



COURSERA DATA ANALYSIS WITH PYTHON



**PRESENTED BY:
MUHAMMAD ABBAS**

Coursera Data Analysis With Python

August 16, 2024

1 Import Libraries

```
[78]: import pandas as pd
import numpy as np
import matplotlib.pyplot as mt
import seaborn as sc
from wordcloud import WordCloud as wc
```

2 Import Dataset

```
[79]: ds=pd.read_csv("datasetuc.csv")
```

```
[80]: ds
```

```
[80]:
```

	Course Title	Rating	\
0	Fashion as Design	4.8	
1	Modern American Poetry	4.4	
2	Pixel Art for Video Games	4.5	
3	Distribución digital de la música independiente	NaN	
4	The Blues: Understanding and Performing an Ame...	4.8	
...	
9590	Power Onboarding	4.1	
9591	Implementing Parent Feedback with Google Forms	4.9	
9592	"Making" Progress Teach-Out	4.2	
9593	Use Microsoft PowerPoint to Create a Visual Sc...	4.5	
9594	Créer du contenu pour les réseaux sociaux avec...	NaN	

	Level	Duration	Schedule	\
0	Beginner level	20 hours (approximately)	Flexible schedule	
1	Beginner level	Approx. 34 hours to complete	Flexible schedule	
2	Beginner level	9 hours (approximately)	Flexible schedule	
3	Beginner level	Approx. 8 hours to complete	Flexible schedule	
4	Beginner level	Approx. 11 hours to complete	Flexible schedule	
...	
9590	Beginner level	Approx. 7 hours to complete	Flexible schedule	
9591	Beginner level	NaN	NaN	
9592	Beginner level	Approx. 5 hours to complete	Flexible schedule	

9593	Beginner level	1 hour	NaN
9594	Beginner level	NaN	NaN

	Review	What you will learn \
0	2,813 reviews	NaN
1	100 reviews	NaN
2	227 reviews	NaN
3	NaN	NaN
4	582 reviews	Students will be able to describe the blues as...
...
9590	36 reviews	NaN
9591	77 reviews	Creating engaging parent surveys using Google ...
9592	10 reviews	NaN
9593	11 reviews	Learn about choice boards and their use within...
9594	NaN	Vous découvrirez les fonctionnalités de Prezi ...

	Skill gain \
0	['Art History', 'Art', 'History', 'Creativity']
1	[]
2	[]
3	[]
4	['Music', 'Chord', 'Jazz', 'Jazz Improvisation']
...	...
9590	[]
9591	['Education', 'online tools', 'engagement', 'e...]
9592	[]
9593	['Teaching tools', 'elementary teaching', 'eng...]
9594	['Social Media Marketing', 'Social Media', 'Co...]

	Modules \
0	['Introduction', 'Heroes', 'Silhouettes', 'Cou...]
1	['Orientation', 'Module 1', 'Module 2', 'Modul...]
2	['Week 1: Introduction to Pixel Art', 'Week 2:...
3	['Semana 1', 'Semana 2', 'Semana 3', 'Semana 4']
4	['Blues Progressions - Theory and Practice ', ...]
...	...
9590	['Week 1 - Get Ready', 'Session 2 - Get Set', ...]
9591	['Learn step-by-step']
9592	['Week 1', 'WEEK 2']
9593	['Learn step-by-step']
9594	['Learn step-by-step']

	Instructor \
0	['Anna Burckhardt', 'Paola Antonelli', 'Michel...]
1	['Cary Nelson']
2	['Andrew Dennis', 'Ricardo Guimaraes']
3	['Eduardo de la Vara Brown.']

4	['Dariusz Terefenko']
...	...
9590	['William J. White']
9591	['Rachel Lindsay']
9592	['Pablo Palomino']
9593	['Rachel Lindsay']
9594	['Sara Nefzaoui']

	Offered By	Keyword \
0	['The Museum of Modern Art']	Arts and Humanities
1	['University of Illinois at Urbana-Champaign']	Arts and Humanities
2	['Michigan State University']	Arts and Humanities
3	['SAE Institute México']	Arts and Humanities
4	['University of Rochester']	Arts and Humanities
...
9590	['Northwestern University']	Social Sciences
9591	['Coursera Project Network']	Social Sciences
9592	['Emory University']	Social Sciences
9593	['Coursera Project Network']	Social Sciences
9594	['Coursera Project Network']	Social Sciences

	Course Url	Number Of Review \
0	https://www.coursera.org/learn/fashion-design	2813
1	https://www.coursera.org/learn/modern-american...	100
2	https://www.coursera.org/learn/pixel-art-video...	227
3	https://www.coursera.org/learn/distribucion-di...	NaN
4	https://www.coursera.org/learn/the-blues	582
...
9590	https://www.coursera.org/learn/power-onboarding	36
9591	https://www.coursera.org/projects/education-fe...	77
9592	https://www.coursera.org/learn/makingprogress	10
9593	https://www.coursera.org/projects/use-microsof...	11
9594	https://www.coursera.org/projects/creer-conten...	NaN

	Duration.1
0	20.0
1	34.0
2	9.0
3	8.0
4	11.0
...	...
9590	7.0
9591	NaN
9592	5.0
9593	1.0
9594	NaN

```
[9595 rows x 15 columns]
```

3 Information About Data

3.1 Dtypes

```
[81]: ds.dtypes
```

```
[81]: Course Title      object
      Rating           float64
      Level            object
      Duration          object
      Schedule          object
      Review            object
      What you will learn object
      Skill gain         object
      Modules           object
      Instructor         object
      Offered By         object
      Keyword            object
      Course Url         object
      Number Of Review   object
      Duration.1         float64
      dtype: object
```

3.2 Shape

```
[82]: ds.shape
```

```
[82]: (9595, 15)
```

3.3 Head()

```
[83]: ds.head()
```

```
[83]:
```

	Course Title	Rating	Level \
0	Fashion as Design	4.8	Beginner level
1	Modern American Poetry	4.4	Beginner level
2	Pixel Art for Video Games	4.5	Beginner level
3	Distribución digital de la música independiente	NaN	Beginner level
4	The Blues: Understanding and Performing an Ame...	4.8	Beginner level

	Duration	Schedule	Review \
0	20 hours (approximately)	Flexible schedule	2,813 reviews
1	Approx. 34 hours to complete	Flexible schedule	100 reviews
2	9 hours (approximately)	Flexible schedule	227 reviews
3	Approx. 8 hours to complete	Flexible schedule	NaN

4 Approx. 11 hours to complete Flexible schedule 582 reviews

What you will learn \

0 NaN

1 NaN

2 NaN

3 NaN

4 Students will be able to describe the blues as...

Skill gain \

0 ['Art History', 'Art', 'History', 'Creativity']

1 []

2 []

3 []

4 ['Music', 'Chord', 'Jazz', 'Jazz Improvisation']

Modules \

0 ['Introduction', 'Heroes', 'Silhouettes', 'Cou...]

1 ['Orientation', 'Module 1', 'Module 2', 'Modul...]

2 ['Week 1: Introduction to Pixel Art', 'Week 2:...

3 ['Semana 1', 'Semana 2', 'Semana 3', 'Semana 4']

4 ['Blues Progressions - Theory and Practice ', ...]

Instructor \

0 ['Anna Burckhardt', 'Paola Antonelli', 'Michel...]

1 ['Cary Nelson']

2 ['Andrew Dennis', 'Ricardo Guimaraes']

3 ['Eduardo de la Vara Brown.']

