

파이썬, SQL, R을 사용한 마케팅을 위한 데이터 분석: 비회원을 회원으로 전환하기



1. 데이터 클리닝 과정

2. 데이터 분석을 통한 인사이트



1.데이터 클리닝 과정

1.데이터 클리닝 과정

누락된 값 확인, 결과는 다음과 같습니다.

```
print(data.isnull().sum())
```

- start_station_name 29283
- start_station_id 29283
- end_station_name 31158
- end_station_id 31158



누락된 데이터 채우기

- 출발지와 도착지의 고유 위도와 경도를 사용하여 고유한 정류장 정보 추출
- 출발지와 도착지의 이름 및 ID가 누락된 행을 찾아 해당 위도/경도에 해당하는 고유 정류장 정보로 채움

분하여 유일한 역 정보를 추출

```
na(subset=['start_station_name', 'start_station_id'])  
(subset=['end_station_name', 'end_station_id'])  
  
:  
on_name']) or pd.isnull(row['start_station_name']) or pd.isnull(row['start_station_id']) or  
ions[(unique_start_stations['start_lat'] == row['start_lat'] && unique_start_stations['start_lon'] == row['start_lon'])]  
_station_name'] = match.iloc[0]['start_station_name']  
_station_id'] = match.iloc[0]['start_station_id']  
  
:  
_name']) or pd.isnull(row['end_station_name']) or pd.isnull(row['end_station_id']) or  
ns[(unique_end_stations['end_lat'] == row['end_lat'] && unique_end_stations['end_lon'] == row['end_lon'])]  
station_name'] = match.iloc[0]['end_station_name']  
station_id'] = match.iloc[0]['end_station_id']
```

Made with Gamma

데이터 완성 결과

- start_station_name 18896
- start_station_id 18896
- end_station_name 20230
- end_station_id 20230

10,387개의 데이터 복구



시작 및 종료 시간 데이터를 날짜/시간 형식으로 변환하여 trip_duration 계산

- `data['started_at'] = pd.to_datetime(data['started_at'])`

일부 행 추가

- `['trip_duration'] = data['ended_at'] - data['started_at']`
데이터 변환을 기반으로 한 trip_duration 계산. 음수인 trip_duration 제거
- `['distance'] , ['Day of week'], ['hour_of_day'], ['Frequency'],` 등등

