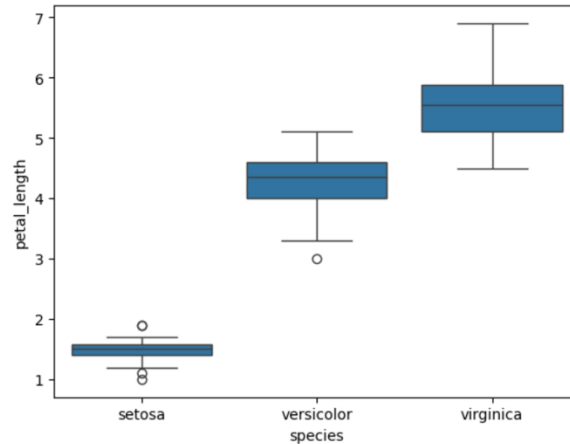


# 1. Iris 데이터셋을 활용해 클래스별 변수 평균 차이를 검정

## 3. 시각화



Boxplot을 봤을 때 setosa, versicolor, virginica 순으로 petal\_length의 평균이 높아진다. 또한 setosa와 versicolor 각각에서 이상치가 보인다.

## 4. 정규성 검정

H0(귀무가설) : 데이터가 정규분포를 따른다.

H1(대립가설) : 데이터가 정규분포를 따르지 않는다.

만약 p-value가 0.05이상이면 귀무가설을 기각한다.

<p-value>

setosa : 약 0.0548 → 정규분포를 따른다.

versicolor: 약 0.1585 → 정규분포를 따른다.

virginica: 약 0.1098 → 정규분포를 따른다.

## 5. 등분산성 검정

H0(귀무가설) : 세 그룹의 분산이 똑같다.

H1(대립가설) : 세 그룹의 분산이 다르다.

만약 p-value가 0.05이상이면 귀무가설을 기각한다.

p-value: 3.1287566394085344e-08 → 세 그룹의 분산이 다르다.

## 6. 가설 수립

H0: 3 개 Species 간 Petal Length 의 평균이 모두 같다.

H1: 적어도 한 쌍의 Species 간 평균은 다르다.

## 7. ANOVA 실행

F: 1180.161182252981

p-value: 2.8567766109615584e-91 → 귀무가설 기각, 적어도 한 쌍의 Species 간 평균은 다르다.

## 8. 사후검정(Tukey HSD)

사후검정 결과 모두 reject 가 True 로 나왔다. 이는 3 개 Species 간 Petal Length 의 평균이 모두 다르다는 뜻이다. 즉, 세 쌍 모두 간에 유의미한 차이가 존재한다.

## 9. 결과 요약

ANOVA 와 사후검정을 통해 세 그룹의 평균이 모두 다르다는 것을 알 수 있다. 또한, 등분산성 검정을 통해 세 그룹의 분산도 모두 다르다는 것을 알 수 있다. 세 그룹은 petal length 값에서 통계적으로 유의미한 차이를 갖고 있다. Boxplot 을 보면 setosa 의 petal\_length 값이 통계적으로 유의하게 짧으며 virginica 의 petal\_length 값이 통계적으로 유의하게 길다.

## 2. 실제 신용카드 사기 데이터셋을 활용해 클래스 불균형 상황에서 분류 모델을 학습

### 5. SMOTE 적용

현재 학습 데이터의 클래스 불균형(사기 거래가 적음)으로 인해 모델이 정상 거래에 치우친 예측을 할 수 있다. 클래스 불균형을 해결하기 위해서는 소수 클래스(사기 거래)를 생성하여 균형을 맞춰줘야 한다. 이를 위해 SMOTE를 사용하는 것이다.

### 7. 최종 성능 평가

	precision	recall	f1-score	support
0	0.99	1.00	1.00	2001
1	0.95	0.89	0.92	98
accuracy			0.99	2099
macro avg	0.97	0.94	0.96	2099
weighted avg	0.99	0.99	0.99	2099

PR-AUC: 0.9538

Class 0 을 보면 recall 값은 1.00, f1-score 은 1.00 으로 조건을 충족하였다.  
Class 1 을 보면 recall 값은 0.89, f1-score 은 0.92 으로 조건을 충족하였다.  
또한 PR-AUC 는 0.9538 로 역시 조건을 충족하였다.