

**KANDULA SRINIVASA REDDY MEMORIAL COLLEGE OF ENGINEERING
(AUTONOMOUS)**

KADAPA-516003. AP

(Approved by AICTE, Affiliated to JNTUA, Ananthapuramu, Accredited by NAAC)

(An ISO 9001-2008 Certified Institution)

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING



**CERTIFICATION COURSE
ON
“PYTHON FOR DATA SCIENCE”**

**Resource Persons : Dr. G.HemaLatha, HoD Dept. of ECE &
Mr.V.Gopi Tilak, Assistant Professor, Dept. of ECE, KSRMCE**

**Coordinator : Mrs. T.Umamaheswari, Assistant Professor, Dept. of
ECE, KSRMCE**

Duration : 29/06/2022 to 29/07/2022



K.S.R.M. COLLEGE OF ENGINEERING

(UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution

Lr./KSRMCE/ECE/ Certificate course on python for Data Science/2022-23/



Date: 20.06.2022

To

The Principal,
K.S.R.M. College of Engineering,
Kadapa.

Respected Sir,

Sub : KSRMCE – (Department of ECE) – Permission to conduct certification course on: "Python for Data Science" Request for granting permission to conduct -Reg.

Respected Sir,

With reference to the cited, the Department of ECE is planning to conduct a certification course on Python for Data Science for B.Tech IV SEM students from **29/06/2022 to 29/07/2022** at DSP lab. In this regard, we kindly request you to grant permission to conduct a certification course. This is submitted for your kind perusal.

Thanking you sir,

Yours Faithfully
Coordinator

(T Umamaheswari)

*Forwarded to the
Principal Sir
S. V. M.*

Permitted
V. S. S. Murthy
PRINCIPAL
K.S.R.M. COLLEGE OF ENGINEERING
KADAPA-516005, (A.P.)



K.S.R.M. COLLEGE OF ENGINEERING

(UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Lr./KSRMCE/ECE/ Certificate course on python for Data Science/2022-23/

Date: 22.06.2022

Circular

All the IV sem students of KSRMCE are hereby informed that the ECE Department is going to organize the Certification course on “**Python for Data Science**” from 29/06/2022 to 29/07/2022 at DSP Lab.

In this connection, all the students who are interested to participate can register by using the following link: <https://forms.gle/eBR5543SW1ypYAf9>

Coordinator:

T. Umamaheswari, Assistant Professor, Department of ECE.

Co-Coordinator:

K.Divya Lakshmi, Assistant Professor, Department of ECE.

V. S. S. Murthy

Principal

PRINCIPAL

K.S.R.M. COLLEGE OF ENGINEERING
KADAPA-516005, (A.P.),

Cc to:

The Management /Deans/HoDs/IQAC for information



/krmce.ac.in

Follow Us:



/krmceofficial

KSRM COLLEGE OF ENGINEERING (AUTONOMOUS)

Registration form of Python for data science

* Indicates required question

1. NAME OF THE STUDENT *

2. Roll Number *

3. Department *

4. section *

1 point

Mark only one oval.

A

B

C

5. MAIL-Id *

This content is neither created nor endorsed by Google.

Google Forms



K.S.R.M. COLLEGE OF ENGINEERING

(UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.
An ISO 14001:2004 & 9001: 2015 Certified Institution



Registered students for Certification Course on Python for Data Science(29/06/2022 – 29/07/2022)

S.NO	Roll Number	NAME OF THE STUDENT	Department	section	MAIL-Id
1	209Y1A0404	A Uday Kumar Reddy	ECE	A	209y1a0404@ksrmce.ac.in
2	209Y1A0406	B.MANASWITHA	ECE	A	209y1a0406@ksrmce.ac.in
3	209y1a0407	B sai krishna achari	ECE	A	209y1a0407@ksrmce.ac.in
4	209y1a0410	Bharath simha reddy	ECE	A	209y1a0410@ksrmce.ac.in
5	209Y1A0414	B. RAVI KANTH REDDY	ECE	A	209y1a0414@ksrmce.ac.in
6	209y1a0417	B.sangeetha	ECE	A	209y1a0417@ksrmce.ac.im
7	209Y1A0418	Botta.kalyani	ECE	A	209y1a0418@ksrmce.ac.in
8	209Y1A0423	Anuhya. C	ECE	A	209y1a0423@ksrmce.ac.in
9	209Y1A0427	Chavva Rajeshwar reddy	ECE	A	209y1a0427@ksrmce.ac.in
10	209Y1A0429	Chennuru Karthik Reddy	ECE	A	209y1a0428@ksrmce.ac.in
11	209Y1A0431	C Rajasekhar naidu	ECE	A	209y1a0431@ksrmce.ac.in
12	209Y1A0434	C siva pavani	ECE	A	209Y1A0434@ksrmce ac. In
13	209y1a0441	D.swetha	ECE	A	209y1a0441@ksrmce.ac
14	209Y1A0448	G. Raghu	ECE	A	209y1a0448@ksrmce.ac.in
15	209y1a0450	G Nithin kumar reddy	ECE	A	209y1a0450@ksrmce.ac.in
16	209y1a0453	G.Nagendra prasad	ECE	A	209y1A0453@ksrm.ac.ce
17	209Y1A0458	G Guru Prasad	ECE	A	209y1a0458@ksrmce.ac.in
18	209Y1A0462	G. Hari babu	ECE	B	209y1a0462@ksrmce.ac.in
19	209Y1A0471	K.Ramya sree	ECE	B	209Y1A0471@ksrmce.ac.in
20	209Y1A0473	K. Deepa Asritha	ECE	B	209Y1A0473@ksrmce.ac.in
21	209Y1A0489	M.Avinash reddy	ECE	B	209Y1A0489@ksrmce.ac.in
22	209y1a04a8	Pandla Chandra Sekhar	ECE	B	209y1a04a8@ksrmce.ac.in
23	209Y1A04B4	Peram Roja Reddy	ECE	B	209y1a04b4@ksrmce.ac.in
24	209y1a04b7	Ramireddy gari Bharathi	ECE	B	209y1a04b7@ksrmce.ac.in
25	209y1a04d5	S siva ruchitha	ECE	C	209y1a04d5@gmail.com
26	209y1a04e8	v harshavardhan	ECE	C	209y1a04e8@ksrmce.ac.in
27	219Y5A0410	C Lokesh Yadav	ECE	C	219y5a0410@ksrmce.ac.in

Python for Data Science

Course Objectives:

To understand libraries of Python and dealing with missing data.

- To know the various classification and regression models.
- To analyze practical cases.

