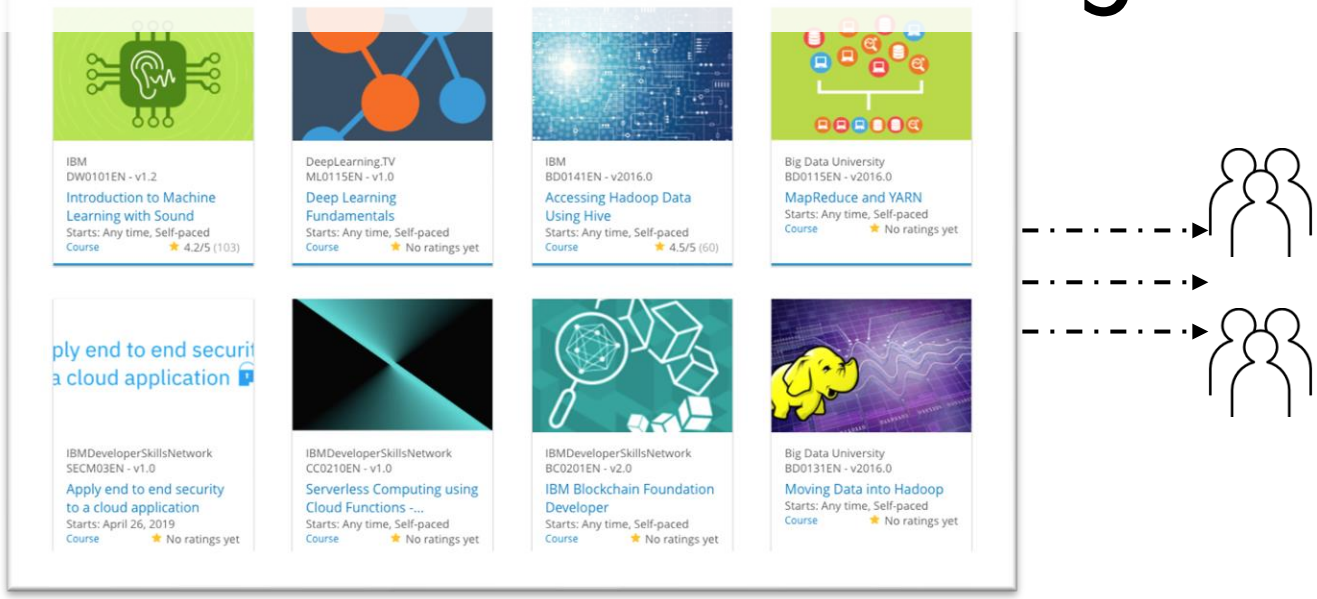
Build a Personalized Online Course Recommender System with Machine Learning

Vinicius Torres
15/12/2022



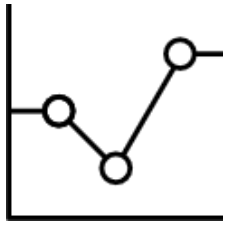
Outline

- Introduction and Background
- Exploratory Data Analysis
- Content-based Recommender System using Unsupervised Learning
- Collaborative-filtering based Recommender System using Supervised learning
- Conclusion
- Appendix

