

HR: odejde, nebo zůstane? Predikce odchodu zaměstnance z firmy

Na možný odchod zaměstnance firmy mají vliv různé proměnné. Cílem HR je právě zjistit či predikovat fluktuaci a zaměřit se tak na možné faktory, které může firma vůči zaměstnanci změnit.

Na základě odborných zdrojů z oblasti psychologie a řízení lidských zdrojů lze identifikovat několik faktorů ovlivňujících fluktuaci zaměstnanců. Obecně můžeme rozdělit do několika skupin – týkající se ohodnocení, prostředí, vztahů a samotné vykonávané pozice. Odborné zdroje nejčastěji hovoří o těchto faktorech:

Odměňování a benefity: Nedostatečná mzda nebo nepřiměřený systém odměn může vést k fluktuaci zaměstnanců (Trevor, Gerhart & Boudreau, 1997).

Vztahy na pracovišti: Konflikty s kolegy nebo nadřízenými mohou zhoršit pracovní atmosféru a přispět k odchodu (Griffeth, Hom & Gaertner, 2000).

Pracovní podmínky: Nepříznivé pracovní prostředí, nadměrná pracovní zátěž nebo nejasně definované pracovní povinnosti (Mobley, 1977).

Možnosti kariérního růstu: Chybějící příležitosti k profesnímu rozvoji a postupu často vedou k rozhodnutí změnit zaměstnání (Hom & Griffeth, 1995).

Work-life balance: Nepřiměřený tlak na pracovní výkon a nemožnost sladění pracovního a osobního života vedou ke zvýšené fluktuaci (Greenhaus & Beutell, 1985).

Nicméně dataset, se kterým pracuji, nesleduje všechny výše uvedené proměnné. Pokud vyjdu z proměnných datasetu, ráda bych ještě zmínila další faktory, které dle odborných zdrojů mohou fluktuaci ovlivnit (závorce je již uveden název sledované proměnné):

Míra spokojenosti zaměstnance (satisfaction_level) – Nízká pracovní spokojenost je často spojována s vyšší mírou fluktuace zaměstnanců (Judge et al., 2001).

Poslední hodnocení výkonu (last_evaluation) – Pravidelné a konstruktivní hodnocení výkonu může zvýšit pracovní spokojenost a snížit tendenci k odchodu ze zaměstnání (Aguinis, 2013).

Počet projektů, na kterých zaměstnanec pracoval (number_project) – Nadměrné pracovní zatížení nebo naopak nedostatek výzev mohou vést k nespokojenosti a zvýšené fluktuaci (Hackman & Oldham, 1976).

Průměrný měsíční počet odpracovaných hodin (average_monthly_hours) – Nerovnováha mezi pracovním a osobním životem, způsobená nadměrným počtem odpracovaných hodin, může negativně ovlivnit spokojenost zaměstnanců a zvýšit jejich odchodovost (Greenhaus & Beutell, 1985).

Počet let ve firmě (time_spend_company) – Délka zaměstnání ve firmě může ovlivňovat loajalitu; novější zaměstnanci mohou být náchylnější k odchodu, zatímco dlouhodobí zaměstnanci mohou vykazovat vyšší míru setrvání (Griffeth, Hom & Gaertner, 2000).

Pracovní úraz (work_accident) – Výskyt pracovních úrazů může vést k nespokojenosti a zvýšené fluktuaci, zvláště pokud zaměstnanci vnímají nedostatečnou péči ze strany zaměstnavatele (Shaw et al., 1998).

Povýšení za posledních 5 let (promotion_last_5years) – Nedostatek kariérního postupu a rozvoje může vést k pocitu stagnace a následně k odchodu zaměstnanců (Ng, Eby, Sorensen & Feldman, 2005).

Úroveň platu (salary) – Nespokojenost s platovým ohodnocením je jedním z hlavních důvodů fluktuace; konkurenceschopné mzdy mohou zvýšit retenci zaměstnanců (Gerhart & Rynes, 2003).

Dataset

Pro tento úkol jsem zvolila volně dostupná data z [Kaggle.com](https://www.kaggle.com) týkající se různých proměnných měření u 15000 zaměstnanců firmy.

