

▾ Nigeria Antibiotic Susceptibility DS

quick EDA of the dataset over at <https://data.mendeley.com/datasets/nd4w9wnmzc>

```
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt

from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

url = '/content/drive/MyDrive/dataset.xlsx'
df = pd.read_excel(url)
df
```

	S\N	SEX	WARD\CLINIC	SPECIMEN	CULTURE	Unnamed: 5	ORGANISM	AMOXY CLAVL SEN
0	1	M	Outpatient	Urine	Positive	NaN	Klebsiella sp	
1	2	M	Outpatient	Urine	Positive	NaN	Klebsiella sp	
2	3	M	Outpatient	Urine	Positive	NaN	Klebsiella sp	
3	4	M	Outpatient	Urine	Positive	NaN	Klebsiella sp	
4	5	M	Outpatient	Urine	Positive	NaN	Klebsiella sp	
...
3607	3737	F	In-patient	urine	Negative	NaN	NaN	
3608	3738	F	In-patient	urine	Negative	NaN	NaN	
3609	3739	F	In-patient	urine	Negative	NaN	NaN	
3610	3740	F	In-patient	urine	Negative	NaN	NaN	
3611	3741	F	In-patient	urine	Negative	NaN	NaN	

```
df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3612 entries, 0 to 3611
Data columns (total 28 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   S\N                                    3612 non-null   object
1   SEX                                    3612 non-null   object
2   WARD\CLINIC                           3612 non-null   object
3   SPECIMEN                               3612 non-null   object
4   CULTURE                                3612 non-null   object
5   Unnamed: 5                             0 non-null      float64
6   ORGANISM                               808 non-null    object
7   AMOXYCILLIN CLAVULANATE SENSITIVE      120 non-null    object
8   AMOXYCILLIN CLAVULANATE INTERMEDIATE   65 non-null     object
9   AMOXYCILLIN CLAVULANATE RESISTANT       421 non-null    object
10  CEFTAZIDIME SENSITIVE                   281 non-null    object
11  CEFTAZIDIME INTERMEDIATE                64 non-null     object
12  CEFTAZIDIME RESISTANT                   259 non-null    object
13  CIPROFLOXACIN SENSITIVE                 330 non-null    object
14  CIPROFLOXACIN INTERMEDIATE              57 non-null     object
15  CIPROFLOXACIN RESISTANT                 217 non-null    object
16  MEROPENEM SENSITIVE                     524 non-null    object
```

```

17  MEROPENEM INTERMEDIATE          17 non-null    object
18  MEROPENEMRESISTANT              64 non-null    object
19  NITROFURANTOIN SENSITIVE         333 non-null   object
20  NITROFURANTOIN INTERMEDIATE      34 non-null    object
21  NITROFURANTOIN RESISTANT          238 non-null   object
22  PIP TAZOBACTAM SENSITIVE          264 non-null   object
23  PIP TAZOBACTAM INTERMEDIATE       44 non-null    object
24  PIP TAZOBACTAM RESISTANT          298 non-null   object
25  CEFTRIAXONE SENSITIVE             112 non-null   object
26  CEFTRIAXONE INTERMEDIATE         59 non-null    object
27  CEFTRIAXONE RESISTANT            168 non-null   object
dtypes: float64(1), object(27)
memory usage: 790.2+ KB

