

# SPSS를 통한 IRIS데이터 분석

2023.04.04  
배상일

# 목차

1. 의사결정트리 및 인공 신경망 비교 분석
  - \* 의사결정트리 모델 생성
  - \* 인공 신경망 모델 생성
2. 군집분석(K-means)모델 만들기

# 의사결정트리 / 인공 신경망 생성

The first screenshot shows the 'iris.csv' file being loaded into the software. The 'Species' field is highlighted in the '필드' (Fields) list. The '종에서 12' (From 12) button is also highlighted. The 'Species' field is set as the target variable in the '속성' (Attributes) table.

The second screenshot shows the 'Species' field being set as the target variable. The '속성' (Attributes) table shows the 'Species' field as the target variable.

The third screenshot shows the 'Species' field being set as the target variable. The '속성' (Attributes) table shows the 'Species' field as the target variable.

The fourth screenshot shows the 'Species' field being set as the target variable. The '속성' (Attributes) table shows the 'Species' field as the target variable.

**iris데이터를 불러온 후 유형 탭에서 종속변수인 Species를 명목형 / 목표로 수정**

**분포 확인을 통해 균형 노드(증폭) 생성이 필요 없음을 확인함**

**Train data, Test data 생성**

# 의사결정트리 / 인공 신경망 생성

Species x \$N-Species 교차표 #3

교차표 외형 주석

\$N-Species

Species		setosa	versicolor	virginica	합계
setosa	빈도	24	1	0	25
	행 %	96.000	4.000	0.000	100
	열 %	100.000	3.704	0.000	33.333
	합계 %	32.000	1.333	0.000	33.333
versicolor	빈도	0	24	1	25
	행 %	0.000	96.000	4.000	100
	열 %	0.000	88.889	4.167	33.333
	합계 %	0.000	32.000	1.333	33.333
virginica	빈도	0	2	23	25
	행 %	0.000	8.000	92.000	100
	열 %	0.000	7.407	95.833	33.333
	합계 %	0.000	2.667	30.667	33.333
합계	빈도	24	27	24	75
	행 %	32	36	32.000	100
	열 %	100	100	100	100
	합계 %	32	36	32.000	100

셀 포함 필드의 교차 표(결측값 포함)

카이 제곱 = 127.806, df = 4, 확률 = 0

확인

인공 신경망 모델 Summary  
- 오차를 확인할 수 있다

[Species] 분석 #3

분석 주석

모두 접기 (C) 모두 펼치기 (E)

출력 필드 Species의 결과

\$N-Species와(과) Species 비교

	정확	틀림	합계
정확	71	94.67%	
틀림	4	5.33%	
합계	75		

확인

인공 신경망 모델 분석 내용  
- **94.67%**의 정확도를 보인다

# 의사결정트리 / 인공 신경망 생성

Species x \$C-Species 교차표

교차표 외형 주석

\$C-Species

Species		setosa	versicolor	virginica	합계
setosa	빈도	23	2	0	25
	행 %	92.000	8.000	0.000	100
	열 %	100.000	6.897	0.000	33.333
	합계 %	30.667	2.667	0.000	33.333
versicolor	빈도	0	24	1	25
	행 %	0.000	96.000	4.000	100
	열 %	0.000	82.759	4.348	33.333
	합계 %	0.000	32.000	1.333	33.333
virginica	빈도	0	3	22	25
	행 %	0.000	12.000	88.000	100
	열 %	0.000	10.345	95.652	33.333
	합계 %	0.000	4.000	29.333	33.333
합계	빈도	23	29	23	75
	행 %	30.667	38.667	30.667	100
	열 %	100	100	100	100
	합계 %	30.667	38.667	30.667	100

셀 포함 필드의 교차 표(결측값 포함)

카이 제곱 = 118.192, df = 4, 확률 = 0

확인

의사결정트리 모델 Summary  
- 오차를 확인할 수 있다

[Species] 분석 #1

분석 주석

모두 접기 (C) 모두 펼치기 (E)

