

Relatório de Sistemas de Informação Analíticos
Sistema Integrado de Data Warehousing para Análise de Filmes e
Séries Mais Vistos na Netflix



2023/2024

Unidade Curricular: Sistemas de Informação Analíticos
Docente: Professor Carlos Miguel Francisco

Grupo:

Karina Kozyar n^a120160
Leonor Santana n^o120178
Matilde Alves n^o 120172

Data de entrega 1: 19 de Abril 2024
Data de entrega 2: 24 de Maio 2024

Índice

Objetivo:	3
Descrição da empresa e do negócio:	3
1. Recolha de Dados:	4
2. Modelação de Dados:	5

Objetivo:

Desenvolver e implementar um conjunto de dados robusto que consolide informações sobre filmes e programas de TV disponíveis na Netflix. Este sistema permitirá análises detalhadas sobre o catálogo da Netflix, incluindo categorias, países onde está disponível, classificações e outros dados relevantes, facilitando a compreensão das preferências dos espectadores e o potencial de sucesso de diferentes tipos de conteúdo.

Descrição da empresa e do negócio:

A Netflix é uma empresa inovadora especializada em análise e consultoria na indústria do entretenimento em streaming, dedicada a ajudar produtores e distribuidores a compreender e otimizar o desempenho das suas séries através de análises de dados avançadas. Fundada num momento em que o streaming de conteúdo se tornou uma forma dominante de consumo de entretenimento, a Netflix posiciona-se na interseção entre tecnologia, análise de dados e criação de conteúdo.

A missão da Netflix é capacitar produtores e distribuidores com dados e insights precisos para promover decisões informadas na criação e distribuição de séries. Através da análise de grandes volumes de dados de visualização, feedback dos espectadores e tendências de mercado, a empresa visa facilitar a produção de conteúdo de alta qualidade que seja relevante para o público-alvo e impulse o sucesso comercial.

Os serviços oferecidos incluem análises de audiência avançadas, consultoria de conteúdo, previsão de popularidade de séries e análises comparativas de desempenho.

A Netflix está comprometida com a inovação tecnológica. A empresa desenvolveu uma plataforma de armazenamento de dados proprietária que integra dados de diversas fontes, incluindo análises de audiência em tempo real, dados demográficos dos espectadores e informações de produção das séries. Esta plataforma permite análises complexas e a geração de modelos preditivos que oferecem uma visão clara das tendências de visualização e preferências do público.

Os clientes da Netflix variam desde pequenos estúdios independentes até grandes produtoras e distribuidoras de conteúdo, todas tentando alcançar o máximo de sucesso com as suas séries na plataforma da Netflix e outras plataformas de streaming.

A Netflix destaca-se como um parceiro essencial para empresas que procuram não apenas criar conteúdo de qualidade, mas também entender e atender às necessidades do público-alvo, impulsionando assim o sucesso comercial e artístico das suas séries.

Etapas:

1. Recolha de Dados:

O conjunto de dados "Filmes e Séries da Netflix" disponível no Kaggle foi compilado por Diego Enrique, um utilizador da plataforma. Este conjunto de dados contém informações sobre filmes e programas de TV disponíveis na Netflix, incluindo detalhes como título, tipo (filme ou programa de TV), realizador, elenco, país de produção, data de lançamento na Netflix, classificação, duração, entre outros em dois datasets designados de "titles" e "credits".

Data Explorer

6.08 MB

credits.csv

titles.csv

Figura 1 - Datasets

Quanto à fonte dos dados específicos, é importante salientar que o Kaggle é uma plataforma que permite aos utilizadores partilharem conjuntos de dados de diversas origens. Assim, o autor do conjunto de dados provavelmente reuniu informações de várias fontes, como APIs públicas, sites da Netflix, bases de dados de filmes e séries, entre outros.

