

# 택시운행분석 데이터셋 활용가이드

2015년 1월



서울시 정보기획단

※ 본 데이터셋은 서울시내 택시들의 과거 1년간 운행정보에 대한 통계를 바탕으로 제작되었으며, 택시 운행이 많은 위치를 예측하는 데 활용하실 수 있습니다.

# 목차

1. 개요
  - 1.1 데이터셋 구축 목적
  - 1.2 데이터셋의 활용
  - 1.3 데이터 처리 흐름도
  - 1.4 업데이트 주기
2. 데이터셋 설명
  - 2.1 파일 정보
  - 2.2 분석 자료
  - 2.3 데이터셋 예시
  - 2.4 용어 설명
  - 2.5 데이터셋 필드 설명
  - 2.6 파일 데이터 구성정보
3. 활용 방안
  - 3.1 데이터셋 활용 가이드
  - 3.2 Open API 활용 가이드
  - 3.3 활용 시나리오

# 1. 개요

## 1.1 데이터셋 구축 목적

택시 이용승객과 택시기사에게 정보 제공

- 서울 시민에게는 택시 타기 좋은 위치 정보
- 서울시 택시기사에게는 승객이 많은 위치 정보



## 1.2 데이터셋의 활용

‘택시 타기 좋은 곳’ 정보제공서비스, ‘승객 많은 곳’ 정보제공서비스 등을 개발할 수 있도록 조건별 택시 승하차 정보 데이터셋 제공



데이터셋  
다운로드



민간 서비스 개발  
(포털사이트 업체, 내비게이션 업체,  
개인 개발자 등)



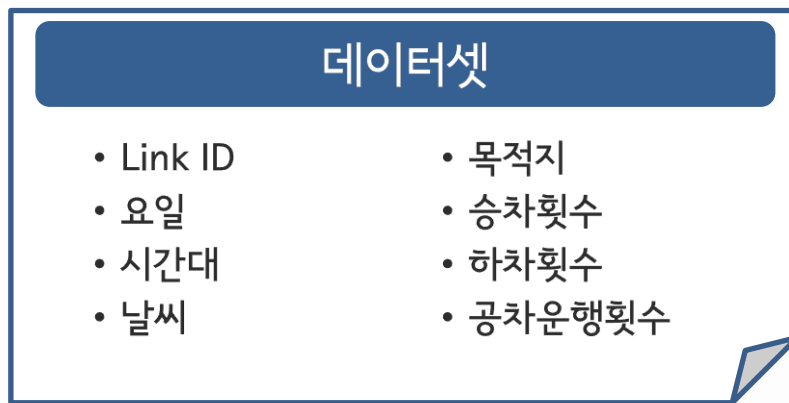
서울시민 / 택시기사  
서비스 이용



택시 타기 좋은 곳 & 승객 많은 곳 정보 획득

# 1. 개요

## 1.3 데이터 처리 흐름도



## 1.4 업데이트 주기

한 달에 한번 (매월 말)

# 2. 데이터셋 설명

## 2.1 파일 정보 (내역 → 6페이지)

분류	파일명	용량
데이터셋	TaxiMatch_Link_Dataset_Full_YYYYMM.zip	~500MB (압축)
목적지 코드표	Dest_Code_YYYYMM.mdb	~1MB
표준노드링크 ID 매핑 테이블	KSLink_ID_MappingTable_YYYYMM.mdb	~20MB
활용가이드	택시운행분석_활용가이드_YYYYMM.pdf	~2MB
150M 링크정보	Link_WGS84_Link_Info_150M.txt	~5MB
서울지역 도로링크 SHP	Seoul_150M_Only.zip	~4MB

## 2.2 분석 자료

항목	분석자료
택시 운행분석 데이터	과거1년치 택시 운행 데이터를 분석에 활용하며, 2014년 9월 이전은 법인택시만, 10월 이후는 개인택시도 포함하여 분석 (예: 2014.12 → 2013.12~ 2014.11 1년치데이터 2015.01 → 2014.01~ 2014.12 1년치데이터 )
표준노드링크	<a href="http://nodelink.its.go.kr">http://nodelink.its.go.kr</a> 의 최신 표준노드링크 사용 (현, 2014년 9월 10일 버전)

## 2.3 데이터셋 예시

T_Link_ID	Day	Time	Weather	Dest <sup>주1)</sup>	CntOn	CntOff	CntEmp
링크ID	요일	시간대	날씨	목적지	승차 횟수	하차 횟수	공차운행 횟수

