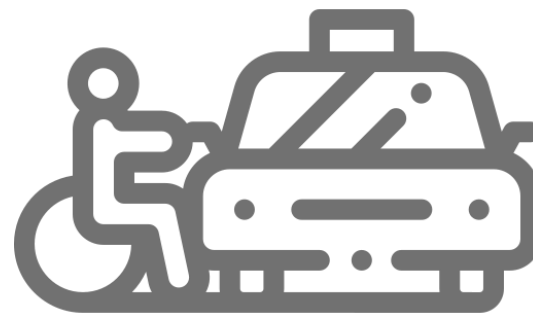


Hadoop-Hive를 이용한 장애인 콜택시 운행 패턴 분석



박진원 임원기 류경민

■ 목차



- 문제정의
- 분석과제
- 데이터 수집 및 요약
- 데이터 저장
- 데이터 탐색적 분석

문제 정의

수행년도	제목	수행기관
2009	장애인 콜택시 운영 실태조사 및 개선방안 연구	서울시복지재단
2015	장애인 콜택시 자동배차시스템 개발	서울시설공단 카리스트
2017	장애인 콜택시 운영분석	서울시
2018	장애인콜택시 지능형 운영체계 알고리즘 개발 및 고객센터 개선 방안 공모사업	서울디지털재단



다양한 연구에도 불구하고 ...
여전히 배차대기시간 문제가
해결되지 않아 ..

[2016 국감] 서울시 장애인콜택시, 대기시간·불친절 민원 여전

배차는 복불복? 장애인 콜택시

4시간 기다려도 감감 "예약 무슨 소용 있나" 항의에 통합시스템 부재 탓만

입력 2016-04-11 10:03:33



댓글쓰기



이 더위에... 2시간째 안 오는 장애인 콜택시

조선일보 | 박상현 기자 | 김승구 인턴기자(UCLA 정치외교학과 졸업)

장애인 콜택시 2시간 대기는 기본... 중도 포기 많아

노인환 기자 | 승인 2019.01.15 18:21 | 댓글 9

1시간 대기 '장애인콜택시' ...7월부터 타기 더 뻑뻑할
듯

장애등급 폐지로 모두 탑승 가능
증차 없어 서울 1대당 800명꼴
이용자 늘어 대기시간 증가 불보듯

분석 내용

1. 구별, 시간대별 대기시간 패턴 분석
2. 콜택시 수요 분석
3. 구별 콜택시 이동 패턴 분석

데이터 수집 및 요약



장애인콜시스템 데이터 (2018, API)

필드명	설명	도메인
car_no	차량고유번호	7815, 1254
car_type	차량유형	중형승합, 승용 ...
exp_time	예상시간	yyyy-mm-dd H:M:S
set_time	배차시간	yyyy-mm-dd H:M:S
ride_time	승차시간	yyyy-mm-dd H:M:S
dept_gu	출발지(구)	강남구, 서초구 ...
dept_dong	출발지(동)	상계동, 하계동 ...
dest_gu	목적지(구)	강남구, 서초구 ...
dest_dong	목적지(동)	상계동, 하계동 ...
ride_exp	승차시간 - 예상시간(분)	0 ~ 500
ride_set	승차시간 - 배차시간(분)	0 ~ 500
set_exp	배차시간 - 예상시간(분)	0 ~ 500
month	월	1~12
day	일	1~31
hour	시	0~24
weekday	요일	월~일

기
존
변
수


파
생
변
수

데이터 저장


자료형 결정

컬럼명	데이터 타입
car_no	String
car_type	String
exp_time	Date
set_time	Date
ride_time	Date
dept_gu	String
dept_dong	String
dest_gu	String
dest_dong	String
ride_exp	int
ride_set	Int
set_exp	Int
month	Int
day	Int
hour	Int
weekday	string

스키마 생성



```
CREATE TABLE cst_18_utf (  
  car_no string,  
  car_type date,  
  exp_time date,  
  set_time date,  
  ride_time string,  
  dept_gu string,  
  dept_dong string,  
  dest_gu string,  
  dest_dong string,  
  ride_exp int,  
  ride_set int,  
  set_exp int,  
  month int,  
  day int,  
  hour int,  
  weekday string)  
ROW FORMAT DELIMITED  
FIELDS TERMINATED BY ','  
LINES TERMINATED BY '\n'  
STORED AS TEXTFILE;
```

