



Program Studi Sains Data
Institut Teknologi Sumatera

Studi Kasus Python Courses on coursera

Kelompok 7 - Basis Data RA



Dosen Pembimbing:

Riksa Meidy Karim, S.Kom., M.Si., M.Sc.

Amalya Citra Pradana, S.Kom., M.Si., M.Sc.

Tirta Setiawan, S.Pd., M.Si.

Anggota Kelompok:

Dimas Wahyu Saputro (120450081)

Armanda Tantra Kumara (120450037)

Taj Shavira H. (120450059)

Fadia Dilla Sabine (120450011)

Khoirunnisa (120450029)



★ Table of contents



01

Pendahuluan

Latar belakang, tujuan, asal data, deskripsi data.

02

Normalisasi

Mengurangi redudansi data dan meningkatkan integritas data.

03

DDL - DML

Membuat struktur, dan mengisi data.

04

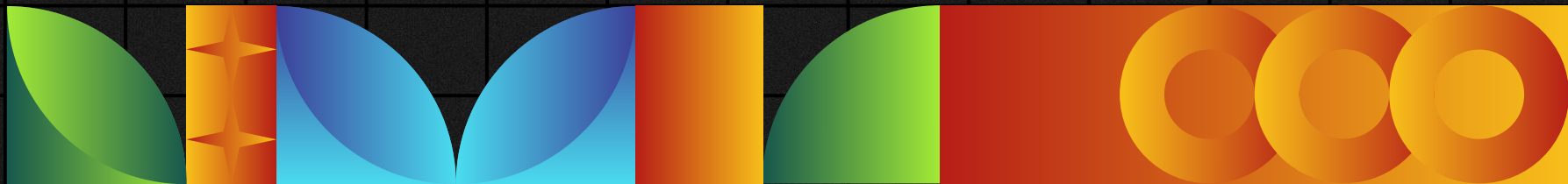
Insight

Apa yang bisa kita dapatkan?

01.

Pendahuluan

Latar belakang, tujuan, asal data, deskripsi data.





Python
ramai
digunakan



Banyak yang
menyediakan



Coursera

Kaggle

Membedah
dan Ambil
Insight

Python Courses on Coursera

This dataset includes scrapped data from coursera with a term 'Python'

coursera
Python Courses EDA

Data Code (1) Discussion (0) Metadata

About Dataset

Context

This is my first trial on selenium scrapping. While I was learning web scrapping, I got a problem fetching the dynamic content from the site. It's where I came to know about selenium.

Content

When you search the term 'Python' coursera brings you all courses involving that term from their database. Now, while I searched that (in June 2021) there were about ~950 courses in total spread across multiple pages. So I crawled in all those pages and got the relevant information.

Acknowledgements

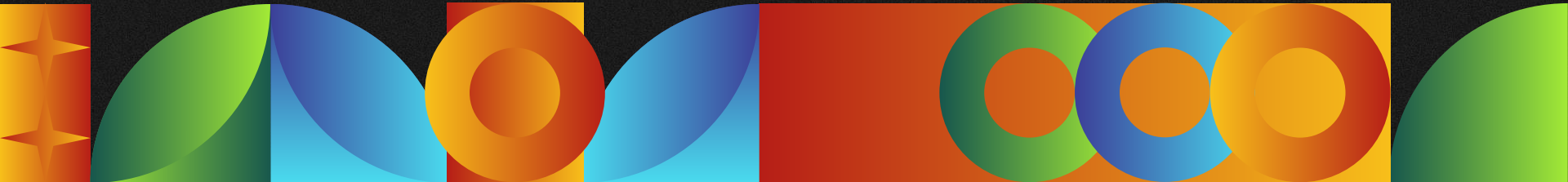
I have provided my approach to find some insights and perspectives on this data (as a notebook). I would love to see yours.
I have made a site in flask you can view here: [GitHub link](#)

Usability ⓘ
10.00

License
Other (specified in description)

Expected update frequency
Never

02. Normalisasi dan Relational Schema



✦ Kenapa harus Normalisasi?