출력 필드 Species의 결과

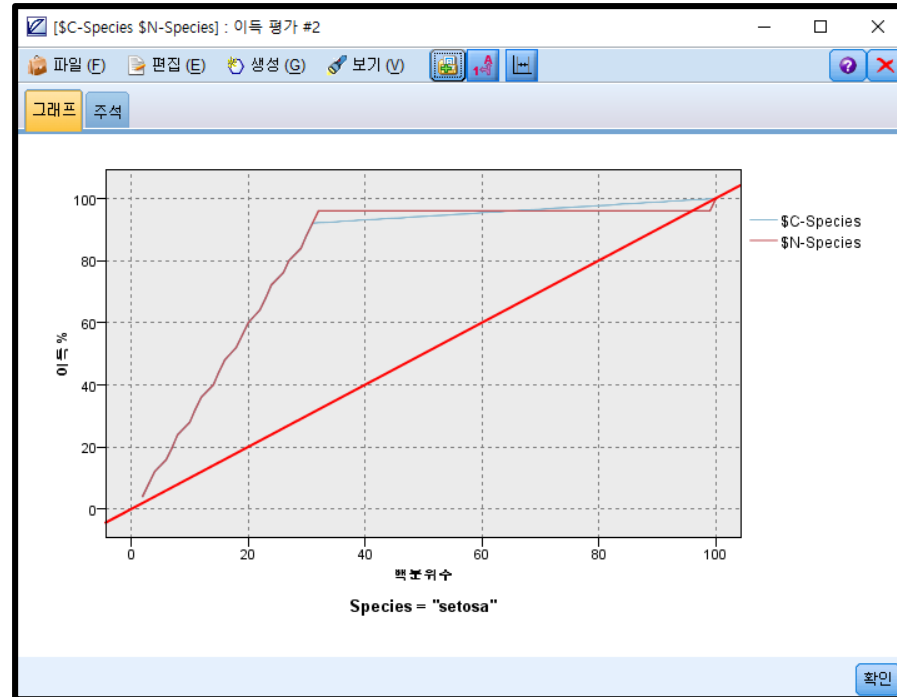
\$C-Species와(과) Species 비교

	정확	틀림	합계
	69	6	75
	92%	8%	

확인

의사결정트리 모델 분석 내용  
- **92.0%**의 정확도를 보인다

# 의사결정트리 / 인공 신경망 생성



[Species] 분석 #2

분석 주석

모두 접기 (C) 모두 펼치기 (E)

