

Importer les packages à utiliser et les données sur l'agriculture

```
In [1]:
import pyforest
from Utils import Remove
import warnings
warnings.filterwarnings('ignore')
```

```
In [3]:
Agriculture_Data = pd.read_csv('data/agriculture_data/Agriculture_Data.csv')
```

```
In [4]:
Agriculture_Data.head()
```

Out[4]:

	region	indicateur	Unit	Date	Value
0	Dakar	Superficie moyenne des parcelles par ménage (Ha)	Ha	2017	0.462637
1	Dakar	Superficie moyenne des parcelles par ménage (Ha)	Ha	2018	0.465369
2	Dakar	Ménages exploitant 3 à 5 parcelles (%)	%	2018	0.000000
3	Dakar	Ménages exploitant moins de 3 parcelles (%)	%	2017	96.799999
4	Dakar	Ménages exploitant moins de 3 parcelles (%)	%	2018	100.000000

```
In [5]:
np.unique(Agriculture_Data['indicateur'])
```

Out[5]:

```
array(['Ménages exploitant 3 à 5 parcelles (%)',
      'Ménages exploitant moins de 3 parcelles (%)',
      'Ménages exploitant plus de 6 parcelles (%)',
      'Nombre moyen de parcelles par ménage',
      'Superficie moyenne des parcelles par ménage (Ha)',
      'Taille moyenne des ménages'], dtype=object)
```

Lecture et transformation des données sur l'agriculture

Ménages exploitant moins de 3 parcelles pour toutes les régions

```
In [6]:
data1 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Ménages exploitant moins de 3 parcelles (%)'].reset_index(drop = True)
data1.head()
```

Out[6]:

	region	indicateur	Unit	Date	Value
0	Dakar	Ménages exploitant moins de 3 parcelles (%)	%	2017	96.799999
1	Dakar	Ménages exploitant moins de 3 parcelles (%)	%	2018	100.000000
2	Diourbel	Ménages exploitant moins de 3 parcelles (%)	%	2017	12.540621
3	Diourbel	Ménages exploitant moins de 3 parcelles (%)	%	2018	25.740260
4	Fatick	Ménages exploitant moins de 3 parcelles (%)	%	2017	37.248814

```
In [7]:
data1 = data1[['region', 'Date', 'Value']].reset_index(drop = True)
```

Ménages exploitant 3 à 5 parcelles pour toutes les régions

In [8]:

```
data2 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Ménages exploitant 3 à 5 parcelles (%)']
data2.head()
```

Out[8]:

	region	indicateur	Unit	Date	Value
2	Dakar	Ménages exploitant 3 à 5 parcelles (%)	%	2018	0.000000
12	Diourbel	Ménages exploitant 3 à 5 parcelles (%)	%	2017	64.680761
13	Diourbel	Ménages exploitant 3 à 5 parcelles (%)	%	2018	57.751203
24	Fatick	Ménages exploitant 3 à 5 parcelles (%)	%	2017	52.979088
25	Fatick	Ménages exploitant 3 à 5 parcelles (%)	%	2018	38.961223

In [9]:

```
R = list(np.arange(28))
R = Remove(R,[0,19])
data2 = data2[['Value']]
data2.index = R
```

Ménages exploitant plus de 6 parcelles pour toutes les régions

In [10]:

```
data3 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Ménages exploitant plus de 6 parcelles (%)']
data3.head()
```

Out[10]:

	region	indicateur	Unit	Date	Value
5	Dakar	Ménages exploitant plus de 6 parcelles (%)	%	2018	0.000000
16	Diourbel	Ménages exploitant plus de 6 parcelles (%)	%	2017	22.778620
17	Diourbel	Ménages exploitant plus de 6 parcelles (%)	%	2018	16.508539
28	Fatick	Ménages exploitant plus de 6 parcelles (%)	%	2017	9.772102
29	Fatick	Ménages exploitant plus de 6 parcelles (%)	%	2018	9.772102

In [11]:

```
R = list(np.arange(28))
R = Remove(R,[0,10,11, 16])
data3 = data3[['Value']]
data3.index = R
```

Nombre moyen de parcelles par ménage pour toutes les régions

In [12]:

```
data4 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Nombre moyen de parcelles par ménage'].reset_index(drop = True)
data4.head()
```

Out[12]:

	region	indicateur	Unit	Date	Value
0	Dakar	Nombre moyen de parcelles par ménage	Unités	2017	1.208000
1	Dakar	Nombre moyen de parcelles par ménage	Unités	2018	1.041970
2	Diourbel	Nombre moyen de parcelles par ménage	Unités	2017	4.337563
3	Diourbel	Nombre moyen de parcelles par ménage	Unités	2018	3.816028
4	Fatick	Nombre moyen de parcelles par ménage	Unités	2017	3.295946

In [13]:

```
data4 = data4[['Value']].reset_index(drop = True)
```

Superficie moyenne des parcelles par ménage pour toutes les régions

In [14]:

```
data5 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Superficie moyenne des parcelles par ménage (Ha)']
].reset_index(drop = True)
data5.head()
```

Out[14]:

	region	indicateur	Unit	Date	Value
0	Dakar	Superficie moyenne des parcelles par ménage (Ha)	Ha	2017	0.462637
1	Dakar	Superficie moyenne des parcelles par ménage (Ha)	Ha	2018	0.465369
2	Diourbel	Superficie moyenne des parcelles par ménage (Ha)	Ha	2017	3.078930
3	Diourbel	Superficie moyenne des parcelles par ménage (Ha)	Ha	2018	4.645847
4	Fatick	Superficie moyenne des parcelles par ménage (Ha)	Ha	2017	2.884257

