Day 76- 90 days of Analytics: Pandas Basics

In today's video, we looked at the basic of Pandas

The following were mentioned

-Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

-To install pandas, we use the command: `pip install pandas`

-Once Pandas is installed, we import it in our applications using the **import** keyword: `import pandas`

-Pandas is usually imported under the **pd** alias. `import pandas as pd`

- A Pandas Series is like a column in a table. It is a one-dimensional array holding data of any type. Example

```
import pandas as pd
a = [1, 7, 2]
myvar = pd.Series(a)
print(myvar)
```

-If nothing else is specified, the values are labeled with their index number. First value has index 0, second value has index 1, …. This label can be used to access a specified value. Example `print(myvar[0])`

-We can also use a key/value object, like a dictionary, when creating a Series. The keys of the dictionary become automatically the labels. Example

```
import pandas as pd

temperature = {"day1": 420, "day2": 380, "day3": 390}

myvar = pd.Series(temperature)

print(myvar)
```

-Data sets in Pandas are usually multi-dimensional tables, called DataFrames. Series is like a column, a DataFrame is the whole table.

-A Pandas DataFrame is a 2 dimensional data structure, like a 2 dimensional array, or a table with rows and columns. Example

```
import pandas as pd
data = {
  "calories": [420, 380, 390],
  "duration": [50, 40, 45]
}
#load data into a DataFrame object:
df = pd.DataFrame(data)
print(df)
```

Pandas use the **loc** attribute to return one or more specified row(s). Example

Return row 0: `print(df.loc[0])` which is a pandas series.

-A simple way to store big data sets is to use CSV files (comma separated files). To read the content of a csv file, we use the read_csv() method. Example

```
df = pd.read_csv('data.csv')
```

To read the content of a Excel file, we use the read_excel() method. Example

```
df = pd.read_excel('data.xlsx','Sheet1')
```

-The head() method returns the first five rows of the data frame

-The tail() method returns the last five rows of the data frame

-To get specific columns from a data frame, we pass them as a list to the pandas object

-We can get information from the data frame using methods such as min(), max(), … and using relational operators such as ==, >, <, >=, …

-The describe() method gives a statistical description of a give series from the data frame

-The set_index() method is used to set index of the data frame from default to our preference

-The **inplace** attribute of the set_index() method set to True permit us to definitively set our index.

-We can also reset the index using the reset_index() method with the inplace attribute set to True

-To write to csv and Excel files, we use the to_csv() and the to_excel() methods respectively

Link to the YouTube Recording: https://www.youtube.com/watch?v=ckLZ7RyyuQI

#90daysofanalytics #community #dataanalysis #dataanalyst #microsoft #msexcel #SQL #powerbi #pythonprogramming #numpy