

사회적가치 창출 관점의 장애인 콜택시 운영 최적화를 위한 빅데이터 분석

서창갑 (동명대학교 경영학과 교수, 주저자 armour@tu.ac.kr)
정대현 (부산대학교 경영학과 강사, 교신저자 jdjh@pusan.ac.kr)

… Abstract …

교통약자의 이동수단에 관한 사회적 배려가 절실하다. 따라서 본 연구는 빅데이터 분석을 활용하여 장애인 콜택시 승차대기에 영향을 미치는 요인을 분석함으로써 교통약자들의 사회적가치 창출 관점에서 운영 최적화를 위한 대안이 무엇인지 답문을 제시하고자 하였다. 장애인 콜택시 운영에서 축적된 두리발 관련 빅데이터 7십만여건의 분석을 통하여 현안 문제점을 파악한 후 개선책 제안이 주 목적이었다. 분석툴은 R, Excel, Power BI, QGIS, SPSS를 활용하여 로(raw)데이터의 결측치 제거, 전처리 과정, 빅데이터 분석, 제안에 이르기까지 모든 진행 과정을 서술하였다. 주요 제안으로는 고객취소 데이터 관리의 문제점, 다크(dark)데이터 최소화 방안, 분산 관리중인 데이터의 통합 데이터베이스화의 필요성, 추가분석에 필요한 시스템 업그레이드 내용, 인공지능의 머신러닝을 적용한 콜 추이 분석 도입의 필요성, 중요도만족도분석(IPA)을 통한 고객만족도 제고 방안 제시, 탄력적 근무시간 시뮬레이션 효과 검증 등을 기술하였다.

Key Words : 빅데이터, 사회적가치, 교통약자, 승차대기, ESG

I. 서 론

ESG 경영은 의미상으로 “환경적 요인, 사회적책임, 투명경영”으로 의역되어야 한다(도서출판 블록체인, 2022). 따라서 장애인 콜택시 운영 최적화를 통한 사회적가치 창출은 궁극적으로 ESG의 S(social) 관점에서 공동체의 사회적책임 일환이다. 우리나라 총인구 51,638천명 가운데 등록된 장애인 인구는 2,644,700명으로 총인구 대비 5%에 해당한다(국가통계포털, 2022). 장애인복지법 제23조(편의시설)에 근거한 사회적약자들의 이동 편의성을 지원하기 위한 여러 대책에도 불구하고

하고 실질적으로 이용자들이 느끼는 복지의 한계성이 큰 것으로 파악된다(이병화·양희택, 2017).

따라서 본 연구는 빅데이터 분석을 활용하여 장애인 콜택시 승차대기에 영향을 미치는 요인과 문제점을 분석함으로써 교통약자들의 편의증진을 위한 시사점을 제시하고자 한다.

빅데이터 구성요소 5V(volume, velocity, variety, value, variable) 중 어떠한 가치를 창출할 수 있는지가 중요한 이유는 빅데이터 분석을 통하여 근본적으로 해결하려는 목적이 무엇인지 명확해야 한다는 관점이다(정대현·박광오, 2021). 이러한 목적 부합성 측면에

서 본 연구는 장애인 콜택시 승차대기 시간을 분석하여 교통약자들의 이용편의 증대를 위한 합의점 도출에 기여하고자 한다.

부산광역시는 장애인 특별교통수단 ‘두리발’을 운영함으로써 교통약자의 이동편의를 제공하고 있다. 두리발에 축적된 7십만여건의 빅데이터 분석을 통한 사용자 실태조사, 고객만족도 제고 및 개선을 위한 문제점 진단, 서비스 품질 고도화 실현을 위한 전략적 제언 등이 본 연구의 목적이다. 특히 장애인 콜택시 사용을 위한 사용자들의 접수시간, 기사 배차를 위한 배차시간, 기사와 사용자 만남을 위한 승차시간에 이르기까지 총대기시간의 집중 분석을 통하여 교통 사회적 약자들의 총대기시간 최적화 방안이 무엇인지를 파악함으로써 두리발의 운영 효율성 제고를 통한 사회적가치 창출 방안을 모색하고자 한다.