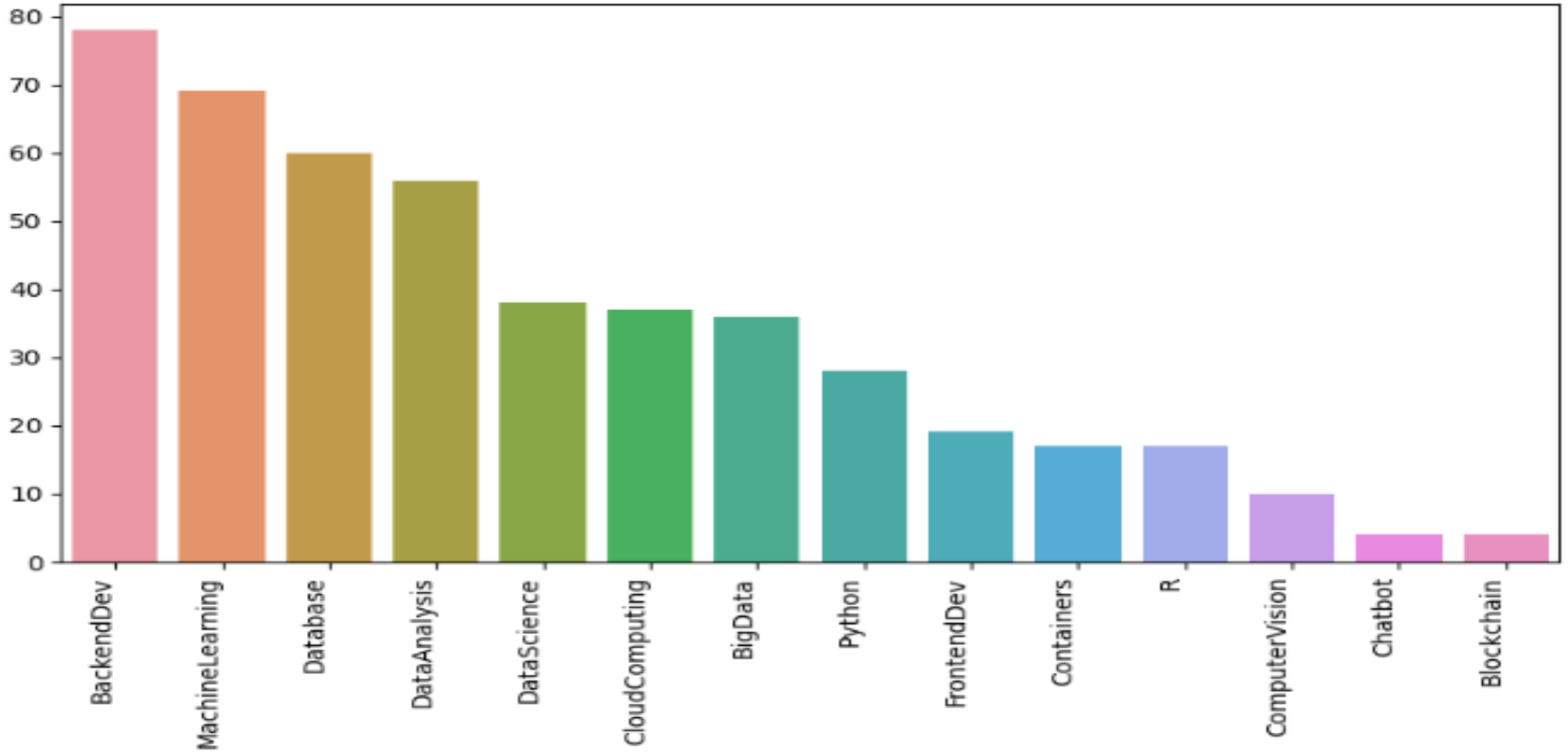
Introduction

- The main objective of this analysis was to provide Data Science course's recommendations based on recorded information on the user's preferences using both, supervised and unsupervised learning.

