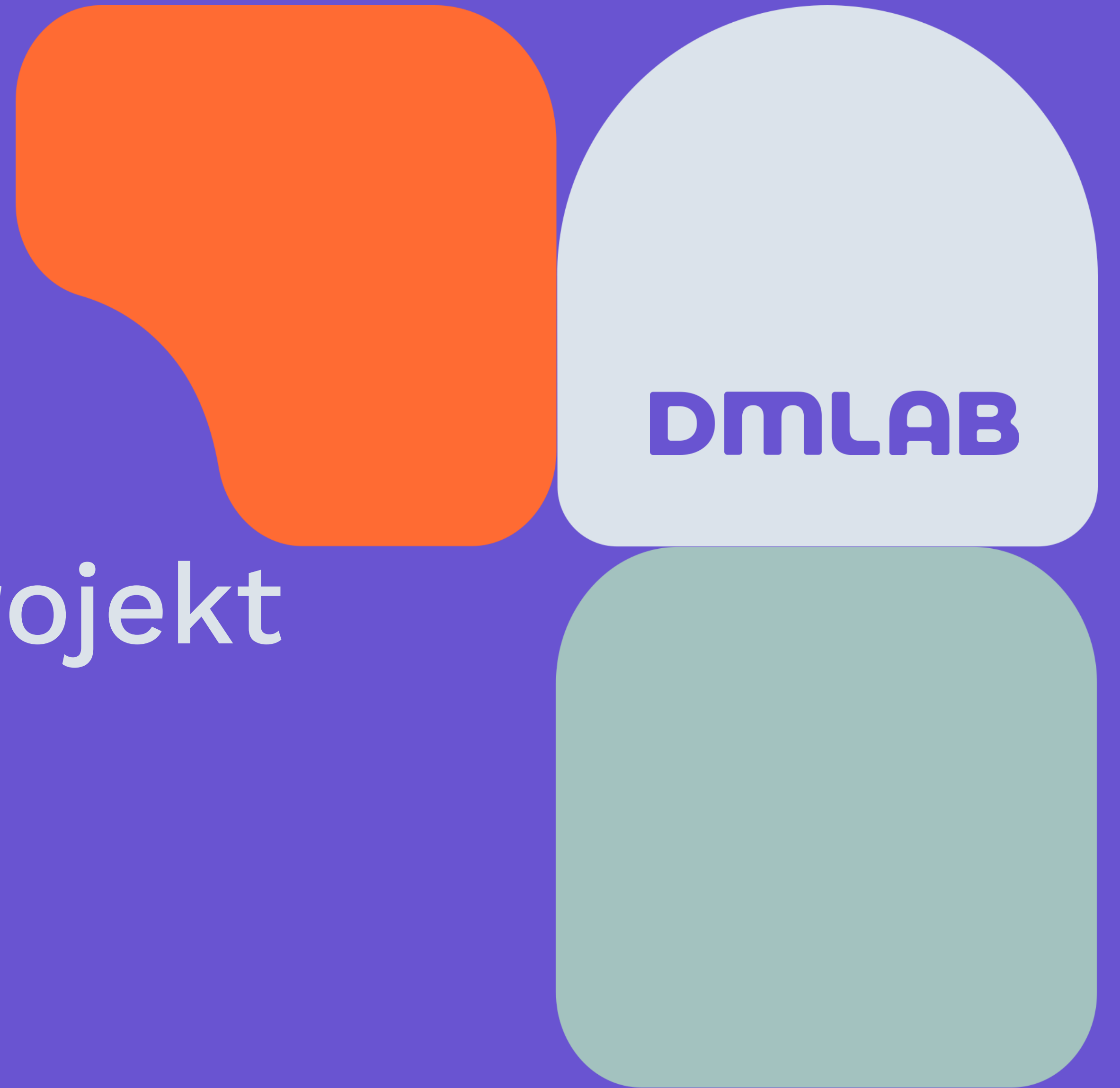
