

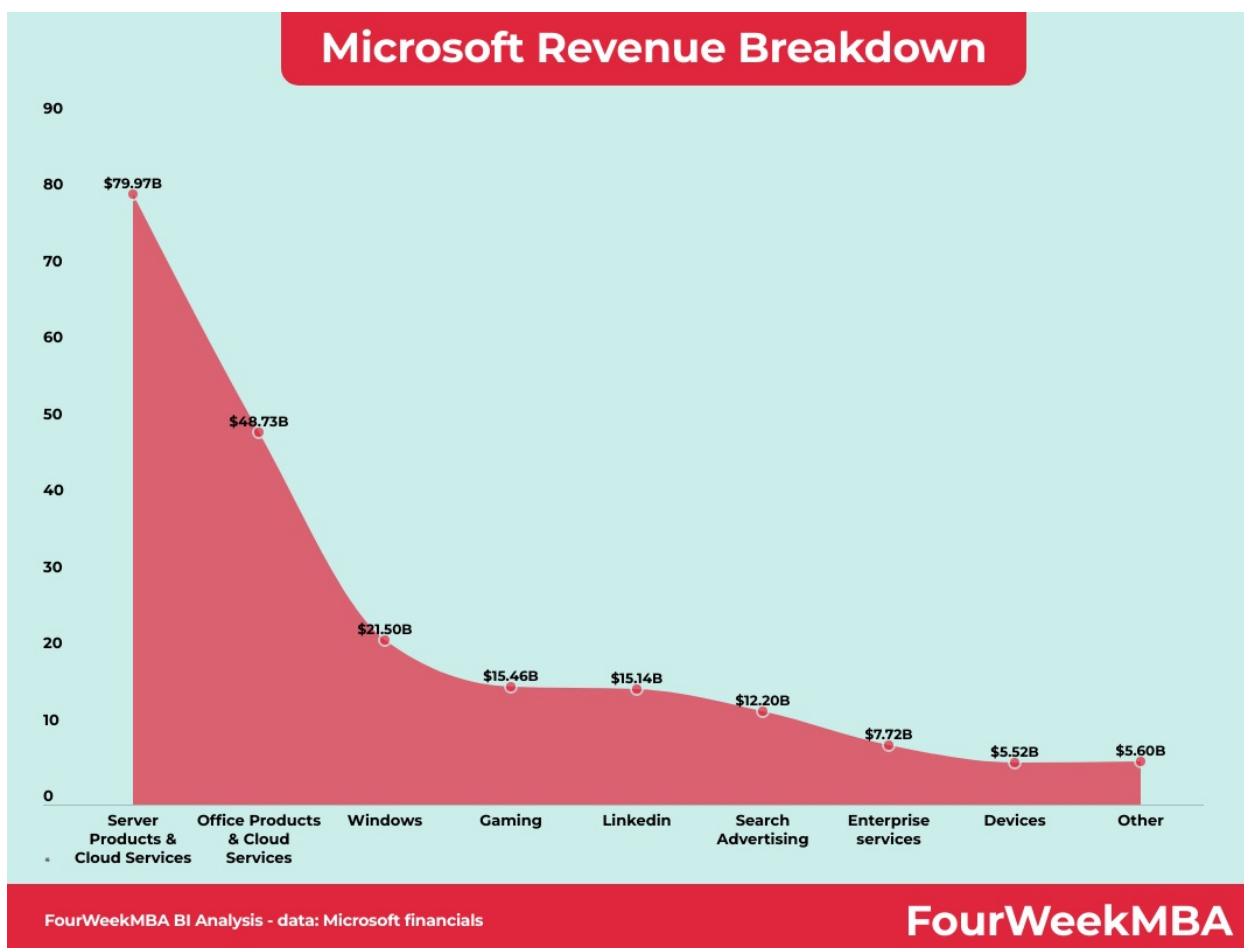
Praca domowa 3.

Weronika Orzechowska (327461)

Wizualizacja źródłowa

Analizowanym wykresem jest Podział przychodów firmy Microsoft opublikowany 20 marca 2024 roku przez FourWeekMBA.

