**버스별 수요도 분석**

**목차**

**1. 프로젝트 개요**

**2. 데이터 수집**

**3. 데이터 전처리**

**4. 데이터 분석 및 시각화**

**5. 결론**

**1. 프로젝트 개요**

**1.1 배경 및 목표**

버스 노선마다 배차 간격이 다름에 따라 수요도 차이로 인해 이러한 현상이 발생하는지 데이터 분석을 통해 확인하고자 함.

**1.2 선정 이유**

**"버스승객 수요가 버스 운행 간격에 직접적인 영향을 미치는가?"**

**1.3 데이터 수집 및 분석 개요**

**수집 및 처리한 데이터:**

* 데이터는 공공데이터 포털 API를 활용해 노선별 시간대별 승객 수 데이터를 수집
* 그 중 성남시만 운행하는 39개 노선의 버스만 Pandas를 활용하여 가공 및 정제
* 시각화 분석도구 Matplotlib, Seaborn, Folium.

**2. 데이터 수집**

**2.1 주요 자료**

1. 성남시 경유 노선 목록 데이터

2. 노선별 이용객 시계열 데이터

3. 노선별 정류소 데이터

4. 정류소 간 통과 노선 데이터

**2.2 수집 도구 및 방식**

- 공공데이터 API활용

- 교통카드 빅데이터 시스템

**2.3 데이터 정리 및 처리**

- 간단한 정리는 Excel VBA를 활용하여 정리 및 처리

- Pandas를 활용한 복잡한 정리 및 처리

텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명텍스트, 폰트, 일렉트릭 블루, 로고이(가) 표시된 사진

자동 생성된 설명API 수집

버스노선현황 리스트, 노선별 이용객 데이터 Excel텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명텍스트, 폰트, 스크린샷, 로고이(가) 표시된 사진

자동 생성된 설명텍스트, 폰트, 로고, 그래픽이(가) 표시된 사진

자동 생성된 설명

**3. 데이터 전처리**

**3.1 성남시 경유 노선 목록 데이터**

* 원본 데이터 : 총 197개의 버스 노선 목록
* 처리 과정:
  1. 성남시를 벗어나는 버스 노선 제거 -> 72개 노선 남음
  2. 고속버스, 마을버스 등 기타버스 제외하고 일반 버스만을 필터링 -> 최종 39개의 노선 확보
* 결과 : 성남시 내 일반 버스 39개의 노선 목록 데이터 완성

**3.2 노선별 이용객 시계열 데이터**

* 원본 데이터 : 23년 기준 버스 이용객 시계열 데이터
* 처리 과정 :
  + 전처리된 39개 버스 노선 목록을 기준으로 관련 없는 데이터 제거
* 결과 : 39개 노선의 이용객 시계열 데이터 생성

**3.3 노선별 정류소 데이터**

* 원본 데이터 : 197개 노선의 경유 정류소 데이터(API활용)
* 처리 과정 :
  1. 1차 API요청으로 197개 노선의 정류소 데이터 수집
  2. 노선 목록 수정(39개 노선으로 축소) 후 다시 API요청
* 결과 : 39개 노선에 대한 정류소 데이터 처리 완료

**3.4 정류소 간 통과 노선 데이터**

* 원본 데이터 : 경기도 기준 정류소 간 통과 노선 데이터
* 처리 과정 :
  1. 성남시 외부 정류소 데이터를 모두 제거
  2. 성남시 내 정류소만 남긴 데이터로 필터링
* 결과 : 성남시 내 정류소 간 통과 노선 데이터 생성

**3.5 추가 데이터 통합**

* 배차 간격 데이터:
  + 노선별 노선 ID를 기준으로 이용객 시계열 데이터(2번)와 병합
* 위치 데이터 통합:
  + 경기도 정류소별 데이터에서 위치 정보를 추출
  + 노선별 정류소 데이터(3번)에 위치 데이터 병합
* 결과 : 정류소와 노선 데이터를 통합한 최종 데이터 생성

**3.1.1 성남시 경유 노선 목록 데이터**

텍스트, 영수증, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

**3.1.2 API를 활용하여 노선번호와 노선 ID 매치**



텍스트, 폰트, 스크린샷, 보라색이(가) 표시된 사진

자동 생성된 설명

XML로 응답을 받고

regionName이 성남인것과

routeType이 일반형시내일경우만

텍스트, 폰트, 스크린샷, 번호이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 영수증, 폰트, 화이트이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷, 화이트이(가) 표시된 사진

자동 생성된 설명**3.2.1 노선별 이용객 시계열 데이터**

경기도 교통 정보 데이터 센터에서 Excel 다운 후 csv 변환

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명평일 주말로 나눔

텍스트, 흑백, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

**3.3.1 노선별 정류소 데이터 추출**

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

Xml을 통하여서 응답값 받아옴

**3.4.1 정류소간 통과 노선 데이터**

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명경기도 교통 정보 데이터 센터에서 추출

**4. 데이터 시각화 및 분석**

**4.1 월별 노선별 수요도(평일, 주말) 비교 시각화**

* 목표:월별로 각 노선의 평일과 주말 이용객 수를 비교
* 시각화 방법:
  + Line Plot을 이용하여 X축은 시간대, Y축은 이용개 수로 설정
  + 노선별로 데이터를 시각화하여 평일과 주말의 수요 차이를 분석

