



# DÁTOVÁ VEDA V ŠTÁTE

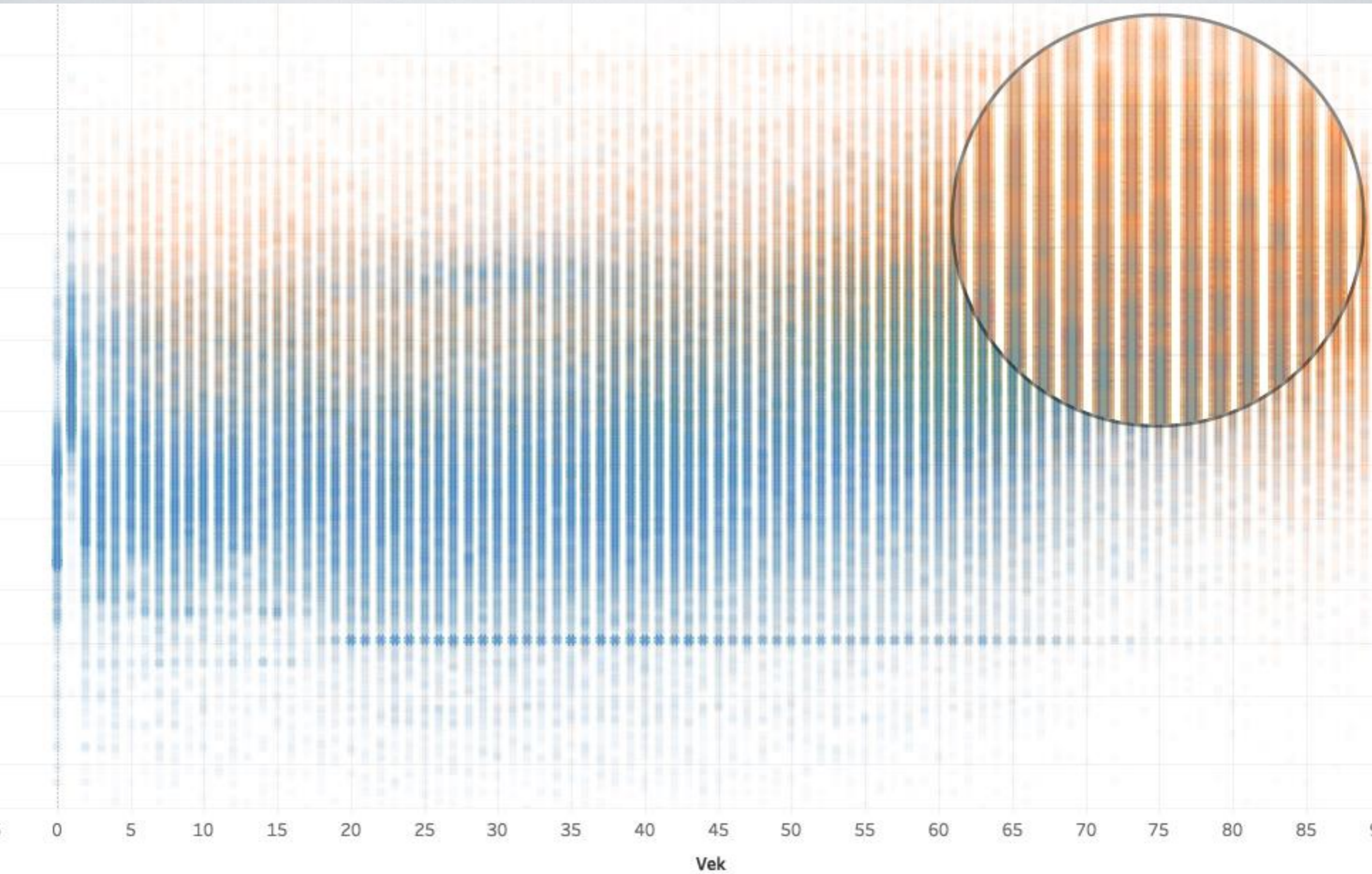
Ako sa dá použiť dátová veda a Big Data v analýzach, ktoré potrebuje štát a občan

# KDE BOLO TAM BOLO ...

- Alebo príbeh o úlohe
- Je možné lepšie naplánovať časť rozpočtu pre dávky poberateľov opakovaných peňažných príspevkov na kompenzáciu ŤZP?



