

Explorační analýza dat narození v ČR

Explorační analýza dat (EDA) je proces analýzy a sumarizace datové sady za účelem lepšího pochopení jejích charakteristik a vztahů. Je to zásadní krok v procesu analýzy dat, protože pomáhá identifikovat jakékoli problémy nebo anomálie v datech a pomáhá vytvářet hypotézy o datech, které lze dále testovat a zpřesňovat. EAD se obvykle provádí na začátku procesu analýzy dat, poté, co byla data shromážděna, a před provedením modelování nebo testování hypotéz.

Provedení EDA na začátku procesu analýzy umožní lépe porozumět datům a identifikovat jakékoli potenciální problémy, které mohou ovlivnit přesnost analýzy. Pomáhá také při výběru statistických technik a modelů, které mohou být pro data vhodné, a může pomoci určit směr analýzy.

Existuje mnoho technik, které lze použít pro průzkumnou analýzu dat, včetně:

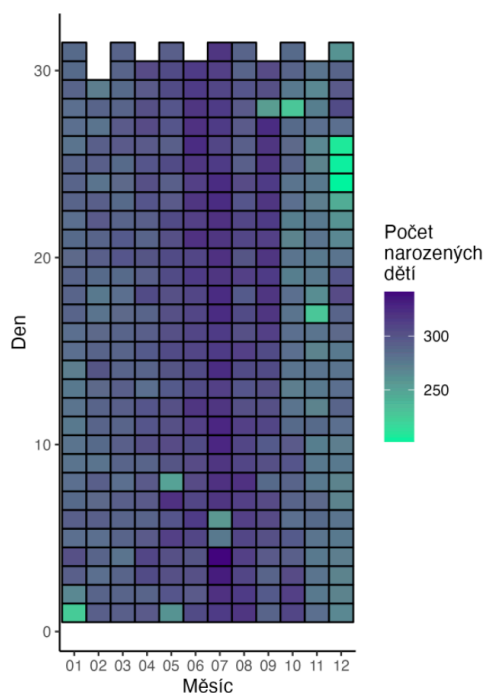
Deskriptivní statistika: Jedná se o výpočet souhrnných statistik, jako je průměr, medián, modus, směrodatná odchylka a kvartily, které pomáhají popsat celkové charakteristiky dat.

Vizualizace: Zahrnuje vytváření grafů a grafů pro vizualizaci dat, což může pomoci identifikovat vzory a trendy v datech. Mezi běžné typy vizualizací patří bodové grafy, spojnicové grafy, sloupcové grafy a histogramy.

Transformace dat: Zahrnuje použití matematických nebo statistických technik na data, aby se extrahovaly užitečné informace nebo aby byla data vhodnější pro analýzu.

Aplikace využívá sady dat o počtu narozených dětí v ČR v letech 2000 až 2020. Na této sadě může zkoumat některé potenciální otázky:

- Existují nějaké trendy v počtu narozených dětí v čase, jako je sezónnost?
- Jsou nějaké rozdíly v počtu narozených dětí v různé dny v týdnu nebo během různých měsíců v roce?
- Existují vnější faktory, jako jsou ekonomické podmínky nebo sociální trendy, souvisí s počtem narozených dětí?
- Obsahuje sada extrémní hodnoty, které mohou naznačovat, že jde o chyby?



Data zobrazují průměr dny v letech 2000 až 2020

Explorační analýza dat je proces analýzy a datové sady za účelem lepšího pochopení jejích charakteristik. Pomáhá identifikovat problémy nebo anomálie v datech a pomáhá vytvářet hypotézy o datech, které lze dále testovat a zpřesňovat. EAD se obvykle provádí na začátku procesu analýzy dat.

Zde jsou data o počtu narozených dětí v ČR v letech 2000 - 20. Může zkoumat například otázky: Existují trendy v počtu narozených dětí, jako je sezónnost? Jsou nějaké rozdíly v počtu narozených dětí v různé dny v týdnu nebo během různých měsíců v roce? Rodí se méně dětí v pátek 13? Vlevo je souhrn dat dle dnů v roce. Pomocí tlačítka lze zvolit jiný pohled na data

Vyberte náhled

Dny v týdnu

