Firma dysponuje bazą 80 000 klientów i udostępniła ich dane dotyczące:

* roku urodzenia,
* płci,
* roku otrzymania prawa jazdy,
* marki samochodu,
* typu samochodu (nadwozia),
* roku produkcji samochodu,
* pojemności silnika,
* mocy silnika,
* wartości rynkowej samochodu,
* lat historii ubezpieczeniowej,
* liczby wypadków/stłuczek w ostatnich pięciu latach przed zawarciem ostatniej polisy ubezpieczeniowej
* czy klient przyniósł szkodę dla firmy ubezpieczeniowej (zmienna binarna {0,1}; tzn. czy ryzyko wypadku zmaterializowało się).

Twoim zadaniem jest oszacowanie zagrożenia wypadkowego nowych potencjalnych 32 000 klientów. W tym celu zbuduj model predykcyjny.

Uwaga:

Za przypadek pozytywny uznajemy zdarzenie wypadku samochodowego, co jest zakodowane wartością {1}, a brak zdarzenia wypadku wartością {0}.

Zatem Predicted Negatives to oczekiwany przez nas bezwypadkowy klient, a False Negatives, to klient spośród tych oczekiwanych, który faktycznie przyniósł szkodę firmie.