Course Outcomes:

After completion of the course the students are able to

1. Use the various libraries of python.
2. Analyze the various classification and regression models.
3. Apply the models on the data sets

Module 1: Introduction to Python- Introduction to Spyder, Variables and Data types, Operators, Lists, Tuples, Dictionary, .

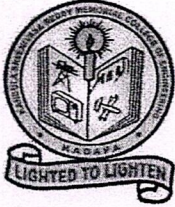
Module 2: Reading data- Pandas Data frames, Control structures & Functions, Exploratory data analysis-Data Visualization, Dealing with missing data.

Module 3: Predictive Modeling- Linear Regression, Model Assessment, Diagnostics to Improve Linear Model Fit, Cross Validation, Classification, Logistic Regression, Performance Measures, K - Nearest Neighbors (kNN), K - means Clustering, Logistic Regression , Decision Trees, Multiple Linear Regression.

Module 4: Case studies- Introduction to Classification Case Study, Case Study on Classification, Introduction to Regression Case Study, Case Study on Regression.

Text Books:

1. Introduction to linear algebra - by Gilbert Strang
2. Applied statistics and probability for engineers – by Douglas Montgomery
3. Mastering python for data science, Samir Madhavan



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - AUTONOMOUS)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA,

Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



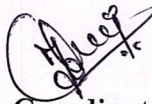
Department of Electronics & Communication Engineering

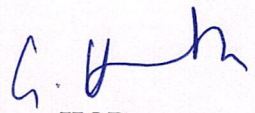
Certificate Course on Python for DataScience (29/06/2022 – 29/07/2022)

Schedule

S.No	Date	Time	Name of the Faculty	Topic
Module-1				
1.	29/06/2022	04:00PM-05:00PM	Dr.G.Hemalatha Mr.V.Gopi Tilak	Introduction to Python- Introduction to Spyder
2.	30/06/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Variables and Data types, Operators
3.	01/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Lists, Tuples, Dictionary
4.	02/07/2022	03:00PM-04:00PM	Mr.V.Gopi Tilak	Sets, Numpy, Matrix,
5.	02/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Linear algebra
Module-2				
6.	04/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Reading data- Pandas Data frames
7.	05/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Control structures & Functions
8.	06/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Exploratory data analysis
9.	07/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Data Visualization
10.	08/07/2022	04:00PM-05:00PM	Mr.V.Gopi Tilak	Dealing with missing data
Module-3				
11.	09/07/2022	03:00PM-04:00PM	Dr.G.Hemalatha	Predictive Modeling-
12.	09/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Linear Regression
13.	11/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Model Assessment
14.	12/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Diagnostics to Improve Linear Model Fit

15.	13/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Cross Validation
16.	14/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Classification
17.	15/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Logistic Regression
18.	16/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Performance Measures
19.	18/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	K - Nearest Neighbors (kNN)
20.	19/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	K - means Clustering
21.	20/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Logistic Regression
22.	21/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Decision Trees
23.	22/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Multiple Linear Regression
Module-4				
24.	23/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Multiple Linear Regression
25.	24/07/2022	03:00PM-04:00PM	Dr.G.Hemalatha	Case studies- Introduction to Classification Case Study
26.	24/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Case Study on Classification
27.	25/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Case Study on Classification
28.	26/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Introduction to Regression Case Study
29.	27/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Introduction to Regression Case Study
30.	28/07/2022	03:00PM-04:00PM	Dr.G.Hemalatha	Case Study on Regression
31.	28/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Case Study on Regression
32.	29/07/2022	03:00PM-04:00PM	Dr.G.Hemalatha	Case Study on Regression
33.	29/07/2022	04:00PM-05:00PM	Dr.G.Hemalatha	Case Study on Regression


Coordinator


HOD
Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003



**K.S.R.M. COLLEGE OF ENGINEERING
(UGC-AUTONOMOUS)**

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.
An ISO 14001:2004 & 9001: 2015 Certified Institution



Week 1 Attendance for Certification Course on Python for Data Science

From 29-06-2022 to 02-07-2022

S.NO	Roll Number	NAME OF THE STUDENT	29/6/2022	30/6/2022	1/7/2022	2/7/2022
1	209Y1A0404	A Uday Kumar Reddy	Uday	Uday	Uday	Uday
2	209Y1A0406	B.MANASWITHA	Manaswita	Manaswita	Manaswita	Manaswita
3	209y1a0407	B sai krishna achari	B.sai krishna	B.sai krishna	B.sai krishna	B.sai krishna
4	209y1a0410	Bharath simha reddy	Bharath	Bharath	Bharath	Bharath
5	209Y1A0414	B. RAVI KANTH REDDY	BRKR	BRKR	BRKR	BRKR
6	209y1a0417	B.sangeetha	Sangeetha	Sangeetha	Sangeetha	Sangeetha
7	209Y1A0418	Botta.kalyani	Kalyani	Kalyani	Kalyani	Kalyani
8	209Y1A0423	Anuhya. C	Anuhya	Anuhya	Anuhya	Anuhya
9	209Y1A0427	Chavva Rajeshwar reddy	C.Rajeshwar	C.Rajeshwar	C.Rajeshwar	C.Rajeshwar
10	209Y1A0429	Chennuru Karthik Reddy	Karthik	Karthik	Karthik	Karthik
11	209Y1A0431	C Rajasekhhar naidu	Rajasekhhar	Rajasekhhar	Rajasekhhar	Rajasekhhar
12	209Y1A0434	C siva pavani	Pavani	Pavani	Pavani	Pavani
13	209y1a0441	D.swetha	Swetha	Swetha	Swetha	Swetha
14	209Y1A0448	G. Raghu	Raghu	Raghu	Raghu	Raghu
15	209y1a0450	G Nithin kumar reddy	Nithin	Nithin	Nithin	Nithin
16	209y1a0453	G.Nagendra prasad	Nagendra	Nagendra	Nagendra	Nagendra
17	209Y1A0458	G Guru Prasad	Guru	Guru	Guru	Guru
18	209Y1A0462	G. Hari babu	Hari	Hari	Hari	Hari
19	209Y1A0471	K.Ramya sree	Ramya	Ramya	Ramya	Ramya
20	209Y1A0473	K. Deepa Asritha	Deepa	Deepa	Deepa	Deepa
21	209Y1A0489	M.Avinash reddy	Avinash	Avinash	Avinash	Avinash
22	209y1a04a8	Pandla Chandra Sekhar	Chandra	Chandra	Chandra	Chandra
23	209Y1A04B4	Peram Roja Reddy	Roja	Roja	Roja	Roja
24	209y1a04b7	Ramireddy gari Bharathi	Bharathi	Bharathi	Bharathi	Bharathi
25	209y1a04d5	S siva ruchitha	Ruchitha	Ruchitha	Ruchitha	Ruchitha
26	209y1a04e8	v harshavardhan	Harsha	Harsha	Harsha	Harsha
27	219Y5A0410	C Lokesh Yadav	Lokesh	Lokesh	Lokesh	Lokesh



