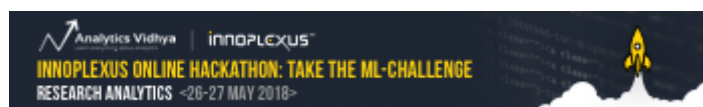


**f** (<https://www.facebook.com/AnalyticsVidhya>)

**t** (<https://twitter.com/analyticsvidhya>)

**g+** (<https://plus.google.com/+Analyticsvidhya/posts>)

**in** (<https://www.linkedin.com/groups/Analytics-Vidhya-Learn-everything-about-5057165>)



([https://datahack.analyticsvidhya.com/contest/innoplexus-hiring-hackathon/?utm\\_source=AVhometop](https://datahack.analyticsvidhya.com/contest/innoplexus-hiring-hackathon/?utm_source=AVhometop))

Home (<https://www.analyticsvidhya.com/>) > Infographics (<https://www.analyticsvidhya.com/blog/category/infographics/>) > Cheat...

# CheatSheet: Data Exploration using Pandas in Python

INFOGRAPHICS (<https://www.analyticsvidhya.com/blog/category/infographics/>)

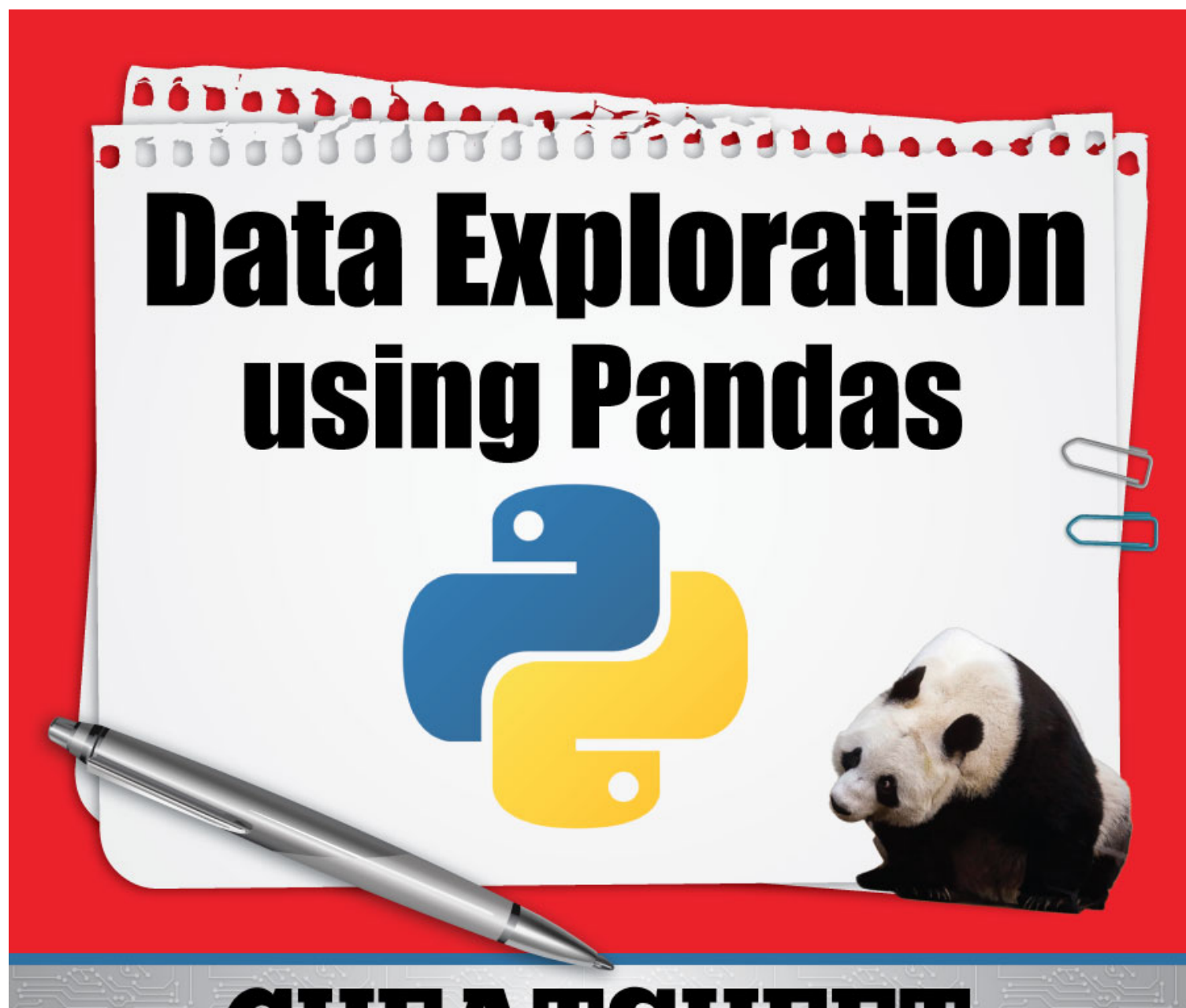
SHARE **f** (<http://www.facebook.com/sharer.php?u=https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/&t=CheatSheet:%20Data%20Exploration%20using%20Pandas%20in%20Python>) **t** (<https://twitter.com/home?status=CheatSheet:%20Data%20Exploration%20using%20Pandas%20in%20Python+https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/>) **g+** (<https://plus.google.com/share?url=https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/>) **p** (<http://pinterest.com/pin/create/button/?url=https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/&media=https://www.analyticsvidhya.com/wp-content/uploads/2015/07/11111.jpg&description=CheatSheet:%20Data%20Exploration%20using%20Pandas%20in%20Python>)

