

Budapest Data Fórum 2021

Nyílt este megnyitó

Arató Bence
arato@biconsulting.hu



2021. JÚNIUS 23-24.

BUDAPEST

DATA FÓRUM

DATA WAREHOUSING | BIG DATA | DATA SCIENCE



BUDAPEST DATA FÓRUM

Online adatmenedzsment
és analitikai konferencia

2021. június 23-24.

Konferenciáink 2021-ben

BUDAPEST

ML

GÉPI TANULÁS | DATA SCIENCE | MESTERSÉGES INTELLIGENCIA

ÁPRILIS

DATA

ADATTÁRHÁZAK | STREAMING DATA | BIG DATA

JÚNIUS

BI

ADATVIZUALIZÁCIÓ | ÜZLETI INTELLIGENCIA | DATA-DRIVEN BUSINESS

NOVEMBER

PLATINA TÁMOGATÓNK

CLOUDERA

ARANY TÁMOGATÓINK

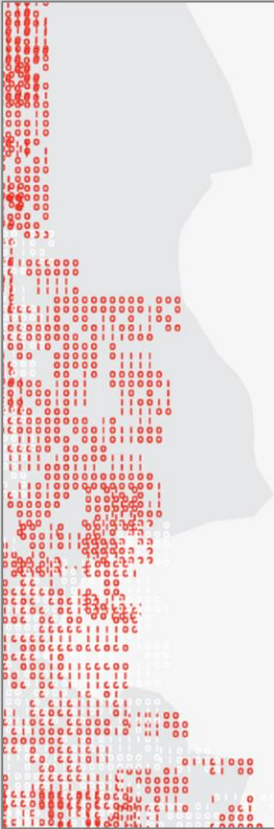
The logo for epam, featuring the word "epam" in a dark blue sans-serif font, with a light blue chevron pointing left before the 'e' and a light blue chevron pointing right after the 'm'.

meta consulting

The logo for dremio, featuring a stylized blue and white unicorn head icon to the left of the word "dremio" in a dark blue sans-serif font.
dremio

EZÜST TÁMOGATÓNK

Morgan Stanley



BUDAPEST DATA FÓRUM NYÍLT ESTE

- 2021 legfontosabb data és álláspiaci trendjei
- Toborzó partnerek bemutatkozásai
- Mit érdemes tanulni egy adatmérnöknek?
- Kerekasztal a toborzási trendekről
- Kerekasztal a modern adatplatformokról

2021. június 24. (csütörtök) · 18:00-20:00 · online

budapestdata.hu

Napirend

18:00 Megnyitó, hazai adatos trendek, Data Salary Survey eredmények

18:30 Toborzó cégek bemutatkozásai

18:50 Hogyan legyen adatmérnök?

19:10 Toborzási trendek kerekasztal

19:10 Modern Adatplatformok kerekasztal

A két kerekasztal párhuzamosan kerül megrendezésre.

Trendek

Trendek

- Álláspiaci trendek
- Technológiai trendek

Technológiai trendek

- **Változatos architektúrák**

- Hagyományos adattárház
 - Csak on-prem
 - On-prem és felhő támogatás is
 - Kizárólag felhőből elérhető DW szolgáltatások
- Data Lake
 - Klasszikus data lake
 - Data lakehouse

- **Cloud támogatás**

- Különböző megközelítések
 - On-prem + cloud
 - cloud-first és cloud only
 - Single cloud vs multi-cloud
- Dedikált szerver nélküli platformok
 - A „Serverless” előnye a könnyű skálázhatóság
 - Kizárólag használat alapú fizetés

- **Független komponensekből építkezés**

- Storage és Compute szétválasztása
 - Adatok tárolása HDFS vagy object storage rétegben (pl. Amazon S3)
 - Nyílt, optimalizált adatformátumok használata (Parquet)
- Többféle alternatív eszköz
 - SQL, streaming, workflow, metaadat, data quality
 - Általában nagy tech cégek belső fejlesztései

