

Lesson 38: BoW and TF-IDF

Data Science with Python – BSc Course

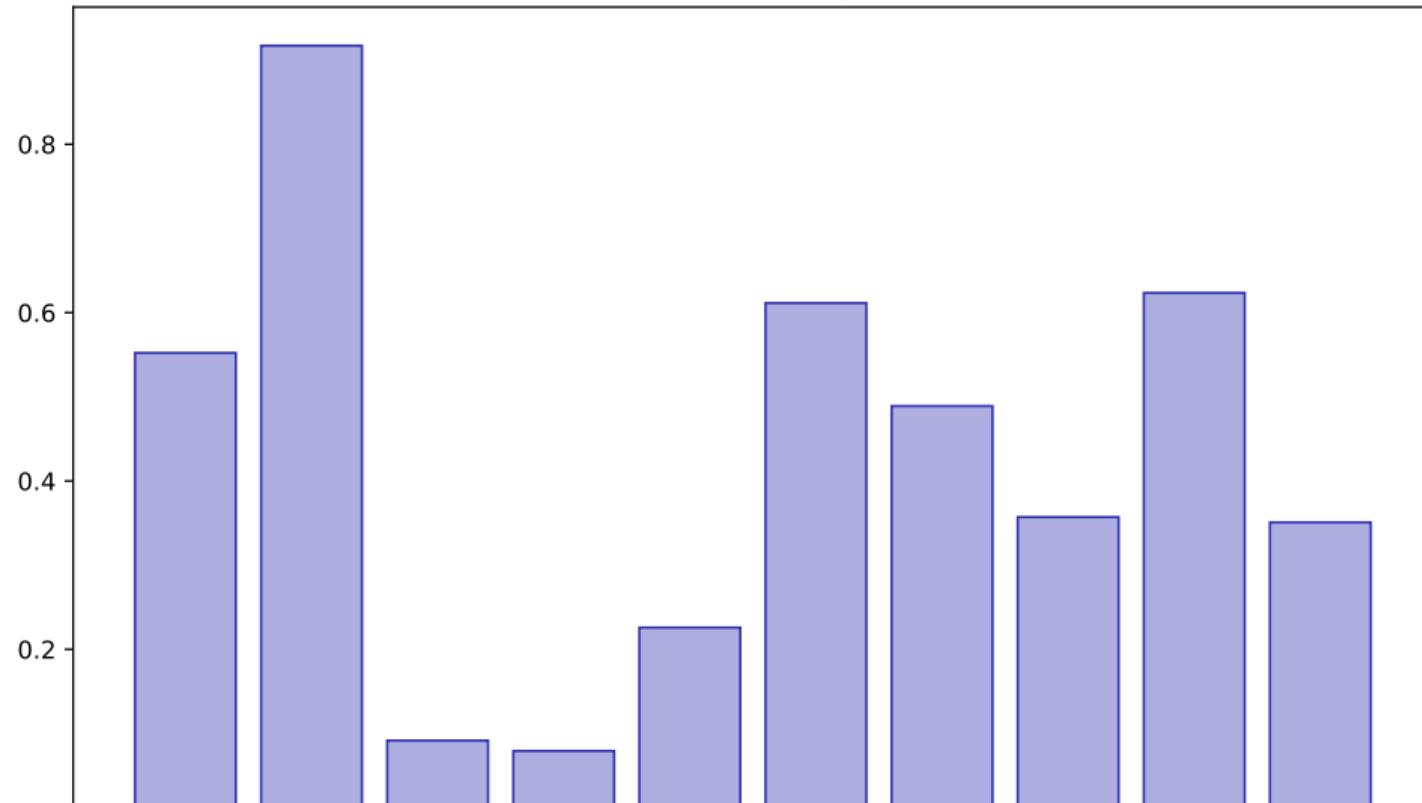
45 Minutes

After this lesson, you will be able to:

- Create bag of words
- Calculate TF-IDF weights
- Build document vectors
- Apply to text classification