출력 필드 Species의 결과

개별 모델

\$C-Species와(과) Species 비교

정확	69	92%
틀림	6	8%
합계	75	

\$N-Species와(과) Species 비교

정확	71	94.67%
틀림	4	5.33%
합계	75	

\$C-Species \$N-Species 간의 일치

일치	69	92%
불일치	6	8%
합계	75	

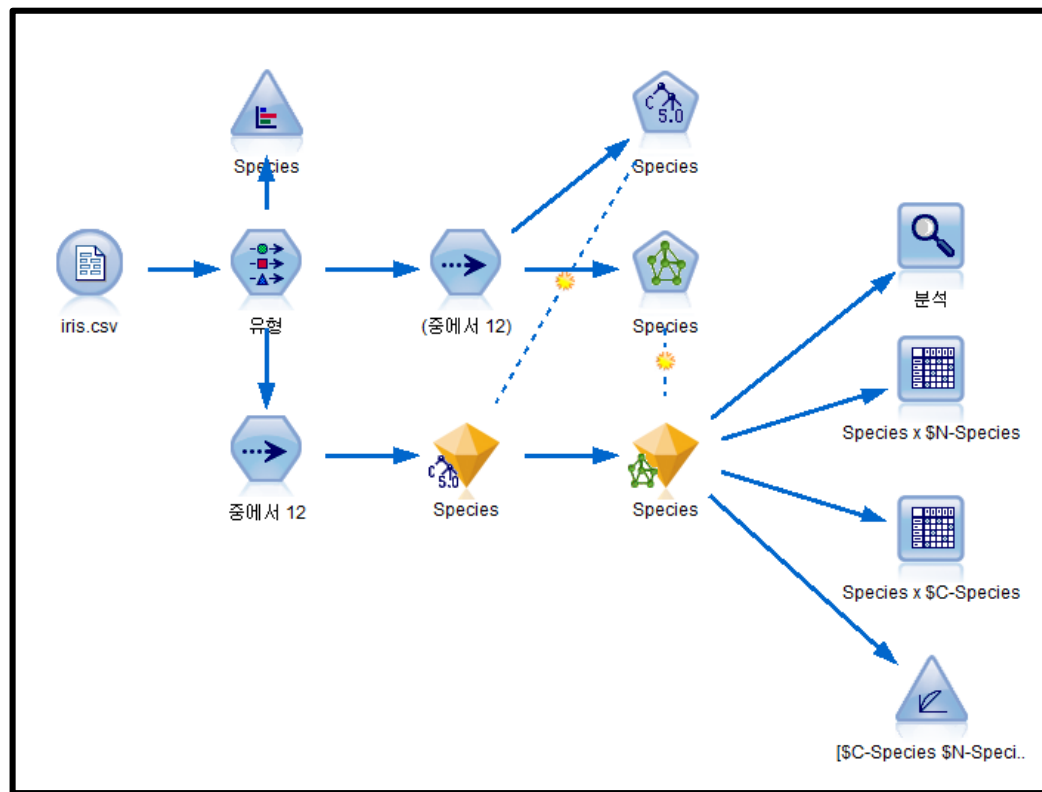
일치와(과) Species 비교

정확	67	97.1%
틀림	2	2.9%
합계	69	

확인

결과적으로 **인공 신경망 모델**이 더 높은 정확도를 갖는다.

# 의사결정트리 / 인공 신경망 생성



의사결정트리 및 인공 신경망 모델 생성의 스트림 최종 모형이다.

# 군집분석(K-means)모델 만들기

테이블 (4개 필드, 150개 레코드)

파일(F) 편집(E) 생성(G)

테이블 주석

iris2.csv

	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width
1	5.100	3.500	1.400	0.200
2	4.900	3.000	1.400	0.200
3	4.700	3.200	1.300	0.200
4	4.600	3.100	1.500	0.200
5	5.000	3.600	1.400	0.200
6	5.400	3.900	1.700	0.400
7	4.600	3.400	1.400	0.300
8	5.000	3.400	1.500	0.200
9	4.400	2.900	1.400	0.200
10	4.900	3.100	1.500	0.100
11	5.400	3.700	1.500	0.200
12	4.800	3.400	1.600	0.200
13	4.800	3.000	1.400	0.100
14	4.300	3.000	1.100	0.100
15	5.800	4.000	1.200	0.200
16	5.700	4.400	1.500	0.400
17	5.400	3.900	1.300	0.400
18	5.100	3.500	1.400	0.300
19	5.700	3.800	1.700	0.300
20	5.100	3.800	1.500	0.300

확인

K-평균

필드 모델 고급 주석

모델 이름: ☐ 자동 (O) ☐ 사용자 정의 (M)

☒ 파티션된 데이터 사용

군집 수:

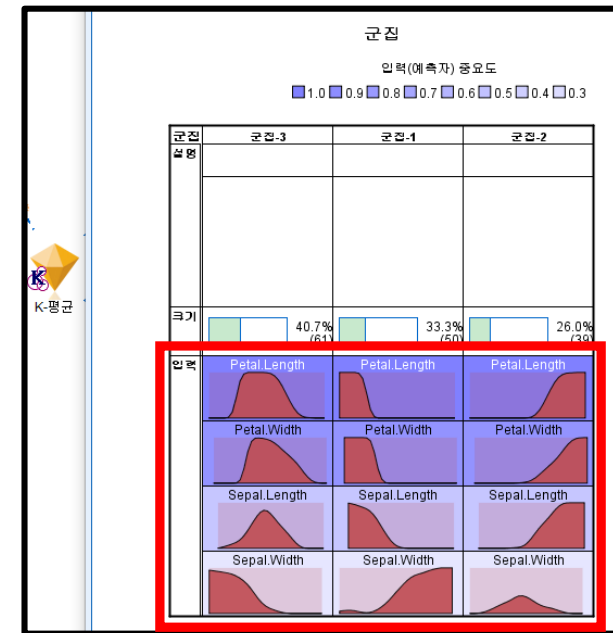
☐ 거리 필드 생성

군집 설명: ☐ 문자열 ☐ 숫자

설명 접두어:

최적화: ☐ 속도 ☒ 메모리

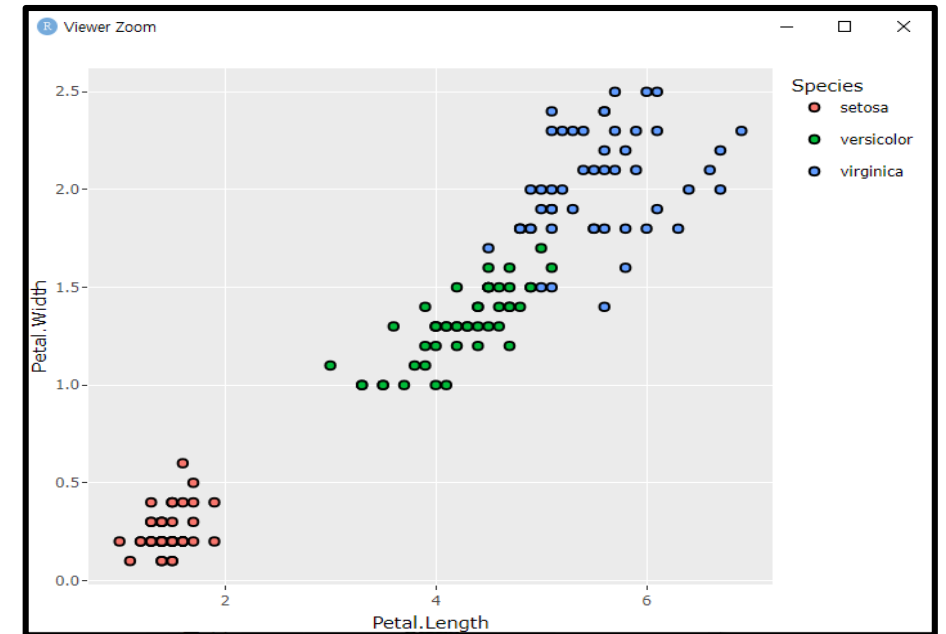
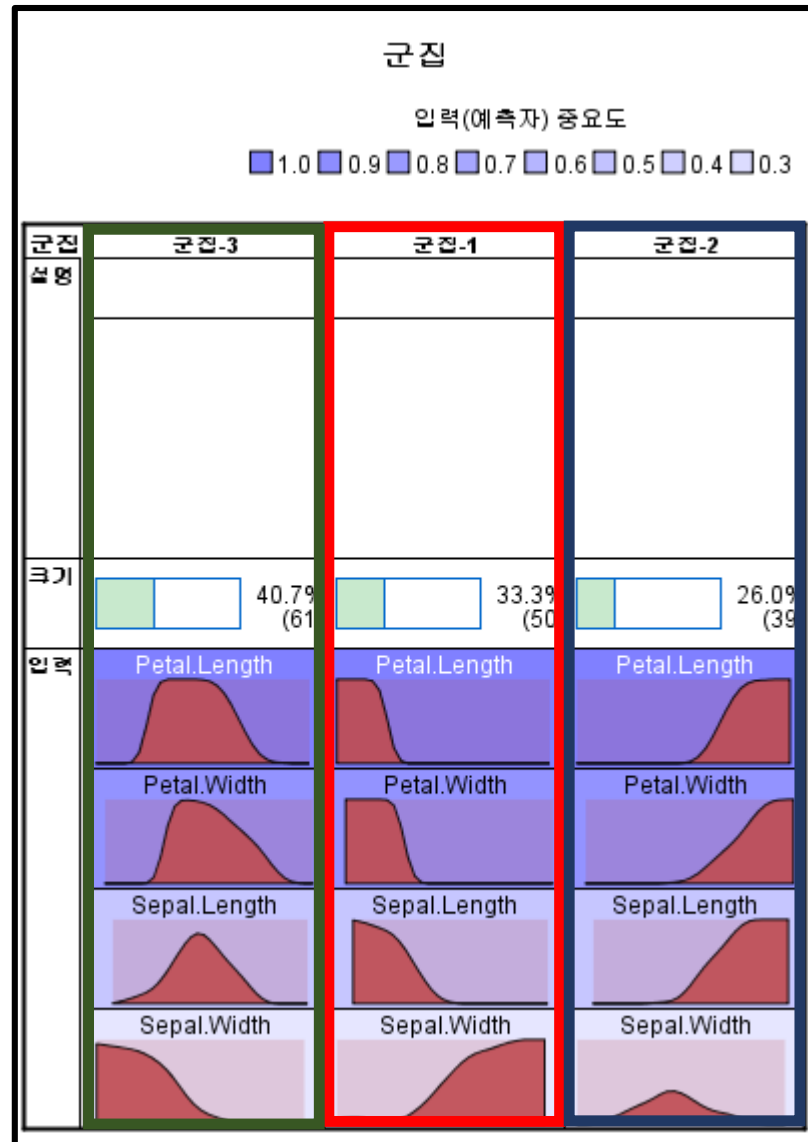
확인 실행 취소 적용(A) 재설정(R)