4 ['Dariusz Terefenko']

	Offered By	Keyword \
0	['The Museum of Modern Art']	Arts and Humanities
1	['University of Illinois at Urbana-Champaign']	Arts and Humanities
2	['Michigan State University']	Arts and Humanities
3	['SAE Institute México']	Arts and Humanities
4	['University of Rochester']	Arts and Humanities

	Course Url	Number Of Review \
0	https://www.coursera.org/learn/fashion-design	2813
1	https://www.coursera.org/learn/modern-american...	100
2	https://www.coursera.org/learn/pixel-art-video...	227
3	https://www.coursera.org/learn/distribucion-di...	NaN
4	https://www.coursera.org/learn/the-blues	582

Duration.1

0 20.0

1 34.0

```

2          9.0
3          8.0
4          11.0

```

3.4 Tail()

```
[84]: ds.tail()
```

```

[84]:
          Course Title  Rating \
9590                Power Onboarding    4.1
9591  Implementing Parent Feedback with Google Forms    4.9
9592                "Making" Progress Teach-Out    4.2
9593  Use Microsoft PowerPoint to Create a Visual Sc...    4.5
9594  Créer du contenu pour les réseaux sociaux avec...    NaN

          Level          Duration          Schedule \
9590  Beginner level  Approx. 7 hours to complete  Flexible schedule
9591  Beginner level                NaN                NaN
9592  Beginner level  Approx. 5 hours to complete  Flexible schedule
9593  Beginner level                1 hour                NaN
9594  Beginner level                NaN                NaN

          Review          What you will learn \
9590  36 reviews                NaN
9591  77 reviews  Creating engaging parent surveys using Google ...
9592  10 reviews                NaN
9593  11 reviews  Learn about choice boards and their use within...
9594    NaN  Vous découvrirez les fonctionnalités de Prezi ...

          Skill gain \
9590                []
9591  ['Education', 'online tools', 'engagement', 'e...
9592                []
9593  ['Teaching tools', 'elementary teaching', 'eng...
9594  ['Social Media Marketing', 'Social Media', 'Co...

          Modules          Instructor \
9590  ['Week 1 - Get Ready', 'Session 2 - Get Set', ...  ['William J. White']
9591                ['Learn step-by-step']    ['Rachel Lindsay']
9592                ['Week 1', 'WEEK 2']    ['Pablo Palomino']
9593                ['Learn step-by-step']    ['Rachel Lindsay']
9594                ['Learn step-by-step']    ['Sara Nefzaoui']

          Offered By          Keyword \
9590  ['Northwestern University']  Social Sciences
9591  ['Coursera Project Network']  Social Sciences
9592                ['Emory University']  Social Sciences

```

```
9593 ['Coursera Project Network'] Social Sciences
9594 ['Coursera Project Network'] Social Sciences
```

```

                                Course Url Number Of Review \
9590  https://www.coursera.org/learn/power-onboarding          36
9591  https://www.coursera.org/projects/education-fe...        77
9592  https://www.coursera.org/learn/makingprogress           10
9593  https://www.coursera.org/projects/use-microsof...        11
9594  https://www.coursera.org/projects/creer-conten...       NaN

```

```

Duration.1
9590      7.0
9591     NaN
9592      5.0
9593      1.0
9594     NaN

```

3.5 Columns

```
[85]: ds.columns
```

```
[85]: Index(['Course Title', 'Rating', 'Level', 'Duration', 'Schedule', 'Review',
           'What you will learn', 'Skill gain', 'Modules', 'Instructor',
           'Offered By', 'Keyword', 'Course Url', 'Number Of Review',
           'Duration.1'],
          dtype='object')
```

3.6 Size

```
[86]: ds.size
```

```
[86]: 143925
```

3.7 Info

```
[87]: ds.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 9595 entries, 0 to 9594
Data columns (total 15 columns):
#   Column              Non-Null Count  Dtype
---  -
0   Course Title        9595 non-null   object
1   Rating              8156 non-null   float64
2   Level              8330 non-null   object
3   Duration            9333 non-null   object
4   Schedule            8912 non-null   object
5   Review              8152 non-null   object

```



```

6   What you will learn  4984 non-null  object
7   Skill gain          9595 non-null  object
8   Modules             9595 non-null  object
9   Instructor          9595 non-null  object
10  Offered By          9595 non-null  object
11  Keyword             9595 non-null  object
12  Course Url          9595 non-null  object
13  Number Of Review    7913 non-null  object
14  Duration.1          9331 non-null  float64

```

dtypes: float64(2), object(13)

memory usage: 1.1+ MB

4 Data Cleaning

```
[88]: ds.head(4)
```

```

[88]:
           Course Title  Rating  Level \
0           Fashion as Design    4.8  Beginner level
1      Modern American Poetry    4.4  Beginner level
2      Pixel Art for Video Games    4.5  Beginner level
3  Distribución digital de la música independiente    NaN  Beginner level

```

```

           Duration  Schedule  Review \
0    20 hours (approximately)  Flexible schedule  2,813 reviews
1  Approx. 34 hours to complete  Flexible schedule    100 reviews
2     9 hours (approximately)  Flexible schedule   227 reviews
3  Approx. 8 hours to complete  Flexible schedule         NaN

```

```

What you will learn  Skill gain \
0           NaN  ['Art History', 'Art', 'History', 'Creativity']
1           NaN                                     []
2           NaN                                     []
3           NaN                                     []

```

```

           Modules \
0  ['Introduction', 'Heroes', 'Silhouettes', 'Cou...
1  ['Orientation', 'Module 1', 'Module 2', 'Modul...
2  ['Week 1: Introduction to Pixel Art', 'Week 2:...
3  ['Semana 1', 'Semana 2', 'Semana 3', 'Semana 4']

```

```

           Instructor \
0  ['Anna Burckhardt', 'Paola Antonelli', 'Michel...
1                                     ['Cary Nelson']
2  ['Andrew Dennis', 'Ricardo Guimaraes']
3                                     ['Eduardo de la Vara Brown.']

```

```

           Offered By  Keyword \

```

```

0          ['The Museum of Modern Art']  Arts and Humanities
1  ['University of Illinois at Urbana-Champaign']  Arts and Humanities
2          ['Michigan State University']  Arts and Humanities
3          ['SAE Institute México']  Arts and Humanities

```

```

                                Course Url Number Of Review \
0      https://www.coursera.org/learn/fashion-design          2813
1  https://www.coursera.org/learn/modern-american...         100
2  https://www.coursera.org/learn/pixel-art-video...        227
3  https://www.coursera.org/learn/distribucion-di...        NaN

```

```

Duration.1
0      20.0
1      34.0
2       9.0
3       8.0

```

4.1 Checking The Null Values

```
[89]: ds.isnull()
```

```

[89]:      Course Title  Rating  Level  Duration  Schedule  Review \
0          False    False   False    False    False    False
1          False    False   False    False    False    False
2          False    False   False    False    False    False
3          False     True    False    False    False     True
4          False    False   False    False    False    False
...          ...      ...      ...      ...      ...      ...
9590         False    False   False    False    False    False
9591         False    False   False     True     True    False
9592         False    False   False    False    False    False
9593         False    False   False    False     True    False
9594         False     True    False     True     True     True

```

```

      What you will learn  Skill gain  Modules  Instructor  Offered By \
0          True          False    False    False    False
1          True          False    False    False    False
2          True          False    False    False    False
3          True          False    False    False    False
4         False          False    False    False    False
...          ...      ...      ...      ...      ...
9590         True          False    False    False    False
9591         False          False    False    False    False
9592         True          False    False    False    False
9593         False          False    False    False    False
9594         False          False    False    False    False

```

	Keyword	Course Url	Number Of Review	Duration.1
0	False	False	False	False
1	False	False	False	False
2	False	False	False	False
3	False	False	True	False
4	False	False	False	False
...
9590	False	False	False	False
9591	False	False	False	True
9592	False	False	False	False
9593	False	False	False	False
9594	False	False	True	True

[9595 rows x 15 columns]