II. 연구 개요

두리발은 2명이 함께 한다는 ‘둘이’와 이동수단 ‘발’의 합성어로서 교통약자 특별교통수단에 근거 2006년 10대로 출범하였다. 2022년 8월 기준 총 187대의 차량을 24시간 운영형태로 슬로프 탑승방식 183대와 리프트 방식 4대로 구성되어 있다. 휠체어 이용고객은 3명, 비휠체어 이용고객은 2명 동승을 허용하고 애완견 및 애완묘 등 케이지나 이송용 가방을 지참하는 경우 탑승이 가능하다. 그리고 두리발 이용 제한 지침은 3회 이상 콜센

터에 사전연락 없이 배차를 취소하는 경우 1개월 이용 제한이 가능토록 규정하고 있다.

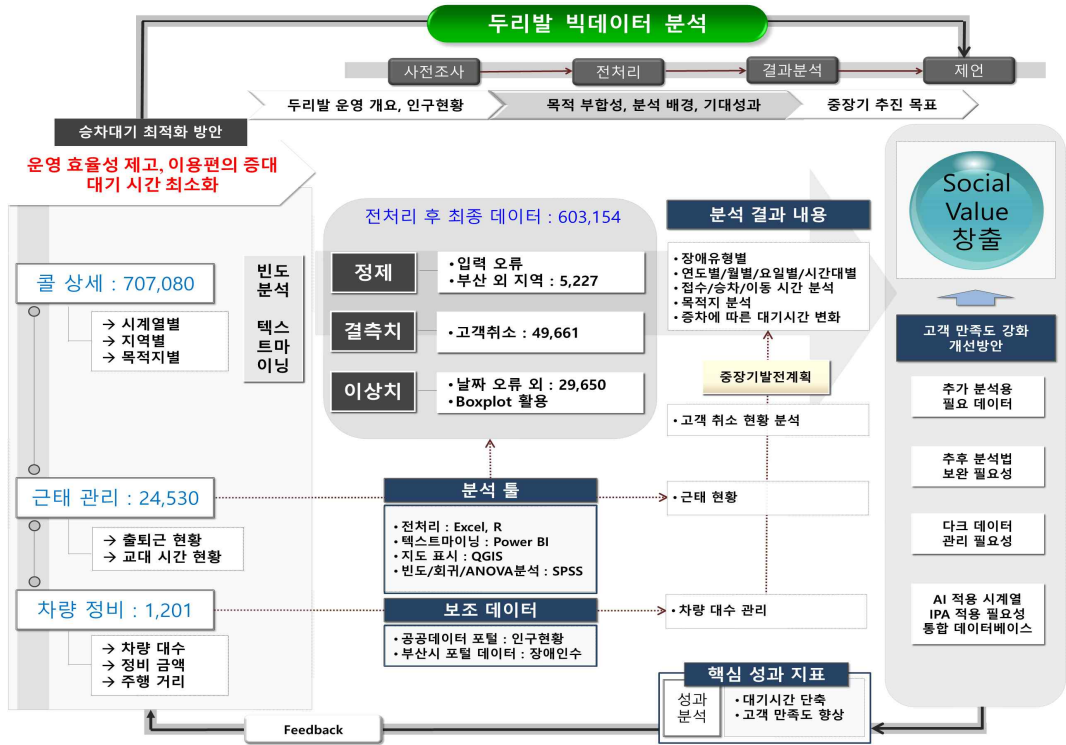
전체 연구 프로세스를 요약하면 <그림 1>과 같이 도식화할 수 있다. 두리발에 축적된 2년 7개월치의 빅데이터 분석을 통하여 교통약자의 사회적가치를 창출하기 위한 운영 효율성 제고, 이용편의 증대, 대기시간 최소화 등의 목적을 실현하고자 하였다. 본 연구에서는 콜상세 데이터 7십만여건의 시계열별, 지역별, 목적지별 빅데이터 분석을 위하여 분석툴은 Excel, R, Power BI, QGIS, SPSS를 활용하여 분석 목적에 부합될 수 있도록 전처리 및 시각화를 진행하였다. 그리고 우리나라 전체 인구수와 장애인 현황 파악을 위하여 공공데이터 포털 및 부산시 홈페이지에 공지된 데이터 등을 보조적으로 활용하였다.

국가통계포털 2021년 12월 기준 우리나라 총인구는 감소추세이지만 장애인 등록 인구 및 65세이상 노령인구 비율은 점차 증가추세를 보이고 있다<그림 2>. 따라서 교통약자들의 편익을 제공하기 위한 두리발의 정책들이 사회적가치를 달성할 수 있도록 하는 빅데이터 분석이 필요한 시점이다.

