

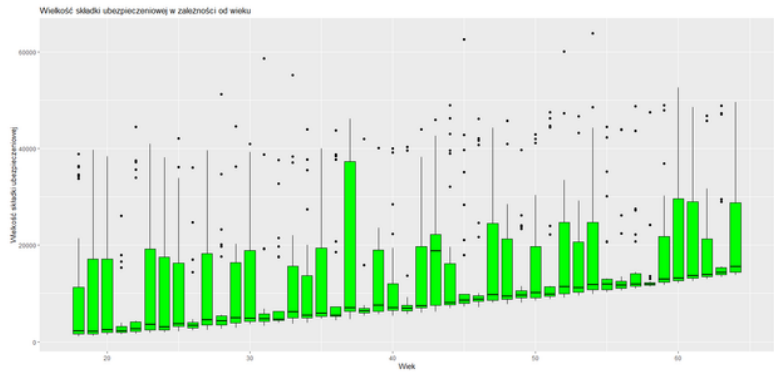
# PODSTAWOWA ANALIZA DANYCH

TEN ZESTAW DANYCH ZAWIERA SZCZEGÓŁOWE INFORMACJE O KLIENTACH UBEZPIECZENIOWYCH, W TYM ICH WIEK, PŁEĆ, INDEKS MASY CIAŁA (BMI), LICZBĘ DZIECI, STATUS PALENIA I REGION ORAZ WYSOKOŚĆ SKŁADEK UBEZPIECZENIOWYCH. POCHODZI ZE STRONY KAGGLE.COM.

PO OCZYSZCZENIU DANYCH Z NIEPEŁNYCH REKORDÓW I PRZEKSZTAŁCENIU CZĘŚCI INFORMACJI OTRZYMANO 1338 REKORDY I 3 KOLUMNY: INDEX, AGE, CHARGES, NA ICH PODSTAWIE BĘDZIEMY ANALIZOWAĆ JAK WIEK MA WPLYW NA WYSOKOŚĆ SKŁADEK UBEZPIECZENIOWYCH.

## BOXPLOT

### I. Wielkość składki w zależności od wieku - boxplot

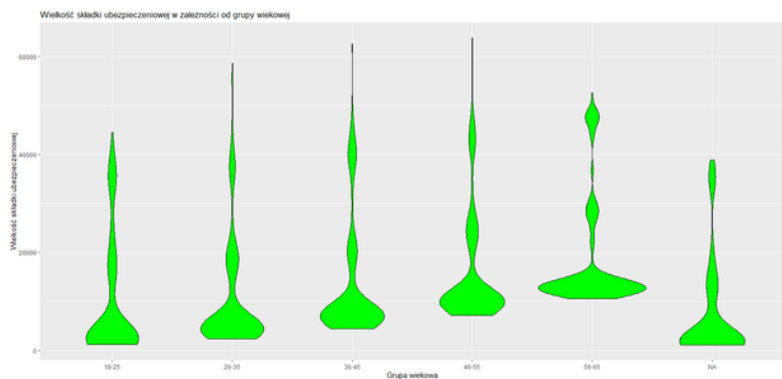


Dane zostały podzielone na grupy w zależności od wieku i każda z nich została wyświetlona na osobnym boxplocie.

Możemy zauważyć, że dla starszych klientów składki są zwykle wyższe niż dla młodszych. Wartości odstające mogą oznaczać, że klienci z tym wiekiem mieli znacznie wyższe składki niż pozostali klienci w tej grupie wiekowej. Wysoki zielony słupki na boxplocie wieku oznaczają, że w tej grupie wiekowej wielkość składki jest wyższa niż w pozostałych grupach wiekowych. Może to sugerować, że klienci w tej grupie wiekowej są bardziej narażeni na ryzyko i dlatego płacą wyższe składki ubezpieczeniowe.

