



أكاديمية سدايا  
SDAIA Academy

# TED's Talk using Topic Modelling

Banan Alhethlool

Ghadah Alharbi



أكاديمية سدايا  
SDAIA Academy

# TED's Talk using Topic Modelling

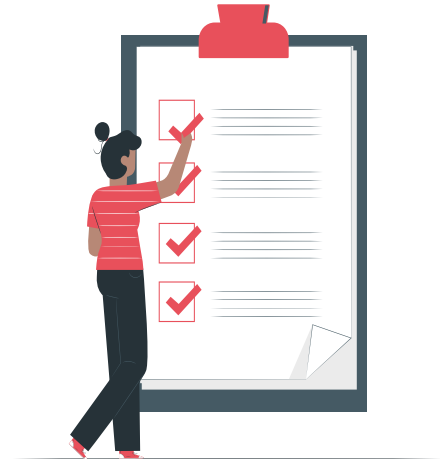
Banan Alhethlool

Ghadah Alharbi

# TABLE OF CONTENTS

---

- Introduction
- Dataset
- Approach & Methodology
- Results



# Introduction

---



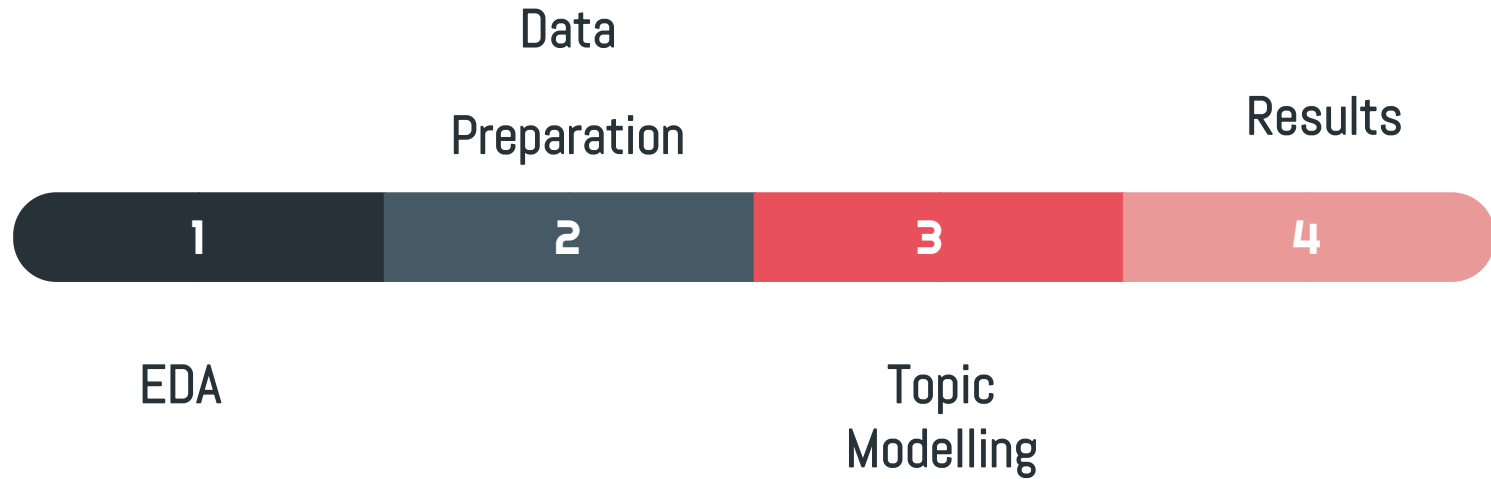
# Dataset

talk_id	title	speaker_1	all_speakers	occupations	about_speakers	views	recorded_date	published_date	event	native_lang	available_lang	comments	duration	topics	related_talks	url	description	transcript	
0	1	Averting the climate crisis	Al Gore	{0: 'Al Gore'}	{0: ['climate advocate']}	Laureate Al Gore focused the world...	3523392	2006-02-25	2006-06-27	TED2006	en	['ar', 'bg', 'cs', 'de', 'el', 'en', 'es', 'fa...	272	977	['alternative energy', 'cars', 'climate change...	{243: 'New thinking on the climate crisis', 54...	<a href="https://www.ted.com/talks/al_gore_averting_the...">https://www.ted.com/talks/al_gore_averting_the...</a>	With the same humor and humanity he exuded in ...	Thank you so much, Chris. And it's truly a gre...