**Anomali
penambahan**



**Anomali
penghapusan**



**Anomali
Modifikasi**



✦ Tahap Normalisasi

UNF



1NF



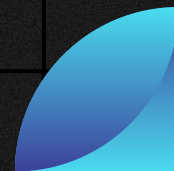
2NF



★ Data Awal

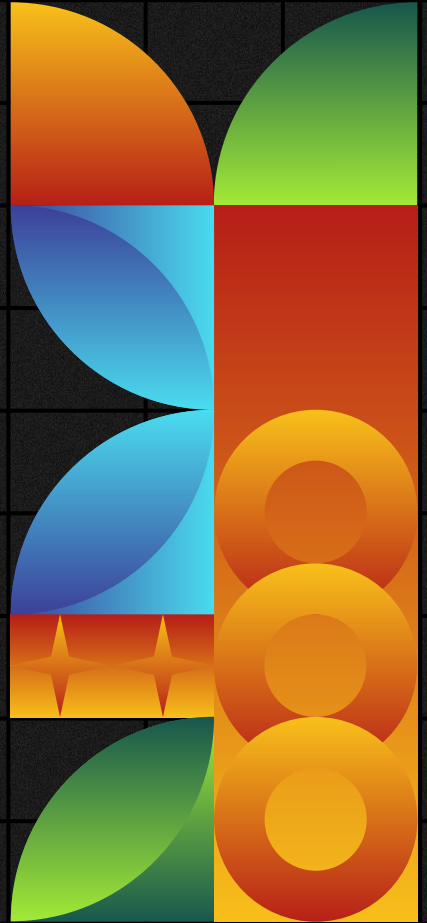
	university	course	type	review	votes	students	difficulty
0	University of Michigan	Python for Everybody	SPECIALIZATION	4.8	225236	240000	Beginner
1	Google	Google IT Automation with Python	PROFESSIONAL CERTIFICATE	4.7	21768	410000	Beginner
2	University of Michigan	Python 3 Programming	SPECIALIZATION	4.7	15904	290000	Beginner
3	University of Michigan	Applied Data Science with Python	SPECIALIZATION	4.5	29358	730000	Intermediate
4	Università di Napoli Federico II	Data Science con Python e R	SPECIALIZATION				Intermediate
5	Google	Crash Course on Python	COURSE	4.8	17595	350000	Beginner
6	IBM	IBM Data Science	PROFESSIONAL CERTIFICATE	4.6	74747	830000	Beginner
7	University of Michigan	Programming for Everybody (Getting Started with Python)	COURSE	4.8	189604	2200000	Mixed
8	IBM	Data Science Fundamentals with Python and SQL	SPECIALIZATION	4.6	37454	450000	Beginner
9	IBM	IBM Data Analyst	PROFESSIONAL CERTIFICATE	4.6	36275	510000	Beginner
10	IBM	Python for Data Science, AI & Development	COURSE	4.6	21662	300000	Beginner
11	University of Pennsylvania	Introduction to Programming with Python and Java	SPECIALIZATION	4	179	11000	Beginner
12	Moscow Institute of Physics and Technology	Python for Data Science, AI & Development	SPECIALIZATION	4.5	1695	71000	Intermediate
13	University of Michigan	Statistics with Python	SPECIALIZATION	4.6	2359	92000	Beginner
14	University of California, Irvine	IELTS Preparation	SPECIALIZATION	4.5	50	4000	Beginner
15	Rice University	Introduction to Scripting in Python	SPECIALIZATION	4.7	3460	100000	Beginner
16	University of California San Diego	Data Structures and Algorithms	SPECIALIZATION	4.6	13335	490000	Intermediate
17	University of Colorado Boulder	Data Structures and Algorithms	SPECIALIZATION	4.4	8	1600	Advanced
18	Pontificia Universidad Católica de Chile	Introducción a la programación en Python I: Aprendiendo	COURSE	4.4	2768	170000	Beginner
19	Codio	Programming in Python: A Hands-on Introduction	SPECIALIZATION	4.7	23	1600	Beginner
20	University of Michigan	Python Basics	COURSE	4.8	12882	250000	Beginner

—Apakah ini UNF, 1NF, 2NF?