Exploratory Data Analysis

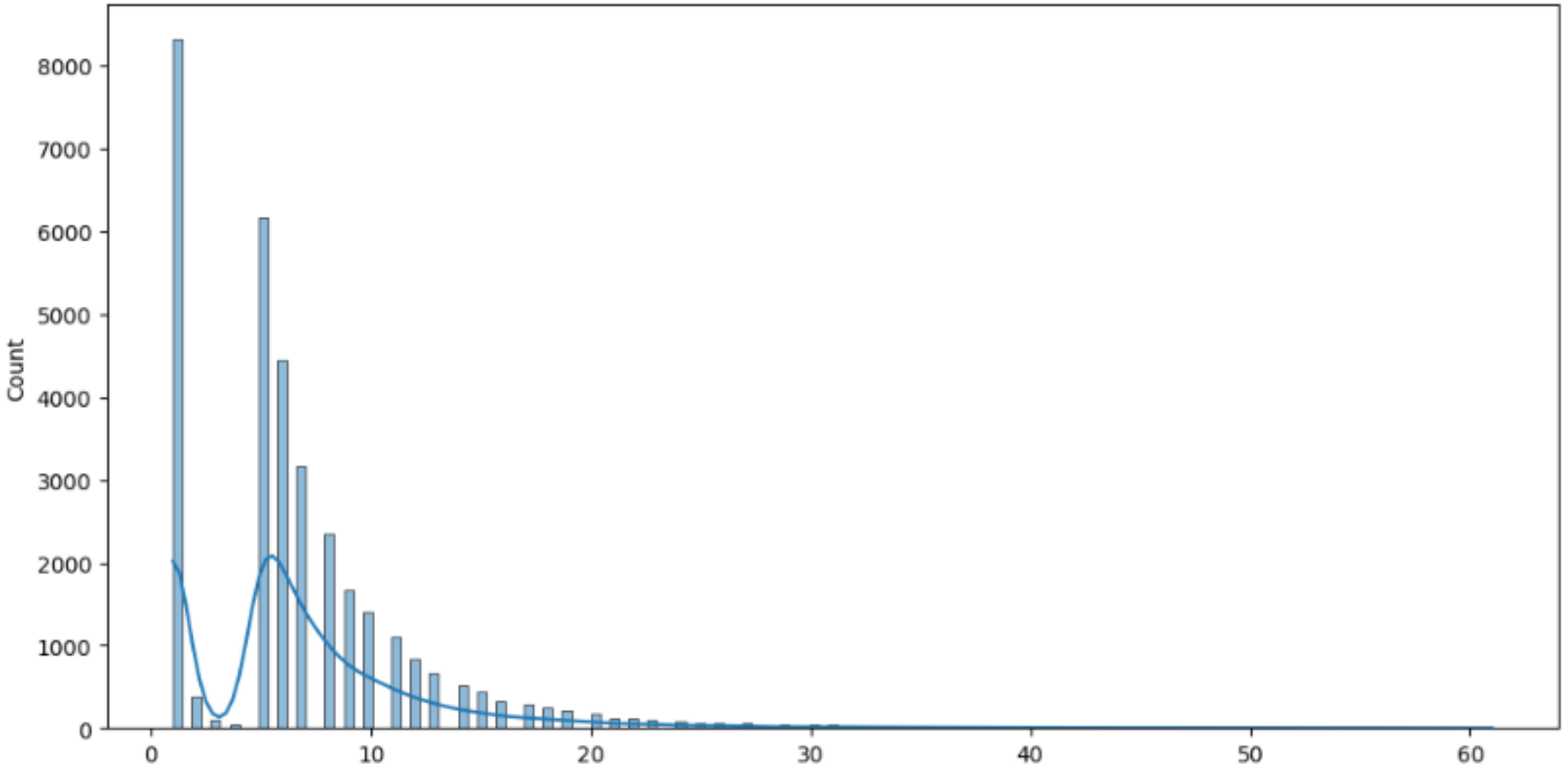


Course counts per genre



- BackEndDev, Machine Learning and Database are the topics with the higher number of courses.

Course enrollment distribution



- Most courses have lower enrollment (less than 10) and fewer courses have higher enrollment.