- **Open Source (újra) népszerűsége**

- Ismét egyre több a jól használható open source eszköz
- hagyományos open core + fizetős modell
- Egyre népszerűbb a teljesen open source szoftver + fizetős felhőszolgáltatás modell is

Álláspiaci trendek

- Egyre növekvő kereslet adatos szakemberekre
 - Ez általában is növekvőben lévő terület
 - Továbbra is érkeznek Budapestre fejlesztői és egyéb központok
 - Jelentős az export/külföld/remote hatás is
- A (felső)oktatás jelenleg nem tart ezzel lépést
 - De azért egyre inkább felpörögnek az intézmények
 - Egyre több magán képzés is indul
- Online tanulási lehetőség rengeteg van
 - De hogy mennyit ér, az kérdéses
 - Portfolió is nagyon fontos

Trendek

- Álláspiaci trendek -> Toborzási kerekasztal
- Technológiai trendek -> Adatplatformok kerekasztal



A Data Salary Survey

Mi az a Data Salary Survey?



Mi az a Data Salary Survey?

- Először 2019-ben készítettük el
- A piacon egyébként elérhető anyagok mellett (pl. Hays, Grafton) szerettünk volna egy kifejezetten adatos felmérést
- Célok
 - Többdimenziós adatfelvétel
 - Pozíció
 - Tapasztalat
 - Csapattagság
 - Munkahely típusa
 - “Privacy first” szemlélet
 - Minimális számú kérdés (pl. nincs iparági kérdés)
 - Legyen lehetséges a teljesen anonim kitöltés

Mi az a Data Salary Survey?

Data Salary Survey 2021

1 Oldal 1 2 Oldal 2

1) Mivel foglalkozol?

Milyen területen dolgozol? *

☐ Adattárház, big data ☐ BI, riporting, adatvizualizáció ☐ Data Science, ML, AI ☐ Több területen is

☐ Egyéb

Add meg, hogy a felsorolt területek közül melyik jellemzi leginkább a munkádat.

Ha ezek közül több is jellemző, akkor válaszd a "Több területen is" opciót.

Ha egyik sem jellemző, akkor add meg te a jellemző területet az Egyéb/ Other opciónál.

Mi az általad betöltött pozíció megnevezése? *

Hány év szakmai tapasztalattal rendelkezel ezen a területen? *

Megadhatasz közelítő értéket is, például 15+.

Milyen jellegű pozícióban dolgozol?

☐ Egyedül dolgozom ☐ Egy csapat tagjaként dolgozom ☐ Egy kisebb csapatot irányítok ☐ Egy nagyobb területért felelek

Minta a kérdőívből

Mi az a Data Salary Survey?

2) Mennyi a jövedelmed?

Milyen formában dolgozol? *

☒ Munkaviszonyban dolgozom ☐ Vállalkozóként, számlázással dolgozom ☐ Gyakornok vagyok

Munkaviszony

Milyen időkeretben dolgozol? *

☐ Teljes munkaidős munkaviszony ☐ 6 órás részmunkaidő ☐ 4 órás részmunkaidő ☐ Egyéb

Mennyi a HAVI bruttó fizetésed? *

Forint/ hó

Csak maga a fizetés, extra pénzek, jutalmak és cafeteria nélkül.

Egyéb pénzbeli jövedelmek (ÉVES) *

Forint/ év összesen

Éves kapott jutalom, 13. havi fizetés, szintén bruttó.

