[팀 프로젝트 결과보고서]



|  |  |
| --- | --- |
| **■과 목 명:** | 데이터분석언어 |
| **■담당교수:** | 안용학 교수님 |
| **■제 출 일:** | 2023.12.16 |
| **■학 과:** | 데이터사이언스융합학과 |
| **■ - 2조 -** | |
| **■성 명:** | 민병윤(조장), 이상민, 최다인 |

**주제: 수원시 내 광역버스 승차 행태 변화 분석 (입석금지 Issue를 바탕으로 진행)**

**1. 내용: 팀프로젝트 결과보고서**

**(1) 분석 개요 및 설명**

최근 시행된 광역버스의 입석금지 조치(실질적 조치 시행: 2022.11.21.)로 인해 승객이 많이 몰리는 시간대의 일부 정류장 이후부터는 만차로 탑승 자체가 불가능하여, 이전 정류장으로 거슬러 올라가야 탑승이 가능한 불편한 상황이 생겨나고 있습니다. 이러한 문제를 해결하기 위해 경기도에서는 예약 버스 좌석시스템인 miri 앱을 도입하였으나 아직 일부 시간대 버스 노선에만 적용되고 있고 버스당 40명의 탑승객 제약으로 인원수용에 한계가 있기 때문에 아직도 서울로 출퇴근하는 많은 도민들이 불편을 겪고 있습니다.

이에 경기도 중에서 인원이 가장 많은 수원시의 광역버스 이용 실태를 분석하여 승차 승객이 가장 많은 버스 정류장을 기점으로 이용객 수 분석을 통해 예약 시스템으로 도입해야 할 노선을 발굴하여 예약서비스를 확대하고자 합니다.

**(2) 분석대상 데이터**

**- 데이터 수집 대상, 수집 범위, 데이터 명세**

**(2-1) 수원시 광역버스 노선별 정류장별 시간별 승차인원 데이터**

출처: 교통카드 빅데이터 통합정보 시스템

수집 범위: 수원시 21개 노선 ‘22년 9월, ‘23년 9월, 7770번 노선의 22년 3월,6월,9월,12월, 23년 3월, 6월, 9월 시간대별 승차인원

데이터 명세

1) 노선: 광역버스 번호

2) 기종점: 버스 종점(차고지 정보)

3) 일자: 승차인원 집계 일자

4) 정류장순번: 버스 운행 노선 순번 (첫 번째 정류장부터 1,2,3 …. 순서)

5) 정류장: 버스 노선에 해당하는 정류장

6) 시간: 해당 시간대 정류장에서 탑승한 승차 인원수의 합계

**데이터 수집 특이 사항**

시간별 승차 인원 Data에 대해 장기간의 Data를 확보하지 못한 이슈가 발생했습니다.

조회 조건이 최대 5일로 1회 추출 시 평균 5분 이상이 소요되어 데이터 추출에 긴 시간이 소요됨

데이터를 요청하기 위해 메일을 보냈으나 별도 공문이 있어야 제공이 가능하다는 답변을 받음

1년의 Data를 활용하려 했으나 시간적 제약, 데이터 확보 제약으로 수집 기간을 축소하여 분석을 진행했습니다.

**(2-2) 전국 정류장 위치 정보 데이터**

-출처: 국토교통부

수집 범위: 전국 2022년 정류장 기준 정류장이 수원시, 서울시이거나 관리도시가 경기, 서울인 버스 정류장

데이터 명세

1) 정류장번호: 정류장의 고유번호

2) 정류장명: 버스 정류장명

3~4) 위도, 경도: 정류장에 해당하는 위도, 경도 정보

5) 정보수집일시: 정보 수집 일자

6) 관리도시명: 해당 정류장이 소속된 도시명

**(2-3) 기상관측 데이터**

출처: 기상청 기상장료 개방포털

수집 범위: 수원시의 ‘22년 3,6,9,12월, ‘23년 3,6,9월 일별 기온, 강수량, 풍속, 습도 데이터

데이터 명세

1) 지점명: 관측 지점(지역명)