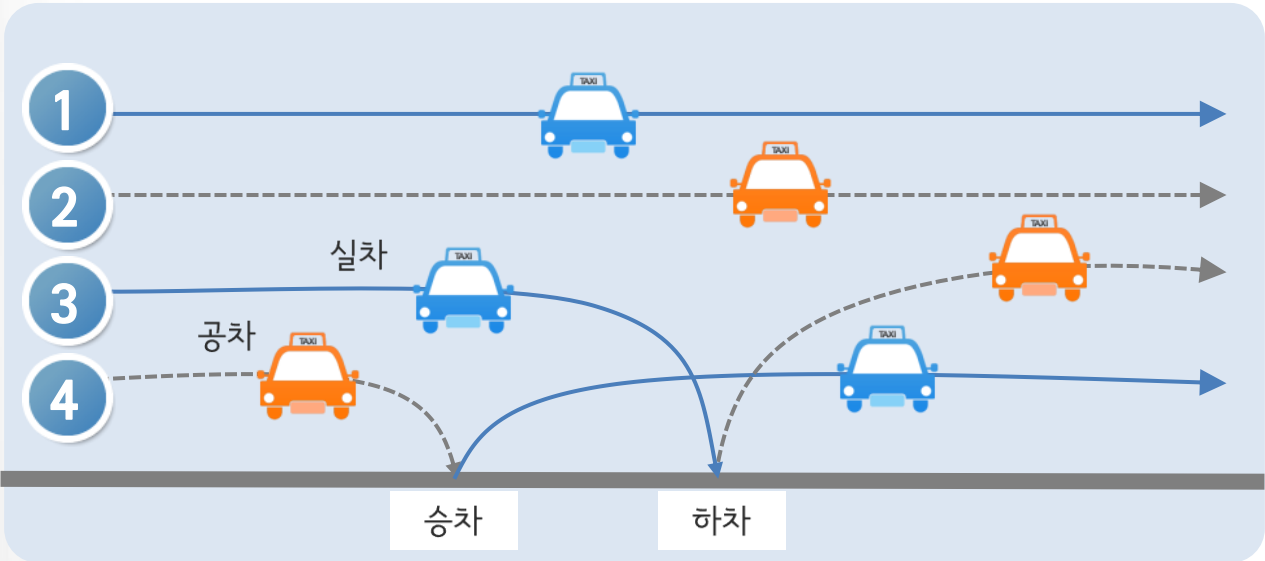
Primary Key

T\_Link\_ID,Day,Time,Weather, Dest,CntOn,CntOff,CntEmp  
T10001,1,0,1,1114,512,0,0  
T10001,1,0,1,1123,254,0,0  
T10001,1,0,1,1168,312,0,0  
T10001,1,0,1,,112,15685

Example

주1) 목적지(Dest)는 승차지점이 아닌 하차지점의 행정구역(시군구)을 의미, 따라서 승차횟수에만 해당되며 하차횟수와 공차운행횟수의 경우는 Null

## 2.4 용어 설명



	상태	승차	하차	공차운행	실차운행
1	실차운행				1
2	공차운행			1	
3	실차운행 → 하차 → 공차운행		1	1	1
4	공차운행 → 승차 → 실차운행	1		1	1

- 승차 : 공차 상태로 있다가 손님을 태운 시점의 상태
- 하차 : 실차 상태로 있다고 손님이 내린 시점의 상태
- 공차운행 : 공차 상태로 주행 (정지상태 포함)
- 실차운행 : 실차 상태로 주행 (정지상태 포함)

## 2.5 데이터셋 필드 설명 (TaxiMach\_Link\_Dataset\_Full\_YYYYMM.zip)

T\_Link\_ID

해당지점의 링크 ID

Day

일(1)

월(2)

화(3)

수(4)

목(5)

금(6)

토(7)

Time

(24시간/ 30분 단위  
⇒ 48개)

00

(00시~00시30분)

01

(00시 30분 ~01시)

02

(01시~01시30분)

...