Teljes éves bruttó jövedelem

0

Részesülsz jelentősebb nem pénzbeli juttatásokban is? *

☐ Igen ☐ Nem

Módszertan

Módszertani kérdések

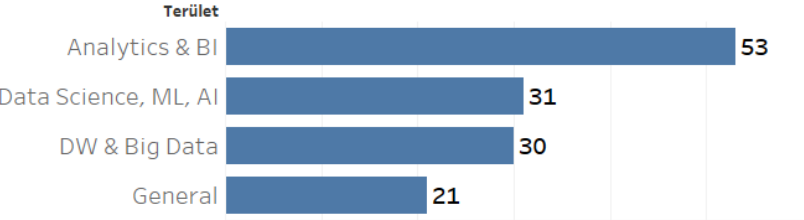
- Nettó/bruttó és családi adókedvezmények
- Lokáció, Remote, Home Office
- Skillek, technológiák
- Piaci szűrés
 - Magyar vagy külföldi piacról van tulajdonképpen szó?
 - Vannak tiiszta esetek és árnyalatok

Előzetes eredmények

- **Figyelmeztetések – Előzetes eredmények**
 - A felmérés kis mintájú és NEM reprezentatív
 - Vannak mintavételi torzítások
 - Vannak kérdőívtechnikai vagy módszertani hibák is
 - Adatvalidáció még csak részleges

Válaszok áttekintése

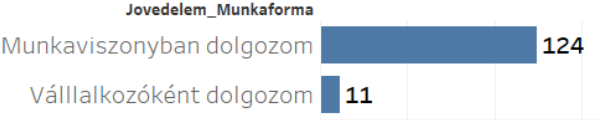
Stat - Terület



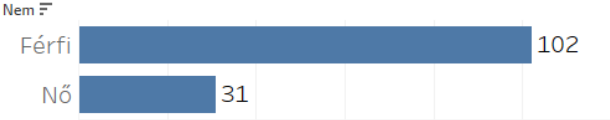
Stat - Szenioritás



Stat - Jelleg



Stat - Nem



A válaszok megoszlása néhány fontosabb jellemző szerint



Big Data Borat

@BigDataBorat

Follow



In Data Science, 80% of time spent prepare data, 20% of time spent complain about need for prepare data.