# Dopredu identifikovať potenciálnych poberateľov s nárokom na príspevok ŽZP



# ČO POTREBUJEME - DÁTA

Ľubovolný formát



- Predchadzajúce príspevky ŽZP zo sociálnej oblasti
- dáta o príspevkoch pre ŽZP osoby - len informácia áno, nie

- Dáta zo zdravotnej oblasti

- dáta o jednotlivých agregovaných platbách poisťovne na pacienta cez diagnózu, výkon, lieky



# PRÍPRAVA A ČISTENIE DÁT

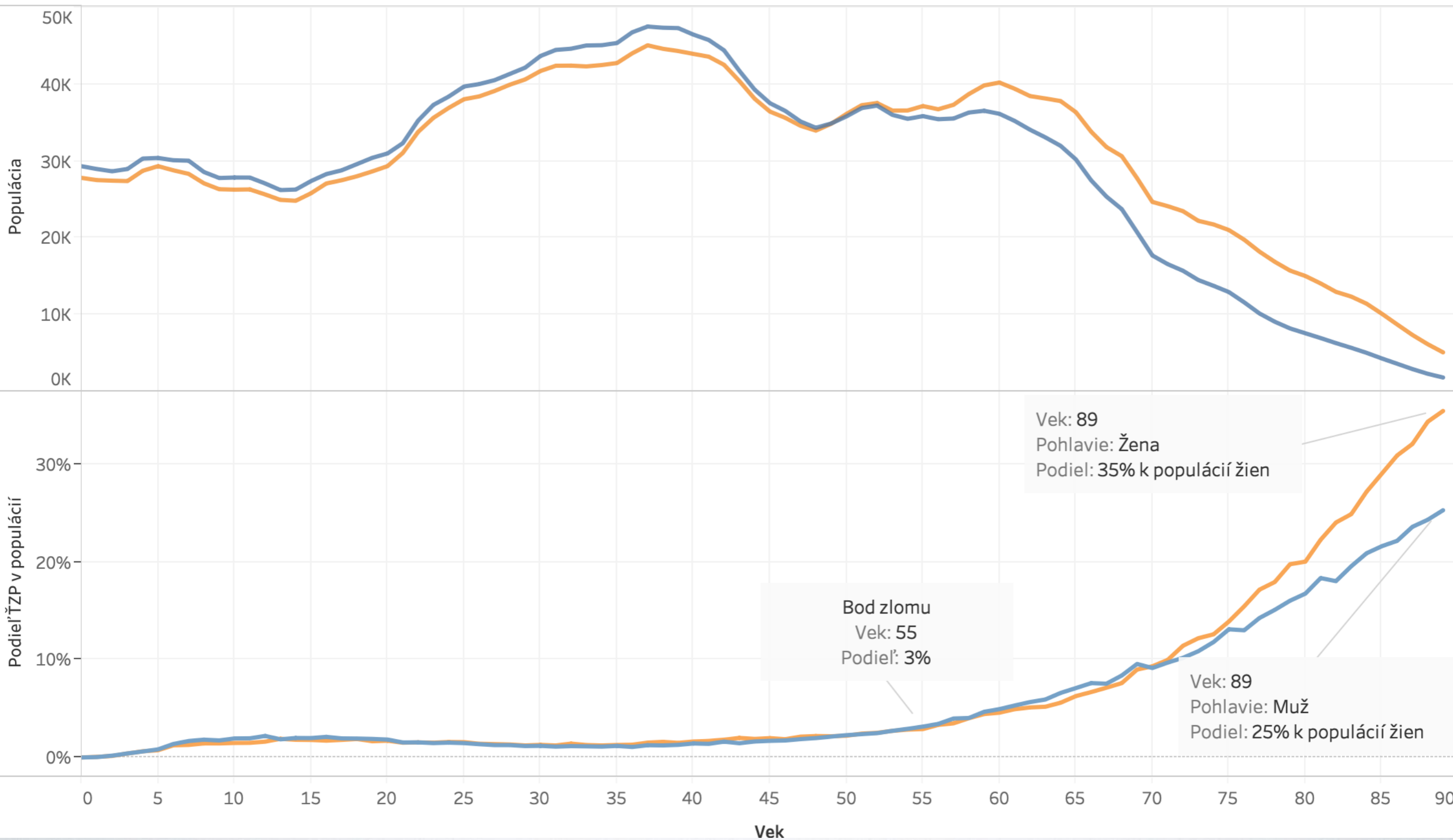
Nudná, ale časovo  
náročná časť

- Načítanie zo zdrojov
- Pochopenie sémantiky
- Transformácia
- Odstránenie "chýb",  
identifikácia minimálnych  
pozorovaní (Outlier)