In [15]:

```
data5 = data5[['Value']].reset_index(drop = True)
```

Taille moyenne des ménages pour toutes les régions

In [16]:

```
data6 = Agriculture_Data.loc[Agriculture_Data['indicateur'] == 'Taille moyenne des ménages'].reset_index(drop = True)
data6.head()
```

Out[16]:

	region	indicateur	Unit	Date	Value
0	Dakar	Taille moyenne des ménages	Individus	2017	4.904000
1	Dakar	Taille moyenne des ménages	Individus	2018	5.363169
2	Diourbel	Taille moyenne des ménages	Individus	2017	9.519451
3	Diourbel	Taille moyenne des ménages	Individus	2018	10.407418
4	Fatick	Taille moyenne des ménages	Individus	2017	7.974183

In [17]:

```
data6 = data6[['Value']].reset_index(drop = True)
```

Assemblage des indicateurs sur l'agriculture pour en faire un seul jeu de donnée

In [18]:

```
Agriculture_Data = pd.concat([data1, data2, data3, data4, data5, data6], axis=1)
```

In [19]:

```
Agriculture_Data.columns= ['region', 'Date', 'Ménages exploitant moins de 3 parcelles (%)',
'Ménages exploitant 3 à 5 parcelles (%)',
'Ménages exploitant plus de 6 parcelles (%)',
'Nombre moyen de parcelles par ménage',
'Superficie moyenne des parcelles par ménage (Ha)',
'Taille moyenne des ménages']
```

In [20]:

```
Agriculture_Data.head()
```

Out[20]:

	region	Date	Ménages exploitant moins de 3 parcelles (%)	Ménages exploitant 3 à 5 parcelles (%)	Ménages exploitant plus de 6 parcelles (%)	Nombre moyen de parcelles par ménage	Superficie moyenne des parcelles par ménage (Ha)	Taille moyenne des ménages
0	Dakar	2017	96.799999	NaN	NaN	1.208000	0.462637	4.904000
1	Dakar	2018	100.000000	0.000000	0.000000	1.041970	0.465369	5.363169
2	Diourbel	2017	12.540621	64.680761	22.778620	4.337563	3.078930	9.519451
3	Diourbel	2018	25.740260	57.751203	16.508539	3.816028	4.645847	10.407418
4	Fatick	2017	37.248814	52.979088	9.772102	3.295946	2.884257	7.974183

Décrire les données

In [21]:

```
Agriculture_Data.describe()
```

Out[21]:

	Date	Ménages exploitant moins de 3 parcelles (%)	Ménages exploitant 3 à 5 parcelles (%)	Ménages exploitant plus de 6 parcelles (%)	Nombre moyen de parcelles par ménage	Superficie moyenne des parcelles par ménage (Ha)	Taille moyenne des ménages
count	28.000000	28.000000	26.000000	24.000000	28.000000	28.000000	28.000000
mean	2017.500000	53.930653	40.360239	9.933108	2.770551	2.685132	9.306321
std	0.509175	28.576675	20.333939	7.777486	1.062623	1.762347	1.947172
min	2017.000000	12.540621	0.000000	0.000000	1.041970	0.462637	4.904000
25%	2017.000000	29.616007	32.247736	4.920405	2.156988	1.027505	8.530413
50%	2017.500000	54.395512	42.953219	9.456790	2.665345	2.874809	9.246027
75%	2018.000000	71.796966	57.275926	13.710308	3.550553	3.756447	10.538056
max	2018.000000	100.000000	67.115271	24.925579	4.492590	8.034989	12.344740

Vérifier l'existence des valeurs manquantes

In [22]:

```
Agriculture_Data.isnull().sum()
```

Out[22]:

region	0
Date	0
Ménages exploitant moins de 3 parcelles (%)	0
Ménages exploitant 3 à 5 parcelles (%)	2
Ménages exploitant plus de 6 parcelles (%)	4
Nombre moyen de parcelles par ménage	0
Superficie moyenne des parcelles par ménage (Ha)	0
Taille moyenne des ménages	0
dtype: int64	

On constate que quatre valeurs manquent pour l'indicateur 'Ménages exploitant plus de 6 parcelles (%)' et deux pour l'indicateur 'Ménages exploitant 3 à 5 parcelles (%)'.

Remplacement des valeurs manquantes

Les valeurs manquantes des indicateurs 'Ménages exploitant plus de 6 parcelles (%)' et 'Ménages exploitant 3 à 5 parcelles (%)' sont remplacées par la moyenne de la colonne. Le remplacement des valeurs manquantes nous permet de mettre en place un jeu de donnée utile dans le futur pour la mise en place d'un model de machine learning qui fonctionne.