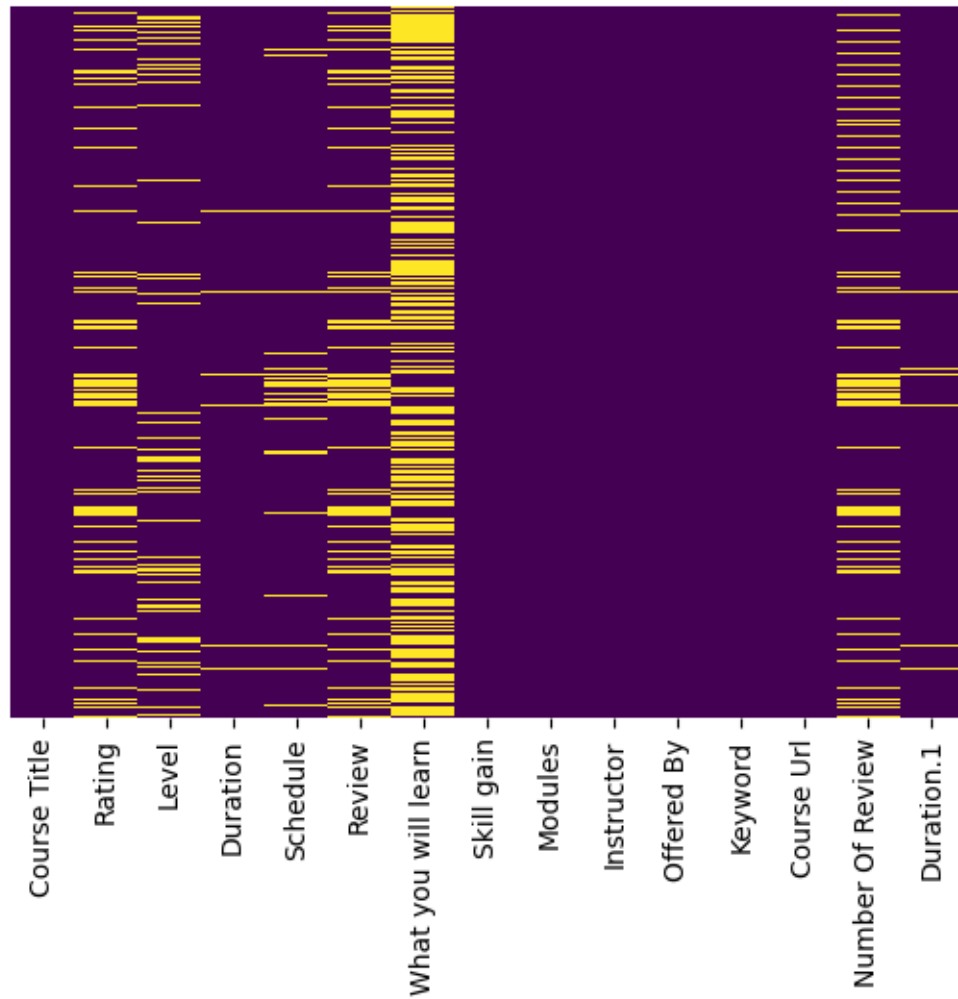
```
[90]: ds.isnull().sum()
```

```
[90]: Course Title          0
      Rating                1439
      Level                1265
      Duration              262
      Schedule             683
      Review               1443
      What you will learn  4611
      Skill gain           0
      Modules              0
      Instructor           0
      Offered By           0
      Keyword              0
      Course Url           0
      Number Of Review     1682
      Duration.1           264
      dtype: int64
```

4.2 Checking The Null Values Through Graph

```
[91]: sc.heatmap(ds.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
[91]: <Axes: >
```



4.3 By Handling The Null Values

```
[92]: dsc=ds.fillna(0)
```

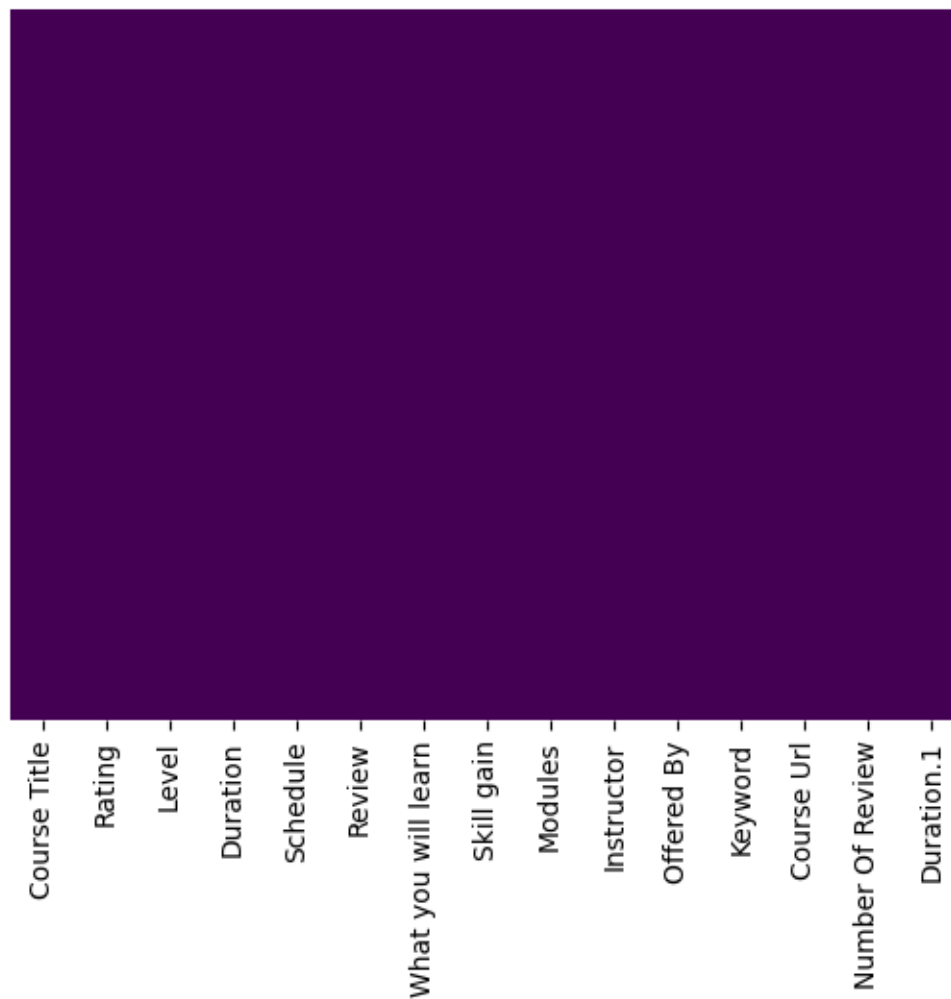
```
[93]: dsc.isnull().sum()
```

```
[93]: Course Title      0
      Rating           0
      Level            0
      Duration          0
      Schedule          0
      Review            0
      What you will learn 0
      Skill gain         0
      Modules           0
```

```
Instructor          0
Offered By          0
Keyword             0
Course Url          0
Number Of Review    0
Duration.1          0
dtype: int64
```

```
[94]: sc.heatmap(dsc.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
[94]: <Axes: >
```



4.4 Now Checked The Duplicates Values In DataSets

```
[95]: dsc.duplicated().sum()
```

```
[95]: 900
```

```
[96]: dub_count=dsc.duplicated(keep=False)
```

```
[97]: heatmap_data=dub_count.astype(int).values.reshape(-1,1)
```

```
[98]: sc.  
      ↪heatmap(heatmap_data,cbar=False,cmap='viridis',yticklabels=False,xticklabels=False)
```

```
[98]: <Axes: >
```



4.5 By Deleting The Duplicates Values

```
[99]: fdsc=dsc.drop_duplicates()
```

4.6 After Removing Duplicates Values Now Checked

```
[100]: fdsc.duplicated().sum()
```

```
[100]: 0
```

```
[101]: dub_count=fdsc.duplicated(keep=False)

[102]: heatmap_data=dub_count.astype(int).values.reshape(-1,1)

[103]: sc.
↳heatmap(heatmap_data,cbar=False,cmap='viridis',yticklabels=False,xticklabels=False)

[103]: <Axes: >
```



4.7 Drop The Unnecessary Columns

```
[104]: fd=fdsc.drop(['Review','Duration'],axis=1)

[105]: fd.columns

[105]: Index(['Course Title', 'Rating', 'Level', 'Schedule', 'What you will learn',
          'Skill gain', 'Modules', 'Instructor', 'Offered By', 'Keyword',
          'Course Url', 'Number Of Review', 'Duration.1'],
          dtype='object')
```

4.8 Handle The Unnecessary Values In Data

```
[106]: fd1=fd.replace([''], 'No Specified')

[107]: fd2=fd1.rename(columns={'Duration.1': 'Duration to complete'})

[108]: fd2.columns

[108]: Index(['Course Title', 'Rating', 'Level', 'Schedule', 'What you will learn',
          'Skill gain', 'Modules', 'Instructor', 'Offered By', 'Keyword',
          'Course Url', 'Number Of Review', 'Duration to complete'],
          dtype='object')

[109]: fd2['Level']=fd2['Level'].replace(0, 'No Specified')

[110]: fd2['Schedule']=fd2['Schedule'].replace(0, 'Fixed Schedule')
```

5 Data Analysis & Visualization

```
[111]: fd2.head()
```

```
[111]:
```

	Course Title	Rating	Level	\
0	Fashion as Design	4.8	Beginner level	
1	Modern American Poetry	4.4	Beginner level	
2	Pixel Art for Video Games	4.5	Beginner level	
3	Distribución digital de la música independiente	0.0	Beginner level	
4	The Blues: Understanding and Performing an Ame...	4.8	Beginner level	

	Schedule	What you will learn	\
0	Flexible schedule		0
1	Flexible schedule		0
2	Flexible schedule		0
3	Flexible schedule		0
4	Flexible schedule	Students will be able to describe the blues as...	

