

A03 - SPOTIFY

Analiza popularności pisanej twórczości internetowej

Zuzanna Mróz, Kacper Staroń, Jakub Szypuła

JAK ZARABIAĆ NA MUZYCE?

- Ważnym elementem budowania marki jest nie tylko ogólna popularność, ale też liczba oddanych fanów.
- Niestety analiza tych drugich nie jest taka prosta i trudna do skwantyfikowania ogólniejszymi metodami.
- Rozwiązanie: analiza twórczości.

CEL BIZNESOWY

- Rozwiązanie big data pozwalające pobierać, przechowywać, przetwarzać i analizować dane dotyczące popularności twórczości internetowej.
- Skierowane na informacje o konkretnych zespołach.

ŹRÓDŁA



1. Spotify – popularny serwis do streamingu muzyki

2. A03 – serwis na którym można zamieszczać autorskie opowiadania (tzw. Fanfiction)



ARCHITEKTURA ROZWIĄZANIA

TABELE - SPOTIFY

Nazwa kolumny	Opis	Typ	Format
BandName	Nazwa zespołu	String	tekst
Followers	Liczba osób która podąża za danym artystą na Spotify	Int	liczba całkowita
Popularity	Wskaźnik popularności wg spotify (od 1 do 100)	Int	liczba całkowita

Nazwa kolumny	Opis	Typ	Format
BandName	Nazwa zespołu	String	tekst
Genre	Nazwa gatunku muzycznego	String	tekst

Nazwa kolumny	Opis	Typ	Format
BandName	Nazwa zespołu	String	tekst
AlbumName	Tytuł albumu	String	tekst
ReleaseDate	Data wydania albumu	Date	YYYY-MM-DD

