

```
In [26]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [27]: df = pd.read_csv("C:/Users/user/Desktop/My learning/ClinSoft/expenses.csv")
```

```
In [28]: df.head()
```

Out[28]:

	age	sex	bmi	children	smoker	region	charges
0	19.0	female	27.900	0.0	yes	southwest	16884.92400
1	18.0	male	33.770	1.0	no	southeast	1725.55230
2	28.0	male	33.000	3.0	no	southeast	4449.46200
3	33.0	male	22.705	0.0	no	northwest	21984.47061
4	32.0	male	28.880	0.0	no	northwest	3866.85520

```
In [29]: print(df.isna())
```

	age	sex	bmi	children	smoker	region	charges
0	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False
...
1333	False	False	False	False	False	False	False
1334	False	False	False	False	False	False	False
1335	False	False	False	False	False	False	False
1336	False	False	False	False	False	False	False
1337	False	False	False	False	False	False	False

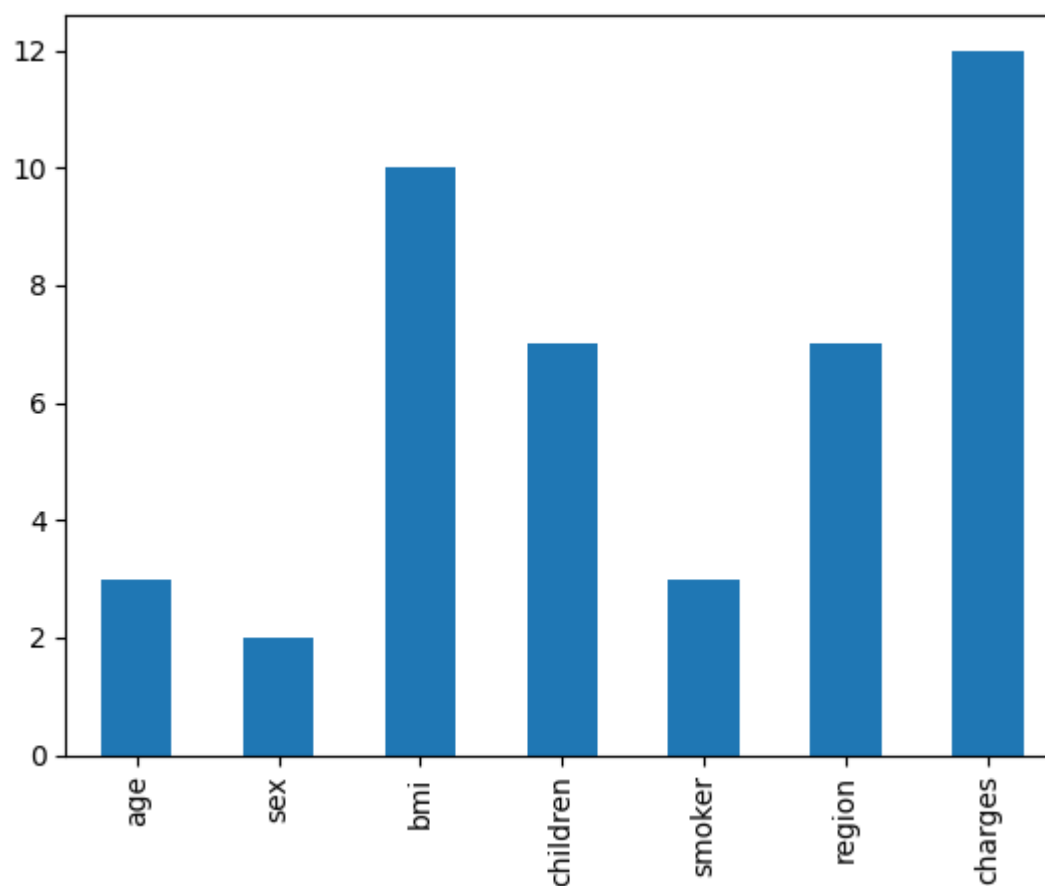
[1338 rows x 7 columns]

```
In [30]: print(df.isna().any())
```

```
age      True
sex      True
bmi      True
children True
smoker   True
region   True
charges  True
dtype: bool
```

```
In [31]: df.isna().sum().plot(kind = 'bar')
```

```
Out[31]: <Axes: >
```



```
In [32]: df.shape
```

```
Out[32]: (1338, 7)
```

```
In [33]: df.dropna().shape
```