K.S.R.M. COLLEGE OF ENGINEER (UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

Week 5 Attendance for Certification Course on Python for Data Science

From 25-07-2022 to 29-07-2022



S.NO	Roll Number	NAME OF THE STUDENT	25/7/2022	26/7/2022	27/7/2022	28/7/2022	29/7/2022
1	209Y1A0404	A Uday Kumar Reddy	Uday	Uday	Uday	Uday	Uday
2	209Y1A0406	B.MANASWITHA	Manasa	Manasa	Manasa	Manasa	Manasa
3	209y1a0407	B sai krishna achari	B. Sai Krishna	B. Sai Krishna	B. Sai Krishna	B. Sai Krishna	B. Sai Krishna
4	209y1a0410	Bharath simha reddy	Bharath	Bharath	Bharath	Bharath	Bharath
5	209Y1A0414	B. RAVI KANTH REDDY	BRKR	BRKR	BRKR	BRKR	BRKR
6	209y1a0417	B.sangeetha	Sangeetha	Sangeetha	Sangeetha	Sangeetha	Sangeetha
7	209Y1A0418	Botta.kalyani	Kalyani	Kalyani	Kalyani	Kalyani	Kalyani
8	209Y1A0423	Anuhya. C	Anuya	C. Anuya	C. Anuya	C. Anuya	C. Anuya
9	209Y1A0427	Chavva Rajeshwar reddy	C. Raju	C. Raju	C. Raju	C. Raju	C. Raju
10	209Y1A0429	Chennuru Karthik Reddy	C. Karthik	C. Karthik	C. Karthik	C. Karthik	C. Karthik
11	209Y1A0431	C Rajasekhar naidu	Rajasekh	Rajasekh	Rajasekh	Rajasekh	Rajasekh
12	209Y1A0434	C siva pavani	Pavani	Pavani	Pavani	Pavani	Pavani
13	209y1a0441	D.swetha	Swetha	Swetha	Swetha	Swetha	Swetha
14	209Y1A0448	G. Raghu	Raghu	Raghu	Raghu	Raghu	Raghu
15	209y1a0450	G Nithin kumar reddy	Nithin	Nithin	Nithin	Nithin	Nithin
16	209y1a0453	G.Nagendra prasad	Nagendra	Nagendra	Nagendra	Nagendra	Nagendra
17	209Y1A0458	G Guru Prasad	Guru	Guru	Guru	Guru	Guru
18	209Y1A0462	G. Hari babu	Hari	Hari	Hari	Hari	Hari
19	209Y1A0471	K.Ramya sree	Ramya	Ramya	Ramya	Ramya	Ramya
20	209Y1A0473	K. Deepa Asritha	Deepa	Deepa	Deepa	Deepa	Deepa
21	209Y1A0489	M.Avinash reddy	Avinash	Avinash	Avinash	Avinash	Avinash
22	209y1a04a8	Pandla Chandra Sekhar	Chandra	Chandra	Chandra	Chandra	Chandra
23	209Y1A04B4	Peram Roja Reddy	Roja	Roja	Roja	Roja	Roja
24	209y1a04b7	Ramireddy gari Bharathi	Bharathi	Bharathi	Bharathi	Bharathi	Bharathi
25	209y1a04d5	S siva ruchitha	Ruchitha	Ruchitha	Ruchitha	Ruchitha	Ruchitha
26	209y1a04e8	v harshavardhan	Harsha	Harsha	Harsha	Harsha	Harsha
27	219Y5A0410	C Lokesh Yadav	Lokesh	Lokesh	Lokesh	Lokesh	Lokesh



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - Autonomous)

Kadapa, Andhra Pradesh, India- 516 003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.



KSNR

lives on..



DEPARTMENT OF ECE

CERTIFICATION COURSE on

" PYTHON FOR DATA SCIENCE "



Department of ECE-KSRMCE in association
with IETE & IEEE



29-06-2022 TO
29-07-2022



SJ-IOT LAB

Coordinator

Smt.T.Umamaheswari,
Asst. Prof, Department of ECE

CO-Coordinator

Smt. K.Divya Lakshmi,
Asst. Prof, Department of ECE

Resource persons

Dr. G.HemaLatha,
Professor & HOD, ECE

V. Gopi Tilak,
Asst. Prof, Department of ECE

<https://forms.gle/oPcnkNBpZ62op469>

Dr. G. Hemalatha
(HOD)

Dr. V.S.S. Murthy
(Principal)

Dr. Kandula Chandra Obul Reddy
(Managing Director)

Smt. K.Rajeswari
(Correspondent Secretary, Treasurer)

Sri K. Madan Mohan Reddy
(Vice - Chairman)

Sri K. Raja Mohan Reddy
(Chairman)

f @ t > [ksrmceofficial](#)

www.ksrmce.ac.in

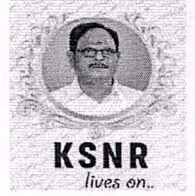
8143731980, 8575697569



K.S.R.M. COLLEGE OF ENGINEERING (UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003
Approved by AICTE, New Delhi & Affiliated to JNTUA,
Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution



Lr./KSRMCE/ECE/ Certificate course on python for Data Science/2022-23/

ACTIVITY REPORT

Certification Course

On

“Python for Data Science”