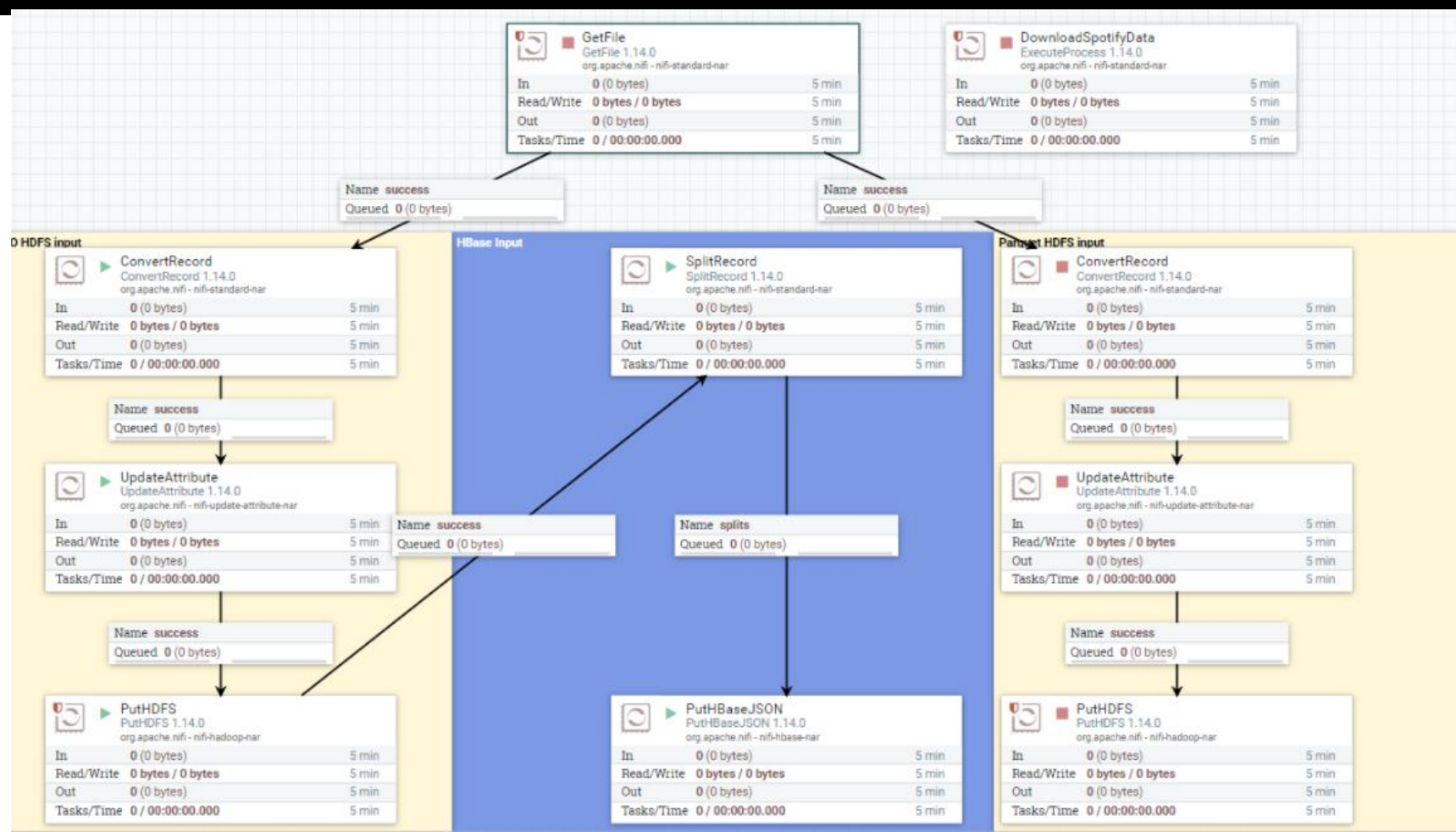
TABELE - TEKSTY

Nazwa kolumny	Opis	Typ	Format
id	ID tekstu	int	liczba całkowita
date_updated	Data zamieszczenia tekstu w serwisie	datetime	YYYY-MM-DD hh:mm:ss
bookmarks	Liczba “zakładek” odnoszących się do tekstu	int	liczba całkowita
nchapters	Liczba rozdziałów	int	liczba całkowita
complete	Informacja czy tekst został zakończony, czy nie	Bool	True/False
comments	Liczba komentarzy	int	liczba całkowita
hits	Liczba wyświetleń	int	liczba całkowita
kudos	Liczba “polubień”	int	liczba całkowita
language	Język w jakim napisano tekst	String	tekst
rating	Klasyfikacja wiekowa	String	tekst
status	Status tekstu	String	tekst
words	Liczba słów	int	Liczba całkowita