```

```
antibiotics = df.columns[7:]
```

```
df.drop(columns=['Unnamed: 5'], inplace=True)
```

```
def look_unique(df):
    for column in df:
        print(column, ': ', df[column].unique())
```

```
look_unique(df)
```

```

S\N : [1 2 3 ... 3739 3740 3741]
SEX : ['M' 'F']
WARD\CLINIC : ['Outpatient' 'In-patient']
SPECIMEN : ['Urine' 'urine']
CULTURE : ['Positive' 'positive' 'Negative']
ORGANISM : ['Klebsiella sp' 'E.coli' 'Pseudomonas sp' 'Proteus spp'
'Staphylococcus aureus' 'Staphylococcus epidermidis'
'Staphylococcus saprophyticus' 'Enterococcus' 'Candida spp' nan]
AMOXYCILLIN CLAVULANATE SENSITIVE : ['S' nan 'NON']
AMOXYCILLIN CLAVULANATE INTERMEDIATE : [nan 'I']
AMOXYCILLIN CLAVULANATE RESISTANT : [' ' nan 'R']
CEFTAZIDIME SENSITIVE : ['S' nan]
CEFTAZIDIME INTERMEDIATE : [nan 'I']
CEFTAZIDIME RESISTANT : [nan 'R']
CIPROFLOXACIN SENSITIVE : ['S' nan]
CIPROFLOXACIN INTERMEDIATE : [nan 'I']
CIPROFLOXACINRESISTANT : [nan 'R']
MEROPENEM SENSITIVE : ['S' nan]
MEROPENEM INTERMEDIATE : [nan 'I' 'NON']
MEROPENEMRESISTANT : [nan 'R']
NITROFURANTOIN SENSITIVE : ['S' nan]
NITROFURANTOIN INTERMEDIATE : [nan 'I' 'NON']
NITROFURANTOIN RESISTANT : [nan 'R']
PIP TAZOBACTAM SENSITIVE : ['S' nan]
PIP TAZOBACTAM INTERMEDIATE : [nan 'I' 'NON']
PIP TAZOBACTAM RESISTANT : [nan 'R']
CEFTRIAXONE SENSITIVE : [nan 'S']
CEFTRIAXONE INTERMEDIATE : [nan 'I']
CEFTRIAXONE RESISTANT : [nan 'R']

```

```

df[antibiotics] = df[antibiotics].replace(['NON', ' '], np.nan)
df.replace('positive', 'Positive', inplace=True)
look_unique(df)

```

```

S\N : [1 2 3 ... 3739 3740 3741]
SEX : ['M' 'F']
WARD\CLINIC : ['Outpatient' 'In-patient']
SPECIMEN : ['Urine' 'urine']
CULTURE : ['Positive' 'Negative']
ORGANISM : ['Klebsiella sp' 'E.coli' 'Pseudomonas sp' 'Proteus spp'
'Staphylococcus aureus' 'Staphylococcus epidermidis'
'Staphylococcus saprophyticus' 'Enterococcus' 'Candida spp' nan]
AMOXYCILLIN CLAVULANATE SENSITIVE : ['S' nan]
AMOXYCILLIN CLAVULANATE INTERMEDIATE : [nan 'I']
AMOXYCILLIN CLAVULANATE RESISTANT : [nan 'R']
CEFTAZIDIME SENSITIVE : ['S' nan]
CEFTAZIDIME INTERMEDIATE : [nan 'I']
CEFTAZIDIME RESISTANT : [nan 'R']
CIPROFLOXACIN SENSITIVE : ['S' nan]
CIPROFLOXACIN INTERMEDIATE : [nan 'I']
CIPROFLOXACINRESISTANT : [nan 'R']
MEROPENEM SENSITIVE : ['S' nan]
MEROPENEM INTERMEDIATE : [nan 'I']
MEROPENEMRESISTANT : [nan 'R']
NITROFURANTOIN SENSITIVE : ['S' nan]
NITROFURANTOIN INTERMEDIATE : [nan 'I']

```

```

NITROFURANTOIN RESISTANT : [nan 'R']
PIP TAZOBACTAM SENSITIVE : ['S' nan]
PIP TAZOBACTAM INTERMEDIATE : [nan 'I']
PIP TAZOBACTAM RESISTANT : [nan 'R']
CEFTRIAXONE SENSITIVE : [nan 'S']
CEFTRIAXONE INTERMEDIATE : [nan 'I']
CEFTRIAXONE RESISTANT : [nan 'R']

