One Line Code for Data Examination

Introduction

Whenever we start to examine any dataset. We generally take a look at **Shape of the data, Top, Bottom and Random Sample of the observations, Data type, data statistics and the missing values** available in the dataset. In order to examine everything I mentioned, we have to write some code, specifically finding the data missing percentage. I try to make this code. With this, a user can just run this one line of code in their IPython session and can examine without writing any other code. Below, I have described step by step procedure to run this code.

The code will help user's to examine their data by just running this external code. By running this external code, a user can examine,

Guidence required

Hello Everyone, this is something I did because I'm trying to improve my python skills. I beleive there could be some fault in my code or any mistake. I will appreciate your response and view on it. If you think there is possibility to improve this code. Please help me with your guidence.

If you want to use this code, you are allowed to use it with appropriate reference.

Get your environment set up

The first thing you'll need to do is load in the libraries and datasets you'll be using. However, my code already have all prerequisite libraries, but still you need pandas library to read you data into pandas dataframe.

In [ ]:
```
#Provides a DataFrame structure to store data in memory and work with it easily and efficiently.
import pandas as pd

# save filepath and read file into pandas from a URL for easier access
pd.set_option('display.max_colwidth', 200)
url  = 'https://raw.githubusercontent.com/justmarkham/pycon-2016-tutorial/master/data/yelp.csv'
# read the data and store data in DataFrame titled yelp
yelp = pd.read_csv(url)
```

Step 1 (Most Important one)

As different user titled their data frame differently but my code only understands one title name "dataset". So In order to use my code you need to do this 👇👇👇.

In [ ]:
```
#Specify your titled datframe as dataset.
dataset = yelp
```

Or if you don't want to do this just titled your data frame as the 'dataset'. Like this 👇👇👇

In [ ]:
```
#Provides a DataFrame structure to store data in memory and work with it easily and efficiently.
import pandas as pd

# save filepath and read file into pandas from a URL for easier access
pd.set_option('display.max_colwidth', 200)
url  = 'https://raw.githubusercontent.com/justmarkham/pycon-2016-tutorial/master/data/yelp.csv'
# read the data and store data in DataFrame titled dataset
dataset= pd.read_csv(url)
```

Step 2 Running External Code: %run

As you begin developing more extensive code, you will likely find yourself working in both IPython for interactive exploration, as well as a text editor to store code that you want to reuse. Rather than running this code in a new window, it can be convenient to run it within your IPython session. This can be done with the %run magic.

My External code name is One_Line_Code.ipynb

You can execute this code from your IPython session as follows:

In [3]:
```
# save filepath and read file into pandas from a URL for easier access
pd.set_option('display.max_colwidth', 200)
url  = 'https://raw.githubusercontent.com/justmarkham/pycon-2016-tutorial/master/data/yelp.csv'
# read the data and store data in DataFrame titled yelp
yelp = pd.read_csv(url)
```

In [4]:
```
#Specify your titled datframe as dataset.
dataset = yelp
```

In [ ]:
```
#Run the external code
%run One_Line_Code.ipynb
```

Examine the External code output

After you've run this script, all functions defined in it are available for use in your IPython session.

The shape of the data

Dataset_info:

Number_of_variables	5
Number_of_observations	150

Shape of the data:
Input data has 150 observations and 5 variables

Top observations of the data

My code is flexible. In order to examine the top observations, the code will not show you the top 5, 10 or 20 observations but it is up to user how many top rows/observation he want to see. The code asks to user to fill the number. The number is basically tell the the number of top observations he want to see.

```
Top observations of the data:

Enter the number, to see the top observations:
```

After filling the number, **PRESS ENTER**, the Top rows will look like this. (I will fill 5.)

```
Top observations of the data:
Enter the number, to see the top observations: 5
Head:
```