Dataset obsahuje následující proměnné:

- **satisfaction_level**: míra spokojenosti zaměstnance (0–1)
- **last_evaluation**: poslední hodnocení výkonu (0–1)
- **number_project**: počet projektů, na kterých zaměstnanec pracoval
- **average_monthly_hours**: průměrný měsíční počet odpracovaných hodin
- **time_spend_company**: počet let ve firmě
- **Work_accident**: zda zaměstnanec měl pracovní úraz (0 = ne, 1 = ano)

- **left**: cílová proměnná (0 = zůstal, 1 = odešel)
- **promotion_last_5years**: zda zaměstnanec dostal povýšení za posledních 5 let (0 = ne, 1 = ano)
- **Department**: oddělení, ve kterém pracuje
- **salary**: úroveň platu (low, medium, high)

Analýza dat

Pro analýzu dat jsem zvolila logistickou regresi, kde závisle proměnná je *left*, tedy jestli zaměstnanec zůstal, nebo firmu opustil. Data jsem analyzovala v jazyce *R*, v online nástroji *Google Colab*

Níže uvádím hodnoty a interpretaci hlavních koeficientů.

Tady je tabulka regresních koeficientů:

| Proměnná | Koeficient | p-hodnota |
|------------------------|------------|-----------|
| Intercept | -1,657 | < 0,001 |
| Satisfaction level | -4,136 | < 0,001 |
| Last evaluation | 0,731 | < 0,001 |
| Number of projects | -0,315 | < 0,001 |
| Average monthly hours | 0,004 | < 0,001 |
| Time spent in company | 0,268 | < 0,001 |
| Work accident | -1,530 | < 0,001 |
| Promotion last 5 years | -1,430 | < 0,001 |
| Department_R&D | -0,402 | 0,003 |
| Department_Accounting | 0,181 | 0,139 |
| Department_HR | 0,413 | 0,001 |
| Department_Management | -0,268 | 0,078 |
| Department_Marketing | 0,169 | 0,166 |
| Department_ProductMng | 0,028 | 0,818 |
| Department_Sales | 0,142 | 0,109 |
| Department_Support | 0,231 | 0,017 |
| Department_Technical | 0,251 | 0,007 |
| Salary_Low | 1,944 | < 0,001 |
| Salary_Medium | 1,413 | < 0,001 |

Tabulka 1: Regresní koeficienty

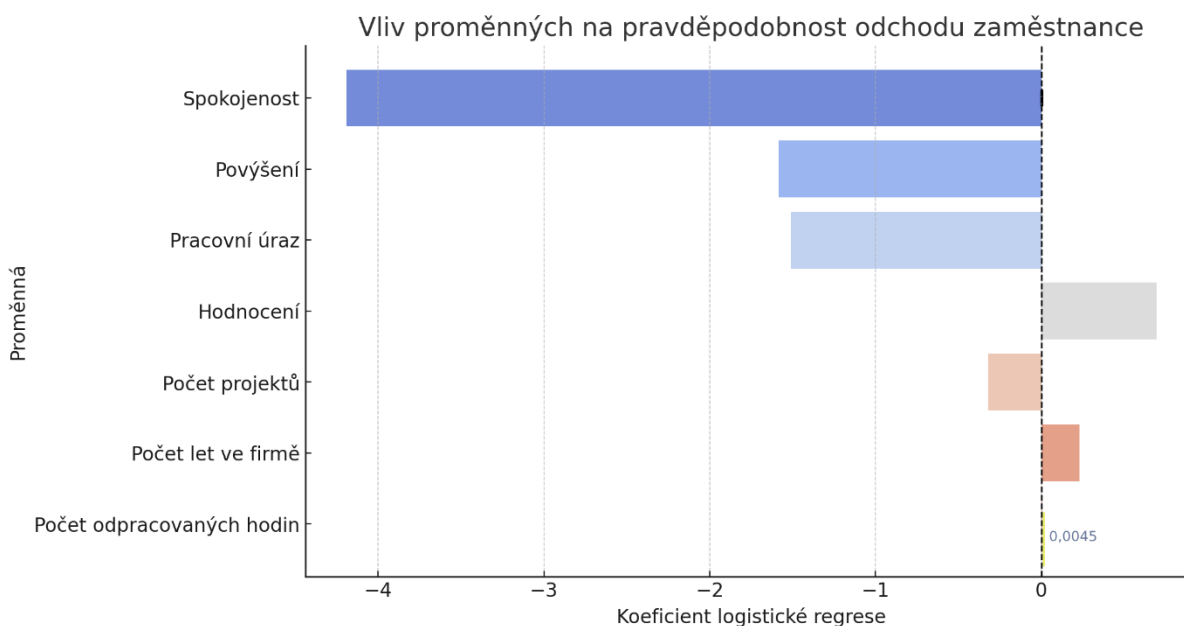
✓ **Pozitivní vliv na odchod (zvyšují pravděpodobnost odchodu)**