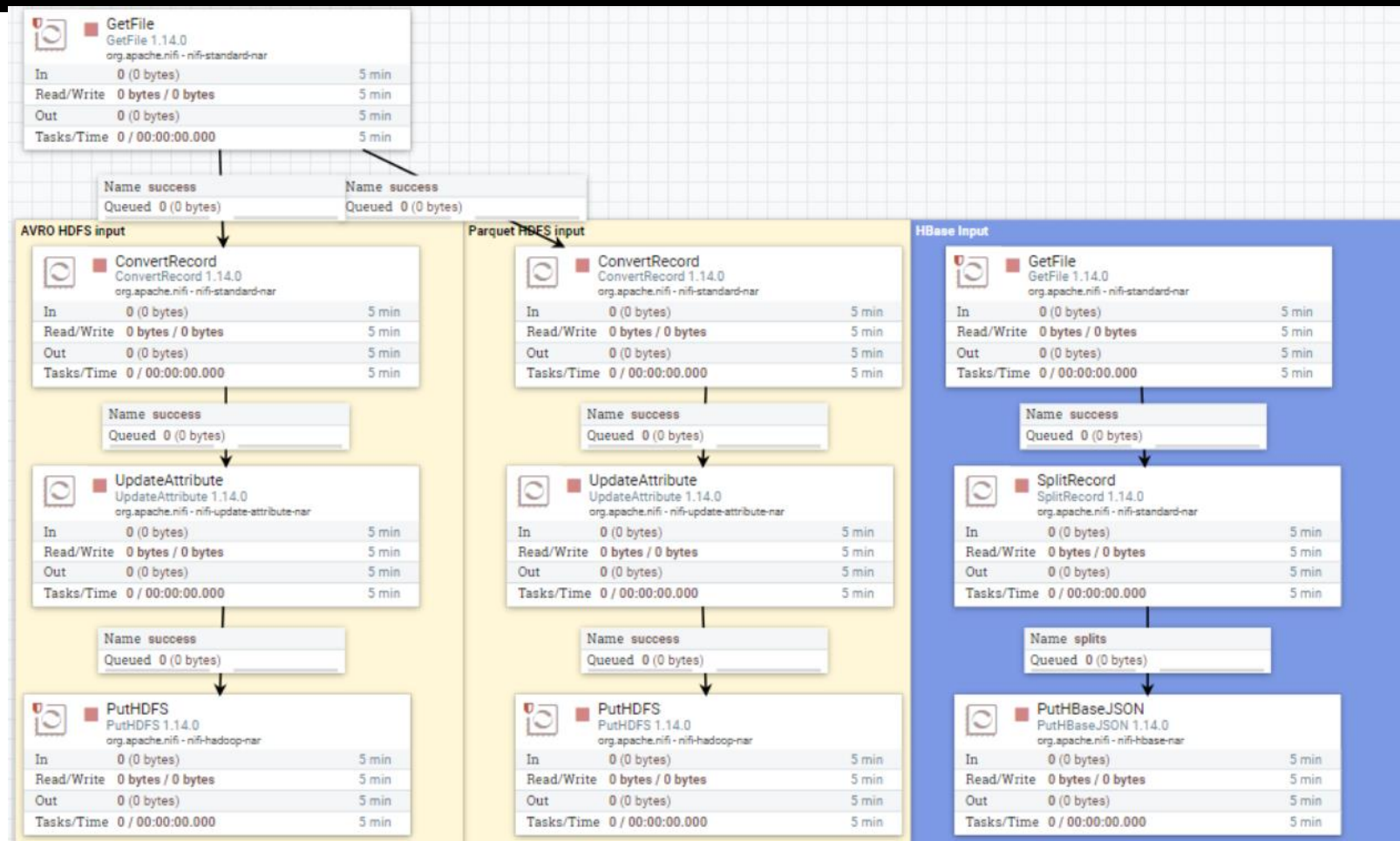
TABELE - TAGI

Nazwa kolumny	Opis	Typ	Format
id	ID tekstu	int	liczba całkowita
meta_name	Typ metadanych zawartych w wierszu	String	tekst ze zbioru: authors, categories, characters, fandoms, relationships, series, tags, warnings
meta_val	Wartość metadanej	String	tekst

NIFI - SPOTIFY



NIFI - A03



HBASE

```
connection.create_table(  
    'fic_meta',  
    {  
        '_id': dict(), #id, title, summary, language  
        '_status': dict(), #date_updated, status, n_chapters, complete, words  
        '_rating': dict() #rating, kudos, hits, comments, bookmarks  
    }  
)  
  
connection.create_table(  
    'tag_meta',  
    {  
        'tag_meta': dict()  
    }  
)
```