29/06/2022 to 29/07/2022

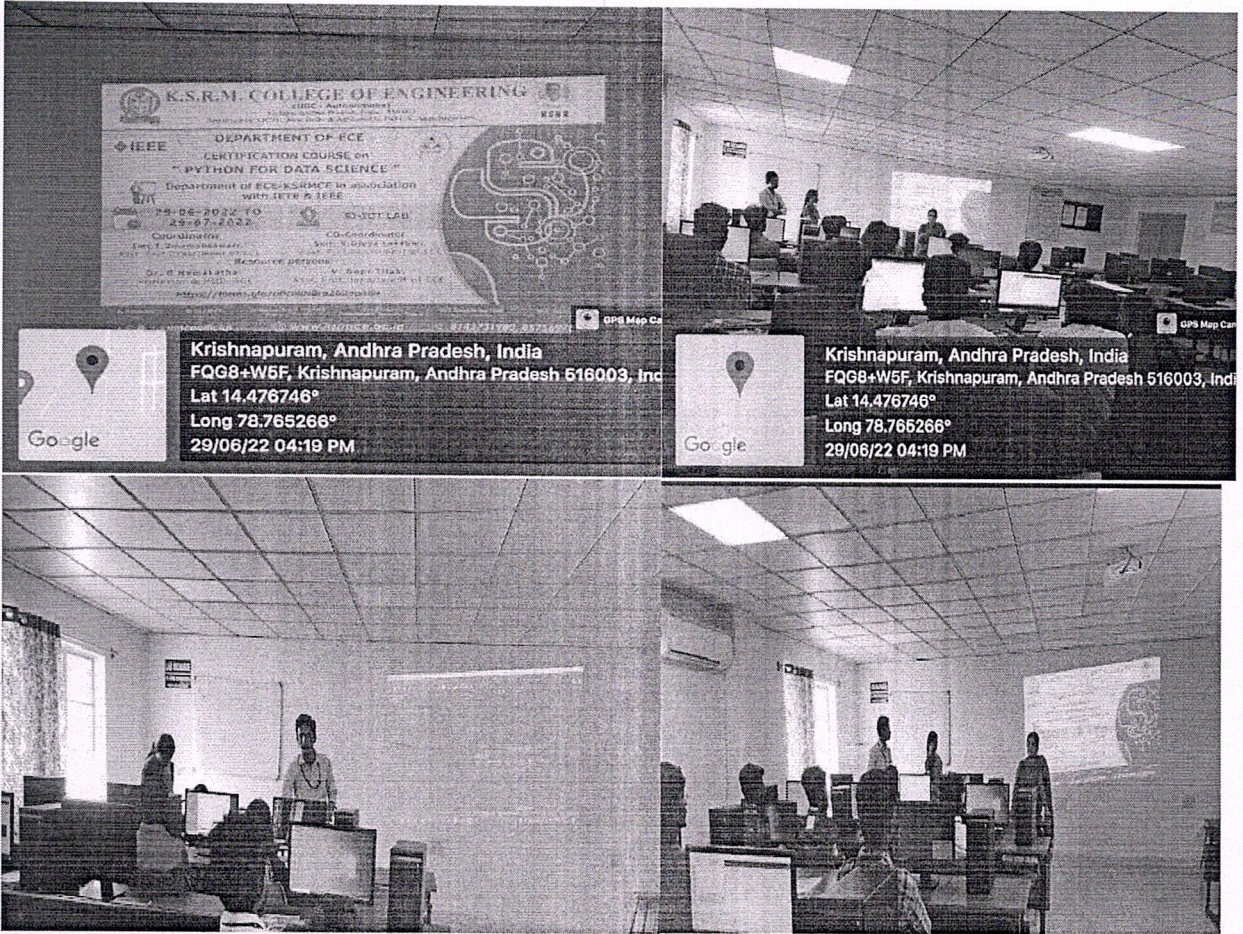
Target Group	:	Students
Details of Participants	:	27 Students
Coordinators	:	Smt.T.Umamaheswari, Asst. Professor in ECE Dept. Smt.K.Divya lakshmi, Asst. Professor in ECE Dept.
Organizing Department	:	Department of Electronics & Communication Engineering
Venue	:	DSP LAB

Description:

Certificate course on **“Python for Data Science”** was organized by Dept. of ECE from **29/06/2022 to 29/07/2022** in DSP Lab. Dr. G.HemaLatha, HOD, Dept. of ECE, Sri V.Gopi Tilak acted as Course instructors. The main aim of the course is to create awareness among students on the fundamental ideas behind Python for data science. Python is a great choice for the backend side of development as well as the software development of devices. Some of the many advantages of working with Python for Data Science are the speed at which you can develop code and a large number of libraries for all kinds of platforms. Thirty Hours course was successfully completed and participation certificates were provided to the participants.

Photo

:



Krishnapuram, Andhra Pradesh, India
FQG8+W5F, Krishnapuram, Andhra Pradesh 516003, India
Lat 14.476746°
Long 78.765266°
29/06/22 04:19 PM



Krishnapuram, Andhra Pradesh, India
FQG8+W5F, Krishnapuram, Andhra Pradesh 516003, India
Lat 14.476746°
Long 78.765266°
29/06/22 04:19 PM

Coordinators

HoD-ECE
Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003

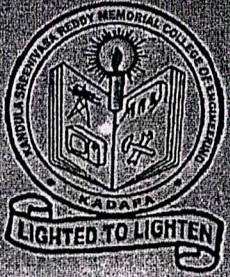


/ksrmce.ac.in

Follow Us:



/ksrmceofficial



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - Autonomous)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.



KSNR
lives on.

Certificate of Participation

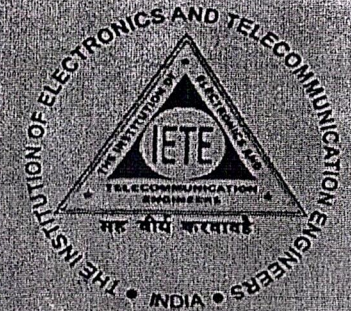
This is to certify that



Mr/Ms B. Manaswitha

with

Roll.No. 209Y1A0406



has attended the Certification course on " Python for Data Science" from 29-06-2022 to 29-07-2022 organized by Dept. of Electronics and Communication Engineering in Association with IETE-ISF and IEEE

Dr. G. Hemalatha
HOD, ECE

Prof V S S Murthy
Principal



K.S.R.M. COLLEGE OF ENGINEERING

(UGC - Autonomous)

Kadapa, Andhra Pradesh, India - 516003

Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.



KSNR
lives on.

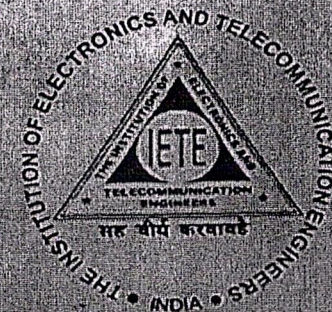
Certificate of Participation

This is to certify that



Mr/Ms Chavva Rajasekhar Reddy with

Roll.No. 209Y1A0427



has attended the Certification course on " Python for Data Science" from 29-06-2022 to 29-07-2022 organized by Dept. of Electronics and Communication Engineering in Association with

IETE-ISF and IEEE

Dr. G. Hemalatha

Prof V S S Murthy

Principal

feedback form of python for DS

* Indicates required question

1. Roll Number *

2. Name of the Student *

3. Organization of Course and session planning by instructor. *

Mark only one oval.

- Excellent
- Very good
- Good
- Fair
- Poor

4. Clarity in content delivery. *

Mark only one oval.

- Excellent
- Very good
- Good
- Fair
- Poor

5. Content is relevant and useful *

Mark only one oval.

Excellent

Very good

Good

Fair

Poor

6. Adequate opportunity to interact with trainer *

Mark only one oval.

Excellent

Very good

Good

Fair

Poor

7. Overall rating

Mark only one oval.

Excellent

Very good

Good

Fair

Poor

This content is neither created nor endorsed by Google.