	Skill gain	\
0	['Art History', 'Art', 'History', 'Creativity']	
1	No Specified	
2	No Specified	
3	No Specified	
4	['Music', 'Chord', 'Jazz', 'Jazz Improvisation']	

	Modules	\
0	['Introduction', 'Heroes', 'Silhouettes', 'Cou...']	
1	['Orientation', 'Module 1', 'Module 2', 'Modul...']	
2	['Week 1: Introduction to Pixel Art', 'Week 2:...	
3	['Semana 1', 'Semana 2', 'Semana 3', 'Semana 4']	


```

4  ['Blues Progressions - Theory and Practice ', ...

                                Instructor \
0  ['Anna Burckhardt', 'Paola Antonelli', 'Michel...
1                                ['Cary Nelson']
2                                ['Andrew Dennis', 'Ricardo Guimaraes']
3                                ['Eduardo de la Vara Brown.']
4                                ['Dariusz Terefenko']

                                Offered By      Keyword \
0                                ['The Museum of Modern Art'] Arts and Humanities
1  ['University of Illinois at Urbana-Champaign'] Arts and Humanities
2                                ['Michigan State University'] Arts and Humanities
3                                ['SAE Institute México'] Arts and Humanities
4                                ['University of Rochester'] Arts and Humanities

                                Course Url Number Of Review \
0      https://www.coursera.org/learn/fashion-design      2813
1  https://www.coursera.org/learn/modern-american...      100
2  https://www.coursera.org/learn/pixel-art-video...      227
3  https://www.coursera.org/learn/distribucion-di...      0
4      https://www.coursera.org/learn/the-blues      582

Duration to complete
0      20.0
1      34.0
2      9.0
3      8.0
4      11.0

```

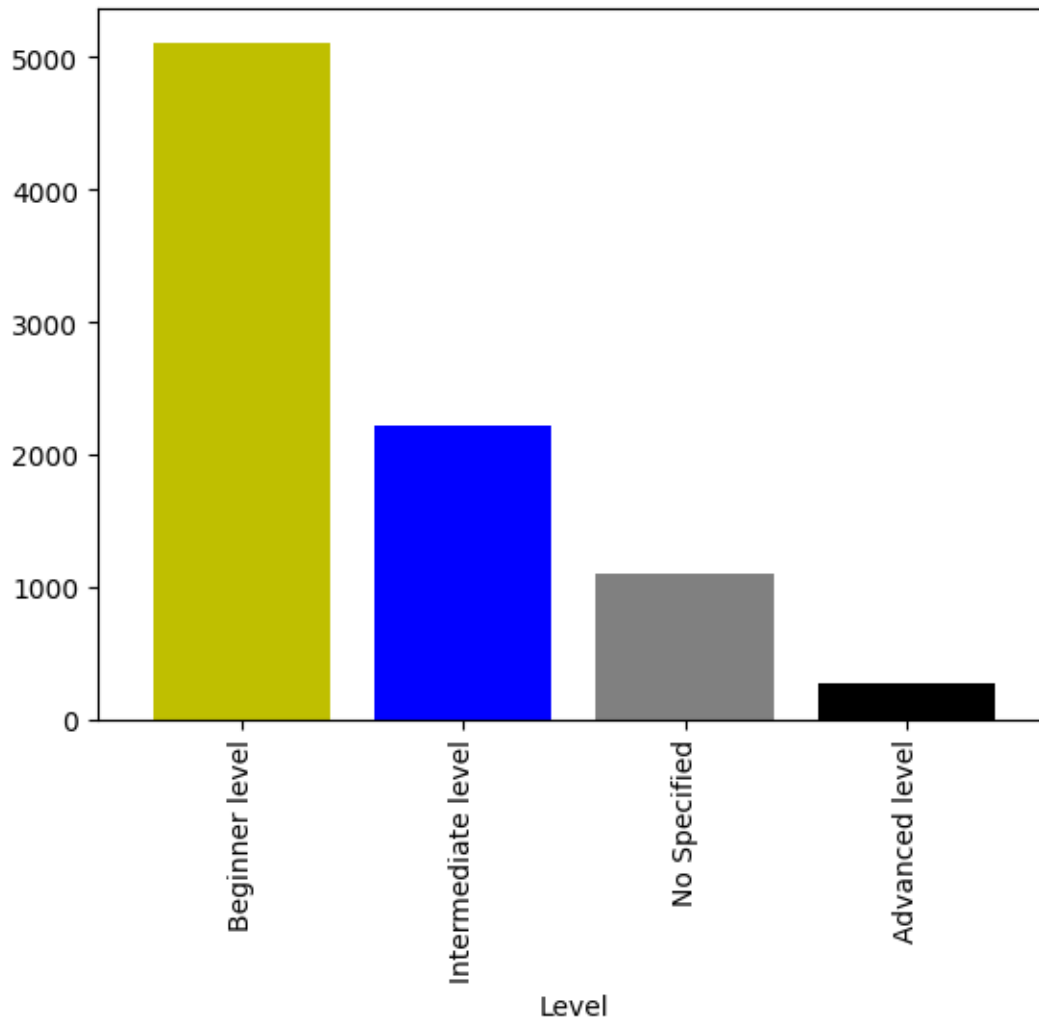
5.1 Checking The Courses By Level

```
[112]: fd2['Level'].value_counts()
```

```
[112]: Level
Beginner level      5112
Intermediate level  2211
No Specified        1106
Advanced level       266
Name: count, dtype: int64
```

```
[113]: fd2['Level'].value_counts().
        plot(kind='bar',color=['y','b','gray','black'],width=0.8)
```

```
[113]: <Axes: xlabel='Level'>
```



5.2 The most popular courses based on their reviews

```
[114]: fd2['Number Of Review']=pd.to_numeric(fd2['Number Of Review'],errors='coerce')
```

```
[115]: sort_data=fd2.sort_values(by='Number Of Review', ascending=False)
```

```
[116]: top_review=sort_data[['Course Title','Number Of Review']].head(10)
```

```
[117]: top_review
```

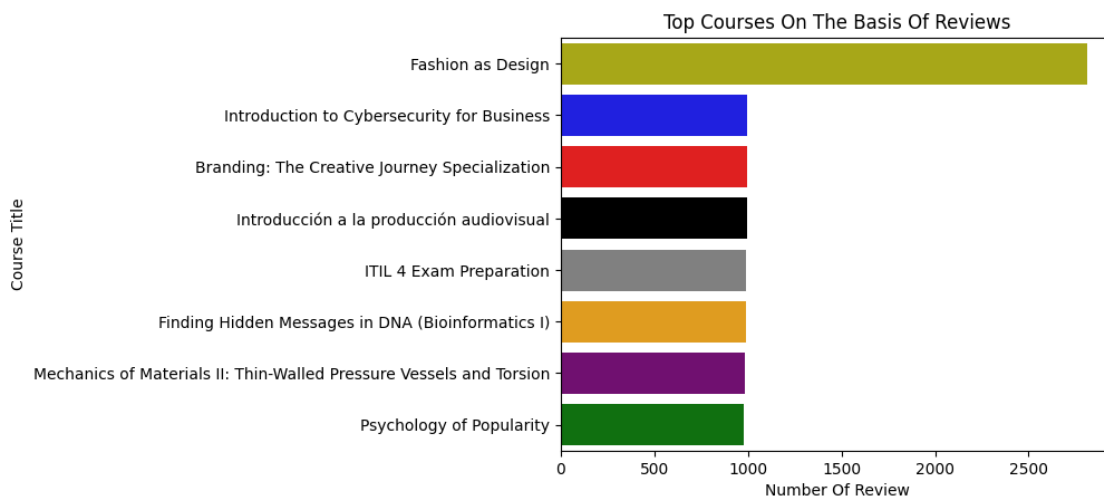
```
[117]:
```

	Course Title	Number Of Review
0	Fashion as Design	2813.0
4995	Introduction to Cybersecurity for Business	998.0
6084	Branding: The Creative Journey Specialization	995.0
8161	Introducción a la producción audiovisual	994.0