1NF

Suatu tabel dikatakan 1NF jika dan hanya jika setiap atribut dari data tersebut hanya memiliki nilai tunggal dalam satu baris.



2NF

- Tabel dalam keadaan 1 NF.
- Setiap atribut bukan primary key memiliki ketergantungan fungsional penuh pada primary key.

DetailCourse							
id	course	university_id	type_id	difficulty_id	review	votes	students
1	Python for Everybody	U1	T1	D1	4.8	225236	2400000
2	Google IT Automation with Python	U2	T2	D1	4.7	21768	410000
3	Python 3 Programming	U1	T1	D1	4.7	15904	290000
4	Applied Data Science with Python e R	U1	T1	D2	4.5	29358	730000
5	Data Science con Python e R	U3	T1	D2			
6	Crash Course on Python	U2	T3	D1	4.8	17595	350000

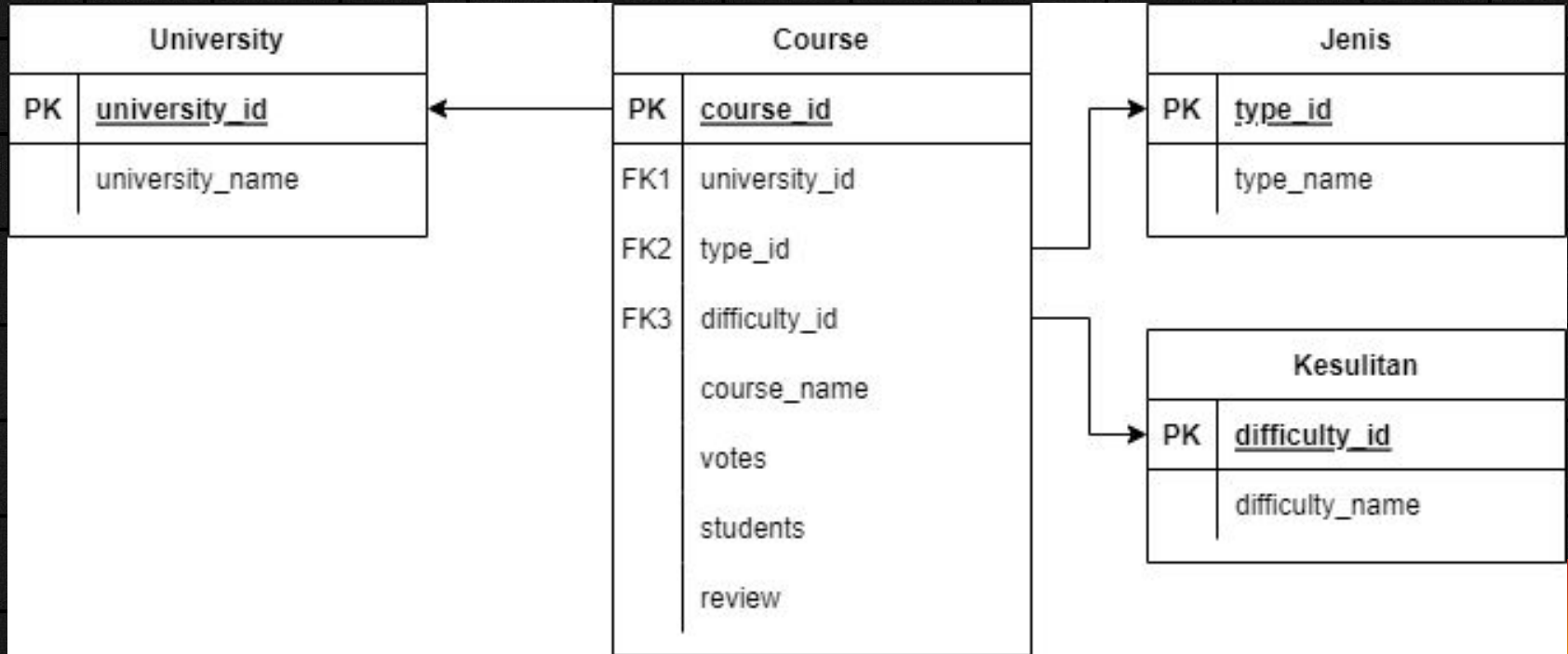
University	
university_id	University
U1	University of Michigan
U2	Google
U3	Università di Napoli Federico II
U4	IBM