6:47 PM - 26 Feb 2013

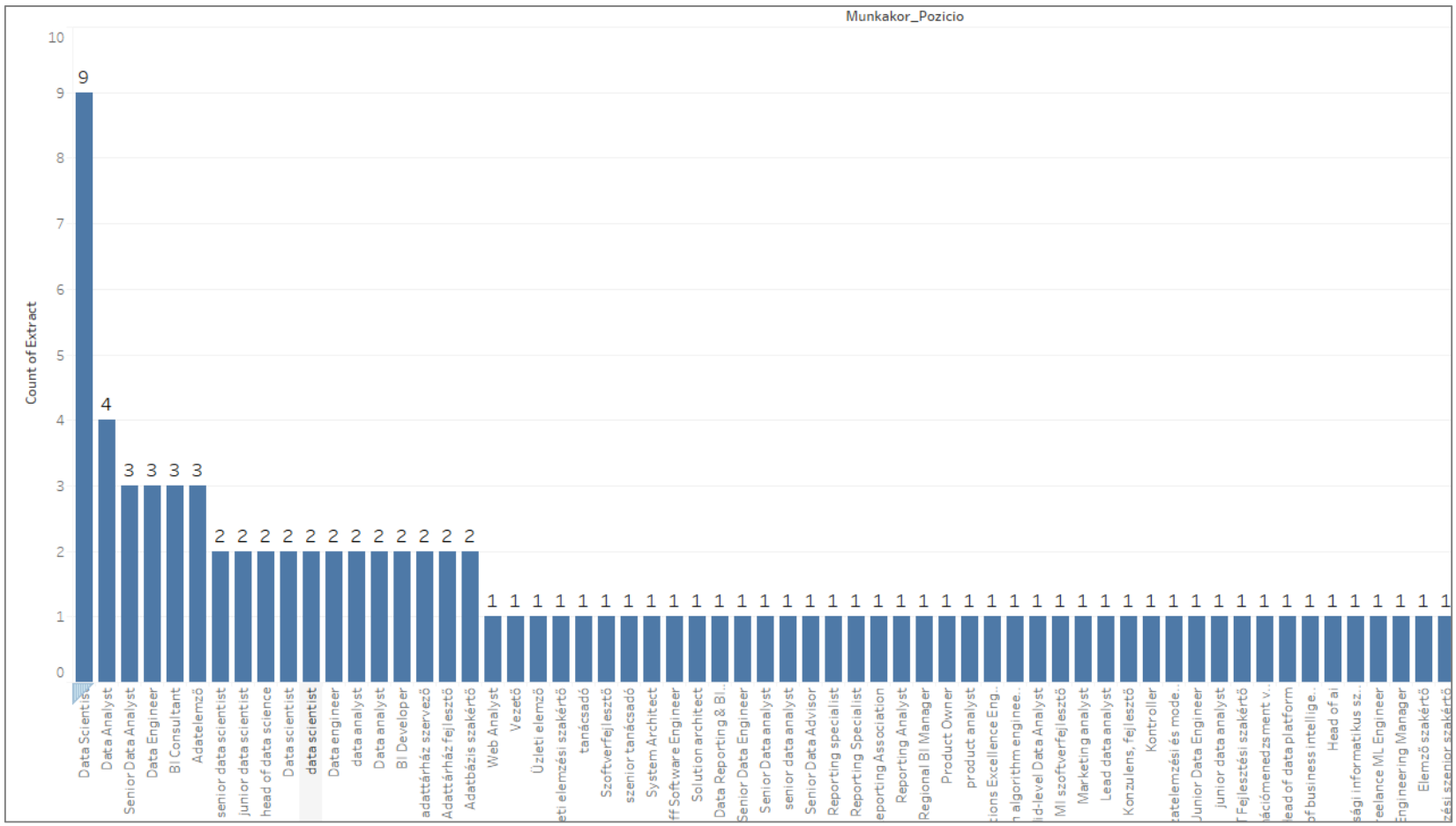
537 Retweets 343 Likes



Adatelőkészítés

- Adatelőkészítési kérdések
 - **Pozíciók standardizálása és kategorizálása**
 - **Kiugró/hibás értékek szűrése**
 - Nem pénzbeli juttatások vizsgálata
 - Vállalkozói konformizált jövedelem számítása