- **Nízký plat** (*coef* = 1.94, *p* < 0.001) → Zaměstnanci s nízkým platem mají vyšší pravděpodobnost odchodu než ti s vysokým platem.
- **Střední plat** (*coef* = 1.43, *p* < 0.001) → I zaměstnanci se středním platem častěji odcházejí ve srovnání s těmi s vysokým platem.
- **Délka zaměstnání (time_spend_company)** (*coef* = 0.267, *p* < 0.001) → Čím déle je zaměstnanec ve firmě, tím vyšší je pravděpodobnost jeho odchodu.
- **Průměrný měsíční počet hodin (average_monthly_hours)** (*coef* = 0.004, *p* < 0.001) → Vyšší počet odpracovaných hodin lehce zvyšuje pravděpodobnost odchodu.
- **Poslední hodnocení (last_evaluation)** (*coef* = 0.730, *p* < 0.001) → Vyšší hodnocení výkonu souvisí s vyšší pravděpodobností odchodu.
- **Oddělení:**
 - HR (*coef* = 0.413, *p* = 0.001) → Zaměstnanci z HR oddělení častěji odcházejí.
 - Technické oddělení (*coef* = 0.250, *p* = 0.007) → Rovněž vyšší pravděpodobnost odchodu.
 - Podpora (*coef* = 0.230, *p* = 0.017) → Podobná situace jako u technického oddělení.

✗ **Negativní vliv na odchod (snižují pravděpodobnost odchodu)**

- **Spokojenost zaměstnance (satisfaction_level)** (*coef* = -4.135, *p* < 0.001) → Čím vyšší spokojenost, tím nižší pravděpodobnost odchodu.
- **Počet projektů (number_project)** (*coef* = -0.315, *p* < 0.001) → Více projektů snižuje pravděpodobnost odchodu.
- **Pracovní úraz (Work_accident)** (*coef* = -1.529, *p* < 0.001) → Zaměstnanci, kteří měli pracovní úraz, méně často odcházejí.
- **Povýšení za posledních 5 let (promotion_last_5years)** (*coef* = -1.430, *p* < 0.001) → Ti, kteří byli povýšeni, zůstávají ve firmě častěji.
- **Oddělení R&D (výzkum a vývoj)** (*coef* = -0.401, *p* = 0.003) → Zaměstnanci v R&D oddělení méně často odcházejí.

Koeficienty si může zobrazit graficky pro přehlednost (proměnná *Oddělení* zde není, jelikož je kategorická).