2) 일시: 관측 일자 및 시간

3) 기온/강수량/풍속/습도: 기상 관측값

**(3) 개발환경**

- 사용 언어 및 환경: python (google colab 사용)

- 사용 라이브러리: pandas, folium, numpy, matplotlib, seaborn

pandas: csv파일이나 엑셀파일의 분석대상 데이터 소스를 읽고, 데이터를 정렬 및 필터링하여 필요한 정보 추출에 사용

numpy: 승객 수 데이터의 평균, 표준편차, 중앙값 등 통계적 분석이나 노선간 차이를 파악하기 위한 연산을 위해 사용

Folium : 광역버스 노선의 시간대별 이용 승객수를 지도상에 시각적으로 나타내기 위해 사용

Matplotlib & Seaborn: 데이터를 시각화하여 패턴이나 추세를 쉽게 파악하기 위해 사용

**(4) 분석 방안**

**- 대상 데이터에 대한 분석 방법 및 내용**

전체적인 분석 방안은 분석에 대한 시나리오를 설정하고 데이터 전처리 후 시각화, 분석결과 도출 연구 결과 도출 순으로 진행했습니다.

대중교통 이용 수 데이터를 통해 혼잡 정류장, 혼잡 시간, 혼잡 요일 분석, 버스 정류장을 지도에 시각화하여 밀집된 구역을 한눈에 파악해 보고, 수치 데이터들은 Heat map, 그래프를 통한 시각화를 진행하여 전체적인 이용객 수 분석을 진행했습니다.

**데이터 전처리**

①수원시 구별 승차분석을 위해 정류장의 위도/경도에 해당하는 주소를 불러와 시와 구로 분할

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

②평일과 주말의 승차분석을 위해 버스 운영 시간대인 4~24시의 total 승차인원과 요일에 따라 주말/평일을 구분하는 변수를 생성하여 분석

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명

**데이터 분석 결과**

**①수원시 구별 승차 행태 분석**

평일, 주말 모두 장안구의 탑승객이 가장 않았으며, 그 다음 순위는 평일과 주말이 다른 패턴을 보였습니다. 평일엔 영통구가, 주말은 팔달구가 더 많은 것으로 보이는데 그 이유는 평일 영통구에는 출퇴근 인원, 학생들이 많이 탑승하고 주말에 팔달구는 나들이객이 많아 평일/주말 순위의 변동이 있었습니다.

■ 수원시 구별 승차 행태 분석

도표, 텍스트, 스크린샷, 그래프이(가) 표시된 사진

자동 생성된 설명**② 날씨에 따른 승차 행태 분석**

날씨 유형 강수(비), 기온(춥고 더운 날씨), 습도(꿉꿉한 날씨), 풍속(태풍)별로 승차인원 분석을 진행했습니다. 평일에는 날씨에 따른 승차인원 변화가 크지 않았으며, 주말에는 날씨에 영향을 받는 것으로 분석되었습니다. 이는 평일 출퇴근은 무조건 해야 하는 행위이며, 주말 나들이는 취소가 가능하여 변동성이 있는 것으로 보입니다.

**③ 평일과 주말 승차 행태 분석**

평일 출퇴근시간대, 주말은 느지막한 오후에 승차인원이 집중되었으며, 버스 노선별로 유사한 패턴을 보이는 것으로 확인했습니다. 평일은 화요일, 수요일, 주말은 토요일에 승차인원이 집중되는 것을 알 수 있었습니다.

상세 결과는 다음 장에 설명되어 있습니다.