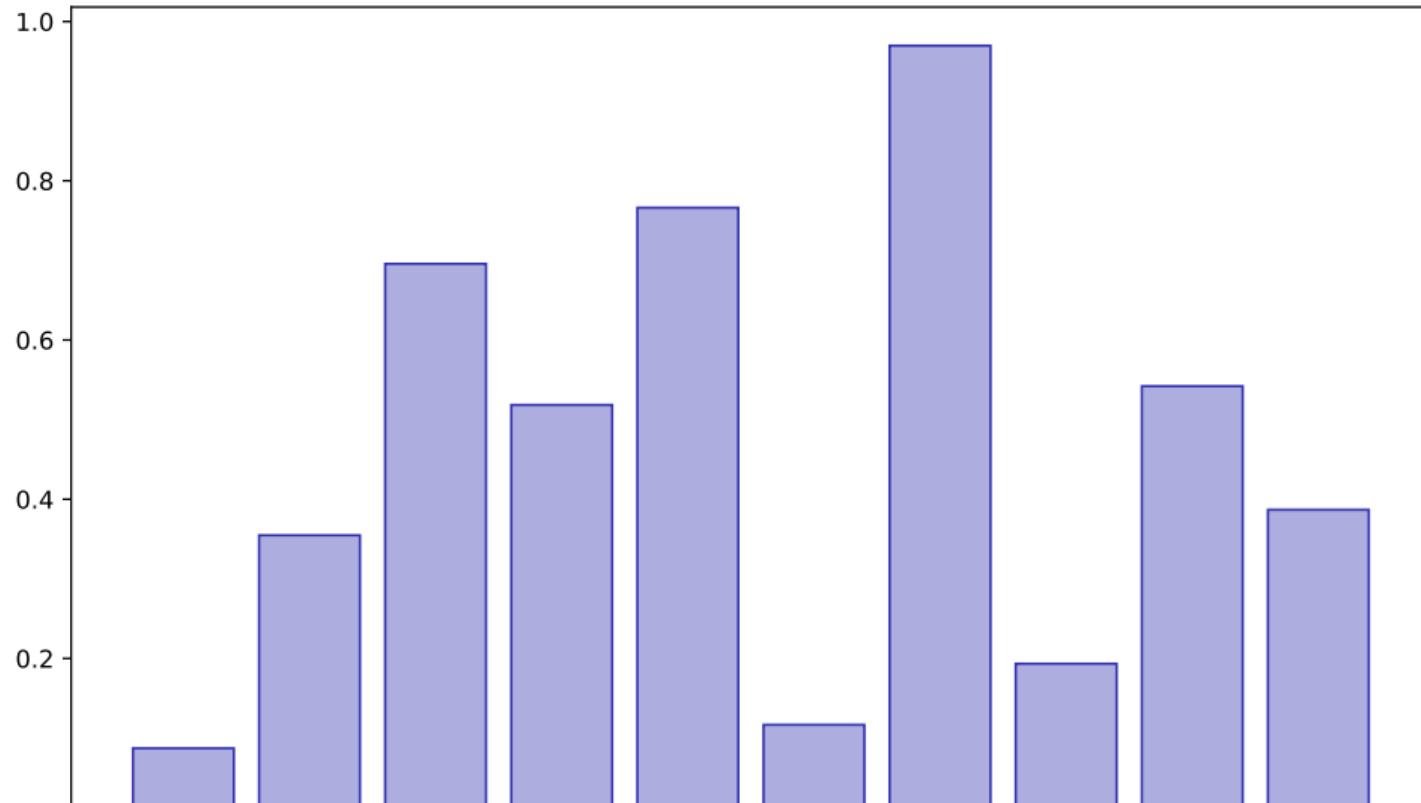
Building towards your final project

Bow Concept

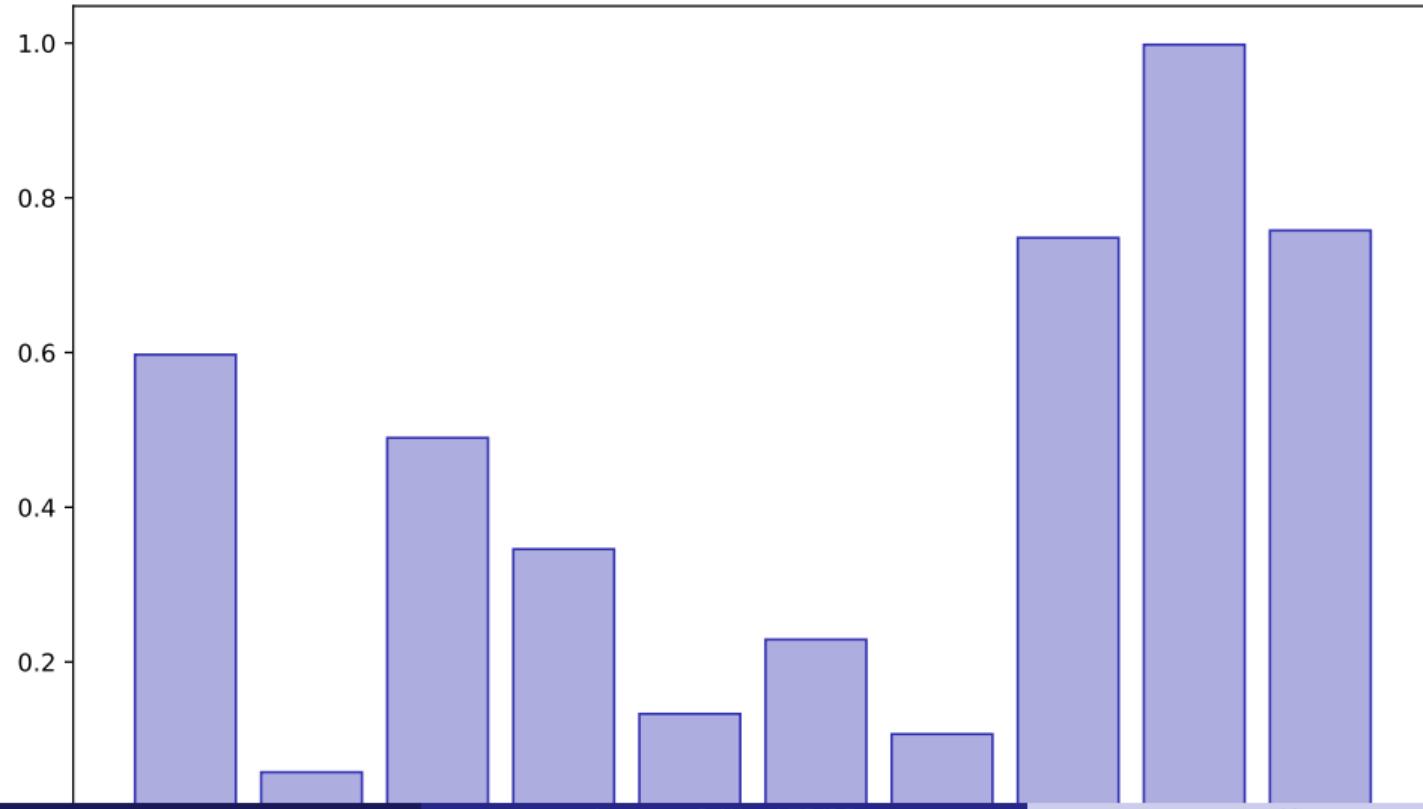


