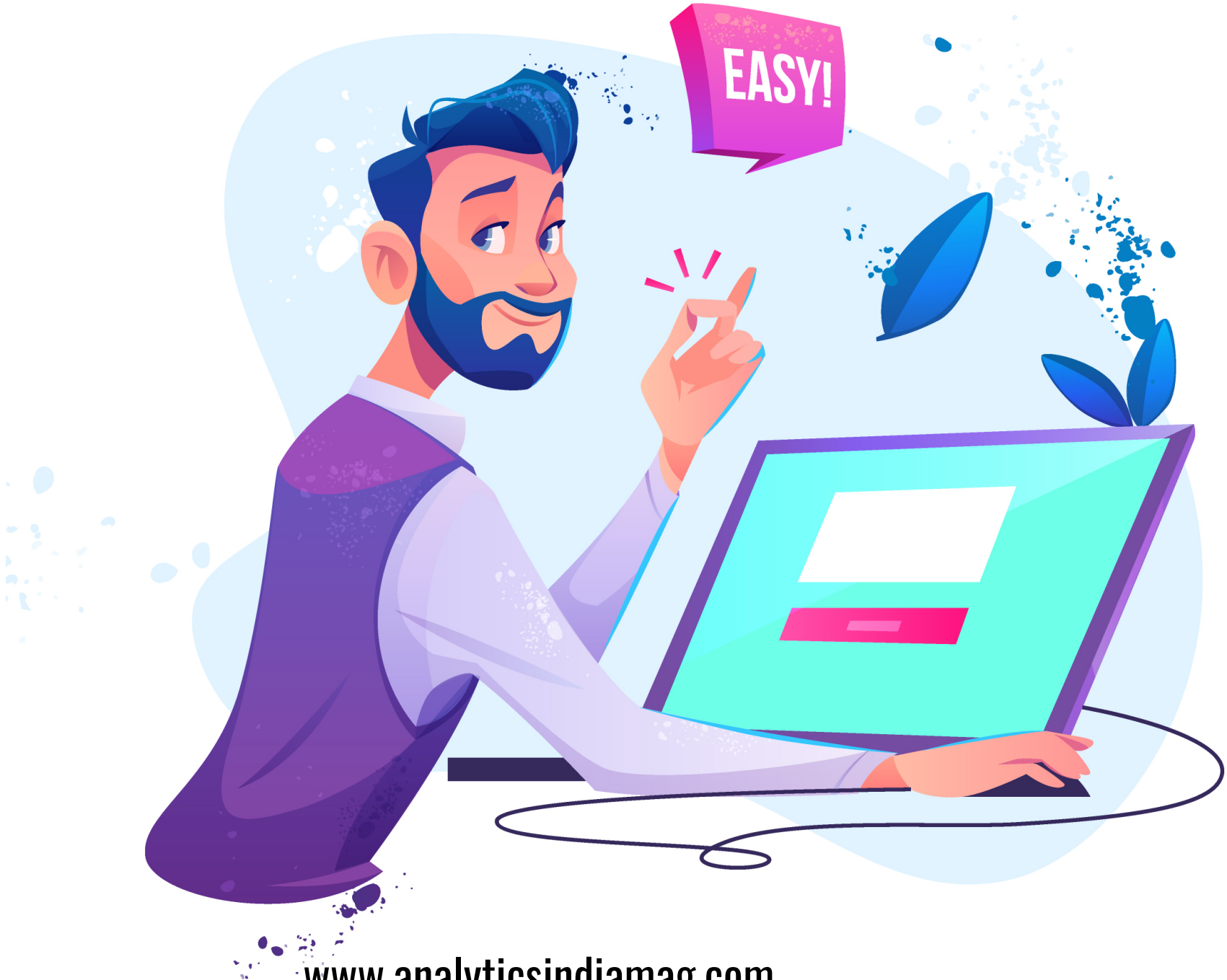
A Python Library

HANDS ON GUIDE TO FAKER



www.analyticsindiamag.com

What is Faker?

A Python Library

Faker is an open-source python library

that allows you to create your own dataset i.e you can generate random data with random attributes like name, age, location, etc. It supports all major locations and languages which is beneficial for generating data based on locality.

Faker data can be used to tune machine learning models, for stress testing a model, etc. Depending upon your need you can generate data that best fits your demand. Faker data can also be used for learning purposes like performing different operations on different types of data types.

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

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()
```

Implementation

Now we will use this variable to generate different attributes.