## Introduction

If some one would ask me to mention 2 most important libraries in Python for data science, I'll probably name "pandas" and "scikit-learn". Pandas for the capability to read datasets in DataFrames, exploring and making them ready for modeling / machine learning and Scikit-learn for actually learning from these features created in Pandas.

While there are quite a few cheat sheets to summarize what scikit-learn brings to the table, there isn't one I have come across for Pandas. Hence, we thought of creating a cheat sheet for common data exploration operations in Python using Pandas. If you think we have missed any thing in the cheat sheet, please feel free to mention it in comments.

The PDF version of the sheet can be downloaded from [here](http://discuss.analyticsvidhya.com/t/download-cheatsheet-for-data-exploration-using-pandas-in-python/2378) (http://discuss.analyticsvidhya.com/t/download-cheatsheet-for-data-exploration-using-pandas-in-python/2378) (so that you can copy paste codes)



# CHEATSHEET

## 1. Reading and Writing Data

### a. Reading a CSV file

```
>>>df=pd.read_csv('AnalyticsVidhya.csv')
```

### b. Writing content of data frame to CSV file

```
>>>df.to_csv('AV.csv')
```

### c. Reading an Excel file

```
>>>df=pd.read_excel('AV.xlsx','sheet1')
```

### d. Writing content of data frame to Excel file

```
>>>df.to_excel('AV2.xlsx',sheet_name='sheet2')
```



## 2. Getting Preview of Dataframe

### a. Looking at top n records

```
>>>df.head(5)
```

### b. Looking at bottom n records

```
>>>df.tail(5)
```

### c. View columns name

```
>>>df.columns
```

## 3. Rename Columns of Data Frame

### a. Rename method helps to rename column of data frame.

```
>>>df2=df.rename(columns={'old_columnname':'new_columnname'})
```

This statement will create a new data frame with new column name.

### b. To rename the column of existing data frame, set inplace=True.

```
>>>df.rename(columns={'old_columnname':'new_columnname'}, inplace=True)
```

## 4. Selecting Columns or Rows

### a. Accessing sub data frames

```
>>>df[['column1','column2']]
```

### b. Filtering Records

```
>>>df[df['column1']>10]
```

```
>>>dff[(dff['column1']>10) & dff['column2']==30]
```



```
>>>df[(df['column1']>10) | df['column2']==30]
```

## 5. Handling Missing Values

This is an inevitable part of dealing with data. To overcome this hurdle, use `dropna` or `fillna` function.