텍스트, 도표, 폰트, 지도이(가) 표시된 사진

자동 생성된 설명

**4.2 월별 노선 수요도 순위 비교 시각화**

* 목표: 각 노선의 월별 이용객 수에 따른 순위를 비교
* 시각화 방법:
  + Plot을 사용하여 X축은 시간대, Y축은 이용객 수를 기준
  + 상위 3개, 하위 3개 노선에 대한 비교 시각화
  + 순위 변동을 시각적으로 파악할 수 있도록 상위/하위 노선의 변화 추이를 나타냄

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

**4.3 월별 노선별 순위에 따른 배차 간격 비교 시각화**

* 목표: 상위 3개 및 하위 3개 노선의 배차 간격 차이를 분석
* 시각화 방법:
  + Plot을 사용하여 X축은 시간대, Y축은 배차 간격(분)을 기준으로 상위/하위 노선의 배차 간격 비교
  + 상위 노선일수록 배차 간격이 비교적 짧고, 하위 노선은 길다는 패턴을 시각적으로 비교

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

**4.4 월별 노선 수요도에 따른 정류소 지도 시각화**

* 목표: 상위 3개 및 하위 3개 노선이 지나가는 정류소의 트래픽을 비교
* 시각화 방법:
  + Folium 지도 시각화를 사용하여 각 노선이 지나가는 정류소를 Line으로 표시
  + 이용객 수가 많고 배차 간격이 짧은 노선의 라인을 두껍게 처리하여 트래픽을 시각적으로 표현
  + 지도 상에서 각 노선의 위치와 통행량을 한눈에 파악할 수 있도록 시각화

지도, 텍스트, 아틀라스이(가) 표시된 사진

자동 생성된 설명

**4.4.1.1 월별 노선별 수요도(평일, 주말)비교 시각화**

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명시각화 전 처리 과정 250번 노선을 기준으로 진행

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

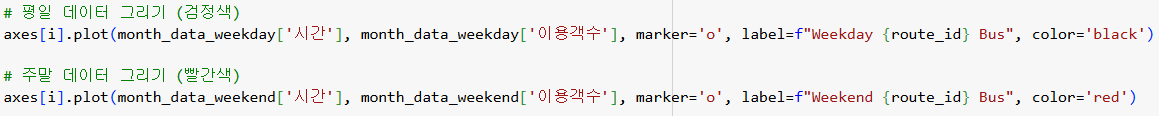
자동 생성된 설명

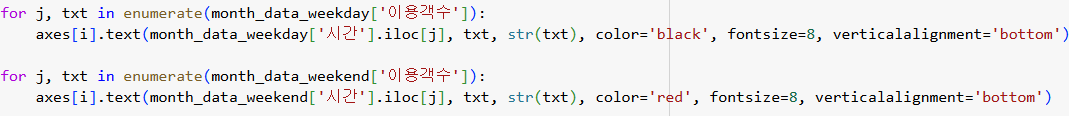
텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명

시각화 진행텍스트, 스크린샷, 폰트, 라인이(가) 표시된 사진

자동 생성된 설명





텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 도표, 폰트, 지도이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명**4.4.1.2 평일 및 주말 250번 계절별로 시각화**

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

평일도표, 라인, 그래프, 텍스트이(가) 표시된 사진

자동 생성된 설명 250번 계절별 시각화

주말 250번 계절별 시각화

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

**4.4.2 월별 노선 수요도 순위 비교 시각화**

텍스트, 폰트, 라인, 화이트이(가) 표시된 사진

자동 생성된 설명시각화 전 처리 과정

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명



텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

평일텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명 기준 1월달 상위3 하위3 노선 시간대별 이용객 시각화

주말 기준 1월달 상위3 하위3 노선 시간대별 이용객 시각화

텍스트, 도표, 라인, 그래프이(가) 표시된 사진

자동 생성된 설명

**4.4.3 노선 수요 순위에 따른 배차 간격 비교**

시각화 전 처리 과정

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명평일

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명주말(토요일, 일요일)

텍스트, 폰트, 스크린샷이(가) 표시된 사진

자동 생성된 설명

평일 시각화 1월

텍스트, 폰트, 스크린샷, 라인이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 번호, 폰트이(가) 표시된 사진

자동 생성된 설명

주말(토요일, 일요일) 시각화 1월

텍스트, 스크린샷, 라인, 도표이(가) 표시된 사진

자동 생성된 설명

**4.4.4 월별 노선 수요도에 따른 정류소 지도 시각화**

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명시각화 전 처리 과정

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 라인, 화이트이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트이(가) 표시된 사진