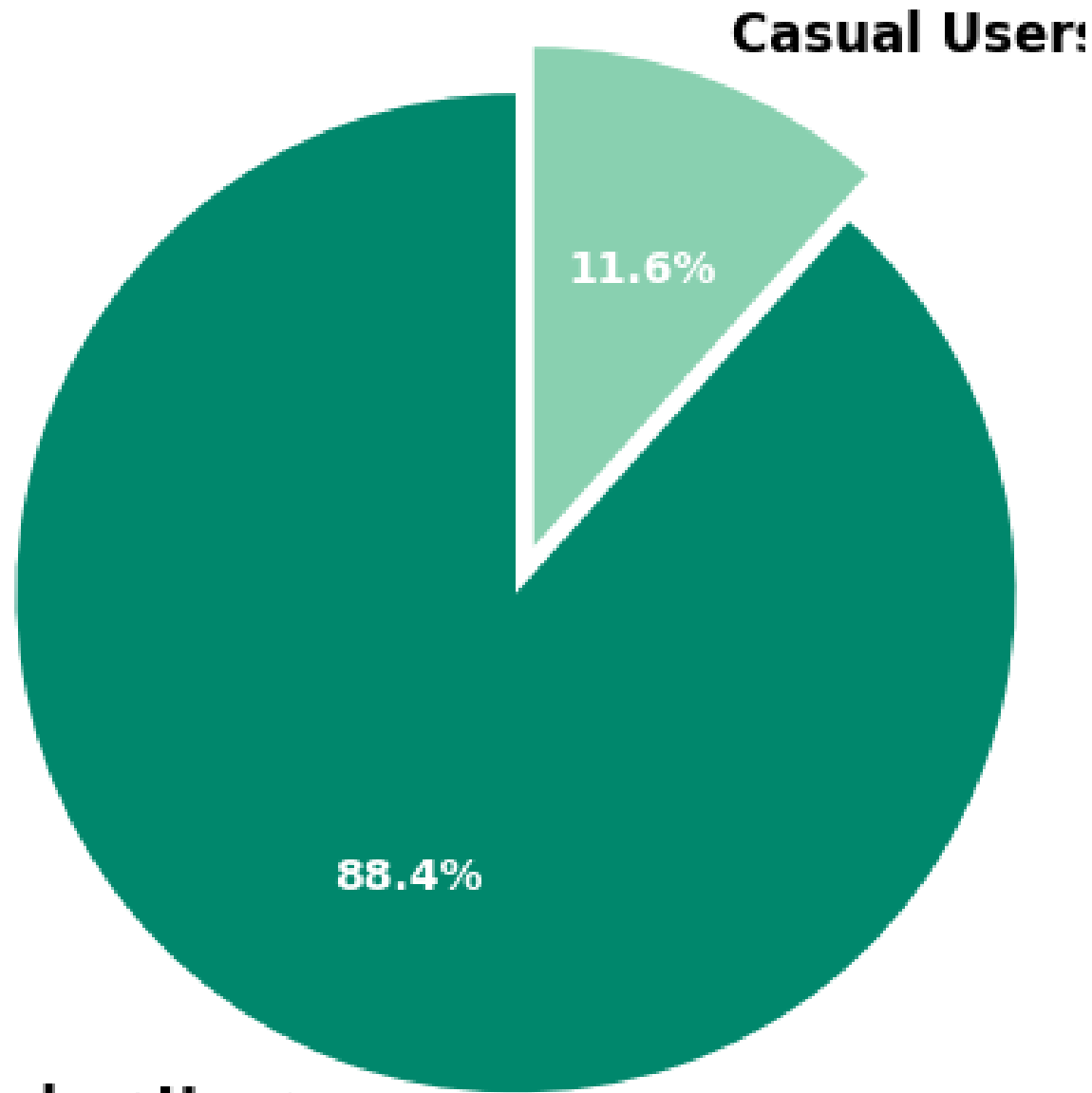
2.데이터 분석을 통한 인사이트

- 회원과 비회원 간의 패턴 차이
- 비회원 리더를 위한 마케팅 전략

-회원과 비회원 간의 패턴 차이

- 사용자 유형별 매출 비교
- 요일별 사용 빈도
- 시간대별 사용 분포: 회원 대 비회원

Revenue Share by User Type