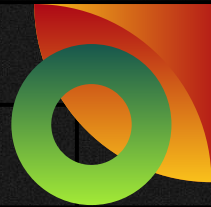
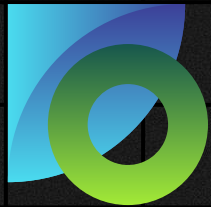
DifficultyCourse	
difficulty_id	Difficulty Name
D1	Beginner
D2	Intermediate
D3	Advanced
D4	Mixed

Type Course	
type_id	Type Name
T1	SPECIALIZATION
T2	PROFESSIONAL CERTIFICATE
T3	COURSE
T4	GUIDED PROJECT
T5	DEGREE
T6	MASTER TRACK CERTIFICATE



Relational Schema





03.

DDL - DML

Membuat Struktur dan Mengisi Data.



★ DDL (Data Definition Language)



Create



Rename




Alter



Drop

```
4  # =====
5  # TABEL UNIVERSITY
6  • CREATE TABLE IF NOT EXISTS university(
7      university_id VARCHAR(5) NOT NULL PRIMARY KEY,
8      university_name VARCHAR(255)
9  );
10 • DESCRIBE university;
```

Result Grid						
Filter Rows:						
Export:  Wrap Ce						
	Field	Type	Null	Key	Default	Extra
▶	university_id	varchar(5)	NO	PRI	NULL	
	university_name	varchar(255)	YES		NULL	

★ DML (Data Manipulation Language)



Insert



Select



Update



Delete

```
61  # =====
62  # DML ke tabel jenis
63  • INSERT INTO jenis ( type_id, type_name)
64    VALUES ( "T1", "SPECIALIZATION"),
65            ( "T2", "PROFESSIONAL CERTIFICATE"),
66            ( "T3", "COURSE"),
67            ( "T4", "GUIDED PROJECT"),
68            ( "T5", "DEGREE"),
69            ( "T6", "MASTER TRACK CERTIFICATE")
70  ;
71  • SELECT * FROM jenis;
```

	type_id	type_name
▶	T1	SPECIALIZATION
	T2	PROFESSIONAL CERTIFICATE
	T3	COURSE
	T4	GUIDED PROJECT
	T5	DEGREE
	T6	MASTER TRACK CERTIFICATE
•	NULL	NULL

★ DML (Data Manipulation Language)



Insert



Select



Update



Delete

university_id	university_name
U1	University of Michigan
U10	Codio
U11	The Hong Kong University of Science and Techn...
U12	DeepLearning.AI
U13	Coursera Project Network
U14	Infosec
U15	HSE University
U16	Wesleyan University
U17	Universidad Austral
U18	EDHEC Business School
U19	Universidad de los Andes
U2	Google
U20	University of Toronto
U21	Google Cloud
U22	Amazon Web Services
U23	University of London
U24	University of Alberta
U25	University of Queensland

```
145 # implementasi DML - Update
146 • UPDATE university
147 SET university_name = "Institut Teknologi Sumatera"
148 WHERE university_id = "U1";
```

university_id	university_name
U1	Institut Teknologi Sumatera
U10	Codio
U11	The Hong Kong University of Science and Techn...
U12	DeepLearning.AI
U13	Coursera Project Network
U14	Infosec
U15	HSE University
U16	Wesleyan University
U17	Universidad Austral
U18	EDHEC Business School
U19	Universidad de los Andes
U2	Google
U20	University of Toronto
U21	Google Cloud
U22	Amazon Web Services

