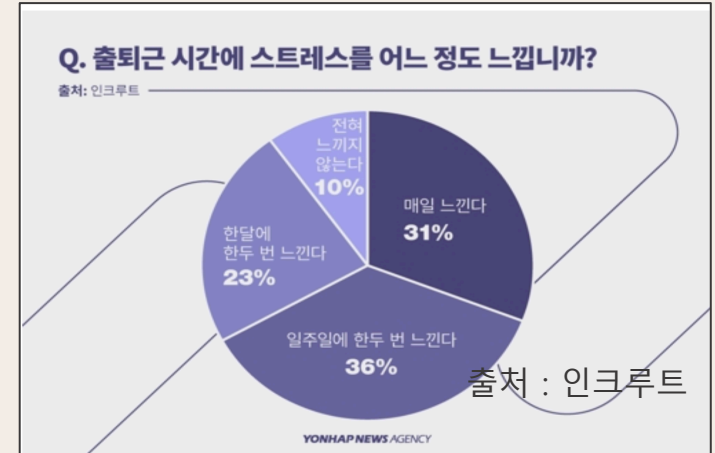




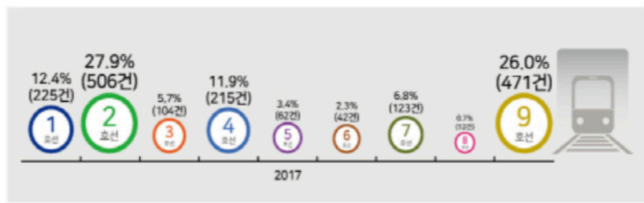
# 날씨에 따른 서울 지하철 2호선 혼잡도 예측

# 1. 공모 배경

지하철은 시민들이 가장 많이  
이용하는 교통 수단 중 하나  
그러나 혼잡한 지하철은 스트레스의 유발 요인



서울지하철 호선별 성범죄 발생현황은? 2·9호선에서 가장 많이 발생



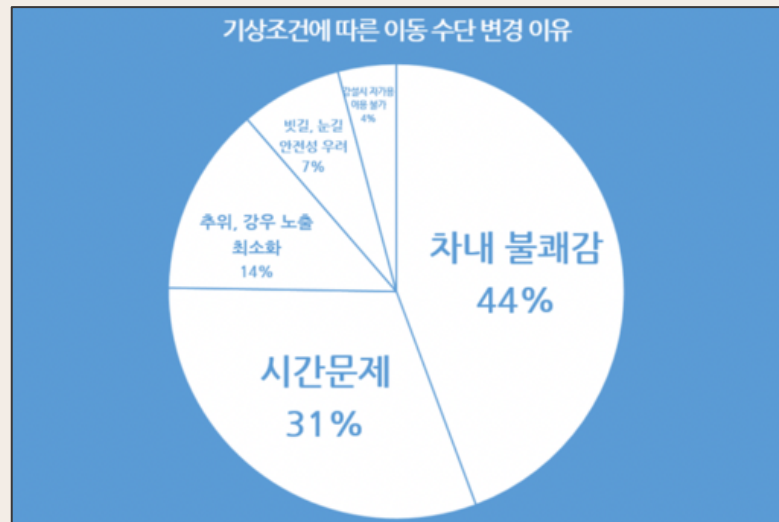
2017년 서울지하철에서 성범죄가 가장 많이 발생한 노선은 2호선, 9호선, 1호선 순으로 나타났습니다.

출처 : 이데일리

유동인구가 많은 호선 일수록 지하철 범죄가  
많다  
(2호선 > 9호선 > 1호선)

# 1. 공모 배경

평소 버스를 이용하던 승객들이 비가 내리고  
체감온도가 낮은 날에는 지하철을 타는 경향성



출처 : 기상청 공식 블로그

날씨에 따라 지하철 혼잡도를 예측 할 수 있다면  
지하철 범죄 문제를 예방하고, 시민들의 교통수단 선택에 도움이 될 것이다.

## 2. 활용 데이터

### 1. 지하철 데이터

서울 열린데이터 광장(<https://data.seoul.go.kr/>)

- '서울 교통공사 연도별 일별 시간대별 역별 승하차 인원' 데이터

date	날짜
line	호선 정보
station_code	역 코드
station_name	역 이름
승하차	승 하차 인원 구분
05.~06	5시~6시 사이의 인원
:	:
:	:
24.~25	24시~25시 사이의 인원
sum	합계

시간대별, 호선별, 역별, 시간대별 승 하차 인원이 기록되어 있다.

## 2. 활용 데이터

### 2. 기상 데이터

기상자료개방포털(<https://data.kma.go.kr/>)

#### ● 서울 지역 기상 데이터

date	관측 일시
humid	평균 상대 습도(%)
station_name	역 이름
rain	일 강수량(mm)
snow	최심적설량(cm)
temperature	일 평균기온(c)
wind	일 평균 풍속(m/s)

일별 기상 정보가 기록되어 있다.

### 3. 데이터처리 방향 및 분석 계획

	②		③		④		⑤	
date	holiday	day	rush_user	notrush_user	mean_rush_user	mean_notrush_user	rush_busylevel	notrush_busylevel
2017-01-01	T	Sunday	0	739831	0	1902	0	3
2017-01-02	F	Monday	901914	676281	3391	3060	3	7
2017-01-03	F	Tuesday	935820	721445	3518	3264	3	7
2017-01-04	F	Wednesday	946006	747858	3556	3384	3	7
2017-01-05	F	Thursday	940673	764994	3536	3462	3	7
2017-01-06	F	Friday	983864	807237	3699	3653	3	7

#### ① Outlier 제거

공휴일 데이터는 날씨와 상관없이 사람들이 대중교통을 이용하지 않는 경향성이 있었음. 따라서 공휴일을 제외한 날을 분석.

#### holiday 변수 추가

날씨 뿐만 아니라 출퇴근 시간 따라 지하철 인원에 크게 변동이 있었음.  
구분하여 분석하기 위해 holiday 변수를 추가

#### ③ rush\_user, notrush\_user 변수 추가

하루의 지하철 이용 승객을 출퇴근 시간과 아닌 시간대의 승객으로 구분하여 총 합을 구함  
휴일에는 rush\_user(출퇴근 승객)이 없다고 보고 모두 notrush\_user로 둠

### 3. 데이터처리 방향 및 분석 계획

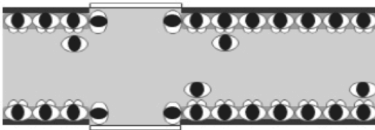
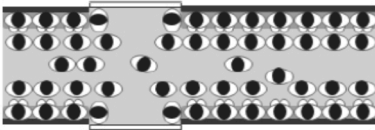
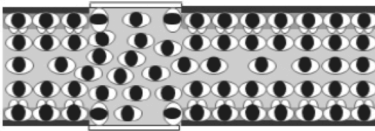
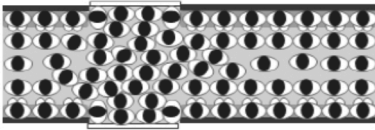
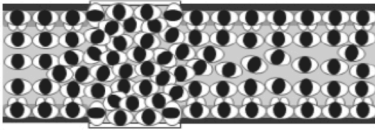
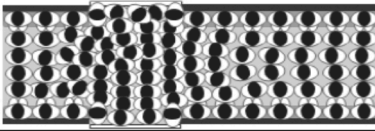
	②	③	④	⑤					
①	date	holiday	day	rush_user	notrush_user	mean_rush_user	mean_notrush_user	rush_busylevel	notrush_busylevel
	2017-01-01	T	Sunday	0	739831	0	1902	0	3
	2017-01-02	F	Monday	901914	676281	3391	3060	3	7
	2017-01-03	F	Tuesday	935820	721445	3518	3264	3	7
	2017-01-04	F	Wednesday	946006	747858	3556	3384	3	7
	2017-01-05	F	Thursday	940673	764994	3536	3462	3	7
	2017-01-06	F	Friday	983864	807237	3699	3653	3	7

- ④ mean\_rush\_user, mean\_not\_rush user 변수 생성  
출퇴근, 출퇴근이 아닌 시간의 인구를 그 시간대에 배차된 지하철 수로 나누어 지하철 한 대당 평균 몇 명이 탑승하는지 구함

rush\_busylevel, notrush\_busylevel 변수 추가  
구해진 mean\_rush\_user와 mean\_not\_rush\_user를 통해 혼잡도를 구함.  
구간 정보는 서울시 교통공사 참고(뒷장)