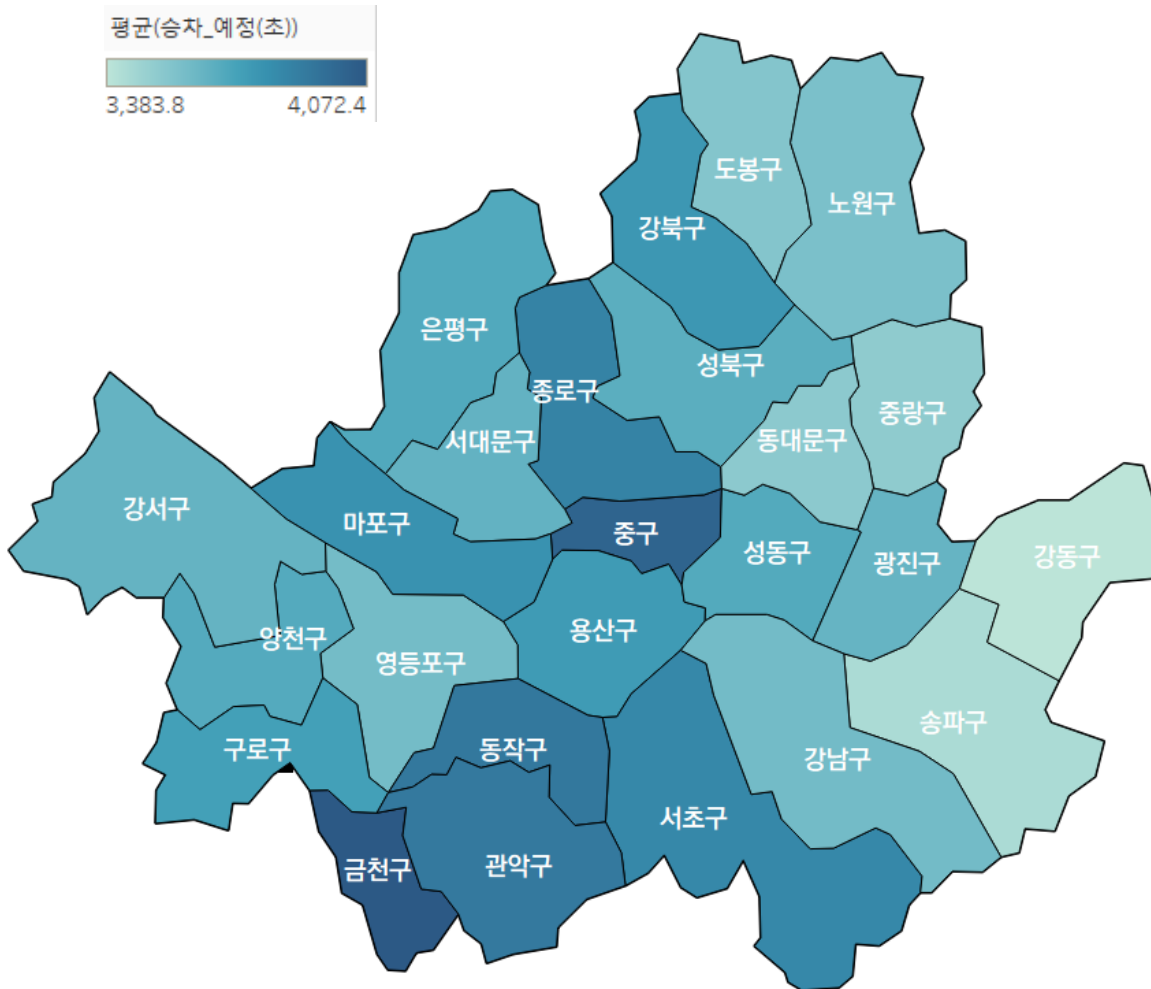


적재

```
LOAD DATA  
LOCAL INPATH  
"cst_18_utf.csv"  
OVERWRITE INTO  
TABLE cst_18_utf;
```

데이터 분석

구별 평균대기시간(분)