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
0	9yKzy9PApeiPPOUJEtnvkg	2011-01-26	fWKvX83p0-ka4JS3dc6E5A	5	My wife took me here on my birthday for breakfast and it was excellent. The weather was perfect which made sitting outside overlooking their grounds an absolute pleasure. Our waitress was excell...	review	rLtl8ZkDX5vH5nAx9C3q5Q	2	5	0
1	ZRJwVLyzEJq1VAihDhYiow	2011-07-27	ljZ33sJrzXqU-0X6U8NwyA	5	I have no idea why some people give bad reviews about this place. It goes to show you, you can please everyone. They are probably griping about something that their own fault...there are many peop...	review	0a2KyEL0d3Yb1V6aivbluQ	0	0	0
2	6oRAC4uyJCsJl1X0WZpVSA	2012-06-14	IESLBzqUCLdSzSqm0eCSxQ	4	love the gyro plate. Rice is so good and I also dig their candy selection :)	review	0hT2KtfLiobPvh6cDC8JQg	0	1	0
3	_1QQZuf4zZOyFCvXc0o6Vg	2010-05-27	G-WvGaISbqqaMHINnByodA	5	Rosie, Dakota, and I LOVE Chaparral Dog Park!!! It's very convenient and surrounded by a lot of paths, a desert xeriscape, baseball fields, ballparks, and a lake with ducks.\n\nThe Scottsdale Park...	review	uZetl9T0NcROGOyFfughhg	1	2	0
4	6ozycU1RpktNG2-1BroVtw	2012-01-05	1uJFq2r5QfJG_6ExMRCaGw	5	General Manager Scott Petello is a good egg!!! Not to go into detail, but let me assure you if you have any issues (albeit rare) speak with Scott and treat the guy with some respect as you state y...	review	vYmM4KTsC8ZfQBg-j5MWkw	0	0	0

Bottom observations of the data

Same like previous one, here user have also freedom to examine the number of bottom observations. He can fill any number.

```
Bottom observations of the data
Enter the number, to see of tail observations: 5
Tail:
```

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
9995	VY_tvNUCCXGXQeSvJl757Q	2012-07-28	Ubyfp2RSDYW0g7Mbr8N3iA	3	First visit...Had lunch here today - used my Groupon. \n\nWe ordered the Bruschetta, Pretzels and Steak & Cheese Calzone.\n\n-We both thought there was WAY too much Balsamic used. Also, we expec...	review	_eqQoPtQ3e3UxLE4faT6ow	1	2	0
9996	EKzMHI1tip8rC1-ZAy64yg	2012-01-18	2XyIOQKbVFb6uXQdJ0RzlQ	4	Should be called house of deliciousness!\n\nI could go on and on about this item, that item, blah blah blah but I dont really waznt to. This is good food, done creatively, and in my mind presented...	review	ROru4uk5SaYc3rg8IU7SQw	0	0	0
9997	53YGfwmbW73JhFiemNeyzQ	2010-11-16	jyznYkIbpqVmlsZxSDSypA	4	I recently visited Olive and Ivy for business last week, and after 3 visits, I am convinced that Fox Restaurants has some of the best establishments in the valley. Olive and Ivy is the Fox Restau...	review	gGbN1aKQHMgfQZkqlsuwzg	0	0	0
9998	9SKdOoDHcFoxK5ZtsgHJoA	2012-12-02	5UKq9WQE1qQbJ0DJbc-B6Q	2	My nephew just moved to Scottsdale recently so me and a bunch of friends brought him here to show him a local bar that he and the girlfriend could come shoot pool, watch football, play volleyball ...	review	0lyVoNazXa20WzUyZPLaQQ	0	0	0
9999	pF7uRzygyZsltbmVpjIyvw	2010-10-16	vWSmOhg2ID1MNZHaWapGbA	5	4-5 locations.. all 4.5 star average.. I think Arizona really has some fantastic Pizza options, and Spinatos is at the top of my pizza fix list.. the semi sweet sauce is addictive, great service, ...	review	KSBFytcdjPKZgXKQnYQdkA	0	0	0

Random observations of the data

For example, here I oly want to see 2 random samples. So enter the value 2.

```
Random samples of observations from the data
Enter the number, to see random observations: 2
Sample:
```