**(5) 분석 결과**

**- 제안서에서의 예상 분석 결과와 실제 분석 결과에 대한 비교 설명**

제안서에서의 예상 분석 결과는 예측으로 표현했으며 각 유형별 예측과 분석결과는 다음과 같습니다.

|  |  |  |  |
| --- | --- | --- | --- |
| No | 1 | 분석 내용 | 강수여부 별 승차인원 |
| 예측 | 비가오는 날에는 버스 이용 승차인원이 증가할 것이다 | | |
| 분석결과 | 비가 오는 날(강수량 = 0)과 비가오지 않는 날(강수량 > 0)의 시간별 탑승인원 분석  평일에는 비가 오는 날이 비가 오지 않는 날의 탑승인원 차이가 거의 없다.  주말에는 비가 오는 날이 비가 오지 않는 날보다 탑승인원이 꽤 많이 적다. | | |

■ 강수여부 별 승차인원

도표, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 2 | 분석 내용 | 기온 별 승차인원 |
| 예측 | 평균 기온이 낮거나 높은 날에 대중교통 이용자가 몰릴 것이다 | | |
| 분석결과 | 기온이 높은 날(24℃ 초과)과 기온이 낮은 날(5℃ 미만)의 시간별 탑승인원 분석  평일과 주말 모두 기온이 높은 날이 기온이 낮은 날보다 탑승인원이 다소 많다.  특히, 날씨가 추워질수록 출근시간대 자차 이용이 많다 | | |

■ 기온 별 승차인원

도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 3 | 분석 내용 | 습도 별 승차인원 |
| 예측 | 습도가 높은 날에 대중교통 이용자가 줄어들 것이다. | | |
| 분석결과 | 습도가 높은 날(80% 초과)과 습도가 낮은 날(45% 미만)의 시간별 탑승인원 분석  평일에는 습도가 높은 날과 습도가 낮은 날의 탑승인원 차이가 거의 없다.  주말에는 습도가 높은 날이 습도가 낮은 날보다 탑승인원이 꽤 많이 적다.  강수여부 별 승차인원 추세와 상당히 유사하다 | | |

■ 습도 별 승차인원

도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 4 | 분석 내용 | 풍속 별 승차인원 |
| 예측 | 평균풍속이 높은 날(강풍)에 대중교통 이용자가 몰릴 것이다 | | |
| 분석결과 | 풍속이 높은 날(3m/s 이상)과 풍속이 낮은 날(1m/s 이하)의 시간별 탑승인원 분석  평일은 풍속이 높은 날이 풍속이 낮은 날보다 출퇴근시간대 탑승인원이 많다.  주말은 풍속 요인이 탑승인원과 큰 연관이 없다. | | |

■ 풍속 별 승차인원

도표, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

**날씨에 따른 분석 종합 결과**

평일 출퇴근 시간대의 따뜻한 기온 또는 강풍이 불 때 다소 몰리지만, 날씨에 따른 승차인원 영향을 거의 받지 않는 것 같습니다. 열심히 일하는 직장인들의 영향이 큰 것으로 보입니다.

주말은, 비가 오면 버스가 한산한 것으로 보이며, 날씨에 따라 약속이 취소되어 한산한 것으로 예상됩니다.

|  |  |  |  |
| --- | --- | --- | --- |
| No | 5 | 분석 내용 | 평일과 주말 승차인원 분석(요일별) |
| 예측 | 월요일, 금요일에 버스 이용객 수가 몰릴 것으로 예상됨 | | |
| 분석결과 | 평일은 화>수>월>목>금 순서, 주말은 토>일 순서로 승차인원이 많았음  평일: 직장인 승객이 집중되어 월, 금 휴가자가 다소 있으며, 화요일에 승차인원이 가장 많음  주말: 여행, 나들이, 개인 일정을 토요일에 집중적으로 하여 승차인원이 일요일 보다 토요일에 집중됨 | | |