Remplacement de NaN dans la colonne 'Ménages exploitant plus de 6 parcelles (%)'

In [23]:

```
Mean = Agriculture_Data[['Ménages exploitant plus de 6 parcelles (%)']].mean(axis = 0)
Agriculture_Data[['Ménages exploitant plus de 6 parcelles (%)']] = Agriculture_Data[['Ménages exploitant plus de 6 parcelles (%)']].replace(np.nan,Mean)
```

Remplacement de NaN dans la colonne 'Ménages exploitant 3 à 5 parcelles (%)'

In [24]:

```
MEan = Agriculture_Data[['Ménages exploitant 3 à 5 parcelles (%)']].mean(axis = 0)
Agriculture_Data[['Ménages exploitant 3 à 5 parcelles (%)']] = Agriculture_Data[['Ménages exploitant 3 à 5 parcelles (%)']].replace(np.nan,MEan)
```

In [25]:

```
Agriculture_Data.isnull().sum()
```

Out[25]:

```
region                0
Date                  0
Ménages exploitant moins de 3 parcelles (%)  0
Ménages exploitant 3 à 5 parcelles (%)      0
Ménages exploitant plus de 6 parcelles (%)  0
Nombre moyen de parcelles par ménage      0
Superficie moyenne des parcelles par ménage (Ha)  0
Taille moyenne des ménages                0
dtype: int64
```

Jeu de donnée final sur l'agriculture

In [26]:

```
Agriculture_Data.head()
```

Out[26]:

	region	Date	Ménages exploitant moins de 3 parcelles (%)	Ménages exploitant 3 à 5 parcelles (%)	Ménages exploitant plus de 6 parcelles (%)	Nombre moyen de parcelles par ménage	Superficie moyenne des parcelles par ménage (Ha)	Taille moyenne des ménages
0	Dakar	2017	96.799999	40.360239	9.933108	1.208000	0.462637	4.904000
1	Dakar	2018	100.000000	0.000000	0.000000	1.041970	0.465369	5.363169
2	Diourbel	2017	12.540621	64.680761	22.778620	4.337563	3.078930	9.519451
3	Diourbel	2018	25.740260	57.751203	16.508539	3.816028	4.645847	10.407418
4	Fatick	2017	37.248814	52.979088	9.772102	3.295946	2.884257	7.974183

Vérifier le type de valeurs pour chaque variable

In [27]:

```
Agriculture_Data.dtypes
```

Out[27]:

```
region                object
Date                  int64
Ménages exploitant moins de 3 parcelles (%) float64
Ménages exploitant 3 à 5 parcelles (%)      float64
Ménages exploitant plus de 6 parcelles (%)  float64
Nombre moyen de parcelles par ménage      float64
Superficie moyenne des parcelles par ménage (Ha) float64
Taille moyenne des ménages                float64
dtype: object
```

Visualisation des données sur l'agriculture

In [28]:

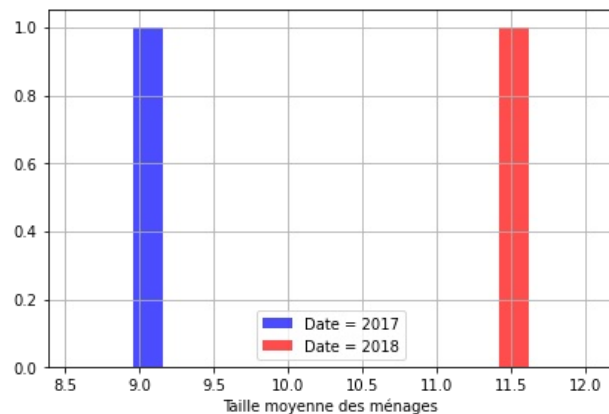
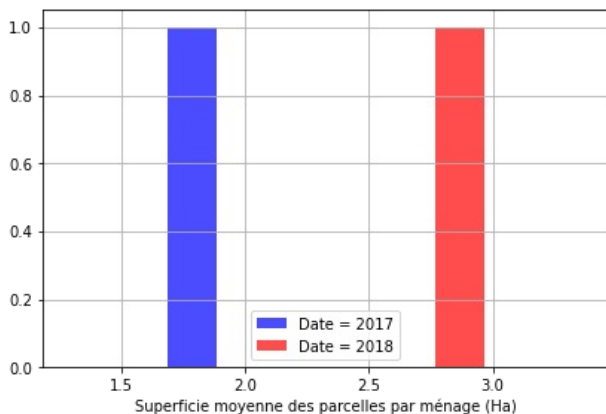
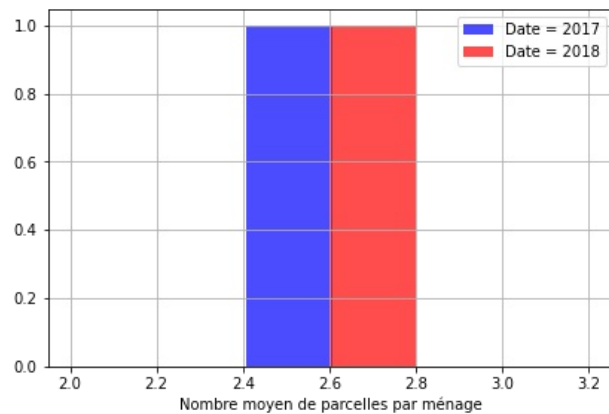
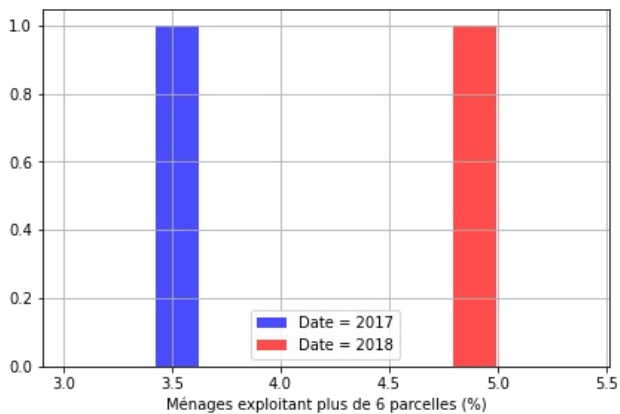
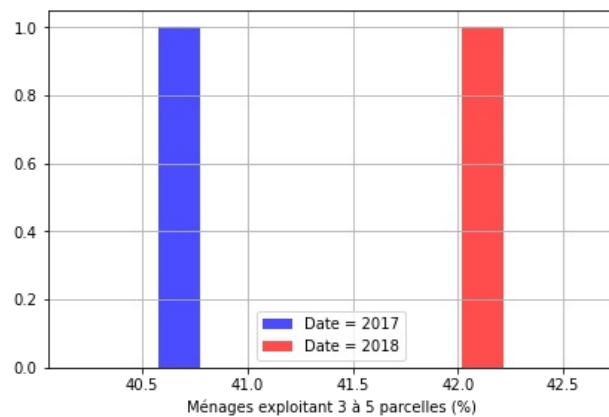
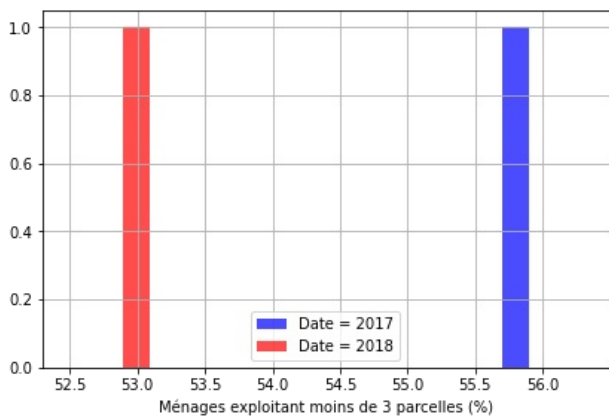
```
continous_columns = Agriculture_Data.iloc[:, 2:].columns
```

