U.S. Medical Insurance Costs

Project Overview

Codecademy lesson excerpt:

> For this project, you will be investigating a medical insurance costs dataset in a .csv file using the Python skills that you have developed. This dataset and its parameters will seem familiar if you have done any of the previous Python projects in the data science path.
>
> However, you are now tasked with working with the actual information in the dataset and performing your own independent analysis on real-world data! We will not be providing step-by-step instructions on what to do, but we will provide you with a framework to structure your exploration and analysis.

Purpose

The purpose of this project is to demonstrate understanding of basic python programming skills.

The data provided will be analyzed to answer the following question:

> Figure out what the average age is for someone who has at least one child in this dataset.

Analysis

In [1]:
```
import pandas as pd
```

Based upon initial import and inspection of data, the following fields are available for use:

- age
- sex
- bmi
- children
- smoker
- region
- charges

In [2]:
```
df = pd.read_csv('insurance.csv')

print(df.head())
```

```
   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 order to confirm if any data values are invalid, each column will be checked for null values. Based on the initial question `Figure out what the average age is for someone who has at least one child in this dataset,` these columns will to be checked for valid data:

- age
- children

Number of errors in the `age` column:

In [3]:
```
print(len(df[df.age.isnull()]) )
```

```
0
```

Number of errors in the `children` column:

In [4]:
```
print(len(df[df.children.isnull()]) )
```

```
0
```

Total population with children:

In [5]:
```
population_with_kids = df[df.children >= 1]
number_population_with_kids = len(df[df.children >= 1])

print(number_population_with_kids)
```

```
764
```

Average age of the population with children:

In [6]:
```
avg_age_with_kids = df[(df.children >= 1)].age.mean()

print(avg_age_with_kids)
```

```
39.78010471204188
```

Result

The average age of the surveyed population with at least one child is **39 years old.**