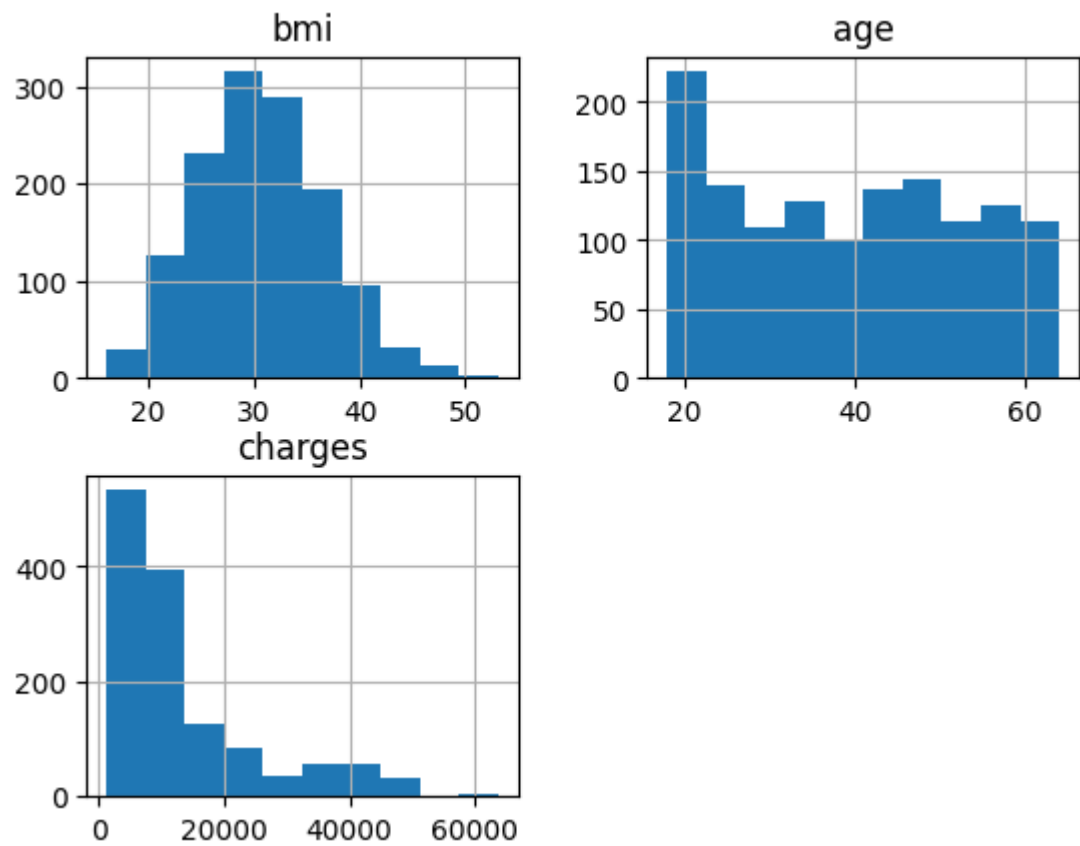
```
Out[33]: (1297, 7)
```

```
In [34]: print(df.isna().any())
```

```
age      True
sex      True
bmi      True
children True
smoker   True
region   True
charges  True
dtype: bool
```

```
In [35]: ▶ cols_with_missing = ['bmi', 'age', 'charges']
```