ANALIZY SPARKOWE

```
logger.info("3. Count")
print('fic count', df_fic.count())
print('tag count', df_tag.count())
print('spotify band count', band.count())

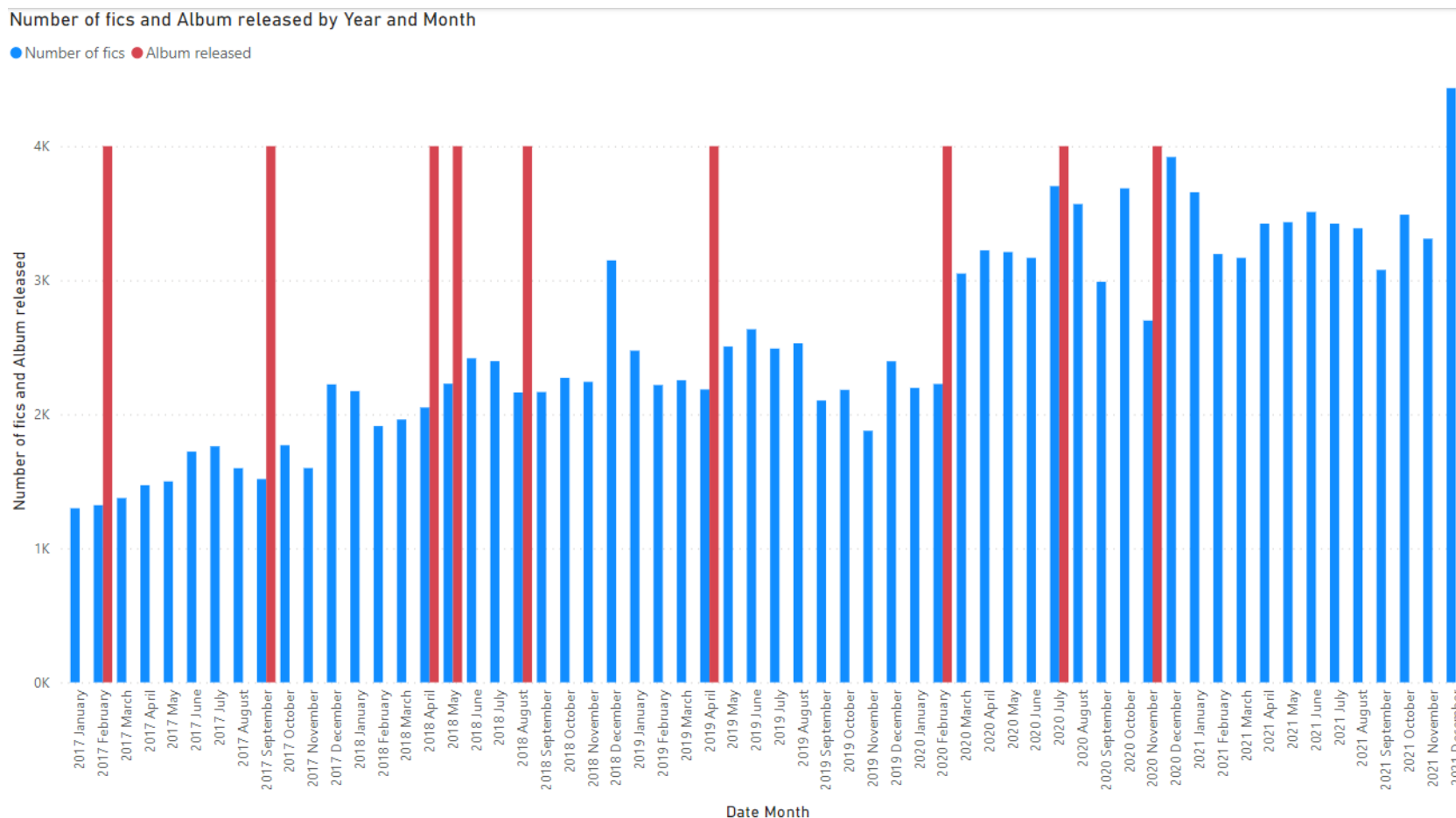
logger.info("4. Tag names")
df_tag.select("meta_name").distinct().show()

logger.info("5. Top 10 Fandoms")
join1 = df_fic.join(df_tag.filter(f.col('meta_name') == 'fandoms'))
join1 = join1.distinct().groupBy('meta_val').agg({'count': sum('count')})
join2 = join1.withColumn("band", f.col("meta_val"))
join2.show()
join2.createOrReplaceTempView('Top10Fandoms')

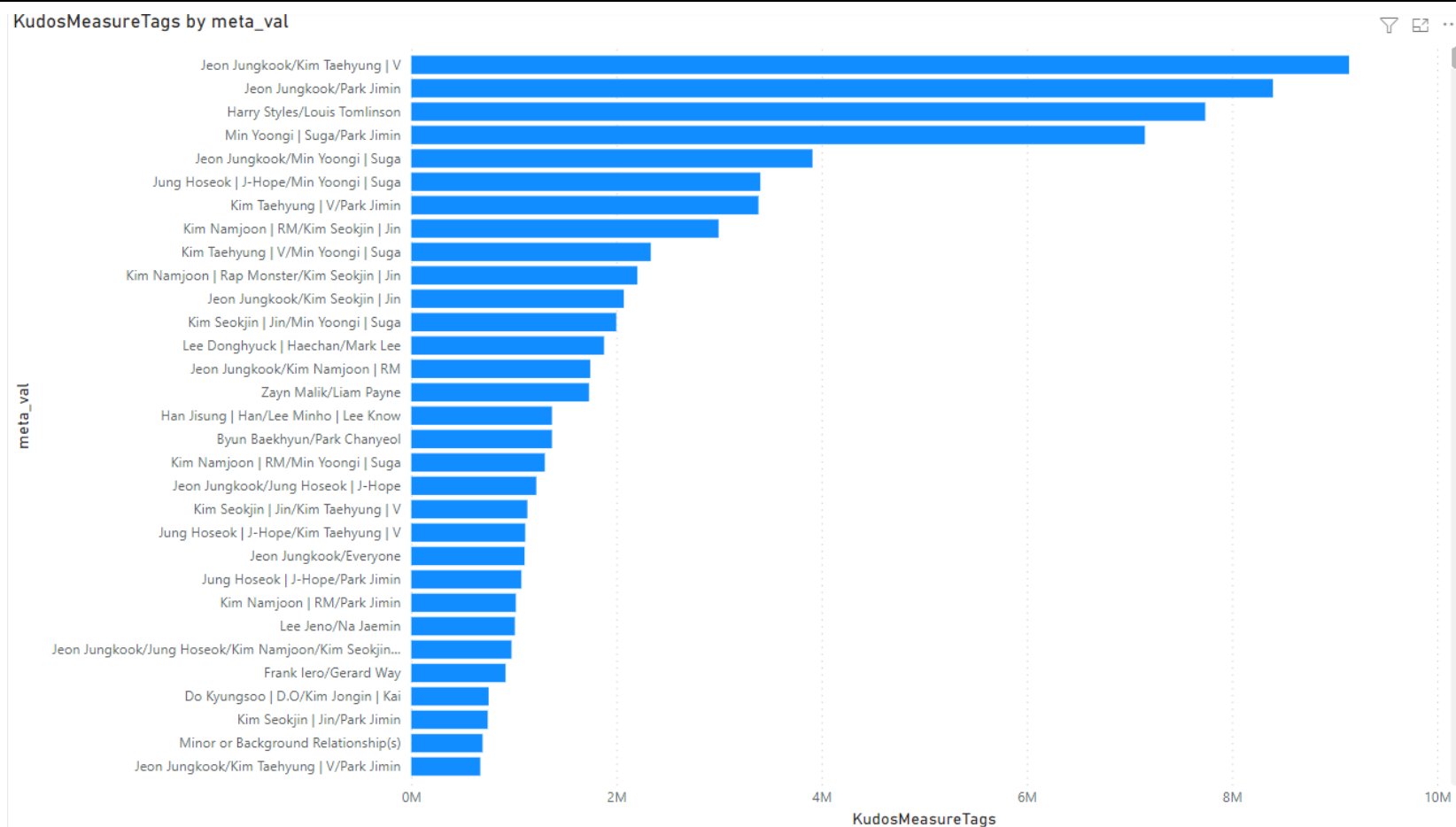
logger.info("6. Top Ships")
w = Window.partitionBy('relationship')
# sprawdzanie w jakim fandomie najczęściej znajdują się relacje
join3 = df_tag.filter(f.col('meta_name') == 'fandoms')
join3 = join3.groupby("fandom", "relationship").count()
join3 = join3.withColumn('maxF', f.max('count').over(w))
# liczenie statystyk dla danego shipu i doklejanie
join4 = df_fic.join(df_tag.filter(f.col('meta_name') == 'fandoms'))
join4 = join4.distinct().withColumn("relationship", f.col("meta_val"))
join4 = join4.join(join3, on='relationship', how='left')
join4.show(truncate=False)
join4.createOrReplaceTempView('TopShips')
```