- 군집을 만들기 위해 종속변수가 되는 Species열을 제거한 후 데이터를 불러온다
- K평균 탭에서 **군집 수**를 바꿔가며 **군집 모형**을 **확인**하고 가장 적합한 군집 수를 알아낸다



# 군집분석(K-means)모델 만들기



R-Studio에서 IRIS 데이터에 대한 plot을 생성하여 Species 별 분포를 확인하고, SPSS 분석 모델로 분류한 군집 모형과 비교를 한다.

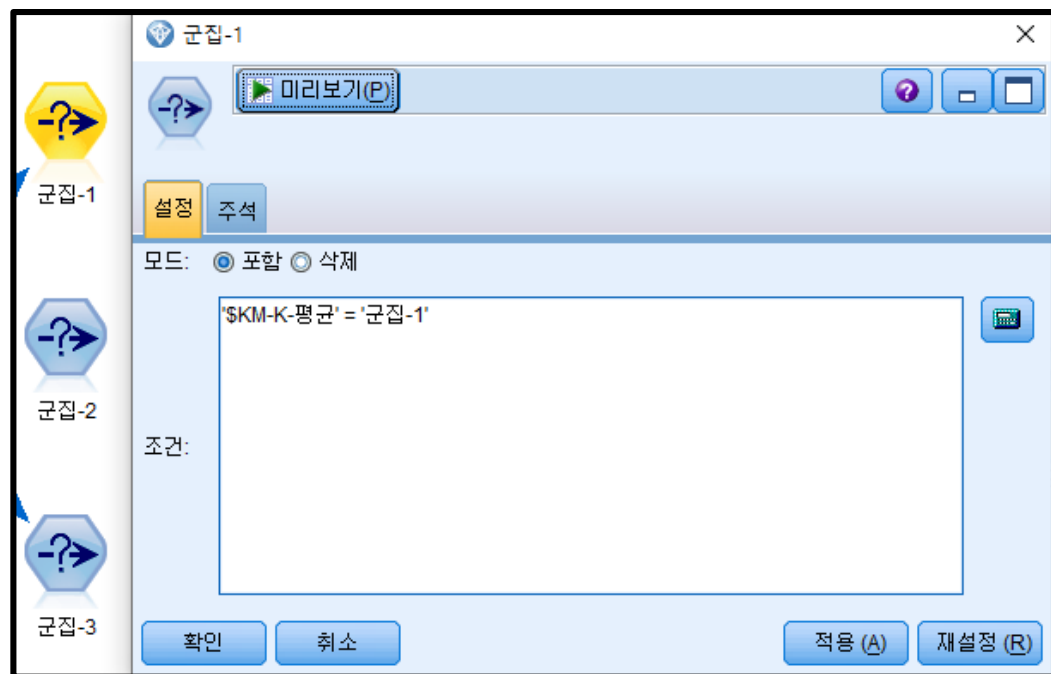
비교 결과(추정)