Adatelőkészítés



A pozíciónevek nagyon heterogének, rengeteg a különböző érték

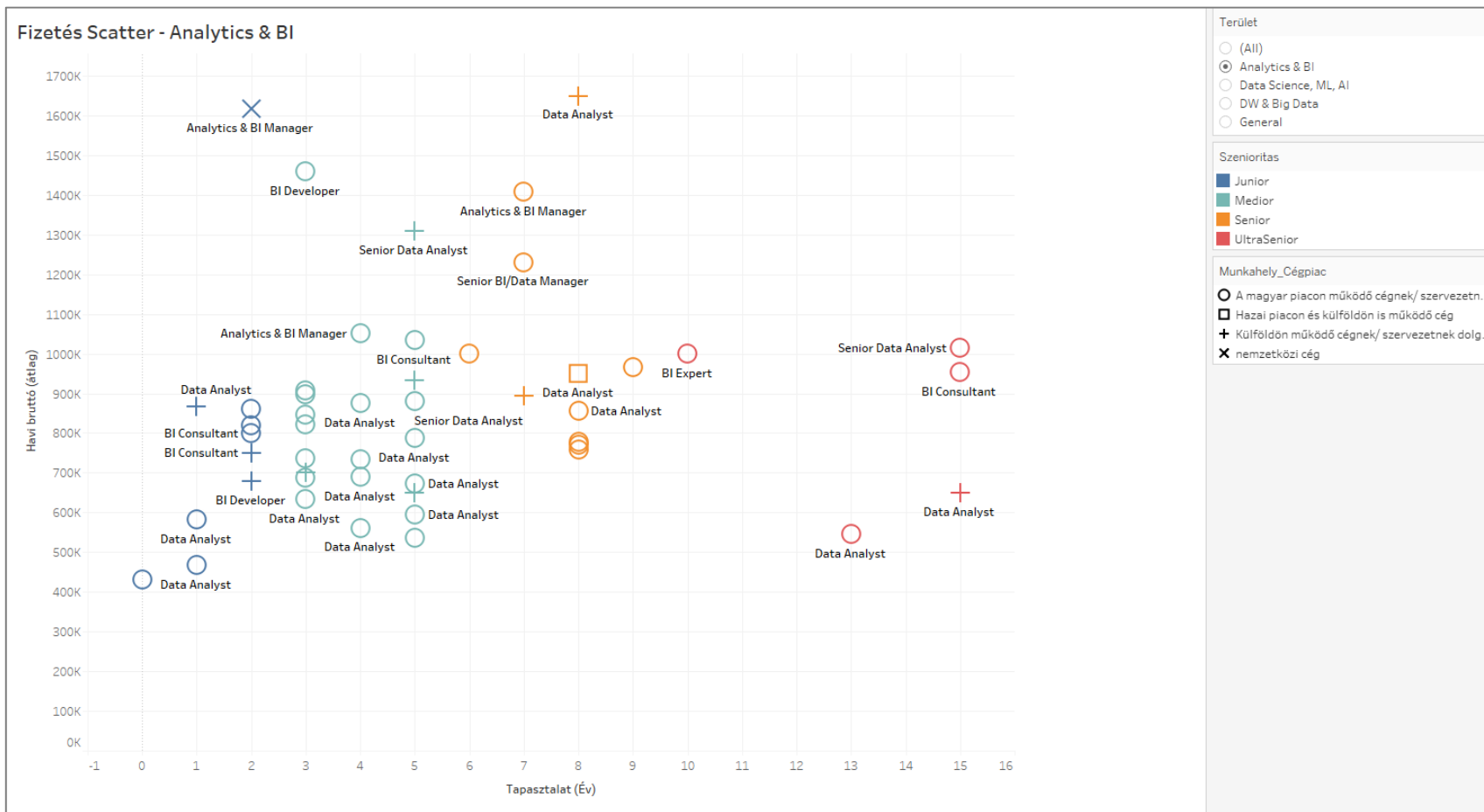
Adatelőkészítés

Terület	Pozicio LVL	Pozicio RIP	Pozicio Std
Data Science, ML, AI	Junior	Junior Data Scientist	Junior Data Scientist
	Manager	Data Science Manager	Data & AI Team Lead
			Data Science Manager
			Head of AI
			Head of Data Science
	Senior	Senior Data Scientist	Senior Data Scientist
	Standard	Data Scientist	Data Scientist
		ML Engineer	ML Engineer

Standardizált és kategorizált pozíciónevek (részlet)

