

```
In [4]: import pandas as pd
ds=pd.read_csv("Downloads/student-mat.csv")
ds.head(400)
```

```
Out[4]:
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	f
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...
390	MS	M	20	U	LE3	A	2	2	services	services	...	5	
391	MS	M	17	U	LE3	T	3	1	services	services	...	2	
392	MS	M	21	R	GT3	T	1	1	other	other	...	5	
393	MS	M	18	R	LE3	T	3	2	services	other	...	4	
394	MS	M	19	U	LE3	T	1	1	other	at_home	...	3	

395 rows × 33 columns



```
In [8]: import pandas as pd
import numpy as np
import seaborn as sbs
import matplotlib.pyplot as plt
ds1=pd.read_csv("Downloads/student-mat.csv",sep=',')
ds1.head()
```

```
Out[8]:
```

	school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	...	famrel	free
0	GP	F	18	U	GT3	A	4	4	at_home	teacher	...	4	
1	GP	F	17	U	GT3	T	1	1	at_home	other	...	5	
2	GP	F	15	U	LE3	T	1	1	at_home	other	...	4	
3	GP	F	15	U	GT3	T	4	2	health	services	...	3	
4	GP	F	16	U	GT3	T	3	3	other	other	...	4	

5 rows × 33 columns



```
In [12]: #for binary attributes
ds=ds1[['schoolsup', 'famsup', 'paid', 'activities', 'nursery', 'higher', 'internet']
ds.head()
```

```
Out[12]:
```

	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic
0	yes	no	no	no	yes	yes	no	no
1	no	yes	no	no	no	yes	yes	no
2	yes	no	yes	no	yes	yes	yes	no
3	no	yes	yes	yes	yes	yes	yes	yes
4	no	yes	yes	no	yes	yes	no	no

```
In [13]: ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 395 entries, 0 to 394
Data columns (total 8 columns):
#   Column          Non-Null Count  Dtype
---  -
0   schoolsup       395 non-null    object
1   famsup          395 non-null    object
2   paid            395 non-null    object
3   activities      395 non-null    object
4   nursery         395 non-null    object
5   higher          395 non-null    object
6   internet        395 non-null    object
7   romantic        395 non-null    object
dtypes: object(8)
memory usage: 24.8+ KB
```

```
In [16]: ds=ds.replace('no', 0)
ds=ds.replace('yes', 1)
ds.head()
```

```
Out[16]:
```

	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic
0	1	0	0	0	1	1	0	0
1	0	1	0	0	0	1	1	0
2	1	0	1	0	1	1	1	0
3	0	1	1	1	1	1	1	1
4	0	1	1	0	1	1	0	0

In [17]: `ds.info`

Out[17]: <bound method DataFrame.info of  

	ser	higher	internet	romantic	schoolsup	famsup	paid	activities	nur
0		1	0	0	0	1	1	0	0
1		0	1	0	0	0	1	1	0
2		1	0	1	0	1	1	1	0
3		0	1	1	1	1	1	1	1
4		0	1	1	0	1	1	0	0
..		...	...	...	...	...	...	...	...
390		0	1	1	0	1	1	0	0
391		0	0	0	0	0	1	1	0
392		0	0	0	0	0	1	0	0
393		0	0	0	0	0	1	1	0
394		0	0	0	0	1	1	1	0

  
[395 rows x 8 columns]>

In [19]: `n=np.array(ds[['schoolsup','famsup']])`  
`n=n.reshape(-1,2)`  
`n.shape`

Out[19]: (395, 2)

In [21]: `m=np.array(ds[['romantic','internet']])`  
`m=m.reshape(-1,2)`  
`m.shape`

Out[21]: (395, 2)

In [24]: `from scipy.spatial import distance`  
`dist_matrix = distance.cdist(n,m)`  
`from scipy.spatial import distance`  
`dist_matrix = distance.cdist(n,m)`  
`dist_matrix.shape`  
`sbs.heatmap(dist_matrix)`  
`plt.show()`

Out[24]: (395, 395)

```
In [49]: sbs.heatmap(dist_matrix)
sns.color_palette("blend:#7AB,#EDA", as_cmap=True)
plt.show()
```

-----  
**NameError** Traceback (most recent call last)

Cell In[49], line 2

```
1 sbs.heatmap(dist_matrix)
----> 2 sns.color_palette("blend:#7AB,#EDA", as_cmap=True)
3 plt.show()
```

**NameError**: name 'sns' is not defined



```
In [38]: numeric=ds1[['age', 'Medu', 'Fedu', 'traveltime', 'studytime', 'failures']]
numeric.head()
```

```
Out[38]:
```

	age	Medu	Fedu	traveltime	studytime	failures
0	18	4	4	2	2	0
1	17	1	1	1	2	0
2	15	1	1	1	2	3
3	15	4	2	1	3	0
4	16	3	3	1	2	0

```
In [39]: numeric.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 395 entries, 0 to 394  
Data columns (total 6 columns):  
#   Column      Non-Null Count  Dtype  
---  -  
0   age         395 non-null    int64  
1   Medu        395 non-null    int64  
2   Fedu        395 non-null    int64  
3   traveltime  395 non-null    int64  
4   studytime   395 non-null    int64  
5   failures    395 non-null    int64  
dtypes: int64(6)  
memory usage: 18.6 KB
```