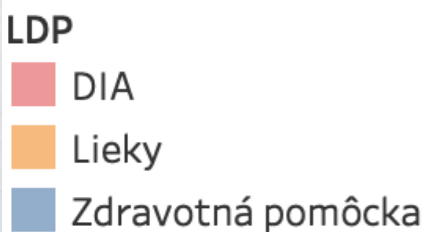
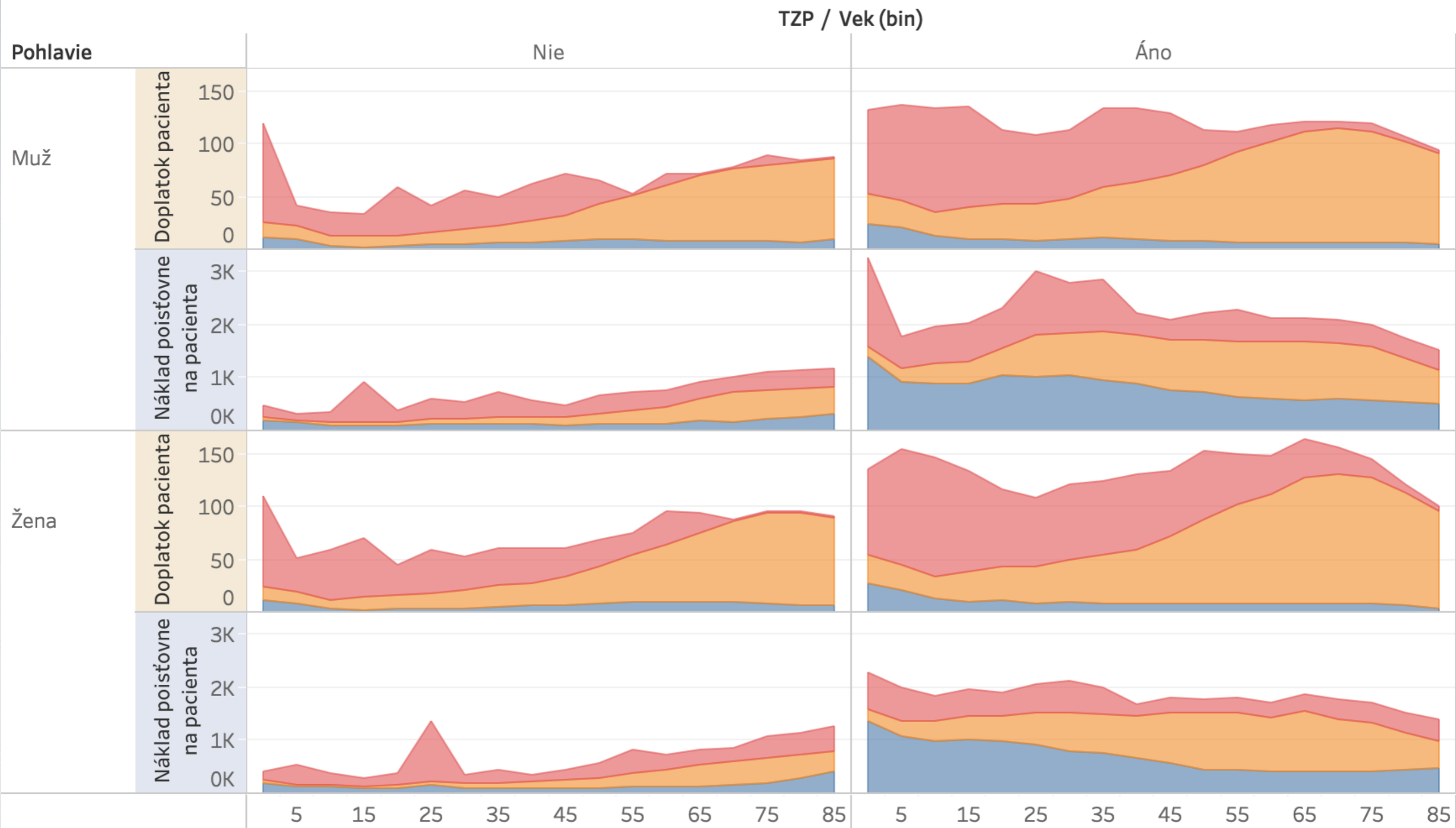