Google Forms



K.S.R.M. COLLEGE OF ENGINEERING (UGC-AUTONOMOUS)

Kadapa, Andhra Pradesh, India- 516 003



Approved by AICTE, New Delhi & Affiliated to JNTUA, Ananthapuramu.

An ISO 14001:2004 & 9001: 2015 Certified Institution

Department of Electronics & Communication Engineering


Feed back of certification Course on Python for Data Science (29/06/2022 – 29/07/2022)

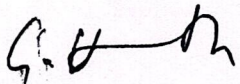
S.no.	Roll Number	Name of the Student	Organization of Course and session planning by instructor.	Clarity in content delivery.	Content is relevant and useful	Adequate opportunity to interact with trainer	Overall rating
1	209Y1A0404	Reddy	Good	Very good	Good	Good	Good
2	209Y1A0406	A	Excellent	Excellent	Excellent	Excellent	Excellent
3	209y1a0407	achari	Excellent	Excellent	Excellent	Excellent	Excellent
4	209y1a0410	reddy	Excellent	Very good	Excellent	Excellent	Excellent
5	209Y1A0414	B.Ravi kantha Reddy	Very good	Fair	Very good	Very good	Very good
6	209y1a0417	B.sangeetha	Good	Good	Good	Good	Good
7	209Y1A0418	Botta.kalyani	Excellent	Excellent	Excellent	Excellent	Excellent
8	209Y1A0423	Anuhya. C	Excellent	Excellent	Excellent	Excellent	Excellent
9	209Y1A0427	C. Rajeshwar reddy	Very good	Excellent	Very good	Excellent	Excellent
10	209Y1A0429	Chennuru Karthik Reddy	Excellent	Very good	Very good	Excellent	Excellent
11	209Y1A0431	C Rajasekhar naidu	Very good	Very good	Very good	Very good	Very good
12	209Y1A0434	C siva pavani	Excellent	Poor	Poor	Poor	Poor
13	209y1a0441	D.swetha	Excellent	Excellent	Excellent	Excellent	Excellent
14	209Y1A0448	G. Raghu	Excellent	Very good	Very good	Very good	Excellent
15	209y1a0450	G Nithin kumar reddy	Excellent	Excellent	Excellent	Fair	Excellent
16	209y1a0453	G.Nagendra prasad	Excellent	Excellent	Excellent	Excellent	Excellent
17	209Y1A0458	G Guru Prasad	Excellent	Excellent	Very good	Very good	Excellent
18	209Y1A0462	G. Hari babu	Excellent	Very good	Excellent	Very good	Excellent
19	209Y1A0471	K.Ramya sree	Very good	Very good	Good	Very good	Excellent
20	209Y1A0473	K. Deepa Asritha	Excellent	Very good	Very good	Excellent	Excellent
21	209Y1A0489	M.Avinash reddy	Good	Excellent	Very good	Very good	Excellent
22	209y1a04a8	Pandla Chandra Sekhar	Good	Good	Very good	Good	Very good
23	209Y1A04B4	Peram Roja Reddy	Very good	Very good	Excellent	Good	Good
24	209y1a04b7	R. Bharathi	Excellent	Excellent	Excellent	Excellent	Excellent
25	209y1a04d5	S siva ruchitha	Very good	Very good	Very good	Very good	Very good
26	209y1a04e8	v harshavardhan	Very good	Very good	Very good	Very good	Very good
27	219Y5A0410	C Lokesh Yadav	Very good	Excellent	Excellent	Excellent	Excellent

Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
VALUE ADDED/CERTIFICATE COURSE ON
PYTHON FOR DATA SCIENCE FROM 29/06/2022 TO 29/07/2022
AWARD LIST

S.No	Roll Number	Name of the Student	Marks Obtained
1.	209Y1A0404	A Uday Kumar Reddy	10
2.	209Y1A0406	B.MANASWITHA	12
3.	209y1a0407	B sai krishna achari	14
4.	209y1a0410	Bharath simha reddy	9
5.	209Y1A0414	B. RAVI KANTH REDDY	10
6.	209y1a0417	B.sangeetha	8
7.	209Y1A0418	Botta.kalyani	8
8.	209Y1A0423	Anuhya. C	8
9.	209Y1A0427	Chavva Rajeshwar reddy	13
10.	209Y1A0429	Chennuru Karthik Reddy	11
11.	209Y1A0431	C Rajasekhar naidu	11
12.	209Y1A0434	C siva pavani	9
13.	209y1a0441	D.swetha	9
14.	209Y1A0448	G. Raghu	7
15.	209y1a0450	G Nithin kumar reddy	8
16.	209y1a0453	G.Nagendra prasad	9
17.	209Y1A0458	G Guru Prasad	10
18.	209Y1A0462	G. Hari babu	12
19.	209Y1A0471	K.Ramya sree	10
20.	209Y1A0473	K. Deepa Asritha	12
21.	209Y1A0489	M.Avinash reddy	12
22.	209y1a04a8	Pandla Chandra Sekhar	13
23.	209Y1A04B4	Peram Roja Reddy	9
24.	209y1a04b7	Ramireddy gari Bharathi	8
25.	209y1a04d5	S siva ruchitha	9
26.	209y1a04e8	v harshavardhan	10
27.	219Y5A0410	C Lokesh Yadav	7


Coordinator


HoD
Professor & H.O.D.
Department of E.C.E.
K.S.R.M. College of Engineering
KADAPA - 516 003

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
PYTHON FOR DATA SCIENCE FROM 29/06/2022 TO 19/07/2022
ASSESSMENT TEST

Roll Number: _____ Name of the Student: _____

Time: 15 Min

(Objective Questions)

Max.Marks: 15

Note: Answer the following Questions and each question carries one mark.