4706	ITIL 4 Exam Preparation	993.0
3401	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
3967	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
5442	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
7426	Mechanics of Materials II: Thin-Walled Pressur...	986.0
4064	Psychology of Popularity	981.0

```
[118]: colors=['y','b','red','k','gray','orange','purple','green']
sc.barplot(data=top_review,x='Number Of Review',y='Course_
↪Title',palette=colors,hue='Course Title')
mt.title("Top Courses On The Basis Of Reviews")
```

```
[118]: Text(0.5, 1.0, 'Top Courses On The Basis Of Reviews')
```



5.3 Instructor Impact On Course Success

```
[119]: top_instructors=sort_data[['Instructor','Number Of Review','Course Title']].
↪head(5)
```

```
[120]: top_instructors
```

```
[120]:
```

	Instructor	Number Of Review \
0	['Anna Burckhardt', 'Paola Antonelli', 'Michel...	2813.0
4995	['Greg Williams']	998.0
6084	['MARÍA EIZAGUIRRE DIÉGUEZ', 'Michael Thompson...	995.0
8161	['Vicente Hinojosa Alarcón', 'Julio César Pére...	994.0
4706	['Kelley Conrad', 'Sabrina Moore', 'Allyson Pi...	993.0

	Course Title
0	Fashion as Design

```

4995      Introduction to Cybersecurity for Business
6084 Branding: The Creative Journey Specialization
8161      Introducción a la producción audiovisual
4706      ITIL 4 Exam Preparation

```

```

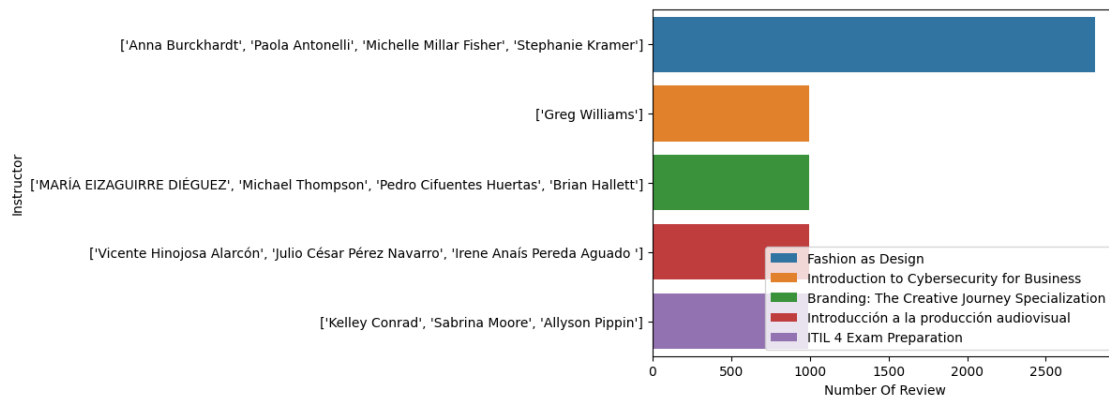
[121]: sc.barplot(data=top_instructors,x='Number Of Review',y='Instructor',hue='Course_
      ↪Title')
      mt.legend(loc='lower right')

```

```

[121]: <matplotlib.legend.Legend at 0x2274475f410>

```



5.4 How Course Duration Impact The Course Rating And Student Satisfaction

```

[122]: fd2['Rating']=pd.to_numeric(fd2['Rating'],errors='coerce')

```

```

[123]: sort_rat=fd2.sort_values(['Duration to complete'],ascending=False)

```

```

[124]: top_rat=sort_rat[['Course Title','Duration to_
      ↪complete','Rating','Schedule','Number Of Review','Offered By','Level']].
      ↪head(10)

```

```

[125]: top_rat

```

```

[125]:
      Course Title  Duration to complete \
6687      Fundamentals of Web Development      133.0
55      Write Your First Novel      127.0
2356      Database Management Essentials      122.0
7438  Análisis estático lineal y no lineal de estruc...      103.0
4026      Toxicology 21: Scientific Applications      95.0
298      Modern & Contemporary American Poetry ("ModPo")      94.0
51      Write A Feature Length Screenplay For Film Or ...      93.0
5795  Build a Modern Computer from First Principles:...      89.0
1622  Build a Modern Computer from First Principles:...      89.0

```

3095 Build a Modern Computer from First Principles:...

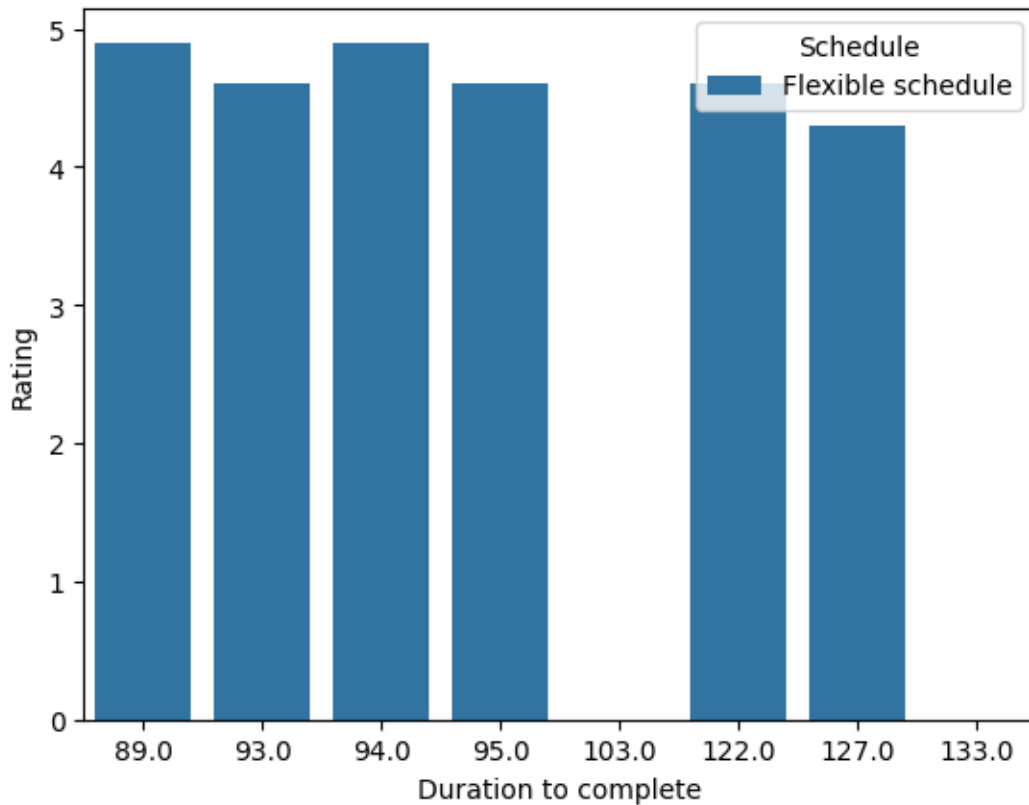
89.0

	Rating	Schedule	Number Of Review \
6687	0.0	Flexible schedule	0.0
55	4.3	Flexible schedule	0.0
2356	4.6	Flexible schedule	0.0
7438	0.0	Flexible schedule	0.0
4026	4.6	Flexible schedule	76.0
298	4.9	Flexible schedule	21.0
51	4.6	Flexible schedule	66.0
5795	4.9	Flexible schedule	496.0
1622	4.9	Flexible schedule	179.0
3095	4.9	Flexible schedule	496.0

	Offered By	Level
6687	['Illinois Tech']	Beginner level
55	['Michigan State University']	Beginner level
2356	['University of Colorado System']	Intermediate level
7438	['Universidad de los Andes']	Advanced level
4026	['Johns Hopkins University']	Intermediate level
298	['University of Pennsylvania']	Beginner level
51	['Michigan State University']	Beginner level
5795	['Hebrew University of Jerusalem']	No Specified
1622	['Hebrew University of Jerusalem']	No Specified
3095	['Hebrew University of Jerusalem']	No Specified

```
[126]: sc.barplot(data=top_rat,x='Duration to complete',y='Rating',hue='Schedule')
```

```
[126]: <Axes: xlabel='Duration to complete', ylabel='Rating'>
```



```
[127]: low_rat=sort_rat[['Course Title','Duration to complete','Rating','Schedule']].
      ↪tail(10)
```

```
[128]: low_rat
```

