upGrad

Summary

Pandas

Pandas is a data analysis library built on NumPy.

Introduction to Pandas

Pandas:

Pandas

Pandas is an **open source** library which is known for its **rich applications** and utilities for all kinds of mathematical, financial and statistical functions

The different functions:

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

upGrad

Operator	Explanation	Example
pd.Series()	Create a pandas series from array-like objects	s = pd.Series([2, 4, 5, 6, 9]) print(s)
pd.date_range()	Create a series of type date-time	date_series = pd.date_range(start = '11-09-2017', end = '12-12-2017') date_series
Indexing using pd.Series()	Used for indexing explicitly	pd.Series([0, 1, 2], index = ['a', 'b', 'c'])
.Dataframe()	Create dataframes from dictionaries	df = pd.DataFrame({'name': ['Vinay', 'Kushal', 'Aman', 'Saif'], 'age': [22, 25, 24, 28], 'occupation': ['engineer', 'doctor', 'data analyst', 'teacher']}) df
pd.read_csv()	Read a CSV file as a dataframe	market_df = pd.read_csv("global_sales_data/market_fact.csv")
.head()	Look at the top and bottom entries of a dataframe	market_df.head()
.info()	Look at the data types of each column	market_df.info()
.describe()	Gives you a summary of all numeric columns	market_df.describe()
.columns	Gives you column names	market_df.columns
.shape	Gives the number of rows and columns	market_df.shape
.values	Extract the values of a dataframe as a NumPy array	market_df.values
.set_index()	Changing the index	market_df.set_index('Ord_id', inplace = True)

Loading, Indexing and Iterating data

Indexing and iterating through data:

Operator	Explanation
df.iloc	Position based indexing
df.loc	Label based indexing

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

upGrad

Operator	Explanation	Example
df[start_index:end_index]	Select the rows between the start and end indices	market_df[2:7]
df['column_name']	Select a single column from a dataframe	sales = market_df['Sales']
df.column_name	Select a single column from a dataframe	sales = market_df.Sales
df[['column_1', 'column_2', 'column_n']]	Select multiple columns in a dataframe	market_df[['Cust_id', 'Sales', 'Profit']].head()

Filtering data in a dataframe

Filtering

Filtering essentially is a way of pulling out **specific datasets** or parts of the datasets to get the **relevant data** for analysis

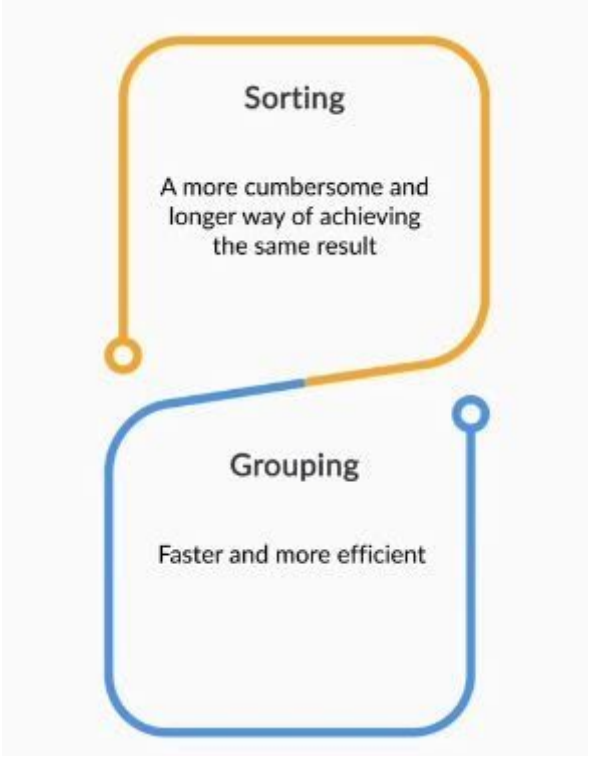
Sorting in a dataframe

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

Sorting

A way of **arranging** data as per some **criteria**

How it differs from grouping:



Editing Dataframes

Functions for editing in a dataframe:

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

upGrad

Operator	Explanation	Example
pd.merge()	Merge multiple dataframes using common columns/keys	df_1 = pd.merge(market_df, customer_df, how='inner', on='Cust_id')
pd.concat()	Concatenate two dataframes	pd.concat([df1, df2], axis = 0)

Grouping Data

Grouping

A way of efficiently picking a **part of the dataset** based on some criteria, looking at some specific elements within that and then providing it into a **new or existing** dataset

Functions for grouping:

Operator	Explanation
df.groupby()	Grouping data using a categorical variable
.sum()	Concatenate two dataframes
.sort_values()	Sorting values

Lambda functions and Pivot tables

Lambda functions and pivot tables in Pandas:

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

upGrad

Operator	Explanation	Example
.apply()	Apply lambda function to a row	master_df['Profit'] = master_df['Profit'].apply(lambda x: round(x, 1))
.pivot_table()	Create a pivot table	master_df.pivot_table(values = 'Sales', index = 'Customer_Segment', aggfunc = 'mean')

Outputting data

Outputting data to different file formats in Pandas:

Operator	Explanation	Example
.to_csv()	Output to csv	master_df.to_csv('master_sheet.csv')
.to_html()	Output to html	master_df.to_html('test.html')
.to_pickle()	Output as a pickle file	master_df.to_pickle('dummy.pkl')

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved

upGrad

Disclaimer: All content and material on the UpGrad website is copyrighted material, either belonging to UpGrad or its bonafide contributors and is purely for the dissemination of education. You are permitted to access print and download extracts from this site purely for your own education only and on the following basis:

- You can download this document from the website for self-use only.
- Any copies of this document, in part or full, saved to disc or to any other storage medium may only be used for subsequent, self-viewing purposes or to print an individual extract or copy for non-commercial personal use only.
- Any further dissemination, distribution, reproduction, copying of the content of the document herein or the uploading thereof on other websites or use of content for any other commercial/unauthorized purposes in any way which could infringe the intellectual property rights of UpGrad or its contributors, is strictly prohibited.
- No graphics, images or photographs from any accompanying text in this document will be used separately for unauthorised purposes.
- No material in this document will be modified, adapted or altered in any way.
- No part of this document or UpGrad content may be reproduced or stored in any other web site or included in any public or private electronic retrieval system or service without UpGrad's prior written permission.
- Any rights not expressly granted in these terms are reserved.

© Copyright 2019. UpGrad Education Pvt. Ltd. All rights reserved