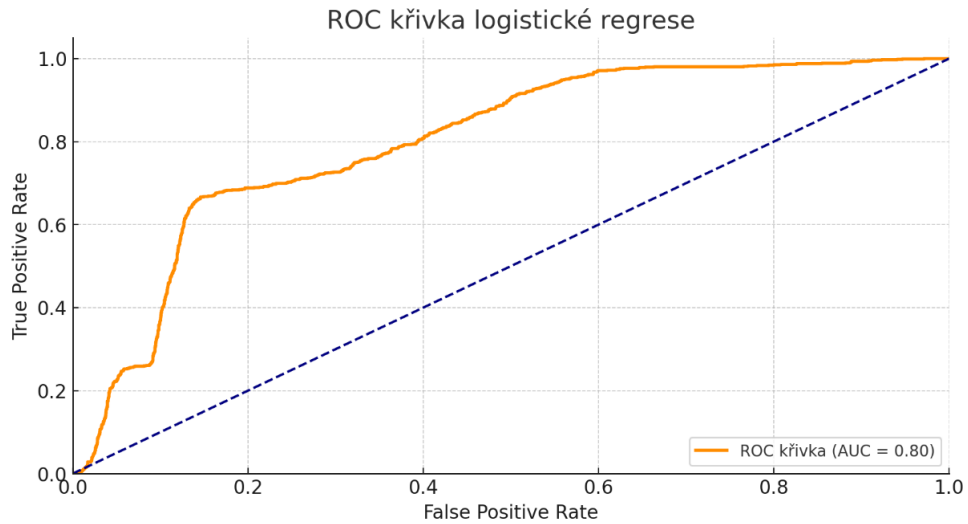
Graf 1: Regresní koeficienty

Celková kvalita modelu

Získali jsme $R^2 = 0,32195$ podle Cox & Snell. To znamená, že náš model vysvětluje přibližně 22 % variability v odchodu zaměstnanců. To není špatné, ale zároveň to ukazuje, že odchod zaměstnanců je ovlivněn i dalšími faktory, které nejsou v modelu zahrnuty (např. osobní důvody, firemní kultura apod.). Model můžeme zlepšit tak, že přidáme další relevantní proměnné nebo použijeme jiný typ modelu.

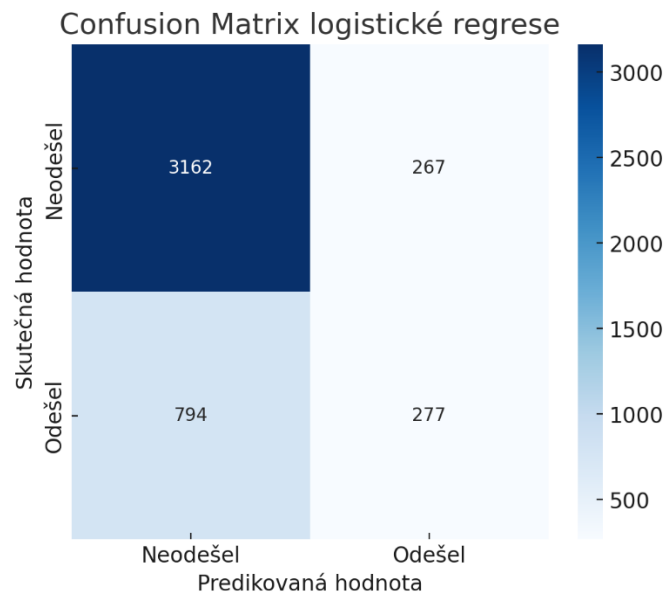
Kromě Cox & Snell existují další metody. McFadden R^2 vychází 0,192 a Nagelkerke R^2 0,286. Tyto hodnoty naznačují, že model má určitou prediktivní sílu, ale stále existuje prostor pro zlepšení, například přidáním dalších relevantních proměnných.

Pro zajímavost si vykreslíme ROC křivku a Confusion matrix pro vyhodnocení modelu.



Graf 2: ROC křivka

ROC křivka ukazuje výkon modelu. AUC (plocha pod křivkou) je kolem 0,76, což znamená, že model má slušnou prediktivní schopnost, ale není perfektní.

