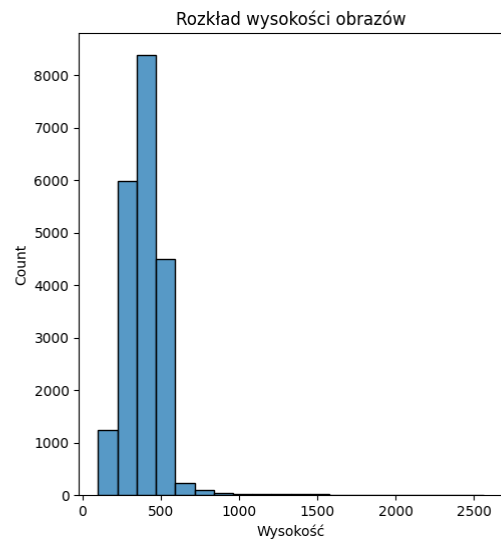
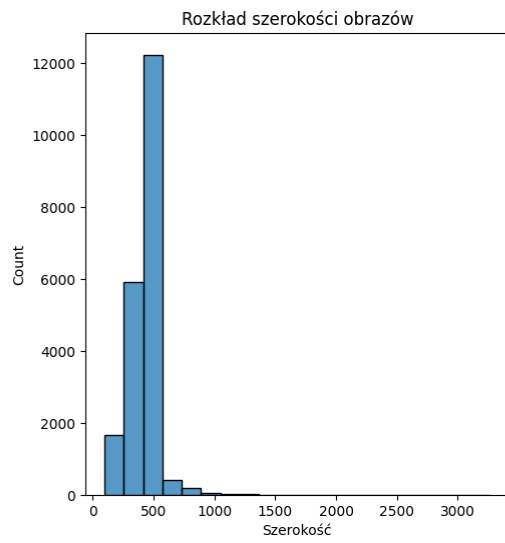


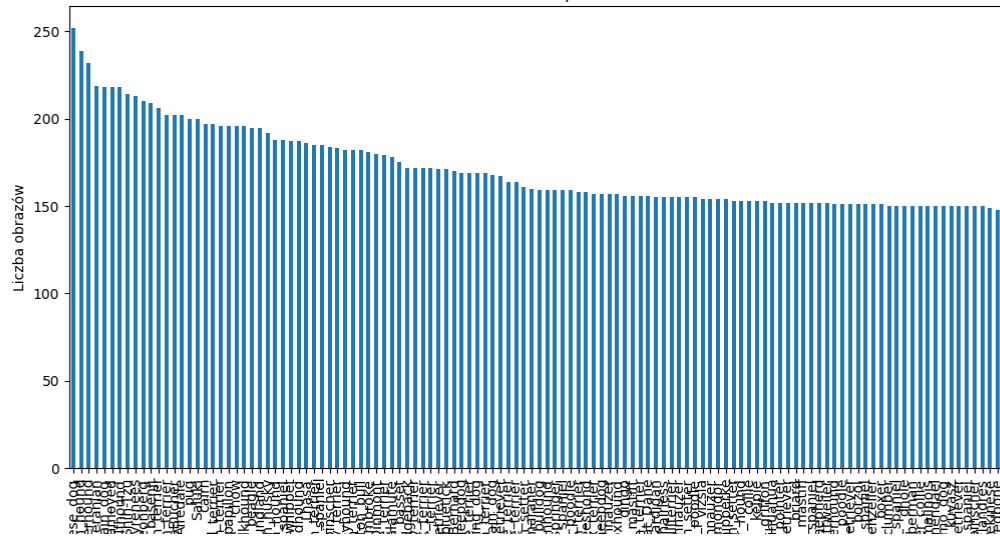
1. **Struktura danych:** Dane zawierają 9 kolumn, z czego większość to zmienne liczbowe (szerokość, wysokość, głębokość obrazu oraz współrzędne prostokąta ograniczającego xmin, ymin, xmax, ymax). Kolumna breed reprezentuje kategorię, czyli rasę psa, którą będziemy przewidywać.
2. **Brakujące wartości:** Wszystkie kolumny są kompletne, co oznacza, że nie ma brakujących wartości. Jest to duży plus, ponieważ brakujące dane nie będą wpływać na trening modelu ani nie będą wymagać dodatkowych zabiegów imputacji.
3. **Macierz korelacji:** Zgodnie z macierzą korelacji:
 - Zauważalna jest wysoka korelacja między szerokością (width) a xmax oraz między wysokością (height) a ymax. Jest to naturalne, gdyż xmax i ymax wyrażają położenie graniczne, które często zależy od wymiarów obrazu.
 - depth jest zmienną o niskiej korelacji z pozostałymi cechami, co może sugerować jej mniejsze znaczenie przy klasyfikacji ras psów.
4. **Analiza zmiennych numerycznych:**
 - Rozkład szerokości i wysokości obrazów jest podobny, jednak widocznych jest wiele wartości odstających, szczególnie dla szerokości.
 - Wykres pudełkowy wskazuje, że większość obrazów ma szerokość i wysokość w stosunkowo wąskim zakresie, ale niektóre obrazy są znacznie większe, co może sugerować potencjalne problemy z homogenicznością rozmiaru obrazów.
5. **Rozkład ras psów:** Rozkład liczby obrazów dla różnych ras jest względnie zrównoważony, choć niektóre rasy mają nieco więcej próbek niż inne. Jest to istotne z punktu widzenia modelu, ponieważ równomierny rozkład klas pomaga uniknąć uprzedzeń modelu względem jednej konkretnej klasy.
6. **Automatyczny raport:** W raporcie wygenerowanym za pomocą narzędzia (np. Pandas Profiling) możesz znaleźć dodatkowe szczegóły na temat statystyk opisowych każdej kolumny, wykrytych wartości odstających oraz potencjalnych problemów z korelacją między cechami. Zaleca się przejrzanie raportu, aby mieć pełny obraz danych przed przystąpieniem do treningu modelu.

Na tej podstawie można stwierdzić, że dane są w stosunkowo dobrej kondycji, ale warto rozważyć:

- Ujednolicenie rozmiaru obrazów (np. przeskalowanie do wspólnego formatu przed treningiem modelu).
- Rozważenie usunięcia zmiennej depth, jeśli okaże się, że jej wpływ na wyniki modelu jest marginalny.



Rozkład ras psów



Wykres pudełkowy szerokości i wysokości obrazów