Visualisation des indicateurs prélevés en 2017 et 2018 dans la région deThies

In [29]:

```
plt.figure(figsize=(15, 15))

for i, column in enumerate(continous_columns, 1):
    plt.subplot(3, 2, i)
    Agriculture_Data[Agriculture_Data["region"] == 'Thies'][Agriculture_Data["Date"] == 2017][column].hist(bins=5
, color='blue', label='Date = 2017', alpha = 0.7)
    Agriculture_Data[Agriculture_Data["region"] == 'Thies'][Agriculture_Data["Date"] == 2018][column].hist(bins=5
, color='red', label='Date = 2018', alpha = 0.7)
    plt.legend()
    plt.xlabel(column)
```



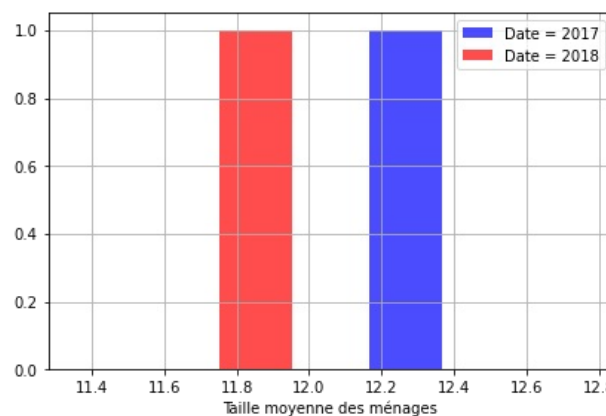
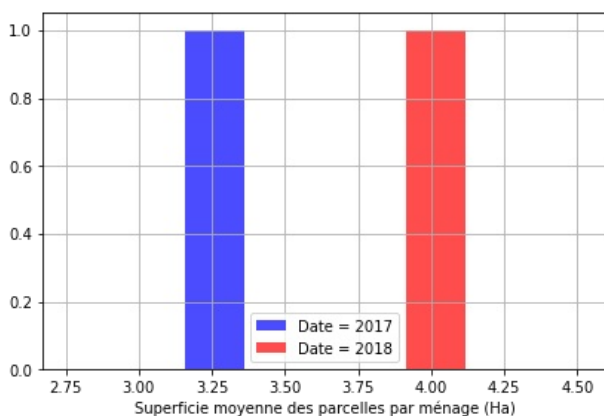
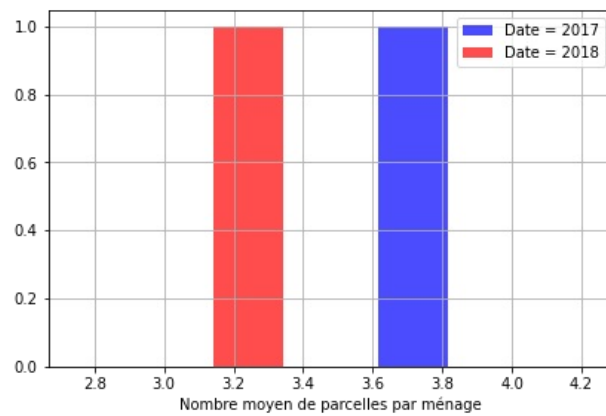
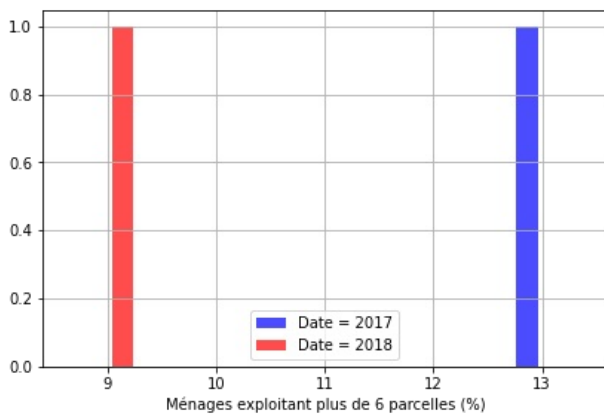
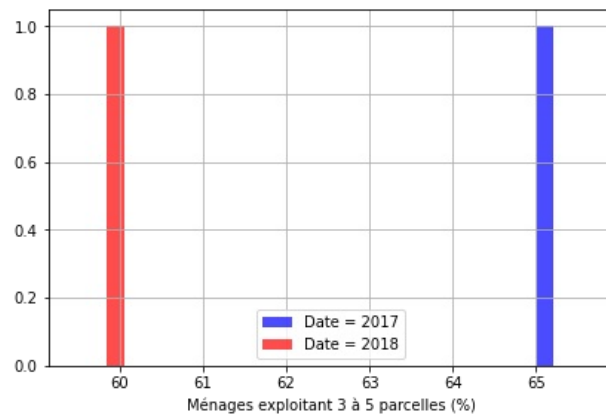
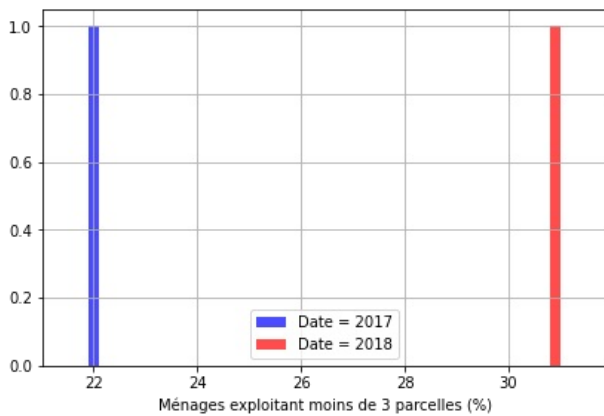
- On observe à travers la première figure que le nombre de ménages exploitant au moins de 3 parcelles dans la région de Thiés en 2018 est en réduction par rapport celui de l'année 2017.
- Pour ce qui est des autres indicateurs, on observe une augmentation en 2018 par rapport aux indicateurs enregistrés en 2017.