<https://fourweekmba.com/microsoft-revenue-breakdown/>



Głównym elementem wymagającym poprawy jest typ wykresu. Zastosowano wykres typu density plot, gdzie zmienne na osi X ewidentnie nie są takiego typu.

Patrząc początkowo na ten wykres można odnieść błędne wrażenie co ma przedstawiać. Dodatkowo kolejne kategorie nie są uporządkowane w kolejności malejącej, pod względem wartości na osi OY.

Na wykresie nie są zaznaczone osie oraz nie są one podpisane, co może doprowadzić do błędnej analizy wyników. Dodatkowo etykiety przy punktach wydają się być mało czytelne.

Kolorystyka jest miła dla oka, więc ją zostawimy na wykresie końcowym. Użyta czcionka jest poprawna, ale ciężko określić jej nazwę, więc zostaniemy przy tej domyślnej.

Przygotowanie danych

Będziemy korzystać z trzech następujących bibliotek:

```
library(dplyr)
library(stringr)
library(ggplot2)
```

dplyr, stringr posłużą nam do edycji danych, natomiast ggplot2 do stworzenia wizualizacji.

Korzystając z dokładnego zaznaczenia wartości w kolejnych punktach, stworzymy zbiór wynikowy, na podstawie którego stworzymy poprawny już wykres.

```
data <- data.frame(category = c("Server Products & Cloud Services",
                                "Office Products & Cloud Services",
                                "Windows",
                                "Gaming",
                                "Linkedin",
                                "Search Advertising",
                                "Enterprise services",
                                "Devices",
                                "Other"),
                    value = c(79.97, 48.73, 21.5, 15.46, 15.14, 12.2, 7.72, 5.52, 5.6))
```

Dla ułatwienia wartości zapisujemy już jako miliardy dolarów.

Żeby później łatwiej było stworzyć wykres zmodyfikujemy jeszcze trochę tą ramkę danych:

- tworzymy osobną kolumnę, która będzie przechowywać etykiety do punktów na wykresie
- sortujemy wartości malejąco po kolumnie value
- ustawiamy zawijanie tekstu w zmiennej category
- faktoryzujemy wartości w zmiennej category

```
data <- data %>% mutate(val_str = paste0("$", as.character(value), "B")) %>%
  arrange(desc(value))
data$category = str_wrap(data$category, width = 15)
data$category <- factor(data$category, levels = unique(data$category))
```