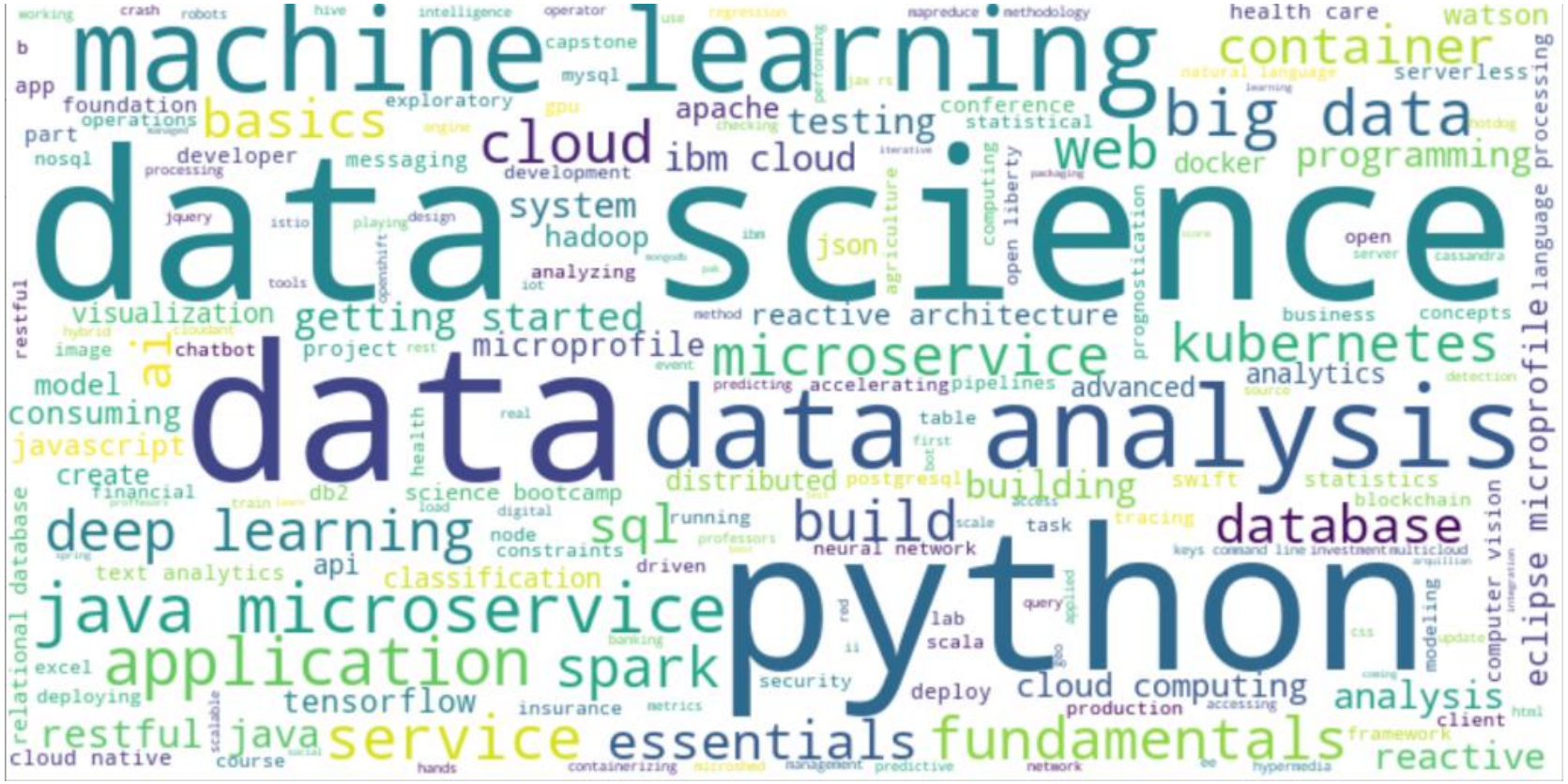
10 most popular courses

[96]:

	TITLE	Enrolls
0	python for data science	14936
1	introduction to data science	14477
2	big data 101	13291
3	hadoop 101	10599
4	data analysis with python	8303
5	data science methodology	7719
6	machine learning with python	7644
7	spark fundamentals i	7551
8	data science hands on with open source tools	7199
9	blockchain essentials	6719
10	data visualization with python	6709

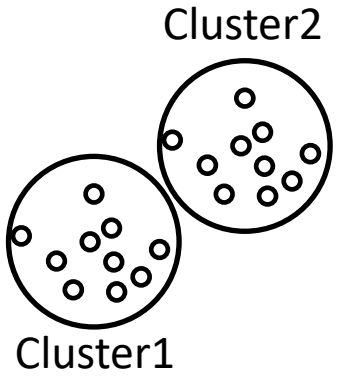
- Courses with a more general propose seems to be more popular than specifics one.

