Health Insurance Prédiction Prediction - EDA

Python notebook utilisant les données de Health Insurance Prédiction · 390 vues · il y a 2 ans

0

Copier et éditer 1

Version 3

3 commit

In [1]:

```
# This Python 3 environment comes with many helpful analytics libraries
 installed
# It is defined by the kaggle/python docker image: https://github.com/ka
ggle/docker-python
# For example, here's several helpful packages to load in

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
sns.set_style("whitegrid")

# Input data files are available in the "../input/" directory.
# For example, running this (by clicking run or pressing Shift+Enter) wi
ll list the files in the input directory

import os
print(os.listdir("../input"))

# Any results you write to the current directory are saved as output.
```

```
['insurance.csv']
```

In [2]:

```
df = pd.read_csv('../input/insurance.csv')
```

In [3]:

```
df.head()
```

Out[3]:

	age	sex	bmi	children	smoker	region	charges
0	19	female	27.900	0	yes	southwest	16884.92400
1	18	male	33.770	1	no	southeast	1725.55230
2	28	male	33.000	3	no	southeast	4449.46200
3	33	male	22.705	0	no	northwest	21984.47061
4	32	male	28.880	0	no	northwest	3866.85520

In [4]:

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1338 entries, 0 to 1337
Data columns (total 7 columns):
age         1338 non-null int64
sex         1338 non-null object
bmi         1338 non-null float64
children    1338 non-null int64
smoker      1338 non-null object
region      1338 non-null object
charges     1338 non-null float64
dtypes: float64(2), int64(2), object(3)
memory usage: 73.2+ KB
```

In [5]:

In [5]:

```
df.describe()
```

Out[5]:

	age	bmi	children	charges
count	1338.000000	1338.000000	1338.000000	1338.000000
mean	39.207025	30.663397	1.094918	13270.422265
std	14.049960	6.098187	1.205493	12110.011237
min	18.000000	15.960000	0.000000	1121.873900
25%	27.000000	26.296250	0.000000	4740.287150
50%	39.000000	30.400000	1.000000	9382.033000
75%	51.000000	34.693750	2.000000	16639.912515
max	64.000000	53.130000	5.000000	63770.428010

In [6]:

```
df.isnull().sum()
```

Out[6]:

```
age         0
sex         0
bmi         0
children    0
smoker      0
region      0
charges     0
dtype: int64
```

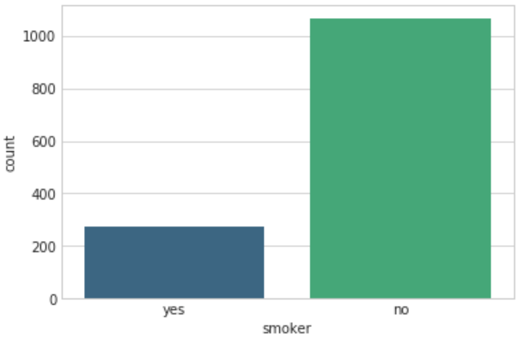
Number of people who are smokers.

In [7]:

```
sns.countplot(x='smoker',data=df,palette='viridis')
```

Out[7]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e9ed3080>
```

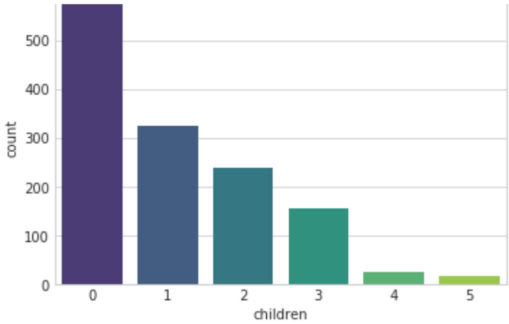


In [8]:

```
sns.countplot(x='children',data=df,palette='viridis')
```

Out[8]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e9efb780>
```



In [9]:

```
df.age.nunique()
```

Out[9]:

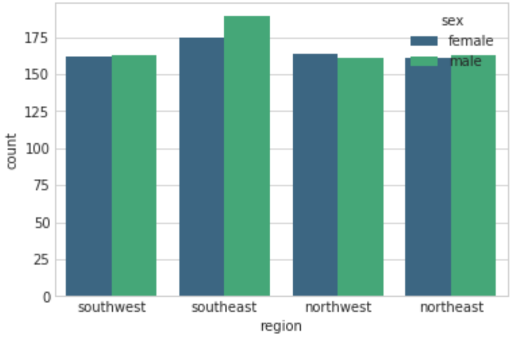
```
47
```

In [10]:

```
sns.countplot(x='region',data=df,hue='sex',palette='viridis')
```

Out[10]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e6b62f98>
```

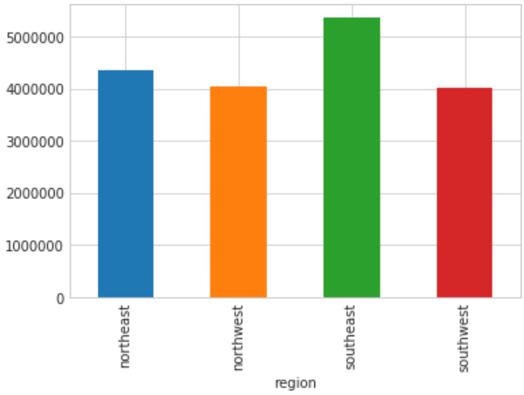


In [11]:

```
by_region = df.groupby('region').charges.sum()
by_region.plot(kind='bar')
```

Out[11]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e6b88550>
```

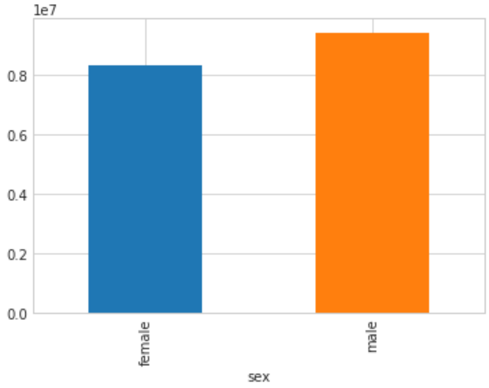


In [12]:

```
by_sex = df.groupby('sex').charges.sum()
by_sex.plot(kind='bar')
```

Out[12]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e6b1fac8>
```

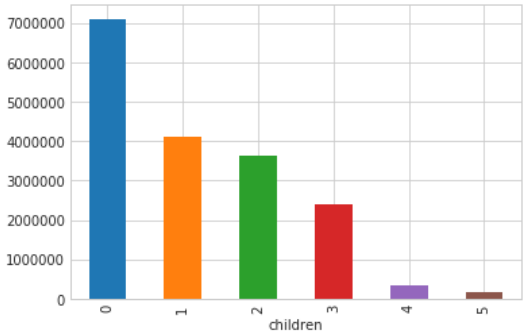


In [13]:

```
by_nofchildren = df.groupby('children').charges.sum()
by_nofchildren.plot(kind='bar')
```

Out[13]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e6a560f0>
```

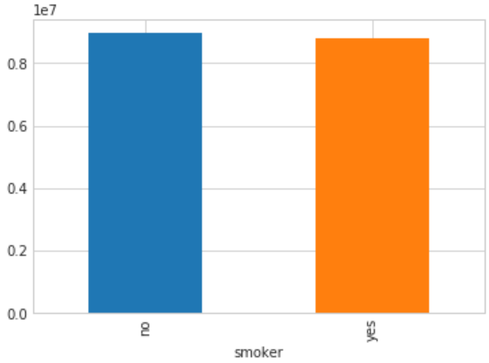


In [14]:

```
by_smoker = df.groupby('smoker').charges.sum()
by_smoker.plot(kind='bar')
```

Out[14]:

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f53e69f8438>
```



To be continued

Ce noyau a été publié sous la licence open source Apache 2.0 .

Avez-vous trouvé ce noyau utile?
Montrez votre appréciation avec un vote positif

▲ 0

Les données

Source d'information

Prévision du coût de l'assurance maladie

assurance.csv 7 colonnes

Prévision du coût de l'assurance maladie

Dernière mise à jour: il y a 2 ans (version 1)

À propos de ce jeu de données

Pas encore de description

Commentaires (0)

Click here to comment...

© 2019 Kaggle Inc

Notre équipe Conditions Confidentialité Contact / Support