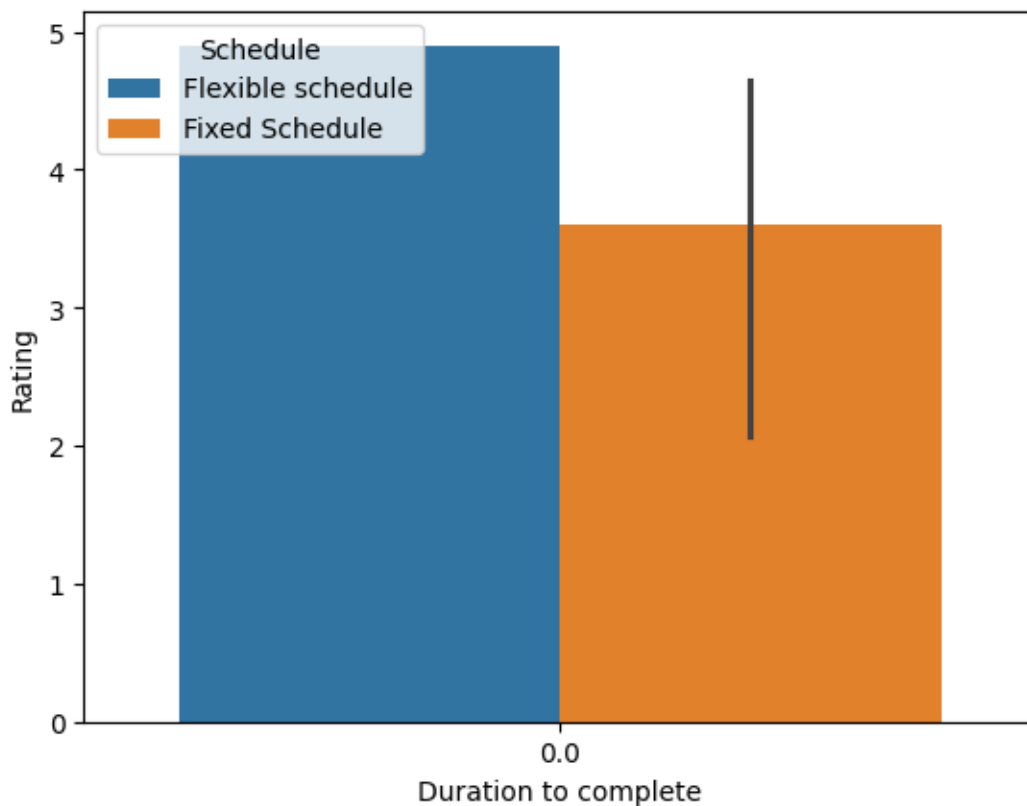
```
[128]:
```

	Course Title	Duration to complete	Rating	Schedule
4364	IBM Business Intelligence (BI) Analyst Profess...	0.0	4.9	Flexible schedule
5285	Advanced SQL Retrieval Queries in SQLiteStudio	0.0	4.5	Fixed Schedule
8479	Creating Interactive Learning Videos with Edpu...	0.0	4.7	Fixed Schedule
920	Data Analysis with SQL: Inform a Business Deci...	0.0		
4474	Introduction to virtual machines in Microsoft ...	0.0		
7483	Building an Ecosystem model with Insight Maker	0.0		
8490	Improving Classroom Management with Class Dojo	0.0		
4478	Google Workspace Admin: Getting Started	0.0		
4500	Creating Database Tables with SQL	0.0		
5125	Configure Replication and Enable Point-in-Time...	0.0		

920	4.5	Fixed Schedule
4474	4.8	Fixed Schedule
7483	4.6	Fixed Schedule
8490	4.7	Fixed Schedule
4478	0.0	Fixed Schedule
4500	4.6	Fixed Schedule
5125	0.0	Fixed Schedule

```
[129]: sc.barplot(data=low_rat,x='Duration to complete',y='Rating',hue='Schedule')
```

```
[129]: <Axes: xlabel='Duration to complete', ylabel='Rating'>
```



5.5 Analyze The Which Skills Is Required For Data Scientist

```
[130]: ds_df=fd2[fd2['Modules'].str.contains('Data Science |Machine_L  
↳Learning|Python|AI',case=False)]
```

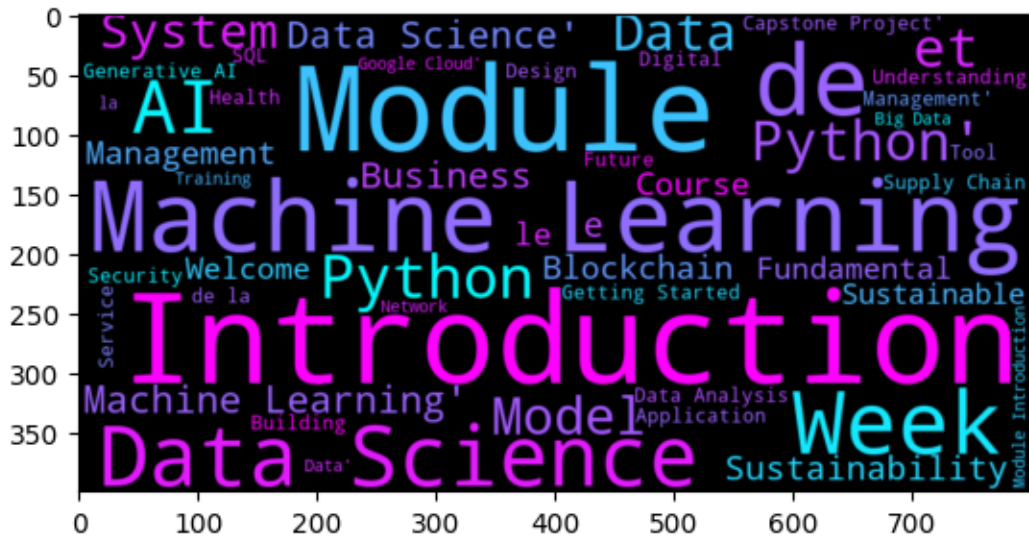
```
[131]: text=' '.join(ds_df['Modules'].tolist())
```

```
[132]: wordc=wc(width=800,height=400,max_words=50,colormap='cool').generate(text)  
mt.Figure(figsize=(15,8))
```

```
[132]: <Figure size 1500x800 with 0 Axes>
```

```
[133]: mt.imshow(wordc,interpolation='bilinear')
```

```
[133]: <matplotlib.image.AxesImage at 0x227404e2f00>
```



5.6 How Different Institute Perform In Terms Of High Rated Courses

```
[134]: sort_data=fd2.sort_values(by=['Number Of Review','Rating'],
    ↪ascending=[False,False])
```

```
[135]: top_institute=sort_data[['Course Title','Number Of Review','Rating','Offered_
    ↪By','Level','Duration to complete']].head(10)
```

```
[136]: top_institute
```

```
[136]:
```