금천구	67.87
중구	66.98
동작구	65.41
관악구	65.33
종로구	64.55
서초구	64.14
마포구	63.35
강북구	62.99
용산구	62.72
구로구	62.34
은평구	61.62
성동구	61.54
양천구	61.46
성북구	61.18
서대문구	60.80
강서구	60.72
광진구	60.70
강남구	60.03
영등포구	59.96
노원구	59.58
도봉구	59.14
동대문구	58.75
중랑구	58.58
송파구	57.18
강동구	56.38

: HiveQL Query

```
select dept_gu,
round(avg(ride_exp),1)
from cst_18_utf
group by dept_gu
oreer by dept_gu;
```

**평균 대기시간이
가장 긴 지역은
금천구 > 중구 > 동작구
관악구 > 종로구
순이다**

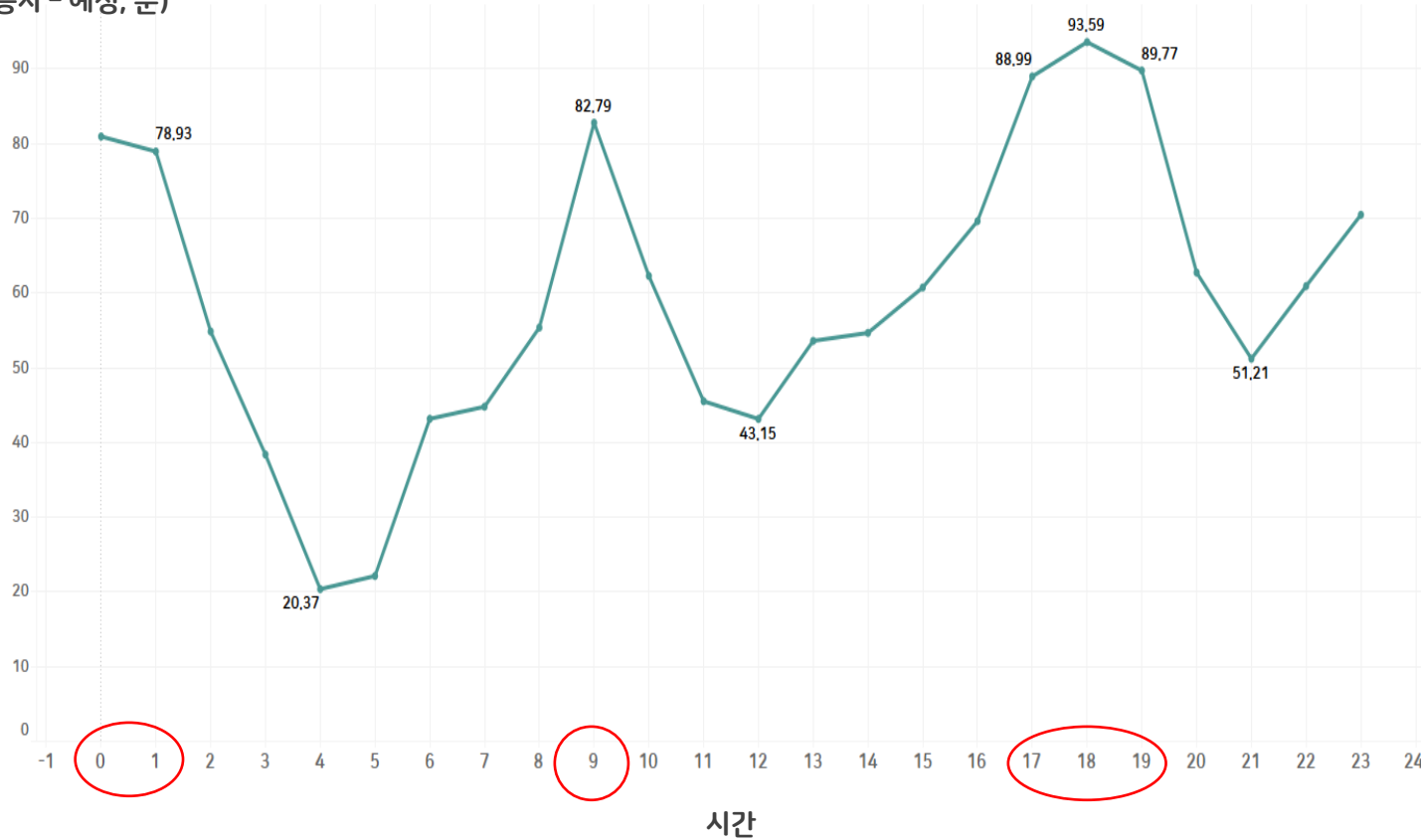
데이터 분석

시간대별 평균 대기시간(분)