a. `dropna`: It is used to drop rows or columns having missing data

```
>>>df1.dropna()
```

b. `fillna`: It is used to fill missing values

```
>>>df2.fillna(value=5) #It replaces all missing values with 5
```

```
>>>mean=df2['column1'].mean()
```

```
>>>df2['column1'].fillna(mean) #It replaces all missing values of column1 with mean of available values
```

## 6. Creating New Columns

New column is a function of existing columns

```
>>>df['NewColumn1']=df['column2'] #Create a copy of existing column2
```

```
>>>df['NewColumn2']=df['column2']+10 #Add 10 to existing column2 then create a new one
```

```
>>>df['NewColumn3']= df['column1'] + df['column2'] #Add elements of column1 and column2 then create new column
```

## 7. Aggregate

a. `Groupby`: `Groupby` helps to perform three operations

- i. Splitting the data into groups
- ii. Applying a function to each group individually
- iii. Combining the result into a data structure

```
>>>df.groupby('column1').sum()
>>>df.groupby(['column1','column2']).count()
```



b. `Pivot Table`: It helps to generate data structure. It has three components  
index, columns and values (similar to excel)

```
>>>pd.pivot_table(df, values='column1', index=['column2','column3'], columns=['column4'])
```



```
>>>pd.pivot_table(df, values= column1 , index=[ column2 , column3 ], columns=[ column4 ])
```

**By default, it shows the sum of values column but you can change it using argument aggfunc**

```
>>>pd.pivot_table(df, values='column1', index=['column2','column3'], columns=['column4'], aggfunc=len)
```

**#it shows count**

**c. Cross Tab: Cross Tab computes the simple cross tabulation of two factors.**

```
>>>pd.crosstab(df.column1, df.column2)
```

## 8. Merging/ Concatenating DataFrames

**It performs similar operation like we do in SQL.**

**a. Concatenating: It concatenate two or more data frames based on their columns.**

```
>>>pd.concat([df1,df2])
```

**b. Merging: We can perform left, right and inner join also.**

```
>>>pd.merge(df1, df2, on='column1', how='inner')
```

```
>>>pd.merge(df1, df2, on='column1', how='left')
```

```
>>>pd.merge(df1, df2, on='column1', how='right')
```

```
>>>pd.merge(df1, df2, on='column1', how='outer')
```

## 9. Applying function to element, column or dataframe

**a. Map: It iterates over each element of a series.**

```
>>>df['column1'].map(lambda x: 10+x) #this will add 10 to each element of column1
```

```
>>>df['column2'].map(lambda x: 'AV'+x) #this will concatenate "AV" at the beginning of
each element of column2 (column format is string)
```

**b. Apply: As the name suggests, applies a function along any axis of the DataFrame.**

```
>>>df[['column1','column2']].apply(sum) #it will returns the sum of all the values of
column1 and column2.
```

**c. ApplyMap: This helps to apply a function to each element of dataframe.**

```
>>>func = lambda x: x+2
```

```
>>>df.applymap(func) #it will add 2 to each element of dataframe (all columns of
dataframe must be numeric type)
```

## 10. Identify unique values

Function unique helps to return unique values of a column.

```
>>>df['Column1'].unique()
```

## 11. Basic Stats

Pandas helps to understand the data using basic statistical methods.

**a. describe:** This returns the quick stats (count, mean, std, min, first quartile, median, third quartile, max) on suitable columns

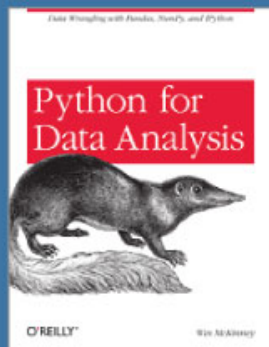
```
>>>df.describe()
```

**b. covariance:** It returns the co-variance between suitable columns.

```
>>>df.cov()
```

**c. correlation:** It returns the co-variance between suitable columns.

```
>>>df.corr()
```



To learn more, we recommend  
Wes McKinney's **Python for Data Analysis**  
Book for Learning Pandas

For more resources on analytics / data science, visit  
**[www.analyticsvidhya.com](http://www.analyticsvidhya.com)**



(<https://www.analyticsvidhya.com/wp-content/uploads/2015/07/cheatsheet.jpg>)

You can keep this cheat sheet handy while performing data exploration. Download the PDF Version here (<http://discuss.analyticsvidhya.com/t/download-cheatsheet-for-data-exploration-using-pandas-in-python/2378/1>).