Utworzona wizualizacja

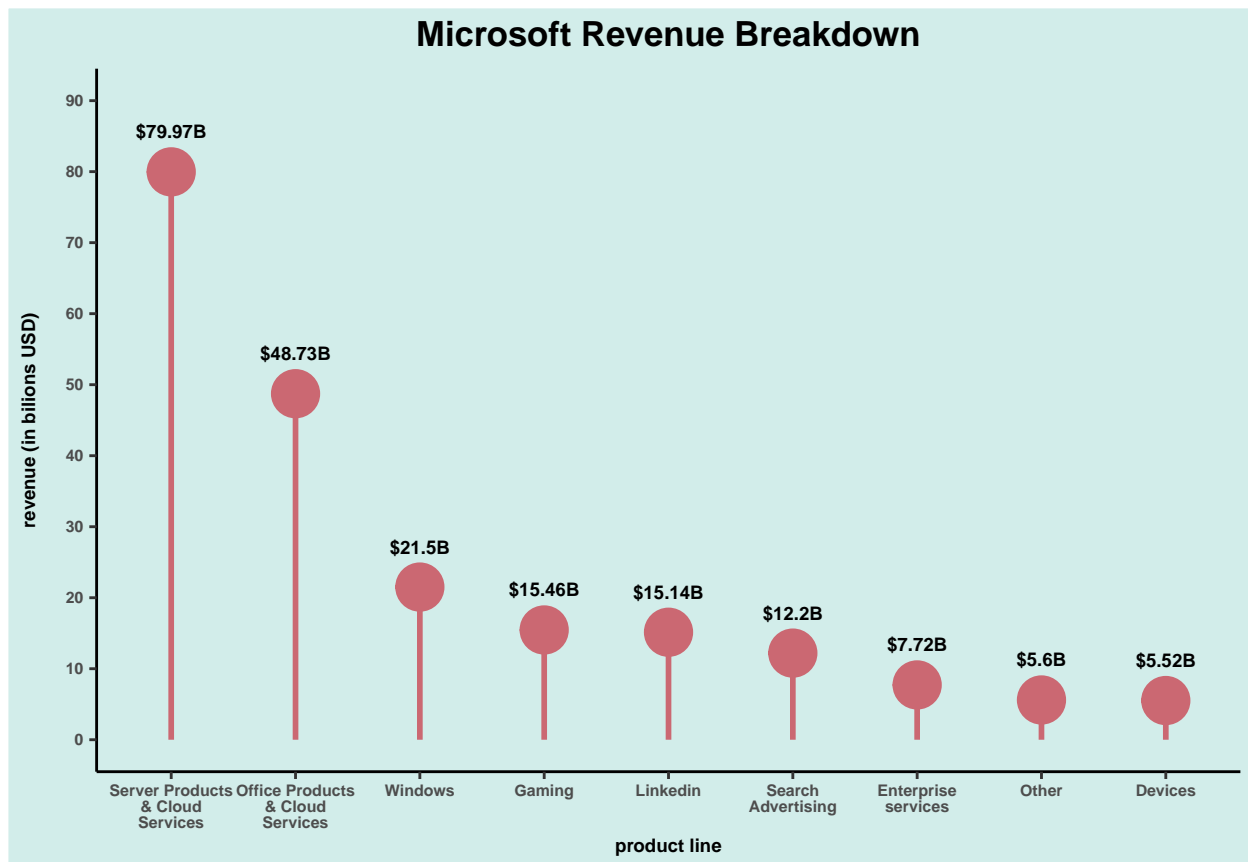
W wizualizacji wykorzystamy wykres typu lolipop. Pozwoli on na zaznaczenie wyraźnego podziału między kolejnymi kategoriami i będzie wskazywał punktowo wartości na osi OY, które dodatkowo podpiszemy tak jak na wykresie pierwotnym.

Na wykresie dodamy również opisy osi, które zostały wywnioskowane na podstawie artykułu, do którego załączony był analizowany wykres.

```

ggplot(data, aes(x=category, y=value)) +
  geom_segment(aes(x=category, xend=category, y=0, yend=value),
    linewidth = 1, color = "#cb6872") +
  geom_point(size=6, color="#cb6872", shape=19, stroke=2) +
  geom_text(aes(label=val_str), vjust=-2.5, hjust=0.5,
    color="black", size = 2.3, fontface = "bold") +
  scale_y_continuous(breaks = seq(0, 90, by = 10), limits = c(0, 90)) +
  labs(x = "product line", y = "revenue (in billions USD)",
    title = "Microsoft Revenue Breakdown") +
  theme_classic() +
  theme(plot.background = element_rect(fill = "#d2edea"),
    panel.background = element_rect(fill = "#d2edea"),
    axis.text.x = element_text(size = 5.5, face = "bold"),
    axis.text.y = element_text(face = "bold", size = 5.5),
    axis.title.x = element_text(face = "bold", size = 7),
    axis.title.y = element_text(face = "bold", size = 7, margin = margin(r=10)),
    plot.title = element_text(face = "bold", hjust = 0.5))

```



Stworzony wykres jest lepszy ponieważ zaznaczony jest podział między kategoriami. Osoba czytająca wykres nie zostanie już zmylona rzekomą ciągłością zmiennych na osi OX. Dodatkowo podpisane są osie, więc dokładnie wiadomo czego dotyczą (nie jest w zasadzie oczywiste czego dotyczy oś OX, patrząc na pierwotny wykres).

Dodatkowo wartości są uporządkowane teraz w kolejności malejącej, więc czytelnik od razu widzi statystyki i zależności między kolejnymi produktami: który jest najbardziej, a który najmniej przychodowy.

Wartości w kolejnych punktach są też teraz bardziej czytelne.