Word cloud of course titles

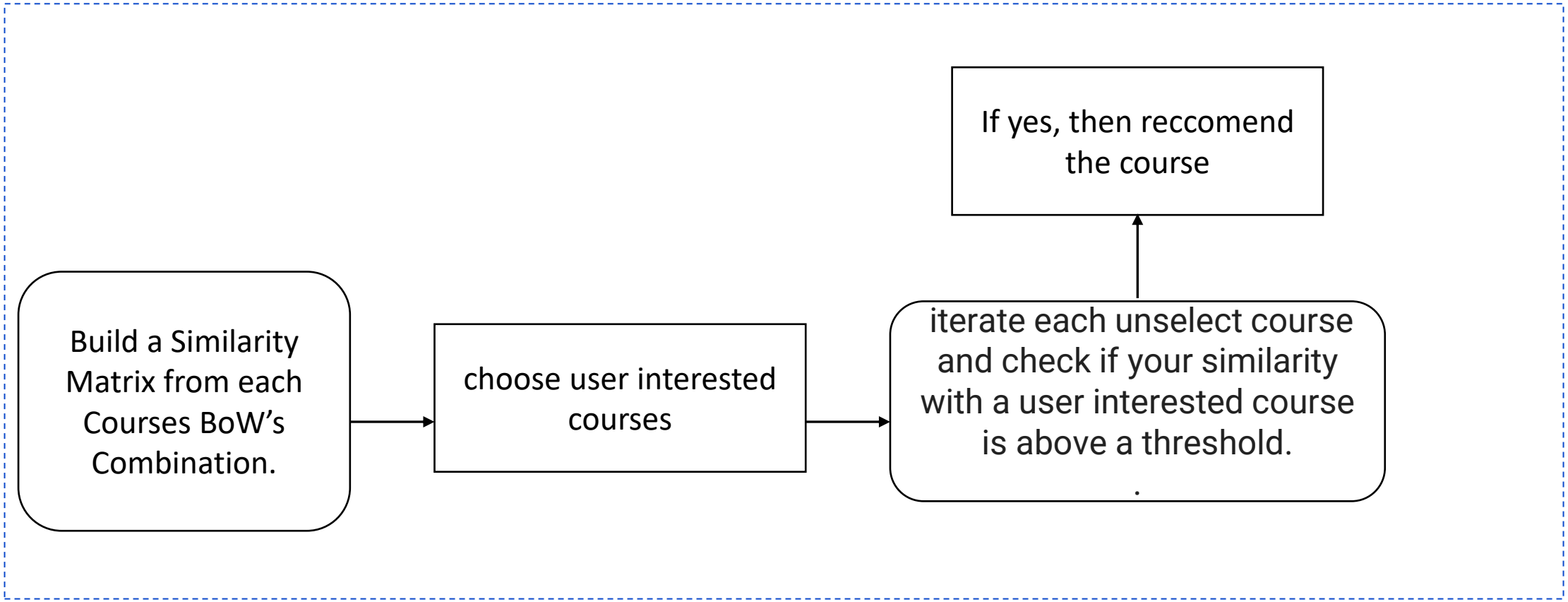


- Data Science, Machine Learning and Python are the most frequents words in the course's titles

Content-based Recommender System using Unsupervised Learning



Flowchart of content-based recommender system using course similarity



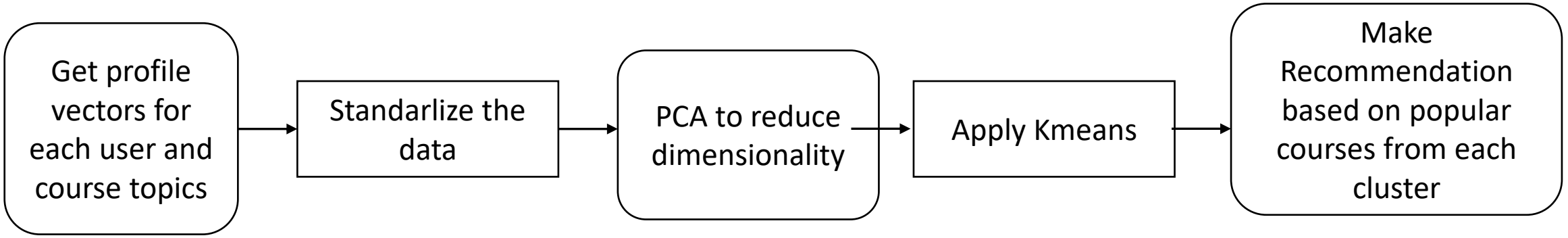
Evaluation results of course similarity based recommender system

- Threshold to consider similar courses: **60%**
- On average, **1.985** new/unseen courses have been recommended to each user

