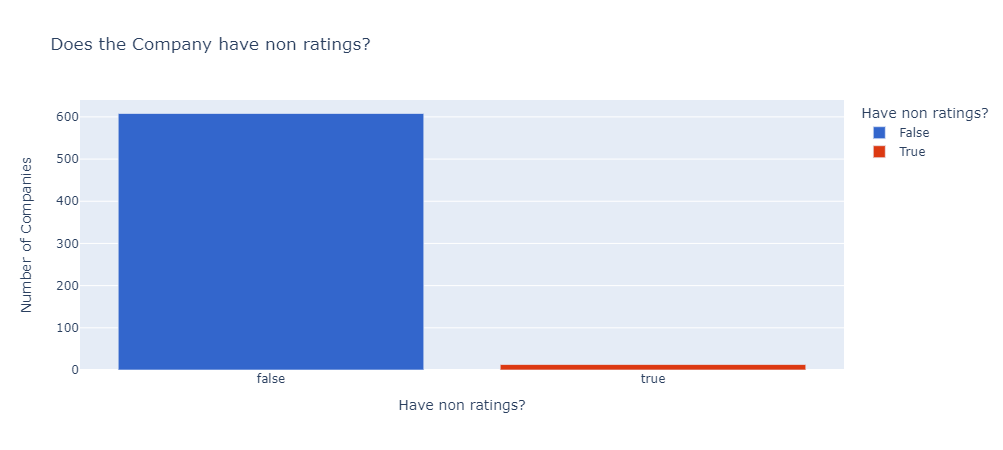
Pobieranie danych

# Analiza EDA

## Czy istnieją spółki, które nie mają żadnych ocen?

Naszą analizę rozpoczęliśmy od sprawdzenie czy istnieją spółki, które ani razu nie zostały ocenione przez ekspertów.

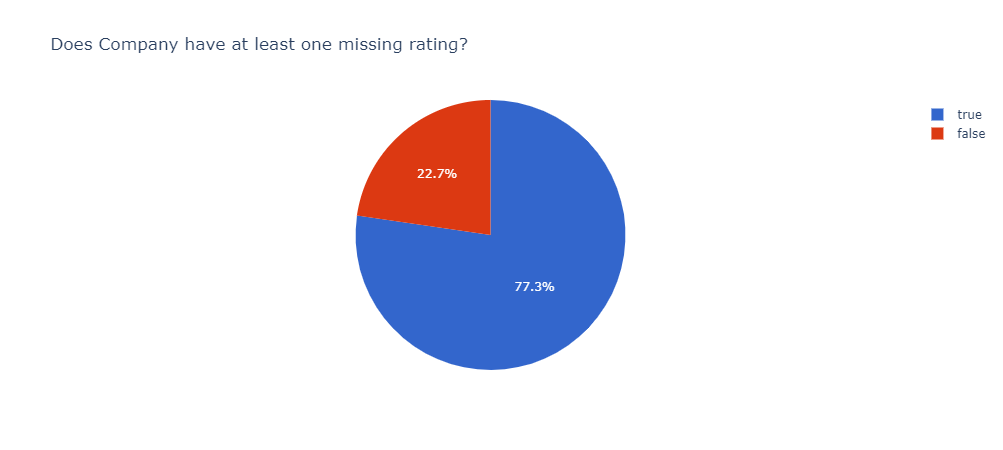


Okazuje się, że istnieje 13 spółek dla których nie ma żadnych ocen. Zostały one przez nas usunięte.

## Czy istnieje duplikaty w danych?

Zidentyfikowaliśmy 115 duplikatów w wierszach i w każdym przypadku postanowiliśmy zostawić pierwszy z nich. Liczba ta może wynikać nie tylko z własności danych pierwotnych, ale także z zaimplementowanych procedur pobierania danych giełdowych.