47

(23시30분~24시)

Weather

맑음(1)

비(2)

눈(3)

정보없음(-1)

Dest

하차지점의 시군구코드 ※ 승차횟수가 입력된 레코드만 해당

CntOn

승차횟수

CntOff

하차횟수

CntEmp

공차운행횟수

※ 목적지 코드 (Dest\_Code\_YYYYMM.mdb)

SiGunGu\_cd

Dest(하차지점)의 시군구 코드

SiGunGu\_name

Dest(하차지점)의 시군구명

※ 표준노드링크 ID 매핑 (KSLink\_ID\_MappingTable\_YYYYMM.mdb)

T\_Link\_ID

표준노드링크를 150m로 분할한 후 재부여한 ID

Link\_ID

표준노드링크 ID

## 2.6 파일 데이터 구성정보

150M 링크정보 (Link_WGS84_Link_Info_150M.txt)	
T_LINK_ID	150M 링크ID
X_MAX	X좌표 끝점
Y_MAX	Y좌표 끝점
X_MIN	X좌표 시작점
Y_MIN	Y좌표 시작점
X_PART	파트X좌표
Y_PART	파트Y좌표

표준노드링크 ID 매핑 테이블 (KSLink_ID_MappingTable_YYYYMM.mdb)	
LINK_ID	링크ID
T_LINK_ID	150M 링크ID

택시운행분석 데이터 (TaxiMatch_Link_Dataset_Full_YYYYMM.zip)	
T_LINK_ID	150M 링크ID
DAY	요일
TIME	시간
WEATHER	날씨
DEST	목적지
CNT_ON	승차건수
CNT_OFF	하차건수
CNT_EMP	공차건수

목적지코드표 (Dest_Code_YYYYMM.mdb)	
SIGunGu_Cd	시군구코드
SIGunGu_Cd	시군구명



# 3. 활용 방안

## 3.1 데이터셋 활용 가이드

### 1) 데이터 다운로드 및 공간 매핑



① 서울열린데이터광장(<http://data.seoul.go.kr>)에서 데이터셋을 다운받아 DB화한다.

- 메뉴 : Dataset > Dataset 목록

② 지능형교통체계관리시스템 표준노드링크 자료실(<http://nodelink.its.go.kr>)에서 표준노드링크를 다운받는다.

- 메뉴 : 표준노드/링크 자료실

③ 다운받은 데이터셋의 LinkID와 표준노드링크의 LinkID를 매핑한다.

# 3. 활용 방안

택시운행분석 데이터셋

## 2) 데이터 추출 및 가공



- ① 주요 조건(위치, 요일, 시간대), 기타 조건(날씨, 목적지) 등 필요한 조건을 사용자가 직접 입력하거나 기기로부터 정보를 받아온다.
- ② 현재위치 또는 원하는 위치에서 가장 가까운 Link를 LinkID 목록에서 가져온다.
- ③ 데이터셋에서 원하는 조건에 해당하는 데이터를 검색해서 가져온다.
- ④ 데이터셋의 데이터는 LinkID별 > 요일별 > 시간대별로 세분화되어 있기 때문에 원하는 조건에 맞게 데이터를 가공한다.

# 3. 활용 방안

## 3) 서비스별 활용 데이터 예시

승객 중심의 서비스의 경우  
(빈 택시가 있을 가능성이 높은 위치 안내)

공차운행횟수에서 승차횟수를 뺀 수치를 활용

택시기사 중심의 서비스의 경우  
(손님을 태울 가능성이 높은 위치 안내)

승차횟수 수치 활용

## 4) 지도 표출 예시



※ 색상 표현단계는 도로간 우선순위 구별을 위해 5단계 이상이 권장됨

## 3.2 Open API 활용 가이드

1) 요청주소

<http://openAPI.seoul.go.kr:8088>

2) 요청인자

변수명	타입	변수설명	값설명
KEY	STRING(필수)	인증키	OpenAPI 에서 발급된 인증키
TYPE	STRING(필수)	요청파일타입	xml : xml, xml파일 : xmlf, 엑셀파일 : xls, json파일 : json
SERVICE	STRING(필수)	서비스명	ListTaxiDrivingDataset
X_LOC	INTEGER(필수)	검색할 X 좌표	정수 입력
Y_LOC	INTEGER(필수)	검색할 Y 좌표	정수 입력
DAY	STRING(필수)	요일	문자형 입력
TIME	STRING(필수)	시간(30분간격)	문자형 입력
WEATHER	STRING(필수)	날씨 (-1:날씨정보없음 1:맑음 2:비 3:눈)	문자형 입력

