# 2. Vizualizace dat

**Zadání:**

V jednom ze cvičení jste probírali práci s moduly pro vizualizaci dat. Mezi nejznámější moduly patří matplotlib (a jeho nadstavby jako seaborn), pillow, opencv, aj. Vyberte si nějakou zajímavou datovou sadu na webovém portále Kaggle a proveďte datovou analýzu datové sady. Využijte k tomu různé typy grafů a interpretujte je (minimálně alespoň 5 zajímavých grafů)​. Příklad interpretace: z datové sady pro počasí vyplynulo z liniového grafu, že v létě je vyšší rozptyl mezi minimální a maximální hodnotou teploty. Z jiného grafu vyplývá, že v létě je vyšší průměrná vlhkost vzduchu. Důvodem vyššího rozptylu může být absorpce záření vzduchem, který má v létě vyšší tepelnou kapacitu.

**Řešení:**

Pro tuto úlohu jsem použil datovou sadu o diskografii kapely The Cure z webu. K analýze a vizualizaci dat využívám knihovnu **pandas** pro manipulaci s daty a **matplotlib** pro tvorbu grafů.

1. Načtení dat a předzpracování

- Data se načtou z CSV souboru a celé datumy z „album\_release\_date“ převedu na rok „album\_release\_year“). Pro mé účely je to dostačující přesnost a navíc se tím i zlepší čitelnost grafů.

2. Vztah mezi tóninou (key) a popularitou skladeb

- Bar plot ukazuje průměrnou popularitu skladeb v různých tóninách. Tento graf nám umožňuje vidět, zda existuje vztah mezi tóninou skladby a její oblíbeností.

3. Histogram délky skladeb

- Histogram zobrazuje rozložení délky skladeb v minutách. Z něj lze vyčíst, jaká je typická délka skladby.

4. Podíl instrumentálních skladeb

- Koláčový graf ukazuje procentuální podíl instrumentálních skladeb oproti skladbám s vokály, což poskytuje přehled o tom, kolik skladeb je čistě instrumentálních.  
  
5. Vztah mezi hlasitostí a rokem vydání

- Bar plot zobrazuje průměrnou hlasitost skladeb v různých letech. Tento graf může naznačovat změny v produkčním stylu kapely během jejich kariéry. Zde předpokládám, že hlasitost skladeb v CSV souboru byla měřena na souborech s normalizovanou hlasitostí.  
  
6. Vztah mezi tanečností (danceability, jak dobře se na danou skladbu tancuje) a popularitou skladeb

- Scatter plot vizualizuje vztah mezi tanečností skladeb a jejich oblíbeností, což může poskytnout informace o tom, zda tanečnost skladeb ovlivňuje jejich popularitu.