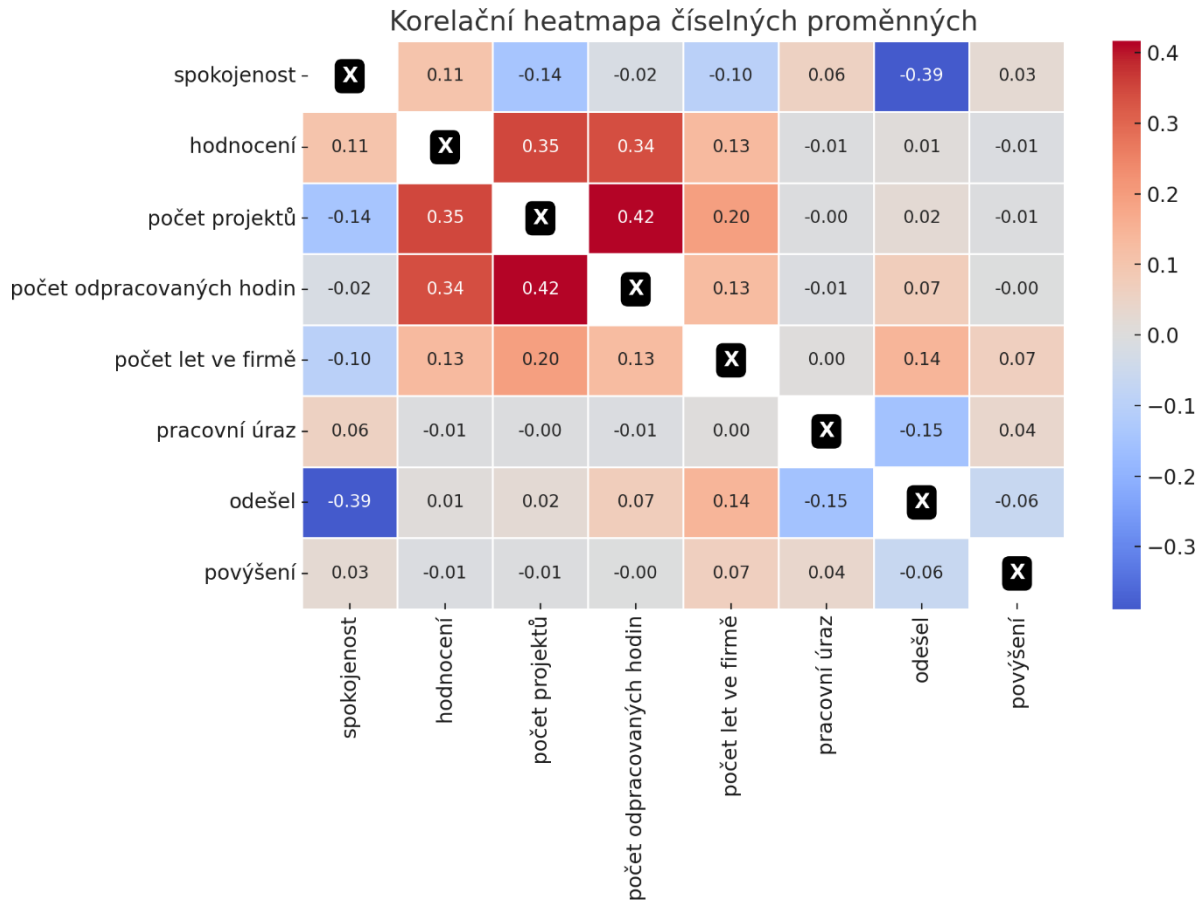


Graf 3: Confusion matrix

Confusion matrix – ukazuje, kolik predikcí bylo správných a kde model dělá chyby. Vidíme, že model má slušnou přesnost, ale může docházet k určitému počtu falešně pozitivních či falešně negativních predikcí.

Vizualizace dat

Nyní pro zajímavost vykreslíme korelační heatmapu, která ukazuje korelace mezi jednotlivými proměnnými.



Graf 4: Korelační heatmapa

Níže se podíváme na některé hodnoty:

Počet projektů × počet odpracovaných hodin

- Čím více projektů zaměstnanec řeší, tím více hodin odpracuje.
- Může naznačovat přetížení zaměstnanců s velkým množstvím úkolů.

Počet let ve firmě × odešel

- Slabší, ale pozitivní vztah – zaměstnanci, kteří jsou ve firmě déle, mají vyšší pravděpodobnost odchodu.
- Může ukazovat na vyhoření nebo stagnaci v kariérním růstu.