1. Which one of these factors does not affect the speed of execution in python?
a)Dataset Size b)Machine Capabilities c)Editor Used d>Loading Technique []
2. Which of the following is the most important language for Data Science?
a)Ruby b)R c)Java d)None []
3. A collection of information about a related topic is referred to as a _____
a)Visualisation b)Analysis c)Conclusion d)Data []
4. Pandas is an open-source _____ Library
a)Ruby b). Javascript c). Java d). Python []
5. Which function overloads the >> operator?
a)more() b)gt() c)re() d)rshift() []
6. What is the output of the following program :
`i = 0`
`while i < 3:`
`print i`
`print i+1`
a)0 2 1 3 2 4 b)0 1 2 3 4 5 c)0 1 1 2 2 3 d)1 0 2 4 3 5 []
7. What is the maximum length of a Python identifier?
a)32 b)64 c) not specified d)128 []
8. What will be the output of the following code snippet?
`print(2**3 + (5 + 6)**(1 + 1))`
a)129 b)121 c)29 d)none []
9. How is a code block indicated in Python?
a)Brackets b)indentation c)key d)none []
10. What is output of $-33 == 33.0$
a)False b)True c)33 d)None of the above []
11. In the following options which are python libraries which are used for data analysis and scientific computations
A - Numpy B - Scipy C - Pandas D - All the above []
12. Python is a _____ object-oriented programming language.
a)Special purpose b)General purpose c)Medium level programming language d)All of the mentioned above []
13. List, tuple, and range are the _____ of Data Types.
a)Sequence Types b)Binary Types c)Boolean Types d)None []
14. Bytes, bytearray, memoryview are type of the _____ data type.
a)Mapping Type b)Boolean Type c)Binary Types d)None []

K.S.R.M. COLLEGE OF ENGINEERING (AUTONOMOUS), KADAPA-516003
DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
VALUE ADDED /CERTIFICATE COURSE ON
PYTHON FOR DATA SCIENCE FROM 29/06/2022 TO 19/07/2022
ASSESSMENT TEST

9/15

Roll Number: 209Y1A0441 Name of the Student: D. Soetha

Time: 15 Min

(Objective Questions)

Max.Marks: 15

Note: Answer the following Questions and each question carries one mark.

- Which one of these factors does not affect the speed of execution in python?
a)Dataset Size b)Machine Capabilities c)Editor Used d>Loading Technique [dgk]
- Which of the following is the most important language for Data Science?
a)Ruby b)R c)Java d)None [b]
- A collection of information about a related topic is referred to as a _____
a)Visualisation b)Analysis c)Conclusion d)Data [a]
- Pandas is an open-source _____ Library
a)Ruby b). Javascript c). Java d). Python [d]
- Which function overloads the >> operator?
a)more() b)gt() c)re() d)rshift() [d]
- What is the output of the following program :

```
i = 0  
while i < 3:  
    print i  
    print i+1
```


a)0 2 1 3 2 4 b)0 1 2 3 4 5 c)0 1 1 2 2 3 d)1 0 2 4 3 5 [bk]
- What is the maximum length of a Python identifier?
a)32 b)64 c) not specified d)128 [c]
- What will be the output of the following code snippet?

```
print(2**3 + (5 + 6)**(1 + 1))
```


a)129 b)121 c)29 d)none [a]
- How is a code block indicated in Python?
a)Brackets b)indentation c)key d)none [ad]
- What is output of $-33 == 33.0$
a)False b)True c)33 d)None of the above [a]
- In the following options which are python libraries which are used for data analysis and scientific computations
A - Numpy B - Scipy C - Pandas D - All the above [d]
- Python is a _____ object-oriented programming language.
a)Special purpose b)General purpose c)Medium level programming language d)All of the mentioned above [b]
- List, tuple, and range are the _____ of Data Types.
a)Sequence Types b)Binary Types c)Boolean Types d)None [a]
- Bytes, bytearray, memoryview are type of the _____ data type.
a)Mapping Type b)Boolean Type c)Binary Types d)None [b]

Python for Data Science

Tutorial Content

- Overview of Python Libraries for Data Scientists
- Reading Data; Selecting and Filtering the Data; Data manipulation, sorting, grouping, rearranging
- Plotting the data
- Descriptive statistics
- Inferential statistics

2

Python Libraries for Data Science

Many popular Python toolboxes/libraries:

- NumPy
- SciPy
- Pandas
- SciKit-Learn

All these libraries are installed on the SCC

Visualization libraries

- matplotlib
- Seaborn

and many more ...

3

Python Libraries for Data Science

NumPy:

- introduces objects for multidimensional arrays and matrices, as well as functions that allow to easily perform advanced mathematical and statistical operations on those objects
- provides vectorization of mathematical operations on arrays and matrices which significantly improves the performance
- many other python libraries are built on NumPy

Link: <http://www.numpy.org/>

4

Python Libraries for Data Science

SciPy:

- collection of algorithms for linear algebra, differential equations, numerical integration, optimization, statistics and more
- part of SciPy Stack
- built on NumPy

Link: <https://www.scipy.org/scipylib/>

5

Python Libraries for Data Science

Pandas:

- adds data structures and tools designed to work with table-like data (similar to Series and Data Frames in R)
- provides tools for data manipulation: reshaping, merging, sorting, slicing, aggregation etc.
- allows handling missing data

Link: <http://pandas.pydata.org/>

6

Python Libraries for Data Science

SciKit-Learn:

- provides machine learning algorithms: classification, regression, clustering, model validation etc.
- built on NumPy, SciPy and matplotlib

Link: <http://scikit-learn.org/>

7

Python Libraries for Data Science

matplotlib:

- python 2D plotting library which produces publication quality figures in a variety of hardcopy formats
- a set of functionalities similar to those of MATLAB
- line plots, scatter plots, barcharts, histograms, pie charts etc.
- relatively low-level; some effort needed to create advanced visualization

Link: <https://matplotlib.org/>

8

Python Libraries for Data Science

Seaborn:

- based on matplotlib
- provides high level interface for drawing attractive statistical graphics
- Similar (in style) to the popular ggplot2 library in R

Link: <https://seaborn.pydata.org/>

9

Login to the Shared Computing Cluster

- Use your SCC login information if you have SCC account
- If you are using tutorial accounts see info on the blackboard

Note: Your password will not be displayed while you enter it.

10

Selecting Python Version on the SCC

view available python versions on the SCC

```
[scc1 ~] module avail python
```

load python 3 version

```
[scc1 ~] module load python/3.6.2
```

11

Download tutorial notebook

On the Shared Computing Cluster

```
[scc1 ~] cp /project/scv/examples/python/data_analysis/dataScience.ipynb .
```

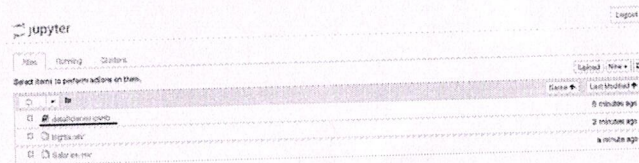
On a local computer save the link:

http://rcs.bu.edu/examples/python/data_analysis/dataScience.ipynb

12

Start Jupyter notebook

On the Shared Computing Cluster
[scc1 ~] jupyter notebook



13

Loading Python Libraries

```
In [ ]: # Import Python Libraries
import numpy as np
import scipy as sp
import pandas as pd
import matplotlib as mpl
import seaborn as sns
```

Press Shift+Enter to execute the *jupyter* cell

14

Reading data using pandas

```
In [ ]: #Read csv file
df = pd.read_csv("http://rcs.bu.edu/examples/python/data_analysis/Salaries.csv")
```

Note: The above command has many optional arguments to fine-tune the data import process.