: HiveQL Query

```
select hour,  
round(avg(ride_exp))  
as avg_ride_exp  
from cst_18_utf  
group by hour;
```

평균 대기시간
(승차 - 예정, 분)



콜이 집중되는 시간은
0~1시, 9시, 17~19시

최대 평균 콜대기시간 **93분**
최소 평균 콜대기시간 **20분**

데이터 분석

지역구-시간대별 평균대기시간(분)

	hour																							
dept_gu	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
강남구	86.9	71.1	63.5	50.3	29.7	25.9	43.3	41.9	61.1	78.1	59.6	49.7	42.4	52.3	52.0	57.2	61.8	86.3	94.0	107.3	79.1	69.3	75.3	78.0
강동구	71.6	68.4	58.4	56.9	22.2	47.5	43.4	39.5	49.9	62.9	44.5	38.0	40.6	51.5	54.2	70.3	81.5	103.6	98.6	96.1	72.2	56.7	59.0	63.6
강북구	95.6	98.3	67.7	44.4	31.2	15.9	47.9	42.3	55.8	79.6	69.5	51.1	45.5	56.4	59.4	67.1	77.4	89.8	84.6	77.9	51.5	45.9	59.1	73.7
강서구	90.2	75.6	52.9	51.1	16.9	25.5	90.4	48.9	63.3	101.2	75.3	51.2	40.7	49.8	47.1	46.1	56.2	74.8	92.4	87.4	49.9	42.3	48.8	55.3
관악구	95.0	83.2	73.1	78.9	17.0	17.8	44.6	79.8	54.2	108.0	83.5	46.4	45.7	54.8	56.7	61.9	68.5	86.9	90.0	82.6	65.7	63.0	65.4	77.2
광진구	56.4	51.4	37.5	38.7	25.8	15.6	27.6	30.5	52.5	63.1	54.7	36.5	37.8	48.0	50.9	62.2	76.5	104.1	107.5	104.9	76.8	60.5	60.3	60.5
구로구	96.1	94.5	72.2	39.1	16.6	44.4	61.0	46.0	57.0	99.7	69.8	45.5	43.6	57.9	57.5	56.9	63.9	80.1	86.3	82.2	65.1	49.3	60.7	67.0
금천구	74.5	74.9	51.9	51.1	43.8	17.6	52.7	48.2	70.9	116.0	69.2	59.5	47.2	54.9	56.2	61.2	70.5	73.0	88.0	98.6	69.5	60.0	71.9	71.2
노원구	59.3	66.6	49.7	24.5	21.7	20.3	31.2	38.7	56.9	81.4	64.0	48.5	44.2	56.5	60.7	63.7	66.3	79.9	77.1	57.4	36.8	31.5	43.1	54.6
도봉구	96.9	69.7	31.9	50.0	38.3	11.1	43.8	55.7	41.8	79.4	66.5	52.4	44.8	56.0	56.5	59.6	64.5	77.0	74.8	59.0	42.5	34.5	55.3	75.6
동대문구	68.0	90.3	41.0	42.2	8.3	32.0	38.7	32.4	50.0	76.1	57.1	46.1	45.7	56.1	54.6	59.4	71.9	84.5	99.1	88.2	56.0	45.7	58.4	76.8
동작구	72.4	63.6	51.9	12.0	27.7	36.2	43.3	55.0	63.1	104.4	72.0	45.9	43.6	53.9	52.6	59.0	71.8	87.4	96.3	93.4	68.3	57.1	69.4	69.5
마포구	93.7	79.7	76.5	56.0	33.8	57.4	24.5	34.3	52.4	71.3	51.7	39.9	39.9	50.2	54.5	67.9	80.7	109.1	112.4	96.5	61.9	53.8	60.4	74.3
서대문구	79.7	99.0	57.8	41.3	24.3	21.1	73.8	36.4	57.3	68.9	49.9	34.6	39.8	53.2	53.3	67.8	85.2	116.3	122.8	114.3	68.6	48.6	62.1	65.6
서초구	87.7	102.6	62.7	55.6	41.5	13.7	31.9	64.0	64.0	94.1	60.6	46.5	45.5	53.3	54.1	54.6	63.7	82.9	97.0	109.1	91.9	69.3	65.9	81.2
성동구	104.3	82.7	52.3	46.1	32.7	49.5	48.0	42.5	54.7	76.7	63.1	42.6	47.5	54.8	53.0	61.2	69.4	96.4	93.4	91.2	62.0	43.7	71.0	89.7
성북구	75.6	70.9	42.9	27.9	9.3	11.1	27.9	43.2	60.4	83.4	63.8	49.2	45.1	57.1	54.6	59.1	70.2	82.0	92.3	78.9	44.2	48.6	60.4	70.9
송파구	101.3	98.2	73.5	44.8	7.3	10.3	41.4	37.6	52.3	72.7	60.0	44.3	38.9	46.3	47.8	55.3	66.6	84.9	94.9	100.2	76.5	53.6	57.2	79.0
양천구	91.4	106.0	70.0	51.0	57.9	47.6	63.3	55.8	59.6	100.4	65.7	45.5	44.4	57.9	52.8	54.5	60.2	76.5	80.9	79.4	61.7	48.0	46.6	71.9
영등포구	78.4	85.0	45.7	44.8	33.4	36.6	25.4	56.4	44.0	80.9	61.9	40.9	40.9	53.0	54.1	58.7	61.7	79.3	86.7	97.1	71.1	57.1	71.9	79.7
용산구	75.4	85.4	42.6	35.8	46.9	33.8	44.7	51.6	57.2	76.2	52.3	41.3	45.5	57.0	62.6	65.8	78.8	108.1	103.0	99.1	73.3	65.8	72.2	75.8
은평구	85.9	77.8	112.1	38.1	12.9	39.3	48.1	43.5	53.8	78.1	57.5	43.6	41.2	46.7	49.5	60.8	73.7	95.4	101.1	99.6	68.2	53.7	76.9	84.9
종로구	81.9	85.7	67.9	70.1	10.0	31.4	65.1	46.1	52.9	67.5	52.6	41.6	43.3	52.0	56.5	61.5	74.9	100.0	108.9	106.7	62.8	59.3	69.6	74.0
중구	67.3	62.5	66.4	54.1	39.5	39.2	69.6	47.0	55.3	74.5	64.4	43.5	42.5	57.0	61.7	66.6	75.8	100.1	105.9	110.2	80.7	62.4	73.8	84.3
중랑구	69.4	71.0	45.7	33.0	12.3	18.5	49.9	33.7	50.1	87.9	66.2	53.2	49.7	57.9	55.7	60.0	62.7	77.6	72.9	65.1	45.3	36.8	48.9	61.5

지역구 모두 유사한 패턴을 보이며,

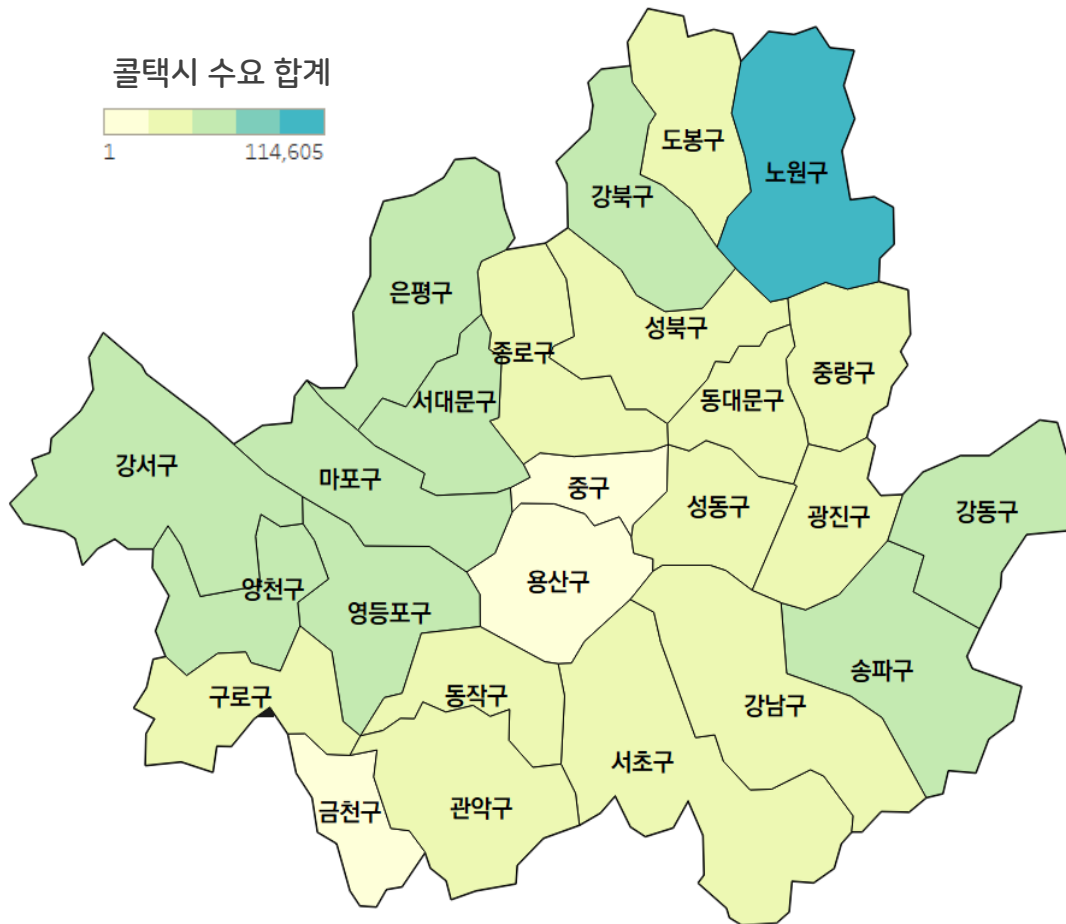
강동, 강북, 마포, 서대문, 용산구는 타 지역구에 비해 16시부터 평균대기시간이 길어진다.

강남, 광진, 서초, 중구는 타 지역구에 비해 19시까지 평균대기시간이 길다.

데이터 분석

구별 콜택시 수요

: HiveQL Query



노원구	114,605
강서구	66,923
서대문구	60,538
은평구	57,534
마포구	56,360
송파구	52,601
강동구	51,585
강북구	51,402
영등포구	48,398
양천구	47,389
중랑구	45,102
강남구	44,762
성북구	43,771
도봉구	40,350
동대문구	39,843
관악구	38,997
구로구	38,284
종로구	34,769
동작구	34,347
서초구	32,926
광진구	30,407
성동구	26,749
중구	18,958
용산구	18,405
금천구	16,080