- bezpośredni dostęp do pobranych tabel
- zliczenie liczby wierszy w poszczególnych tabelach
- lista unikalnych tagów
- top 10 najpopularniejszych zespołów po ilości tekstów
- najczęstsze relacje między członkami zespołu w tekstach na podstawie liczby tekstów
- liczba tekstów dla zespołu BTS dla każdego miesiąca wg daty wydania
- kiedy wychodziły poszczególne albumy dla danych zespołów wg miesięcy

PRZYKŁADOWE WIZUALIZACJE

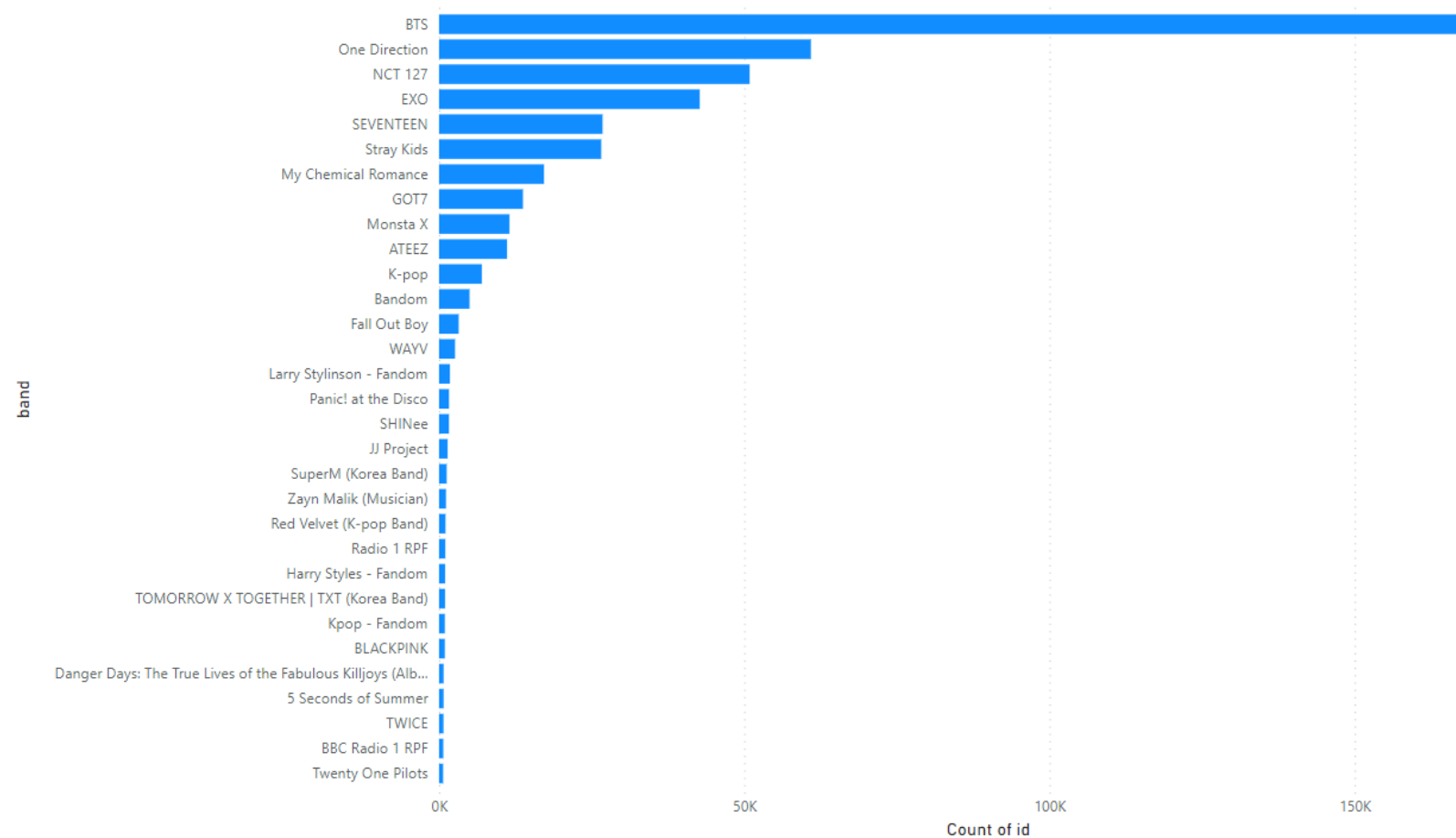


PRZYKŁADOWE WIZUALIZACJE



PRZYKŁADOWE WIZUALIZACJE

Count of id by band



DZIĘKUJEMY ZA UWAGĘ