III. 빅데이터 개요

1. 빅데이터 전처리

본 연구의 핵심인 콜상세 총대기시간 분석에는 603,154건(전체 원시데이터의



<그림 1> 연구 프로세스

85.3%)의 데이터를 사용하였다. 2020년 1월~2022년 7월까지의 전체 데이터 707,080건 중에서 전처리 과정을 통하여

주요 결측치로 판단한 내용은 고객취소 데이터 49,661건(7.02%), 승차 대기시간 부재 및 날짜 오류 데이터(하차시간이 승차시간보다 앞서는 경우 등) 29,650건 (4.19%)을 분석에서 배제시키고 부산 이외의 5,227건을 제외한 최종 분석용 데이터 현황은 <표 1>과 같다.



<그림 2> 총인구 대비 장애인 증가추세

<표 1> 최종 분석용 데이터 현황

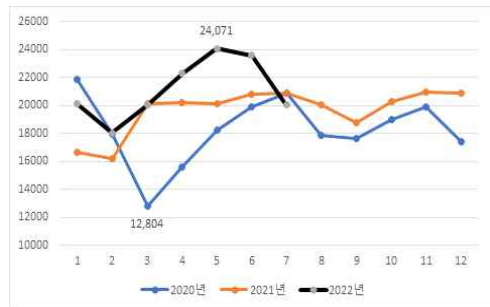
구분	빈도	비율	
소스 데이터	707,080		
결측치	고객취소	49,661	7.02
이상치	날짜오류 외 배차, 승차 대기시간	29,650	4.19
	부산 외	경남, 충북	19,388
부산 외	경남, 충북	5,227	0.74
최종 분석용 데이터	603,154	85.30	

2. 분석 현황

1) 두리발 이용자 분석

두리발 이용자 현황은 장애 33 유형별에서 지체1급(21%), 뇌병변 1급(16%), 65세이상 0급(16%)의 사용 비중이 높은 것으로 조사되었다. 부산광역시의 지역별 장애인수는 부산진구(18,474명, 10.52%), 사하구(17,859명, 10.17%), 해운대구(17,611명, 10.03%) 순이며 중구(2,591명, 1.48%)가 최저 분포로 나타났다(부산시 홈페이지, 2022년6월 기준). 이는 지역별 콜 건수와 비교해 보아도 부산진구-사상구-해운대구-북구 순으로 파악되어 장애인수 및 65세 노령 인구 거주지와 비례함을 알 수 있다.

콜 건수에 대한 빈도분석은 <그림 3>과 같이 COVID19의 영향으로 2020년 3월(12,804건)에 급격하게 줄었으나 이후 전반적으로 꾸준히 증가하는 추세이다. 특히 2022년 4월 거리두기 완전해제가 이루어진 이후 증가추세가 두드러진다.



<그림 3> 연도별 콜건수 추이

2) 총 대기시간 분석

배차 대기시간은 사용자가 콜센터에

접수 후 배차 가능한 기사를 수배하기까지의 소요시간을 뜻한다. 분석 결과 최장 시간대는 06시대(33분19초)이며 최단 시간대는 02시대(3분25초)이다.

승차 대기시간이란 배차 완료 후 기사가 사용자를 승차시키기 전까지의 소요시간을 뜻한다. 최장은 03시대(26분58초) 최단은 14시대(17분17초)이다. 따라서 배차+승차 총대기시간은 오전 6시대(51분31초)가 최장이고 오전 1시대(26분07초)가 최단으로 파악되었다.

분석 결과 근무 교대 시간대인 8H, 12H, 15H, 17H, 21H에서 모두 평균을 상회하고 있다. 9H~12H 시간대의 콜 비율이 40.38%인 반면 현행 근무자 편성내용은 22.62% 배치 중이다. 이를 탄력적으로 콜 건수 비율로 조정해 보면 콜 집중 시간대에서 현행 83명을 146명으로