```
print('Name: ', exp.name())
print('Address: ',exp.address())
print('DOB: ',exp.date_of_birth())
```

```
Name:  Jason Peterson
Address:  546 Johnson Falls
Jamesstad, NV 73984
DOB:  2003-06-12
```

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())
```

```
小泉 陽一
प्रबोध बादामी
हनुमान् कुण्डा
स्वरा बसु
ज़ाकिर बसु
```

Implementation

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

```
exp.text()
```

```
'午前アクセルペダル憲法両替する運見落とすバケツ。証言する極端なマリン索引バーゲンオークション式。\n癌デッドは動物柔らかい。\n脊椎主婦花瓶ビック組かい転倒スキーム文言。移動メニューバスサンプル教会ヘアオークション。尊敬するバーゲン発生する採査器官移動細かい。\n彼今欠乏キャビネット。クロスジャーナルニュースストレージ特徴電話。シュガーバスケット装置ブラケット頼む。'
```

```
'Magnam consequatur in.'
```

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)
```

```
'Abhishek are why.'
```

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.

Implementation

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()
{'job': 'Logistics and distribution manager',
'company': 'हासन-बाबू',
'ssn': '112-42-5270',
'residence': '01/245 महाजन\nदालखोला-601846',
'current_location': (Decimal('37.167661'), Decimal('137.470529')),
'blood_group': 'A-',
'website': ['https://kulkrnnii.com/',
 'http://www.hegdde.com/',
 'https://www.abbaasii.net/',
 'https://www.dhaaliivaal-aacaary.com/'],
'username': 'dduaa',
'name': 'ज़ोया कृष्णन',
'sex': 'M',
'address': '4737 लाला\nफतेहगढ-394778',
'mail': 'amitaa94@hotmail.com',
'birthdate': datetime.date(1947, 10, 9)}
```

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.

	job	company	ssn	residence	current_location	blood_group	website	username	name	sex	
0	Advertising account executive	गावित Inc	748-24-2892	15/3734 अली\nअहमदनगर 281317	(89.674630, 40.708613)	AB-	[http://mdn.biz/, https://jmaant.com/, https:/...	qkhaan	बसु, निखिल	F	लूथरा\nआदि
1	Bonds trader	साया, चौहान and कृष्णा	087-43-3430	46 मुकेश मंडल\nबहादुरगढ-376730	(13.4081605, -175.961357)	O-	[http://aahuujaa-mhaaviir.org/]	bnaaaavyaa	जया ड़ाल	F	8/7 हुसैन\n
2	Herbalist	कुलकर्णी-चौधरी	074-65-8783	637 श्रीविमल\nचित्रकूट-569986	(48.6561425, 121.793856)	AB-	[https://ddaanii.info/, https://www.mnni-vphaa...	fdttaa	मदन, हासन	F	31 बालासु
3	Colour technologist	डानी, आहूजा and महाजन	777-83-6265	28/79 शिरोळे\nअहमदाबाद 968076	(3.1681525, -143.240180)	O-	[http://www.gaaykvaadd.com/, http://www.hegdde...	sdaaraa	सरस्वती जमानत	F	जोशी\nचव
4	Surveyor, quantity	बालकृष्णन Ltd	869-57-6543	96/422 अनुपम मंगत\nजिससेबेलारी 671438	(-6.848585, 46.998456)	O-	[http://khaan-bhnddaarii.info/, https://www.ba...	duubenikhil	इशान कुमार	F	कृष्णमूर्ति
...	...	...	...	...	...	...	...	...	...	...	
95	Minerals surveyor	छाबरा Group	264-79-9275	5/3 अखिल जोशी\nअलीगढ 488479	(-19.960429, 95.327806)	O+	[http://www.haasn.info/, https://shirole.net/,...	chaabraaarjun	गणेश, ज़ाकिर	M	शिरोळे\nआदि
96	Phytotherapist	लाल and Sons	736-11-7559	15 गुप्ता\nचित्रदुर्ग-183124	(-56.8229475, -60.267723)	A-	[https://mhaadev-dhaaliivaal.info/, https://ww...	raamllaavivek	अरोड़ा, नरेन्द्र	M	2 पुष्पा
97	Neurosurgeon	बसु, लोकनाट्यों and आचार्य	160-31-6462	6 लीला वफादार\nचिरमिरी 145255	(44.438855, -114.508031)	O+	[http://www.aacaary.biz/, http://www.baaptt.in...	ahluvaaliyaaadvait	गावित, ऐश्वर्या	M	51 वाल\n
98	Lobbyist	आचार्य PLC	006-98-9791	46 इशान मगर\nदादरा और नगर हवेली 651750	(-89.490709, 33.241764)	O-	[https://www.vaal.org/]	rcnaapusskr	रिज़वान मजूमदार	F	93/0058 नू नग
99	Homeopath	अरोड़ा LLC	107-93-0207	5161 अरोड़ा\nचित्तूर 519000	(56.7408205, 29.805982)	B+	[http://www.agrvaal.com/, http://aacaary.com/]	kirnn44	कृष्णा, ज़ाकिर	M	लाला\nइल

100 rows × 13 columns

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, etc

Follow us

Read the complete tutorial

Link in Description

www.analyticsindiamag.com