Visualisation des indicateurs prélevés en 2017 et 2018 dans la région de Kolda

In [30]:

```
plt.figure(figsize=(15, 15))

for i, column in enumerate(continous_columns, 1):
    plt.subplot(3, 2, i)
    Agriculture_Data[Agriculture_Data["region"] == 'Kolda'][Agriculture_Data["Date"] == 2017][column].hist(bins=5,
    color='blue', label='Date = 2017', alpha = 0.7)
    Agriculture_Data[Agriculture_Data["region"] == 'Kolda'][Agriculture_Data["Date"] == 2018][column].hist(bins=5,
    color='red', label='Date = 2018', alpha = 0.7)
    plt.legend()
    plt.xlabel(column)
```



- Concernant les indicateurs enregistrés dans la région de Kolda en 2018, la grande majorité sont en réduction par rapport aux indicateurs enregistrés en 2017. Seuls les indicateurs 'Ménages exploitant moins de 3 parcelles (%)' et 'Superficie moyenne des parcelles par ménage (Ha)' qui sont en hausse.

Exploitation des données sur l'urbanisation

Importer les données

In [32]:

```
Urbanization_Data = pd.read_csv('data/urbanization_data/Urbanisation_Data.csv')
```

In [33]:

```
Urbanization_Data.head()
```

Out[33]:

	region	indicator	Unit	Date	Value
0	Dakar	Taux d'urbanisation	NaN	1970	96.188240
1	Dakar	Taux d'urbanisation	NaN	1971	96.205776
2	Dakar	Taux d'urbanisation	NaN	1972	96.223231
3	Dakar	Taux d'urbanisation	NaN	1973	96.240606
4	Dakar	Taux d'urbanisation	NaN	1974	96.257900

In [34]:

```
np.unique( Urbanization_Data['indicator'])
```

Out[34]:

```
array(['Effectif de la population', 'Nombre de ménages ruraux',  
      'Nombre de ménages urbains', 'Population rurale',  
      'Population urbaine', 'Taille moyenne des ménages',  
      'Taille moyenne des ménages ruraux',  
      'Taille moyenne des ménages urbains', 'Taux d'urbanisation'],  

```

Visualisation de la variation du taux d'urbanisation dans quelques régions

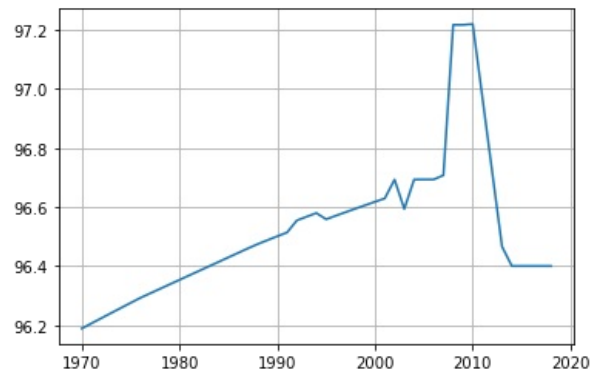
Visualisation de la variation du taux d'urbanisation dans la région de Dakar de 1970 à 2020

In [35]:

```
dat1 = Urbanization_Data[Urbanization_Data['indicator'] == "Taux d'urbanisation"][Urbanization_Data['region'] == 'Dakar']
```

In [36]:

```
plt.plot(dat1['Date'], dat1['Value'])  
plt.grid()  
plt.show()
```