<표 2> 총 대기시간 분석

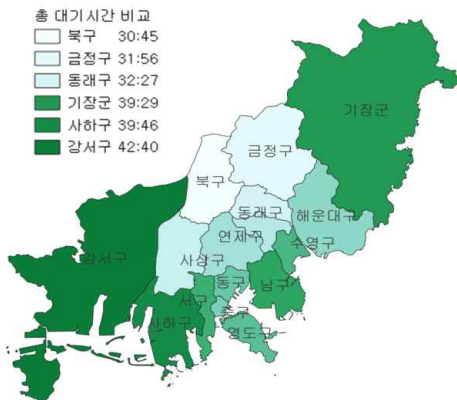
시간대	콜건수		대기시간 (분:초)			근무인원	
	건수	비율	배차	승차	배차+승차	현재	조정(안)
0 H	1,232	0.20	5:53	22:36	28:29		
1 H	787	0.13	3:40	22:26	26:07		
2 H	548	0.09	3:25	22:59	26:25		
3 H	1,280	0.21	4:17	26:58	31:16	2	2
4 H	2,166	0.36	5:30	25:53	31:24		
5 H	2,874	0.48	31:23	19:49	51:13		
6 H	14,358	2.38	33:19	18:11	51:31		
7 H	35,055	5.81	14:35	20:31	35:07	44	50
8 H	50,054	8.30	23:21	19:37	42:59		
9 H	54,291	9.00	15:05	18:36	33:43		
10 H	58,064	9.63	10:44	17:27	28:13	83	146
11 H	72,021	11.94	15:48	17:43	33:32		
12 H	58,180	9.65	23:48	17:49	41:38		
13 H	46,080	7.64	13:28	18:03	31:32		
14 H	61,443	10.19	10:01	17:17	27:19	121	97
15 H	56,127	9.31	16:44	19:04	35:49		
16 H	33,504	5.55	12:43	20:17	33:01	77	32
17 H	18,229	3.02	13:08	22:22	35:31		
18 H	11,280	1.87	14:16	21:43	36:00		
19 H	10,332	1.71	10:04	19:42	29:47		
20 H	8,445	1.40	5:19	18:09	23:29	38	38
21 H	3,284	0.54	24:13	23:38	47:52		
22 H	1,966	0.33	17:30	24:35	42:06		
23 H	1,554	0.26	9:37	23:52	33:30	2	2
계/평균	603,154	100%	15:47	18:46	34:34	367	367

배치하고 121명을 97명으로 줄이는 검토안을 고려해 볼 필요성이 있다.

<표 3> 차량대수별 대기시간 비교

차량대수	운영기간		콜건수		운행/대당	대기시간		
	년/월	개월	계	월평균		배차	승차	계
160	2020/1월~	7	127,287	18,184	114	13:51	18:35	32:26
181	2020/8월~	16	306,863	19,179	106	16:05	18:50	34:55
187	2021/12월~	8	169,004	21,126	113	15:46	18:44	34:30

증차에 따른 대기시간의 변화 정도를 분석한 결과 <표 3>과 같이 160대, 181대, 187대 기준 배차 및 승차 대기시간과 차량대수에 따른 ANOVA를 이용한 집단비교는 P값 0.000에서 통계적으로 유의한 평균 차이를 보였다. 그러나 대기시간 단축이라는 근본적인 문제점 해결에는 큰 영향을 미치지 못하는 것으로 판단된다. 코로나 창궐로 인한 이동제한 등의 영향이 존재하며 콜 건수에 비례하여 배차 및 승차 대기시간이 연동한다고 볼 수 있기 때문에 추후 추가적인 분석이 필요할 것이다.



<그림 4> 총 대기시간 지역별 현황

<그림 4>의 총대기시간 지역별 분포에서 부산 외곽 지역인 강서구, 사하구,

기장군 순으로 평균 총 대기시간(34분34초)을 상회하고 있어 기사 대기장소와 다소 차이가 발생하는 것으로 예측된다.

<표 4> 요일별/대기 및 이동 시간

요일	접수	승차	접수+승차	이동
월	18:09	18:42	36:52	23:47
화	17:43	18:46	36:30	24:10
수	16:54	18:41	35:36	24:03
목	16:13	18:41	34:55	24:00
금	17:31	18:54	36:26	24:25
토	8:19	19:18	27:37	24:43
일	5:46	18:23	24:10	23:48
평균	15:47	18:46	34:34	24:07

요일별 콜 빈도가 수-화-금-목-월-토-일 순을 감안할 때 평일 총대기시간의 상관관계는 크지 않았으며 이동 평균 시간은 24분07초로 분석되었다<표 4>.