Count Vectorizer

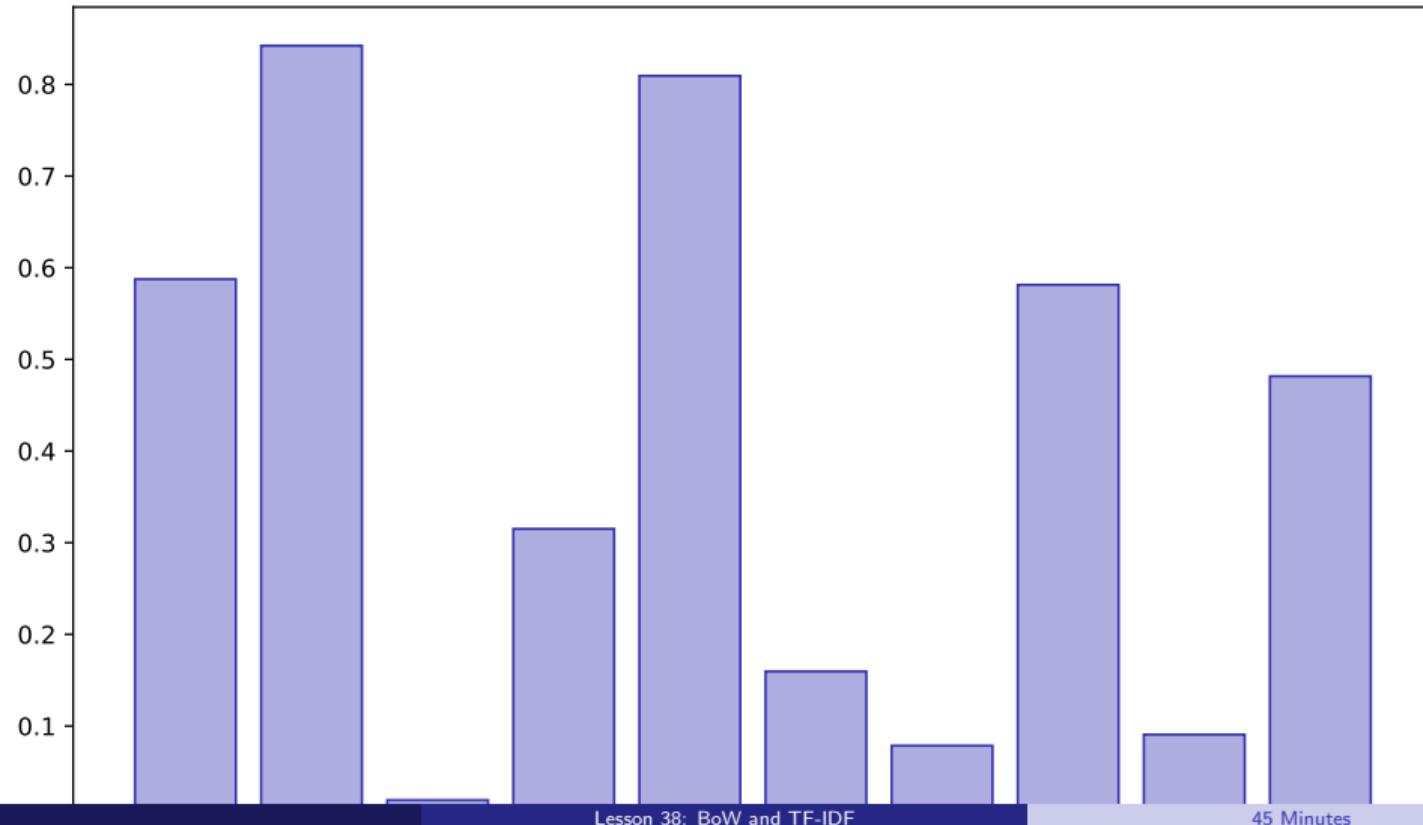
Count Vectorizer