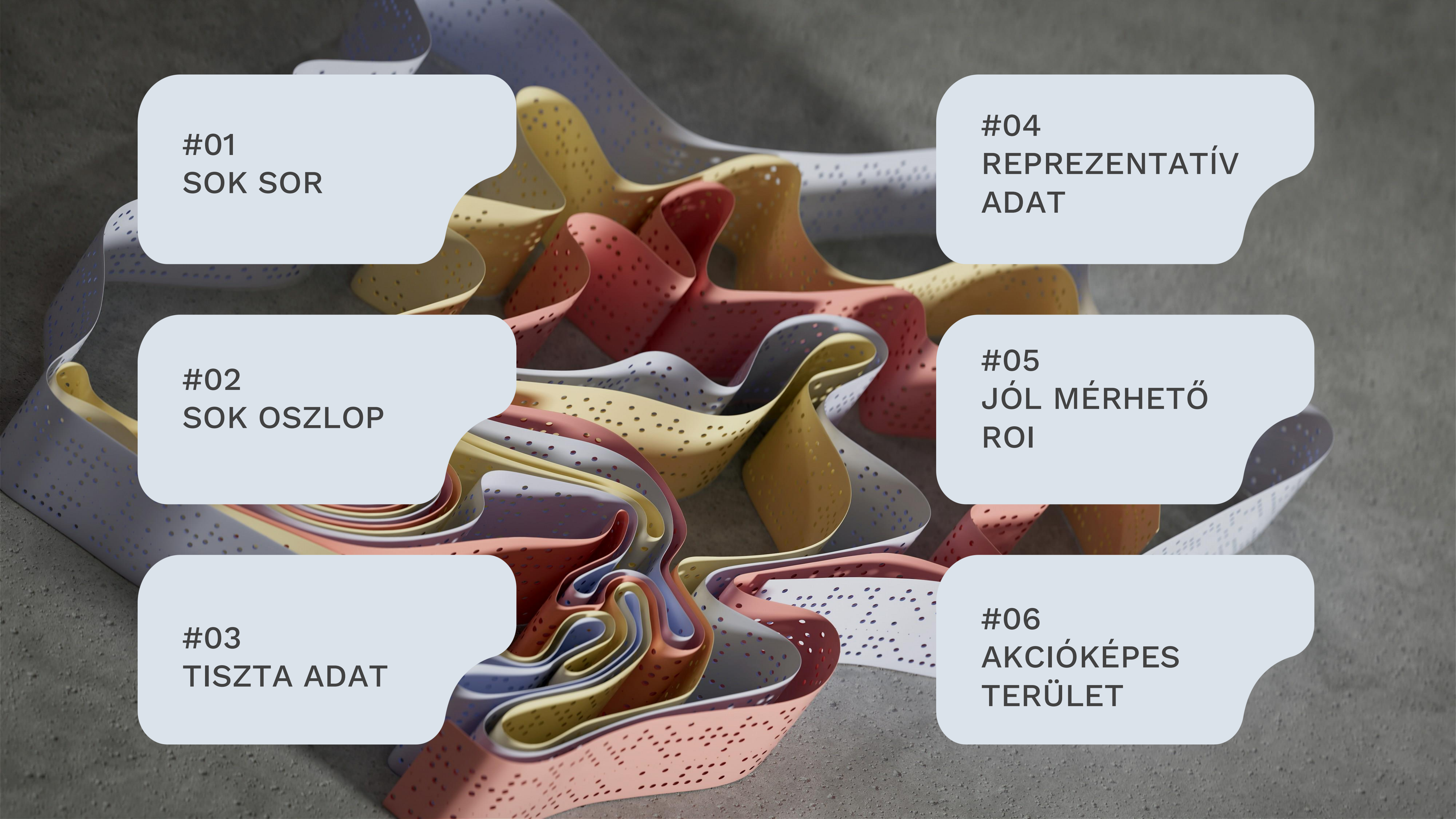




