

# Etat\_des\_lieux\_collèges

January 21, 2023

## #État des lieux des collèges

CSV utilisés :

- Résultats détaillés au DNB
- Label numérique collèges académie de Montpellier
- Indices de position sociale dans les collèges de France métropolitaine et DROM

## 0.1 Analyses des indicateurs de résultats des collèges en occitanie

### 0.1.1 Imports des packages utiles

```
[1]: # importing necessary packages
import pandas as pd
```

```
[2]: # import des librairies pour les graphiques
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import plotly.graph_objects as go
from plotly.subplots import make_subplots
```

```
[3]: # Import necessary libraries for image
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import folium # for maps
import plotly.graph_objects as go
!pip install -U kaleido
import kaleido
from IPython.display import Image
import io
from PIL import Image as ImagePIL
```

Requirement already satisfied: kaleido in  
/home/ubuntu/anaconda3/lib/python3.9/site-packages (0.2.1)

### 0.1.2 ETL du CSV Résultats détaillés au DNB

```
[4]: # importing the csv files
df = pd.read_csv("fr-en-resultats-detailles-au-dnb.csv", sep = ";")
df
```

```
[4]:
```

	année	académie	code région	libellé région	série	\
0	2011	MONTPELLIER	76	Occitanie	Serie college	
1	2011	MONTPELLIER	76	Occitanie	Serie college	
2	2011	MONTPELLIER	76	Occitanie	Serie professionnelle	
3	2011	MONTPELLIER	76	Occitanie	Serie professionnelle	
4	2011	MONTPELLIER	76	Occitanie	Serie professionnelle	
...	...	...	...	...	...	
1723	2021	TOULOUSE	76	Occitanie	Serie professionnelle	
1724	2021	TOULOUSE	76	Occitanie	Serie generale	
1725	2021	TOULOUSE	76	Occitanie	Serie generale	
1726	2021	TOULOUSE	76	Occitanie	Serie professionnelle	
1727	2021	TOULOUSE	76	Occitanie	Serie professionnelle	

	inscrits	inscrits filles	présents	présents filles	admis	...	\
0	21265	10994.0	21057.0	10900.0	19223.0	...	
1	21265	10994.0	21057.0	10900.0	19223.0	...	
2	514	221.0	483.0	211.0	429.0	...	
3	514	221.0	483.0	211.0	429.0	...	
4	514	221.0	483.0	211.0	429.0	...	
...	...	...	...	...	...	...	
1723	1479	558.0	1335.0	500.0	983.0	...	
1724	91	35.0	65.0	26.0	36.0	...	
1725	91	35.0	65.0	26.0	36.0	...	
1726	164	56.0	135.0	49.0	91.0	...	
1727	164	56.0	135.0	49.0	91.0	...	

	admis mention assez bien	admis mention bien	admis mention très bien	\
0	6113.0	4197.0	1808.0	
1	6113.0	4197.0	1808.0	
2	138.0	41.0	2.0	
3	138.0	41.0	2.0	
4	138.0	41.0	2.0	
...	...	...	...	
1723	330.0	187.0	42.0	
1724	10.0	2.0	2.0	
1725	10.0	2.0	2.0	
1726	27.0	15.0	7.0	
1727	27.0	15.0	7.0	

	genre	âge	présence	admissibilité	\
0	Garçons	15 ans ou moins	Non présents	Non présents	

1	Garçons	15 ans ou moins	Présents	Admis
2	Filles	15 ans ou moins	Non présents	Non présents
3	Garçons	15 ans ou moins	Non présents	Non présents
4	Filles	15 ans ou moins	Présents	Non admis
...	...	...	...	...
1723	Garçons	16 ans	Présents	Admis
1724	Filles	17 ans ou plus	Présents	Admis
1725	Garçons	17 ans ou plus	Non présents	Non présents
1726	Filles	17 ans ou plus	Présents	Admis
1727	Filles	17 ans ou plus	Présents	Admis

	admissibilité détails	Résultats	Effectifs
0	f. Non présents	non_presents_garcons	114
1	b. Admis mention bien	admis_mention_bien_garcons	1803
2	f. Non présents	non_presents_filles	10
3	f. Non présents	non_presents_garcons	21
4	e. Non admis	non_admis_filles	18
...	...	...	...
1723	b. Admis mention bien	admis_mention_bien_garcons	118
1724	b. Admis mention bien	admis mention bien filles	0
1725	f. Non présents	non_presents_garcons	17
1726	d. Admis sans mention	admis sans mention filles	17
1727	a. Admis mention très bien	admis mention très bien filles	3

[1728 rows x 22 columns]

```
[5]: # afficher les infos des colonnes
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1728 entries, 0 to 1727
Data columns (total 22 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   année                                1728 non-null   int64
1   académie                             1728 non-null   object
2   code région                          1728 non-null   int64
3   libellé région                       1728 non-null   object
4   série                                1728 non-null   object
5   inscrits                             1728 non-null   int64
6   inscrits filles                      1728 non-null   float64
7   présents                             1728 non-null   float64
8   présents filles                     1728 non-null   float64
9   admis                               1728 non-null   float64
10  admis filles                         1728 non-null   float64
11  admis sans mention                   1728 non-null   float64
12  admis mention assez bien             1728 non-null   float64
```

```

13 admis mention bien          1728 non-null    float64
14 admis mention très bien    1728 non-null    float64
15 genre                      1728 non-null    object
16 âge                        1728 non-null    object
17 présence                   1728 non-null    object
18 admissibilité              1728 non-null    object
19 admissibilité détails      1728 non-null    object
20 Résultats                  1728 non-null    object
21 Effectifs                  1728 non-null    int64
dtypes: float64(9), int64(4), object(9)
memory usage: 297.1+ KB

```

```

[6]: # fonction pour transformer le type des colonnes d'un df
def formaterCol(dataframe, col, na, type):
    dataframe[col] = dataframe[col].fillna(na).astype(type)
    return dataframe

```

```

[7]: # transformer le type des colonnes float en int
# je choisis de fill des NaN par 0 puisque ça n'impactera pas les analyses
# inscrits filles
df = formaterCol(df, "inscrits filles", 0, int)
# présents
df = formaterCol(df, "présents", 0, int)
# présents filles
df = formaterCol(df, "présents filles", 0, int)
# admis
df = formaterCol(df, "admis", 0, int)
# admis filles
df = formaterCol(df, "admis filles", 0, int)
# admis sans mention
df = formaterCol(df, "admis sans mention", 0, int)
# admis mention assez bien
df = formaterCol(df, "admis mention assez bien", 0, int)
# admis mention bien
df = formaterCol(df, "admis mention bien", 0, int)
# admis mention très bien
df = formaterCol(df, "admis mention très bien", 0, int)

```

```

[8]: # infos des colonnes post transformation
df.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1728 entries, 0 to 1727
Data columns (total 22 columns):
#   Column                Non-Null Count  Dtype
---  -
0   année                 1728 non-null   int64
1   académie              1728 non-null   object

```

2	code région	1728 non-null	int64
3	libellé région	1728 non-null	object
4	série	1728 non-null	object
5	inscrits	1728 non-null	int64
6	inscrits filles	1728 non-null	int64
7	présents	1728 non-null	int64
8	présents filles	1728 non-null	int64
9	admis	1728 non-null	int64
10	admis filles	1728 non-null	int64
11	admis sans mention	1728 non-null	int64
12	admis mention assez bien	1728 non-null	int64
13	admis mention bien	1728 non-null	int64
14	admis mention très bien	1728 non-null	int64
15	genre	1728 non-null	object
16	âge	1728 non-null	object
17	présence	1728 non-null	object
18	admissibilité	1728 non-null	object
19	admissibilité détails	1728 non-null	object
20	Résultats	1728 non-null	object
21	Effectifs	1728 non-null	int64

dtypes: int64(13), object(9)  
memory usage: 297.1+ KB

```
[9]: # drop des colonnes pas très utiles
# drop code région (76 partout)
df = df.drop(columns = "code région")
# drop libellé région (Occitanie partout)
df = df.drop(columns = "libellé région")
# vérifications
df.info()
df
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1728 entries, 0 to 1727
Data columns (total 20 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   année                                1728 non-null   int64
1   académie                             1728 non-null   object
2   série                                1728 non-null   object
3   inscrits                             1728 non-null   int64
4   inscrits filles                      1728 non-null   int64
5   présents                             1728 non-null   int64
6   présents filles                     1728 non-null   int64
7   admis                                1728 non-null   int64
8   admis filles                        1728 non-null   int64
9   admis sans mention                  1728 non-null   int64
10  admis mention assez bien            1728 non-null   int64
```

11	admis mention bien	1728 non-null	int64
12	admis mention très bien	1728 non-null	int64
13	genre	1728 non-null	object
14	âge	1728 non-null	object
15	présence	1728 non-null	object
16	admissibilité	1728 non-null	object
17	admissibilité détails	1728 non-null	object
18	Résultats	1728 non-null	object
19	Effectifs	1728 non-null	int64

dtypes: int64(12), object(8)

memory usage: 270.1+ KB

```
[9]:
```

	année	académie	série	inscrits	inscrits filles	\
0	2011	MONTPELLIER	Serie college	21265	10994	
1	2011	MONTPELLIER	Serie college	21265	10994	
2	2011	MONTPELLIER	Serie professionnelle	514	221	
3	2011	MONTPELLIER	Serie professionnelle	514	221	
4	2011	MONTPELLIER	Serie professionnelle	514	221	
...	...	...	...	...	...	
1723	2021	TOULOUSE	Serie professionnelle	1479	558	
1724	2021	TOULOUSE	Serie generale	91	35	
1725	2021	TOULOUSE	Serie generale	91	35	
1726	2021	TOULOUSE	Serie professionnelle	164	56	
1727	2021	TOULOUSE	Serie professionnelle	164	56	

	présents	présents filles	admis	admis filles	admis sans mention	\
0	21057	10900	19223	10088	7105	
1	21057	10900	19223	10088	7105	
2	483	211	429	193	248	
3	483	211	429	193	248	
4	483	211	429	193	248	
...	...	...	...	...	...	
1723	1335	500	983	370	424	
1724	65	26	36	16	22	
1725	65	26	36	16	22	
1726	135	49	91	33	42	
1727	135	49	91	33	42	

	admis mention assez bien	admis mention bien	admis mention très bien	\
0	6113	4197	1808	
1	6113	4197	1808	
2	138	41	2	
3	138	41	2	
4	138	41	2	
...	...	...	...	
1723	330	187	42	
1724	10	2	2	

1725		10	2	2
1726		27	15	7
1727		27	15	7

	genre	âge	présence	admissibilité \
0	Garçons	15 ans ou moins	Non présents	Non présents
1	Garçons	15 ans ou moins	Présents	Admis
2	Filles	15 ans ou moins	Non présents	Non présents
3	Garçons	15 ans ou moins	Non présents	Non présents
4	Filles	15 ans ou moins	Présents	Non admis
...	...	...	...	...
1723	Garçons	16 ans	Présents	Admis
1724	Filles	17 ans ou plus	Présents	Admis
1725	Garçons	17 ans ou plus	Non présents	Non présents
1726	Filles	17 ans ou plus	Présents	Admis
1727	Filles	17 ans ou plus	Présents	Admis

	admissibilité détails	Résultats	Effectifs
0	f. Non présents	non_presents_garcons	114
1	b. Admis mention bien	admis_mention_bien_garcons	1803
2	f. Non présents	non_presents_filles	10
3	f. Non présents	non_presents_garcons	21
4	e. Non admis	non_admis_filles	18
...	...	...	...
1723	b. Admis mention bien	admis_mention_bien_garcons	118
1724	b. Admis mention bien	admis mention bien filles	0
1725	f. Non présents	non_presents_garcons	17
1726	d. Admis sans mention	admis sans mention filles	17
1727	a. Admis mention très bien	admis mention très bien filles	3

