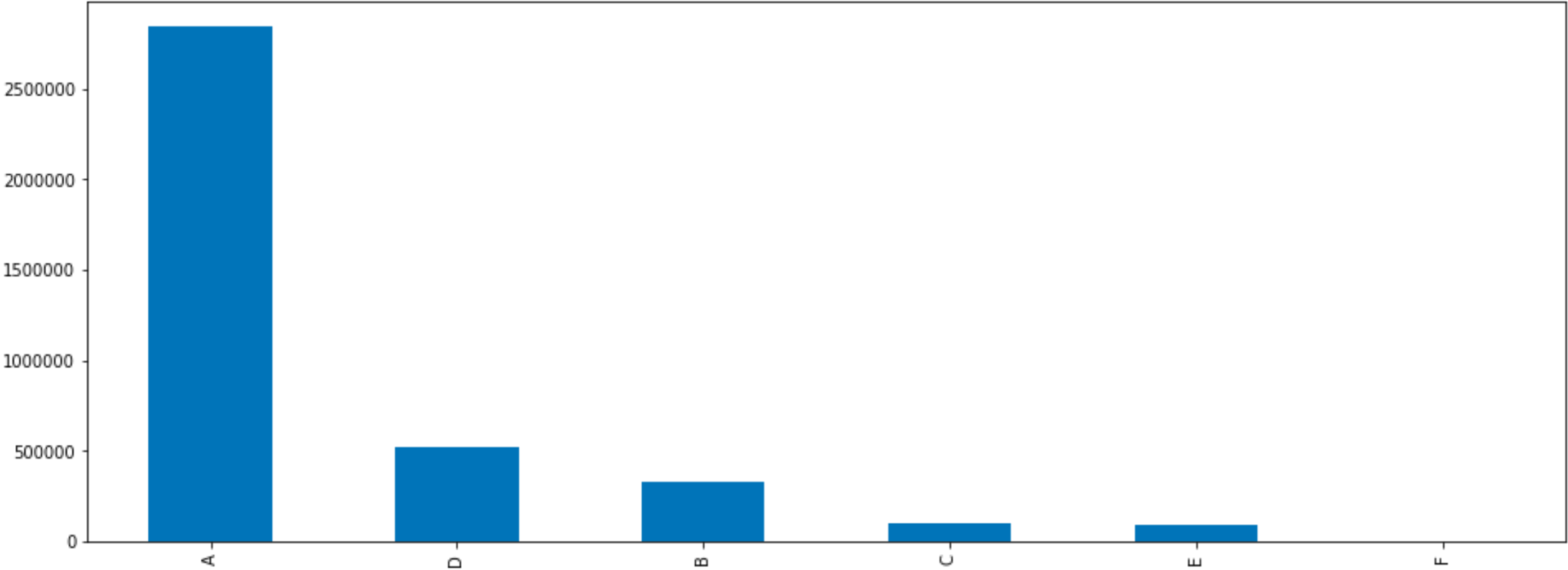


Q027 - Em que tipo de escola você frequentou o Ensino Médio?

A	Somente em escola pública.
B	Parte em escola pública e parte em escola privada SEM bolsa de estudo integral.
C	Parte em escola pública e parte em escola privada COM bolsa de estudo integral.
D	Somente em escola privada SEM bolsa de estudo integral.
E	Somente em escola privada COM bolsa de estudo integral.
F	Não frequentei a escola

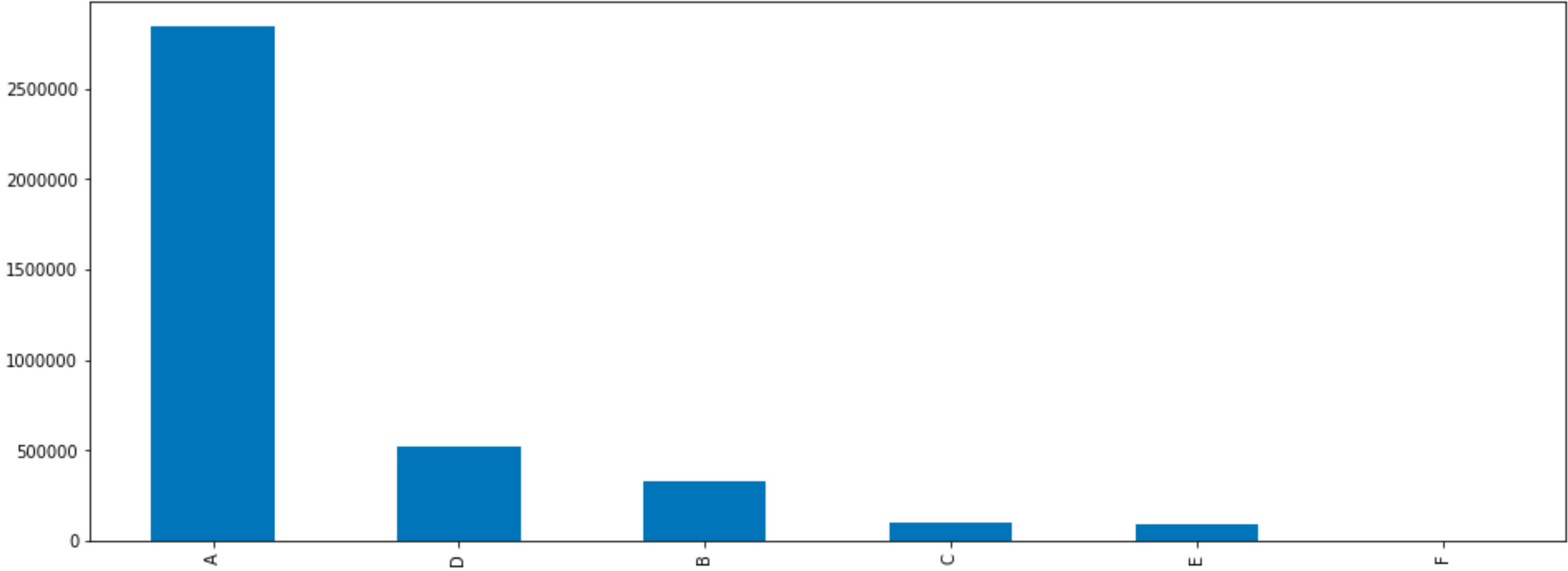
Q027 - Em que tipo de escola você frequentou o Ensino Médio?