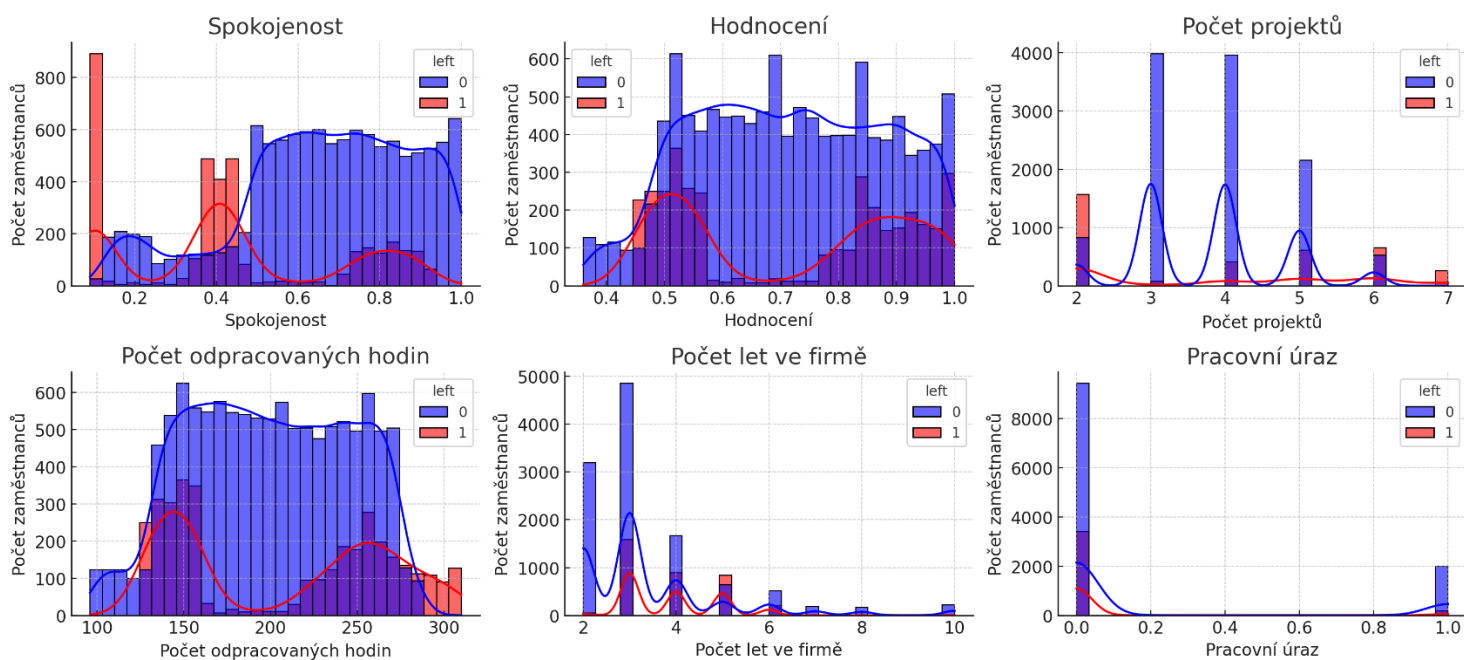
Spokojenost × odešel

- Nejvýraznější negativní vztah – nespokojení zaměstnanci častěji odcházejí.
- Potvrzuje, že udržení zaměstnanců souvisí se spokojeností na pracovišti.

Počet odpracovaných hodin × spokojenost

- Zaměstnanci s více hodinami mají obecně nižší spokojenost.
- Může signalizovat problém s work-life balance.

Jako další se podíváme na histogramy.



Graf 5: Histogramy

Histogramy ukazují rozložení číselných proměnných podle toho, zda zaměstnanec **odešel (červená)** nebo **zůstal (modrá)**. Klíčové poznatky:

1. Spokojenost

- Zaměstnanci, kteří odešli, měli častěji velmi nízkou nebo naopak velmi vysokou spokojenost.
- Ti, co zůstali, mají spokojenost častěji ve středních hodnotách.

2. Hodnocení výkonu

- Zaměstnanci, kteří odešli, mají více rozložené hodnocení – od nižších po velmi vysoké.
- U zůstávajících zaměstnanců je rozložení rovnoměrnější.

3. Počet projektů

- Zaměstnanci, kteří odešli, často pracovali buď na **málo projektech (1-2)**, nebo naopak na **hodně projektech (6-7)**.
 - U zůstávajících zaměstnanců je počet projektů mezi 3 až 5.
4. **Počet odpracovaných hodin**
- Zaměstnanci, kteří odešli, měli tendenci pracovat buď **výrazně méně**, nebo **výrazně více** hodin než ti, kteří zůstali.
 - Znamená to, že extrémní pracovní vytížení může vést k odchodům.
5. **Počet let ve firmě**
- Nejvíce odchodů se děje mezi zaměstnanci s **5 a více lety praxe**.
 - Kratší doba zaměstnání souvisí spíše se stabilnější pracovní silou.
6. **Pracovní úrazy a povýšení**
- Pracovní úraz nemá významný vliv na odchod zaměstnanců.
 - Zaměstnanci, kteří **nebyli povýšeni za posledních 5 let**, častěji odcházejí, což může naznačovat frustraci z nedostatku kariérního růstu.