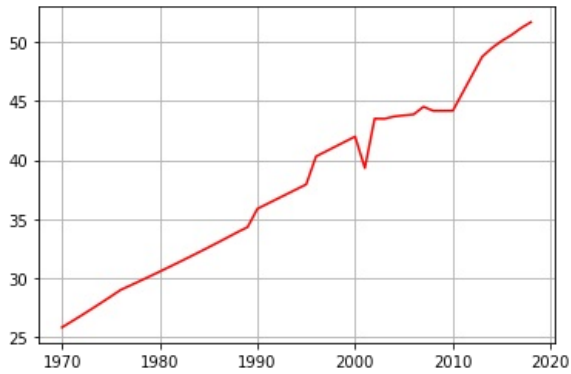
Visualisation de la variation du taux d'urbanisation dans la région de Thies de 1970 à 2020

In [37]:

```
dat1 = Urbanization_Data[Urbanization_Data['indicator'] == "Taux d'urbanisation"][Urbanization_Data['region'] == 'Thies']
```

In [38]:

```
plt.plot(dat1['Date'], dat1['Value'], color = 'r')  
plt.grid()  
plt.show()
```



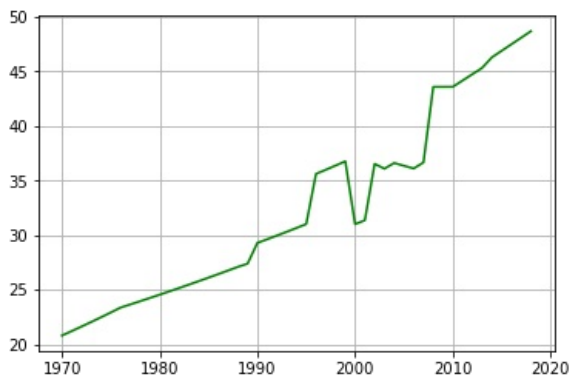
Visualisation de la variation du taux d'urbanisation dans la région de Saint-Louis de 1970 à 2020

In [39]:

```
dat1 = Urbanization_Data[Urbanization_Data['indicator'] == "Taux d'urbanisation"][Urbanization_Data['region'] == 'Saint Louis']
```

In [40]:

```
plt.plot(dat1['Date'], dat1['Value'], color = 'g')  
plt.grid()  
plt.show()
```



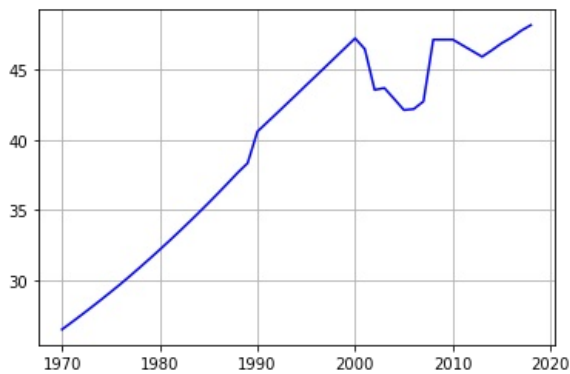
Visualisation de la variation du taux d'urbanisation dans la région de Ziguinchor de 1970 à 2020

In [41]:

```
dat1 = Urbanization_Data[Urbanization_Data['indicator'] == "Taux d'urbanisation"][Urbanization_Data['region'] == 'Ziguinchor']
```

In [42]:

```
plt.plot(dat1['Date'], dat1['Value'], color = 'b')
plt.grid()
plt.show()
```



- On observe à travers les figures ci-dessous que le taux d'urbanisation ne cesse d'augmenter dans les régions de Dakar, Thies, Saint Louis et Ziguinchor depuis les années 1970 jusqu'en 2020. En ne considérant que les régions Thies, Saint Louis et Ziguinchor, on voit que le taux d'urbanisation a augmenté respectivement d'environ 25 %, 30 % et 22 % dans ces régions. L'augmentation du taux d'urbanisation dans ces régions entraîne une augmentation du menace de déforestation.
- En ce qui concerne la région de Dakar, elle enregistre le taux d'urbanisation le plus élevé (plus de 90 %) depuis 1970 jusqu'en 2020 faisant d'elle la région la plus menace par la déforestation.

Sélectionner les indicateurs prélevés en 2017 et 2018

In [43]:

```
array = [2017, 2018]
Urbanization_Data = Urbanization_Data.loc[Urbanization_Data['Date'].isin(array)]
```

In [44]:

```
Urbanization_Data.head()
```

Out[44]:

	region	indicator	Unit	Date	Value
45	Dakar	Taux d'urbanisation	NaN	2017	9.640002e+01
46	Dakar	Taux d'urbanisation	NaN	2018	9.639996e+01
92	Dakar	Effectif de la population	NaN	2017	3.529300e+06
93	Dakar	Effectif de la population	NaN	2018	3.630324e+06
139	Dakar	Population urbaine	NaN	2017	3.402246e+06

Lecture des données et Transformation des données

In [45]:

```
np.unique(Urbanization_Data[['indicator']])
```

Out[45]:

```
array(['Effectif de la population', 'Population rurale',
      'Population urbaine', 'Taux d'urbanisation'], dtype=object)
```

Effectif de la population dans chaque région

In [46]:

```
Data1 = Urbanization_Data.loc[Urbanization_Data['indicator'] == 'Effectif de la population'].reset_index(drop = True)
Data1.head()
```

Out[46]:

	region	indicator	Unit	Date	Value
0	Dakar	Effectif de la population	NaN	2017	3529300.0
1	Dakar	Effectif de la population	NaN	2018	3630324.0
2	Diourbel	Effectif de la population	NaN	2017	1692967.0
3	Diourbel	Effectif de la population	NaN	2018	1746496.0
4	Fatick	Effectif de la population	NaN	2017	813542.0

