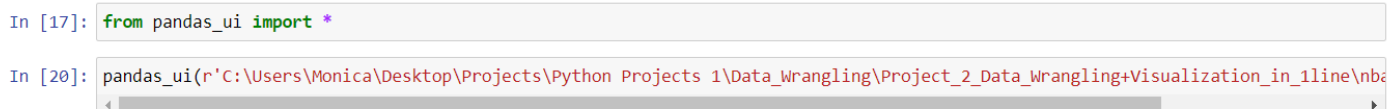


Project Name: Data-Wrangling Visualization with Bamboolib OR Pandasui in 1 line

Table of Contents

Demo
Overview
Motivation
Technical Aspect
Installation
Run/How to Use/Steps
Directory Tree/Structure of Project
To Do/Future Scope
Technologies Used/System Requirements/Tech Stack
Credits

Demo



```
In [17]: from pandas_ui import *  
  
In [20]: pandas_ui(r'C:\Users\Monica\Desktop\Projects\Python Projects 1\Data Wrangling\Project_2_Data_Wrangling+Visualization_in_1line\nb2')
```

Overview

This is a Data Wrangling with pandas-ui library.

Data wrangling involves processing the data in various formats like - merging, grouping, concatenating etc. for the purpose of analysing or getting them ready to be used with another set of data. Python has built-in features to apply these wrangling methods to various data sets to achieve the analytical goal.

This repository contains the code for Data Wrangling using python's various libraries.

It used pandas and pandas-ui libraries.

These libraries help to perform individually one particular functionality.

Pandas objects rely heavily on Numpy objects.

Here I am not using Bamboolib library because its paid so alternative to that is pandas-ui.

pandas_ui helps you wrangle & explore your data and create custom visualizations without digging through StackOverflow.

The purpose of creating this repository is to gain insights into how to wrangle or transform data.

These python libraries raised knowledge in discovering these libraries with practical use of it.

It leads to growth in my ML repository.

This above screenshot will help you to understand flow of output.

Motivation

The reason behind building this is, I know that wrangling costs analytics professionals as much as 80% of their time therefore, to become ML Professional this is one of the skillsets necessary in bucket-list. One very common truth of life is that, when we hear proverbs or saying, we do not understand it much or completely rather we can absorb it fully when we go through it. Right. That means, here information from advice format to practical self-experience format got transferred and then we could understand it completely. Similarly, in Data Wrangling is the process of converting and mapping data from its raw form to another format with the purpose of making it more valuable and appropriate for advance tasks such as Data Analytics and Machine Learning. For companies, their clients are important and so for employee because it's all about business and decisions taken for business. Data Wrangling exactly here comes for rescue as Data wrangling is the art of providing the right information to business analysts to make the right decision on time. Data wrangling also provides organisations with the right information in a short span of time to access the right information thereby helping make strategic decisions for the business. One of the most common steps taken in data science work is data wrangling to solve complex business problems. One of the important goal of Data Wrangling is to drive better decisions based on data in short time span and that is what I am trying to achieve by doing this project and that is also one of the important qualities that employee should have. Performing data wrangling in right direction will rescue an organization 6 on 10 times from drowning. When analyst can reach to precise decision through this process then it is good deal. Hence, I continue to gain knowledge while practicing the same and spread literary wings in tech-heaven.

Technical Aspect

Pandas_ui will automatically perform the specified operation on the interactive pandas dataframe and generates the python code simultaneously. It Generates statistics report powered by pandas profiling report. Dozens of panda's operations just a few clicks away using this library.

One of the most important reason for selecting pandas-ui library is because it has One of the most powerful functionalities in this tool is history which allows you to reverse, redo your actions on python-pandas code generated and the code can be saved for later use.

And hidden secret is, you can create powerful visual presentations with no effort and the code can be copied for later use.

Another reason for selecting and working with this library is, because it is easy to use means Quick onboarding with a minimal learning curve and it provides Private, secure, local functionality means All your data remains private and secure. Use it on your local machine

without the need for cloud access. Satisfies the requirements of your IT department. It provides flexibility as Seamless integration into your existing working environment. And it reduces 90% of coding time.

Pandas module mainly works with the tabular data. It contains Data Frame and Series. Pandas is 18 to 20 times slower than Numpy. Pandas is seriously a game changer when it comes to cleaning, transforming, manipulating and analyzing data.

Installation

Using intel core i5 9th generation with NVIDIA GFORCE GTX1650.

Windows 10 Environment Used.

Already Installed Anaconda Navigator for Python 3.x

The Code is written in Python 3.8.

If you don't have Python installed then please install Anaconda Navigator from its official site.

If you are using a lower version of Python you can upgrade using the pip package, ensuring you have the latest version of pip, *python -m pip install --upgrade pip and press Enter.*

Run/How to Use/Steps

Keep your internet connection on while running or accessing files and throughout too.

Follow this when you want to perform from scratch.

Open Anaconda Prompt, Perform the following steps:

```
cd <PATH>
```

```
pip install pandas
```

```
pip install pandas-ui
```

You can also create requirement.txt file as, `pip freeze > requirements.txt`
run files.

Follow this when you want to just perform on local machine.

Download ZIP File.

Right-Click on ZIP file in download section and select Extract file option, which will unzip file.

Move unzip folder to desired folder/location be it D drive or desktop etc.

Open Anaconda Prompt, write `cd <PATH>` and press Enter.

eg: `cd C:\Users\Monica\Desktop\Projects\Python Projects 1\13\Data_Wrangling\Project_2_Data_Wrangling+Visualization_in_1line`

In Anconda Prompt, `pip install -r requirements.txt` to install all packages.

Open in Jupyter Notebook, `<filename>.ipynb`

That is,

Open in Jupyter Notebook,

1)DataWrangling_Visualization_Bamboolib_OR_pandasui_in_1line.ipynb

This takes nba-player-stats-2019.csv file as input dataset.

Please be careful with spellings or numbers while typing filename and easier is just copy filename and then run it to avoid any silly errors.

Note: cd <PATH>

[Go to Folder where file is. Select the path from top and right-click and select copy option and paste it next to cd one space <path> and press enter, then you can access all files of that folder] [cd means change directory]

Directory Tree/Structure of Project

Folder: 13)Data_Wrangling > Project_2_Data_Wrangling+Visualization_in_1line

1)DataWrangling_Visualization_Bamboolib_OR_pandasui_in_1line.ipynb

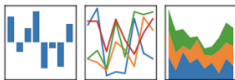
To Do/Future Scope

Can try another new library.

Technologies Used/System Requirements/Tech Stack

pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



Credits

Data Professors Channel