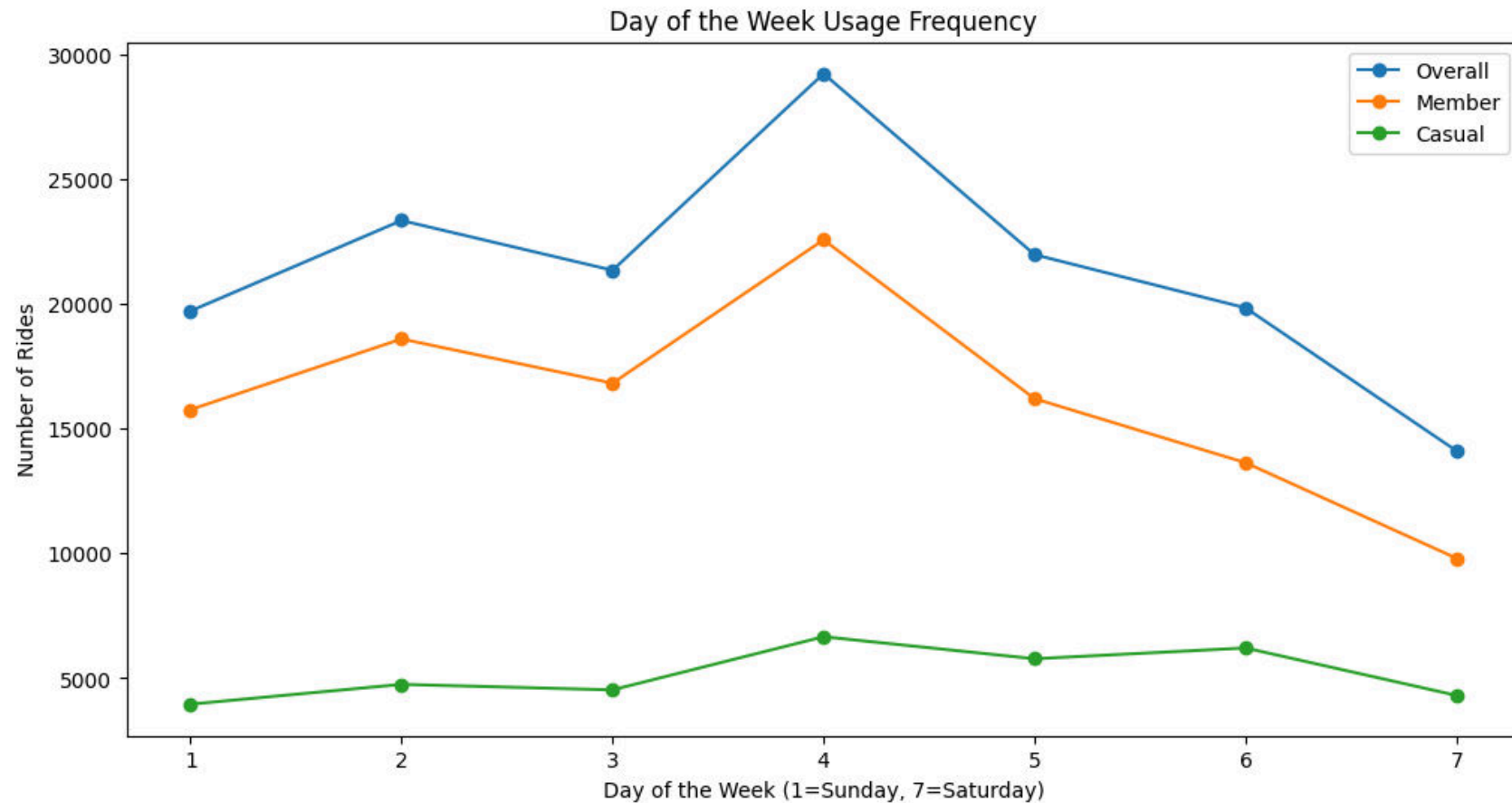


사용자 유형별 매출

원형 차트를 통해 연간 회원이 비회원에 비해 더 많은 수익을 생성한다는 것을 확인

이를 바탕으로 비회원을 회원으로 전환하는 마케팅 전략 수립이 총 수익 증가에 기여할 것이라 결론

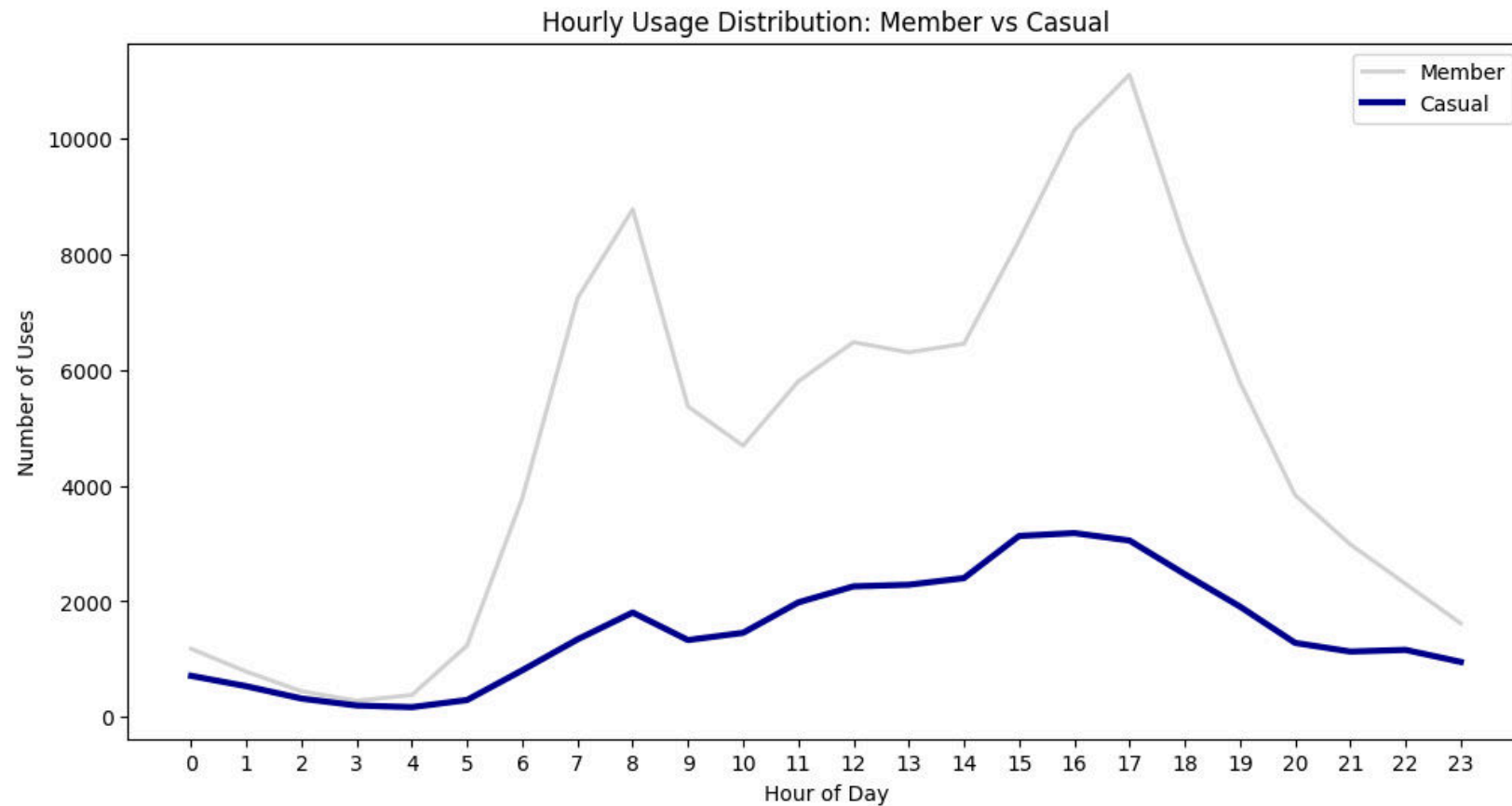
Member Users



