

DATENANALYSE - UDEMY COURSES DATASET

Dilanur Gider - 190503060
Deniz Özdemirli - 190503029

In dieser Untersuchung haben wir den Datensatz von UdeMy-Kursen verwendet. Unser Projekt besteht aus fünf Phasen: Lesen des Datensatzes, die Bereinigung der Daten, die Analyse der Daten, Datenvisualisierung (Plotting) und endlich Interpretation von Diagrammen und Ziehen von Schlussfolgerungen .

EINFÜHRUNG

Dieser Datensatz enthält Informationen über UdeMy-Kurse, einschließlich Kurstitel, URL, Preis, Anzahl der Abonnenten, Anzahl der Bewertungen, Anzahl der Vorlesungen, Kursniveau, Bewertung, Inhaltsdauer, Zeitstempel der Veröffentlichung und Thema. Die Kategorien dieser Kurse sind Business, Design, Webentwicklung und Musik.

DATA ANALYSE

level	subject	num_subscribers
All Levels	Business Finance	633
	Graphic Design	335
	Musical Instruments	324
	Web Development	633
Beginner Level	Business Finance	399
	Graphic Design	184
	Musical Instruments	266
	Web Development	422
Expert Level	Business Finance	25
	Graphic Design	7
	Musical Instruments	12
	Web Development	14
Intermediate Level	Business Finance	134
	Graphic Design	76
	Musical Instruments	78
	Web Development	134

Free/Paid	
Free	11532.318328
Paid	2429.103120

num_subscribers	
year_published	
2013	268923
2015	121584
2014	120291

DATA CLEANING

```
df.isnull().sum()
df.dropna(inplace=True)

# Search for any duplicates
df.duplicated().sum()
df.drop_duplicates(inplace=True)
```

```
# Convert published_timestamp datatype to datetime
df['published_timestamp'] = df['published_timestamp'].str.replace("T", " ").str.replace("Z", "")
df['published_timestamp'] = pd.to_datetime(df['published_timestamp'])
df['year_published'] = df['published_timestamp'].dt.year
del df['Date']
```

Als erstes haben wir überprüft, ob die Tabelle Nullwerte und Duplikate enthält. Und wir haben diese Werte gelöscht. Dann haben wir 'published_timestamp' in 'datetime' Datentyp konvertiert. Addiert neue Spalte mit nur Jahrwert. Als letztes löschte wir 'Date' spalte, da Tabelle enthält bereits 2 Date Information.

Bevor der Data Cleaning

	course_id	course_title	url	price	num_subscribers	num_reviews	num_lectures	level	rating	content_duration	published_timestamp	subject	Date	Free/Paid
0	41295	Learn HTML5 Programming From Scratch	https://www.udemy.com/learn-html5-programming-...	0	268923	8629	45	Beginner Level	0.82	10.5	2013-02-14T07:03:41Z	Web Development	2013-02-14	Free
1	59014	Coding for Entrepreneurs Basic	https://www.udemy.com/coding-for-entrepreneurs-...	0	161029	279	27	Expert Level	0.69	3.5	2013-06-09T15:51:55Z	Web Development	2013-06-09	Free

Nach der Data Cleaning (mit Sortierung nach Jahr)

	course_id	course_title	url	price	num_subscribers	num_reviews	num_lectures	level	rating	content_duration	published_timestamp	subject	Free/Paid	year_published
3668	1282064	Cryptocurrency (BTC & ETH) Investment & Tradin...	https://www.udemy.com/cryptocurrency-btc-eth-...	20	0	0	7	All Levels	0.95	2.5	2017-07-06 21:46:30	Business Finance	Paid	2017
3649	1187084	Forex how traders beat the market with little...	https://www.udemy.com/forex-how-traders-beat-t...	100	0	0	16	All Levels	0.15	1.0	2017-07-06 21:16:13	Business Finance	Paid	2017

ERGEBNISSE

Bei der Datenanalyse festgestellt, dass die höchste Zahl von Abonnenten im Allgemeinen bei kostenlosen Kursen zu verzeichnen war. Es wurde auch festgestellt, dass die meisten Kurse im Jahr 2016 veröffentlicht wurden. Obwohl die Zahl der Webentwicklungskurse bis 2016 rapide anstieg, begann diese Zahl nach 2016 zu sinken. Der Kurs mit den meisten Teilnehmern war "Learn HTML5 From Scratch" aus der Kategorie Webentwicklung.

GRAFIKEN