[1728 rows x 20 columns]

```
[10]: # Je veux connaître le nombre d'admis par année, région et série
# Je vais retirer dupliquer le dataframe et prendre que les colonnes utiles_
      ↳ pour cette analyse
# duplication df
df2 = df[['année', 'académie', 'série', 'admis', 'inscrits']]
# retire les duplicates de résultats
df2 = df2.drop_duplicates(subset=['académie', 'inscrits'], keep='last')
# Trie la colonne année par année (ascendant : 2011 à 2021)
df2 = df2.sort_values(by="année")
# Transforme la colonne en string afin de pouvoir l'utiliser pour les graphiques
df2["année"] = df2["année"].apply(str)
# group by la région, l'année et la série afin d'avoir le nombre d'admis
result = df2.groupby(['année', 'académie', 'série', 'admis'])['admis'].count()
# print(result)
result
```

```
[10]: année  académie      série      admis
      2011  MONTPELLIER  Serie college  220      1
                                   2992      1
                                   19223     1
                                   Serie professionnelle  180      1
                                   429      1
                                   ..
      2021  TOULOUSE    Serie generale  1250      1
                                   27517     1
                                   Serie professionnelle  91      1
                                   983      1
                                   1679      1
```

Name: admis, Length: 143, dtype: int64

```
[11]: import re
aNegliger = re.compile("Serie college",re.I)
subst = "Serie generale"
def apply_reg_serie_college(str_):
    '''Fonction pour remplacer les formations de "Serie college" par "serie_
    ↳ generale"
        - paramètres :
            - str_ : le nom de la colonne dans laquelle la fonction va chercher_
    ↳ la pattern à nettoyer
        - Return la chaine de caractère nettoyé pour la colonne '''
    if pd.notna(str_):
        return re.sub(aNegliger, subst, str(str_))
    else:
        return str_

# copie du df2 pour appliquer les modifications sur la colonne
df2_filtre = df2
df2_filtre['série'] = df2['série'].apply(apply_reg_serie_college)

aNegliger2 = re.compile("Serie technologique")
def apply_reg_serie techno(str_):
    '''Fonction pour remplacer les formations de "Serie technologique" par_
    ↳ "Serie generale"
        - paramètres :
            - str_ : le nom de la colonne dans laquelle la fonction va chercher_
    ↳ la pattern à nettoyer
        - Return la chaine de caractère nettoyé pour la colonne '''
    if pd.notna(str_):
        return re.sub(aNegliger2, subst, str(str_))
    else:
        return str_

```



```
# copie du df2 pour appliquer les modifications sur la colonne
df2_filtre['série'] = df2['série'].apply(apply_reg_serie_techno)
```

```
df2_filtre
```

```
[11]:
```

	année	académie	série	admis	inscrits
1394	2011	MONTPELLIER	Serie generale	402	466
1431	2011	TOULOUSE	Serie generale	193	314
1392	2011	MONTPELLIER	Serie professionnelle	429	514
1389	2011	MONTPELLIER	Serie generale	19223	21265
1401	2011	MONTPELLIER	Serie generale	713	994
...	...	...	...	...	...
1698	2021	MONTPELLIER	Serie professionnelle	1505	1892
1696	2021	MONTPELLIER	Serie generale	25305	28530
1725	2021	TOULOUSE	Serie generale	36	91
1709	2021	MONTPELLIER	Serie generale	41	132
1727	2021	TOULOUSE	Serie professionnelle	91	164

```
[143 rows x 5 columns]
```

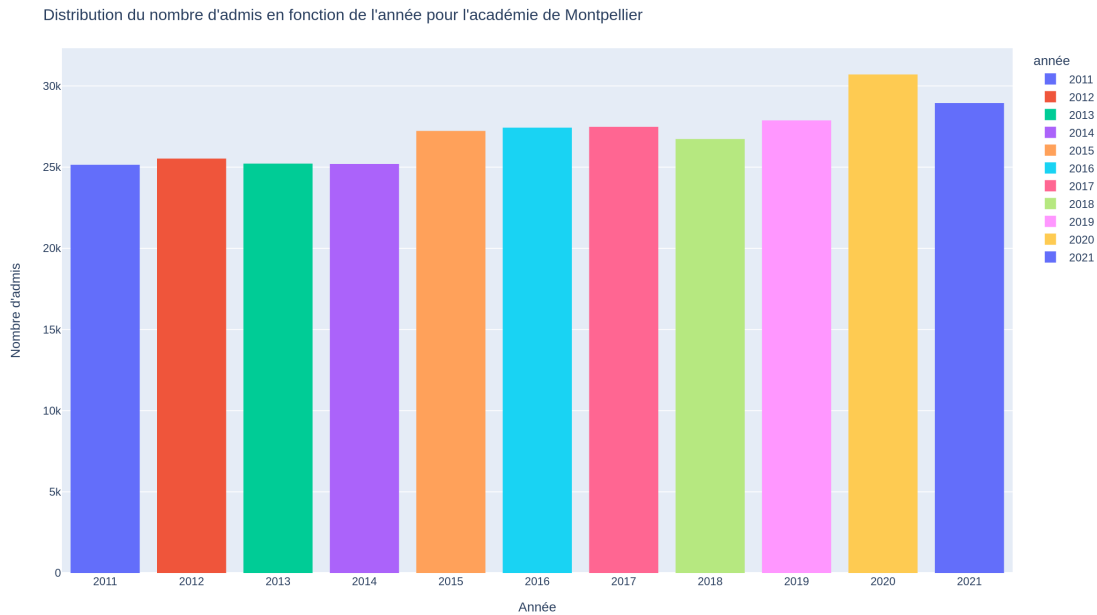
### 0.1.3 Analyses des résultats des collèges pour chaque académie

1 - Montpellier Nombre d'admis en fonction de l'année pour l'académie de Montpellier :

```
[12]: # Division par académie
# Montpellier
df2M = df2[df2['académie'] == 'MONTPELLIER']
# Total admis pour toutes les séries
fig1 = px.histogram(df2M, x="année", y="admis", color="année")
fig1.update_layout(title_text="Distribution du nombre d'admis en fonction de l'année pour l'académie de Montpellier")
fig1.update_yaxes(title_text="Nombre d'admis")
fig1.update_xaxes(title_text="Année")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

```
[12]:
```



On peut voir que le nombre d'élèves admis **augmente chaque année**. On remarque également **un pic à l'année 2020** qui pourrait être expliqué par l'allègement des examens dûs aux difficultés liées à la covid

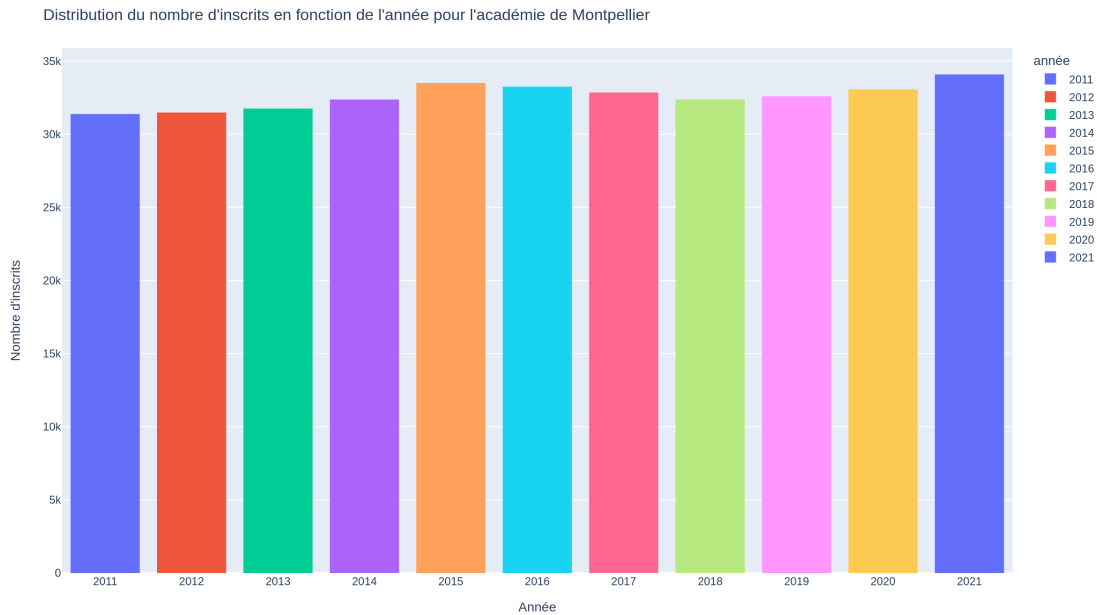
---

**Nombre total d'inscrits en fonction de l'année pour l'académie de Montpellier :**

```
[13]: # Total d'inscrits pour toutes les séries
fig2 = px.histogram(df2M, x="année", y="inscrits", color="année")
fig2.update_layout(title_text="Distribution du nombre d'inscrits en fonction de l'année pour l'académie de Montpellier")
fig2.update_yaxes(title_text="Nombre d'inscrits")
fig2.update_xaxes(title_text="Année")
fig2.show()

image = fig2.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[13]:

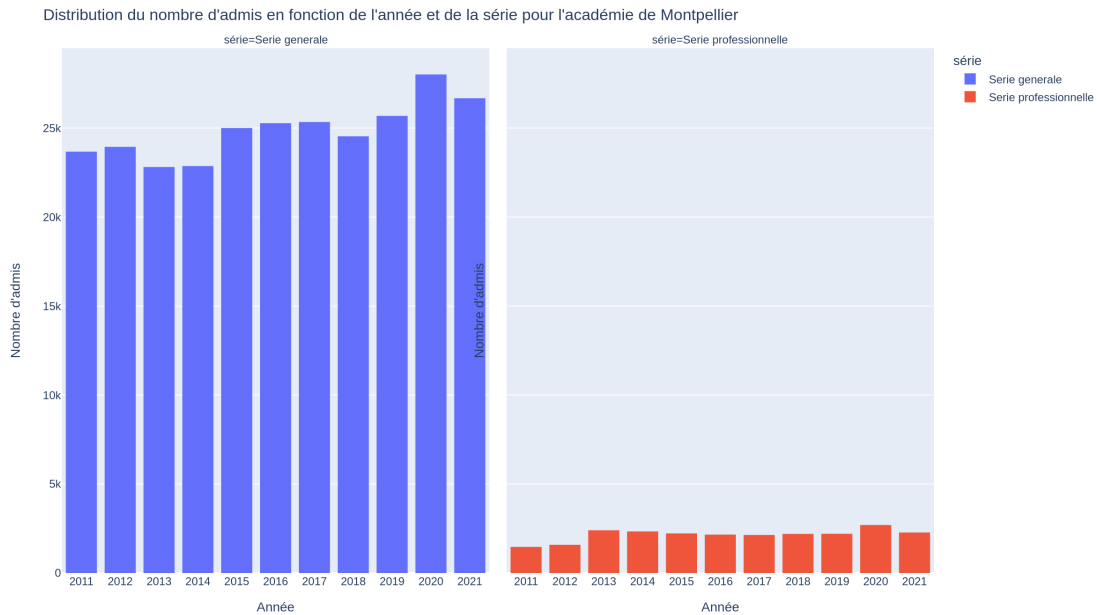


Nombre d'admis en fonction de l'année et de la série pour l'académie de Montpellier :

```
[14]: # Montpellier : nb admis en fct de l'année et série
fig3 = px.histogram(df2M, x="année", y="admis", color="série",
    ↪ facet_col="série")
fig3.update_layout(title_text="Distribution du nombre d'admis en fonction de
    ↪ l'année et de la série pour l'académie de Montpellier")
fig3.update_yaxes(title_text="Nombre d'admis")
fig3.update_xaxes(title_text="Année")
# fig3.show()

image = fig3.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[14]:



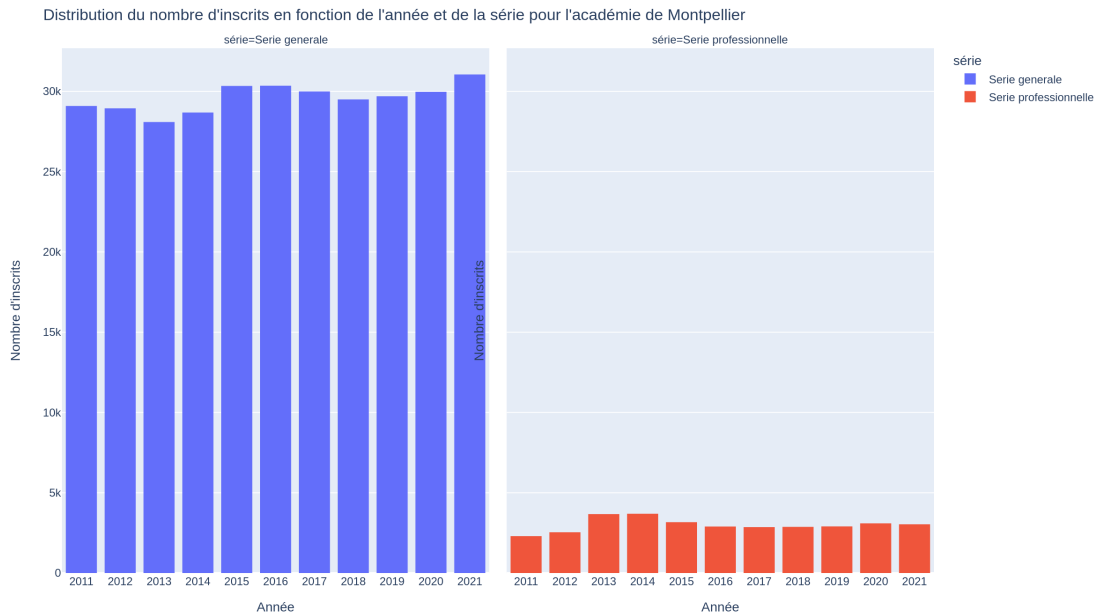
On constate que le nombre d'élèves admis en série générale est **nettement supérieur** aux élèves en série professionnelle. Il n'y a pas autant de lycée de chaque série.

Nombre d'inscrits en fonction de l'année et de la série pour l'académie de Montpellier :

```
[15]: # Montpellier : nb d'inscrits en fct de l'année et série
fig4 = px.histogram(df2M, x="année", y="inscrits", color="série",
                    facet_col="série")
fig4.update_layout(title_text="Distribution du nombre d'inscrits en fonction de
                    l'année et de la série pour l'académie de Montpellier")
fig4.update_yaxes(title_text="Nombre d'inscrits")
fig4.update_xaxes(title_text="Année")
# fig4.show()

image = fig4.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[15]:



On peut clairement voir **l'évolution du nombre d'élèves inscrits** par rapport aux **admis au fil des années**, suivant leur **série**.

On peut voir que pour l'académie de Montpellier, les élèves en **série générale** :

- le nombre **d'admis** en **2011** est de **23 671**
- le nombre **d'admis** en **2021** est de **26 669**
- le nombre **d'inscrits** en **2011** est de **29 084**
- le nombre **d'inscrits** en **2021** est de **31 050**

Et pour les élèves en **série professionnelle** :

- le nombre **d'admis** en **2011** est de **1469**
- le nombre **d'admis** en **2021** est de **2279**
- le nombre **d'inscrits** en **2011** est de **2299**
- le nombre **d'inscrits** en **2021** est de **3036**

Au total :

- le nombre **d'admis** en **2011** est de **25 140**
- le nombre **d'admis** en **2021** est de **28 948**
- le nombre **d'inscrits** en **2011** est de **31 383**
- le nombre **d'inscrits** en **2021** est de **34 086**

```
[16]: # ratio admis série générale
rasg = (26669 - 23671)/23671
print(rasg)
# ratio inscrits série générale
risg = (31050 - 29084)/29084
```

```

print(risg)
# ratio admis série professionnelle
rasp = (2279 - 1469)/1469
print(rasp)
# ratio inscrits série professionnelle
risp = (3036 - 2299)/2299
print(risp)
# ratio admis total
rat = (28948 - 25140)/25140
print(rat)
# ratio inscrit total
rit = (34086 - 31383)/31383
print(rit)

```

```

0.12665286637657894
0.06759730435978545
0.5513955071477196
0.32057416267942584
0.15147175815433572
0.08612943313258771

```

On constate que pour l'académie de Montpellier, les élèves en **série générale**, en **10 ans** (2011 à 2021) :

- le nombre d'admis a augmenté de **12,66%**
- le nombre d'inscrits a augmenté de **6,76%**

Et pour les élèves en **série professionnelle** :

- le nombre d'admis a augmenté de **55,14%**
- le nombre d'inscrits a augmenté de **32,06%**

Enfin, au total on a :

- le nombre d'admis a augmenté de **15,15%**
- le nombre d'inscrits a augmenté de **8,61%**

Etant donné que **le nombre d'admis augmente énormément par rapport au nombre d'inscrits**, on en déduit que le niveau global des collégiens s'est amélioré en 10 ans.

---

## Nombre d'inscrits en séries générales et technologiques à Montpellier en 2011

```

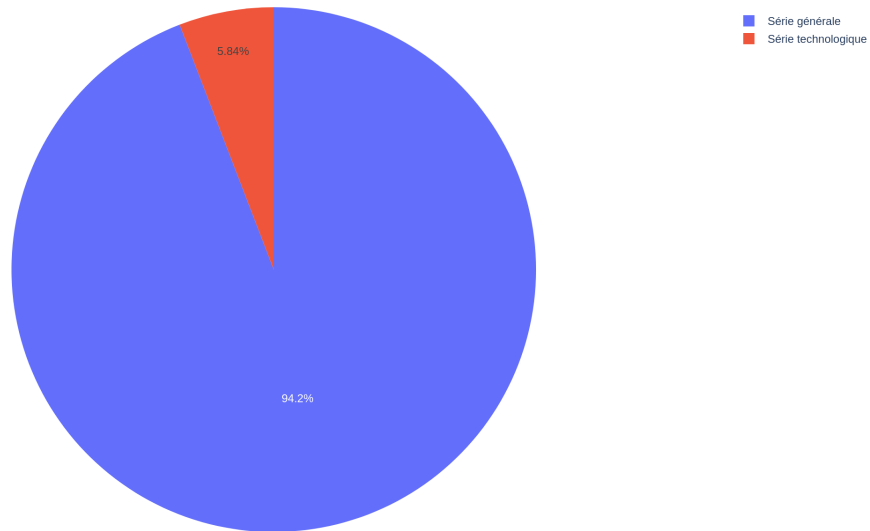
[17]: # ratio inscrits série générale / série technologique pour 2011
asgvals = [1469, 23671]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'inscrits en séries_
↳générales et technologiques pour l'académie de Montpellier en 2011")
# fg1.show()

```

```
image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[17]:

Distribution du nombre d'inscrits en séries générales et technologiques pour l'académie de Montpellier en 2011



Le nombre d'inscrits en série générale est **largement plus élevé** par rapport aux inscrits en série technologique en 2011.

---

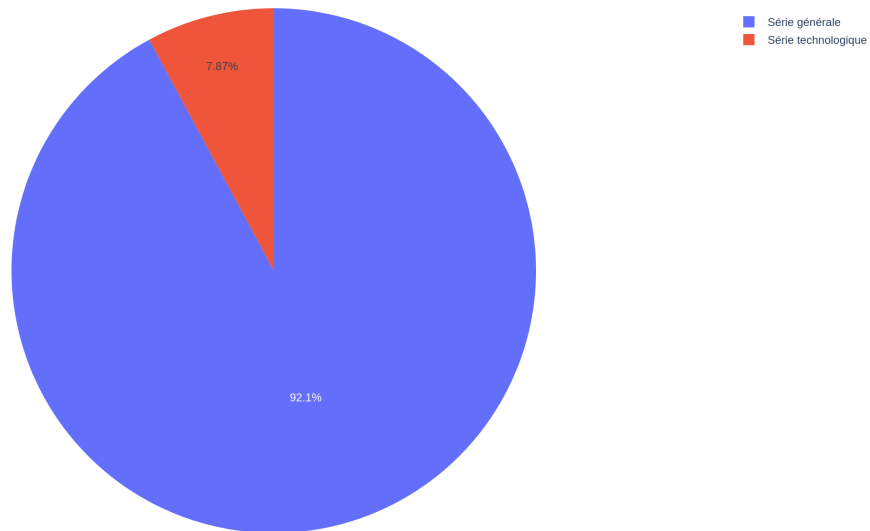
### Nombre d'inscrits en séries générales et technologiques à Montpellier en 2021

```
[18]: # ratio inscrits série générale / série technologique pour 2021
asgvals = [2279, 26669]
asglabels = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabels)
fg1.update_layout(title_text="Distribution du nombre d'inscrits en séries_
↳générales et technologiques pour l'académie de Montpellier en 2021")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[18]:

Distribution du nombre d'inscrits en séries générales et technologiques pour l'académie de Montpellier en 2021



Il y'a une **plus grande proportion d'élèves inscrits en série technologique en 2021** qu'en **2011** même si elle reste plutôt faible.

---

#### Nombre d'admis en séries générales et technologiques à Montpellier en 2011

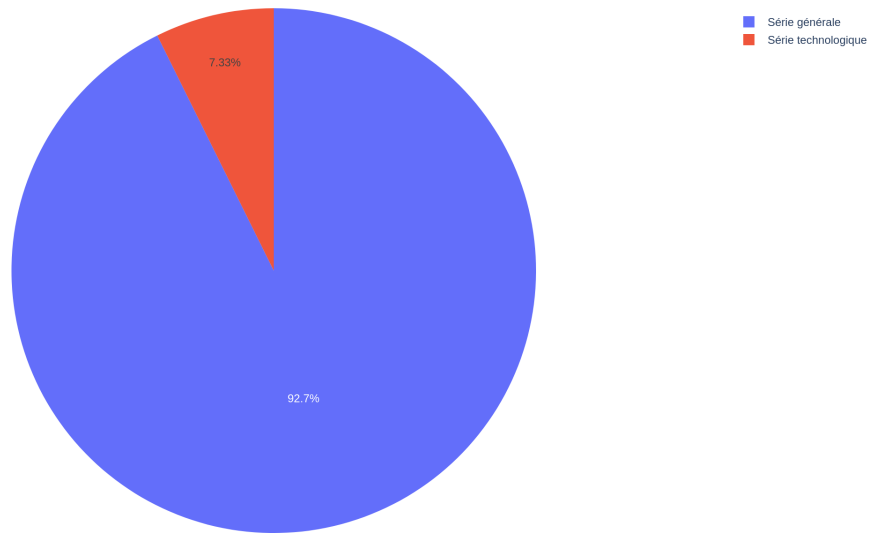
```
[19]: # ratio admis série générale / série technologique pour 2011
asgvals = [2299, 29084]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'admis en séries_
↳générales et technologiques pour l'académie de Montpellier en 2011")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[19]:



Distribution du nombre d'admis en séries générales et technologiques pour l'académie de Montpellier en 2011



---

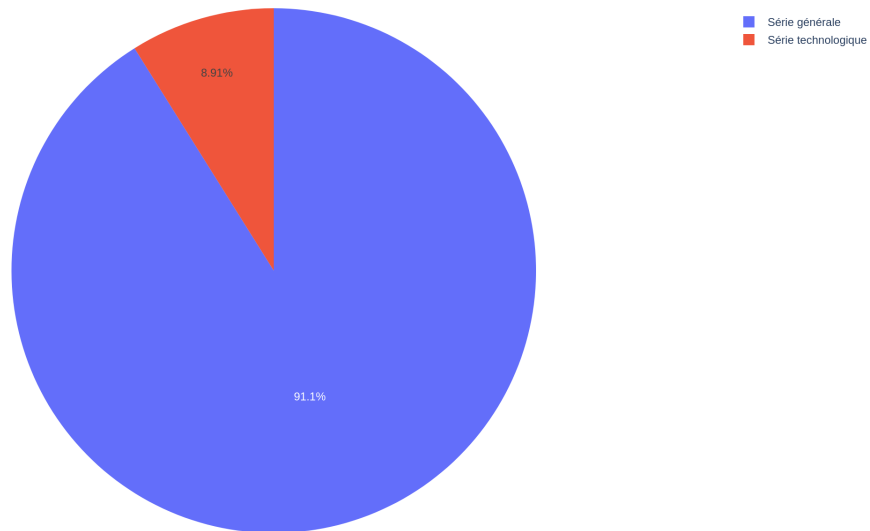
### Nombre d'admis en séries générales et technologiques à Montpellier en 2021