# DESKRIPTÍVNA ANALYTIKA

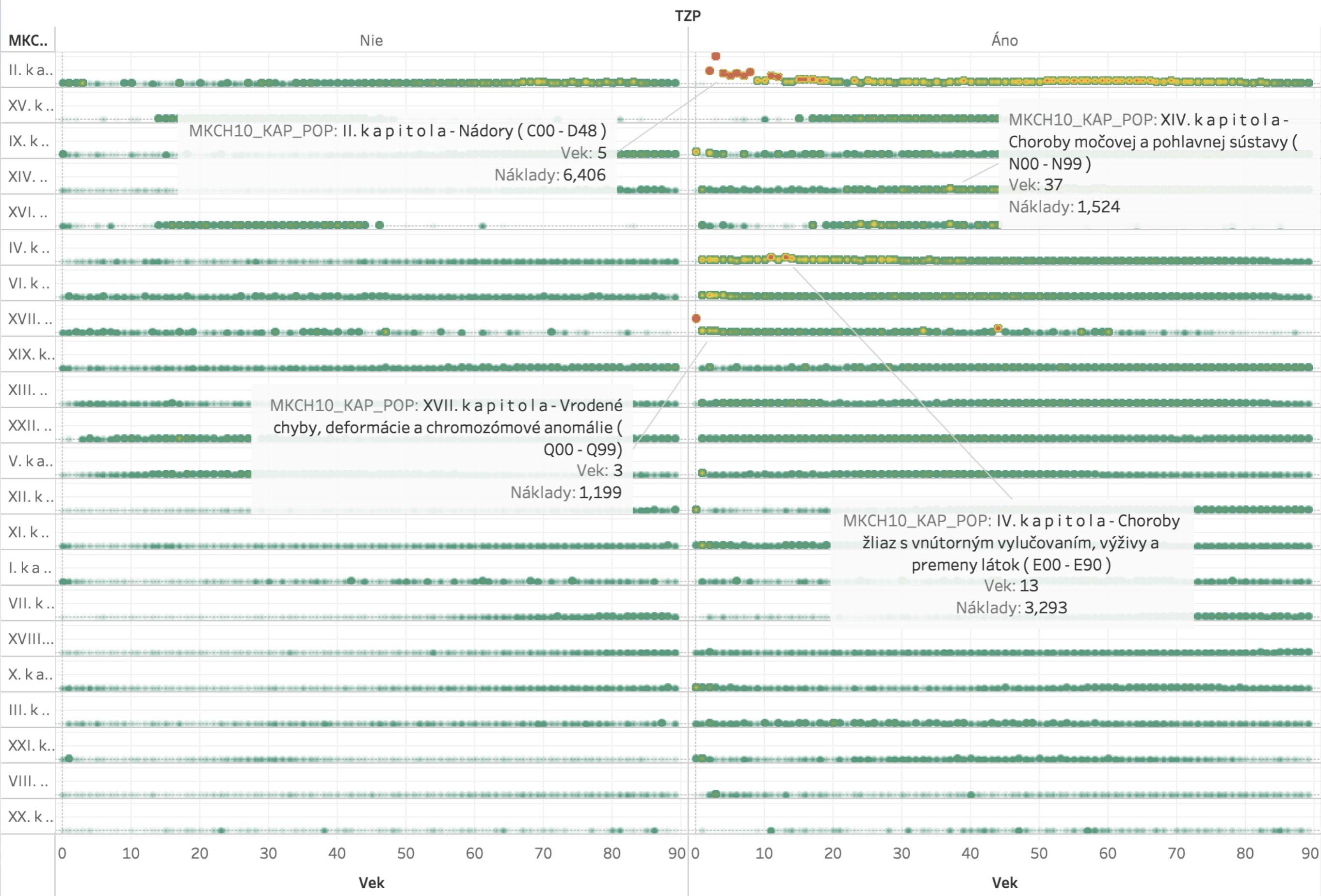
Podiel ŤZP v populácií podľa veku a pohlavia