Egy data science projekt lépésről lépésre

Data Science Klub

Mitől **sikeres** egy
data science
projekt?



#01
SOK SOR

#02
SOK OSZLOP

#03
TISZTA ADAT

#04
REPREZENTATÍV
ADAT

#05
JÓL MÉRHETŐ
ROI

#06
AKCIÓKÉPES
TERÜLET

Hogyan
kezdünk bele?

model.fit()



ÜZLET MEGÉRTÉSE

1

Üzleti célok

Mit akarunk elérni üzletileg?
Mik azok, amik befolyásolják a projekt kimenetelét?
Jó válaszok a rossz kérdésre...

Helyzetértékelés

Erőforrások, korlátok, feltételek és egyéb tényezők

Adatelemzési célok

Üzleti cél üzleti terminológiával
Adatelemzési cél technológiai nyelven
Üzleti sikerkritériumok vs. hatékonysági kiértékelés

01 Üzleti probléma

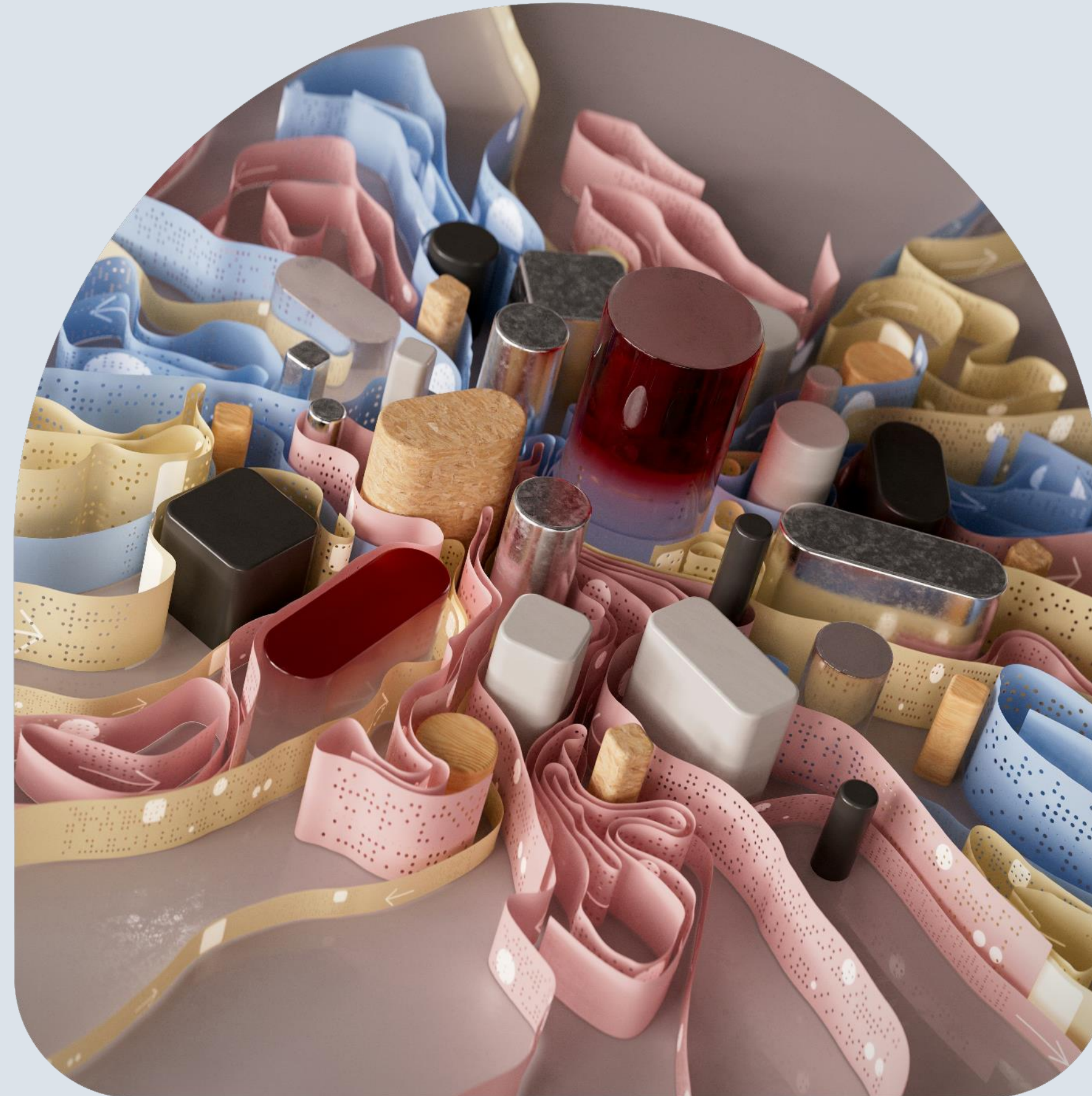
Félévente az ügyfelek 1%-a elvándorol.
Mondd meg ki fog elvándorolni?