```
[20]: # ratio inscrits série générale / série technologique pour 2021
asgvals = [3036, 31050]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'admis en séries_
↳générales et technologiques pour l'académie de Montpellier en 2021")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[20]:

Distribution du nombre d'admis en séries générales et technologiques pour l'académie de Montpellier en 2021



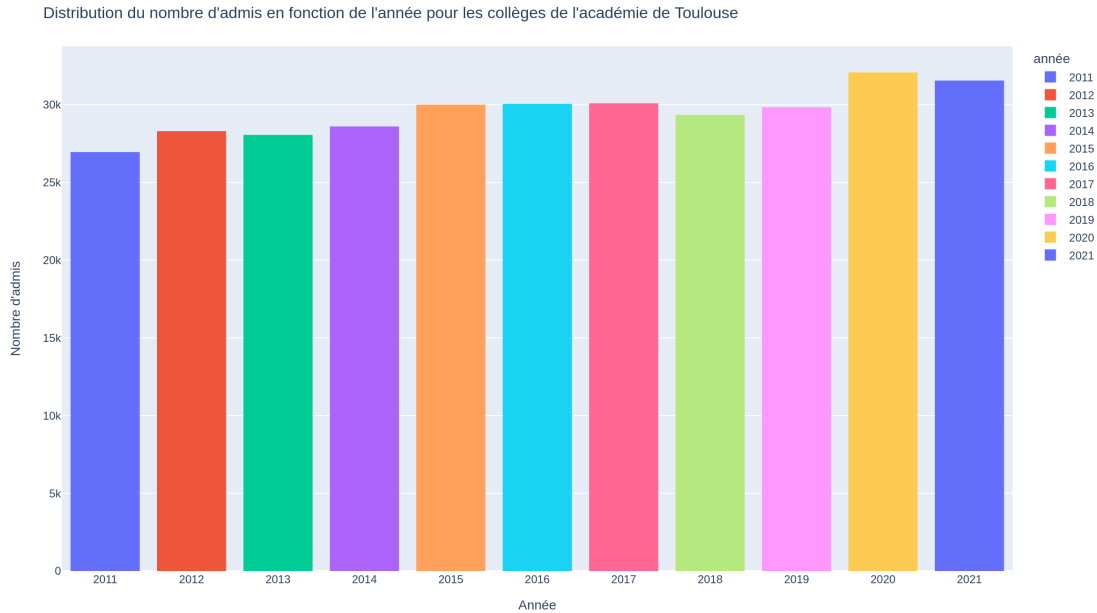
Même remarques pour les admis

## 2 - Toulouse Nombre d'admis en fonction de l'année au collège à Toulouse

```
[21]: # Division par académie
# Toulouse
df2T = df2[df2['académie'] == 'TOULOUSE']
# Total admis pour toutes les séries
fig1 = px.histogram(df2T, x="année", y="admis", color="année")
fig1.update_layout(title_text="Distribution du nombre d'admis en fonction de l'année pour les collèges de l'académie de Toulouse")
fig1.update_yaxes(title_text="Nombre d'admis")
fig1.update_xaxes(title_text="Année")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[21]:



Même remarque que pour les résultats de l'académie de Montpellier.

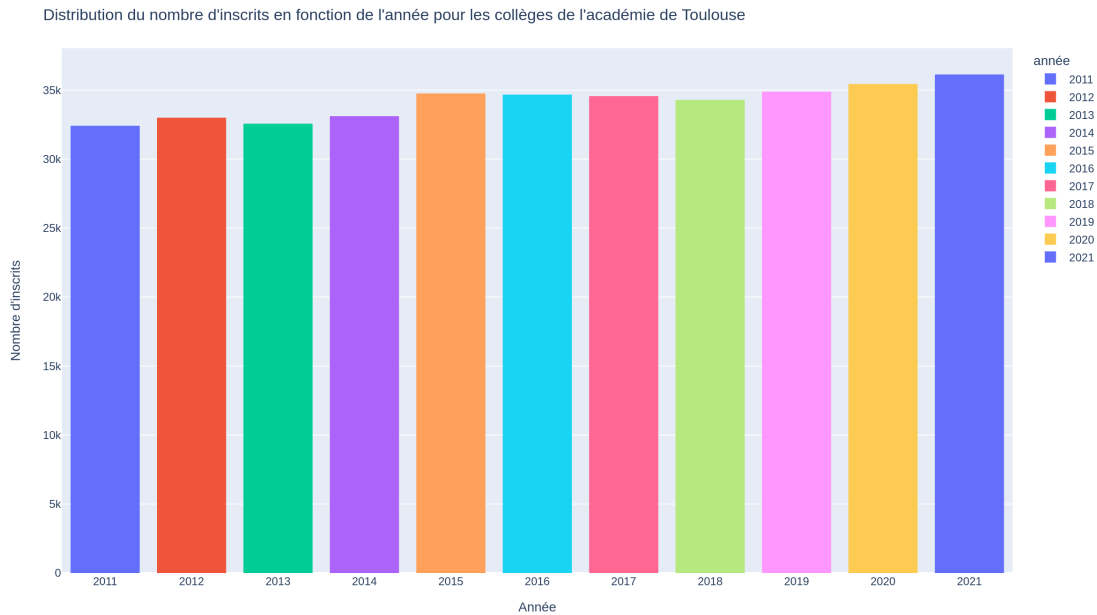
On peut voir que le nombre d'élèves admis **augmente chaque année**. On remarque également **un pic à l'année 2020** qui pourrait être expliqué par l'allègement des examens dûs aux difficultés liées à la covid.

### Nombre total d'inscrits en fonction de l'année au collège à Toulouse

```
[22]: # Total d'inscrits pour toutes les séries
fig2 = px.histogram(df2T, x="année", y="inscrits", color="année")
fig2.update_layout(title_text="Distribution du nombre d'inscrits en fonction de l'année pour les collèges de l'académie de Toulouse")
fig2.update_yaxes(title_text="Nombre d'inscrits")
fig2.update_xaxes(title_text="Année")
# fig2.show()

image = fig2.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[22]:

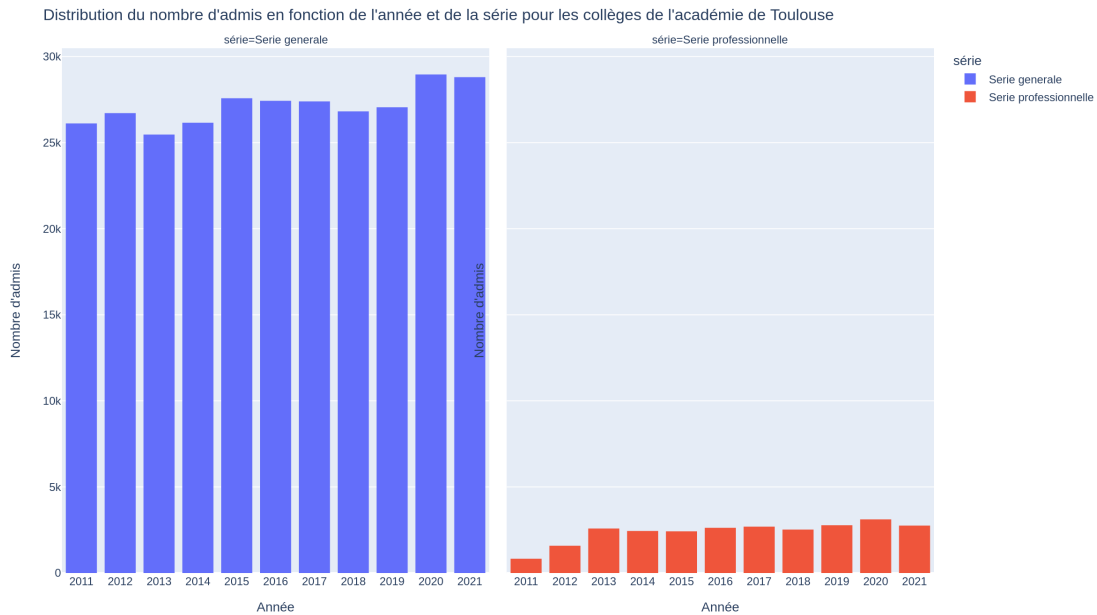


### Nombre d'admis en fonction de l'année et de la série à Toulouse

```
[23]: # Toulouse : nb admis en fct de l'année et série
fig3 = px.histogram(df2T, x="année", y="admis", color="série",
    ↪ facet_col="série")
fig3.update_layout(title_text="Distribution du nombre d'admis en fonction de
    ↪ l'année et de la série pour les collèges de l'académie de Toulouse")
fig3.update_yaxes(title_text="Nombre d'admis")
fig3.update_xaxes(title_text="Année")
# fig3.show()

image = fig3.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[23]:

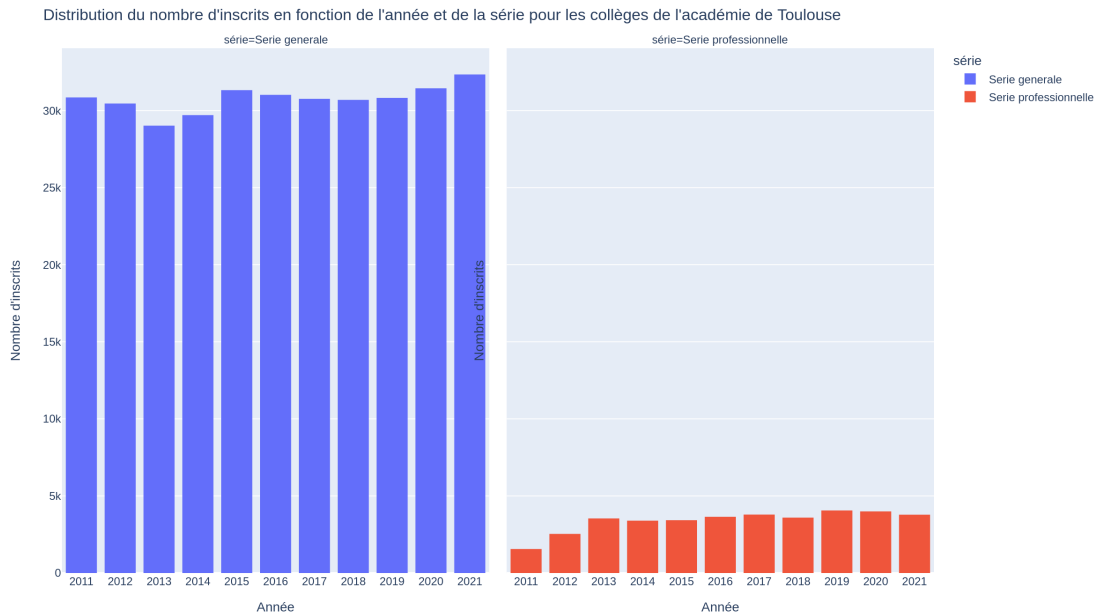


### Nombre d'inscrits en fonction de l'année et de la série à Toulouse