군집-1 : setosa

군집-2 : versicolor

군집-3 : virginica

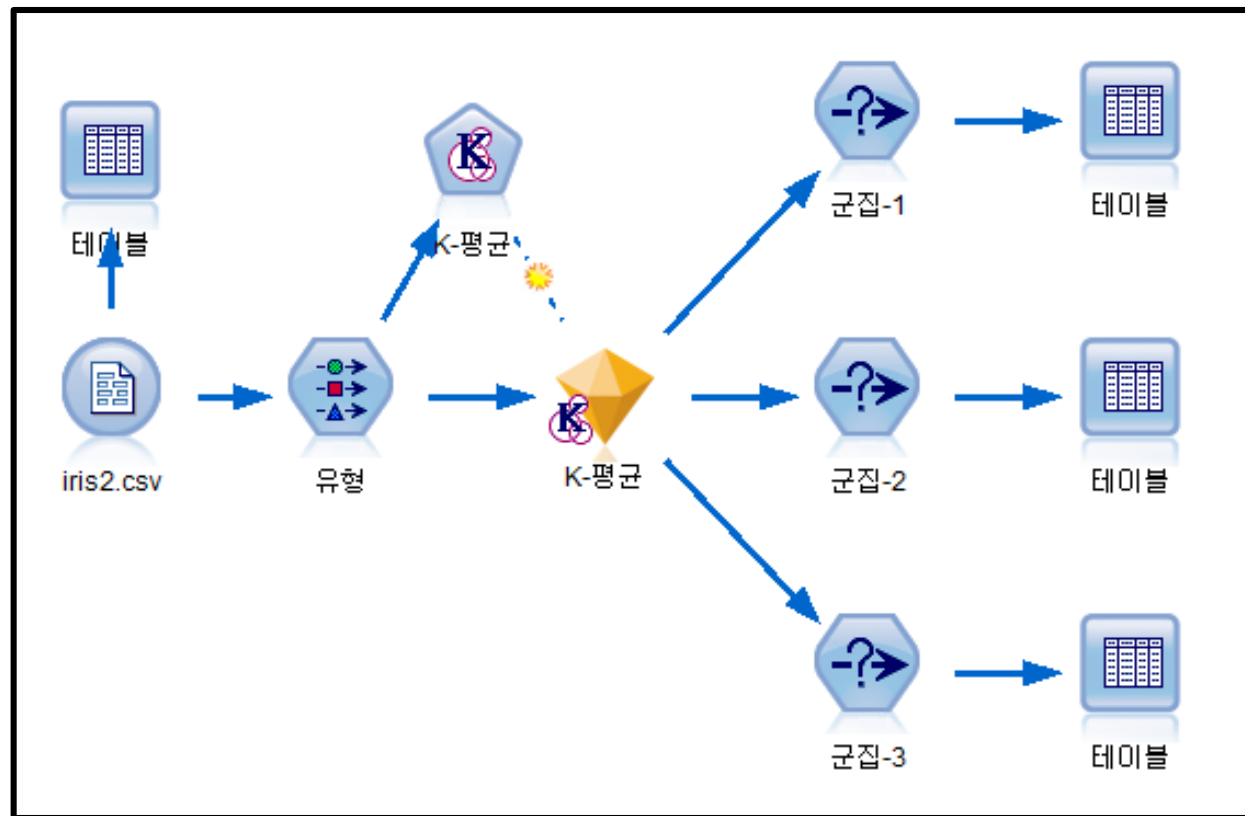
# 군집분석(K-means)모델 만들기



	Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	\$KM-K-평균
1	5.100	3.500	1.400	0.200	군집-1
2	4.900	3.000	1.400	0.200	군집-1
3	4.700	3.200	1.300	0.200	군집-1
4	4.600	3.100	1.500	0.200	군집-1
5	5.000	3.600	1.400	0.200	군집-1
6	5.400	3.900	1.700	0.400	군집-1
7	4.600	3.400	1.400	0.300	군집-1
8	5.000	3.400	1.500	0.200	군집-1
9	4.400	2.900	1.400	0.200	군집-1
10	4.900	3.100	1.500	0.100	군집-1
11	5.400	3.700	1.500	0.200	군집-1
12	4.800	3.400	1.600	0.200	군집-1
13	4.800	3.000	1.400	0.100	군집-1
14	4.300	3.000	1.100	0.100	군집-1
15	5.800	4.000	1.200	0.200	군집-1
16	5.700	4.400	1.500	0.400	군집-1
17	5.400	3.900	1.300	0.400	군집-1
18	5.100	3.500	1.400	0.300	군집-1
19	5.700	3.800	1.700	0.300	군집-1
20	5.100	3.800	1.500	0.300	군집-1

선택 탭을 사용하여 군집을 분류하고 각 군집 별 테이블을 생성한다.

# 군집분석(K-means)모델 만들기



군집분석(K-means) 모델의 스트림 최종 모형이다.

A solid blue vertical bar is positioned on the left side of the slide, extending from the top to the bottom.

**Thank you**