02 Üzleti probléma

Mondd meg, hogy ki mekkora valószínűséggel vándorol el.

03 Üzleti probléma

Mennyi embert keressek meg? Hol van ennek az optimuma?



01 Megoldás

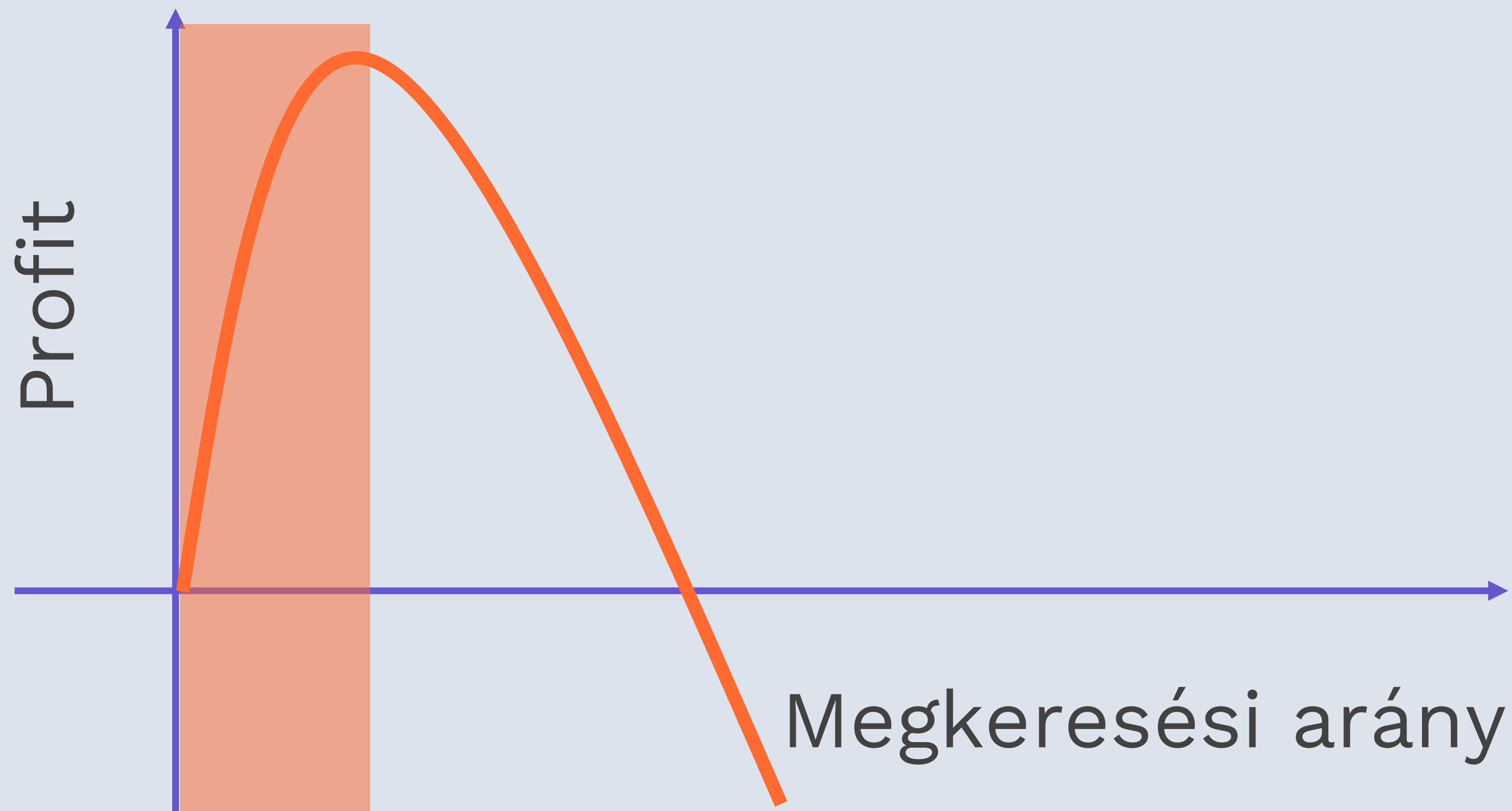
Minden felhasználóra: Igen/Nem.
Modell pontossága: 99%.
Mindenkire azt mondjuk: Nem

