Jeśli chodzi o sprawozdanie z wnioskami dotyczących obrazków wygenerowanych w ramach badań eksploracyjnych (EDA), to należy zwrócić uwagę na następujące kwestie:

1. Histogramy dla cech typu numerycznego:

* pokazują rozkład wartości danej cechy wśród wszystkich próbek danych
* pozwalają na ocenę rozrzutu danych oraz na wychwycenie ewentualnych odstępstw od normy (np. duża ilość wartości odstających)

1. Wykresy słupkowe dla typów jakościowych:

* pokazują, jaka część próbek danych należy do danej kategorii (np. jaki procent liter jest "A", a jaki "B")
* pozwalają na porównanie udziału poszczególnych kategorii w zbiorze danych

1. Diagram analizy korelacji:

* pokazuje, jakie są powiązania między poszczególnymi cechami danych
* pozwala na zidentyfikowanie cech, które silnie ze sobą korelują (co może być istotne przy wyborze cech do modelowania)

1. Wykresy pudełkowe dla atrybutów typu liczbowego:

* pokazują rozrzut danych dla danej cechy
* pozwalają na ocenę, czy dane rozkładają się równomiernie czy są skupione wokół jakiegoś średniego (a także ewentualnych odstępstw od normy)

1. Wykresy rozrzutu dla cech typu numerycznego w zależności od klas:

* pokazują, jak dana cecha wpływa na przynależność próbki do danej klasy
* pozwalają na ocenę, czy dana cecha jest dobrym predytorem klasy

1. Wykresy rozkładu empirycznego (gęstości) dla cech typów liczbowych w zależności od klas (kontynuacja):

* pokazują, jak dana cecha wpływa na rozkład wartości wśród próbek danej klasy
* pozwalają na ocenę, czy dana cecha jest dobrym predytorem klasy

Przy opracowywaniu sprawozdania z wnioskami dotyczących wygenerowanych obrazków, należy zwrócić uwagę na to, czy dane zawarte w obrazkach potwierdzają nasze hipotezy dotyczące danych, czy też wychodzą poza nasze oczekiwania. Ważne jest również uwzględnienie ewentualnych problemów z danymi (np. brakujące wartości, odstępstwa od normy), które mogą wpływać na wyniki modelowania.

Przykładowe wnioski dotyczące poszczególnych obrazków mogą wyglądać następująco:

* Histogramy dla cech typu numerycznego:
  + Cecha X rozkłada się normalnie, ale z dużym rozrzutem wartości. Może to oznaczać, że ta cecha nie będzie dobrym predytorem klasy.
  + Cecha Y ma dużo wartości odstających od reszty. Możliwe, że mamy do czynienia z błędami w danych lub wartościami odstającymi są istotne dla modelu.
* Wykresy słupkowe dla typów jakościowych:
  + W zbiorze danych dominuje klasa A, ale klasa B również ma duży udział. Może to oznaczać, że model będzie miał trudności z dokładnym przewidywaniem klasy B.
* Diagram analizy korelacji:
  + Zauważamy silną korelację pomiędzy cechami X i Y. Możliwe, że jedna z nich będzie zbędna dla modelu.

Wykresy pudełkowe dla atrybutów typu liczbowego:

Cecha X ma rozrzut danych skupiony wokół średniej, ale z dużymi wartościami odstającymi. Możliwe, że te wartości odstające są istotne dla modelu lub też mamy do czynienia z błędami w danych.

Cecha Y ma rozrzut danych skupiony wokół średniej i niewiele wartości odstających. Może to oznaczać, że ta cecha będzie dobrym predytorem klasy.

Wykresy rozrzutu dla cech typu numerycznego w zależności od klas:

Dla cechy X widzimy, że próbki klasy A i B rozkładają się w podobny sposób. Możliwe, że ta cecha nie będzie dobrym predytorem klasy.

Dla cechy Y widzimy, że próbki klasy A i B rozkładają się znacząco inaczej. Może to oznaczać, że ta cecha będzie dobrym predytorem klasy.

* Wykresy rozkładu empirycznego (gęstości) dla cech typów liczbowych w zależności od klas:
  + Dla cechy X widzimy, że gęstość rozkładu dla próbek klasy A i B są podobne. Możliwe, że ta cecha nie będzie dobrym predytorem klasy.
  + Dla cechy Y widzimy, że gęstość rozkładu dla próbek klasy A i B są znacząco różne. Może to oznaczać, że ta cecha będzie dobrym predytorem klasy.

Wniosek końcowy powinien opierać się na analizie wszystkich obrazków oraz na porównaniu ich wyników. Możliwe, że niektóre cechy okażą się lepszymi predyktorem niż inne, a także możliwe, że niektóre algorytmy będą lepiej radzić sobie z klasyfikacją danych niż inne. Ważne, by zwrócić uwagę na ewentualne ograniczenia i wyzwania związane z danymi (np. brakujące wartości, odstępstwa od normy) oraz na to, jak te ograniczenia i wyzwania mogą wpływać na wyniki modelowania.

Wniosek końcowy powinien również zawierać sugestie dotyczące dalszych kroków, takich jak ewentualne przygotowanie danych (np. usunięcie wartości odstających lub brakujących wartości), uwzględnienie dodatkowych cech lub testowanie innych algorytmów. Ważne jest również, by zapisać wnioski i sugestie w formie przystępnej dla odbiorców, tak aby inni mogli w przyszłości skorzystać z naszych doświadczeń i wiedzy.