In [47]:

```
Data1 = Data1[['region', 'Date', 'Value']].reset_index(drop = True)
```

Population rurale dans chaque région

In [48]:

```
Data2 = Urbanization_Data.loc[Urbanization_Data['indicator'] == 'Population rurale'].reset_index(drop = True)
Data2.head()
```

Out[48]:

	region	indicator	Unit	Date	Value
0	Dakar	Population rurale	NaN	2017	127054.0
1	Dakar	Population rurale	NaN	2018	130693.0
2	Diourbel	Population rurale	NaN	2017	1418707.0
3	Diourbel	Population rurale	NaN	2018	1463564.0
4	Fatick	Population rurale	NaN	2017	676053.0

In [49]:

```
Data2 = Data2[['Value']].reset_index(drop = True)
```

Population urbaine dans chaque région

In [50]:

```
Data3 = Urbanization_Data.loc[Urbanization_Data['indicator'] == 'Population urbaine'].reset_index(drop = True)
Data3.head()
```

Out[50]:

	region	indicator	Unit	Date	Value
0	Dakar	Population urbaine	NaN	2017	3402246.0
1	Dakar	Population urbaine	NaN	2018	3499631.0
2	Diourbel	Population urbaine	NaN	2017	274260.0
3	Diourbel	Population urbaine	NaN	2018	282932.0
4	Fatick	Population urbaine	NaN	2017	137489.0

In [51]:

```
Data3 = Data3[['Value']].reset_index(drop = True)
```

Taux d'urbanisation dans chaque région

In [52]:

```
Data4 = Urbanization_Data.loc[Urbanization_Data['indicator'] == "Taux d'urbanisation"].reset_index(drop = True)
Data4.head()
```

Out[52]:

	region	indicator	Unit	Date	Value
0	Dakar	Taux d'urbanisation	NaN	2017	96.400023
1	Dakar	Taux d'urbanisation	NaN	2018	96.399963
2	Diourbel	Taux d'urbanisation	NaN	2017	16.199961
3	Diourbel	Taux d'urbanisation	NaN	2018	16.199980
4	Fatick	Taux d'urbanisation	NaN	2017	16.900049

In [53]:

```
Data4 = Data4[['Value']].reset_index(drop = True)
```

Assemblage des indicateurs sur l'urbanisation pour en faire un seul jeu de données

In [54]:

```
Urbanization_Data = pd.concat([Data1, Data2, Data3, Data4], axis = 1)
```

In [55]:

```
Urbanization_Data.columns = ['region', 'Date', 'Effectif de la population', 'Population rurale', 'Population urbaine', 'Taux d'urbanisation']
```

Jeu de données final sur l'urbanisation

In [56]:

```
Urbanization_Data.head()
```

Out[56]:

	region	Date	Effectif de la population	Population rurale	Population urbaine	Taux d'urbanisation
0	Dakar	2017	3529300.0	127054.0	3402246.0	96.400023
1	Dakar	2018	3630324.0	130693.0	3499631.0	96.399963
2	Diourbel	2017	1692967.0	1418707.0	274260.0	16.199961
3	Diourbel	2018	1746496.0	1463564.0	282932.0	16.199980
4	Fatick	2017	813542.0	676053.0	137489.0	16.900049

Décrire les données

In [57]:

```
Urbanization_Data.describe()
```

Out[57]:

	Date	Effectif de la population	Population rurale	Population urbaine	Taux d'urbanisation
count	28.000000	2.800000e+01	2.800000e+01	2.800000e+01	28.000000
mean	2017.500000	1.106515e+06	5.910026e+05	5.155124e+05	34.535694
std	0.509175	8.374225e+05	3.301001e+05	8.639206e+05	21.036519
min	2017.000000	1.724820e+05	1.262570e+05	4.622500e+04	16.199961
25%	2017.000000	6.550860e+05	4.121172e+05	1.422680e+05	22.274875
50%	2017.500000	8.128085e+05	5.469675e+05	2.224640e+05	26.349929
75%	2018.000000	1.094949e+06	6.981380e+05	4.073462e+05	47.874969
max	2018.000000	3.630324e+06	1.463564e+06	3.499631e+06	96.400023

Vérifier l'existence des valeurs manquantes

In [58]:

```
Urbanization_Data.isna().sum()
```

Out[58]:

```
region                0
Date                  0
Effectif de la population  0
Population rurale      0
Population urbaine     0
Taux d'urbanisation    0
dtype: int64
```

Vérifier les types des données

In [59]:

```
Urbanization_Data.dtypes
```

Out[59]:

```
region                object
Date                  int64
Effectif de la population  float64
Population rurale      float64
Population urbaine     float64
Taux d'urbanisation    float64
dtype: object
```

Visualisation des données

In [60]:

```
continous_columns = Urbanization_Data.iloc[:, 2:].columns
continous_columns
```

Out[60]:

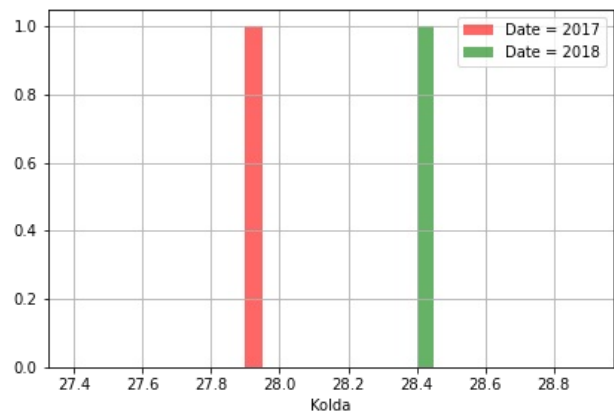
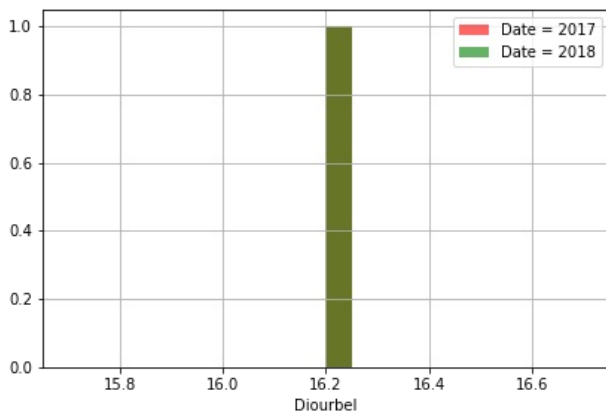
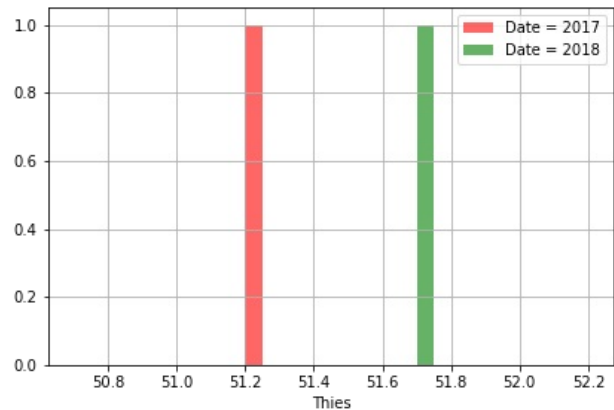
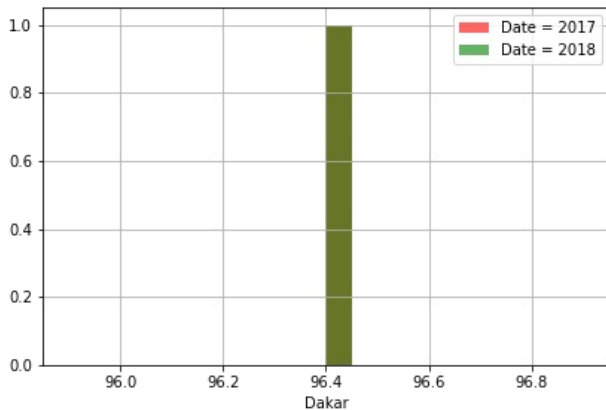
```
Index(['Effectif de la population', 'Population rurale', 'Population urbaine',
      'Taux d'urbanisation'],
      dtype='object')
```

Visualisation du taux d'urbanisation des régions de Dakar, Thies, Diourbel et Kolda en 2017 et 2018

In [61]:

```
plt.figure(figsize=(15, 15))

for i, region in enumerate(['Dakar', 'Thies', 'Diourbel', 'Kolda'], 1):
    plt.subplot(3, 2, i)
    Urbanization_Data[Urbanization_Data["region"] == region][Urbanization_Data["Date"] == 2017]["Taux d'urbanisation"].hist(bins=20, color='red', label='Date = 2017', alpha=0.6)
    Urbanization_Data[Urbanization_Data["region"] == region][Urbanization_Data["Date"] == 2018]["Taux d'urbanisation"].hist(bins=20, color='green', label='Date = 2018', alpha=0.6)
    plt.legend()
    plt.xlabel(region)
```



- On voit à travers les figures 1 et 3 que le taux d'urbanisation des les régions de Dakar et Diourbel pour les années 2017 et 2018 est resté le même. Contrairement aux régions Thiès et Kolda leurs taux d'urbanisation pour l'année 2018 est en hausse par rapport aux ceux de l'année 2017.
- Un autre constat est que la région de Dakar affiche le plus grand taux d'urbanisation avec une population à plus 90 % urbaine.

Concaténation des deux jeux de données

In [62]:

```
Data1 = Urbanization_Data[continous_columns]
Data_concat = pd.concat([Agriculture_Data, Data1], axis = 1)
```

In [63]:

```
Data_concat.head()
```

Out[63]:

	region	Date	Ménages exploitant moins de 3 parcelles (%)	Ménages exploitant 3 à 5 parcelles (%)	Ménages exploitant plus de 6 parcelles (%)	Nombre moyen de parcelles par ménage	Superficie moyenne des parcelles par ménage (Ha)	Taille moyenne des ménages	Effectif de la population	Population rurale	Population urbaine	Taux d'urbanisation
0	Dakar	2017	96.799999	40.360239	9.933108	1.208000	0.462637	4.904000	3529300.0	127054.0	3402246.0	96.400023
1	Dakar	2018	100.000000	0.000000	0.000000	1.041970	0.465369	5.363169	3630324.0	130693.0	3499631.0	96.399963
2	Diourbel	2017	12.540621	64.680761	22.778620	4.337563	3.078930	9.519451	1692967.0	1418707.0	274260.0	16.199961
3	Diourbel	2018	25.740260	57.751203	16.508539	3.816028	4.645847	10.407418	1746496.0	1463564.0	282932.0	16.199980
4	Fatick	2017	37.248814	52.979088	9.772102	3.295946	2.884257	7.974183	813542.0	676053.0	137489.0	16.900049

In []: