

# 서식 1

## 2020 빅데이터 분석 및 아이디어 공모전 참가신청서

				접수번호	
지원형태	<input type="checkbox"/> 개인 <input type="checkbox"/> 팀		지원부문	<input type="checkbox"/> 분 석 <input type="checkbox"/> 아이디어	
신청(대표)자명	국대전		생년월일		
주 소					
전화번호			E-Mail		
소 속	<input type="checkbox"/> 학생(학교 : 국민대학교) <input type="checkbox"/> 직장(부서: ) <input type="checkbox"/> 기타( )				
구 분	팀원(1)	팀원(2)	팀원(3)	팀원(4)	
성 명	김	기세현	김	원	
소속기관	국민대학교	국민대학교	국민대학교	국민대학교	
휴 대 폰					

신청인은 『빅데이터 분석 및 아이디어』 공모전 참가신청서를 첨부와 같이 제출하며  
 유의사항을 충분히 숙지하였으며 공모전 진행을 위한 요청사항에 대한 수행과 관련  
 법령 및 규정, 운영요령을 준수할 것을 동의 합니다.

2020 년 05월 29일

신청인(대표자)

**[재)대전정보문화산업진흥원장 귀하**

개인정보 수집이용 동의서 및 참가 서약서			
개인정보 수집·이용 동의	<p>1. 정보 수집·이용 범위와 사용목적</p> <ul style="list-style-type: none"> <li>■ 수집·이용범위 : 성명, 소속, 생년월일, (주민등록등본상)주소, 연락처</li> <li>■ 사용목적 : 『2020년 빅데이터 분석 및 아이디어 공모전』 참가신청 및 참가자 관리, 수상자 선정, 후속지원, 자료제출, 변경사항 관리 등 공모전 운영·관리 목적 제공받는 자 : 공모전 주최/주관 기관</li> <li>■ 보유 및 이용기간 : 제공된 날로부터 동의철회 시(최대5년) 보유</li> </ul> <p>2. 개인정보 수집·이용에 동의하지 않을 권리 및 동의하지 않을 경우의 불이익</p> <ul style="list-style-type: none"> <li>■ 정보주체는 『2020년 빅데이터 분석 및 아이디어 공모전』에 개인정보 수집·이용의 동의를 거부할 권리가 있습니다.</li> <li>■ 개인정보 수집·이용에 동의하지 않을 경우에는 본 공모전에 참가신청이 불가합니다.</li> </ul> <p>3. 이용자 개인정보보호를 위하여 수집된 개인정보는 업무 이외의 다른 목적으로 사용하지 않습니다.</p> <p style="text-align: center;"> <input type="checkbox"/> 동의합니다.                     <input type="checkbox"/> 동의하지 않습니다.                 </p>		
	<p>참가 서약서</p> <p>상기 본인은 『2020년 빅데이터 분석 및 아이디어 공모전』에 출품함에 있어 개인정보 수집·이용에 동의하며, 다음 각 호의 규정을 준수할 것을 서약합니다.</p> <ol style="list-style-type: none"> <li>1. 본 공모의 제반규정을 준수하며, 이를 준수하지 않을 경우 어떠한 조치도 감사한다.</li> <li>2. 출품작이 이미 발표된 작품이거나, 타 공모전에 수상한 작품일 경우 심사대상에서 제외하고, 당선작 발표 이후라도 수상을 취소하며, 상금은 환수조치 될 수 있다.</li> <li>3. 출품작에 대한 저작권으로 인하여 발생하는 민·형사상 책임은 출품자에게 있다.</li> <li>4. 공모전의 선정방법 및 절차, 그리고 최종결과에 대해 추후 어떠한 이의를 제기하지 않는다.</li> <li>5. 공모전 진행에 필요한 요구사항 및 향후 지원관리를 위한 성과조사에 성실히 응한다.</li> </ol> <p>참가서약서 및 신청서 내용이 사실임을 확인하며, 허위사실 기재 등으로 인하여 어떠한 문제가 발생했을 시 모든 책임은 본인에게 있음 확인합니다.</p>		
<p>2020년 05월 29일</p>			
	서약인:	(팀장)	
		팀원 I	기 세현
		팀원 II	
		팀원 III	
<p><b>[재)대전정보문화산업진흥원장 귀하</b></p>			

## I. 참가자 정보

공 모 명	대전 심야버스 노선 입지분석 및 선정		
성 명	국대전	지원부분	분석
소 속	국민대학교		
연 락 처	휴대폰	E-mail	

## II. 세부 내용

### ○ 분석 개요

#### 1-1. 분석 목적

최근 대전 광역시에 관한 기사 및 ‘시민의 소리’ 페이지를 통해서 시민들의 민원을 분석해본 결과 교통에 대한 불편이 높은 비중을 차지하고 있습니다. 이후 교통에 대한 세부적인 주세 선정을 위해 시민들을 대상으로 한 설문조사 결과 버스에 관한 불편과 그 중에서도 야간 버스 운행에 관한 불편이 많은 것을 확인한 후 본 팀은 회귀모델과 GIS 분석을 통해 대전시의 심야 버스 운행에 대한 입지 분석 및 노선 선정을 주제로 선정하였습니다.

#### 1-2. 배경 및 필요성

대전 광역시는 대한민국 정중앙에 위치해 있으며 과거부터 육상교통의 요지로서 내륙의 교통 허브 역할을 담당해왔습니다. 이에 따라 대전은 ‘대중교통 중심도시’를 기조로 각종 정책을 추진해왔고 최근 BRT, 트램 등 다양한 대중교통 사업과 정책을 진행중입니다.

하지만 주제 선정 단계에서 대전 시민들이 교통 불편에 대한 민원이 많은 것을 알 수 있었습니다. 본 팀은 심층적인 원인 분석을 위해 대전 시민들(55명)을 대상으로 설문조사를 실시해 보았습니다. 그 결과로 교통에 대한 불편중에서 이른 막차시각으로 인한 불편감에 대한 응답이 제일 높은 것을 확인 할 수 있었고 이를 바탕으로 대전시 심야버스 노선 입지 분석 및 선정으로 주제를 결정했습니다.

분석에 앞서 본 팀은 부족한 기초 자료를 보완하기 위해 대전 시내 교통과 관련된 기사를 수집하고 시내 교통에 대한 논문들을 찾아보았습니다. 그중에서 분석 주제와 관련이 있는 몇 개의 자료를 선정하여 다음과 같은 정보를 얻게 되었습니다.

18년 3월 23일 충청뉴스에 실린 기사<sup>(1)</sup>에 따르면 2018년 당시 대전시의 대다수 버스 노선의 막차 시각이 오후 11시 이전이었다는 것을 알 수 있었습니다. 그러나 현재도 대부분의 버스가 오후 11시 이전에 운행이 중단되는 것을 추가적인 조사를 통해 확인하였습니다. 밤 늦게 버스를 이용할 일이 많은 대학생 및 직장인들에게는 심야시간대의 버스 운영에 대한 필요성이 높다고 할 수 있습니다. 또한 대전시에서도 빅데이터를 이용한 심야버스로의 전환도 내부적으로 검토했다고 언급하고 있습니다. 이러한 이유로 본 분석의 주제가 시의적절하다고 판단하였습니다.

2013년 대한산업공학회 추계학술대회 논문집에 실린 『Analytic Hierarchy Process(계층분석과정)를 이용한 심야버스 순환노선 설계』<sup>(2)</sup>에 따르면 학업 및 회식 등으로 인해 상대적으로 늦게 귀가하는 시민들과 야간근로자 등에게 심야 버스가 많은 혜택을 주고 있다는 것을 알 수 있었습니다. 이에 본 팀 또한 심야버스가 대전 시민들의 야간 경제활동 보장은 물론 소득약자에게 적합한 수단이 될 수 있다고 판단하였습니다. 추가적으로 AHP(계층분석과정)의 결과 산출에서 아이디어를 착안하여 수학적 모델을 활용한 심야버스 수요 예측을 통해 정교하고 효율적인 심야버스 운영이 가능할 것이라고 판단하였습니다.

2015년에 발표된 『경기도 심야버스 운행체계 개선방안 연구』<sup>(3)</sup>에 따르면 심야버스 운영에 대해 서는 심야시간대에 승차인원 수요를 고려하여 적정 수준의 심야버스 서비스가 탄력적으로 제공될 수 있도록 심야버스의 운행체계의 조정이 필요하다는 것을 확인했습니다. 이에 따라 본 팀은 회귀분석을 통한 수학적 모델을 구축해 심야시간대의 승차인원 수요가 높을 것으로 예측되는 행정동 및 정류소를 우선 도입 지점으로 선정하였습니다.

2019년에 발표된 『서울시 심야버스 정책은 어떻게 성공적으로 도입되었는가?: 정치적 관리의 관점을 중심으로』<sup>(4)</sup>는 서울시의 성공적인 심야전용 올빼미 버스 도입 및 운영이 정치적 관리 측면에서 어떻게 가능했는지 언급하고 있습니다. ‘대중교통 중심도시’를 꿈꾸는 대전시가 심야버스 도입 및 운영 이후 나아가야 할 정치적 관리의 방향을 알 수 있었습니다.

위와 같은 기사와 논문자료의 내용을 바탕으로 본 팀은 모델 구축 이전에 대전 시내의 교통 문제 현황과 시내의 여러 가지 조건들에 대한 분석을 먼저 진행하였습니다.

## ○ 분석 결과 상세 내용

### 1. 시내 교통 문제 현황과 대전광역시 분석

#### 1-1. 교통 분야 민원 키워드 분석

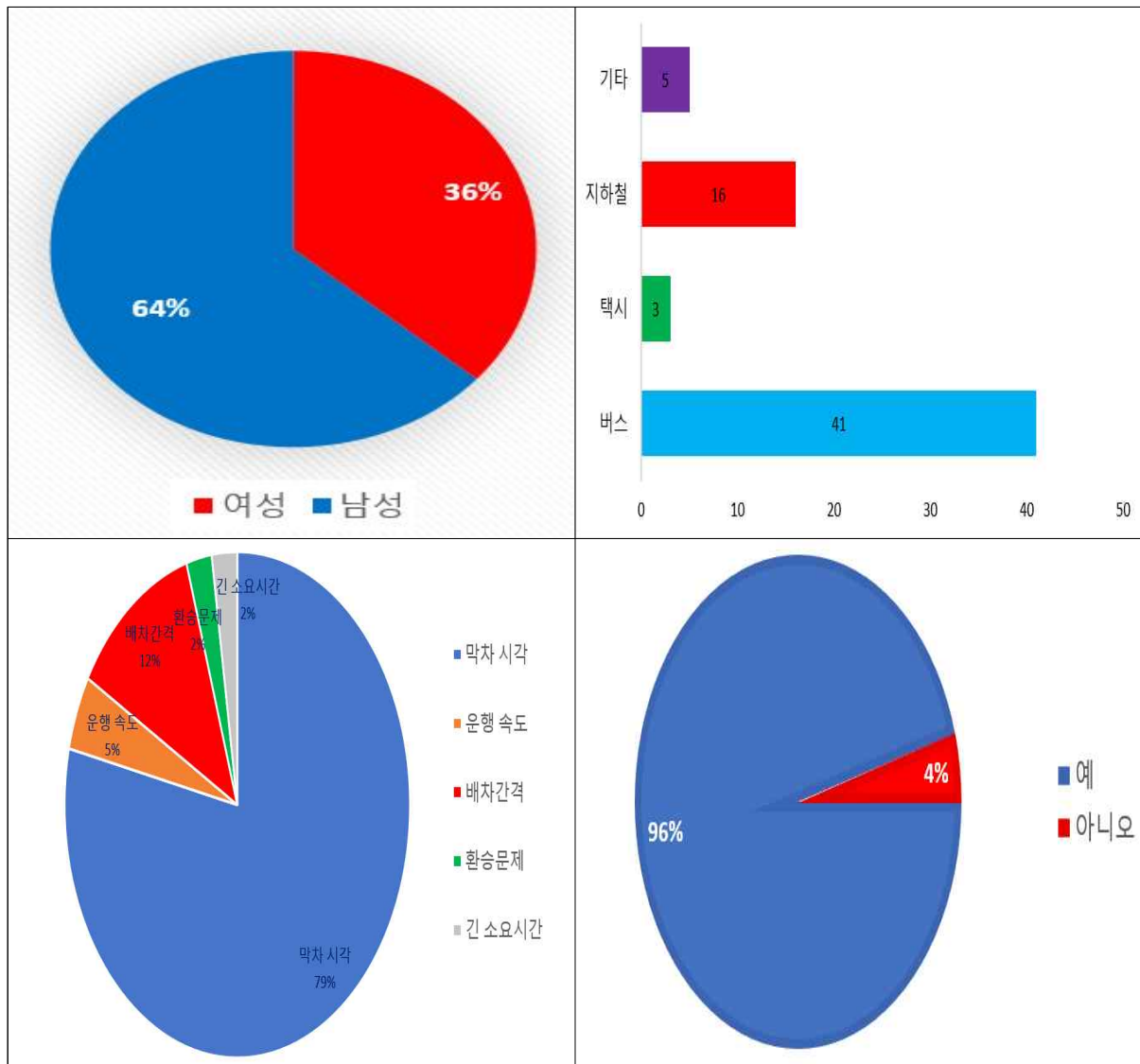


[그림 1]

본 팀은 분석 진행 전 주제 선정을 위해 대전시청 홈페이지의 ‘시민의 소리’에서 대전 시민들의 불편사항을 확인하였습니다. 민원에 관한 내용을 키워드화하여 워드클라우드 형태로 시각화하였습니다. 그 결과로 위의 [그림1]과 같이 버스, 지하철, 도로, 자전거 등의 교통과 관련된 키워드들이 많이 나타나는 것을 확인하였습니다.

이후 민원들의 구체적인 내용을 확인하기 위해 ‘시민의 소리’ 교통 분야 민원 페이지를 추가적으로 살펴봤습니다. 실제로 앞서 언급한 키워드들에 대한 게시글이 많은 것을 알 수 있었습니다. 따라서 분석의 큰 주제를 대전시 교통 분야에 대한 문제점 해결로 선정하게 되었습니다.

## 1-2. 대전시민 대상 설문조사



[그림 2]

추가적으로 대전 시내의 교통문제에 대한 심층적인 분석을 하기 위해서 대전 시민들(55명)을 대상으로 구글폼을 이용하여 설문조사를 실시하였습니다. 설문조사 결과는 위 [그림 2]와 같으며 다음과 같은 결과를 얻어낼 수 있었습니다.

첫째 대전시의 대중교통 수단 중에서 버스에 대한 불편함(중복응답 허용)이 가장 높은 것을 알 수 있었습니다.

둘째 버스 이용에 불편을 겪은 경험이 있는 시민들의 가장 큰 불편 사항은 ‘이른 막차시각(73%)’과 ‘배차간격(12%)’이었습니다.

셋째 만약 대전시에서 심야 버스를 운행한다면 이용할 계획이 있는가에 대한 물음에도 96%의 긍정적인 응답을 확인할 수 있었습니다.

이러한 결과를 바탕으로 교통 분야 관련 민원중에서 시내버스에 대한 ‘이른 막차 시각’의 해결 방안으로 대전시 심야버스 노선 운영에 대한 입지분석 및 선정을 최종 주제로 선정하게 되었습니다.

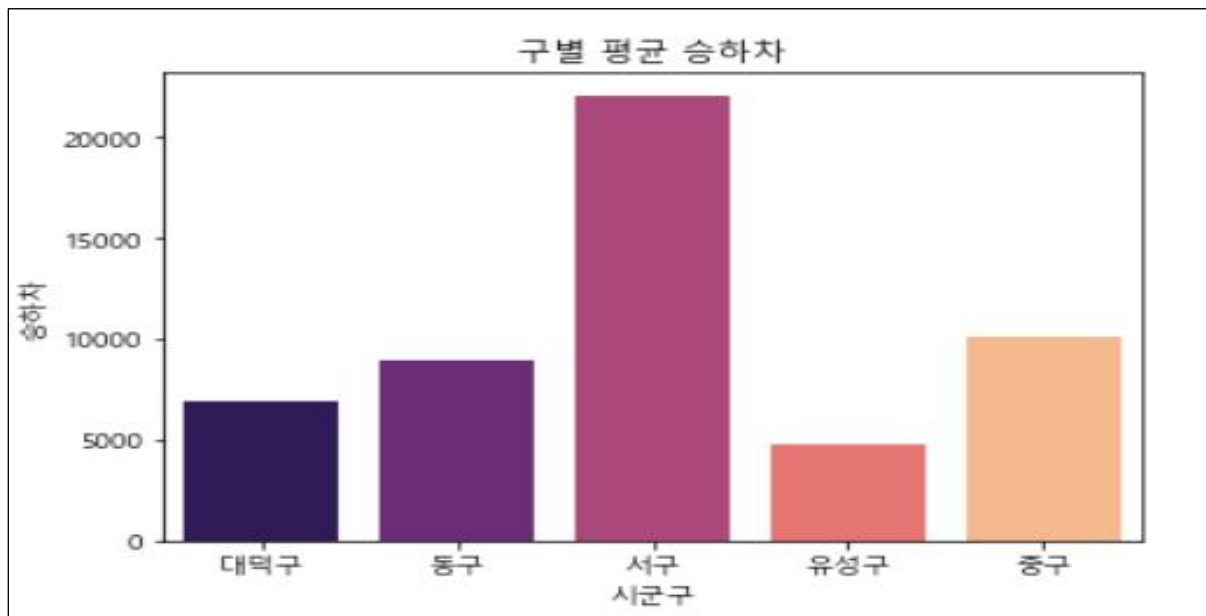
### 1-3. 대전광역시 분석

심야 버스 운영에 대한 세부적인 분석을 진행하기 전에 대전광역시의 각 구와 행정동별로 어떠한 특징을 가지고 있는지 대략적인 분석을 진행해보았습니다. 우선 대전시의 각 구별 2019년 유동인구수를 분석해본 결과는 아래의 [그림 3]과 같습니다.



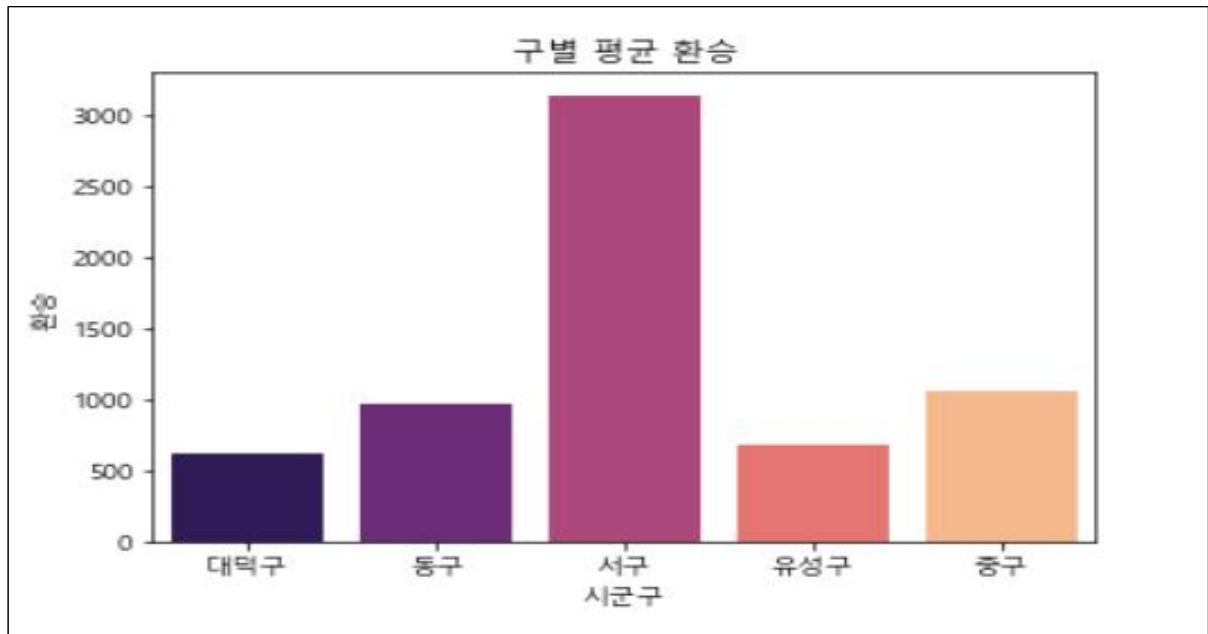
[그림 3]

대전시의 각 구별 평균 유동 인구수는 비슷한 양상을 보였지만 유성구는 다른 구와 달리 유독 높은 유동인구수를 보이는 것을 알 수 있었습니다. 따라서 유동인구와 승하차 인원이 비례할 것이라는 가정을 한 후 분석을 진행하였습니다.



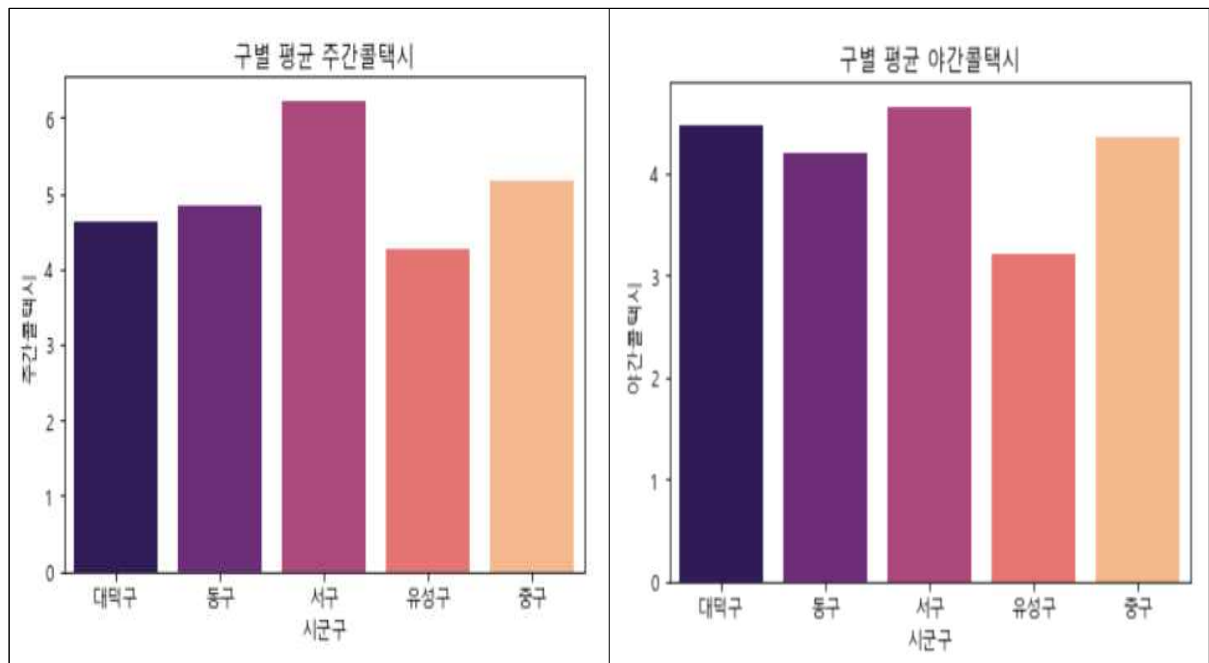
[그림 4]

위의 [그림 4]는 대전시의 하루 평균 구별 버스 승하차 인원 관련 것으로 서구의 승하차 인원이 많은 것을 확인할 수 있었습니다. 이는 다른 행정동에 비해서 버스 승하차 인원이 매우 많은 둔산1, 2, 3동이 모두 서구에 포함(이상치)되어 있기 때문이라고 판단하였습니다.



[그림 5]

위의 [그림 5]는 구별 하루 평균 환승 인원에 관한 그림으로 서구가 압도적으로 많은 것을 알 수 있었습니다. 탑승과 직접적으로 관련이 있는 승하차 인원과 평균 환승 인원은 유동인구의 양상과는 다른 형태를 띠는 것을 확인할 수 있었습니다.



[그림 6]

위의 [그림 6]은 구별 콜택시 이용 건수(건수/시간)를 주/야간으로 나누어 시각화한 것으로 구별 양상은 크게 다르지 않는 것으로 확인을 하였습니다. 주간에 비해 야간의 콜택시 이용 건수는 적었지만 콜택시 이용은 서구가 가장 많고 유성구가 가장 적은 것은 동일하였습니다.

위의 분석과정을 거치면서 일반적으로 유동인구가 많을 경우 대중교통 이용 인원이 많은 것으로 예상했지만 유동인구 양상과는 오히려 다른 양상을 보이는 것을 알 수 있었습니다. 이에 따라 본 팀은 버스 승하차 인원에는 유동인구 뿐만 아니라 다른 변수들이 영향을 미칠 수 있을 것이라고 판단하였고 다양한 데이터를 수집하여 향후 분석을 진행하게 되었습니다.



## 2. 회귀분석

### 2-1. 다변량 회귀 모델

대전시 버스 승하차 인원을 종속변수로 선정하고 승하차 인원에는 영향을 줄 것으로 예상되는 변수들을 탐색해 모델의 독립변수로 선정하였습니다. 해당 독립변수는 각 행정동에 대한 인프라 변수와 주/야간에 따른 변화를 확인할 수 있는 변수로 구분하였습니다. 공공데이터 포털을 이용하거나 크롤링 등을 통해 총 57개의 데이터를 수집하였고 모든 독립변수로 사용했을 때의 summary 결과는 [표 1], [표 2], [표 3], [표4]와 같습니다.

변수명	계수	표준오차	t-value	p-value
(Intercept)	14741.24	2240.01	6.581	1.62E-06
주간유동인구	5351.8	5966.1	0.897	0.37987
주간매출액	3347.05	3158.37	1.06	0.3013
주간콜택시	3627.15	1888.26	1.921	0.06843
유동인구_주간비율	-53.15	913.43	-0.058	0.95415
매출액_주간비율	-1746.94	765.74	-2.281	0.03307
콜택시_주간비율	262.91	561.46	0.468	0.64442

[표 1]

먼저 위의 [표 1]은 주/야간에 따라 값이 달라지는 변수들과 절편의 summary 결과입니다. 해당 범주에서의 유의한 변수는 절편, 주간콜택시, 매출액\_주간 비율로 나타났습니다.

변수명	계수	표준오차	t-value	p-value
대형상권	-483.91	922.01	-0.525	0.60519
매출액_건축관련업	-5505.57	2592.85	-2.123	0.04577
매출액_기타제조/도매	4969.3	2512.75	1.978	0.06124
매출액_레저/스포츠	-186.57	953.14	-0.196	0.84669
매출액_문화/취미	958.51	1630.12	0.588	0.5628
매출액_미용	-4344.9	2498.16	-1.739	0.09663
매출액_사무기기	111.37	603.92	0.184	0.85546
매출액_소매	134.55	1215.01	0.111	0.91288
매출액_숙박	911.9	1258.08	0.725	0.47655
매출액_신변감화	1055.74	1100.32	0.959	0.34823
매출액_용역서비스	785.06	1156.08	0.679	0.50451
매출액_유아교육기관	1682.49	882.42	1.907	0.07034
매출액_유흥	-1126.25	2000.51	-0.563	0.57941
매출액_음식료품	198.56	1043.19	0.19	0.85087
매출액_의료기관/제약	1456.57	2751.18	0.529	0.60206
매출액_의류	133.27	991.43	0.134	0.89435
매출액_일반음식점	-354.9	2898.08	-0.122	0.9037
매출액_자동차정비/유지	-521.89	826.83	-0.631	0.53472
매출액_전자제품	574.71	1142.73	0.503	0.62025
매출액_주방용품	-920.26	621.85	-1.48	0.15375
매출액_주유	-511.54	637.97	-0.802	0.43163
매출액_학습자재	-1105.66	1214.08	-0.911	0.37279
매출액_학원	-934.67	1151.02	-0.812	0.42588
매출액_휴게음식점	1306.51	937.09	1.394	0.17783
유동인구_고령	-1093.74	2390.11	-0.458	0.65193
유동인구_20대	-5111.45	4614.04	-1.108	0.28047

[표 2]

위의 [표 2]는 행정동별 인프라 변수(매출액 / 유동인구 관련)들로 유의한 변수들은 매출액\_건축관련업, 매출액\_기타제조/도매, 매출액\_미용, 매출액\_유아교육기관, 매출액\_휴게음식점으로 나타났습니다.



변수명	계수	표준오차	t-value	p-value
버스환승	15085.08	1257.76	11.994	7.35E-11
지하철역	-1600.12	590.92	-2.708	0.01318
타슈_대여	13875.04	4870	2.849	0.00961
타슈_반납	-13661.2	4813.87	-2.838	0.00985
자동차등록대수	-1398.71	2183.61	-0.641	0.52874
콜택시	-3652.54	2187	-1.67	0.10974
도서관	471.79	1169.04	0.404	0.69061
대학교	341.37	739.1	0.462	0.64893
고등학교	-1333.87	987.31	-1.351	0.19107
중학교	1139.75	1181.93	0.964	0.34587
초등학교	-627.8	1734.28	-0.362	0.72097
유치원	2639.26	1300.2	2.03	0.05523
인구수	391.61	2465.73	0.159	0.87533
인구_남성비율	-239.87	742.26	-0.323	0.74977
인구_20대비율	1611.73	829.94	1.942	0.06567
인구_고령자비율	-31.97	1188.25	-0.027	0.97879
인구_고소득자비율	-483.32	1012.76	-0.477	0.63813
인구_밀도	629.36	665.12	0.946	0.35479
세대수	709.41	1521.83	0.466	0.6459
공시지가	1456.29	1601.31	0.909	0.37344
토지면적	-330.73	768.99	-0.43	0.67151
대덕구	-1962.01	3190.51	-0.615	0.54519
동구	-602.8	2758.37	-0.219	0.82912
서구	-2862.28	3610.83	-0.793	0.43682
유성구	-9683.39	4668.21	-2.074	0.05053

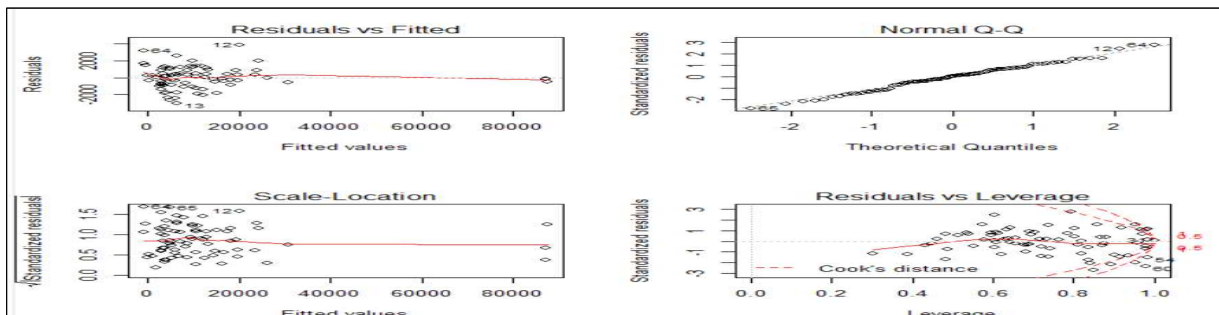
[표 3]

위의 [표 3]은 행정동별 인프라 변수(교통 / 학업 / 인구 / 토지 / 지역)들로 유의한 변수들은 버스환승, 지하철역, 타슈\_대여, 타슈\_반납, 유치원, 인구\_20대비율, 유성구(더미변수)로 나타났습니다.

Multiple R-Squared	0.9939	F-통계량	59.55
Adjusted R-Squared	0.9772	p-value	3.98E-15

[표 4]

위의 [표 4]는 앞서 언급된 57개 변수가 모두 독립변수로 들어간 모델의 결정계수, F-통계량, 유의 확률(p-value)을 나타낸 표입니다. 모델 자체는 유의하고 결정계수값이 99.39%이며 수정된 결정계수 값 또한 97.72%로 상당히 높은 것을 알 수 있었습니다. 하지만 유의하지 않은 변수가 존재했고 행정동이 79개이고 변수가 57개라는 점에서 과적합이 발생했다고 판단했습니다.



[그림 7]

위의 [그림 7]은 회귀모델의 정규성, 독립성, 등분산성을 살펴보기 위한 plot으로 시각화 결과 3개의 이상치 값(둔산 1,2,3동)을 발견하여 둔산동에 대한 데이터를 제외한 후 분석을 진행하였습니다.

변수명	VIF값	변수명	VIF값
주간유동인구	444.617	매출액_휴게음식점	10.9689
주간매출액	124.603	유동인구_고령	71.3577
주간콜택시	44.5379	유동인구_20대	265.93
유동인구_주간비율	10.422	버스환승	19.7604
매출액_주간비율	7.32436	지하철역	4.36173
콜택시_주간비율	3.9377	타슈_대여	296.252
대형상권	10.6187	타슈_반납	289.463
매출액_건축관련업	83.9769	자동차등록대수	59.5601
매출액_기타제조/도매	78.8681	콜택시	59.745
매출액_레저/스포츠	11.3478	도서관	17.0712
매출액_문화/취미	33.1927	대학교	6.82362
매출액_미용	77.9553	고등학교	12.1762
매출액_사무기기	4.5558	중학교	17.4498
매출액_소매	18.44	초등학교	37.5702
매출액_숙박	19.7705	유치원	21.1168
매출액_신변감화	15.1231	인구수	75.9443
매출액_용역서비스	16.6949	인구_남성비율	6.88207
매출액_유아교육기관	9.72654	인구_20대비율	8.60385
매출액_유흥	49.9905	인구_고령자비율	17.6369
매출액_음식료품	13.5934	인구_고소득자비율	12.8121
매출액_의료기관/제약	94.5456	인구_밀도	5.52593
매출액_의류	12.278	세대수	28.9291
매출액_일반음식점	104.912	공시지가	32.0298
매출액_자동차정비/유지	8.53951	토지면적	7.38657
매출액_전자제품	16.3113	대덕구	16.5905
매출액_주방용품	4.83024	동구	15.5471
매출액_주유	5.084	서구	34.0418
매출액_학습자재	18.412	유성구	33.0435
매출액_학원	16.549		

[표 5]

과적합을 개선하기 위해서 독립변수 간의 상관성으로 정확한 예측을 방해하는 다중공선성을 제거하였습니다. 위의 [표 5]는 앞서 모델에서 사용한 독립변수의 다중공선성(VIF) 값으로 10 이상의 값을 가지는 변수들은 정제 및 제거 과정을 거쳤습니다. 다중공선성 문제를 해결한 후에는 Adjusted R-squared 값을 기준으로 혼합선택법을 통하여 58개의 변수들 중 18개의 변수들을 선택하였습니다. 최종적으로는 모델을 설명하는데 유의하지 못한 변수들을 제거하여 9개 변수만을 사용한 최종모델을 만들었고 해당 모델의 변수와 summary 결과는 아래 [표 6]과 같습니다.

변수명	계수	표준오차	t-value	p-value
(Intercept)	9942.9	341.6	29.107	< 2E-16
주간유동인구	-1090.7	533.4	-2.045	0.044872
주간매출액	1813.7	526	3.448	0.000988
주간콜택시	610.6	340.1	1.795	0.077171
버스환승	5402.8	351.8	15.358	< 2E-16
지하철역	-1689.5	383.4	-4.407	3.95E-05
매출액_주유	-605.3	325	-1.862	0.067011
학업	2054.3	358	5.737	2.62E-07
인구_20대비율	1328.9	346.3	3.837	0.000281
유성구	-5205.3	1240.5	-4.196	8.28E-05

[표 6]