```
[24]: # Toulouse : nb d'inscrits en fct de l'année et série
fig4 = px.histogram(df2T, x="année", y="inscrits", color="série",
    ↪facet_col="série")
fig4.update_layout(title_text="Distribution du nombre d'inscrits en fonction de
    ↪l'année et de la série pour les collèges de l'académie de Toulouse")
fig4.update_yaxes(title_text="Nombre d'inscrits")
fig4.update_xaxes(title_text="Année")
# fig4.show()

image = fig4.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[24]:



On peut voir que pour l'académie de Toulouse, pour les élèves **en série générale** :

- le nombre d'admis en **2011** est de **26 119**
- le nombre d'admis en **2021** est de **28 803**
- le nombre d'inscrits en **2011** est de **30 865**
- le nombre d'inscrits en **2021** est de **32 345**

Et pour les élèves en **série professionnelle** :

- le nombre d'admis en **2011** est de **803**
- le nombre d'admis en **2021** est de **2753**
- le nombre d'inscrits en **2011** est de **1555**
- le nombre d'inscrits en **2021** est de **3783**

Au **total** :

- le nombre d'admis en **2011** est de **26 949**
- le nombre d'admis en **2021** est de **31 556**
- le nombre d'inscrits en **2011** est de **32 410**
- le nombre d'inscrits en **2021** est de **36 128**

```
[25]: # ratio admis série générale
rasg = (28803 - 26119)/26119
print(rasg)
# ratio inscrits série générale
risg = (32345 - 30864)/30864
print(risg)
# ratio admis série professionnelle
rasp = (2753 - 803)/803
```

```

print(rasp)
# ratio inscrits série professionnelle
risp = (3783 - 1555)/1555
print(risp)
# ratio admis total
rat = (31556 - 26949)/26949
print(rat)
# ratio inscrit total
rit = (36128 - 32410)/32410
print(rit)

```

```

0.10276044258968567
0.047984707102125455
2.428393524283935
1.432797427652733
0.17095253998293072
0.11471767972847886

```

On constate que pour l'académie de Montpellier, les élèves en **série générale**, en **10 ans** (2011 à 2021) :

- le nombre d'admis a augmenté de **10,27%**
- le nombre d'inscrits a augmenté de **4,79%**

Et pour les élèves en **série professionnelle** :

- le nombre d'admis a augmenté de **242,8%**
- le nombre d'inscrits a augmenté de **143,2%**

Enfin, au **total** on a :

- le nombre d'admis a augmenté de **17,09%**
- le nombre d'inscrits a augmenté de **11,47%**

Même constat que pour Montpellier : **le nombre d'admis augmente énormément par rapport au nombre d'inscrits**, on en déduit que **le niveau global des collégiens s'est amélioré en 10 ans**.

---

### Nombre d'inscrits en séries générales et technologiques à Toulouse en 2011

```

[26]: # ratio inscrits série générale / série technologique pour 2011
asgvals = [803, 26119]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'inscrits en séries_
↳générales et technologiques pour les collèges de l'académie de Toulouse en_
↳2011")
# fg1.show()

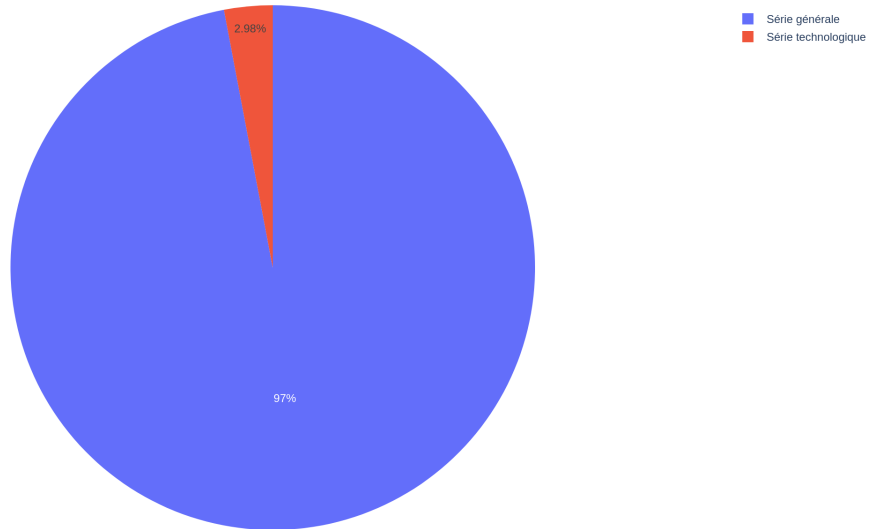
image = fg1.to_image(format='png',width=1200, height=700, scale=2)

```

```
Image(image)
```

[26] :

Distribution du nombre d'inscrits en séries générales et technologiques pour les collèges de l'académie de Toulouse en 2011



---

### Nombre d'inscrits en séries générales et technologiques à Toulouse en 2021

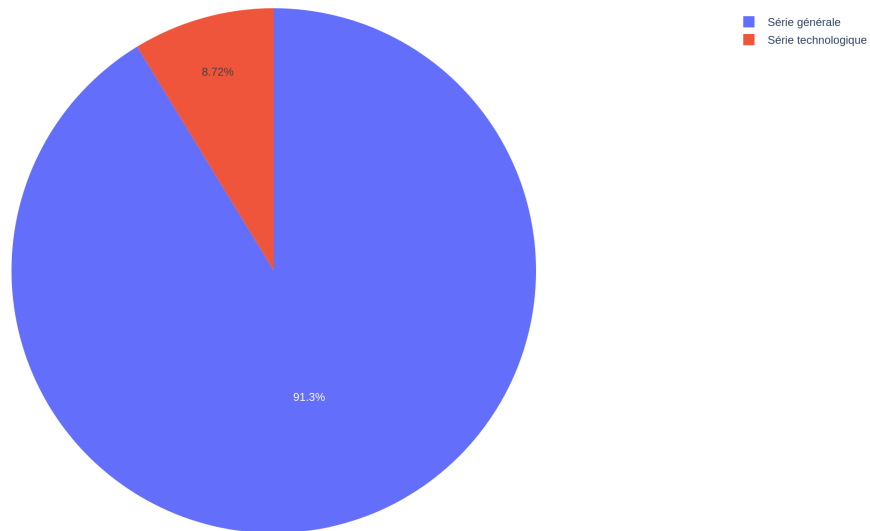
```
[27]: # ratio inscrits série générale / série technologique pour 2021
asgvals = [2753, 28803]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'inscrits en séries_
↳générales et technologiques pour les collèges de l'académie de Toulouse en_
↳2021")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[27] :



Distribution du nombre d'inscrits en séries générales et technologiques pour les collèges de l'académie de Toulouse en 2021



On peut voir que la proportion d'élèves inscrits en série technologique a **nettement augmenté** depuis 2011. On est passés de **2,98%** à **8,72%** ce qui est largement supérieur à l'évolution de l'académie de Montpellier

---

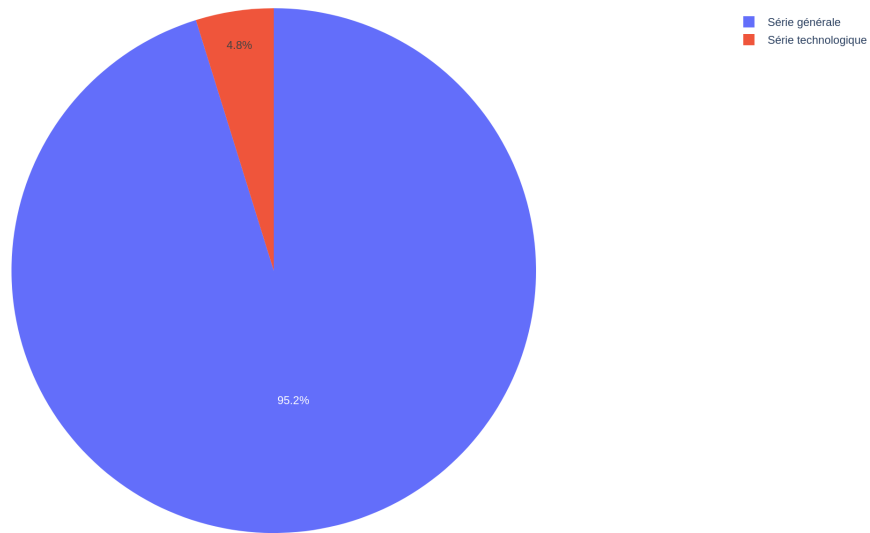
#### Nombre d'admis en séries générales et technologiques à Toulouse en 2011

```
[28]: # ratio admis série générale / série technologique pour 2011
asgvals = [1555, 30865]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'admis en séries_
↳générales et technologiques pour les collèges de l'académie de Toulouse en_
↳2011")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[28]:

Distribution du nombre d'admis en séries générales et technologiques pour les collèges de l'académie de Toulouse en 2011



---

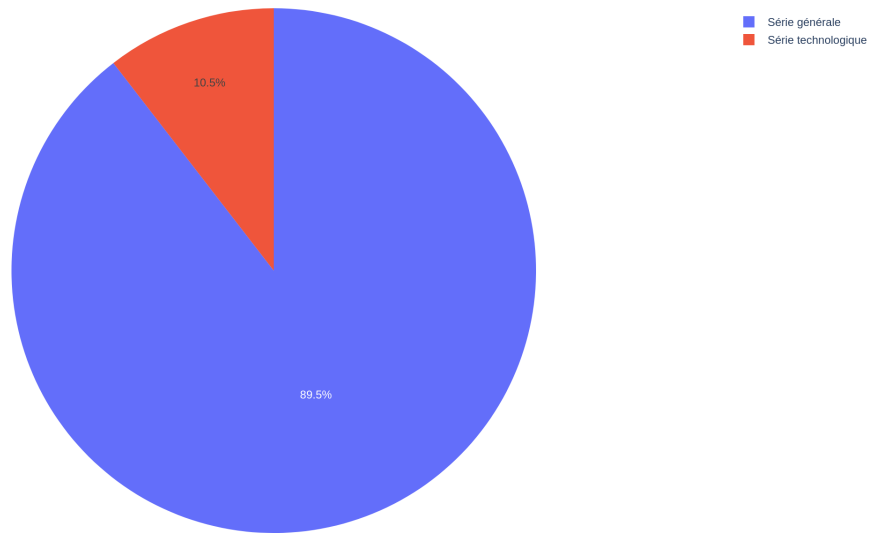
### Nombre d'admis en séries générales et technologiques à Toulouse en 2021

```
[29]: # ratio admis série générale / série technologique pour 2021
asgvals = [3783, 32345]
asglabls = ['Série technologique', 'Série générale']
fg1 = px.pie(values=asgvals, names= asglabls)
fg1.update_layout(title_text="Distribution du nombre d'admis en séries_
↳générales et technologiques pour les collèges de l'académie de Toulouse en_
↳2021")
# fg1.show()

image = fg1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[29]:

Distribution du nombre d'admis en séries générales et technologiques pour les collèges de l'académie de Toulouse en 2021



Même remarque pour les admissions :

On peut voir que la proportion d'élèves admis en série technologique a nettement **augmenté depuis 2011**. On est passés de **4,8%** à **10,5%** ce qui est largement supérieur à l'évolution de l'académie de Montpellier.

#### 0.1.4 ETL du CSV Label numérique collège académie de Montpellier

```
[30]: # importing the csv files
dflabels = pd.read_csv("fr-en-occitanie-ac-montpellier-label-numerique-college.
↳csv", sep = ";")
dflabels
```