1	92	The best stats you've ever seen	Hans Rosling	{0: 'Hans Rosling'}	{0: ['global health expert, data visionary']}	Rosling's hands, data sings. Glob...	14501685	2006-02-22	2006-06-27	TED2006	en	['ar', 'az', 'bg', 'bn', 'bs', 'cs', 'da', 'de...	628	1190	['Africa', 'Asia', 'Google', 'demo', 'economic...	{2056: 'Own your body's data', 2296: 'A visual...	<a href="https://www.ted.com/talks/hans_rosling_the_bes...">https://www.ted.com/talks/hans_rosling_the_bes...</a>	You've never seen data presented like this. Wi...	About 10 years ago, I took on the task to teach...



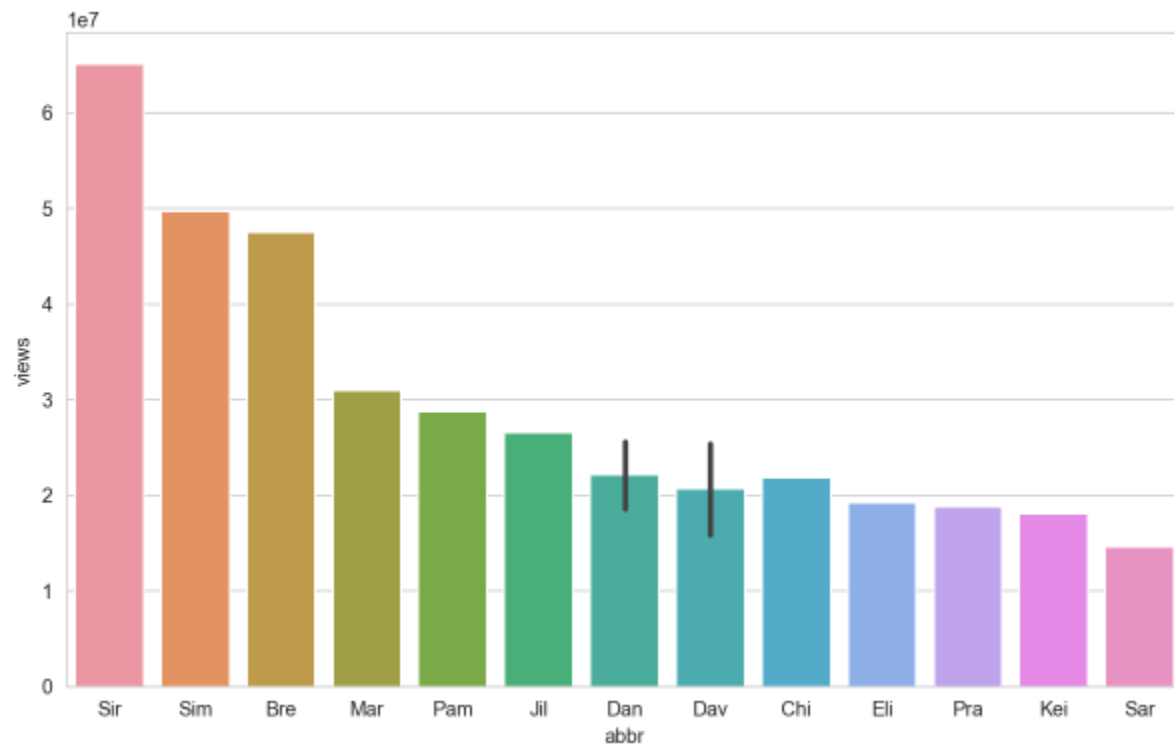
# Approach & Methodology

---



# EDA

---



# Data Preparation

---

- ✓ Handling missing values
- ✓ Handling duplicate values
- ✓ NLTK



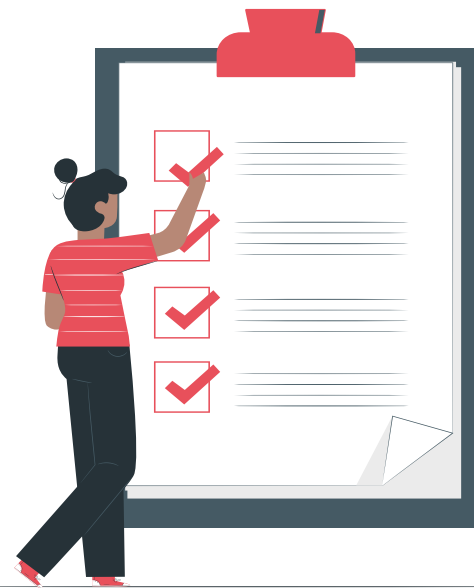


# Topic Modelling

---

TFIDF used for Vectorization:

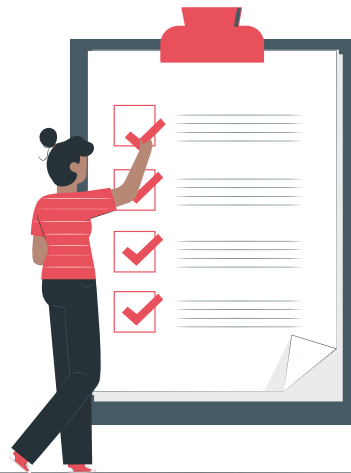
- ✓ **NMF**
- ✓ **LSA**
- ✓ **LDA**



# Topic Modelling

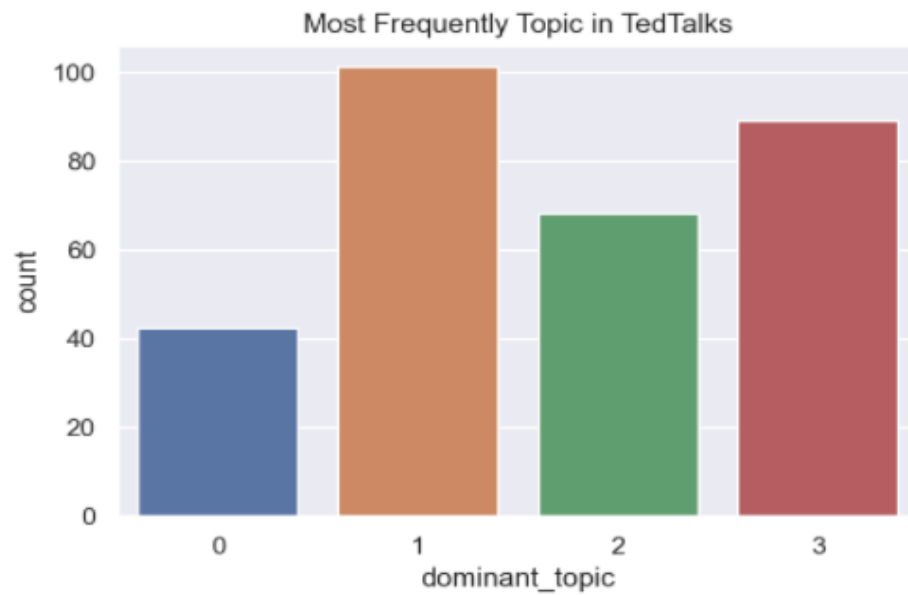
```
{'Geology': 'water, planet, earth, ocean, citi',  
 'stories': 'said, music, love, stori, feel',  
 'Economy': 'countri, africa, percent, dollar, govern',  
 'Computer Scince': 'brain, comput, design, technolog, kind'}
```

	Geology	stories	Economy	Computer Scince	dominant_topic
Doc0	0.032	0.000	0.028	0.175	3
Doc1	0.000	0.238	0.000	0.000	1
Doc2	0.069	0.029	0.000	0.174	3
Doc3	0.004	0.000	0.116	0.056	2
Doc4	0.137	0.041	0.028	0.006	0
...	...	...	...	...	...
Doc693	0.000	0.093	0.000	0.151	3
Doc694	0.072	0.186	0.000	0.051	1
Doc695	0.005	0.113	0.000	0.000	1
Doc696	0.000	0.199	0.060	0.039	1
Doc697	0.077	0.045	0.074	0.000	0



# Results

<b>Geology</b>	water, planet, earth, ocean, citi
<b>stories</b>	said, music, love, stori, feel
<b>Economy</b>	countri, africa, percent, dollar, govern
<b>Computer Science</b>	brain, comput, design, technolog, kind



# Thank you for listening

Banan Alhethloul

Ghadah Alharbi

