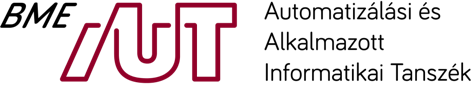
**Házi feladat specifikáció**

Üzleti Intelligencia

Félév: 2023 ősz

**Online videótéka, a Netflix**

**Pethes Dániel Imre (EAQXCB)**

**pethesdani.imre@gmail.com** 

# **Bemutatás**

A projekt során a Netflix cipőjébe bújva, historikus adatokból egy olyan üzleti intelligencia rendszert készítek mely segítségével betekintést nyerhetek a platform teljesítményébe és statisztikákat kaphatok az egész platform és a rajta található tartalmak teljesítményéről. Ezek alapján a jövőbeli tartalmak megválasztásához is próbálok kimutatásokat készíteni.

# Főbb funkciók

* A feladat megoldásához számos forrásból tervezek adatokat nyerni, köztük:
  + Kaggle.com (pl. <https://www.kaggle.com/datasets/imtkaggleteam/netflix> vagy <https://www.kaggle.com/datasets/adnananam/netflix-revenue-and-usage-statistics>, illetve szükség esetén további datasetek)
  + Google Trends (pl. Netflix, esetleges competitorok, akár egyes sorozatok)
  + IMDB, Rotten Tomatoes vagy hasonló rating oldal datasetje
* Az alkalmazás architektúráját Kimball architektúra szerint tervezem megvalósítani, minden adatforrásnak külön staging táblával, illetve egy Data Warehouse-sal
* Megvalósítandó ETL/ELT job-ok.
  + A dataset CSV fájljainak beolvasása
  + A CSV fájlok betöltése a staging area-ba
  + A staging areákról átalakítás és átvitel a DW-ba
* Az adatok végső helyéül egy relációs adatbáziskezelőt tervezek használni.
* Megjelenítési réteg
  + Platform népszerűsége
  + Legjobban teljesítő sorozatok
  + Egyes évek üzleti teljesítménye és az abban az évben felkerült tartalmak
  + Legjövedelmezőbb tartalom fajták
  + Más szolgáltatók népszerűségé, összehasonlítás
  + Egyéb lehetséges nézetek
* Szükség esetén alap predikciók, pl. korrelációk
* Alap ETL tesztek

# Technológiák

* Az adatbázishoz Microsoft SQL Server terveztem használni.
* ETL motorként az ehhez illeszkedő, SQL Server Integration Services-t terveztem használni.
* Az esetleges szükséges scripteket ennek a Visual Studio projektjében terveztem elhelyezni
* Az adatvizualizációhoz Microsoft Power BI-t vagy Tableau-t tervezek használni
* Szükség esetén az egyes nézetekhez R nyelvet tervezek használni.