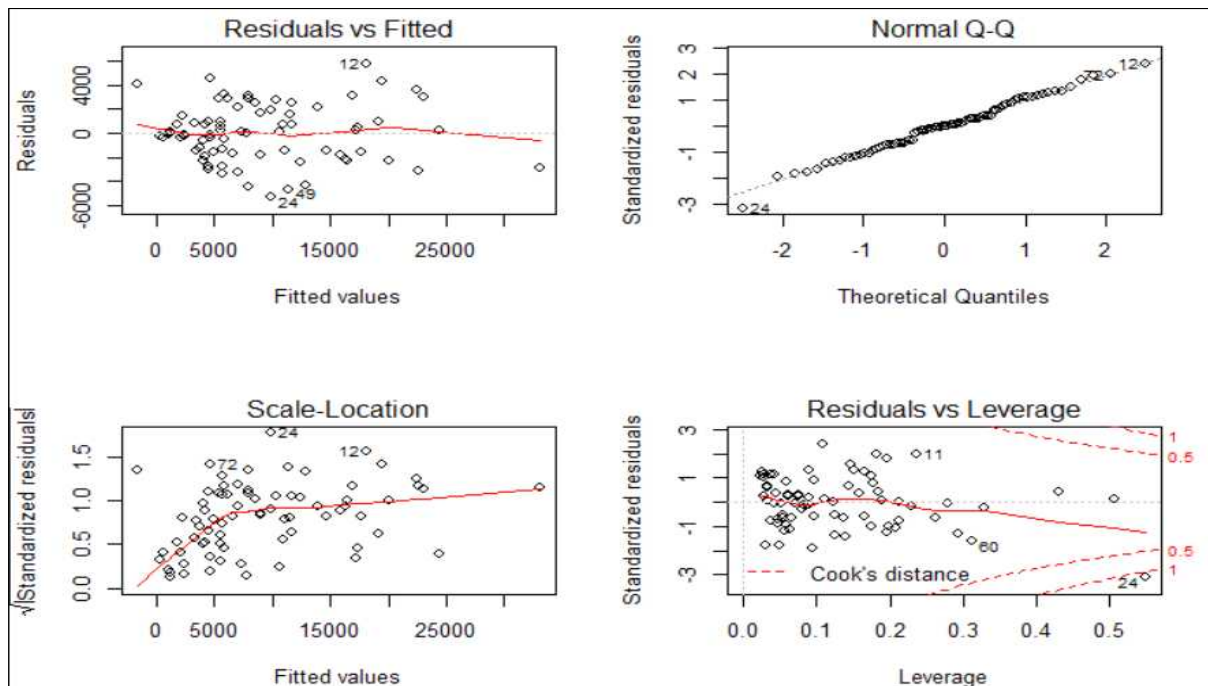
이후 해당 모델에 포함된 변수들의 해석을 진행하였습니다. 주간 유동인구와 버스 승하차 인원이 반비례한다는 것을 알 수 있었습니다. 또한 시간대별 유동인구의 합을 사용해 중복 인원이 존재하게 되어 동일한 장소에 오래 머무르는 경우 유동성이 감소하는 부분으로 해석된 것으로 판단하였습니다.

주간 매출액의 경우 해당 지역 방문이 선행되어야 하기 때문에 버스 승하차 인원 양의 영향을 끼치는 것을 알 수 있었습니다.

지하철은 버스와 같은 대중교통으로써 ‘지하철-버스’ 간의 환승이 아니라면 주로 버스의 대체재로 사용됩니다. 이로 인해 지하철역이 많을수록 버스 승하차 인원이 적다는 결과가 나온 것으로 판단하였습니다.

마지막으로 유성구의 경우는 5개의 구 중에서 버스 승하차 인원이 가장 적은 것으로 드러났는데 이러한 점이 모델에도 일정 부분 작용이 된 것으로 파악했습니다.

유의수준 0.1하에서 모든 변수들이 유의한 것을 알 수 있었고 아래 [그림 8]에서 볼 수 있듯이 초기 모델보다 개선이 된 것을 확인하였습니다.



[그림 8]

최종 모델의 결정계수, F-통계량, 유의확률(p-value)을 확인해본 결과 모델 자체도 유의한 것으로 나타났고 초기 모델보다 결정계수 값은 다소 떨어졌지만 모델의 독립변수가 모두 유의하다는 점을 고려했을 때 훨씬 적합한 모델이라고 판단하였습니다.

<b>Multiple R-Squared</b>	<b>0.8917</b>	<b>F-통계량</b>	<b>60.36</b>
<b>Adjusted R-Squared</b>	<b>0.8769</b>	<b>p-value</b>	<b>&lt;2.2E-16</b>

[표 7]

## 2-2. 모델 학습 및 예측

최종 모델을 선정한 후 야간 승하차 인원을 예측하기 위해 아래 [그림 9]와 같이 주간 데이터로 모델을 학습시키고 야간 데이터로 예측을 하였습니다. 심야 버스 운행 시간에는 버스 운행과 지하철 운행이 중지되므로 ‘야간유동인구’, ‘야간환승’, ‘야간지하철역’, ‘야간콜택시’ 변수가 주/야간에 따라 값이 달라지는 변수이며 나머지 변수들은 인프라 및 인구 관련 변수입니다.

	주간유동인구	주간매출액	주간콜택시	버스환승	지하철역	매출액_주유	학업	인구_20대비율	유성구
0	348519	12886456692	1	2825	1	392579178	0	0.073117	0
1	200896	7515060963	3	134	0	522564761	7	0.102381	0
2	190544	594417680	3	182	0	0	3	0.118735	0
3	170962	2064661752	3	888	1	0	10	0.123917	0
4	45699	650099661	6	1751	1	0	9	0.101048	0

	야간유동인구	야간매출액	야간콜택시	버스환승	지하철역	매출액_주유	학업	인구_20대비율	유성구
0	17375	980620493	1	0	0	392579178	0	0.073117	0
1	15772	711052339	3	0	0	522564761	7	0.102381	0
2	16939	50693805	3	0	0	0	3	0.118735	0
3	9119	189779403	3	0	0	0	10	0.123917	0
4	3239	33855934	5	0	0	0	9	0.101048	0

[그림 9]

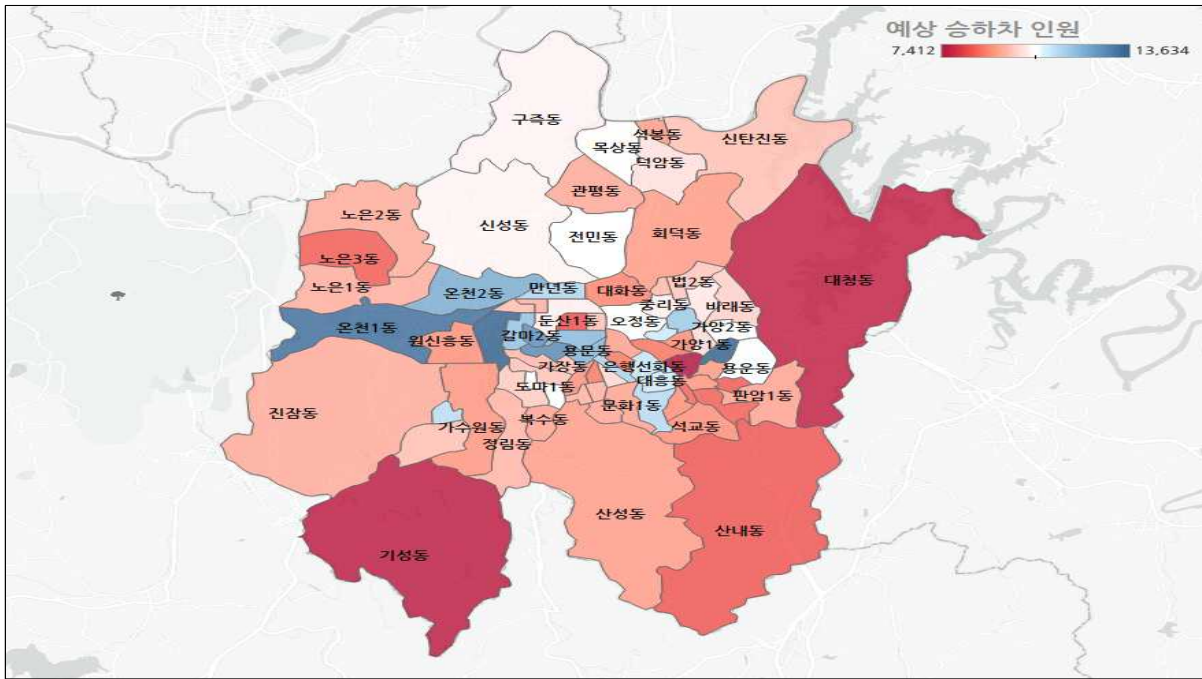
### 3. 입지 분석

#### 3-1. 행정동 선정

아래 [표 8]은 모델을 통해서 예측한 승하차인원 수를 기준으로 선정된 상위 10개동을 나타낸 것으로 아래의 10개 행정동과 이상치로 제외되었던 둔산 1,2,3동을 추가해 최종적으로 13개의 행정동을 선정하였습니다.

행정동	예측 인원	순위	행정동	예측 인원	순위
자양동	13633.8	1	온천2동	12195.9	6
월평1동	13626.12	2	용문동	12024.35	7
온천1동	13390.21	3	탄방동	11918.37	8
괴정동	12871.34	4	갈마1동	11765.96	9
갈마2동	12749.77	5	용전동	11628.82	10

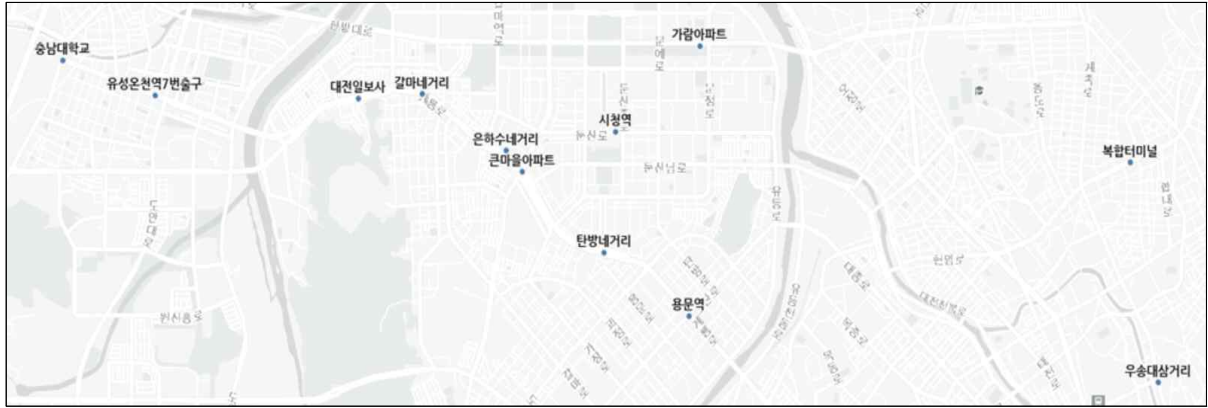
[표 8]



[그림 10]

위 [그림 10]은 예측 인원을 지도에 히트맵 형태로 매핑한 그림으로 대전시의 중앙부를 동서로 가로지르는 형태를 확인한 후 대략적인 심야 버스 운영 노선 형태를 고려해 볼 수 있었습니다. 이후 세부적인 입지 분석을 위해서 선정된 행정동별(13개)로 주간 승하차 인원이 많은 정류소를 1개소씩 선정하였습니다.

### 3-2. 정류소 선정



[그림 11]

위의 [그림 10]은 앞서 언급된 행정동별로 정류소를 1개소씩 선정하여 지도에 표시한 것입니다. 충남대 ~ 우송대 삼거리까지의 노선을 고려하여 아래와 같이 3개의 최종 노선을 선정하게 되었습니다.

### 3-3. 노선 선정



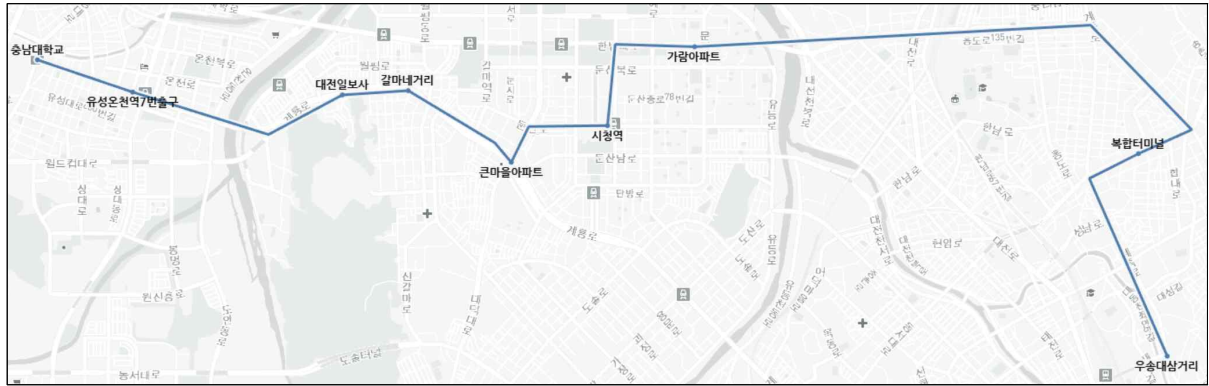
[그림 12]

위의 [그림 12]는 정류소 전체를 순환하는 순환노선입니다.



[그림 13]

위의 [그림 13]은 용문역 방향의 비순환 노선입니다.



[그림 14]

위의 [그림 14]는 시청역 방향의 비순환 노선입니다.

## ○ 결과 해석 및 시사점

‘대중교통 중심도시’를 꿈꾸는 대전시는 작년 기준으로 버스에 대해 시민들의 만족도가 전반적으로 높은 것으로 드러났습니다. 하지만 여전히 시민들이 버스에 대한 불편 사항을 가지는 것을 알 수 있었고 본 팀에서는 추가적인 설문조사를 통해서 시내버스의 ‘이른 막차 시각’ 문제를 최우선 해결과제로 선정하였습니다.

79개 행정동을 기준으로 여러 가지 데이터(사회 인프라, 유동인구, 카드 사용 내역 등)를 정제하여 모델의 독립변수로 지정하였습니다. 버스 승하차 인원을 종속변수로 설정한 후 이상치 제거, 다중공선성 해결, 혼합선택법 등을 활용하여 모델을 간소화하였고 해당 모델을 통해서 심야시간에 대한 버스 승하차 인원을 예측하였습니다. 그 결과를 바탕으로 총 13개 행정동을 추출했고 각 행정동 별로 주간 버스 승하차 인원이 많은 정류소들을 선정하여 최종적으로 3개의 노선을 제안하였습니다.

제한적인 데이터(API 사용 불가, 시계열 데이터 부족 등)를 활용한 수학적 모델을 바탕으로 심야버스 승하차 인원을 예측하고 세부 노선 선정까지 진행했다는 점에서 향후 대전시의 교통 관련 정책의 참고자료로 쓰일 수 있다고 생각합니다. 제한적인 부분을 해결하여 정류소 및 시간대별 데이터를 실시간으로 파악하고 모델링의 변수에 활용한다면 보다 정확하고 효율적인 심야버스 운영이 가능할 것이라고 생각합니다.

마지막으로 통신사와의 업무협약을 통해서 시민들의 수요를 기반으로 정교한 노선을 수립하고 시범운행을 통해 해당 사업 추진의 정당성을 강화할 수 있다고 생각합니다. 시계열에 따른 다양한 데이터들의 활용과 정책이 뒷받침 된다면 진정한 대중교통 중심도시 대전으로 도약할 수 있는 좋은 기회가 될 것이라고 생각합니다.

## ○ 활용방안 및 기대효과

본 분석에서는 대전시내의 심야 버스 노선 운영에 관련된 각종 요인들의 현황을 통해서 버스 노선 선정에 대한 입지분석을 실시하였습니다. 여러 가지 활용방안과 기대효과를 고려해 보았을 때 아래의 활용방안과 기대효과를 확인하였습니다.

우선 본 분석의 활용방안은 아래와 같습니다.

첫째, 실제로 더 다양한 데이터를 추가하여 심야 버스 노선 선정에 대한 입지선정에 활용할 수 있을 것입니다. 본 분석에서는 인구통계, 교통 인프라, 콜택시 등 공개되어 있는 여러 데이터와 유동인구, 매출액 등 제공받은 데이터를 활용했습니다. 만약 시계열, 각종 API 데이터, 버스 이용 로그 데



이터, 정류소 및 시간대별 승하차 인원 데이터들이 본 분석에서 만든 수학적 모델과 결합된다면 더욱 정확하고 현실적인 분석이 가능할 것이라고 판단됩니다.

둘째, 심야 버스 노선 운영 뿐만 아니라 대전시 인구 구성 형태에 따라 다양한 교통 수단을 운영하는 데 본 분석을 참고할 수 있다고 생각합니다. 각 구 및 행정동별로 인구통계학적 데이터와 사회적 인프라 데이터를 취합해 여러 요인들을 고려할 수 있는 모델이라고 생각합니다. 예를 들어 현재 시에서 추진중인 트램 사업의 정류소 선정에 도움이 될 수 있다고 생각합니다. 또한 트램 운영시 발생하는 데이터를 본 분석의 모델에 변수로 추가하여 교통 시스템 운영에 활용을 한다면 대중교통 중심 도시를 꿈꾸는 대전시에게 더할 나위 없이 좋은 기반이 될 것이라고 자신합니다.

셋째, 대전시의 향후 지리정보시스템을 이용한 공간분석 기법 활용을 통해 다양한 정책을 수립한다면 제시한 모델이 좋은 보조자료가 될 것이라고 생각합니다. 본 분석에서는 회귀분석을 활용하여 승하차 인원을 예측하였지만 회귀분석 뿐만 아니라 로지스틱, 군집분석 등 다양한 분석 방법론들로 확장하여 정책 수립시에 기초 자료로써 활용이 가능하다고 판단됩니다.

그리고 본 분석의 기대효과는 다음과 같습니다.

첫째, 대전시의 ‘대중교통 중심도시’ 정책의 시작점이 될 수 있다고 생각합니다. 비단 심야버스 노선 운영 뿐만 아니라 각종 데이터를 활용해 여러 요인들을 분석하여 대전 시민들의 교통 불편을 해소하는 정책에 활용한다면 시민들의 삶의 질을 향상시키는데 기여할 수 있다고 생각합니다.

둘째, 현재 대전시에서 추진 중인 ‘트램’ 운영에 활용할 수 있으며 더불어 이러한 대전시의 정책 실행이 우수한 선례로 남아 다른 지자체의 교통 시스템 운영시 벤치 마킹 모델이 될 수 있다고 생각합니다. ‘트램’ 운영에 따른 여러 요인들을 사전에 분석하여 효율적으로 정책을 실행하고 이에 따라 파생되는 다양한 데이터들이 교통 뿐만 아니라 다양한 정책 운영에 사용된다면 국내 육상교통에 새로운 패러다임을 제시할 수 있을 것이라고 생각합니다.

셋째, 젊고 활기찬 대전의 특성상 야간 경제활동이 활발하게 이루어지는 반면 야간 경제활동에 참여하는 시민들에게 경제활동 시간을 보장은 물론 이동 비용의 부담을 덜어주어 시내 경제활동의 유동성을 보존할 수 있을 것이라고 생각합니다.

대전시는 대한민국의 육상교통의 요지이자 중심도시로서 역동적이고 활기찬 도시입니다. 실제로 2019 대전의 사회지표(e-book)에 따르면 대전 시민들의 시내버스 운영에 대한 만족도는 73%가 만족한다고 응답했으며 대체적으로 시내 교통수단에 대한 만족도가 높은 것으로 나타났습니다. 다만 24시간 경제활동이 일어나는 도시의 특성상 야간 경제활동에 참여하는 시민들 중 일부는 서비스 소비자이자 생산활동자 혹은 학생으로 구성되므로 야간 대중교통의 기여는 그만큼 대전시의 야간 경제활동성에 있어서의 제약이 된다고 생각합니다. 반대로 택시 혹은 대리운전에 의존하는 야간 교통수단의 특성상 비용부담이 젊은 층을 비롯한 소득약자에게는 적절치 못하므로 야간 교통 선택권이 부족한 실정속에서 빅데이터를 활용해 시민들을 위한 야간 버스 운영을 실행한다면 ‘대중교통 중심도시’에 한 걸음 더 가까워 질 것이고 대전시의 위상은 더욱 드높아질 것입니다. 지금까지 본 분석의 보고서를 읽어주셔서 감사합니다.

## ○ 활용데이터 및 참고 문헌 출처 등

### - 기사 및 참고 문헌 출처

- (1) 충청뉴스(정치/행정 > 대전소식) “대전시내버스 막차 시간 늘려줘유”, 2018년 3월 23일
- (2) 대한산업공학회 추계학술대회 논문집 『Analytic Hierarchy Process(계층분석과정)를 이용한 심야버스 순환노선 설계』, 2013년
- (3) 경기연구원 『경기도 심야버스 운행체계 개선방안 연구』, 2015년
- (4) 한국지방정부학회 『서울시 심야버스 정책은 어떻게 성공적으로 도입되었는가? : 정치적 관리의 관점을 중심으로』, 2019년

### - 활용 데이터

#### [대전시 제공]

- 카드매출액
- 유동인구

#### [공공 데이터 포털]

- 2020년\_4월말\_주민등록현황
- 도서관(대전광역시\_동구\_도서관\_20200319, 대전광역시\_유성구\_도서관\_20190906, 대전광역시\_중구\_도서관\_20200304, 대전광역시\_대덕구\_도서관\_20190930, 대전광역시\_서구\_도서관\_20191209)
- 유치원(대전학교정보)
- 초등학교(대전학교정보)
- 중학교(대전학교정보)
- 고등학교(대전학교정보)

#### [교통 데이터 DW 시스템]

- 승하차 인원 및 환승인원(2019.01 ~2019.12)

#### [크롤링]

- 공시지가(국토교통부 표준지공시지가 열람)
- 대학교(네이버)
- 대형상권(네이버)
- 지하철역(네이버)
- 토지면적(나무위키)
- 인구밀도(나무위키)
- 세대수(나무위키)

#### [SKT 빅데이터 허브]

- 대전 콜택시 이용 통화량(2019.01 ~ 2019.06)

#### [대전 광역시 시설 관리공단]

- 타슈 운영정보(2018)