## Jaki procent spółek posiada braki w ocenach?

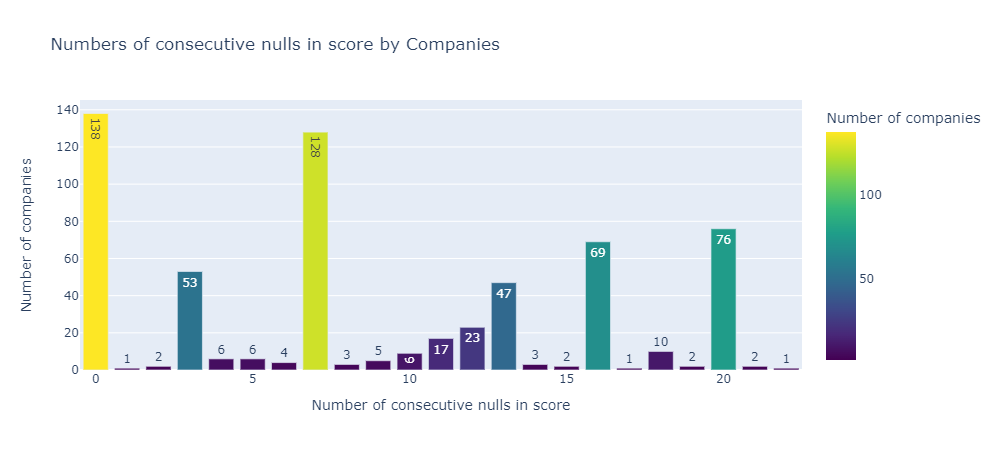


Okazuje się, że zdecydowana większość ze spółek posiada braki w danych. W związku z tym decyzja o usunięciu lub zostawieniu a także możliwa strategia wypełnienia braków danych stanowi istotny element projektu.

Natomiast po usunięciu spółek, które ani razu nie zostały ocenione przez ekspertów okazało się, że wszystkie pozostałych dostępne są przynajmniej częściowe informacje o zwrotach.

## Rozkład następujących po sobie wartości null

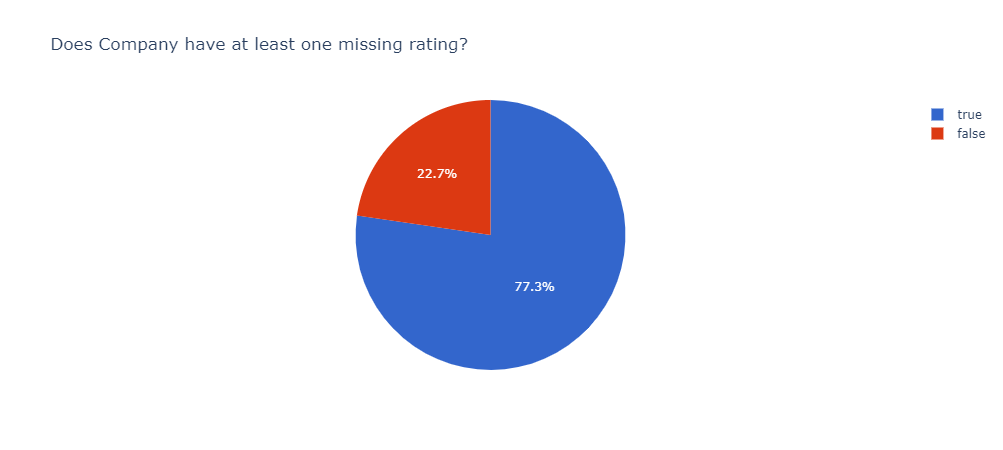
Uznaliśmy, że najbardziej kłopotliwe w dalszej analizie są następujące po sobie braki ocen. W związku z tym dla każdej spółki obliczyliśmy maksymalną liczbę następujących po sobie wartości typu null.



Postanowiliśmy usunąć wszystkie spółki, które nie posiadały ocen 17 lub więcej razy z rzędu.

