Opis analizy danych

# Opis danych

## Oddziały2007.xsl

### Ilość atrybutów: 2

* NrOddz (INT)
* SkrNazwaOdz (STRING)

Każdy z powyższych atrybutów zawiera 45 wartości

### Uwagi:

* Dla optymalniejszego, a także łatwiejszego działania bazy warto by dodać ID oddziału, a także skróty nazw oddziałów.
* Dane wprowadzone w arkuszu, nie zawsze są prawidłowe np. nr 45 – Kierowcy – nie jest to nazwa oddziału.
* Brak stosowania uregulowanych skrótów – np. specjalizacji danych oddziałów, w różnych miejscach są różnie reprezentowane
* W bazie może powstać problem z polskimi znakami
* W cellach przechowywane są dane, które są zbędne np. powtarzanie słowa oddział – skoro cała kolumna mówi o oddziałach, stosowanie słowa oddział jest zbędne

### Proponowane rozwiązania:

* Dodanie ID oraz skrótów oddziałów
* Przefiltrowanie danych znajdujących się w kolumnie SkrNazwaOdz oraz zamiana ich na prawidłowe, np. Kierowcy na Transport, gdzie pełna nazwa byłaby Oddział Transportowy tzn. sekcja szpitala zajmująca się transportem pacjentów itp.
* Jeśli chodzi o polskie znaki, najlepiej byłoby usystematyzować pisownie, do nie używania polskich znaków

## Pracownicy2007.xsl

### Ilość atrybutów: 4

* IdPrac(INT) – 1205 różnych wartości
* RokUr (INT) – 56 różnych wartości
* Wykształcenie (STRING) – 15 wartości
* Płeć (CHAR) – 3 typy – K, M, ‘ ‘ – 3 różne wartości

### Uwagi:

* Mamy to ponownie do czynienia z problemem polskich znaków
* Brak usystematyzowania skrótów wykształcenia – to samo wykształcenie zapisywane inaczej
* 36 osób nie posiada żadnego wykształcenia

### Proponowane rozwiązania:

* Problem polskich znaków, usystematyzować jak w pliku oddzialy.xls
* Usystematyzować sposób zapisy wykształcenia
* W puste komórki uzupełnić słowem null, tak aby było wiadome, że dana osoba nie posiada wykształcenia, tak aby nie mylić z tym, że dana nie została jeszcze uzupełniona.

## Wynagrodzenia2007.xsl

### Ilość atrybutów: 15

* IdPrac(INT) – ponad 1700 różnych wartości
* RokMc ( INT ) – liczba 4 cyfrowa, gdzie pierwsze 2 cyfry oznaczają rok a następne miesiąc, jest tutaj 10 różnych wystąpień
* GrupaZaw (STRING) – 10 różnych wartości
* NrOddz(INT) – 45 wartości (0-44)
* Brutto (DOUBLE) – (0,00 – 59906,40)
* Zasadnicze (DOUBLE) - (-466,17 - 11286,00)
* Nadgodz (DOUBLE) - (0,00 - 4559,94)
* StażPracy (DOUBLE) - (-60,90 - 1102,20)
* Funkcyjny (DOUBLE) - (0,00 - 2204,40)
* DodStop (DOUBLE) - (0,00 - 1575,00)
* Inne (DOUBLE) - (-1128,56 - 56906,40)
* WynChor(DOUBLE) - (-598,20 - 7846,20)
* ZasilkiZUS (DOUBLE) - (0,00 - 6398,84)
* DniZasZUS (INT) - (0 - 111)
* DniChor (INT) - (0 - 33)

### Uwagi:

* W arkuszu Wynagrodzenia2007.xls jest większa liczba pracowników niż w arkuszu Pracownicy2007.xls
* Słabo czytelne grupa zawodowa
* Słabo czytelny rok i miesiąc wynagrodzenia
* Zasadnicze i chorobowe wynagrodzenia przyjmują wartości ujemne
* Brak usystematyzowania w nazwach grup zawodowych
* Dane w arkuszu wpływają na siebie wzajemnie tzn. DnChor wpływają na WynCho, ktre jest dodawane do Inne, które następnie jest dodawane do Brutto, DniZasZUS wpływają na Zasilki ZUS, które nie mają wpływu na resztę wynagrodzeń
* Część pracowników podwójnie otrzymuje wynagrodzenia w tych samych miesiącach, część zaś ma pominięte wynagrodzenia w niektórych miesiącach.

### Proponowane rozwiązania:

* Usystematyzowanie nazewnictwa, skrótów itp.
* Naprawienie wartości ujemnych, przez zmianę ich znaków
* Polepszenie zapisu rok i miesiąc np. na format RRRR-DD, ułatwi to znacznie prace

## ZATR0701.xsl

### Ilość atrybutów: 12

* IdPrac (INT) – około 1000 różnych wartości
* ZatrudOd(DATA)
* ZatrudDo(DATA)
* Stanowisko(STRING)
* E(Pełny etat/G(godziny) (CHAR) – występują dwa typy E lub G
* GodzEtatu (DOUBLE)
* GodzZatr (DOUBLE)
* PłacaEtat(DOUBLE)
* PłacaMies(DOUBLE)
* LatStażu (INT)
* DodFunkc[%] (Double)
* DodZaStopień(za cały Etat) (Double

### Uwagi:

* Puste komórki w ZatrudOd, ZatrudDo ( tu warto wstawić wartość null), GodzZatr
* Błędy w danych, pracownicy, którzy byli zatrudnieni do 2007 roku, w 2008 roku otrzymali wynagrodzenie
* Brakuje nazw stanowisk w komórkach w kolumnie Stanowiska
* Ogólnie występujący problem z polskimi znakami
* Brak usystematyzowania skrótów
* Błędy w zapisie stanowisk

### Proponowane rozwiązania:

* Stworzenie skryptu, który naprawi błędnie zapisane stanowiska
* Usystematyzowanie skrótów
* Zastosowanie wpisów bez polskich znaków, stworzenie skryptu, który będzie konwertował wpisy z polskimi znakami na takie, które nie posiadają wystąpień polskich znaków
* Jeśli chodzi o błędy w zapisie pomiędzy arkuszem Zatr0701 a arkuszem z wynagrodzeniem, dobrze byłoby ustalić, która dana została błędnie wprowadzona.

W pozostałych arkuszach dotyczących zatrudnienia występują analogiczne problemy jak przy ZATR0701.xls