3) 목적지 및 고객취소 사유 분석

두리발 사용자의 목적지 분석은 언급횟수 100회 이하 단어는 제외시키고 57개 주제를 선별하여 텍스트마이닝 기법을 활용하였다. 분석 결과 주거시설, 의료기관, 생활, 복지, 유통, 종교, 관공서, 교육시설, 금융 순으로 조사되어 주로 주거시설(43.8%) 및 의료기관(35.6%) 방문 순으로 조사되었다.

또한 고객취소 사유는 다른 교통수단(59.9%), 탑승지연(9.6%), 통화불능, 고객변심 등의 순으로 파악되었다.

IV. 결론 및 제언

상기 연구 결과를 바탕으로 ESG 사회

적 책임 관점에서 사회적가치를 창출하기 위한 시사점은 다음과 같이 요약할 수 있다. 교통약자의 사회적가치는 대기시간 최소화를 위한 노력과 근본적인 문제점 해결로부터 출발 가능하다. 지속적인 장애인 및 노령인구의 증가 추세에 발맞춰 대기시간 최소화를 위한 정책에 사회구성원들의 관심이 필요할 것이다.

첫째, 총 대기시간은 34분34초가 소요되었다. 부산 외곽지역인 강서구, 사하구, 기장군이 상대적으로 평균 총대기시간을 상회한 것으로 보아 기사 대기장소와의 매칭이 잘 이루어지지 못한 것으로 판단되어 진다. 또한 근무 교대시간대 8H, 12H, 15H, 17H, 21H의 총대기시간도 평균을 상회하였다. 따라서 부산 외곽, 근무 교대시간대, 콜 집중 시간대와 더불어 장애인 다수 거주지인 부산진구, 사상구, 해운대구, 북구 순으로 콜 건수가 비례한점을 감안하여 근무자를 탄력적으로 재배치할 수 있는 대안을 장기적으로 검토해야 할 것이다.

둘째, 고객취소 데이터 사유별 내용에서 다른교통수단 및 승차지연이 다수였다. 이는 총대기시간의 지연에 따른 불만요인과도 연결된다. 또한 고객취소 사유별 분류를 단순 변심, 지연 불만, 대체수단 이용 등의 유효성 검사의 데이터베이스 구축으로 고객만족도 향상을 위한 분석용으로 활용할 필요성이 있다.

셋째, 다크(dark) 데이터 최소화를 위한 보완책 마련이 절실하다. 분석용 데이터로 활용할 수 없는 다크 데이터의 효율적 관리방안을 마련해야 할 것이다.

무엇보다 현행 시스템의 활용을 전제할 때 제때 출퇴근 근태관리가 이루어질 수 있도록 하는 교육이 뒷받침 되어야 할 것이다.

넷째, 분산 관리되고 있는 데이터의 통합 데이터베이스화의 검토를 제안한다. 본 연구에서 다룬 데이터인 콜상세, 근태관리, 차량정비 데이터는 분산 관리 중이다. 특히 근태관리 및 차량정비 현황은 단순한 엑셀 데이터로 관리하는 실정이라 데이터의 참조무결성에 다수의 오류가 파악되었다. 따라서 현재 콜상세 시스템에 기사 근태관리 및 차량정비 데이터를 통합하여 하나의 시스템으로 구축되면 빅데이터 분석 활용도가 높아질 것이다.

다섯째, 추후 분석법 보완의 필요성이 있다. 단순 추세선 분석을 넘어 콜건수 예측에 따른 적절한 사전 대응 전략 수립에 유익한 정보제공이 뒷받침될 수 있도록 Azure ML 활용, ARIMA 모델 및 ETS 지수평활법 등의 적용을 고려해 볼 수 있겠다. 또한 IPA(중요도만족도분석) 분석을 통하여 긴급하게 조치가 필요한 고객의 니즈를 파악함으로써 고객만족도 향상을 꾀할 수 있을 것이다.

끝으로 후속 연구의 확장은 전국 장애인 콜택시 현황의 비교 분석을 통하여 새로운 사회적가치 창출이 가능하도록 정부기관이 주관하는 빅데이터 분석의 필요성을 제안한다. 이로써 좀 더 현실성 있는 교통약자의 근본적 문제점 해결에 도달할 수 있기를 희망해 본다.

(참고문헌 및 기타 세부 분석결과 생략)