There is a number of pandas commands to read other data formats:

```
pd.read_excel('myfile.xlsx', sheet_name='Sheet1', index_col=None, na_values=['NA'])
pd.read_stata('myfile.dta')
pd.read_sas('myfile.sas7bdat')
pd.read_hdf('myfile.h5', 'df')
```

15

Exploring data frames

```
In [3]: #List first 5 records
df.head()
```

```
Out [3]:
```

	rank	discipline	phd	service	sex	salary
0	Prof	B	54	45	Male	164662
1	Prof	A	12	6	Male	93023
2	Prof	A	25	25	Male	112514
3	Prof	A	42	31	Male	132295
4	Prof	B	22	18	Male	124802

16

✍ Hands-on exercises

✓ Try to read the first 10, 20, 50 records;

✓ Can you guess how to view the last few records;

Hint: 

17

Data Frame data types

Pandas Type	Native Python Type	Description
object	string	The most general dtype. Will be assigned to your column if column has mixed types (numbers and strings).
int64	int	Numeric characters. 64 refers to the memory allocated to hold this character.
float64	float	Numeric characters with decimals. If a column contains numbers and NaNs (see below), pandas will default to float64, in case your missing value has a decimal.
datetime64, timedelta[ns]	N/A (but see the <code>datetime</code> module in Python's standard library)	Values meant to hold time data. Look into these for time series experiments.

18

Data Frame data types

```
In [4]: #Check a particular column type
df['salary'].dtype
```

```
Out[4]: dtype('int64')
```

```
In [5]: #Check types for all the columns
df.dtypes
```

```
Out[4]: rank      object
discipline  object
phd          int64
service     int64
sex         object
salary     int64
dtype: object
```

19

Data Frames attributes

Python objects have *attributes* and *methods*.

df.attribute	description
dtypes	list the types of the columns
columns	list the column names
axes	list the row labels and column names
ndim	number of dimensions
size	number of elements
shape	return a tuple representing the dimensionality
values	numpy representation of the data

20

✍ Hands-on exercises

- ✓ Find how many records this data frame has;
- ✓ How many elements are there?
- ✓ What are the column names?
- ✓ What types of columns we have in this data frame?

21

Data Frames methods

Unlike attributes, python methods have *parenthesis*.
All attributes and methods can be listed with a *dir()* function: `dir(df)`

df.method()	description
head([n]), tail([n])	first/last n rows
describe()	generate descriptive statistics (for numeric columns only)
max(), min()	return max/min values for all numeric columns
mean(), median()	return mean/median values for all numeric columns
std()	standard deviation
sample([n])	returns a random sample of the data frame
dropna()	drop all the records with missing values

22

✍ Hands-on exercises

- ✓ Give the summary for the numeric columns in the dataset
- ✓ Calculate standard deviation for all numeric columns;
- ✓ What are the mean values of the first 50 records in the dataset? *Hint:* use `head()` method to subset the first 50 records and then calculate the mean

23

Selecting a column in a Data Frame

Method 1: Subset the data frame using column name:
`df['sex']`

Method 2: Use the column name as an attribute:
`df.sex`

Note: there is an attribute `rank` for pandas data frames, so to select a column with a name "rank" we should use method 1.

24

✍ Hands-on exercises

- ✓ Calculate the basic statistics for the *salary* column;
- ✓ Find how many values in the *salary* column (use *count* method);
- ✓ Calculate the average salary;

25

Data Frames *groupby* method

Using "group by" method we can:

- Split the data into groups based on some criteria
- Calculate statistics (or apply a function) to each group
- Similar to `dplyr()` function in R

```
In [ ]: #Group data using rank
df_rank = df.groupby(['rank'])

In [ ]: #Calculate mean value for each numeric column per each group
df_rank.mean()
```

	phd	service	salary
rank			
AssocProf	15.079929	11.327462	91786.220769
AssistProf	5.052032	2.216420	81392.789474
Prof	27.056217	21.419043	125024.614348

26

Data Frames *groupby* method

Once *groupby* object is create we can calculate various statistics for each group:

```
In [ ]: #Calculate mean salary for each professor rank:
df.groupby('rank')[['salary']].mean()
```

	salary
rank	
AssocProf	91786.220769
AssistProf	81392.789474
Prof	125024.614348

Note: if single brackets are used to specify the column (e.g. salary), then the output is Pandas Series object. When double brackets are used the output is a Data Frame

27

Data Frames *groupby* method

groupby performance notes:

- no grouping/splitting occurs until it's needed. Creating the *groupby* object only verifies that you have passed a valid mapping
- by default the group keys are sorted during the *groupby* operation. You may want to pass `sort=False` for potential speedup:

```
In [ ]: #Calculate mean salary for each professor rank:
df.groupby(['rank'], sort=False)[['salary']].mean()
```

28

Data Frame: filtering

To subset the data we can apply Boolean indexing. This indexing is commonly known as a filter. For example if we want to subset the rows in which the salary value is greater than \$120K:

```
In [ ]: #Calculate mean salary for each professor rank:
df_sub = df[ df['salary'] > 120000 ]
```

Any Boolean operator can be used to subset the data:

> greater; >= greater or equal;
 < less; <= less or equal;
 == equal; != not equal;

```
In [ ]: #Select only those rows that contain female professors:
df_f = df[ df['sex'] == 'Female' ]
```

29

Data Frames: Slicing

There are a number of ways to subset the Data Frame:

- one or more columns
- one or more rows
- a subset of rows and columns

Rows and columns can be selected by their position or label

30

Data Frames: Slicing

When selecting one column, it is possible to use single set of brackets, but the resulting object will be a Series (not a DataFrame):

```
In [ ]: #Select column salary:
df['salary']
```

When we need to select more than one column and/or make the output to be a DataFrame, we should use double brackets:

```
In [ ]: #Select column salary:
df[['rank', 'salary']]
```

31

Data Frames: Selecting rows

If we need to select a range of rows, we can specify the range using ":"

```
In [ ]: #Select rows by their position:
df[10:20]
```

Notice that the first row has a position 0, and the last value in the range is omitted:
 So for 0:10 range the first 10 rows are returned with the positions starting with 0 and ending with 9

32

Data Frames: method loc

If we need to select a range of rows, using their labels we can use method loc:

```
In [ ]: #Select rows by their labels:
df_sub.loc[10:20, ['rank', 'sex', 'salary']]
```

```
Out [ ]:
```

	rank	sex	salary
10	Prof	Male	128255
11	Prof	Male	134778
13	Prof	Male	182200
14	Prof	Male	153750
15	Prof	Male	159480
19	Prof	Male	150500