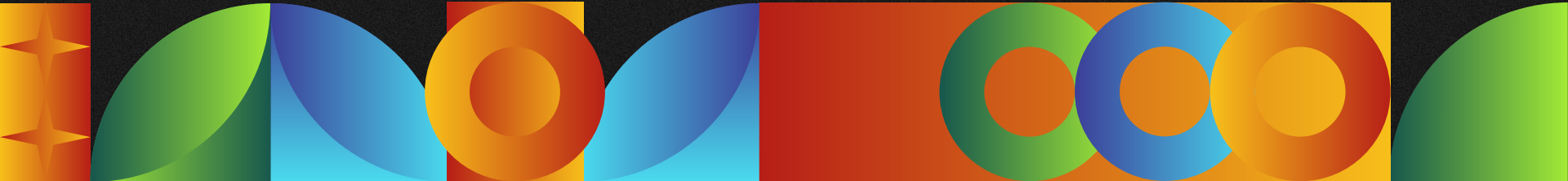


Insight!!

1. Mendapatkan 5 course dengan rating terendah.

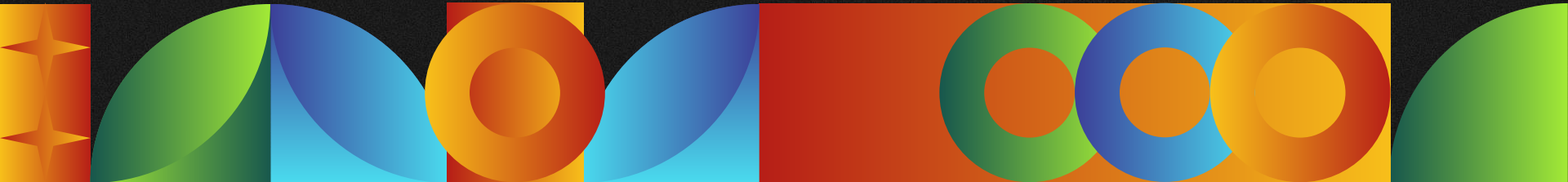
```
3 # Mendapatkan 5 Course dengan Rating terendah
4 • SELECT course.course_name as NamaCourses,
5     university.university_name as Universitas,
6     course.review AS Rating
7 FROM course
8 INNER JOIN university ON course.university_id=university.university_id
9 ORDER BY course.review ASC
10 LIMIT 5;
```

Result Grid			
Filter Rows: <input type="text"/>			
Export:  Wrap Cell Content: <input type="checkbox"/> Fetch rows:			
	NamaCourses	Universitas	Rating
▶	Introduction to Python Programming	University of Pennsylvania	3.9
	Introduction to Programming with Python and J...	University of Pennsylvania	4
	Advanced Data Science with IBM	Moscow Institute of Physics and Technology	4.3
	Python Project for Data Science	IBM	4.3
	Data Structures and Algorithms	University of Colorado Boulder	4.4



2. Mendapatkan 5 course dengan rating tertinggi.

	NamaCourses	Universitas	Rating
▶	Python Data Structures	Institut Teknologi Sumatera	4.9
	Introdução à Ciência da Computação com Pytho...	Institut Teknologi Sumatera	4.9
	Programación en Python	Universidad de los Andes	4.9
	Python for Everybody	Institut Teknologi Sumatera	4.8
	Deep Learning	DeepLearning.AI	4.8





3. Mendapatkan **rata-rata rating** dari setiap pembuat course

Universitas	JumlahCourse	RatingAVG
► Universidad de los Andes	1	4.900000095367432
Institut Teknologi Sumatera	13	4.730769267449012
DeepLearning.AI	1	4.800000190734863
Google	3	4.733333269755046
Rice University	2	4.75
Codio	1	4.699999809265137
Wesleyan University	1	4.699999809265137
University of Toronto	1	4.699999809265137
University of London	1	4.699999809265137
IBM	5	4.579999923706055
University of Pennsylvania	5	4.339999961853027
Moscow Institute of Physic...	4	4.524999976158142
University of California Sa...	1	4.599999904632568
Infosec	1	4.599999904632568
EDHEC Business School	1	4.599999904632568
University of California, Ir...	1	4.5
Coursera Project Network	4	4.549999952316284
Universidad Austral	1	4.5
Google Cloud	1	4.5
Amazon Web Services	1	4.5
University of Colorado Bou...	1	4.400000095367432
The Hong Kong University ...	1	4.400000095367432



Periksa ITERA!