Praktická interpretace výsledků

Z mé zkušenosti z praxe je nutná interpretace výsledků i pro ne-statistiky, kteří často ve firmě (a na vedoucích pozicích) pracují. Níže uvádím tedy výsledky, které může případný manažer použít:

- **Plat je klíčový faktor:** Lidé s nízkými a středními platy častěji odcházejí.
- **Spokojenost je extrémně důležitá:** Nejspokojenější zaměstnanci téměř nikdy neodcházejí.
- **Dlouhá doba ve firmě = vyšší riziko odchodu:** Zaměstnanci, kteří jsou ve firmě déle, mají vyšší pravděpodobnost odchodu.
- **Oddělení má vliv:** HR a technické pozice častěji odcházejí, zatímco výzkumníci zůstávají déle. Zde by stálo zvážit výzkum v rámci oddělení, jak se liší další podmínky práce.
- **Povýšení a pracovní úrazy drží lidi ve firmě:** Lidé, kteří byli povýšeni nebo měli pracovní úraz, mají nižší pravděpodobnost odchodu. Zde se nabízí informace, že tedy máme zaměstnance nechat si ublížit, abychom si je udrželi...? Tento poznatek je v rozporu s teoretickým zakotvením.

Je ale jasné, že výsledky se týkají pouze tohoto datasetu. V jiné firmě by mohly být ukazatele jiné. Nicméně výsledky odpovídají zjištěním, uvedených v teoretickém zakotvení. Zajímavé by bylo se podívat, co vše se schovává pod proměnnou *Spokojenost*, která má velký vliv.

Z mého pohledu to byl velmi zajímavý dataset k testování, doporučuji kolegům pro další zpracování.

Zdroje

Dataset

<https://www.kaggle.com/datasets/mfaisalqureshi/hr-analytics-and-job-prediction?resource=download>

Literatura

- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). "A Meta-Analysis of Antecedents and Correlates of Employee Turnover: Update, Moderator Tests, and Research Implications for the Next Millennium." *Journal of Management*, 26(3), 463–488.
- Hom, P. W., & Griffeth, R. W. (1995). *Employee Turnover*. South-Western College Publishing.
- Mobley, W. H. (1977). "Intermediate linkages in the relationship between job satisfaction and employee turnover." *Journal of Applied Psychology*, 62(2), 237–240.
- Trevor, C. O., Gerhart, B., & Boudreau, J. W. (1997). "Voluntary Turnover and Job Performance: Curvilinearity and the Moderating Influences of Salary Growth and Promotions." *Journal of Applied Psychology*, 82(1), 44–61.
- Aguinis, H. (2013). *Performance management*. Pearson.
- Gerhart, B., & Rynes, S. L. (2003). *Compensation: Theory, evidence, and strategic implications*. SAGE Publications.
- Greenhaus, J. H., & Beutell, N. J. (1985). "Sources of Conflict Between Work and Family Roles." *Academy of Management Review*, 10(1), 76-88.
- Griffeth, R. W., Hom, P. W., & Gaertner, S. (2000). "A meta-analysis of antecedents and correlates of employee turnover: Update, moderator tests, and research implications for the next millennium." *Journal of Management*, 26(3), 463-488.
- Hackman, J. R., & Oldham, G. R. (1976). "Motivation through the design of work: Test of a theory." *Organizational Behavior and Human Performance*, 16(2), 250-279.
- Judge, T. A., Thoresen, C. J., Bono, J. E., & Patton, G. K. (2001). "The job satisfaction-job performance relationship: A qualitative and quantitative review." *Psychological Bulletin*, 127(3), 376-407.

- Ng, T. W. H., Eby, L. T., Sorensen, K. L., & Feldman, D. C. (2005). "Predictors of objective and subjective career success: A meta-analysis." *Personnel Psychology*, 58(2), 367-408.
- Shaw, J. D., Delery, J. E., Jenkins Jr., G. D., & Gupta, N. (1998). "An organization-level analysis of voluntary and involuntary turnover." *Academy of Management Journal*, 41(5), 511-525.