

TECNOLOGIE CLOUD E MOBILE

# SWEETEDx Dreams

ALESSIA LAZZARI 1078863



TRELLO, GITHUB,

## RELATED\_VIDEOS - JOB 1

**Il primo Job implementato compone una collezione che includa tutti i video del dataset, associando a ciascuno di essi i video correlati elencati nel documento "related\_videos". Attraverso l'utilizzo di AWS Glue creo due array, "WatchNext\_id" e "WatchNext\_title", che mostrano i risultati richiesti su MongoDB.**

```
# READ AND MANAGE RELATED VIDEOS
related_videos_path= "s3://alessialazzari-data/related_videos.csv"
related_videos_dataset= spark.read.option("header","true").csv(related_videos_path)
related_videos_dataset=related_videos_dataset.dropDuplicates()

# ADD RELATED VIDEOS per ID TO AGGREGATE MODEL
related_videos_dataset_agg = related_videos_dataset.groupBy(col("id").alias("id_watch_next")).agg(collect_list("related_id").alias
("WatchNext_id"), collect_list("title").alias("WatchNext_title"))

tedx_dataset_agg = tedx_dataset_agg.join (related_videos_dataset_agg,tedx_dataset_agg._id == related_videos_dataset_agg.id_watch_next,"left"
).drop("id_watch_next")
```

## Esempio di risultato ottenuto su MongoDB nella collezione tedx-data.

```
_id: "523970"
id_related: "523970"
slug: "jen_walsh_can_cannabis_help_you_sleep_here_s_the_science"
speakers: "Jen Walsh"
title: "Can cannabis help you sleep? Here's the science"
url: "https://www.ted.com/talks/jen_walsh_can_cannabis_help_you_sleep_here_s..."
description: "Can cannabis actually treat insomnia? It's complicated, says sleep phy..."
duration: "690"
publishedAt: "2024-04-10T14:43:46Z"
image_url: "https://talkstar-photos.s3.amazonaws.com/uploads/3224b0c8-7111-4ac0-b9..."
▶ id_related_list: Array (6)
▶ title_related_list: Array (6)
▶ tags: Array (7)
```

Compare ora la lista degli ID correlati, i titoli della lista dei video correlati e i tag relativi.

```
▼ title_related_list: Array (6)
  0: "Can you change your sleep schedule?"
  1: "The brain science (and benefits) of ASMR"
  2: "Is marijuana bad for your brain?"
  3: "The brain benefits of deep sleep -- and how to get more of it"
  4: "Why do we sleep?"
  5: "What causes insomnia?"
▼ tags: Array (7)
  0: "science"
  1: "health"
  2: "brain"
  3: "humanity"
  4: "mental health"
```

# FUNZIONALITÀ DEL SERVIZIO

**Per lo svolgimento della seconda richiesta, è necessario ricordare le funzionalità dichiarate in precedenza:**

- 1. Raccomandazioni Personalizzate di Contenuti**
- 2. Modalità Notturna**
- 3. Timer di Spegnimento Automatico**
- 4. Raccolta di Feedback e Valutazioni**
- 5. Creazione di Playlist Personalizzate**
- 6. Integrazione con Dispositivi e App di Monitoraggio del Sonno**
- 7. Consigli e Risorse per il Benessere Notturno**

## **Funzionalità scelta**

**01**

**Raccomandazioni  
Personalizzate di  
Contenuti**

# FILTRAGGIO DATI - JOB 2

La seconda funzionalità implementata richiede la creazione di un job per ridurre le dimensioni del dataset.

L'obiettivo è filtrare e selezionare solo i video relativi al sonno e al benessere, consentendo così la visione di contenuti che incentivino il rilassamento serale per tutti gli utenti interessati.

Collection Name	Documents
<a href="#">tedx-data</a>	6363
<a href="#">tedx-versione sonno</a>	45

**SWEETEDx Dreams prevede la personalizzazione dei contenuti relativi al rilassamento serale, offrendo agli utenti una selezione di video mirati per favorire il relax e migliorare la qualità del sonno.**

# FILTRAGGIO DATI

```
# FILTRAGGIO PAROLE SLEEP RELATED
filtered_dataset = tedx_dataset_agg

# FINAL FILTER BASED ON SLEEP-RELATED KEYWORDS
filtered_dataset = tedx_dataset_agg.filter(size(array_intersect(col("tags"),
array(lit("sleep"), lit("dream"), lit("meditation"), lit("bedtime"), lit("insomnia"), lit("melatonin"), lit("REM"), lit("ASMR"), lit("snore"),
    lit("nap"), lit("bed"), lit("bed"), lit("night"))))) > 0)

# Stampa lo schema del dataset filtrato
filtered_dataset.printSchema()
```

**Creo un dataset e applico un filtro per selezionare solo le righe che contengono almeno una delle parole specificate nei "tags".**

**La funzione “array\_intersect()” restituisce un array contenente gli elementi comuni tra la colonna "tags" e l'array di parole specificate nel codice.**

# CRITICITÀ ED EVOLUZIONI

## DATASET LIMITATO

I contenuti ottenuti, una volta implementato il filtro, sono risultati di quantità minore rispetto alle aspettative di partenza.

## AGGIUNTA PAROLE CHIAVE

Allargare il campo di parole chiave potrebbe aumentare la quantità di video utili per gli utenti interessati al servizio.

## FILTRAGGIO DEI CONTENUTI

I contenuti selezionati potrebbero non essere pertinenti alle funzionalità e agli obiettivi che SWEETEDx Dreams si propone di implementare.

## FEEDBACK E VALUTAZIONI

SWEETEDx Dreams dichiara l'implementazione di feedback e valutazioni da parte degli utenti che possono favorire il filtraggio dei video più interessanti.