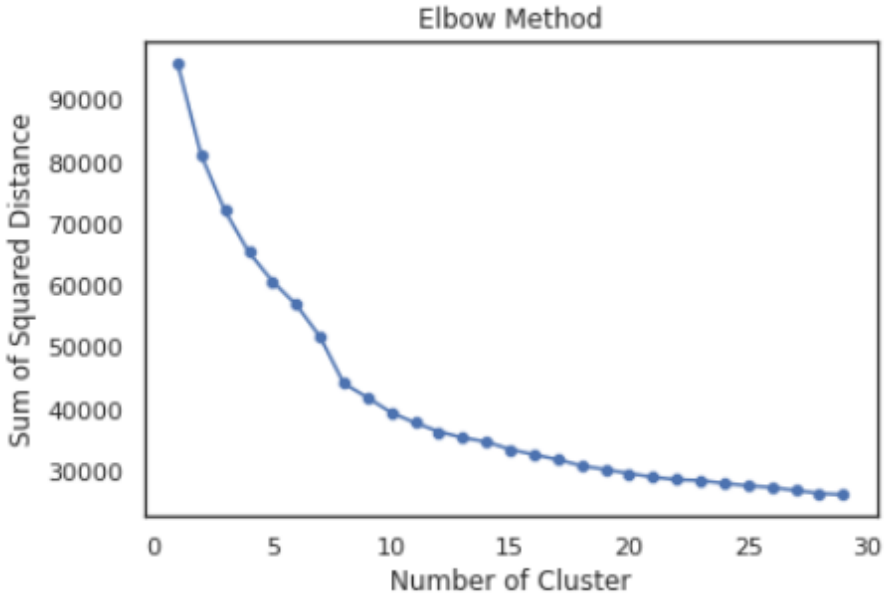
Top Recommended Courses and Frequency

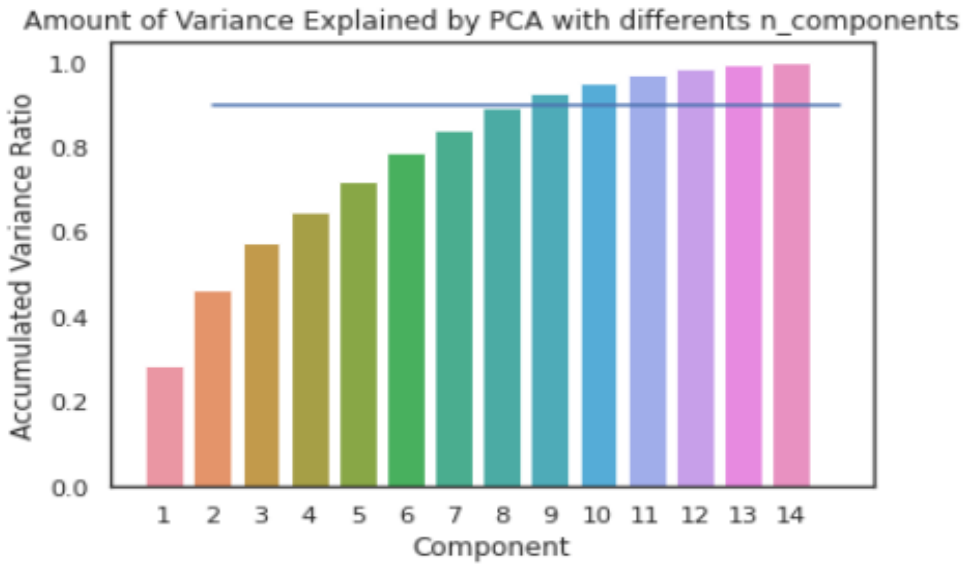
```
PY0101EN      257
excourse22    257
excourse62    257
ST0101EN      146
WA0103EN      104
WA0101EN      104
RP0101EN       71
TA0105         58
TA0105EN       58
SC0101EN       39
Name: course, dtype: int64
```

Flowchart of clustering-based recommender system



Evaluation results of clustering-based recommender system





Top Recommended Courses and Frequency

DS0101EN	157
DV0101EN	124
ML0101ENv3	113
DA0101EN	97
BD0211EN	34
PY0101EN	29
BD0101EN	26
BD0111EN	15
CB0103EN	4

- Number of Clusters for Kmeans: **10**
- Number of Components for PCA: **9** (Explain 90% from the variance)
- Threshold to consider a course popular: **> 100 enrollments**
- On average, **0.599** new/unseen courses have been recommended to each user