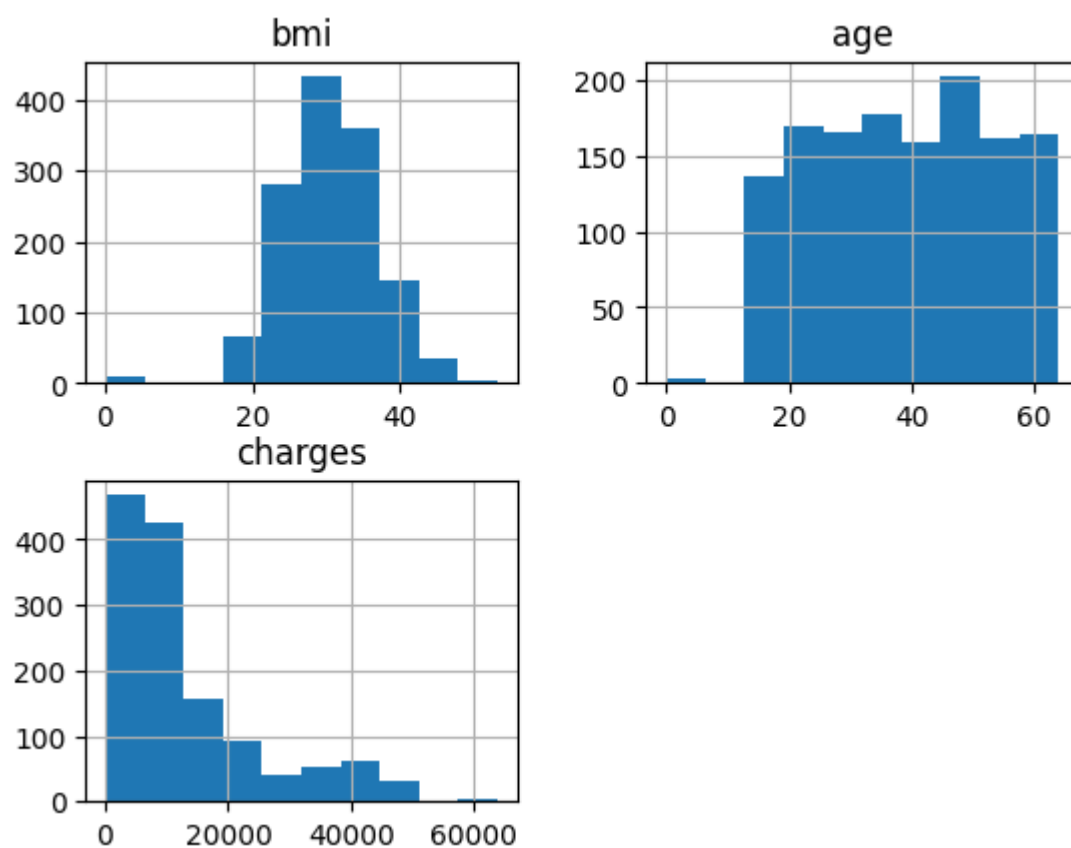
```
df[cols_with_missing].hist()  
plt.show()
```



```
In [36]: ▶ df1 = df.fillna(0)
```

```
In [37]: df1[cols_with_missing].hist()
```

```
Out[37]: array([[<Axes: title={'center': 'bmi'}>, <Axes: title={'center': 'age'}>],  
               [<Axes: title={'center': 'charges'}>, <Axes: >]], dtype=object)
```



In [41]: `df.replace('NA', 0)`

Out[41]:

	age	sex	bmi	children	smoker	region	charges
0	19.0	female	27.900	0.0	yes	southwest	16884.92400
1	18.0	male	33.770	1.0	no	southeast	1725.55230
2	28.0	male	33.000	3.0	no	southeast	4449.46200
3	33.0	male	22.705	0.0	no	northwest	21984.47061
4	32.0	male	28.880	0.0	no	northwest	3866.85520
...
1333	50.0	male	30.970	3.0	no	northwest	10600.54830
1334	18.0	female	31.920	0.0	no	northeast	2205.98080
1335	18.0	female	36.850	0.0	no	southeast	1629.83350
1336	21.0	female	25.800	0.0	no	southwest	2007.94500
1337	61.0	female	29.070	0.0	yes	northwest	29141.36030

1338 rows × 7 columns

```
In [38]: # Create a list of dictionaries with new data
avocados_list = [
    {"date": "2019-11-03", "small_sold": 10376832, "large_sold": 7835071},
    {"date": "2019-11-10", "small_sold": 10717154, "large_sold": 8561348},
]

# Convert List into DataFrame
avocados_2019 = pd.DataFrame(avocados_list)

# Print the new DataFrame
print(avocados_2019)
```

	date	small_sold	large_sold
0	2019-11-03	10376832	7835071
1	2019-11-10	10717154	8561348

```
In [39]: ▶ # Create a dictionary of Lists with new data
avocados_dict = {
    "date": ["2019-11-17", "2019-12-01"],
    "small_sold": [10859987, 9291631],
    "large_sold": [7674135, 6238096]
}

# Convert dictionary into DataFrame
avocados_2019 = pd.DataFrame(avocados_dict)

# Print the new DataFrame
print(avocados_2019)
```

	date	small_sold	large_sold
0	2019-11-17	10859987	7674135
1	2019-12-01	9291631	6238096

```
In [40]: ▶ df1.to_csv("C:/Users/user/Desktop/My learning/ClinSoft/expenses_updated.csv")
# Take a Look at the DataFrame
```

```
In [ ]: ▶
```