A	Somente em escola pública.
B	Parte em escola pública e parte em escola privada SEM bolsa de estudo integral.
C	Parte em escola pública e parte em escola privada COM bolsa de estudo integral.
D	Somente em escola privada SEM bolsa de estudo integral.
E	Somente em escola privada COM bolsa de estudo integral.
F	Não frequentei a escola

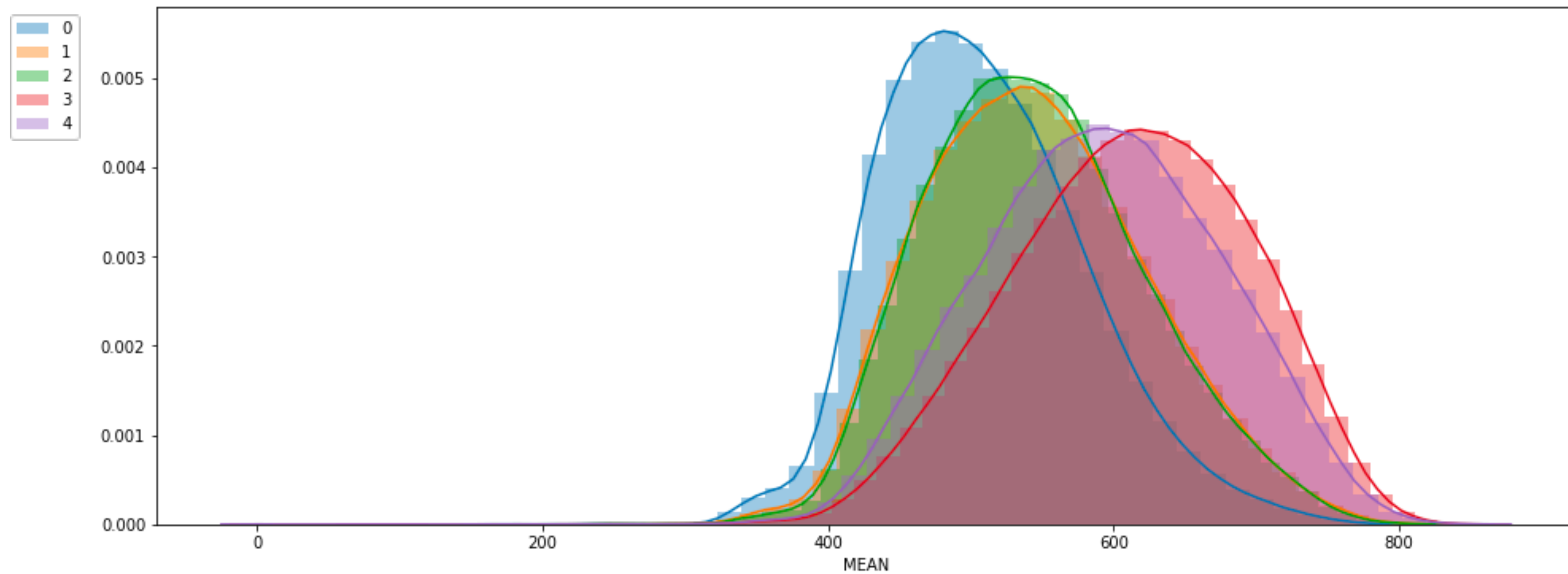


Q027 - Em que tipo de escola você frequentou o Ensino Médio?

A	Somente em escola pública.
B	Parte em escola pública e parte em escola privada SEM bolsa de estudo integral.
C	Parte em escola pública e parte em escola privada COM bolsa de estudo integral.
D	Somente em escola privada SEM bolsa de estudo integral.
E	Somente em escola privada COM bolsa de estudo integral.
F	Não frequentei a escola



Q027 - Em que tipo de escola você frequentou o Ensino Médio?





objetivos

os microdados do ENEM

seleção de características

visualização de características

— **conclusão**



características mais relevantes:

- **renda mensal da família**
- **que tipo de escola frequentou**
- **infraestrutura da residência:**
computador, telefone celular, internet,
banheiro, geladeira, veículo de
transporte
- **escolaridade da mãe**

muito obrigada!

kizzyterra@gmail.com