**If you like what you just read & want to continue your analytics learning, subscribe to our emails (<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>), follow us on twitter (<http://twitter.com/analyticsvidhya>) or like our facebook page (<http://facebook.com/analyticsvidhya>).**

You can also read this article on Analytics Vidhya's Android APP




([play.google.com/store/apps/details?](https://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm_source=blog_article&utm_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1)

[id=com.analyticsvidhya.android&utm\\_source=blog\\_article&utm\\_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1](https://play.google.com/store/apps/details?id=com.analyticsvidhya.android&utm_source=blog_article&utm_campaign=blog&pcampaignid=MKT-Other-global-all-co-prtnr-py-PartBadge-Mar2515-1))

#### Share this:

 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=linkedin&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=facebook&nb=1>)

179

 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=google-plus-1&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=twitter&nb=1>)

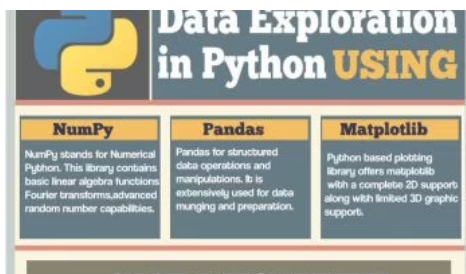
 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=pocket&nb=1>)

 (<https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/?share=reddit&nb=1>)

---

## RELATED

---



(<https://www.analyticsvidhya.com/blog/2017/02/top-28-cheat-sheets-for-machine-learning-data-science-probability-sql-big-data/>)

**Top 28 Cheat Sheets for Machine Learning, Data Science, Probability, SQL & Big Data**

(<https://www.analyticsvidhya.com/blog/2017/02/top-28-cheat-sheets-for-machine-learning-data-science-probability-sql-big-data/>)

February 17, 2017

In "Big data"

(<https://www.analyticsvidhya.com/blog/2015/06/infographic-cheat-sheet-data-exploration-python/>)

**Cheat Sheet for Exploratory Data Analysis in Python**

(<https://www.analyticsvidhya.com/blog/2015/06/infographic-cheat-sheet-data-exploration-python/>)

June 8, 2015

In "Business Analytics"

(<https://www.analyticsvidhya.com/blog/2015/04/pycon-montreal-2015-data-science-workshops/>)

**PyCon Montreal 2015 tutorials - Hands-on way to learn Data Science in Python**

(<https://www.analyticsvidhya.com/blog/2015/04/pycon-montreal-2015-data-science-workshops/>)

April 14, 2015

In "Python"

TAGS: CHEATSHEET (<https://www.analyticsvidhya.com/blog/tag/cheatsheet/>), DATA ANALYSIS (<https://www.analyticsvidhya.com/blog/tag/data-analysis/>), INFOGRAPHIC (<https://www.analyticsvidhya.com/blog/tag/infographic/>), PANDAS (<https://www.analyticsvidhya.com/blog/tag/pandas/>), PYTHON (<https://www.analyticsvidhya.com/blog/tag/python/>)

#### Next Article

**Lead/Manager – Equifax – Mumbai – (5-7 years of Experience)**

(<https://www.analyticsvidhya.com/blog/2015/07/leadmanager-equifax-mumbai-5-7-years-experience/>)

#### Previous Article

**Learning path for Tableau – visualization tool with awesome execution capabilities!**

(<https://www.analyticsvidhya.com/blog/2015/07/learning-path-tableau-worlds-fastest-growing-visualization-tool/>)





(<https://www.analyticsvidhya.com/blog/author/kunalj/>)

Author

## Kunal Jain (<https://www.analyticsvidhya.com/blog/author/kunalj/>)

Kunal is a post graduate from IIT Bombay in Aerospace Engineering. He has spent more than 10 years in field of Data Science. His work experience ranges from mature markets like UK to a developing market like India. During this period he has lead teams of various sizes and has worked on various tools like SAS, SPSS, Qlikview, R, Python and Matlab.