Fizetési körkép

Fizetések tapasztalat szerint

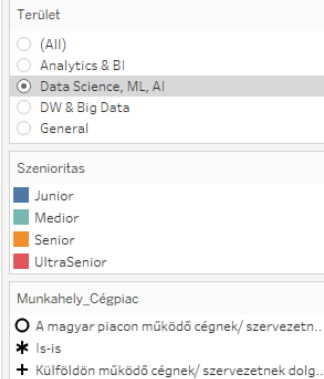


Fizetések tapasztalat szerint

Fizetés Scatter - DW & Big Data

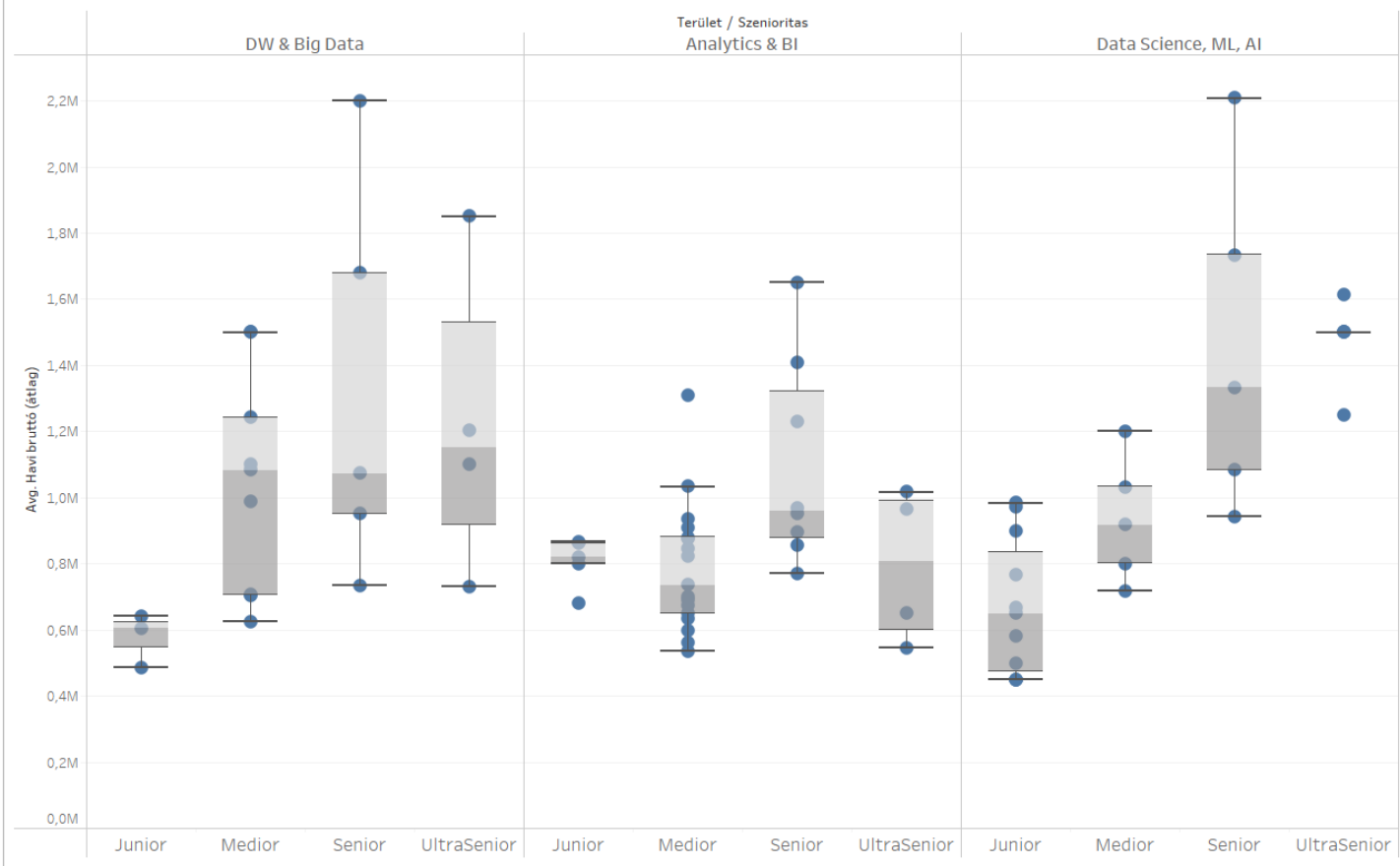


Fizetés Scatter - Data Science, ML, AI

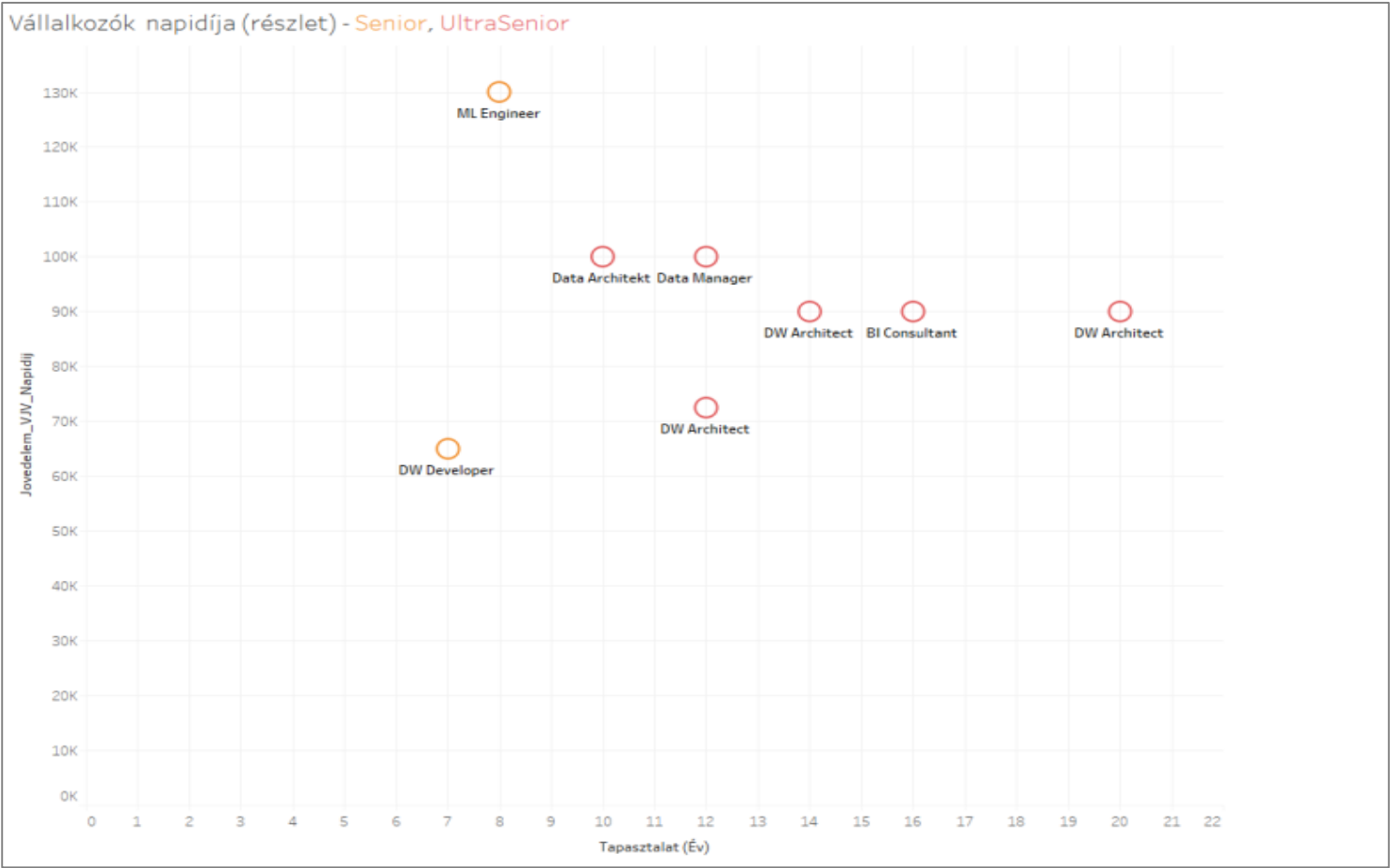


Fizetések tapasztalat szerint

Fizetések - Havi bruttó, bónuszokkal együtt



Vállalkozói napidíjak



Vállalkozói státuszban dolgozók jellemző napidíja (részlet)

Záró gondolatok

- **Figyelmeztetések**
 - A felmérés kis mintás és NEM reprezentatív
 - Jól láthatóak egyes mintavételi torzítások
 - Vannak kérdőívtechnikai vagy módszertani hibák is
- Továbbra is várjuk a kitöltéseket!
 - **DataSalarySurvey.hu**
- A végleges eredmények 2-3 héten belül várhatóak
 - Küldünk email értesítést (azoknak, akik kértek)
 - Konferencia Facebook és LinkedIn oldalán jelezzük
 - Meetup eseményt is tervezünk

Toborzó cégek bemutatkozásai

Toborzó cégek bemutatkozásai

- Cloudera
- EPAM
- Meta Consulting
- Morgan Stanley
- BI Consulting

Hogyan legyek adatmérnök?

Kerekasztalok

Thank You