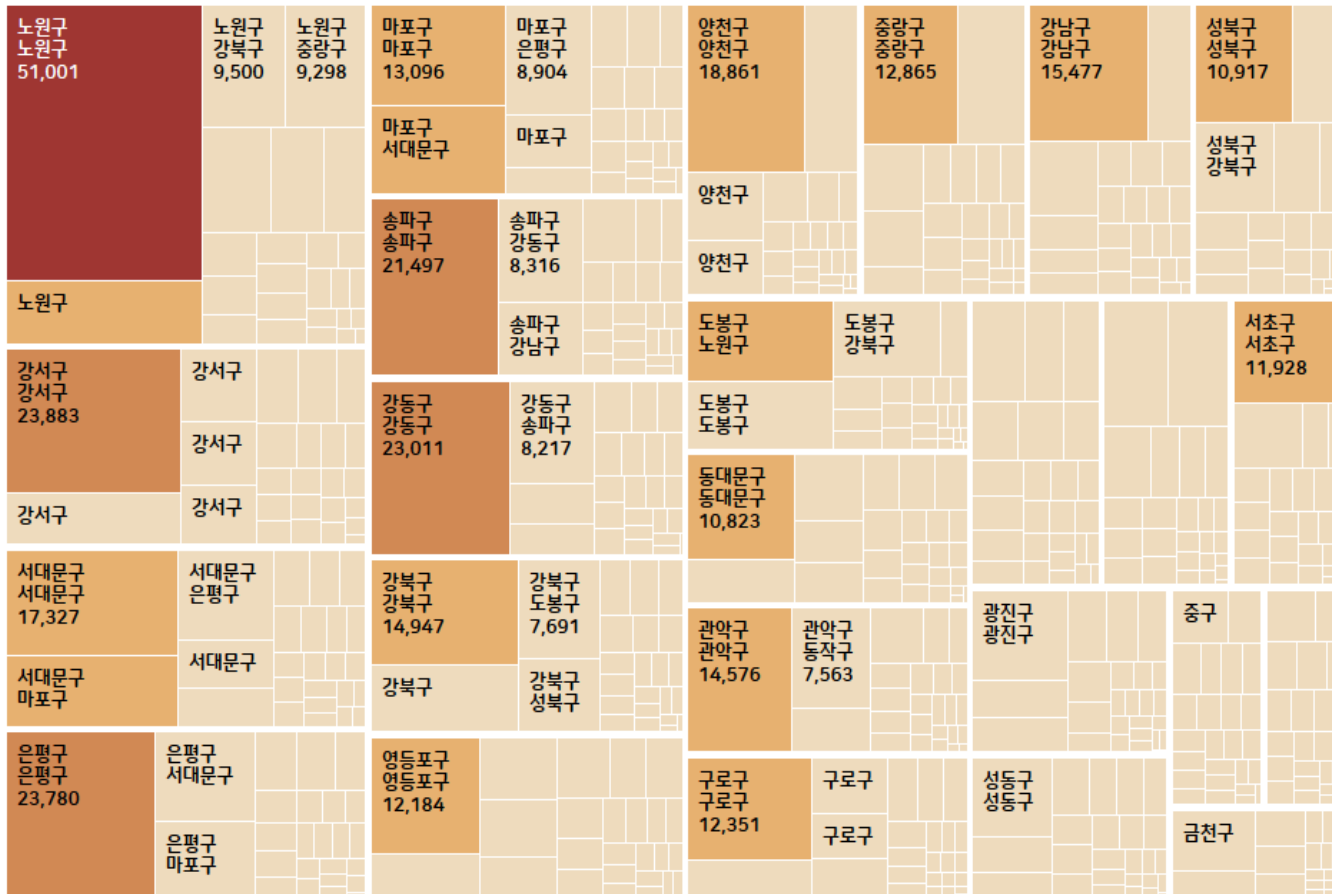
```
select dept_gu,
count(dept_gu)
from cst_18_utf
group by dept_gu
oreer by dept_gu;
```

콜택시 수요가
가장 많은 지역은
노원구 > 강서구 >
서대문구 > 은평구 >
마포구
순이다

데이터 분석

출발지-목적지별 콜택시 수요 (25개 구)