Para obter análises de visualização em tempo real, feedback dos espectadores em redes sociais e informações de produção das séries disponíveis na Netflix, pode ser necessário consultar diversas fontes de dados, como APIs de redes sociais (por exemplo, Twitter, Facebook, Reddit), sites de análises de mídia (por exemplo, IMDb, Rotten Tomatoes), bancos de dados de produção cinematográfica, entre outros.

Em resumo, reunir esses dados relevantes para análises de desempenho e preferências do público pode envolver uma abordagem multifacetada que inclui a consulta a várias fontes de dados e a integração de diferentes conjuntos de informações.

2. Modelação de Dados:

Desenvolver um modelo de dados dimensional que facilite análises rápidas e eficientes, com foco nos indicadores de popularidade, preferências de conteúdo e fatores de sucesso das séries. Este modelo permitirá uma compreensão mais profunda do comportamento dos espectadores e das suas tendências de streaming na Netflix.

Para desenvolver um modelo de dados dimensional que facilite análises rápidas e eficientes, foram criadas tabelas de factos e de dimensão para o esquema Entidade-Relação em floco de neve:

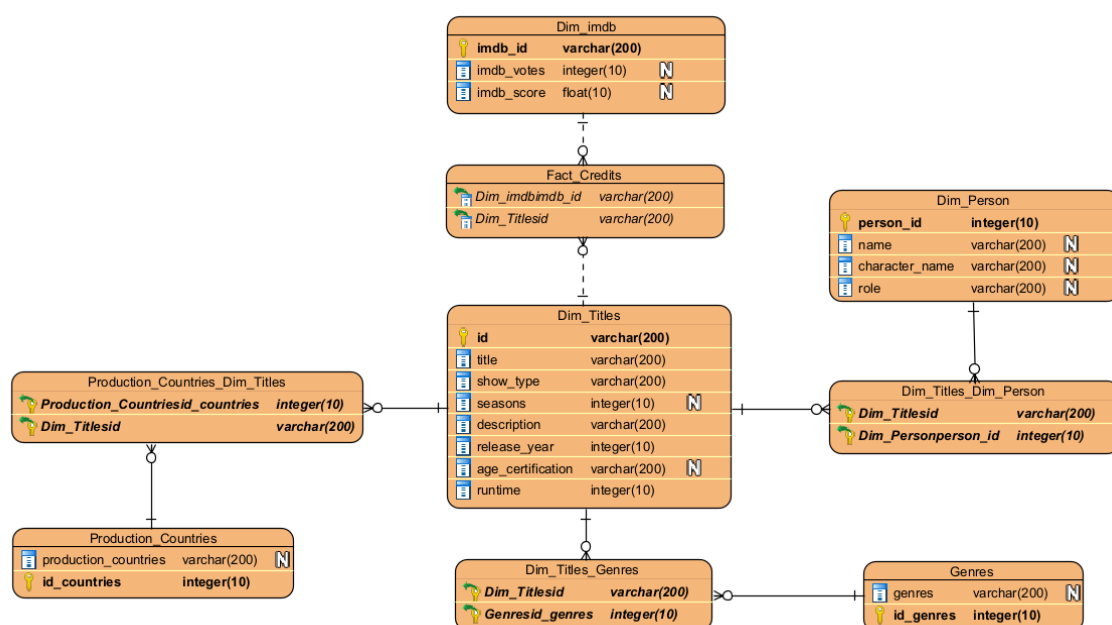


Figura 2 – Modelo ER

A tabela de factos contém duas chaves estrangeiras, sendo estas a chave primária id da tabela dimensão IMDb e a chave primária da tabela dimensão Títulos. A tabela dimensão IMDb, além do código identificador, existem dados da quantidade de votos em determinado filme ou série e a pontuação final realizada. A tabela dimensão Títulos contém dados como: título, tipo (filme/série), número de temporadas (que pode não existir caso seja filme), descrição, data de emissão, certificado de idade recomendada e duração em minutos. As tabelas dimensão dos países produtores, géneros e pessoas, apresentam uma relação de muitos para muitos com a tabela referente aos títulos, assim, foram criadas terceiras tabelas para cada uma destas relações.