Bandingkan dengan nomor 1



★ Mendapatkan Universitas dengan Jumlah Course Lebih dari Tiga

	Universitas	JumlahCourse
▶	Institut Teknologi Sumatera	13
	Coursera Project Network	4
	IBM	5
	University of Pennsylvania	5
	Moscow Institute of Physics and Technology	4

5/30
Penyedia

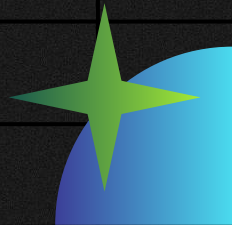
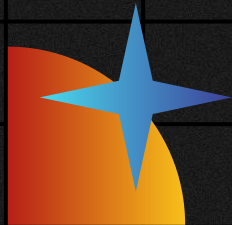


★ Join Seluruh Tabel

	course_id	course_name	university_name	type_name	difficulty_name	votes	students	review
▶	1	Python for Everybody	Institut Teknologi Sumatera	SPECIALIZATION	Beginner	225236	2400000	4.8
	2	Google IT Automation with Python	Google	PROFESSIONAL CERTIFICATE	Beginner	21768	410000	4.7
	3	Python 3 Programming	Institut Teknologi Sumatera	SPECIALIZATION	Beginner	15904	290000	4.7
	5	Crash Course on Python	Google	COURSE	Beginner	17595	350000	4.8
	6	IBM Data Science	IBM	PROFESSIONAL CERTIFICATE	Beginner	74747	830000	4.6
	8	Data Science Fundamentals with Python and SQL	IBM	SPECIALIZATION	Beginner	37454	450000	4.6
	9	IBM Data Analyst	University of Pennsylvania	PROFESSIONAL CERTIFICATE	Beginner	36275	510000	4.6
	10	Python for Data Science, AI & Development	Moscow Institute of Physics and Technology	COURSE	Beginner	21662	300000	4.6
	11	Introduction to Programming with Python and J...	University of Pennsylvania	SPECIALIZATION	Beginner	179	11000	4
	12	Statistics with Python	Institut Teknologi Sumatera	SPECIALIZATION	Beginner	2359	92000	4.6
	13	IELTS Preparation	University of California, Irvine	SPECIALIZATION	Beginner	50	4000	4.5
	14	Introduction to Scripting in Python	Rice University	SPECIALIZATION	Beginner	3460	100000	4.7
	17	Programming in Python: A Hands-on Introduction	Codio	SPECIALIZATION	Beginner	23	1600	4.7
	18	Python Basics	Institut Teknologi Sumatera	COURSE	Beginner	12882	250000	4.8
	21	Introduction to Python	Coursera Project Network	GUIDED PROJECT	Beginner	1634	26000	4.5
	25	IBM Full Stack Cloud Developer	Moscow Institute of Physics and Technology	PROFESSIONAL CERTIFICATE	Beginner	22680	340000	4.6
	26	Python Programming: A Concise Introduction	Wesleyan University	COURSE	Beginner	4265	160000	4.7
	27	Aprende a programar con Python	Universidad Austral	SPECIALIZATION	Beginner	1714	43000	4.5
	29	Introduction to Python Programming	University of Pennsylvania	COURSE	Beginner	129	8100	3.9
	30	Fundamentals of Computing	Rice University	SPECIALIZATION	Beginner	3500	240000	4.8



Yeayyy!





Thanks

Do you have any questions?

youremail@freepik.com

+91 620 421 838

yourcompany.com



CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, and infographics & images by **Freepik**

Please keep this slide for attribution