# Priemerné doplatky a náklady poisťovne na lieky, zdravotné pomôcky a dia potraviny

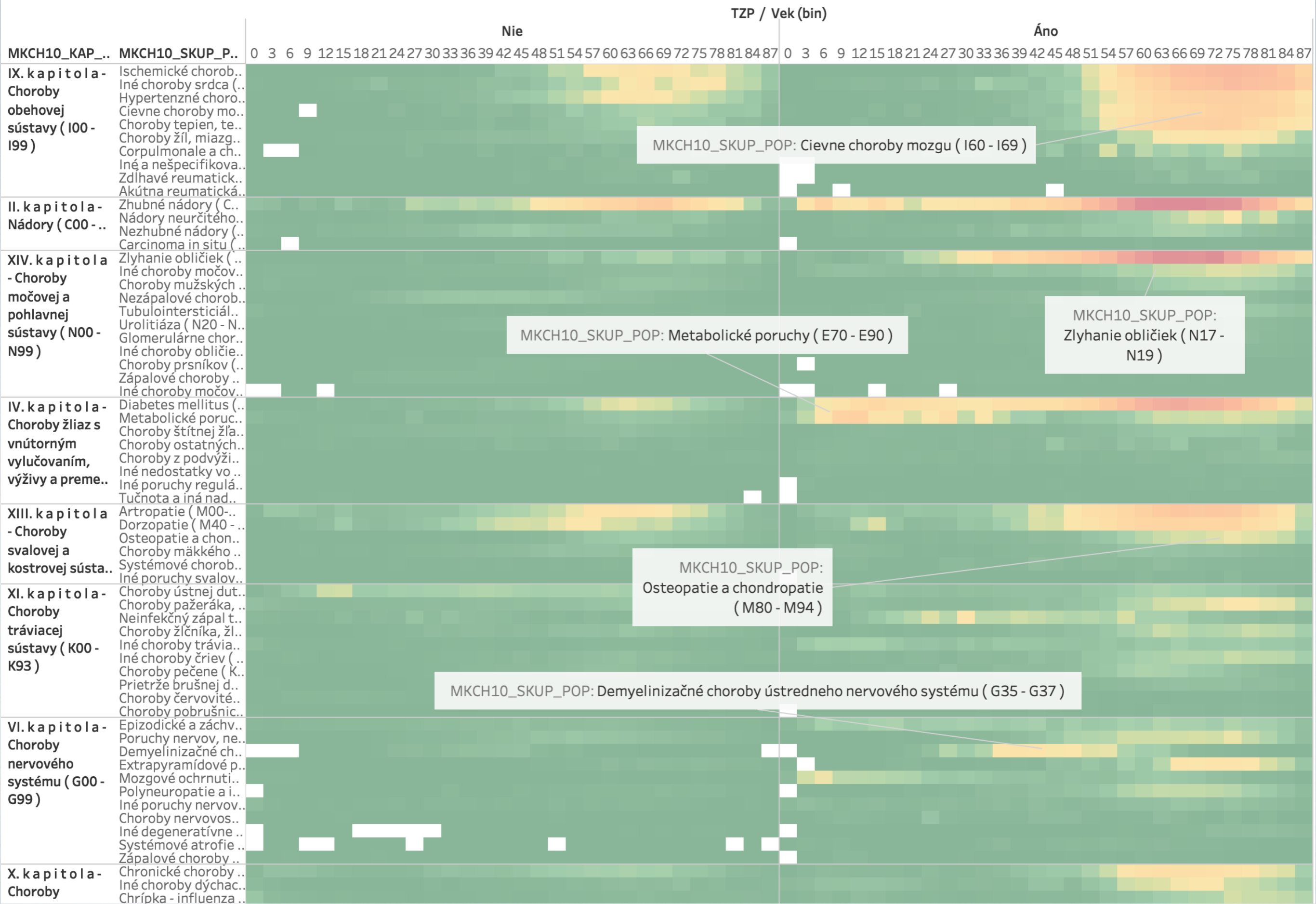


# Priemerné náklady na pacienta podľa kapitoly diagnózy a ŽZP





# Porovnanie ĹZP podľa skupiny diagnózy



# ČO JE TO MACHINE LEARNING



# PRÍKLAD DÔLEŽITÉ VLASTNOSTI (FEATURE IMPORTANCE)

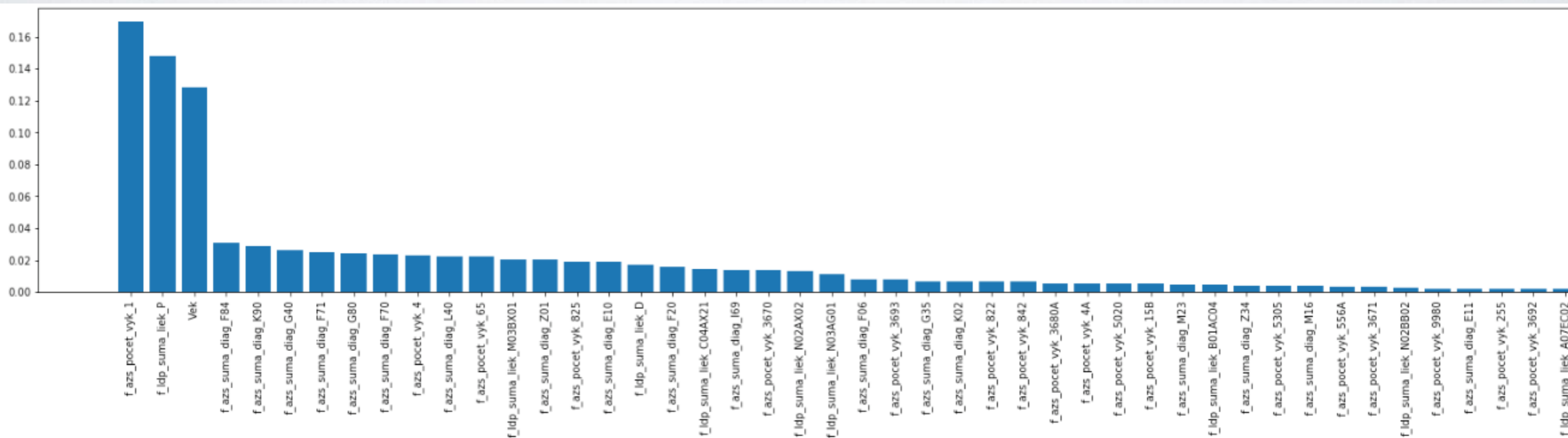
**Výkony 1** - Pod radou sa rozumie poskytnutie informácií pa...

Suma nákladov na zdravotné pomôcky

Vek

**Diagnózy:** Poruchy psychického vývinu, Črevná malabsorpcia, Epilepsia - zrádnik, Stredný stupeň duševnej zaostalost, ...

**Lieky:** Myorelaxanciá - Muskuloskeletárny systém, Periférne vazodilatanciá



# VYHODNOTENIE

Presnosť na testovacej vzorke **83,75%**

Presnosť pri tréovaní **83,69%**

## Confusion matrix

	Predikované <b>negatívne</b>	Predikované <b>pozitívne</b>
Skutočnosť <b>Nie</b>	<b>0,861</b>	0,139
Skutočnosť <b>Áno</b>	0,189	<b>0,811</b>

- Najlepšie predikujeme ľudí, ktorí **nemajú** byť poberateľmi príspevkov pre ŽZP (**86,1%**)
- Presnosť predikcie poberateľov príspevkov pre ŽZP (**81,1%**)
- Z čísel je evidentné, že absencia historických dát má väčší vplyv na presnosť ako absencia príjmu

# ĎALŠIE MOŽNOSTI

- Identifikovanie anomálnych prípadov na preskúmanie
- Predikcia žiadateľov o invalidný dôchodok
- Predikcia finančných dopadov
  - Dopad na rozpočet zo sociálneho poistenia (invalidný dôchodok)
  - Dopad na financovanie kompenzácií ŤZP
- Predikcia nákladov na dlhodobo chorých