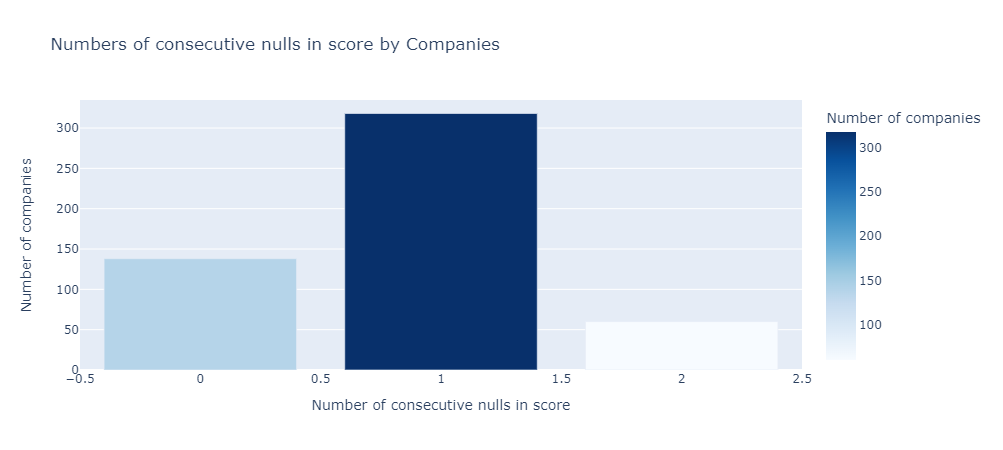
## Jaki procent braków ocen dotyczy wyłącznie jednego przedziału czasowego?

Następnie zweryfikowaliśmy jaki procent wszystkich braków ocen dla każdej spółki dotyczy tylko jednego okresu. Może to świadczyć nie tylko o brakach danych, ale także o tym, że w danym okresie czasu spółka nie była publicznie notowana.



Zdecydowana większość braków ocen dotyczyła jednego konkretnego okresu. Uznaliśmy w takim wypadku za niezasadne uzupełnia tych braków.

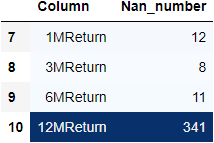
Ostatecznie postanowiliśmy usunąć wszystkie sekwencje trzech lub więcej następujących po sobie braków ocen dla każdej ze spółki. Decyzja ta skutkowała zmniejszenie zbioru dostępnych obserwacji o 23 %.



Zdecydowana większość pozostałych braków dotyczyły pojedynczych ocen. W związku z tym można było dokonać wypełniania brakujących wartości z wykorzystaniem metody forward-fill.

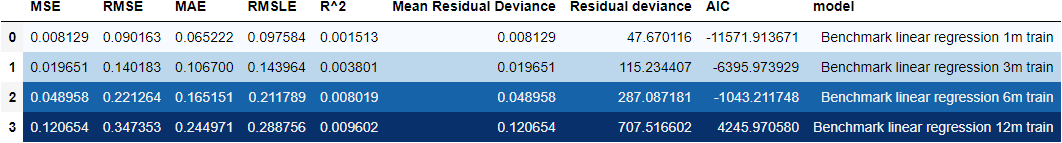
## Braki danych w zwrotach

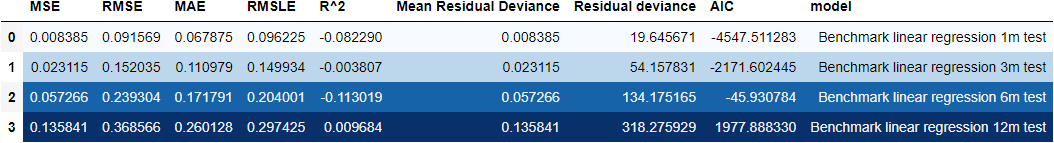
Liczba braków w danych w zwrotach pozyskanych z danych giełdowych okazała się być relatywnie niewielka i związku z tym, wiersze, które zawierały wartości typu Null zostały wyłączone z dalszej analizy.



# Benchmark model

Jako model do przeprowadzenie benchmarku wybrano najprostszy wariat z metod wchodzących w skład uogólnionych modeli liniowych (ang. Generalized regressions models), mianowicie regresję liniową (gausowską). Rezultaty dla wszystkich horyzontów okazały się niezwykle słabe. W niektórych przypadkach R^2 przyjmuje wartości ujemne, co oznacza, że mniejsze wartości błędów można by uzyskać zastępując model najzwyklejszą poziomą linią.





# Wykorzystane metody

# Sposób wyboru modelu

# Wyniki

## Prognozy 3 miesięcznych zwrotów

Słabe wyniki modeli nie zaskakując, gdy weźmiemy pod uwagę, że korelacja pomiędzy ocenami ekspertów a zwrotami jest niewielka.