회원과 비회원의 사용 패턴

회원의 경우 월요일과 수요일에 사용량이 가장 높으며, 목요일부터 토요일까지 사용량이 크게 감소

비회원의 경우 월요일과 화요일에 사용량이 낮지만 수요일부터 금요일까지 사용량이 높아짐



시간대별 사용 분포: 회원 대 비회원

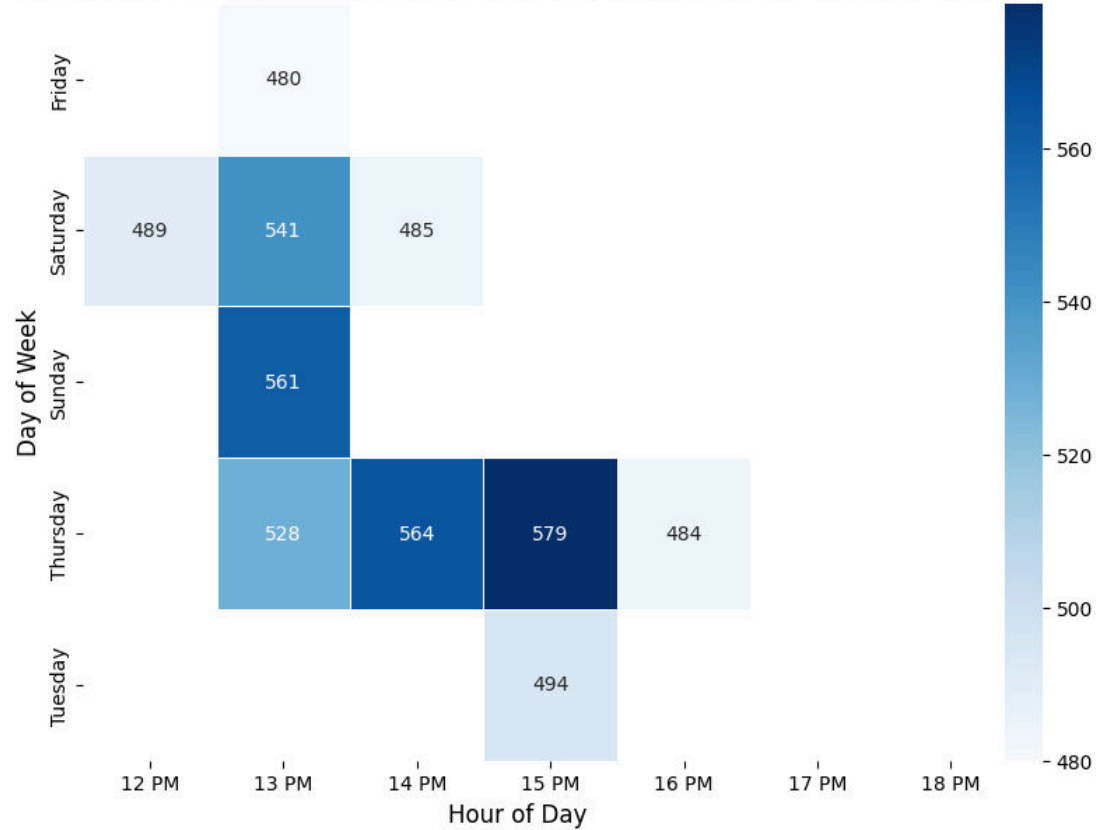
- **회원 사용:** 오전 8시와 오후 5-6시에 사용량이 증가, 통근 패턴을 나타냄
- **비회원 사용:** 오후 3시에 사용량이 증가하여 오후 6시 이후까지 안정적, 여가용 사용 가능성
- **데이터 분석 인사이트:** 다른 사용 패턴을 통해 타겟 마케팅 전략을 제안함
- **마케팅 전략:** 회원에게는 통근 인센티브를, 비회원에게는 여가 시간 프로모션을 제공하여 회원 전환 증가