This article is quite old now and you might not get a prompt response from the author. We would request you to post this comment on Analytics Vidhya **Discussion portal** (<https://discuss.analyticsvidhya.com/>) to get your queries resolved.

## ONE COMMENT

---

**venugopal rao says:**

REPLY

SEPTEMBER 8, 2016 AT 5:18 AM ([HTTPS://WWW.ANALYTICSVIDHYA.COM/BLOG/2015/07/11-STEPS-PERFORM-DATA-ANALYSIS-PANDAS-PYTHON/#COMMENT-115779](https://www.analyticsvidhya.com/blog/2015/07/11-steps-perform-data-analysis-pandas-python/#comment-115779))

All together one place very useful

## LEAVE A REPLY

Your email address will not be published.

Comment


Name (required)





Email (required)

Website

SUBMIT COMMENT

## TOP ANALYTICS VIDHYA USERS

Rank	Name	Points
1	 vopani ( <a href="https://datahack.analyticsvidhya.com/user/profile/Rohan_Rao">https://datahack.analyticsvidhya.com/user/profile/Rohan_Rao</a> )	8714

2		SRK ( <a href="https://datahack.analyticsvidhya.com/user/profile/SRK">https://datahack.analyticsvidhya.com/user/profile/SRK</a> )	8287
3		aayushmnit ( <a href="https://datahack.analyticsvidhya.com/user/profile/aayushmnit">https://datahack.analyticsvidhya.com/user/profile/aayushmnit</a> )	7439
4		mark12 ( <a href="https://datahack.analyticsvidhya.com/user/profile/mark12">https://datahack.analyticsvidhya.com/user/profile/mark12</a> )	6269
5		sonny ( <a href="https://datahack.analyticsvidhya.com/user/profile/sonny">https://datahack.analyticsvidhya.com/user/profile/sonny</a> )	5937

[More Rankings \(http://datahack.analyticsvidhya.com/users\)](http://datahack.analyticsvidhya.com/users)

## POPULAR POSTS

- A Complete Tutorial to Learn Data Science with Python from Scratch (<https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-learn-data-science-python-scratch-2/>)
- Essentials of Machine Learning Algorithms (with Python and R Codes) (<https://www.analyticsvidhya.com/blog/2017/09/common-machine-learning-algorithms/>)
- 7 Types of Regression Techniques you should know! (<https://www.analyticsvidhya.com/blog/2015/08/comprehensive-guide-regression/>)
- Understanding Support Vector Machine algorithm from examples (along with code) (<https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/>)
- A comprehensive beginner's guide to create a Time Series Forecast (with Codes in Python) (<https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/>)
- 6 Easy Steps to Learn Naive Bayes Algorithm (with codes in Python and R) (<https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/>)
- A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python) (<https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in->)

python/)

- A Complete Tutorial on Time Series Modeling in R  
(<https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/>)

## RECENT POSTS

---



(<https://www.analyticsvidhya.com/blog/2018/05/analytics-vidhya-trainings-and-datahack-platform-update/>)

**Exciting updates from Analytics Vidhya – Launch of Trainings and DataHack platform update!**

(<https://www.analyticsvidhya.com/blog/2018/05/analytics-vidhya-trainings-and-datahack-platform-update/>)

KUNAL JAIN , MAY 22, 2018



(<https://www.analyticsvidhya.com/blog/2018/05/deep-learning-faq/>)

**12 Frequently Asked Questions on Deep Learning (with their**

**answers)! (<https://www.analyticsvidhya.com/blog/2018/05/deep-learning-faq/>)**

ANALYTICS VIDHYA CONTENT TEAM , MAY 21, 2018



bengaluru/)

(<https://www.analyticsvidhya.com/blog/2018/05/announcing-datahack-summit-2018-bengaluru/>)

**Announcing DataHack Summit 2018 – Bengaluru, 22 – 24 November 2018**

(<https://www.analyticsvidhya.com/blog/2018/05/announcing-datahack-summit-2018-bengaluru/>)

KUNAL JAIN , MAY 18, 2018



understand-coding/)