```

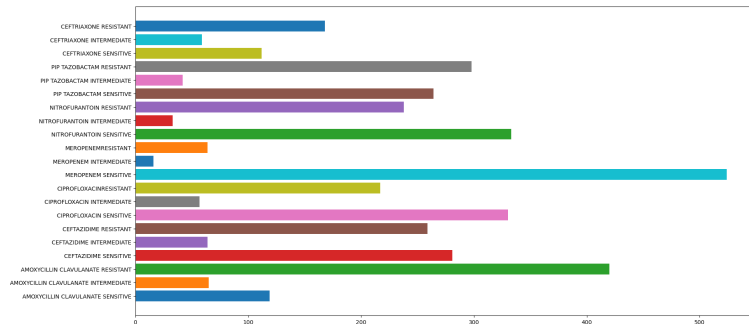
```

plt.figure(figsize=(20,10))

for antibiotic in antibiotics:
    plt.barh(antibiotic, df[antibiotic].count())

plt.show() # TODO ascending order / color code

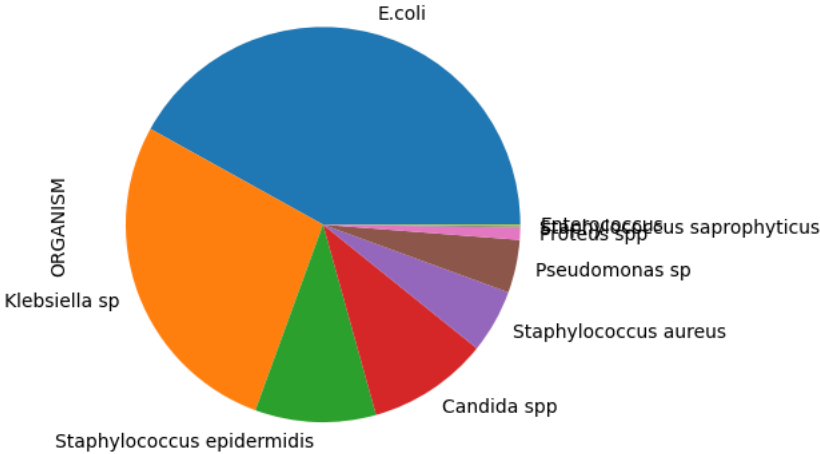
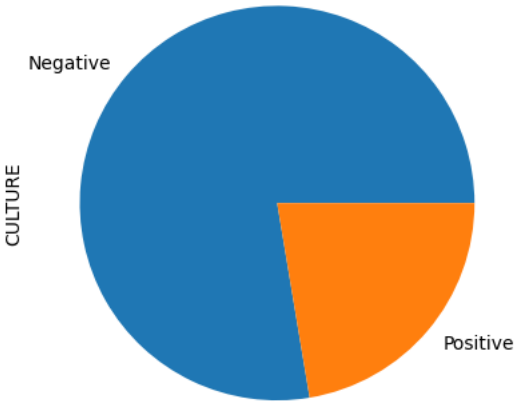
```



```

for categorical in df.columns[1:6]:
    df[categorical].value_counts().plot.pie()
plt.show() # TODO subplots glitching

```



```
numerical_df = df[antibiotics].replace(np.nan, 0).replace(['S', 'R', 'I'], 1)
numerical_df.corr() # requires integers
```

