Learning Objectives

Data Analytics

Pandas

DPhi

Data Analytics: An Analogy

You are An Apple Farmer

- You own an apple farm and you want to know the number of apples you grow but you are too busy with the farm so you hire someone to count them. You sell your apples too, and you get your apple counter to keep a record of the number of apples you have in the beginning and at the end of the day, every day. Many days and months pass and you put sheet after sheet of the apple count together and you discover patterns and trends in the purchasing behaviour of your customers.

- The trends and patterns help you realise that during the colder season, your output of apples are the same but people buy less than during the summer. You then set out to dig deeper into this trend, and find ways to keep the sales of apples consistent throughout the year, beating your competitors at the game and becoming an apple farm tycoon.

Data Analytics: An Analogy

Apples are your data, tracking them is important, *analysis is key*

Apples are your products/assets, knowing their movement is important. For starters, you will know if your supply of apples matches the market's demand, as well as the consistency of the ratio of demand to supply throughout the year.

Reducing the cost of apple production gives you your profit. When you have enough data, you will find trends and patterns in your production. These trends can help you understand your own organisation better, help you reduce inefficiency and therefore reduce your apple production costs.

Data Analytics

- In simple words, data analysis is the process of collecting and organizing data in order to draw helpful conclusions from it. The process of data analysis uses analytical and logical reasoning to gain information from the data.

- The main purpose of data analysis is to find meaning in data so that the derived knowledge can be used to make informed decisions.

Real-life Examples

1. **Economics**
   Analyzing data to form patterns and understanding trends about how the economy in various sectors is growing, is something very essential for economists. Therefore, a lot of economists have started using Python and Pandas to analyze huge datasets.

2. **Recommendation Systems**
   We all have used Spotify or Netflix and been appalled at the brilliant recommendations provided by these sites. The recommendation system is possible only by learning and handling huge masses of data.

3. **Stock Prediction**
   The stock market is extremely volatile. However, that doesn't mean that it cannot be predicted. With the help of Pandas and a few other libraries like NumPy and matplotlib, we can easily make models which can predict how the stock markets turn out.

Why Pandas?

- In real-life, as a Data Scientist, we need to play with data and pandas essentially helps us to give a head start with it.
- Be it reading the data, analyzing /cleaning /changing / transforming. Pandas has a solution for all of it.
- It even helps in calculating basic statistics. If we go on there will be number of use-cases, we have summarized a few of them in the next slide.
- Pandas help in analyzing large volumes of data with ease.

Specific-Use Cases

- **Creating a data:** You can create your own dataset using pandas.

- **Reading data:** Often we get data in various formats such as csv, excel, in sql database, json etc. Pandas have some simple functions to perform this task

- **Analyzing data:** With the help of other libraries like matplotlib, it helps in statistical analysis and visualize data with plots such as histograms, scatter plots etc

- **Cleaning data and handling missingness:** Often Data Scientists get unclean data with a lot missing values and we need to have a solution to deal with it. Pandas facilitate in handling this issue.

Pandas Objects

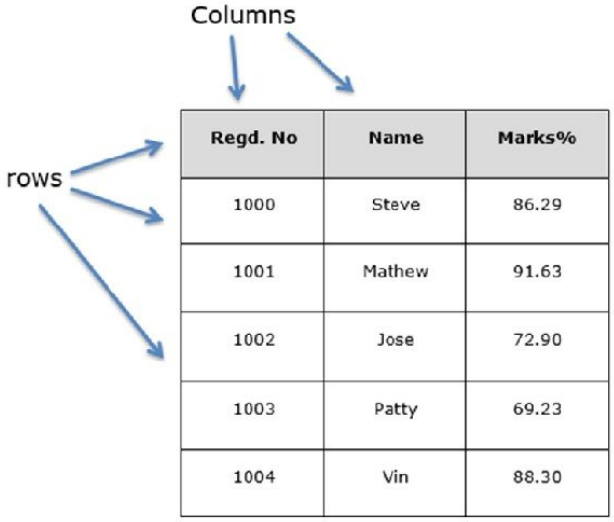
- At the core of the pandas library there are two fundamental data structures/objects:
  - Series
  - Data Frames

- **Series:** stores single column data along with an index. An index is just a way to "number" the Series object.

```
0    0.25
1    0.50
2    0.75
3    1.00
```

Pandas Objects

- **DataFrame:** is a two-dimensional tabular data structure with labeled axes (rows and columns). It is conceptually useful to think of a DataFrame object as a collection of Series objects.



Regd. No	Name	Marks%
1000	Steve	86.29
1001	Mathew	91.63
1002	Jose	72.90
1003	Patty	69.23
1004	Vin	88.30

Reading Material

- **Here is an exhaustive tutorial on pandas with clear use-cases:**

  https://www.learndatasci.com/tutorials/python-pandas-tutorial-complete-introduction-for-beginners/

Notebook for Session

- **Link to the Notebook:** https://dphi.tech/notebooks/852/manish_kc_06/day-1-notebook-introduction-to-pandas?

DPhi

That’s it for this unit. Thank you!

Feel free to post any queries on Discuss.

DPhi