

Pierwszy etap: Zrozumienie problemu + Zrozumienie danych

Oprócz Państwa danych i danych ogólnych (imię, nazwisko, numer indeksu, data, nazwa sprawozdania), sprawozdanie powinno zawierać:

- ogólny opis zbioru danych,
- określenie celu eksploracji i kryteriów sukcesu,
- charakterystyka zbioru danych (pochodzenie, format, liczba przykładów, czy składa się z jednego zbioru, czy z kilku)
- opis atrybutów (nazwy, typy - nominalne/numeryczne, co oznaczają, co oznaczają wartości - jednostka miary, wartości specjalne),
- wyniki eksploracyjnej analizy danych (EDA):
 - ◆ rozkłady wartości atrybutów,
 - ◆ korelacje pomiędzy wartościami atrybutów,
 - ◆ wstępne ustalenia dotyczące zawartości zbioru,
- uwagi nt. jakości danych:
 - ◆ dane brakujące,
 - ◆ punkty oddalone,
 - ◆ dane niespójne,
 - ◆ dane niezrozumiałe,
- opis wyników EDA w odniesieniu do celów eksploracji (czy dane są wystarczające),
- ewentualna rewizja celów.