02 Megoldás

Sorrendezés. Elöl a legveszélyeztetettebbek

03 Megoldás

Profitgörbe





ADATOK MEGÉRTÉSE

2

Adatgyűjtés

Adatforrások és adatgazdák
Adatok megértése üzleti
interjúkon
Kezdeti adatelőkészítés
Integráció több adatforrás esetén

Adatok leírása és feltárása

Riport az adatokról
Alapeloszlások
Összefüggések az adatokban

Adatminőség

Hiányzó értékek
Kiugró értékek
Hibás értékek



Kiválasztás

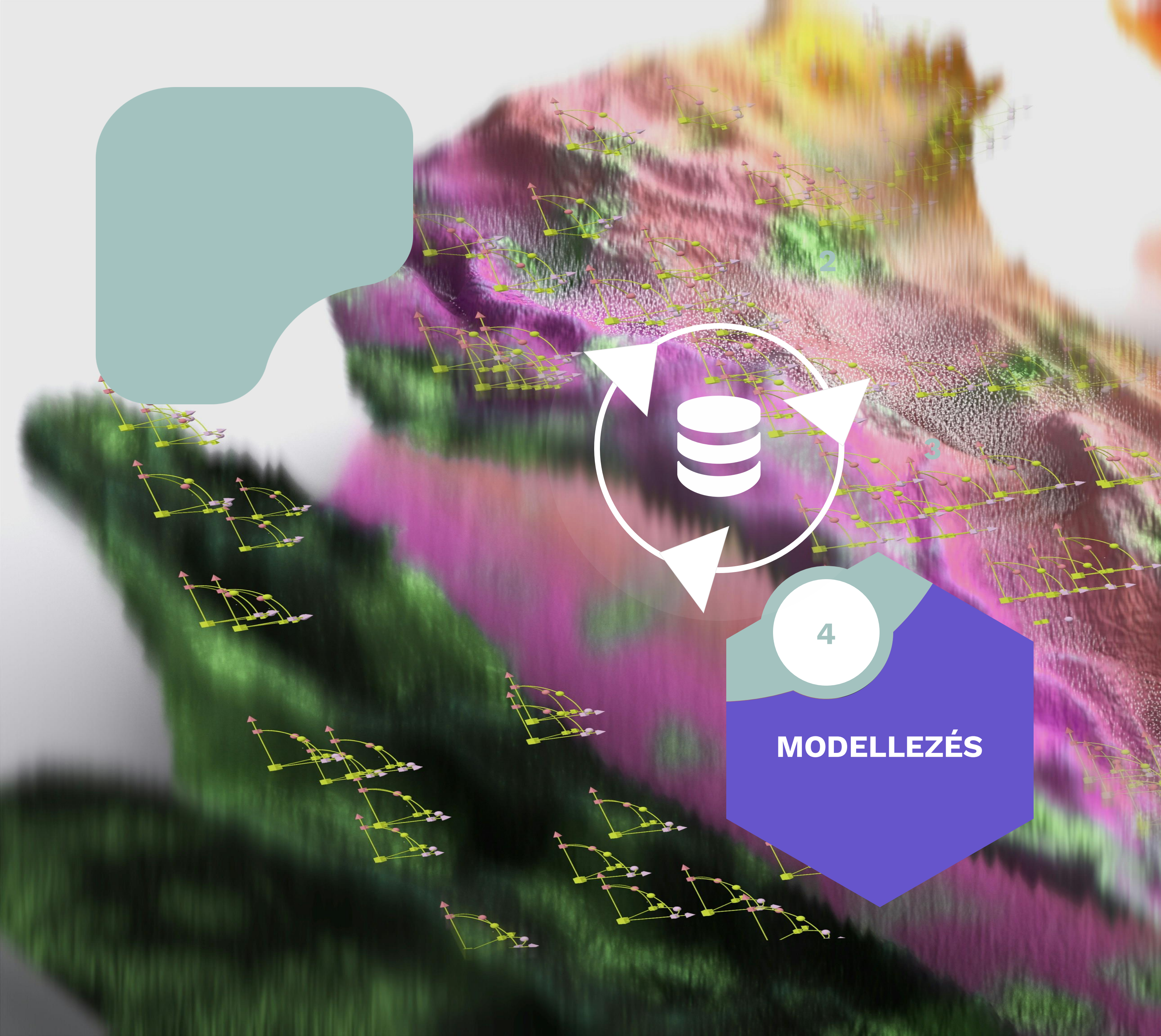
Mit fogunk felhasználni? És mit hagyunk ki?
Kritériumok: adatelemzési célok, adatmennyiségbeli korlátok

Adattisztítás

A feltárt problémák javítása (hogyan a módszereink lefussanak)

Adatgazdagítás

Külső adatforrások bevonása
Új változók képzése
Üzleti logika beépítése



Test design

Hogyan mérjük vissza a modellünk teljesítményét?

A valóságban hogyan kell működnie a modelleknek?

Modellezés

Milyen modellek jöhetnek szóba?

Paraméterek optimalizálása

Különböző módszerek

kombinálása

Kiértékelés

Modellek interpretálása

Technikai kiértékelés

Üzleti kontextus



Üzleti kiértékelés

Teljesültek az üzleti célok?
Ha nem, akkor mi az ami teljesült?
Üzletileg felhasználható a modell?