이동건수 합계



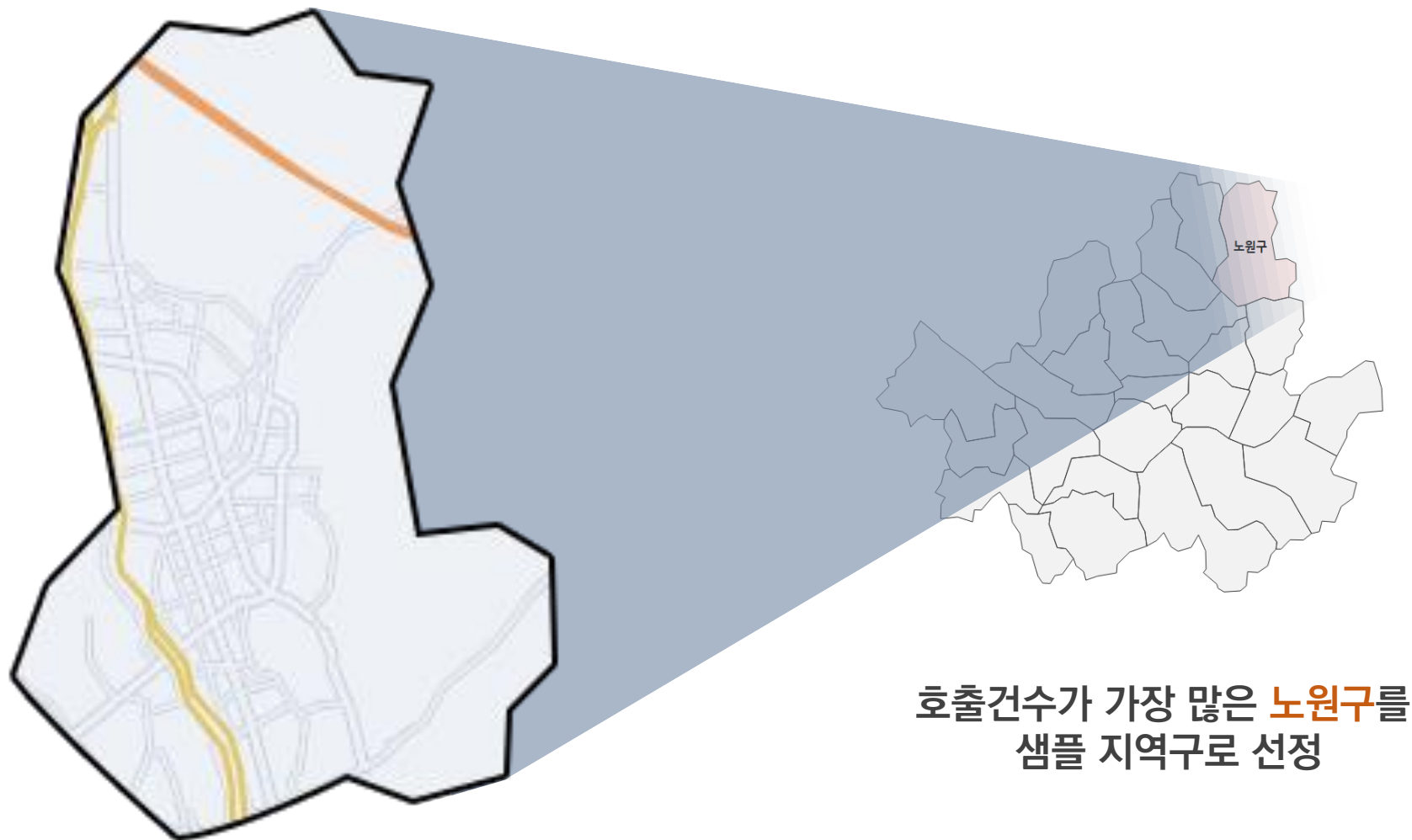
: HiveQL Query

```
select dept_gu, dest_gu,  
count(dept_gu)  
from cst_18_utf  
group by dept_gu,  
dest_gu  
oreer by dept_gu;
```

노원구에서 출발-도착하는
경우가 가장 많으며,
비교적 이동거리가
짧은 패턴이 나타난다.

데이터 분석

샘플 지역구 선정



출발지-목적지별 콜택시 수요 (노원구 內)



```
select dept_dong,
dest_dong,
count(dest_dong) as
cnt_dd
from cst_18_utf
where dept_gu = "노원구"
and dest_gu = "노원구"
group by dept_dong,
dest_dong
order by cnt_dd desc;
```

노원구內출발-도착지는
상계6.7동 -> 하계1동
하계1동 -> 상계 6.7동으로
이동하는 경우가 많으며
비교적 짧은 거리의
이동패턴이 존재한다

결론

평균 대기시간이 가장 긴 지역은 금천구 > 중구 > 동작구 > 관악구 > 종로구 순

콜이 집중되는 시간은 0~1시, 9시, 17~19시, 평균 콜대기시간 **최대93분** **최소20분**

강동, 강북, 마포, 서대문, 용산구는 16~19시,
강남, 광진, 서초, 중구 17~20시에 평균대기시간 ↑

콜택시 수요가 많은 지역은 노원구 > 강서구 > 서대문구 > 은평구 > 마포구 순

노원구에서 출발-도착하는 경우가 가장 많고, 비교적 이동거리가 짧음

노원구內 출발-도착지는 상계6.7동 -> 하계1동, 하계1동 -> 상계 6.7동으로
이동하는 경우가 많으며, 비교적 짧은 거리의 이동패턴이 존재

단거리 고정수요는 콜택시 외의 다른 교통수단으로 대체할 수 있지 않을까?

