

Podstawowa Analiza Danych

Analiza danych to proces pozyskiwania, przetwarzania i interpretowania informacji z różnych źródeł danych. Celem analizy jest znalezienie odpowiedzi na pytania badawcze i zrozumienie zjawisk zachodzących w danych.

W tym projekcie będziemy analizować dane dotyczące nawyków korzystania z ekranów w zależności od dnia tygodnia. Dane pochodzą z platformy Kaggle i zostaną załadowane za pomocą R. Następnie przejdziemy przez proces podstawowej analizy statystycznej. Na końcu przygotowujemy raport z wynikami i wnioskami.

DANE:

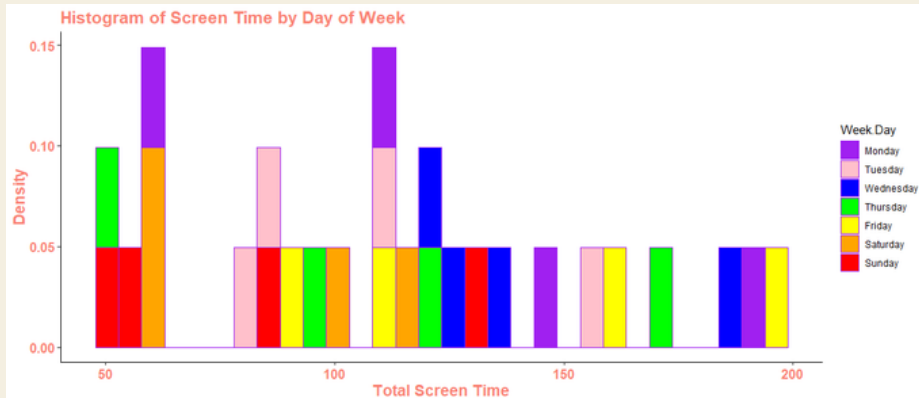
- Data wprowadzenia
- Dzień tygodnia
- Całkowity czas spędzony na telefonie.
- Produktivność: Czas spędzony na działaniach produktywnych, takich jak praca lub zadania szkolne.
- Zdrowie i fitness: Czas spędzony na treningu.

ŹRÓDŁO DANYCH:

www.kagle.com

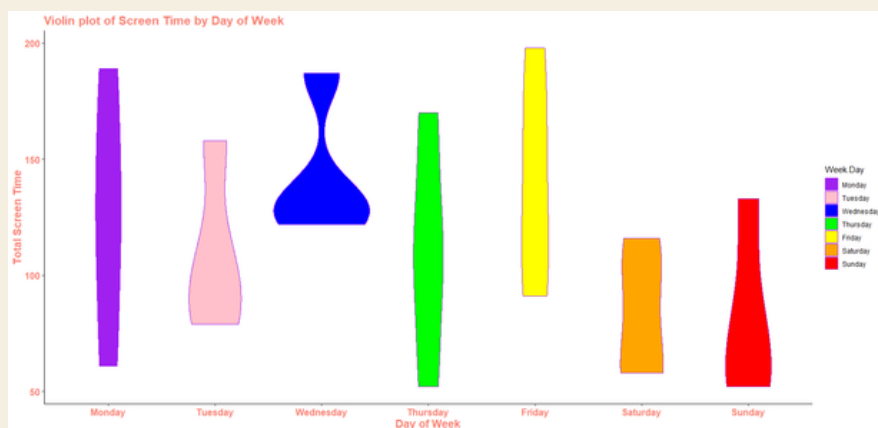
HISTOGRAM

Histogram to wykres, który przedstawia rozkład danych za pomocą słupków. Słupki przedstawiają liczebność danych w określonym przedziale. Histogram pozwala na zobaczenie jak dane są rozłożone oraz zauważenie jakiejś tendencji.



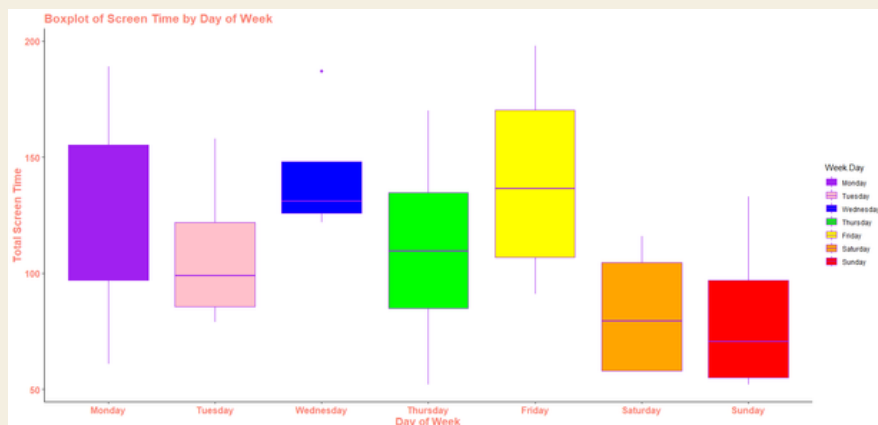
VIOLINPLOT

W violinplocie widzimy prostokąt, który odpowiada przedziałowi między pierwszym i trzecim kwartylem. Po obu stronach prostokąta, widzimy kształt przypominający skrzypce, który przedstawia rozkład danych. Wysokość skrzypiec odpowiada liczebności danych.



BOXPLOT

Jego głównym celem jest przedstawienie rozkładu danych oraz uwypuklenie wartości skrajnych (outlierów) i innych istotnych informacji statystycznych takich jak median, kwartyle, czy minimum i maksimum. Składa się z prostokąta, którego dolna i górna krawędź odpowiada odpowiednio pierwszemu i trzeciemu kwartyłowi. Linia pośrodku prostokąta to mediana.



ESTYMATY PARAMETRÓW

Week.Day	mean	median	sd	IQR
<fct>	<dbl>	<dbl>	<dbl>	<dbl>
Monday	126.	126.	54.2	58.2
Tuesday	109.	99	35.3	36.2
Wednesday	143.	131	30.0	22.2
Thursday	110.	110.	49.4	49.8
Friday	140.	136.	48.3	63.5
Saturday	83.2	79.5	29.8	46.8
Sunday	81.5	70.5	37.3	42

WNIOSKI

Z wykresu boxplot widać, że mediana zmiennej „Total.Screen.Time” jest najwyższa w piątek, następnie w środy, a najniższa w niedzielę. Wykres violinplot potwierdza obserwacje poczynione na wykresie boxplot, a także pokazuje rozkład danych dla każdego dnia tygodnia. Widzimy, że rozkład jest gęstszy wokół mediany w środy, a w innych dniach zdecydowanie bardziej równomierny. W dni powszednie więcej osób spędza więcej czasu przed ekranem niż w weekendy. W piątek jest najwięcej osób, które spędzają dużo czasu przed ekranem, a w niedzielę najmniej. Czas spędzany przed ekranem jest skupiony w niższych wartościach we wtorek i niedzielę, a w wyższych wartościach w piątek. Należy pamiętać, że ta analiza opiera się na małym zbiorze danych i może nie być ogólnie stosowalna dla całej populacji.