(<https://www.analyticsvidhya.com/blog/2018/05/19-data-science-tools-for-people-dont-understand-coding/>)

**19 Data Science and Machine Learning Tools for people who Don't Know Programming**

(<https://www.analyticsvidhya.com/blog/2018/05/19-data-science-tools-for-people-dont-understand-coding/>)

AARSHAY JAIN , MAY 16, 2018





(<http://www.edvancer.in/certified-data-scientist-with-python->

[course?utm\\_source=AV&utm\\_medium=AVads&utm\\_campaign=AVadsnonfc&utm\\_content=pythonav](http://www.edvancer.in/certified-data-scientist-with-python-course?utm_source=AV&utm_medium=AVads&utm_campaign=AVadsnonfc&utm_content=pythonav))

## GET CONNECTED

---



16,095

FOLLOWERS

(<http://www.twitter.com/analyticsvidhya>)



2,826

FOLLOWERS

(<https://plus.google.com/+Analyticsvidhya>)



FOLLOWERS

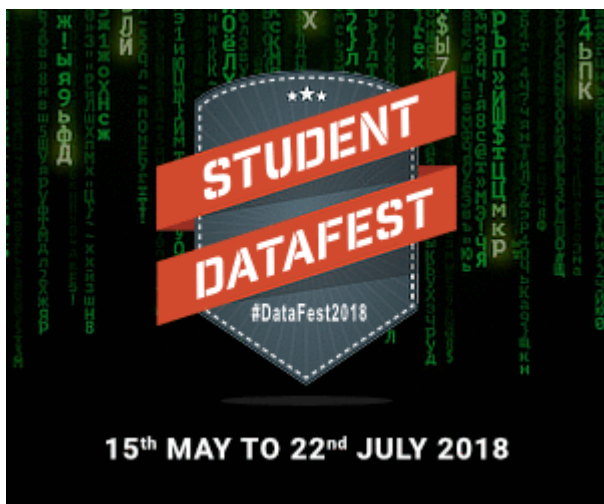
(<http://www.facebook.com/Analyticsvidhya>)



Email

SUBSCRIBE

(<http://feedburner.google.com/fb/a/mailverify?uri=analyticsvidhya>)



(<https://analyticsvidhya.com/student-datafest-2018/?>

[utm\\_source=AVblogbottom](https://analyticsvidhya.com/student-datafest-2018/?utm_source=AVblogbottom))



in-deep-learning/?utm\_source=AV\_Blogbottom)

(https://www.analyticsvidhya.com/about-me/write/)
 (https://www.analyticsvidhya.com/about-

## DATA SCIENTISTSCOMPANIES

### Blog

(https://www.analyticsvidhya.com/blog/)

### Hackathon

(https://datahack.analyticsvidhya.com/contest/)

### Discussions

(https://discuss.analyticsvidhya.com/)

### Apply Jobs

(https://www.analyticsvidhya.com/jobs/)

### Leaderboard

(https://datahack.analyticsvidhya.com/users/)

### Contact Us

(https://www.analyticsvidhya.com/contact/)

### Post Jobs

(https://www.analyticsvidhya.com/corporate/)

### Trainings

(https://datahack.analyticsvidhya.com/course/)

### Hiring

(https://datahack.analyticsvidhya.com/)

### Advertise

(https://www.analyticsvidhya.com/contact/)

### Relists

(https://www.analyticsvidhya.com/users/)

### Relists

(https://www.analyticsvidhya.com/users/)

## JOIN OUR COMMUNITY :

### f

(https://www.facebook.com/analyticsvidhya/)

### t

(https://twitter.com/analyticsvidhya/)

### in

(https://www.linkedin.com/company/analyticsvidhya/)

### g+

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

### in

(https://plus.google.com/+AnalyticsVidhya/)

Subscribe to emailer

>

© Copyright 2013-2018 Analytics Vidhya.

Privacy Policy (https://www.analyticsvidhya.com/privacy-policy/)

Terms of Use (https://www.analyticsvidhya.com/terms/)

Refund Policy (https://www.analyticsvidhya.com/refund-policy/)

Don't have an account? Sign up (l