Interpretáció

Mit csinál a modell? Hogyan dönt?
Intuíció vs modell



**HADRENDBE
ÁLLÍTÁS**

6

Bevezetési terv

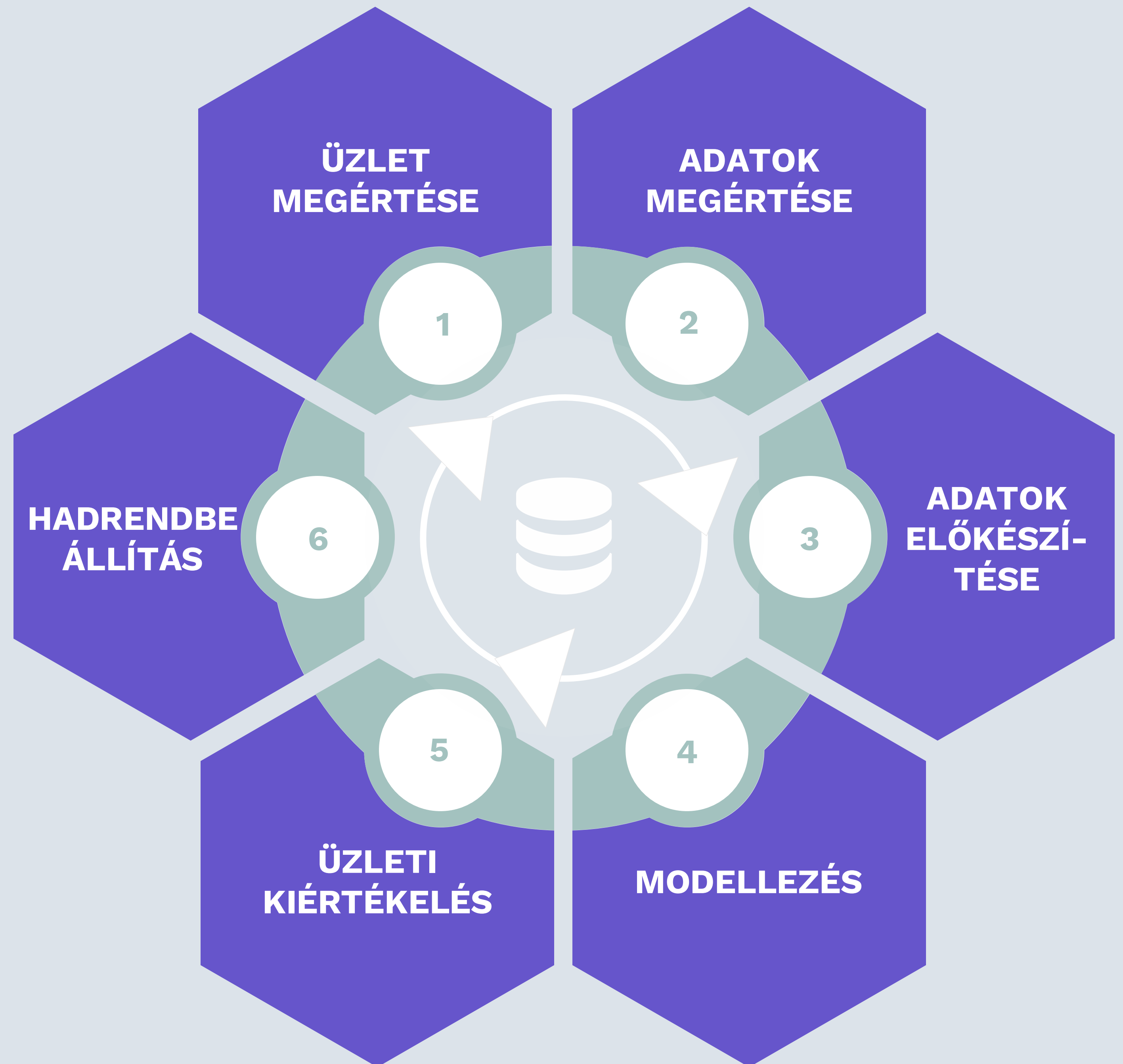
Hogyan használjuk fel?
Automatizáció
Üzleti folyamatba illesztés

Monitoring és karbantartás

Hogyan vesszük észre, ha
elromlik?
Folyamatos visszamérés

CRISP-DM módszertan

- Nem vízesés modell
- Érdeemes pilotolni
- Nem iparágspecifikus
- Keretrendszer, amely
 - egy kevésbé hozzáértő számára is érthetővé tesz egy adatelemzési projektet,
 - megismételhetővé teszi azt,
 - iránymutatást ad arra, hogy a projektnek milyen kimenetei vannak,
 - projekttervezési és -menedzsment gyakorlatot ad.
- További infók [itt](#) és [itt](#)



Van egy adatos problémád?
Keress meg!

Nagy-Rácz István
nagyracz.istvan@dmlab.hu

DMLAB