```
In [41]: num1=np.array(numeric[['age','failures']])  
num1.reshape(-1,2)  
num1.shape
```

```
Out[41]: (395, 2)
```

```
In [42]: num2=np.array(numeric[['Fedu','Medu']])  
num2.reshape(-1,2)  
num2.shape
```

```
Out[42]: (395, 2)
```

```
In [47]: from scipy.spatial import distance
dist_matrix = distance.cdist(num1,num2)
dist_matrix.shape
sbs.heatmap(dist_matrix)
```



```
In [10]: #for nominal attribute
nominal=ds1[['Mjob','Fjob','reason','guardian']]
nominal=nominal.replace('at_home','home')
nominal=(nominal.astype('category'))
```

```
In [11]: from sklearn.preprocessing import LabelEncoder
lb=LabelEncoder()
nominal['guardian']=lb.fit_transform(nominal['guardian'])
```

```
In [20]: nominal['Mjob']=lb.fit_transform(nominal['Mjob'])
nominal['Fjob']=lb.fit_transform(nominal['Fjob'])
nominal['reason']=lb.fit_transform(nominal['reason'])
```

```
In [15]: nominal.head()
```

Out[15]:

	Mjob	Fjob	reason	guardian
0	1	4	0	1
1	1	2	0	0
2	1	2	2	1
3	0	3	1	1
4	2	2	1	0

```
In [16]: nom1=np.array(nominal)
nom1.reshape(-1,2)
```

```
Out[16]: array([[1, 4],
               [0, 1],
               [1, 2],
               ...,
               [0, 1],
               [2, 1],
               [0, 0]])
```

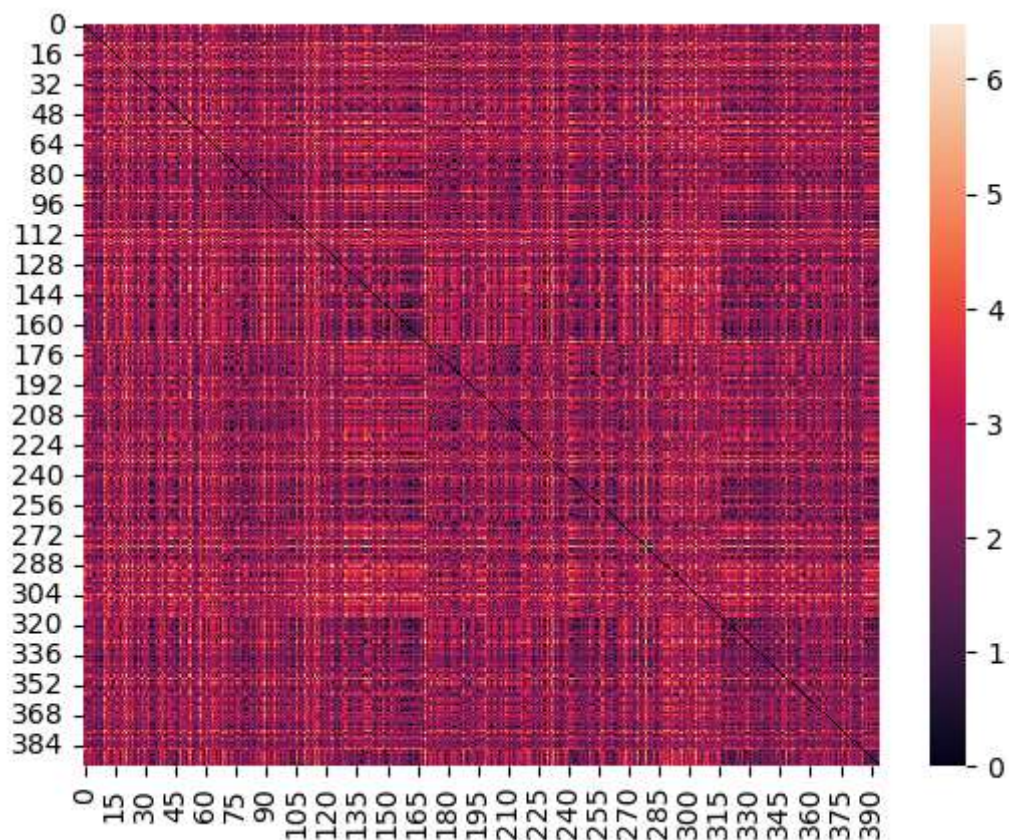
```
In [17]: nom2=np.array(nominal)
nom2.reshape(-1,2)
```

```
Out[17]: array([[1, 4],
               [0, 1],
               [1, 2],
               ...,
               [0, 1],
               [2, 1],
               [0, 0]])
```



```
In [18]: from scipy.spatial import distance
dist_matrix = distance.cdist(nom1,nom2)
sbs.heatmap(dist_matrix)
dist_matrix.shape
```

Out[18]: (395, 395)



```
In [ ]: 78965
```