	AMOXICILLIN CLAVULANATE SENSITIVE	AMOXICILLIN CLAVULANATE INTERMEDIATE	AMOXICILLIN CLAVULANATE RESISTANT	CEFTAZIDIME SENSITIVE	CEFTAZIDIME INTERMEDIATE
AMOXICILLIN CLAVULANATE SENSITIVE	1.000000	-0.024986	-0.066953	0.635489	-0.024986
AMOXICILLIN CLAVULANATE INTERMEDIATE	-0.024986	1.000000	-0.049104	0.466080	-0.066953
AMOXICILLIN CLAVULANATE RESISTANT	-0.066953	-0.049104	1.000000	0.207412	-0.049104
CEFTAZIDIME SENSITIVE	0.635489	0.466080	0.207412	1.000000	-0.039009
CEFTAZIDIME INTERMEDIATE	-0.024790	-0.018181	0.370258	-0.039009	1.000000
CEFTAZIDIME RESISTANT	-0.051299	-0.037623	0.766196	-0.080723	0.370258
CIPROFLOXACIN SENSITIVE	0.582085	0.426912	0.322619	0.894439	-0.051299
CIPROFLOXACIN INTERMEDIATE	-0.023372	-0.017141	0.349080	0.012985	-0.037623
CIPROFLOXACINRESISTANT	-0.046664	-0.034224	0.696974	-0.073430	0.766196
MEROPENEM SENSITIVE	0.448071	0.328624	0.684384	0.705081	-0.046664
MEROPENEM INTERMEDIATE	-0.012312	-0.009030	0.183890	-0.019374	-0.034224
MEROPENEMRESISTANT	-0.024790	-0.018181	0.370258	-0.039009	0.696974
NITROFURANTOIN SENSITIVE	0.579192	0.424790	0.329224	0.911411	-0.024790
NITROFURANTOIN INTERMEDIATE	-0.017724	-0.012999	0.264718	-0.027890	-0.018181
NITROFURANTOIN RESISTANT	-0.049022	-0.035954	0.732188	-0.077140	0.370258
PIP TAZOBACTAM SENSITIVE	0.657302	0.482077	0.163593	0.931075	-0.049022
PIP TAZOBACTAM INTERMEDIATE	-0.020020	-0.014683	0.299018	0.161328	-0.035954
PIP TAZOBACTAM RESISTANT	-0.055349	-0.040594	0.826682	-0.064555	0.732188
CEFTRIAXONE SENSITIVE	0.530714	0.468392	-0.025028	0.615898	-0.049022
CEFTRIAXONE INTERMEDIATE	-0.023785	-0.017444	0.355251	0.329513	-0.035954
CEFTRIAXONE RESISTANT	-0.040766	-0.029898	0.608877	-0.064149	0.732188

highest pearson corrolation values for the given antibiotics

```
numerical_df.corrwith(numerical_df['AMOXICILLIN CLAVULANATE SENSITIVE']).sort_values(ascending=False)[1:10]
```

```
PIP TAZOBACTAM SENSITIVE    0.657302
CEFTAZIDIME SENSITIVE       0.635489
CIPROFLOXACIN SENSITIVE     0.582085
NITROFURANTOIN SENSITIVE    0.579192
CEFTRIAXONE SENSITIVE       0.530714
```

```

MEROPENEM SENSITIVE          0.448071
MEROPENEM INTERMEDIATE      -0.012312
NITROFURANTOIN INTERMEDIATE -0.017724
PIP TAZOBACTAM INTERMEDIATE -0.020020
dtype: float64

```

```
numerical_df.corrwith(numerical_df['CIPROFLOXACIN SENSITIVE']).sort_values(ascending=False)[1:10]
```

```

NITROFURANTOIN SENSITIVE      0.961816
CEFTAZIDIME SENSITIVE         0.894439
PIP TAZOBACTAM SENSITIVE      0.878184
MEROPENEM SENSITIVE           0.769769
AMOXYCILLIN CLAVULANATE SENSITIVE 0.582085
CEFTRIAXONE SENSITIVE         0.564141
AMOXYCILLIN CLAVULANATE INTERMEDIATE 0.426912
CEFTRIAXONE INTERMEDIATE      0.406388
PIP TAZOBACTAM INTERMEDIATE   0.333097
dtype: float64

```

CIPROFLOXACIN SENSITIVE is highly predictable

task

probability of a patient having resistance to AMOXYCILLIN CLAVULANATE SENSITIVE if they have resistance to CIPROFLOXACIN SENSITIVE

- $P(\text{AMOX} \mid \text{CIPR})$

```

total = df.shape[0]
p_cipr = df['CIPROFLOXACIN SENSITIVE'].count() / total
intercept = df[df['CIPROFLOXACIN SENSITIVE'].notna()
               & df['AMOXYCILLIN CLAVULANATE SENSITIVE'].notna()]
intercept = len(intercept) / total
print(f"P(AMOX | CIPR) = {intercept / p_cipr * 100:.3f}%")

```

P(AMOX | CIPR) is 36.061%