# Analityk w HR Consulting

Twój klient jest dużą korporacją wielonarodową i ma 9 dużych działów. Jednym z problemów, z jakimi boryka się Twój klient, jest znalezienie odpowiednich osób do awansu i przygotowanie ich na czas. Obecnie proces jest następujący:

* Najpierw identyfikują grupę pracowników na podstawie rekomendacji/wyników w przeszłości
* Wybrani pracownicy przechodzą osobny program szkoleń i oceny dla każdego pionu. Programy te są oparte na wymaganych umiejętnościach każdej branży
* Na koniec programu, w oparciu o różne czynniki, takie jak wyniki szkolenia, ukończenie KPI (pod uwagę brane są tylko osoby z KPI powyżej 60%) itp., pracownik otrzymuje awans

W przypadku powyższego procesu ostateczne awanse są ogłaszane dopiero po ocenie, co prowadzi do opóźnienia przejścia do nowych ról. Dlatego firma potrzebuje Twojej pomocy w identyfikacji kwalifikujących się kandydatów w określonym punkcie kontrolnym, aby mogli przyspieszyć cały cykl awansu.

# Dane

**Train:**PlikCSVzawierający dane o pracownikiem z nadanym statusem przyznania bądź nie awansu is\_promoted '

|  |  |
| --- | --- |
| **Zmienna** | **Opis** |
| employee\_id | ID pracownika |
| department | Wydział |
| region | Region |
| education | Poziom wykształcenia (Below Secondary, Bachelor's, Master's & above) |
| gender | Płeć (f/m) |
| recruitment\_channel | Kanał rekrutacji (sourcing/referred/other) |
| no\_of\_trainings | Liczba szkoleń w poprzednim roku oceny |
| age | Wiek pracownika |
| previous\_year\_rating | Ocena pracownika w poprzednim roku |
| length\_of\_service | Czas pracy w latach |
| KPI | Czy pracownik spełnił co najmniej 80% kryteriów wskaźników aktywności (Key performance Indicators) (No/ Yes) |
| awards\_won | Czy pracownik dostał nagrody w poprzednim roku (No/ Yes) |
| avg\_training\_score | Średni wynik w szkoleniach w obecnym roku |
| is\_promoted | Rekomendacja do awansu |

# Zadania

Scharakteryzuj zmienne *previous\_year\_rating, length\_of\_service* w zakresie: minimum, pierwszy kwartyl, medina, średnia, trzeci kwartyl, maksimum, odchylenie standardowe, wybrana miarę asymetrii i koncentracji w każdej z grup: według płci (*gender*) i według awansu (*is\_promoted*).

Zlicz ile pracowników jest w każdym departamencie a ilu w każdym regionie (tabele)

Oblicz jak często awans dostawały kobiety a jak często mężczyźni*.* (tabela liczności).

Oblicz jak często awans dostawały osoby wg kanału rekrutacji (*recruitment\_channel*)(tabela liczności).

Dla zmiennych *no\_of\_trainings* i *length\_of\_service* zrób wykres ramka-wąsy (boxplot). Niech zmienna *gender* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla zmiennej *avg\_training\_score* zrób wykres ramka-wąsy (boxplot). Niech zmienna *education* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla pary zmiennych *previous\_year\_rating*, *avg\_training\_score* zrób wykres rozrzutu. Niech zmienna *is\_promoted* będzie tworzyła serie. Dodał tytuł, opis osi, legendę.

Dla pary zmiennych *no\_of\_trainings, age* zrób wykres rozrzutu. Niech zmienna *gender* będzie tworzył serie. Dodał tytuł, opis osi, legendę.

Zbuduj 95% przedział ufności dla średniej i wariancji dla zmiennych *previous\_year\_rating*, *avg\_training\_score* osobno dla kobiet i mężczyzn.

Sprawdź czy średni poziom zmiennej *avg\_training\_score* różni się pomiędzy grupą osób, które otrzymały awans i nie (zmienna *is\_promoted*)

Sprawdź czy wariancja zmiennej *previous\_year\_rating* różni się pomiędzy grupą osób, które otrzymały pożyczkę i nie (zmienna *is\_promoted*)

# Model

Zbuduj model wstępnego decydowania, którym pracownikom firma powinna przyznać awans. Oceń jakość modelu stosując walidację krzyżową, macierz pomyłek i statystyki na jej bazie tj. accuracy, precision, recall i f1-score. Spróbuj różnych modeli.