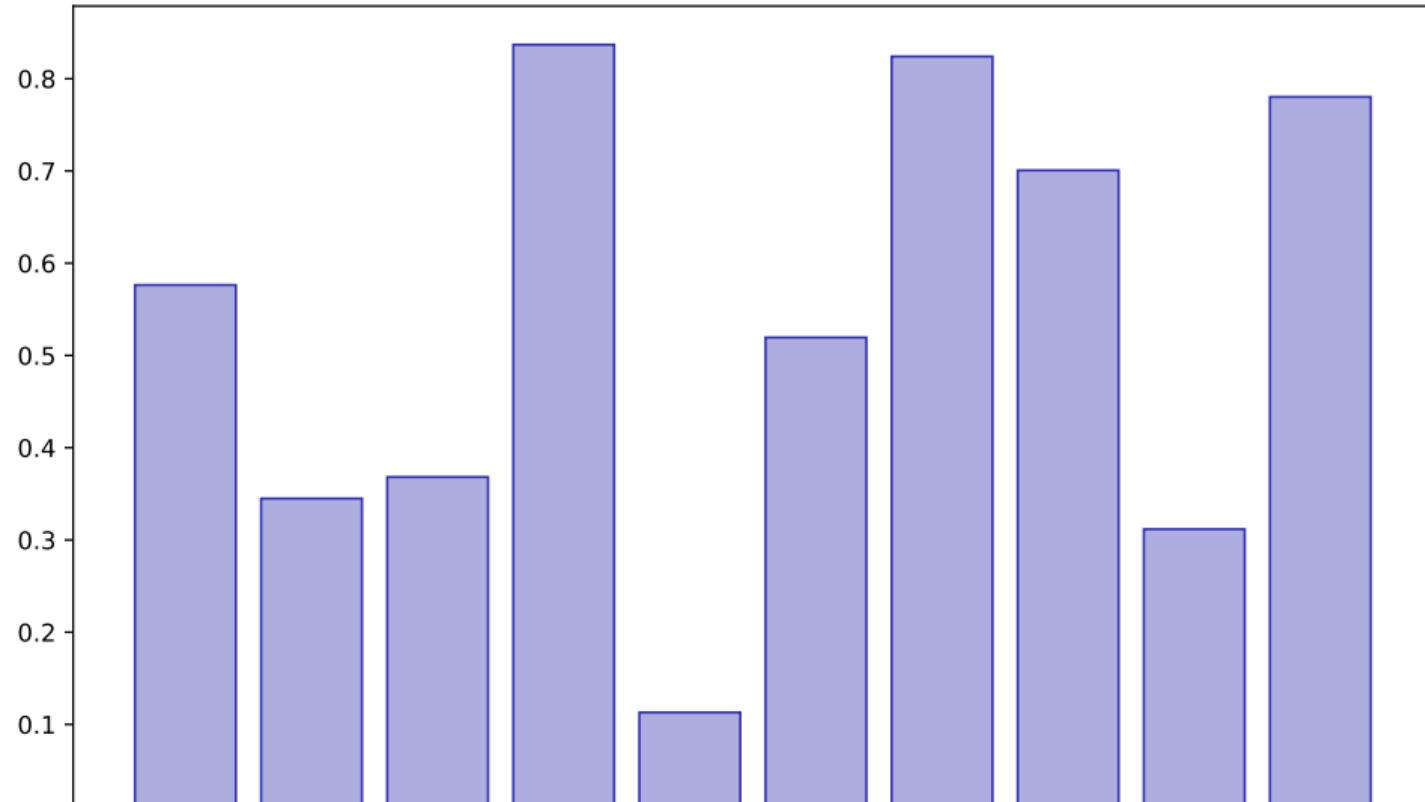
Tfidf Formula



Tfidf Vectorizer

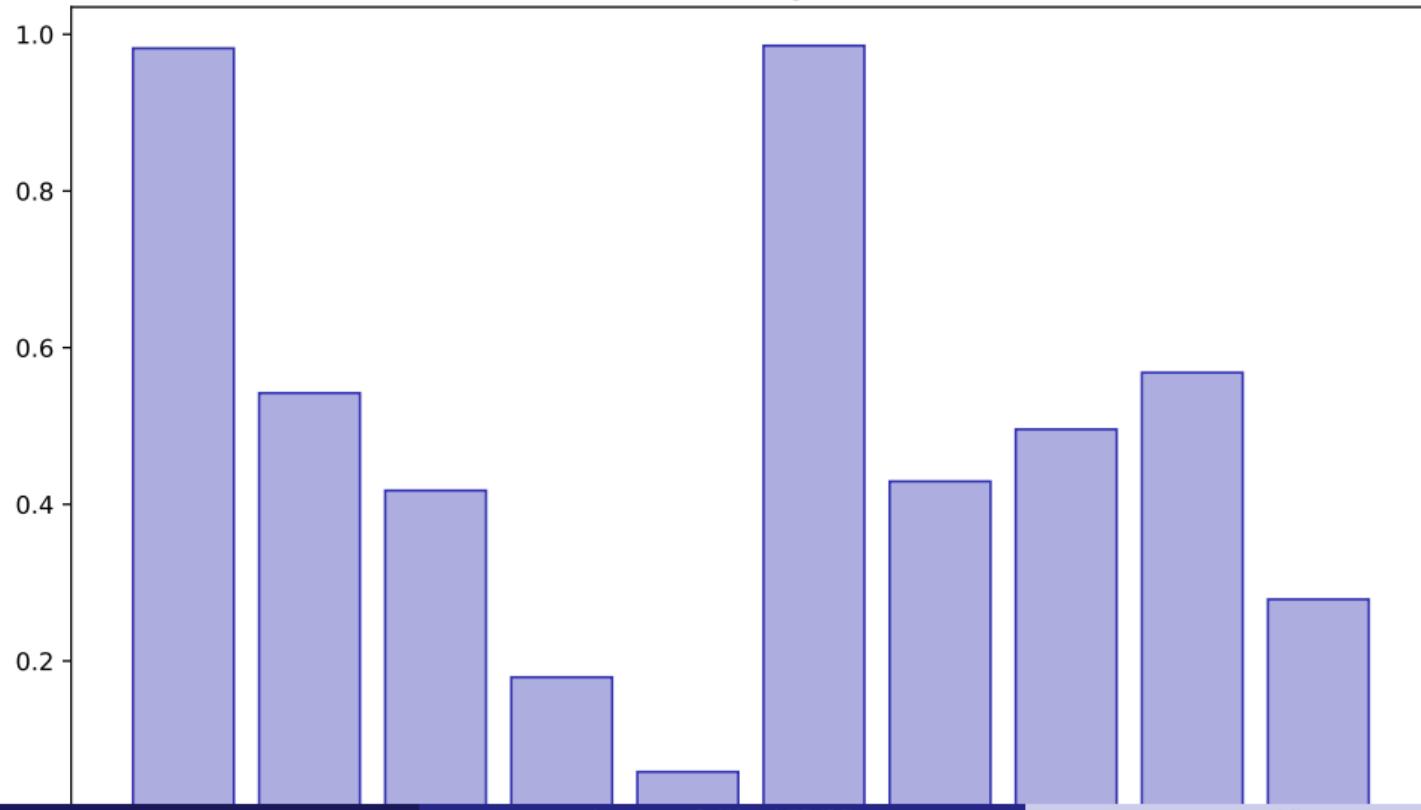