■ 평일과 주말 승차 행태 분석

텍스트, 그래프, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 6 | 분석 내용 | 평일과 주말 승차인원 분석(시간대별) |
| 예측 | 출퇴근 시간에 버스 이용객 수가 몰릴 것으로 예상됨  심야시간의 이용객이 많다면, 회식 후 탑승할 수 있는 심야버스 확대 검토 | | |
| 분석결과 | 평일: 출근시간인 6~8시, 퇴근시간인 17~19시에 승차인원이 집중됨  주말: 느지막이 일정을 시작하는 12시, 16시, 일정을 끝내고 집에 돌아오는 21시에 승차인원이 집중됨  심야시간 이용객은 특별히 몰리지 않아 별도 심야버스 확대는 무의미 할 것으로 판단됨  평일, 주말의 패턴은 특정 버스에만 집중되지 않고 시간대별로 전체적으로 유사한 패턴을 보임 | | |

■ 평일 및 주말 시간대별 승차인원 분석

텍스트, 도표, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 7 | 분석 내용 | 평일과 주말 승차인원 top 노선 |
| 예측 | 주거밀집, 역근처, 회사 근처의 정류강에 버스 이용객 수가 몰릴 것으로 예상됨 | | |
| 분석결과 | 역 근처인 수원역에도 승차인원이 많지만 특히 주거밀집 지역에 승차인원이 몰리는 것으로 확인  주거밀집지역인 수원종합운동장, 수원KT위즈파크, 한일타운의 승차인원이 특히 많음  출퇴근 시간에는 주거 밀집 지역인 선일초교, 남수원 중학교, 경기아트센터의 탑승객이 밀집됨  대학생들의 수업이 끝나는 시간대에는 대학교 정류장인 아주대, 경기대에 승차인원이 밀집됨 | | |
| 예측 | 이용량이 많은 시간대에 배차 간격이 큰 버스가 있다면, 배차 간격 축소를 통한 이용객 수 분산이 필요할 것이며, 반대로 이용량이 적은 정류장에 대해서는 배차 간격을 크게 하여 차량 운영을 효율적으로 할 수 있을 것이라고 생각됨.  버스 노선 중 차고지에서 몇 정거장 지나지 않아 버스가 만석이 되는 경우 승객들의 이용 편의성을 위해 차고지부터 출발하는 게 아닌 가장 혼잡한 노선에서부터 출발하는 버스 도입 검토 | | |
| 분석결과 | 이용량이 가장 많은 시간대(출근)의 버스 배차를 확인한 결과 이미 배차시간 외의 추가 버스를 운영중인 것으로 확인  7770, 8800번 등의 대부분 광역버스의 경우 승차인원이 특히 많은 정류장에서는 항상 탑승인원 만석이 되기 때문에 해당 정류장 위주로 추가 버스를 도입하여 운행한다면 사용자 편의성이 증가할 것으로 판단됨 | | |

■ 평일 및 주말 탑승객이 가장 많은 상위권 노선 분석

도표, 그래프, 텍스트, 라인이(가) 표시된 사진

자동 생성된 설명

■ 7770번 평일

텍스트, 스크린샷, 번호, 라인이(가) 표시된 사진

자동 생성된 설명

■ 8800번 평일

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

**평일/주말 분석 종합 결과**

평일/주말 탑승 인원이 몰리는 시간대 확인 및 차량운영 효율화 제안 및 혼잡 노선 선별하여 증차 제안을 위해 분석했습니다. 그 결과, 이미 혼잡 시간대에 버스 추가 배차 중이였으며, 정류장 별 세분화하여 추가 배차한다면 편의성은 증대할 것으로 예상됩니다.

**- 비교에 따른 최종 결과 및 기대효과 등**

**시간흐름에 따른 승차인원 변화 파악을 위한 시계열 분석 진행**

출퇴근 시간대 주로 이용량이 많은 광역버스는 공휴일마다 승차 인원이 감소하는 특성이 있었으며, 코로나의 영향으로 22년 3월 감소했던 승차 인원이 6월에 회복된 것으로 보입니다.

시간 흐름 및 발생한 주요 이슈들에 따른 승차인원 변화 그래프는 하기와 같습니다.

■ 시간흐름에 따른 승차인원 변화

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

|  |  |  |  |
| --- | --- | --- | --- |
| No | 8 | 분석 내용 | 승차시간별 승차인원에 대한 시계열 분석 |
| 예측 | 코로나가 해제된 2023년에 2022년도보다 승차인원이 증가할 것임 | | |
| 분석결과 | 출근시간대인 6~8시, 퇴근시간대인 17~19시 승차인원이 몰려 있음  시간대별 승차 형태는 코로나가 있었던 22년도 보다 23년 크게 증가한 것을 알 수 있음 | | |

■ 승차시간별 승차인원 비교

도표, 텍스트, 그래프, 라인이(가) 표시된 사진

자동 생성된 설명

**입석금지 시행 전후의 승차인원 변화 분석(입석금지 조치의 영향 유무 분석)**

다음과 같은 조건을 반영하여 입석금지 전후의 승차인원 변화를 분석했습니다.

1) 22년도 9월과 23년도 9월 출근 시간대 정류장별 승차인원을 비교하여 입석금지 조치 영향 파악

2) 출근 시간대(6시~8시) 수원에서 서울로 올라가는 방향의 노선만 추출하여 저장

3) 9월에 있는 추석 연휴의 영향을 제거하기 위해 추석 연휴 날짜는 제거하고 분석 시행

**■ 분석결과 상대적으로 앞에 위치한 정류장의 승차인원이 증가한 것을 알 수 있었습니다.**

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

대표적으로 Gap이 크게 발생한 3000번에 대해서 6~8시 승차인원 분석한 결과는 하기와 같습니다.

■ 입석금지 조치 영향 유무

텍스트, 스크린샷, 그래프, 폰트이(가) 표시된 사진

자동 생성된 설명

**전체적인 승차인원 분석결과 입석금지가 시행되어 많은 사람들이 기존에 타고 있던 정류장에서는 만석이 되어 탑승하지 못한 것으로 보입니다. 버스 탑승을 위해 기존 정류장보다 앞에 있는 정류장으로 이동하여 탑승하는 것으로 보입니다. 그리고 이전 정류장으로 이동하기에는 너무 거리가 먼 정류장에서 탑승하던 사람들은 다른 교통수단으로 대체하여 이동한 것으로 보입니다.**

텍스트, 라인, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

**[결론]**

**예약시스템 도입 노선 발굴에 대한 분석 방법**

1) 출근 시간대 승차인원 파악

2) 2023년 승차인원과 2022년 승차인원 차이를 계산

3) 예약 도입 노선 도출: 가장 gap이 낮은 정류장과 노선예약 시스템 도입이 시급한 노선 도출

**Gap의 값이 가장 작은 정류장 우선순위로 예약시스템 도입**

**2022년에 입석이 허용된 시기에는, 입석해서라도 버스에 승차하여 승차인원 수요가 많은 정류장이었으나, 입석금지로 인해 승차인원이 급감한 정류장이 많이 발생**

■ **입석금지로 인해 22년도 승차 경향에 비해 23년도에 급격하게 변화한 노선**

텍스트, 스크린샷, 메뉴, 번호이(가) 표시된 사진

자동 생성된 설명

■ **예약 시스템 도입이 시급한 노선과 정류장**

|  |  |  |
| --- | --- | --- |
| **No** | **버스번호** | **노선명** |
| 1 | 7770번 | 수원종합운동장, 수원시 체육회관, 아너스빌위즈  수원KT위즈파크, 경기도 청소년, 활동진흥상담복지센터 |
| 2 | 3000번 | 한일타운, 경기일보. 홈플러스, 수원종합운동장. 수원시체육회관, 아너스빌위즈, 수원KT위즈파크, 경기도청소년활동진흥상담복지센터. 수일중학교 |
| 3 | 5100번 | 황골벽산아파트, 청명역 3번출구 |

