The datasets generated can also be used to tune the machine learning model, validate the model, and to test the model.

In this article we will:

1. Explore Faker and its function
2. Create our own Dataset

**Implementation**

In order to explore faker we need to install it using **pip install faker.**

a. Importing useful libraries

We will explore different functions of faker so we will import faker also we will perform some operations on the dataset for which we need to import pandas.

from faker import Faker

import pandas as pd

b. Using different functions

Now we will explore different functions that are there in the Faker library, for this, we need to initiate the Faker function using a variable.

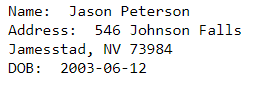
exp = Faker()

Now we will use this variable to generate different attributes.

print('Name: ', exp.name())

print('Address: ',exp.address())

print('DOB: ',exp.date\_of\_birth())

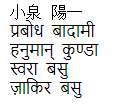


We can generate information according to different regions and localities in different languages. We just need to mention the language we want. Let’s generate some data in the Japanese and Hindi language.

exp = Faker(['ja\_JP', ‘hi\_IN])

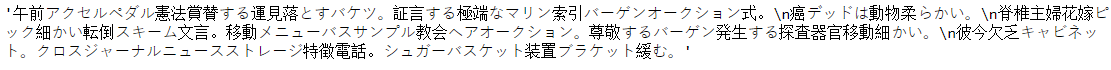
for i in range(5):

         print(exp.name())



We can also create our own sentences using the sentence function and text function.

exp.text()



exp.sentence()

https://analyticsindiamag.com/wp-content/uploads/2020/08/5-1.png

We can also create sentences by using our own defined word library which contains words of our choice and the faker will generate fake sentences using those words.

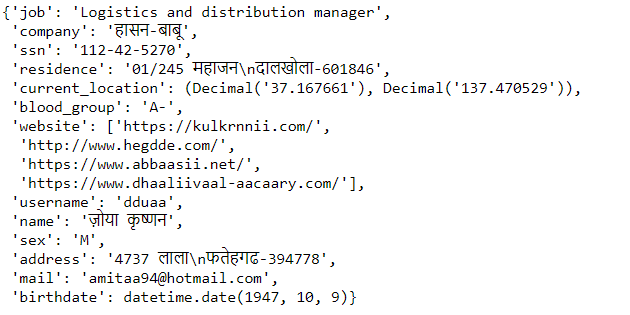
words = ['Hello','Abhishek','all', 'are','where','why',]

exp.sentence(ext\_word\_list=words)

https://analyticsindiamag.com/wp-content/uploads/2020/08/6-1.png

Other than generating names and addresses, we can generate whole profiles for different persons that do not exist. We will use the profile function to generate a fake profile of a person.

exp.profile()



Faker can also generate the random dataset.

c. Create a fake dataset using faker

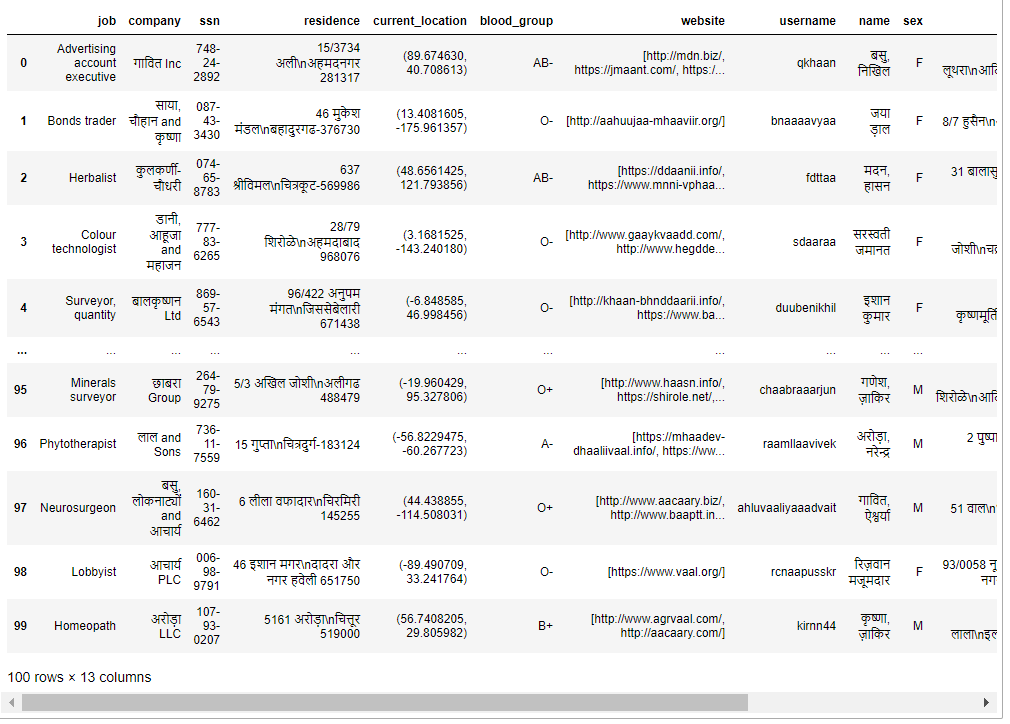
Now we will use the profile function and generate a dataset that contains profiles of 100 unique people that are fake. For this, we will also use pandas to store these profiles into a data frame. We will create these profiles in the Hindi language.

exp = Faker(['hi\_IN'])

data = [exp.profile() for i in range(100)]

df = pd.DataFrame(data)

df



The dataset we have created contains different attributes like residence, location, website, etc. We can use this dataset according to our needs.

We have stored these profiles into a data frame so that we can perform operations on it, like Visualization, Analysis, etc.

**Conclusion**

In this article, we saw how we can use Faker, an open-source python library to generate fake data and how we can create a fake dataset containing profiles of different fake people in different languages, locations, etc. The dataset created can be used for different purposes like training a machine learning model, performing different [operations](https://www.machinehack.com/), etc.