# 3. 활용 방안

## 3) 출력 값

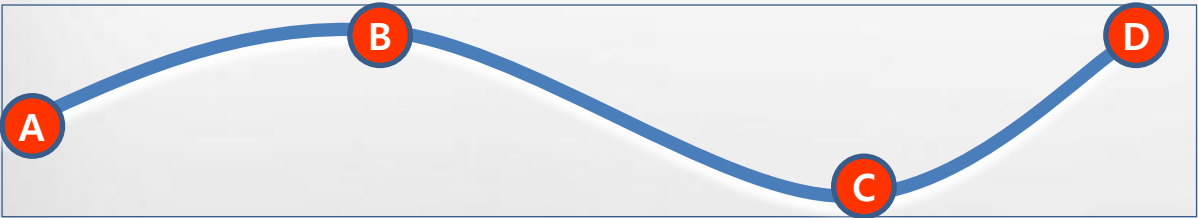
NO	출력명	출력 설명
1	T_LINK_ID	입력좌표의 반경100미터내의 150m기준 링크ID Ex) T_180743
2	X_PART	해당링크의 파트X좌표(링크에서 나누어진 파트의 X좌표) Ex) 127.0336730^127.0335211^127.0334246^127.0333606^
3	Y_PART	해당링크의 파트Y좌표(링크에서 나누어진 파트의 Y좌표) Ex) 37.5291603^37.5287213^37.5282681^37.5278763^
4	DAY	요일
5	TIME	시간(30분간격)
6	WEATHER	날씨 (-1:날씨정보없음 1:맑음 2:비 3:눈)
7	DEST	목적지
8	CNT_ON	승차건수
9	CNT_OFF	하차건수
10	CNT_EMP	공차건수

※ T\_LINK\_ID,X\_PART, Y\_PART 설명

T\_LINK\_ID => A~D(T\_180743)

파트는 T\_LINK\_ID 의 세부화된 여러 개의 위치정보  
각 ‘^’ 기호마다 하나의 파트위치정보

파트 A지점의 좌표 127.0336730, 37.5291603  
파트 B지점의 좌표 127.0335211, 37.5287213  
파트 C지점의 좌표 127.0334246, 37.5282681  
파트 D지점의 좌표 127.0333606, 37.5278763



## 4) 에러 및 정보메세지

에러값	에러설명
INFO-000	정상 처리되었습니다
ERROR-300	필수 값이 누락되어 있습니다. 요청인자를 참고 하십시오.
INFO-100	인증키가 유효하지 않습니다. 인증키가 없는 경우, 열린 데이터 광장 홈페이지에서 인증키를 신청하십시오.
ERROR-301	파일타입 값이 누락 혹은 유효하지 않습니다. 요청인자 중 TYPE을 확인하십시오.
ERROR-310	해당하는 서비스를 찾을 수 없습니다. 요청인자 중 SERVICE를 확인하십시오.
ERROR-331	요청시작위치 값을 확인하십시오. 요청인자 중 START_INDEX를 확인하십시오.
ERROR-332	요청종료위치 값을 확인하십시오. 요청인자 중 END_INDEX를 확인하십시오.
ERROR-333	요청위치 값의 타입이 유효하지 않습니다. 요청위치 값은 정수를 입력하세요.
ERROR-334	요청종료위치 보다 요청시작위치가 더 큼니다. 요청시작조회건수는 정수를 입력하세요.
ERROR-335	샘플데이터(샘플키:sample) 는 한번에 최대 5건을 넘을 수 없습니다. 요청시작위치와 요청종료위치 값은 1 ~ 5 사이만 가능합니다.
ERROR-336	데이터요청은 한번에 최대 1000건을 넘을 수 없습니다. 요청종료위치에서 요청시작위치를 뺀 값이 1000을 넘지 않도록 수정하세요.
ERROR-500	서버 오류입니다. 지속적으로 발생시 열린 데이터 광장으로 문의(Q&A) 바랍니다.
ERROR-600	데이터베이스 연결 오류입니다. 지속적으로 발생시 열린 데이터 광장으로 문의(Q&A) 바랍니다.
ERROR-601	SQL 문장 오류 입니다. 지속적으로 발생시 열린 데이터 광장으로 문의(Q&A) 바랍니다.
INFO-200	해당하는 데이터가 없습니다.



# 3. 활용 방안

## 3.3 활용 시나리오

### 1) 승객 입장 서비스 제공

- 스마트폰 앱을 활용하여 서울시민에게 택시 잡기 쉬운 곳 등의 정보 제공

처음 온 장소, 낯선 장소  
익숙하지 않은 장소,  
**어떡하지?**



### 2) 택시기사 입장 서비스 제공

- 스마트폰 앱과 내비게이션을 활용하여 택시기사에게 승객 많은 곳 등의 정보 제공

공차, 익숙하지 않은 지역,  
되돌아가야 하는 택시  
**어떡하지?**