```
[30]:
```

	Année	RNE	Nom établissement	Localité \
0	2017	0110686L	Collège Joseph Anglade	Lézignan-Corbières
1	2017	0110870L	Collège Antoine Pons	Chalabre
2	2017	0111051H	Collège Rosa Parks	Lézignan-Corbières
3	2017	0300031W	Collège George Ville	Pont-Saint-Esprit
4	2017	0301010K	Collège Jules Vallès	Nîmes
...	...	...	...	...
1003	2022	0480011B	Collège André Chamson	Meyrueis
1004	2022	0480016G	Collège du Trenze	Vialas
1005	2022	0660431K	Collège Le Riberal	Saint-Estève
1006	2022	0660634F	Collège Climatique et Sportif	Font-Romeu-Odeillo-Via
1007	2022	0660925X	Collège Christian Bourquin	Millas

	Département	Label	Position
0	11	2	43.196809464, 2.76527043622
1	11	1	42.9851089045, 2.00497993951
2	11	2	43.184518864, 2.7595799942
3	30	2	44.2562769196, 4.64588054318
4	30	3	43.8588460926, 4.39110910066
...	...	...	...
1003	48	3	44.1799498461, 3.43142941984
1004	48	3	44.3331989353, 3.89597048145
1005	66	3	42.7086982816, 2.83931047208
1006	66	2	42.5150970307, 2.03995955126
1007	66	3	42.6864276173, 2.68759941973

[1008 rows x 7 columns]

```
[31]: # afficher les infos des colonnes
dflabels.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1008 entries, 0 to 1007
Data columns (total 7 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Année                 1008 non-null   int64
1   RNE                   1008 non-null   object
2   Nom établissement    1008 non-null   object
3   Localité              1008 non-null   object
4   Département           1008 non-null   int64
5   Label                 1008 non-null   int64
6   Position              1008 non-null   object
dtypes: int64(3), object(4)
memory usage: 55.2+ KB
```

```
[32]: # Trie la colonne Label par labels (ascendant : de label 1 à label 3)
dflabels = dflabels.sort_values(by="Label")
dflabels = dflabels.sort_values(by="Année")
# dflabels.info()
```

```
[33]: # Transforme la colonne en string afin de pouvoir l'utiliser pour les graphiques
dflabels["Label"] = dflabels["Label"].apply(str)
dflabels["Département"] = dflabels["Département"].apply(str)

# copie temporaire pour les graphiques :
# dflabelsgraph = dflabels

dflabels["Année"] = dflabels["Année"].apply(str)
```

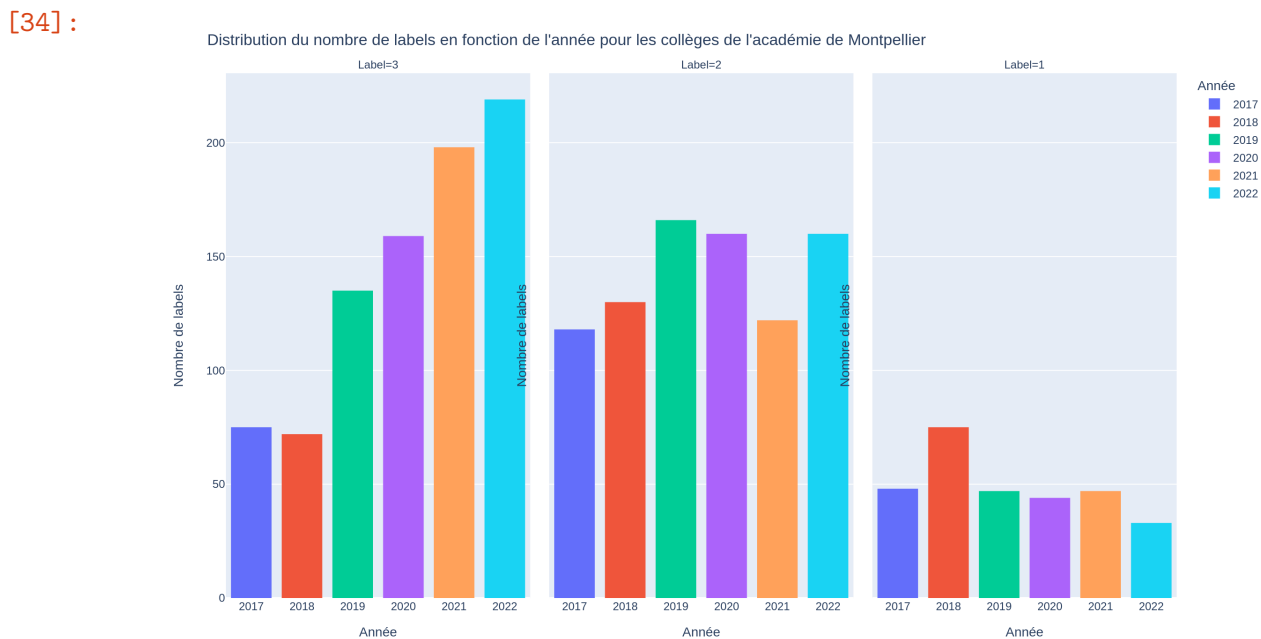
```
# dflabels.info()
```

### 0.1.5 Analyses des labels pour les collèges

Distribution du nombre de labels en fonction de l'année pour les collèges de l'académie de Montpellier

```
[34]: # Nombre de labels en fct de l'année
fig1 = px.histogram(dflabels, x="Année", y="Label", color="Année",
                    facet_col="Label")
fig1.update_layout(title_text="Distribution du nombre de labels en fonction de
                    l'année pour les collèges de l'académie de Montpellier")
fig1.update_yaxes(title_text="Nombre de labels")
fig1.update_xaxes(title_text="Année")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```



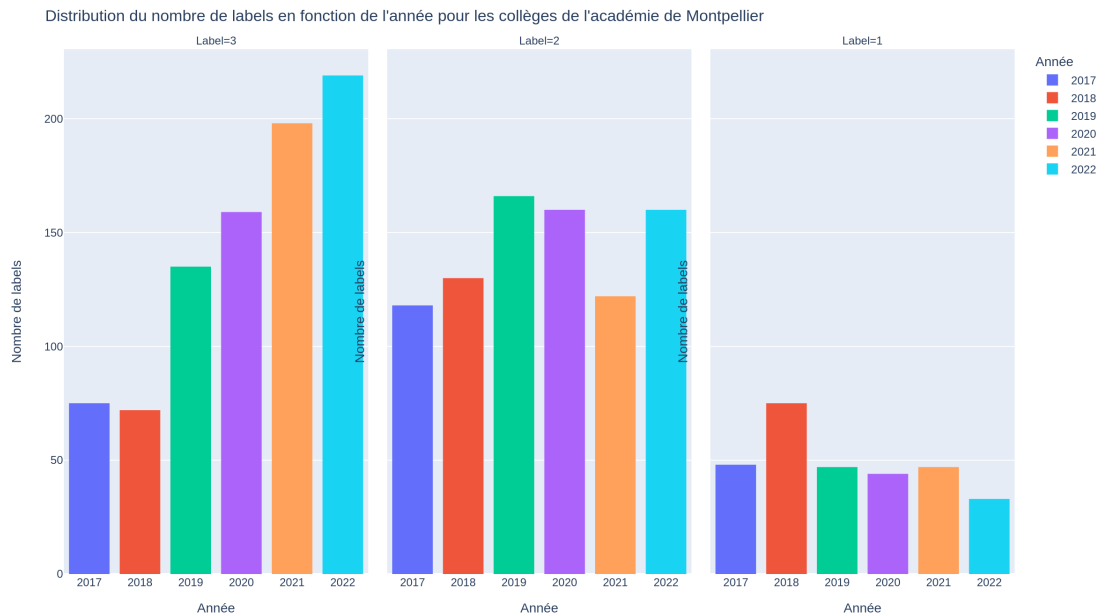
On remarque que le nombre de collèges label 1 diminue au fil des ans tandis que les collèges labels 2 et 3 ne fait qu'augmenter.

Distribution du nombre de labels en fonction de l'année pour les collèges de l'académie de Montpellier

```
[35]: # Nombre de labels en fct de l'année
fig1 = px.histogram(dflabels, x="Année", y="Label", color="Année",
    ↳facet_col="Label")
fig1.update_layout(title_text="Distribution du nombre de labels en fonction de
    ↳l'année pour les collèges de l'académie de Montpellier")
fig1.update_yaxes(title_text="Nombre de labels")
fig1.update_xaxes(title_text="Année")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

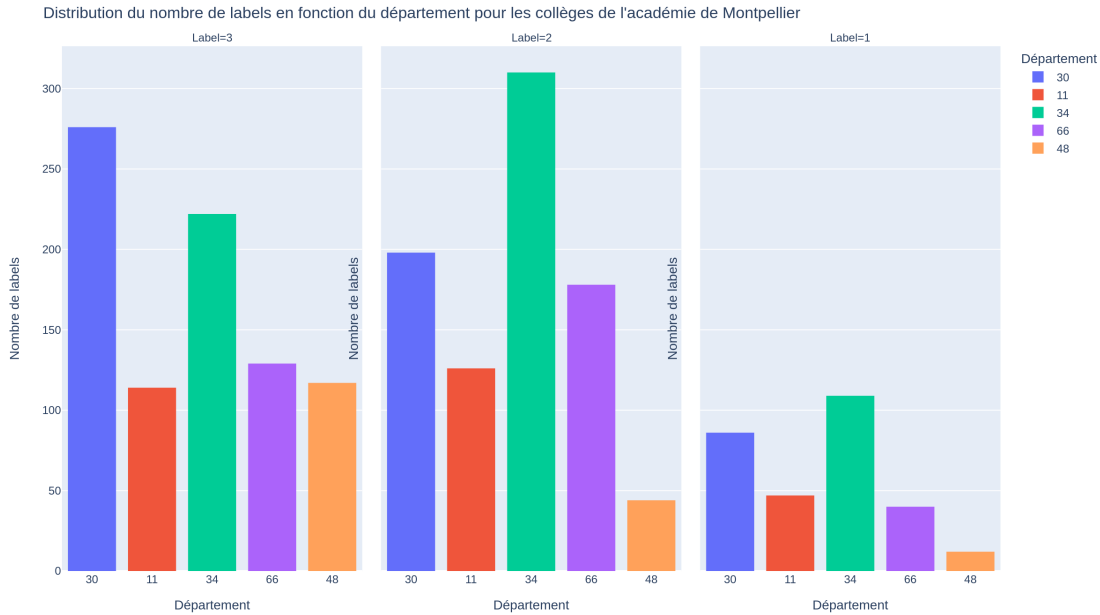
[35]:



```
[36]: # Total admis pour toutes les séries
fig1 = px.histogram(dflabels, x="Département", y="Label", color="Département",
    ↳facet_col="Label")
fig1.update_layout(title_text="Distribution du nombre de labels en fonction du
    ↳département pour les collèges de l'académie de Montpellier")
fig1.update_yaxes(title_text="Nombre de labels")
fig1.update_xaxes(title_text="Département")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[36]:



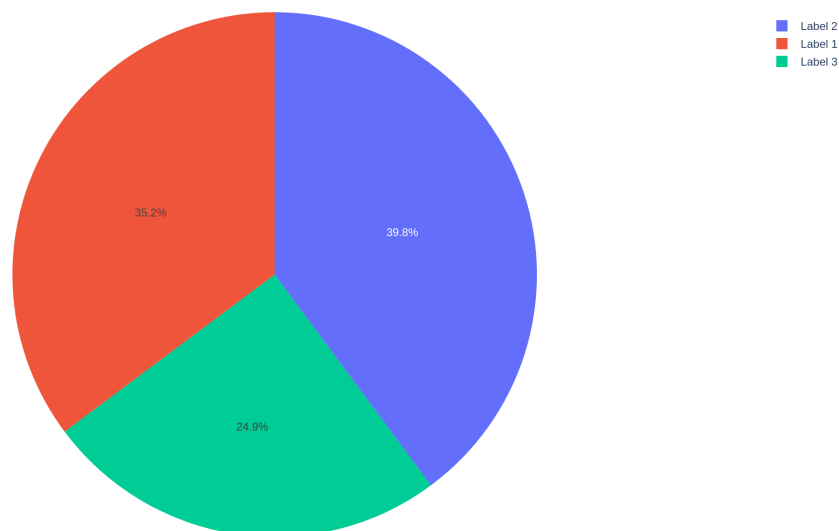
On peut voir que les départements ayant le plus de collèges labellisés **2** et **3** sont les départements de l'Hérault (34), Gard (30) et Pyrénées-Orientales (66)

```
[37]: # Compte du nombre de collèges dont le RNE est unique ayant un label de 1
nblabels1 = dflabels.query("Label == '1'")["RNE"].nunique()
# Compte du nombre de collèges dont le RNE est unique ayant un label de 2
nblabels2 = dflabels.query("Label == '2'")["RNE"].nunique()
# Compte du nombre de collèges dont le RNE est unique ayant un label de 3
nblabels3 = dflabels.query("Label == '3'")["RNE"].nunique()
labelvals = [nblabels1, nblabels2, nblabels3]
labelnames = ['Label 1', 'Label 2', 'Label 3']
fgl1 = px.pie(values=labelvals, names=labelnames)
fgl1.update_layout(title_text="Distribution du nombre de labels pour les_
↳collèges de l'académie de Montpellier")
# fgl1.show()

image = fgl1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[37]:

Distribution du nombre de labels pour les collèges de l'académie de Montpellier



On peut constater que il y'a plus de collèges labéllisés **3** que de collèges labéllisés **1** et **2**.

C'est un résultat assez prédictible étant donné que pour qu'un collège soit labellisé 3, il faut que ce collège réunisse **beaucoup de critères** en terme **d'infrastructure, d'équipements** et autres dans le domaine du numérique.

### 0.1.6 ETL du CSV Indices de position sociale dans les collèges de France métropolitaine et DROM

```
[38]: # importing the csv files
dfIPS = pd.read_csv("fr-en-ips_colleges.csv", sep = ";")
dfIPS
```

```
[38]:
```

	Rentrée scolaire	Académie	Code du département	Département	UAI \
0	2021-2022	LYON	001	AIN	0010039M
1	2021-2022	LYON	001	AIN	0010084L
2	2021-2022	LYON	001	AIN	0010794H
3	2021-2022	LYON	001	AIN	0010796K
4	2021-2022	LYON	001	AIN	0010820L
...	...	...	...	...	...
6962	2021-2022	LA REUNION	974	LA REUNION	9741323V
6963	2021-2022	MAYOTTE	976	MAYOTTE	9760009F
6964	2021-2022	MAYOTTE	976	MAYOTTE	9760167C
6965	2021-2022	MAYOTTE	976	MAYOTTE	9760244L
6966	2021-2022	MAYOTTE	976	MAYOTTE	9760368W

```
Nom de l'établissement Code INSEE de la commune \
```



0	COLLEGE DE L ALBARINE	01384
1	COLLEGE PRIVE SAINT CHARLES	01093
2	COLLEGE DU VALROMEY	01022
3	COLLEGE EMILE CIZAIN	01262
4	COLLEGE GEORGE SAND	01306
...	...	...
6962	COLLEGE BEAUSEJOUR	97418
6963	COLLEGE NELSON MANDELA	97611
6964	COLLEGE DE M TSANGAMOUJI	97613
6965	COLLEGE FREDERIC D ACHERY	97610
6966	COLLEGE DE KWALE	97611

	Nom de la commune	Secteur	IPS
0	SAINT RAMBERT EN BUGHEY	public	98.5
1	CHATILLON SUR CHALARONNE	privé sous contrat	114.6
2	ARTEMARE	public	95.6
3	MONTLUEL	public	115.2
4	PONT DE VEYLE	public	104.2
...	...	...	...
6962	SAINTE MARIE	public	80.3
6963	MAMOUDZOU	public	64.9
6964	M TSANGAMOUJI	public	77.3
6965	KOUNGOU	public	61.7
6966	MAMOUDZOU	public	55.6

[6967 rows x 10 columns]

```
[39]: # afficher les infos des colonnes
dfIPS.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6967 entries, 0 to 6966
Data columns (total 10 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Rentrée scolaire      6967 non-null   object
1   Académie               6967 non-null   object
2   Code du département    6967 non-null   object
3   Département            6967 non-null   object
4   UAI                    6967 non-null   object
5   Nom de l'établissement 6967 non-null   object
6   Code INSEE de la commune 6967 non-null   object
7   Nom de la commune      6967 non-null   object
8   Secteur                6967 non-null   object
9   IPS                    6962 non-null   float64
dtypes: float64(1), object(9)
memory usage: 544.4+ KB
```

```
[40]: # Prendre uniquement les collèges dont l'académie est Montpellier
dfIPSM = dfIPS[dfIPS["Académie"] == 'MONTPELLIER']
#print(dfIPS)
```

```
# les collèges montpellier et toulouse
dfIPST = dfIPS[dfIPS["Académie"] == 'TOULOUSE']
dfIPSALL = pd.concat([dfIPSM,dfIPST])
#print(dfIPSALL)
```

```
[41]: # Transforme la colonne en string afin de pouvoir l'utiliser pour les graphiques
# dfIPS["Code du département"] = dflabels["Code du département"].apply(str)
```

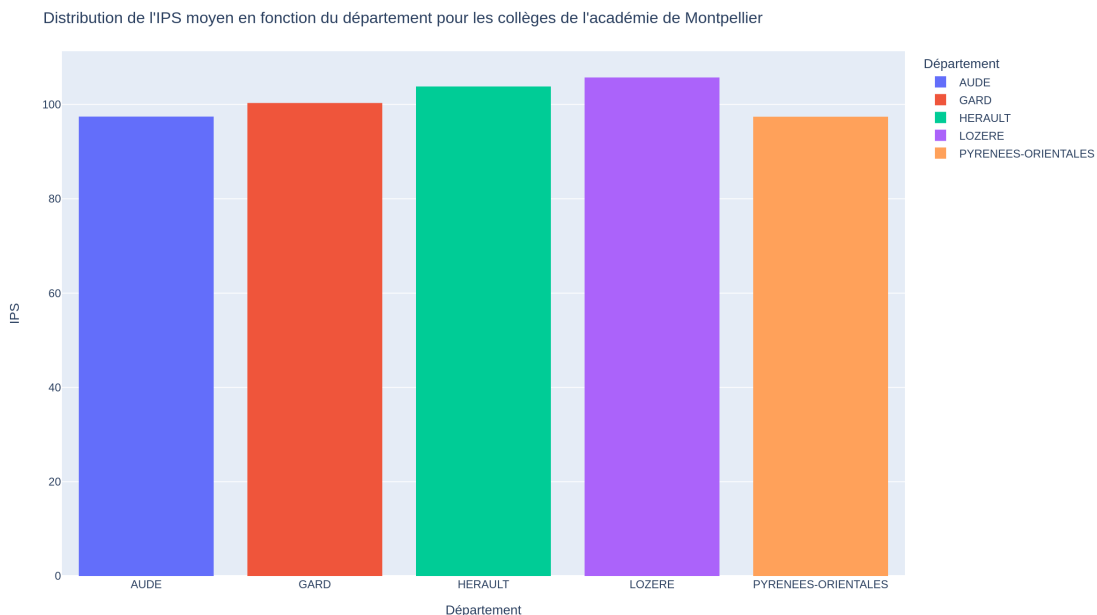
```
[42]: ### Analyses des IPS pour les collèges
```

Distribution de l'IPS moyen en fonction du département pour l'académie de Montpellier

```
[43]: # IPS moyen en fct du département
fig1 = px.histogram(dfIPSM, x="Département", y="IPS", color="Département",
    ↪histfunc="avg", )
fig1.update_layout(title_text="Distribution de l'IPS moyen en fonction du
    ↪département pour les collèges de l'académie de Montpellier")
fig1.update_yaxes(title_text="IPS")
fig1.update_xaxes(title_text="Département")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[43]:



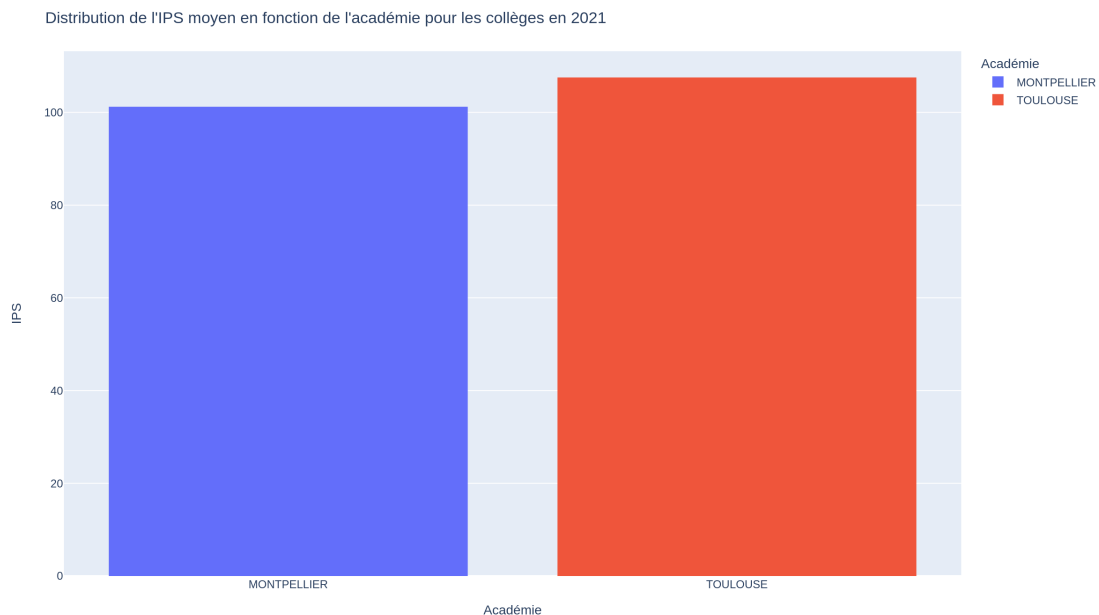
On peut voir que l'IPS moyen ne change pas trop entre les départements de la région occitanie.