⑤

### 3. 데이터처리 방향 및 분석 계획

혼잡도	재차인원	차내상태 설명	혼잡 상태도
		① 입석승객 대상, ② 좌석승객 대상	
50%	80명	① 좌석에 모두 착석하고 간간히 서 있음 ② 앞의 시야가 트임	
100%	160명	① 여유롭게 서 있음 ② 앞에 사람들이 서 있어서 시야가 다소 막힘	
125%	200명	① 지나갈 때 사람과 부딪치게 되는 다소 혼잡한 상태 ② 앞에 사람들이 많이 서 있어서 시야가 막힘	
150%	240명	① 출입문 주변이 혼잡하고 서로 어깨가 밀착됨 ② 앞에 서 있는 사람들이 밀치기도 하여 불편감을 느끼기도 함	
175%	280명	① 출입문 주변이 매우 혼잡하고 서로 몸이 밀착되어 팔을 들 수 없음 ② 앞에 서 있는 사람들과 무릎이 닿기도 하여 불편함	
200%	320명	① 출입문 주변이 매우 혼잡하고 서로 몸과 얼굴이 밀착되어 숨이 막힘 ② 서 있는 사람들이 심하게 밀려 발이 밟히기도 하고 '억' 소리가 나면서 소란스러움	

서울 지하철 2호선의 지하철 한 대당  
10 량이라는 정보를 토대로  
혼잡도 level을 1~7까지 구분

출처 : 서울시 교통공사



### 3. 데이터처리 방향 및 분석 계획

date	holiday	day	rush_user	notrush_user	mean_rush_user	mean_notrush_user	rush_busylevel	notrush_busylevel
2017-01-01	T	Sunday	0	739831	0	1902	0	3
2017-01-02	F	Monday	901914	676281	3391	3060	3	7
2017-01-03	F	Tuesday	935820	721445	3518	3264	3	7
2017-01-04	F	Wednesday	946006	747858	3556	3384	3	7
2017-01-05	F	Thursday	940673	764994	3536	3462	3	7
2017-01-06	F	Friday	983864	807237	3699	3653	3	7

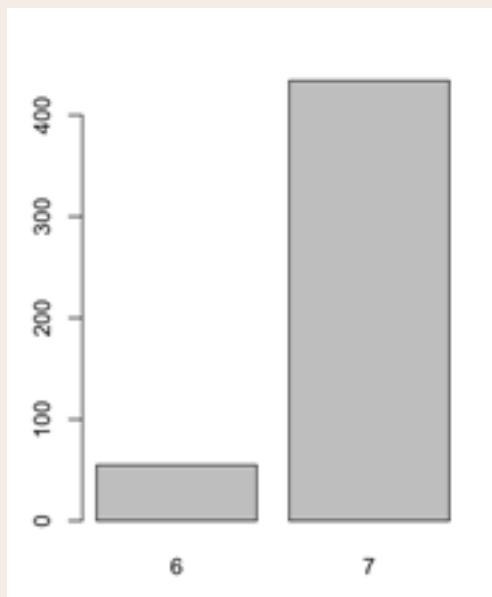
rain	newsnow	snow	meantemp	lowtemp	hightemp	humid	wind
0.0	0	0	2.7	-1.6	6.9	75.9	1.5
0.3	0	0	5.0	1.8	9.2	77.8	2.1
0.0	0	0	2.0	-2.3	7.7	61.8	1.8
0.0	0	0	3.9	1.0	8.9	55.0	1.7
0.0	0	0	3.8	-0.1	7.3	52.3	3.1
0.0	0	0	5.4	2.5	11.4	58.5	2.4

날씨 변수에 따른 **busy\_level** 을 알아내기 위해  
(휴일, 모든시간대), (평일, 출퇴근 시간), (평일, 비 출퇴근 시간)으로  
나누어 **knn** 을 실행