## VIOLINPLOT

### II. Wielkość składki w zależności od wieku - violinplot

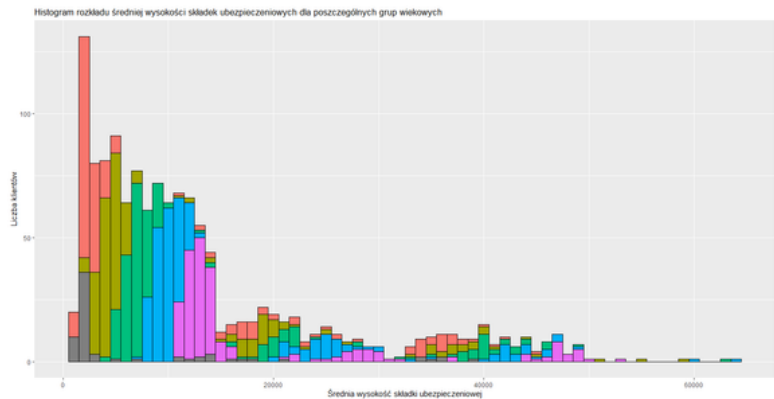


W violinplocie, gęstość rozkładu danych jest przedstawiona przez kształt i rozmiar "skrzypiec". Szerokie miejsca skrzypiec oznaczają, że w danej grupie wiekowej występuje duża liczba klientów o podobnych wartościach składki ubezpieczeniowej. Natomiast wąskie, że w danej grupie jest mniejsza liczba klientów o podobnych wartościach składki.

Długie skrzypce oznaczają, że w danej grupie wiekowej występuje duża rozpiętość wartości składki ubezpieczeniowej i nie ma jednej dominującej wartości. Natomiast krótkie skrzypce oznaczają, że w danej grupie wiekowej występuje mniejsza rozpiętość wartości składki ubezpieczeniowej i jest jedna dominująca wartość.

## HISTOGRAM

### III. Wielkość składki w zależności od wieku - histogram



Ten histogram pozwala na zobrazowanie rozkładu średniej wysokości składek ubezpieczeniowych dla poszczególnych grup wiekowych. Możemy z niego odczytać, jakie są najczęstsze wartości składek dla poszczególnych grup wiekowych oraz jak rozkłada się liczba klientów w poszczególnych przedziałach składek.

Histogram pokazuje nam, że w grupie wiekowej 18-25, średnie składki są najniższe, z kolei w grupie wiekowej 66+ są najwyższe. Również widzimy, że w grupie wiekowej 66+ jest mniej klientów niż w pozostałych grupach. Warto zauważyć, że w każdej z grup wiekowych występuje najwięcej klientów z rozkładem składek ubezpieczeniowych w przedziale 4000-5000\$.

## WNIOSKI:

Kurtoza równa wartości mniejszej niż 3 oznacza, że rozkład jest mniej skupiony niż rozkład normalny. Skośność równa wartości dodatniej oznacza, że rozkład jest skośny w prawo. Analiza danych z tego zestawu sugeruje, że wiek jest jednym z ważniejszych czynników wpływających na wysokość składek ubezpieczeniowych.

Wykresy takie jak boxplot, violinplot i histogram pozwalają na zobrazowanie rozkładu składek dla różnych grup wiekowych. Można zauważyć, że średnie składki dla starszych klientów są wyższe niż dla młodszych, co sugeruje, że starsi klienci płacą wyższe składki niż młodszy.

```
# A tibble: 6 × 7
```

	age_group	median	q0.5	sd	IQR	min	max
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	(18,30]	3578.	3578.	11742.	14195.	1242.	51195.
2	(30,40]	6334.	6334.	11530.	9210.	3260.	58571.
3	(40,50]	8964.	8964.	11693.	11960.	5700.	62593.
4	(50,60]	11938.	11938.	11460.	3466.	8782.	63770.
5	(60,70]	14395.	14395.	11876.	14362.	12558.	49578.
6	<NA>	2198.	2198.	10198.	9639.	1122.	38793.

Variance

Skewness

Kurtosis

146652372

1.512482

1.588954