**(6) 팀인원 역할분담**

Main task는 담당자별로 진행했으며, 자료수집, 문서작성, 발표 등 대부분의 Task는 팀원 모두 함께 진행했습니다.

|  |  |
| --- | --- |
| **담당자** | **Task** |
| 이상민 | ‘22년, ‘23년 주말 데이터 수집 |
| 시계열 데이터 수집 |
| 주말 및 평일 승차 형태 분석 |
| 노선에 따른 승차 변화 분석 |
| 민병윤 | 시계열 데이터 수집 |
| 날씨 데이터 결합 |
| 날씨(강수, 기온, 습도 등)에 따른 승차인원 변화 분석 |
| 최다인 | ‘22년, ‘23년 평일 데이터 수집 |
| 승차 데이터 결합 |
| 구별 승차 형태 분석 |
| 시계열 및 입석금지 영향분석 |

**(7) 팀원 개별 의견**

**(7-1) 이상민**

- 평소 출퇴근하며 이용하는 대중교통에 대해 실제로 데이터를 분석하여 결과를 얻어 보니 현실성 있고 흥미로운 프로젝트가 된 것 같습니다. 데이터 분석에 있어 데이터 수집과 전처리 과정이 분석 전체의 80%이상의 공수를 차지할 만큼 중요한 요소라는 것을 느꼈습니다. 분석을 할 때 팀원들과 논의를 해보며 다양한 인사이트를 얻을 수 있었습니다. 데이터 기간을 길게 하고 다양한 라이브러리, 모델링을 접목시켜 보면 더 좋은 프로젝트가 될 것이라고 생각합니다.

**(7-2) 민병윤**

- 재미있는 주제를 선정하여 평소 불편하다고 생각한 막연한 느낌을 객관적인 데이터를 통해 분석해 볼 수 있었습니다. 프로젝트를 하며 강의 중에 배운 다양한 파이썬 라이브러리의 활용하며 복습할 수 있었습니다. 또한, 데이터 수집의 어려움, 결측치 등 이론적인 내용을 숙달하더라도 발생하는 현실적인 어려움과 같은 데이터라도 조건과 시각화에 따라 분석이 달라질 수 있는 것도 경험했습니다. 이 프로젝트를 통해 좋은 동기들을 알게되었고, 앞으로 수강할 데이터사이언스 과목들에도 꼭 필요한 발판이 된 것 같다고 생각합니다.

**(7-3) 최다인**

- 실제로 입석금지로 인해 3정류장을 거슬러 올라가서 버스를 타는 불편성을 겪고 있어, 이 주제로 분석을 진행 해보고 싶었습니다. 수원에서의 거주기간이 길어 각 구마다의 광역버스의 특성과 정류장 특성을 잘 알고 있었기 때문에 프로젝트 진행이 더 재미있었습니다. 특히 노선 중 1, 2순위로 꼽힌 “수원종합운동장”과 “수원KT위즈파크”의 정류장은 대표적으로 사람들이 많이 이용하는 광역버스 노선이나, 시종착점 중간에 있는 노선이라 만차가 되는 경우가 많은 노선이었기 때문에 입석금지의 영향을 많이 끼칠 것이라고 예상하였는데 이게 실제로 데이터 상으로 분석되니 신기하였습니다. 더 나아가 현재는 수원에서의 출근시간대 위주로 분석을 하였는데, 퇴근시간대 가장 인구 밀집이 높은 사당역 기점의 퇴근 노선도 분석을 하면 퇴근시간대에도 의미있는 결과가 산출될 것 같습니다. 또한, 수업에서 배웠던 파이썬 데이터 분석 코드를 실제로 활용할 수 있어 수업 복습에도 도움이 많이 되었습니다.