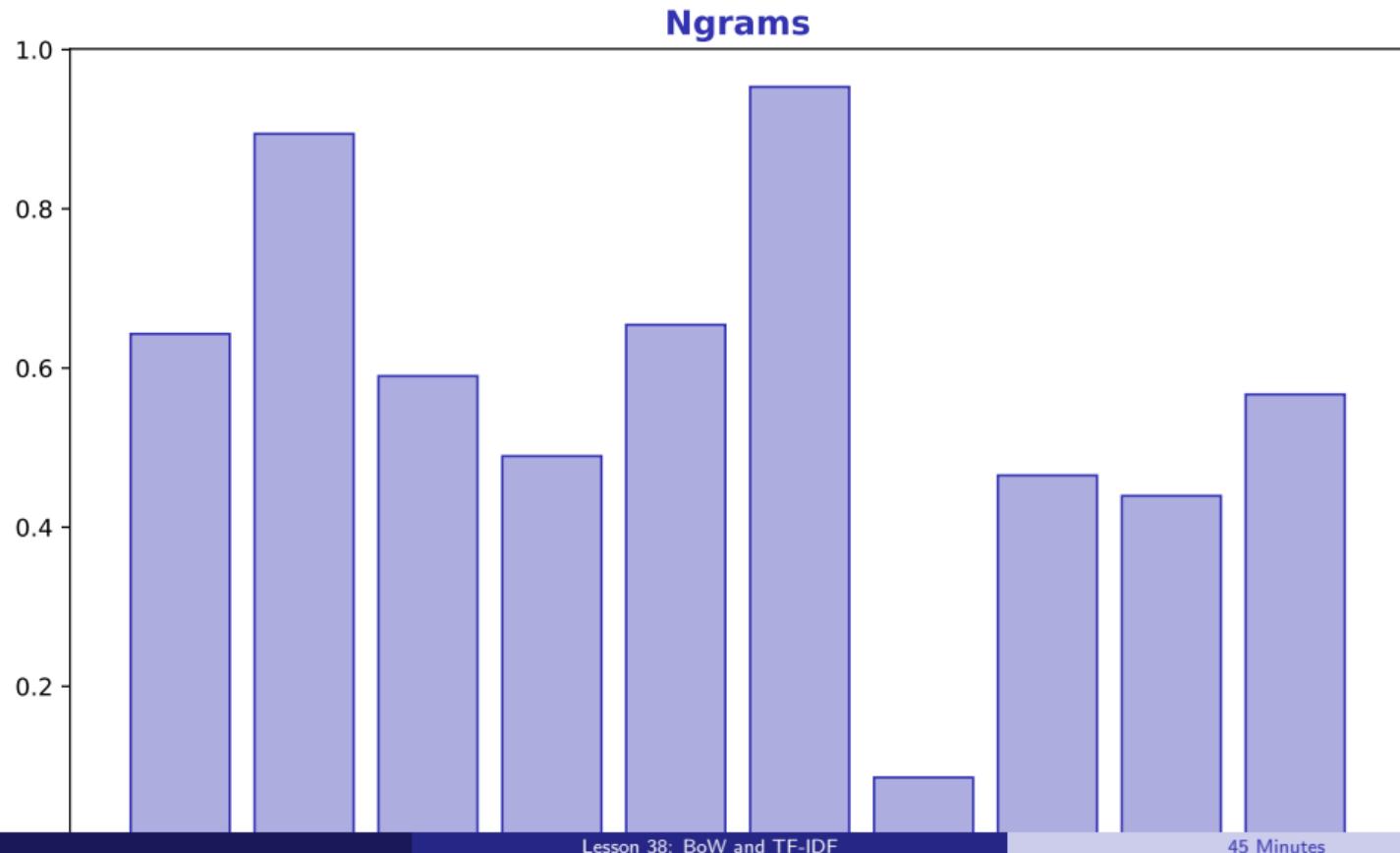
Sparse Matrices



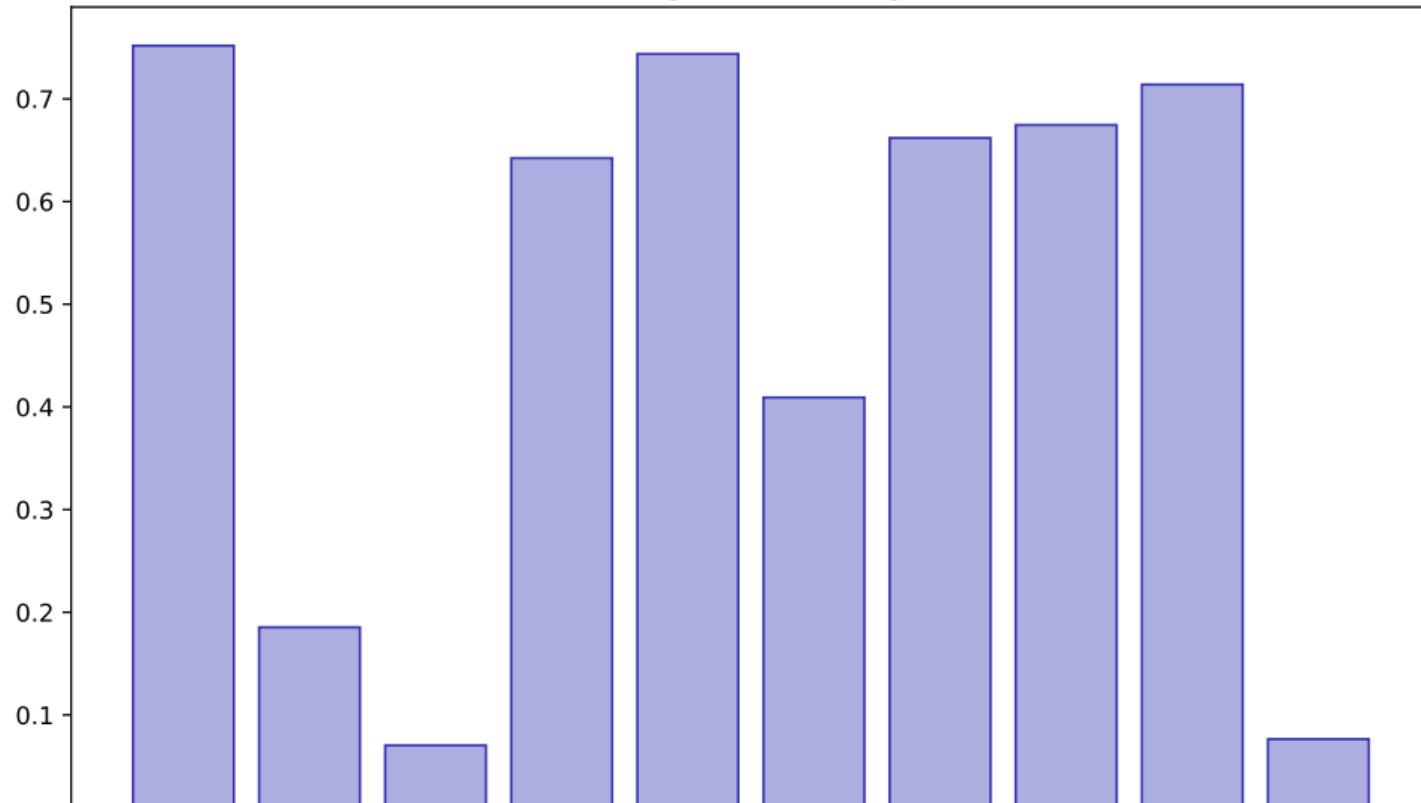
Vocabulary Size

Vocabulary Size





Earnings Call Analysis



Lesson Summary

Key Takeaways:

- Create bag of words
- Calculate TF-IDF weights
- Build document vectors
- Apply to text classification

Apply these skills in your final project