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
1291	OL9p2tme85Sd3d8mCiSOxw	2010-08-03	iGus6wCenfh62PVeQrVcVQ	1	Every time I come here the staff is so rude! Its the closest bar to my gate so I'm kinda stuck. Hey staff, can you please pretend to like people????	review	0o0VMEJeQY0pAAZ9nxErBA	0	0	1
8412	fecYnd2_OTDECk7bd6GOFw	2011-03-19	roCITLIb_RwydpGd9ZO_JA	4	We've eaten at this location quite a few times. We've always enjoyed it. The pizza is great and the employees are friendly. We also threw my son's 6th birthday party here in sept. Our hostess assi...	review	YeK95nGGvYsssVHzcWpmkA	0	1	0

Choice of the observations from the data

Choice of the observation allows user to look into any between of observations he want to see in the dataset. He only has to filled the index numbers.

```
Choice of observations using slice index
Enter the starting observations number: 2
Enter the ending observations number: 4
Choice:
```

	business_id	date	review_id	stars	text	type	user_id	cool	useful	funny
2	6oRAC4uyJCsJl1X0WZpVSA	2012-06-14	IESLBzqUCLdSzSqm0eCSxQ	4	love the gyro plate. Rice is so good and I also dig their candy selection :)	review	0hT2KtfLiobPvh6cDC8JQg	0	1	0
3	_1QQZuf4zZOyFCvXc0o6Vg	2010-05-27	G-WvGaISbqqaMHlNnByodA	5	Rosie, Dakota, and I LOVE Chaparral Dog Park!!! It's very convenient and surrounded by a lot of paths, a desert xeriscape, baseball fields, ballparks, and a lake with ducks.\n\nThe Scottsdale Park...	review	uZetl9T0NcROGOyFfughhg	1	2	0

Data-type of the all features and target variables

From the type of data, the user can identify, features and target variables **Data type and Total non-null values**.

```
 Datatype of the data
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 10 columns):
business_id    10000 non-null object
date           10000 non-null object
review_id      10000 non-null object
stars          10000 non-null int64
text           10000 non-null object
type           10000 non-null object
user_id        10000 non-null object
cool           10000 non-null int64
useful         10000 non-null int64
funny          10000 non-null int64
dtypes: int64(4), object(6)
memory usage: 781.3+ KB
Info:

None
```

Summary statistics of the dataset variables

```
Summary Statistics:
```

	count	unique	top	freq	mean	std	min	25%	50%	75%	max
business_id	10000	4174	JokKtdXU7zXHcr20Lrk29A	37	NaN	NaN	NaN	NaN	NaN	NaN	NaN
date	10000	1995	2011-03-28	21	NaN	NaN	NaN	NaN	NaN	NaN	NaN
review_id	10000	10000	eTa5KD-LTgQv6UT1Zmijmw	1	NaN	NaN	NaN	NaN	NaN	NaN	NaN
stars	10000	NaN	NaN	NaN	3.7775	1.21464	1	3	4	5	5
text	10000	9998	Great service	2	NaN	NaN	NaN	NaN	NaN	NaN	NaN
type	10000	1	review	10000	NaN	NaN	NaN	NaN	NaN	NaN	NaN
user_id	10000	6403	fczQCSmaWF78toLEmb0Zsw	38	NaN	NaN	NaN	NaN	NaN	NaN	NaN
cool	10000	NaN	NaN	NaN	0.8768	2.06786	0	0	0	1	77
useful	10000	NaN	NaN	NaN	1.4093	2.33665	0	0	1	2	76
funny	10000	NaN	NaN	NaN	0.7013	1.90794	0	0	0	1	57

Total missing value percentage of the dataset

In this case, there were no missing value present in the dataset. So It will show 0%.

```
Total 0.0 percentage of values are missing in this data set
```

Missing values sum and percentage in each observation

As we have seen, the data has no missing value, so observations count and percentage is also zero. Hence you will see the message like here. Also you can not see any table now. But, If your data has any missing values in any of the raw. you will see the sum and percentage in the table dataframe.

```
Good Luck!! The data has no missing values
```