## 4. 분석 결과

### 1. 평일 바쁘지 않은 시간대 지하철 혼잡도 분석

지하철 혼잡도 레벨 분포



K=20일때, knn실행 후 얻은 matrix

	Level "6"	Level "7"
Level "6"	13	6
Level "7"	19	48

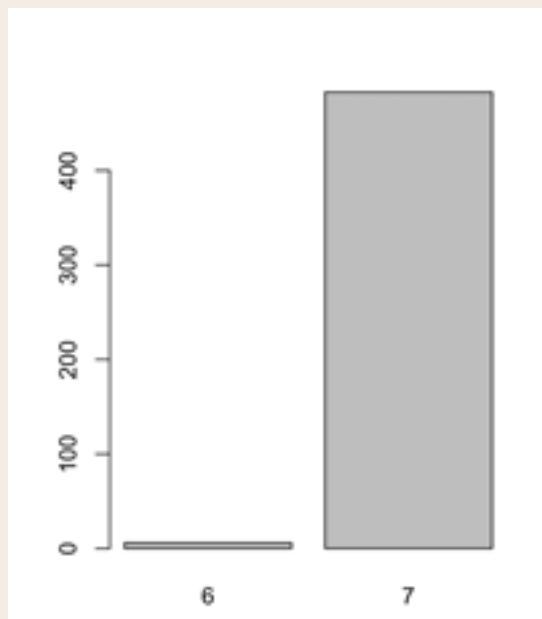
Matrix를 통해 얻은 accuracy

$$\begin{aligned}\text{Overall Accuracy} &= (13+48)/(13+6+19+48) \\ &= \mathbf{0.709}\end{aligned}$$

## 4. 분석 결과

### 2. 평일, 바쁜 시간대 지하철 혼잡도 분석

지하철 혼잡도 레벨 분포



K=20일때, knn실행 후 얻은 matrix

	Level "6"	Level "7"
Level "6"	0	1
Level "7"	0	91

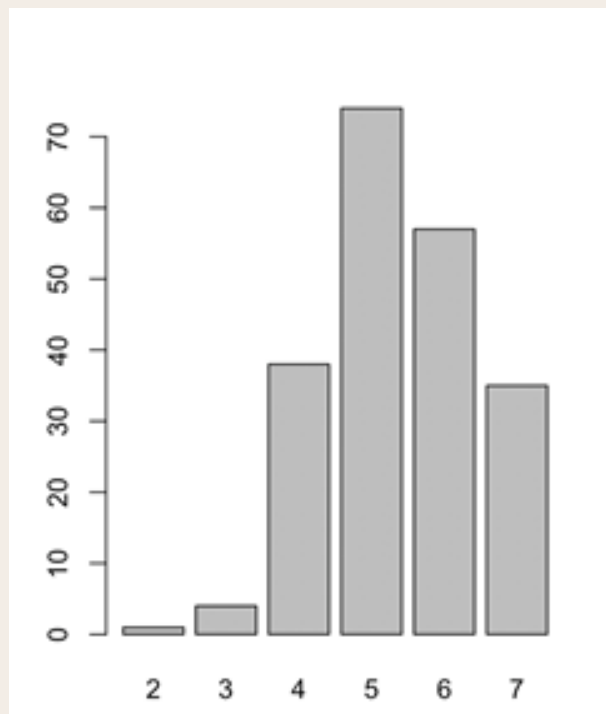
Matrix를 통해 얻은 accuracy

$$\text{Overall Accuracy} = (91)/(92) = \mathbf{0.99}$$

## 4. 분석 결과

### 3. 휴일, 모든 지하철 혼잡도 분석

지하철 혼잡도 레벨 분포



K=13일때, knn실행 후 얻은 matrix

	Level "1"	Level "2"	Level "3"	Level "4"	Level "5"	Level "6"	Level "7"
Level "1"	0	0	0	0	0	0	0
Level "2"	0	0	0	0	0	0	0
Level "3"	0	0	0	0	0	0	0
Level "4"	0	0	0	3	2	1	0
Level "5"	0	0	0	2	4	6	1
Level "6"	0	0	0	2	4	4	0
Level "7"	0	0	0	0	8	0	0

Matrix를 통해 얻은 accuracy

$$\text{Overall Accuracy} = (11)/(37) = \mathbf{0.30}$$

## 4. 활용 방안

위의 분석만으로는, 서울 지하철 2호선의 승객수와 기상 상태가 상관관계가 있는지 알기 힘들다. 그 이유는 지하철 데이터셋을 2018년과 2017년 밖에 확보할 수 없었기 때문이라고 생각된다. 더 많은 데이터로 보완할 수 있다면 지하철 혼잡 지수를 개발 할 수 있을 것이다.

### 지하철 혼잡 지수 활용 방안

- 지하철 역사 내 관리 인원을 조절하여 역사 내 범죄 예방
- 시민들에게는 미리 혼잡도를 알고 대중교통이나 자가용 등 다른 교통수단을 유연하게 선택하는 것에 도움
- 지하철 역사 내의 상인들에게 상품 수요를 예측