### Distribution de l'IPS moyen en fonction de l'académie pour les collèges en 2021

```
[44]: # IPS moyen en fct du département
fig1 = px.histogram(dfIPSALL, x="Académie", y="IPS", color="Académie",
                    histfunc="avg")
fig1.update_layout(title_text="Distribution de l'IPS moyen en fonction de l'académie pour les collèges en 2021")
fig1.update_yaxes(title_text="IPS")
fig1.update_xaxes(title_text="Académie")
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[44]:



On constate que l'**IPS moyen** de l'académie de Toulouse est supérieur de **6,3** points par rapport à celui de Montpellier.

## 0.2 Croisement des données des collèges

```
[45]: # Remodification du CSV pour mise en forme (ETL)
# importing the csv files
dflabels = pd.read_csv("fr-en-occitanie-ac-montpellier-label-numerique-college.
↳ csv", sep = ";")
# Trie la colonne Label par labels (ascendant : de label 1 à label 3)
dflabels = dflabels.sort_values(by="Label")
# Transforme la colonne en string afin de pouvoir l'utiliser pour les graphiques
dflabels["Label"] = dflabels["Label"].apply(str)
dflabels["Département"] = dflabels["Département"].apply(str)
```

```
[46]: # Préparation des données du df des labels afin de prendre en compte que
↳ l'année 2021
# pour matcher avec le fait que pour le df IPS on a des infos que sur 2021/2022
dflabels = dflabels[dflabels["Année"] == 2021]
dflabels
```

```
[46]:      Année      RNE      Nom établissement      Localité \
519    2021    0300945P      Collège Feuchères      Nîmes
950    2021    0342050U      Collège Françoise Giroud      Vendres
951    2021    0342130F      Collège Pierre Deley      Marseillan
955    2021    0660007Z      Collège Irène Joliot Curie      Estagel
513    2021    0110859Z      Collège Joseph Delteil      Limoux
..      ...      ...      ...      ...
540    2021    0660016J      Collège la Garrigole      Perpignan
536    2021    0480002S      Collège Henri Gamala      Le Collet-de-Dèze
531    2021    0340119V      Collège Ferdinand Fabre      Bédarieux
529    2021    0340070S      Collège du Jaur      Saint-Pons-de-Thomières
525    2021    0301670C      Collège Frédéric Desmons      Saint-Geniès-de-Malgoirès
```

```
      Département Label      Position
519             30      1  43.8346647285, 4.36436997917
950             34      1  43.2702388366, 3.23120964337
951             34      1  43.3636032312, 3.52046025283
955             66      1  42.7703572832, 2.69483024839
513             11      1  43.0641830596, 2.21530968734
..      ...      ...      ...
540             66      3  42.6948472397, 2.86727657243
536             48      3  44.2458831779, 3.91986047952
531             34      3  43.6109247059, 3.15462945714
529             34      3  43.485152687, 2.77101015912
525             30      3  43.9446156949, 4.22176967663
```

[174 rows x 7 columns]

```
[47]: # Jointure afin de croiser les RNE du df des labels et les UAI du df des IPS
# pour avoir des infos sur le lien entre les 2
left = dflabels.set_index(['RNE'])
right = dfIPSM.set_index(['UAI'])

dflabels_IPS = left.join(right, lsuffix='_dflabels', rsuffix='_dfIPS')

dflabels_IPS
```

```
[47]:
```

	Année	Nom établissement	Localité \
RNE			
0300945P	2021	Collège Feuchères	Nîmes
0342050U	2021	Collège Françoise Giroud	Vendres
0342130F	2021	Collège Pierre Deley	Marseillan
0660007Z	2021	Collège Irène Joliot Curie	Estagel
0110859Z	2021	Collège Joseph Delteil	Limoux
...	...	...	...
0660016J	2021	Collège la Garrigole	Perpignan
0480002S	2021	Collège Henri Gamala	Le Collet-de-Dèze
0340119V	2021	Collège Ferdinand Fabre	Bédarieux
0340070S	2021	Collège du Jaur	Saint-Pons-de-Thomières
0301670C	2021	Collège Frédéric Desmons	Saint-Geniès-de-Malgoirès

	Département_dflabels	Label	Position \
RNE			
0300945P	30	1	43.8346647285, 4.36436997917
0342050U	34	1	43.2702388366, 3.23120964337
0342130F	34	1	43.3636032312, 3.52046025283
0660007Z	66	1	42.7703572832, 2.69483024839
0110859Z	11	1	43.0641830596, 2.21530968734
...	...	...	...
0660016J	66	3	42.6948472397, 2.86727657243
0480002S	48	3	44.2458831779, 3.91986047952
0340119V	34	3	43.6109247059, 3.15462945714
0340070S	34	3	43.485152687, 2.77101015912
0301670C	30	3	43.9446156949, 4.22176967663

	Rentrée scolaire	Académie	Code du département \
RNE			
0300945P	2021-2022	MONTPELLIER	030
0342050U	2021-2022	MONTPELLIER	034
0342130F	2021-2022	MONTPELLIER	034
0660007Z	2021-2022	MONTPELLIER	066
0110859Z	2021-2022	MONTPELLIER	011
...	...	...	...
0660016J	2021-2022	MONTPELLIER	066
0480002S	2021-2022	MONTPELLIER	048

0340119V	2021-2022	MONTPELLIER	034
0340070S	2021-2022	MONTPELLIER	034
0301670C	2021-2022	MONTPELLIER	030

	Département_dfIPS	Nom de l'établissement \
RNE		
0300945P	GARD	COLLEGE FEUCHERES
0342050U	HERAULT	COLLEGE FRANCOISE GIROUD
0342130F	HERAULT	COLLEGE PIERRE DELEY
0660007Z	PYRENEES-ORIENTALES	COLLEGE IRENE JOLIOT CURIE
0110859Z	AUDE	COLLEGE JOSEPH DELTEIL
...	...	...
0660016J	PYRENEES-ORIENTALES	COLLEGE LA GARRIGOLE
0480002S	LOZERE	COLLEGE HENRI GAMALA
0340119V	HERAULT	COLLEGE FERDINAND FABRE
0340070S	HERAULT	COLLEGE DU JAUR
0301670C	GARD	COLLEGE FREDERIC DESMONS

	Code INSEE de la commune	Nom de la commune	Secteur	IPS
RNE				
0300945P	30189	NIMES	public	87.5
0342050U	34329	VENDRES	public	102.1
0342130F	34150	MARSEILLAN	public	91.4
0660007Z	66071	ESTAGEL	public	100.2
0110859Z	11206	LIMOUX	public	98.3
...	...	...	...	...
0660016J	66136	PERPIGNAN	public	74.9
0480002S	48051	LE COLLET DE DEZE	public	105.1
0340119V	34028	BEDARIEUX	public	94.4
0340070S	34284	SAINT PONS DE THOMIERES	public	97.0
0301670C	30255	SAINT GENIES DE MALGOIRES	public	104.2

[174 rows x 15 columns]

*Ici on a récupéré les données en 2021 des collèges, notamment leur label et leur IPS*

**Distribution de l'IPS moyen en fonction du label pour les collèges de l'académie de Montpellier en 2021**

```
[48]: # IPS cumulé en fct du label
fig1 = px.histogram(dflabels_IPS, x="Label", y="IPS", color="Label",
    ↪histfunc="avg")
fig1.update_layout(title_text="Distribution de l'IPS moyen en fonction du label,
    ↪pour les collèges de l'académie de Montpellier en 2021")
fig1.update_yaxes(title_text="IPS")
fig1.update_xaxes(title_text="Label")
```

```
# fig1.show()

image = fig1.to_image(format='png',width=1200, height=700, scale=2)
Image(image)
```

[48] :



On remarque que l'**IPS moyen** des collèges labellisés **2** est **supérieur** à celui des collèges labellisés **3** ce qui semble incohérent à premiers abords.

Ainsi pour voir s'il y'a réellement un lien entre le label et l'IPS il faudrait effectuer une **ANOVA** pour déterminer **l'effet du label sur l'IPS**.

[49] : *# Préparation des données du df des résultats afin de prendre en compte que ↪ l'année 2021 et 2022*

```
# pour matcher avec le fait que pour le df IPS on a des infos que sur 2021/2022
dfres22 = df[df["année"] == 2022]
# les collèges montpellier et toulouse
dfres21 = df[df["année"] == 2021]
dfres = pd.concat([dfres21,dfres22])
dfres
```

[49] :

	année	académie	série	inscrits	inscrits filles \
333	2021	MONTPELLIER	Serie generale	28530	14455
334	2021	MONTPELLIER	Serie generale	28530	14455
335	2021	MONTPELLIER	Serie generale	28530	14455
336	2021	MONTPELLIER	Serie professionnelle	1892	691
337	2021	MONTPELLIER	Serie professionnelle	1892	691

...	...	...	...	...	...	...
1723	2021	TOULOUSE	Serie professionnelle	1479	558	
1724	2021	TOULOUSE	Serie generale	91	35	
1725	2021	TOULOUSE	Serie generale	91	35	
1726	2021	TOULOUSE	Serie professionnelle	164	56	
1727	2021	TOULOUSE	Serie professionnelle	164	56	

	présents	présents filles	admis	admis filles	admis sans mention	\
333	28178	14272	25305	13116	4691	
334	28178	14272	25305	13116	4691	
335	28178	14272	25305	13116	4691	
336	1794	647	1505	572	480	
337	1794	647	1505	572	480	

...	...	...	...	...	...	...
1723	1335	500	983	370	424	
1724	65	26	36	16	22	
1725	65	26	36	16	22	
1726	135	49	91	33	42	
1727	135	49	91	33	42	

	admis mention assez bien	admis mention bien	admis mention très bien	\
333	5999	7065	7550	
334	5999	7065	7550	
335	5999	7065	7550	
336	563	368	94	
337	563	368	94	

...	...	...	...	...
1723	330	187	42	
1724	10	2	2	
1725	10	2	2	
1726	27	15	7	
1727	27	15	7	

	genre	âge	présence	admissibilité	\
333	Garçons	15 ans ou moins	Présents	Non admis	
334	Garçons	15 ans ou moins	Présents	Admis	
335	Garçons	15 ans ou moins	Présents	Admis	
336	Filles	15 ans ou moins	Présents	Admis	
337	Filles	15 ans ou moins	Présents	Admis	

...	...	...	...	...	...
1723	Garçons	16 ans	Présents	Admis	
1724	Filles	17 ans ou plus	Présents	Admis	
1725	Garçons	17 ans ou plus	Non présents	Non présents	
1726	Filles	17 ans ou plus	Présents	Admis	
1727	Filles	17 ans ou plus	Présents	Admis	

admissibilité détails

Résultats Effectifs



333		e. Non admis	non_admis_garcons	1717
334		d. Admis sans mention	admis_sans_mention_garcons	2514
335		b. Admis mention bien	admis_mention_bien_garcons	3422
336		d. Admis sans mention	admis sans mention filles	175
337	c.	Admis mention assez bien	admis mention assez bien filles	206
...		...	...	...
1723		b. Admis mention bien	admis_mention_bien_garcons	118
1724		b. Admis mention bien	admis mention bien filles	0
1725		f. Non présents	non_presents_garcons	17
1726		d. Admis sans mention	admis sans mention filles	17
1727	a.	Admis mention très bien	admis mention très bien filles	3

[144 rows x 20 columns]

## Import des bibliothèques pour l'ANOVA

```
[50]: # Import des bibliothèques pour l'anova
import statsmodels.api as sa
import statsmodels.formula.api as sfa
import scikit_posthocs as sp
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import scipy.stats as stats
```

## Analyse de la variance

Nous souhaitons répondre à la question suivante : y'a t'il un effet de la labélisation sur l'IPS ?

Pour cela on pose les hypothèses suivantes :

**H0** : La labélisation d'un collège à Montpellier en 2021 n'a pas d'effet sur l'IPS

**H1** : La labélisation d'un collège à a un effet sur l'IPS

Etant donné que Le lien entre des variables en ANOVA est réciproque car les variables indépendantes et dépendantes sont liées entre elles de manière réciproque. Une variable indépendante peut affecter une variable dépendante, et vice-versa. Ainsi on pourra tout autant conclure sur l'effet de l'IPS sur la labélisation.

```
[51]: # ANOVA labels et IPS
lm = sfa.ols('IPS ~ C(Label)', data=dflabels_IPS).fit()
anova = sa.stats.anova_lm(lm)
anova
```

```
[51]:
```

	df	sum_sq	mean_sq	F	PR(>F)
C(Label)	2.0	202.007476	101.003738	0.475485	0.622412
Residual	169.0	35899.379675	212.422365	NaN	NaN

Analyse de l'anova

$P\_value > \alpha (0,05)$ , donc on ne rejette pas  $H_0$  et **on conclut d'une manière significative qu'il n'y a pas d'effet de la labélisation sur l'IPS pour les collèges de Montpellier en**

**2021** et que donc il n'y a pas non plus d'effet de l'IPS sur la labélisation.