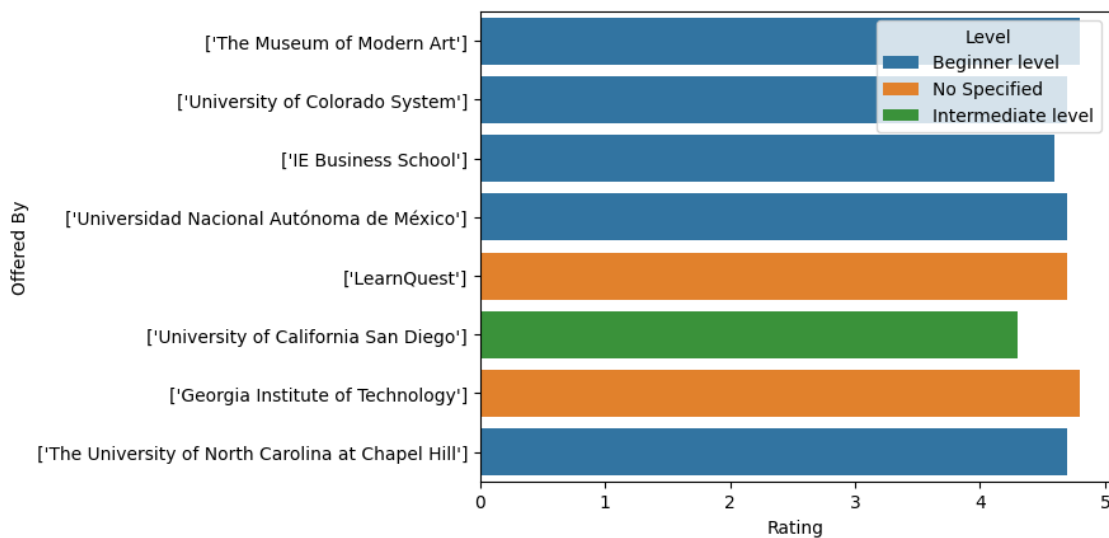
	Course Title	Number Of Review \
0	Fashion as Design	2813.0
4995	Introduction to Cybersecurity for Business	998.0
6084	Branding: The Creative Journey Specialization	995.0
8161	Introducción a la producción audiovisual	994.0
4706	ITIL 4 Exam Preparation	993.0
3401	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
3967	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
5442	Finding Hidden Messages in DNA (Bioinformatics I)	992.0
7426	Mechanics of Materials II: Thin-Walled Pressur...	986.0
4064	Psychology of Popularity	981.0

	Rating	Offered By \
0	4.8	['The Museum of Modern Art']
4995	4.7	['University of Colorado System']
6084	4.6	['IE Business School']
8161	4.7	['Universidad Nacional Autónoma de México']
4706	4.7	['LearnQuest']
3401	4.3	['University of California San Diego']
3967	4.3	['University of California San Diego']
5442	4.3	['University of California San Diego']
7426	4.8	['Georgia Institute of Technology']
4064	4.7	['The University of North Carolina at Chapel H...']

	Level	Duration to complete
0	Beginner level	20.0
4995	Beginner level	13.0
6084	Beginner level	10.0
8161	Beginner level	22.0
4706	No Specified	17.0
3401	Intermediate level	15.0
3967	Intermediate level	15.0
5442	Intermediate level	15.0
7426	No Specified	9.0
4064	Beginner level	7.0

```
[137]: sc.barplot(data=top_institute,x='Rating',y='Offered By',hue='Level')
```

```
[137]: <Axes: xlabel='Rating', ylabel='Offered By'>
```



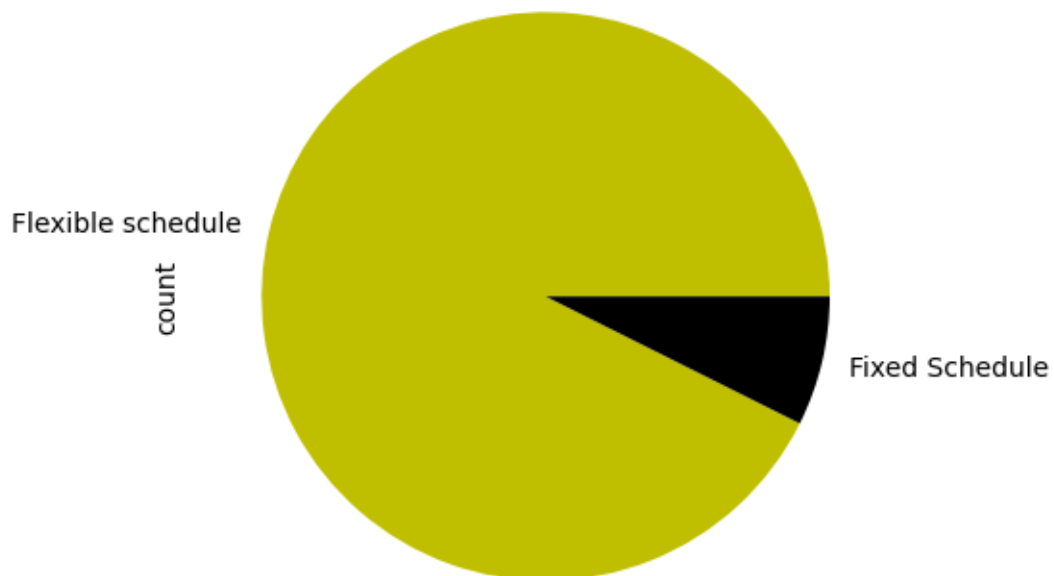
5.7 Identify The Courses With Flexible Or Fixed

```
[138]: fd2['Schedule'].value_counts()
```

```
[138]: Schedule
Flexible schedule      8054
Fixed Schedule         641
Name: count, dtype: int64
```

```
[139]: fd2['Schedule'].value_counts().plot(kind='pie', colors=['y', 'k'])
```

```
[139]: <Axes: ylabel='count'>
```



5.8 How Level Of Course Effect By Duration

```
[140]: sort_data=fd2.sort_values(['Duration to complete'], ascending=False)
```

```
[141]: top_duration=sort_data[['Level', 'Duration to complete']].head(10)
```

```
[142]: top_duration
```

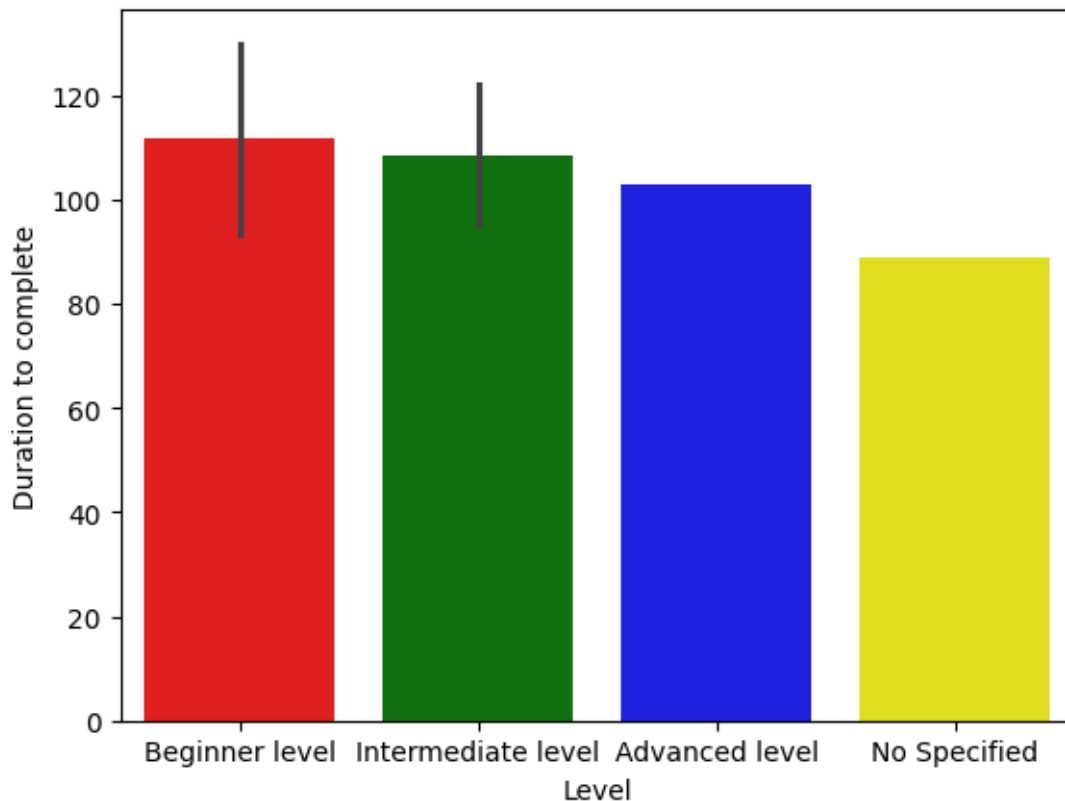
```
[142]:
```

	Level	Duration to complete
6687	Beginner level	133.0
55	Beginner level	127.0

2356	Intermediate level	122.0
7438	Advanced level	103.0
4026	Intermediate level	95.0
298	Beginner level	94.0
51	Beginner level	93.0
5795	No Specified	89.0
1622	No Specified	89.0
3095	No Specified	89.0

```
[143]: colors=['red','green','blue','yellow']
sc.barplot(data=top_duration,x='Level',y='Duration to complete',hue='Level',palette=colors)
```

```
[143]: <Axes: xlabel='Level', ylabel='Duration to complete'>
```