33

Data Frames: method iloc

If we need to select a range of rows and/or columns, using their positions we can use method iloc:

```
In [ ]: #Select rows by their labels:
df_sub.iloc[10:20, [0, 3, 4, 5]]
```

```
Out [ ]:
```

	rank	service	sex	salary
26	Prof	19	Male	148750
27	Prof	83	Male	155850
28	Prof	20	Male	129893
31	Prof	21	Male	155750
25	Prof	23	Male	128933
28	Prof	45	Male	146850
29	Prof	19	Female	129050
48	Prof	28	Female	127000
44	Prof	19	Female	151758
45	Prof	25	Female	140500

34

Data Frames: method iloc (summary)

```
df.iloc[0] # First row of a data frame
df.iloc[i] # (i+1)th row
df.iloc[-1] # Last row
```

```
df.iloc[:, 0] # First column
df.iloc[:, -1] # Last column
```

```
df.iloc[0:7] # First 7 rows
df.iloc[:, 0:2] # First 2 columns
df.iloc[1:3, 0:2] # Second through third rows and first 2 columns
df.iloc[[0,5], [1,3]] # 1st and 6th rows and 2nd and 4th columns
```

35

Data Frames: Sorting

We can sort the data by a value in the column. By default the sorting will occur in ascending order and a new data frame is return.

```
In [ ]: # Create a new data frame from the original sorted by the column Salary
df_sorted = df.sort_values( by='salary')
df_sorted.head()
```

```
Out [ ]:
```

	rank	discipline	phd	service	sex	salary
55	AsstProf	A	2	0	Female	72500
23	AsstProf	A	2	0	Male	89500
42	AsstProf	0	5	0	Female	77500
17	AsstProf	0	4	0	Male	92500
12	AsstProf	0	1	0	Male	89500

36

Data Frames: Sorting

We can sort the data using 2 or more columns:

```
In [ ]: df_sorted = df.sort_values( by=['service', 'salary', ascending = [True, False]]
df_sorted.head(10)
```

```
Out [ ]:
```

	rank	discipline	phd	service	sex	salary
53	Prof	A	12	0	Female	155000
17	AssProf	B	4	0	Male	92000
15	AssProf	B	1	0	Male	84000
23	AssProf	A	2	0	Male	83000
43	AssProf	B	5	0	Female	75000
55	AssProf	A	2	0	Female	72500
37	AssProf	A	3	1	Female	72000
28	AssProf	B	7	2	Male	61000
42	AssProf	B	4	2	Female	60200
66	AssProf	A	4	0	Female	55500

37

Missing Values

Missing values are marked as NaN

```
In [ ]: # Read a dataset with missing values
flights = pd.read_csv("http://rcs.bu.edu/examples/python/data_analysis/flights.csv")
```

```
In [ ]: # Select the rows that have at least one missing value
flights[flights.isnull().any(axis=1)].head()
```

```
Out [ ]:
```

	year	month	day	dep_time	dep_delay	arr_time	arr_delay	carrier	tailnum	flight	origin	dest	air_line	distance	hour	minute
330	2013	1	1	1847.0	20.0	2251.0	NaN	UA	N21412	1228	EWR	DAL	NaN	2425	18.0	7.0
493	2013	1	1	NaN	NaN	NaN	NaN	AA	N287HA	791	LGA	DFW	NaN	1369	NaN	NaN
484	2013	1	1	NaN	NaN	NaN	NaN	AA	N287VA	1525	LGA	MIA	NaN	1556	NaN	NaN
455	2013	1	2	2145.0	10.0	NaN	NaN	UA	N2223	1599	EWR	RSW	NaN	1389	21.0	43.0
826	2013	1	2	NaN	NaN	NaN	NaN	AA	NaN	329	JFK	LAX	NaN	2476	NaN	NaN

38

Missing Values

There are a number of methods to deal with missing values in the data frame:

df.method()	description
dropna()	Drop missing observations
dropna(how='all')	Drop observations where all cells is NA
dropna(axis=1, how='all')	Drop column if all the values are missing
dropna(thresh = 5)	Drop rows that contain less than 5 non-missing values
fillna(0)	Replace missing values with zeros
isnull()	returns True if the value is missing
notnull()	Returns True for non-missing values

39

Missing Values

- When summing the data, missing values will be treated as zero
- If all values are missing, the sum will be equal to NaN
- cumsum() and cumprod() methods ignore missing values but preserve them in the resulting arrays
- Missing values in GroupBy method are excluded (just like in R)
- Many descriptive statistics methods have *skipna* option to control if missing data should be excluded. This value is set to *True* by default (unlike R)

40

Aggregation Functions in Pandas

Aggregation - computing a summary statistic about each group, i.e.

- compute group sums or means
- compute group sizes/counts

Common aggregation functions:

min, max
count, sum, prod
mean, median, mode, mad
std, var

41

Aggregation Functions in Pandas

agg() method are useful when multiple statistics are computed per column:

```
In [ ]: flights[['dep_delay', 'arr_delay']].agg(['min', 'mean', 'max'])
```

```
Out [ ]:
      dep_delay  arr_delay
min  -16.000000  -62.000000
mean   9.384332   2.298675
max   351.000000  389.000000
```

42

Basic Descriptive Statistics

df.method()	description
describe	Basic statistics (count, mean, std, min, quantiles, max)
min, max	Minimum and maximum values
mean, median, mode	Arithmetic average, median and mode
var, std	Variance and standard deviation
sem	Standard error of mean
skew	Sample skewness
kurt	kurtosis

43

Graphics to explore the data

Seaborn package is built on matplotlib but provides high level interface for drawing attractive statistical graphics, similar to ggplot2 library in R. It specifically targets statistical data visualization

To show graphs within Python notebook include inline directive:

```
In [ ]: %matplotlib inline
```

44

Graphics

	description
distplot	histogram
barplot	estimate of central tendency for a numeric variable
violinplot	similar to boxplot, also shows the probability density of the data
jointplot	Scatterplot
regplot	Regression plot
pairplot	Pairplot
boxplot	boxplot
swarmplot	categorical scatterplot
factorplot	General categorical plot

45

Basic statistical Analysis

statsmodel and scikit-learn - both have a number of function for statistical analysis

The first one is mostly used for regular analysis using R style formulas, while scikit-learn is more tailored for Machine Learning.

statsmodels:

- linear regressions
- ANOVA tests
- hypothesis testings
- many more ...

scikit-learn:

- kmeans
- support vector machines
- random forests
- many more ...

See examples in the Tutorial Notebook

46

Conclusion

Thank you for attending the tutorial.

Please fill the evaluation form:

http://scv.bu.edu/survey/tutorial_evaluation.html

Questions:

email: koleinik@bu.edu (Katia Oleinik)

47