자동 생성된 설명

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

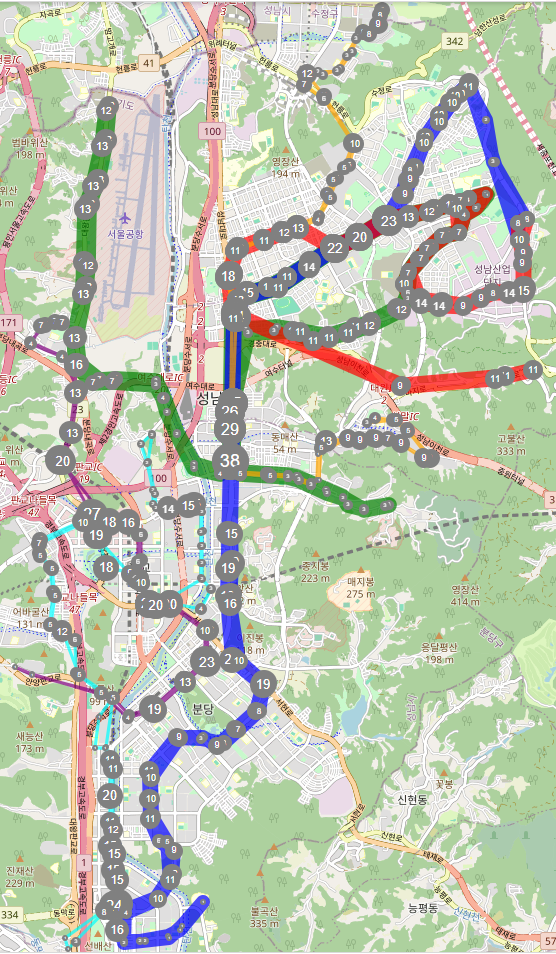
자동 생성된 설명

텍스트, 폰트, 라인, 스크린샷이(가) 표시된 사진

자동 생성된 설명평일 1월 07:00-08:00 시간대 노선별 상위3 하위3 트래픽

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명



텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명토요일 1월 07:00-08:00 시간대 노선별 상위3 하위3 트래픽

지도, 텍스트, 아틀라스이(가) 표시된 사진

자동 생성된 설명

텍스트, 스크린샷, 폰트, 번호이(가) 표시된 사진

자동 생성된 설명일요일 1월 07:00-08:00 시간대 노선별 상위3 하위3 트래픽

지도, 텍스트, 아틀라스이(가) 표시된 사진

자동 생성된 설명

**5. 결론**

**5.1 특정 노선의 시간대별 수요 분석**

* 특정 노선의 시간대별 데이터를 분석한 결과, 평일 출퇴근 시간대에 이용객이 가장 많이 몰리는 현상이 두드러졌다.
* 주말에는 상대적으로 완만한 이용 패턴을 보이며, 평일과 주말 간 그래프의 형태가 달랐다.
  + 예 : 250번 평일에는 출퇴근 시간대에 집중된 M자 형태, 주말에는 완만한 포물선 형태로 나타남.

**5.2 상위/하위 노선별 수요 비교**

* 상위 3개 노선은 평일 기준 1대당 최대 약 6,000명의 이용객을 기록하며, 높은 수요를 보임.
* 하위 3개 노선은 출퇴근 시간대에도 3자리 이용객 수준이며, 대부분의 시간대에서는 평균적으로 2자리 이용객을 넘지 못함.
* 주말에는 상위 노선조차 최대 약 2,500명 수준으로, 주말의 전반적인 수요 감소가 뚜렷하게 나타남.
* 그래프 패턴 역시 평일은 출퇴근 시간대에 집중 M자형태 , 주말은 완만한 포물선 형태의 분포를 보임.

**5.3 상위/하위 노선별 배차간격 비교**

* 상위 노선은 평일 평균 10~15분 간격으로 운행되며, 일부 시간대 수요도가 1등인 노선은 5분 간격으로도 배차됨.
* 하위 노선은 평일에는 15~30분 간격, 주말에는 40~65분 간격으로 배차가 크게 늘어남.
* 주말 배차간격이 상위/하위 노선 모두 평일보다 길어지는 현상이 관찰됨

**5.4 상위/하위 노선의 정류소 및 경로 비교**

* 상위 노선은 많은 정류소를 지나며, 도시 주요 지역 및 지하철역을 연결하는 경향이 있음 -예: 모란역은 60개 가량의 노선이 지나며 주요 교통 거점으로 작용, 출퇴근 시간대 교통 체증 유발 가능성이 높음.
* 하위 노선은 주로 동네 내부만 운행하며, 주요 역과 연결되지 않는 경향이 강함.

**5.5 종합 결론**

* 수요도가 적은 노선일수록 배차간격이 길어지는 현상이 관찰됨.
* 수요도가 높은 노선은 주로 사람들이 많이 사는 지역 및 주요 역을 연결하는 특징을 가짐.