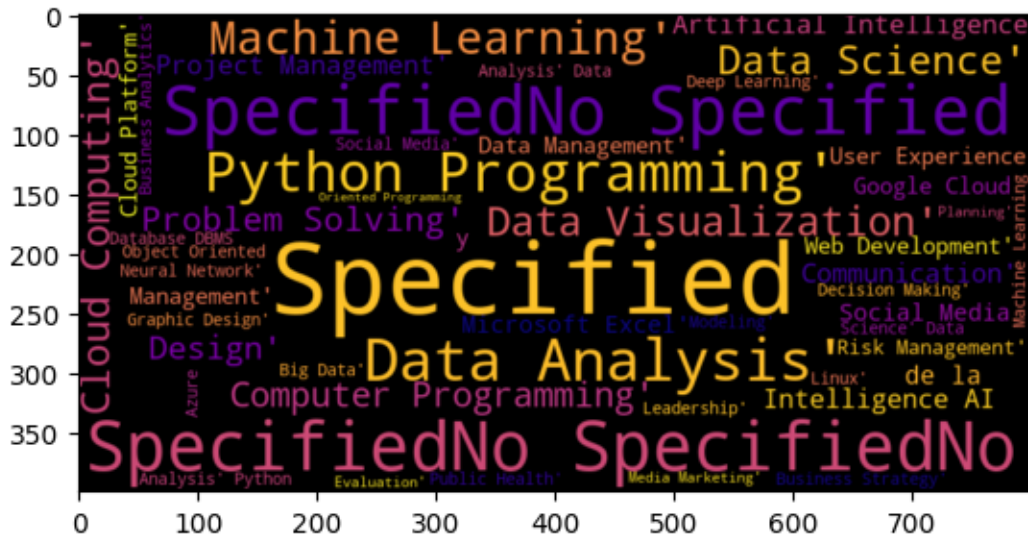
5.9 Showing The Popular Course With Word Clouds

```
[144]: skill_gain_text=''.join(fd2['Skill gain'].tolist())
wordc=wc(width=800,height=400,max_words=50,colormap='plasma').
generate(skill_gain_text)
mt.Figure(figsize=(15,8))
```

```
[144]: <Figure size 1500x800 with 0 Axes>
```

```
[145]: mt.imshow(wordc,interpolation='bilinear')
```

```
[145]: <matplotlib.image.AxesImage at 0x22747723f20>
```



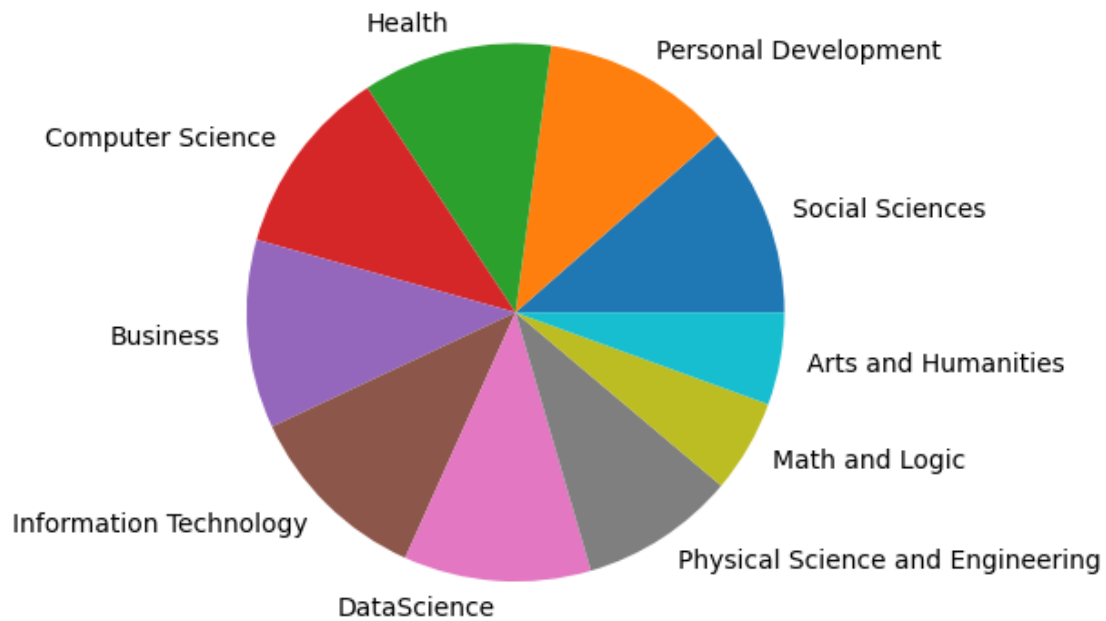
5.10 Analyze The Which Topics On Trends

```
[146]: fd2['Keyword'].value_counts()
```

```
[146]: Keyword
Social Sciences          996
Personal Development     993
Health                   992
Computer Science         988
Business                 985
Information Technology    982
DataScience              979
Physical Science and Engineering  812
Math and Logic           486
Arts and Humanities      482
Name: count, dtype: int64
```

```
[147]: fd2['Keyword'].value_counts().plot(kind='pie',ylabel='')
```

```
[147]: <Axes: >
```



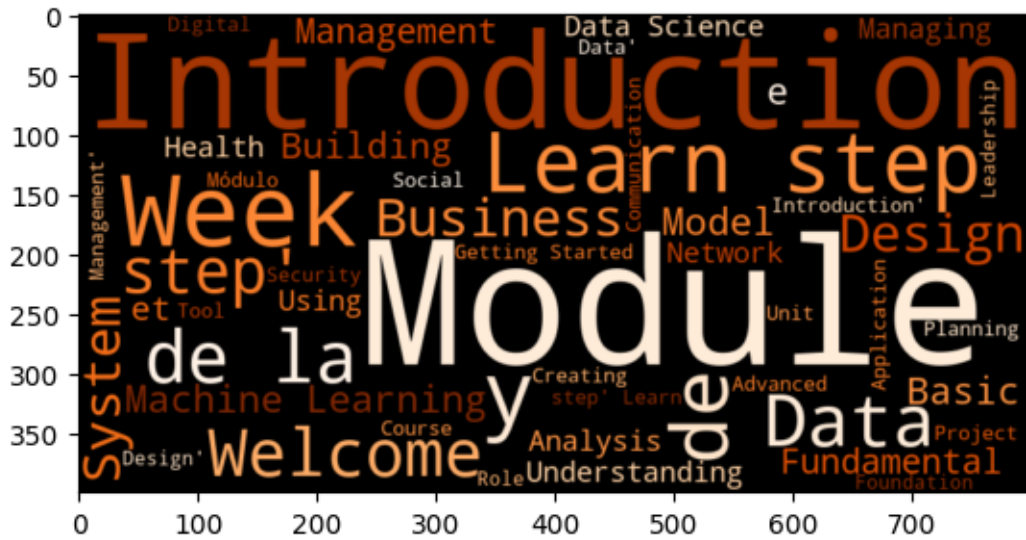
5.11 Mapping Modules Skills With Word Clouds

```
[148]: skill_gain_text=''.join(fd2['Modules'].tolist())
      wordc=wc(width=800,height=400,max_words=50,colormap='Oranges').
      ↪generate(skill_gain_text)
      mt.Figure(figsize=(15,8))
```

```
[148]: <Figure size 1500x800 with 0 Axes>
```

```
[149]: mt.imshow(wordc,interpolation='bilinear')
```

```
[149]: <matplotlib.image.AxesImage at 0x22744a67e60>
```



5.12 Identify The Most Common Learning Path Taken By The Users

```
[150]: fd2['Modules'].value_counts().head(5)
```

```
[150]: Modules
['Learn step-by-step']          542
No Specified                     99
['Week 1', 'Week 2', 'Week 3', 'Week 4']    28
['Module 1', 'Module 2', 'Module 3', 'Module 4']  18
['Week 1', 'Week 2']             13
Name: count, dtype: int64
```

```
[151]: fd2['Modules'].value_counts().head(10).
        plot(kind='barh',color=['yellow','blue','red','green','orange','purple','pink','brown','gray'])
```

```
[151]: <Axes: ylabel='Modules'>
```