-비회원 라이더를 위한 마케팅 전략

- 주요 사용 정류장 식별
- 주요 사용 시간대 파악

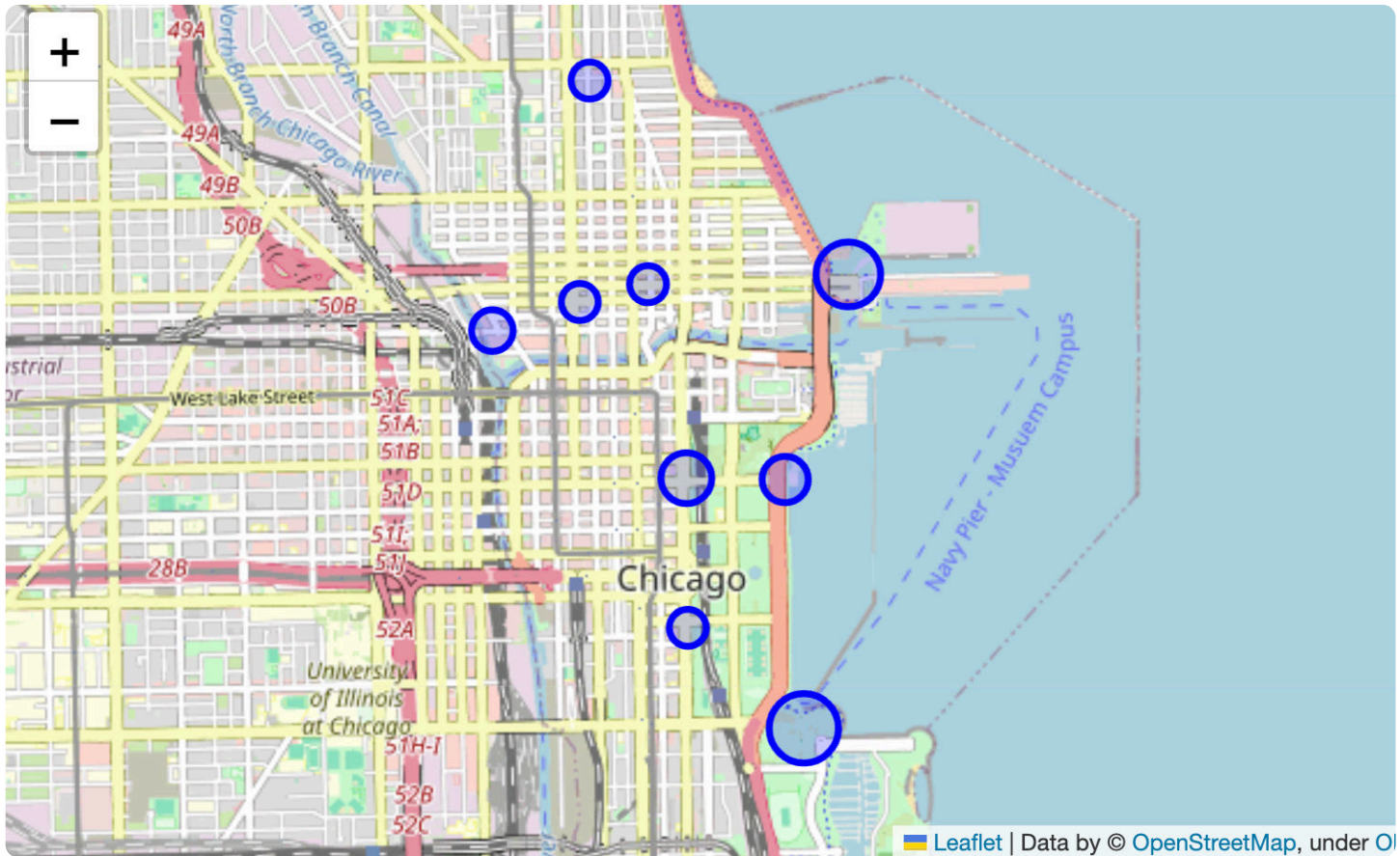
비회원 사용 빈도에 따른 상위 10개 시간대 및 요일

Heatmap of Usage Frequency by Day and Hour for Casual Users



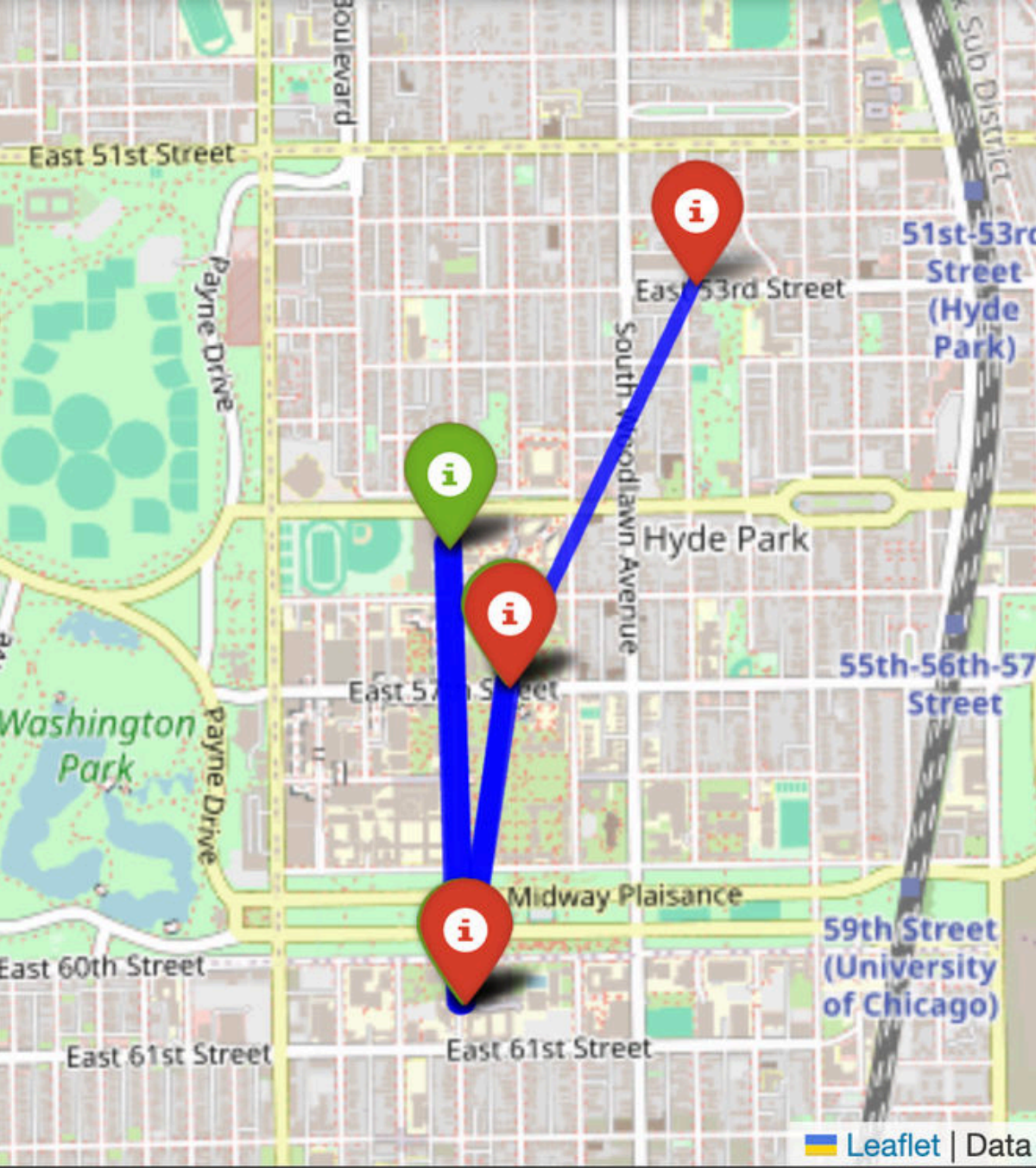
- 가장 높은 사용 빈도는 목요일 오후 3시(15시)로 579회
- 평일 오후 2시부터 4시(14시부터 16시) 사이의 사용량이 다른 시간대에 비해 일관되게 높음

비회원 주요 사용 정류장



각 정류장의 총 트래픽을 계산한 후, 위도와 경도 정보와 함께 지도에 상위 10개 정류장 구현

"이 인사이트를 활용하여, 우리는 주요 정류장에서 GPS 기반 마케팅을 실시하고 피크 시간을 타겟팅함으로써 우리의 효과를 강화할 수 있습니다."



비회원 주요 10대 루트

*Streeter Dr & Grand Ave **to** Streeter Dr & Grand Ave*

*Ellis Ave & 60th St **to** Ellis Ave & 55th St*

*DuSable Lake Shore Dr & Monroe St **to** DuSable Lake Shore Dr & Monroe St*

*Ellis Ave & 55th St **to** Ellis Ave & 60th St*

*DuSable Lake Shore Dr & Monroe St **to** Streeter Dr & Grand Ave*

*Ellis Ave & 60th St **to** University Ave & 57th St*

*University Ave & 57th St **to** Ellis Ave & 60th St*

*University Ave & 57th St **to** Kimbark Ave & 53rd St*

*Streeter Dr & Grand Ave **to** Millennium Park*

*Sheffield Ave & Fullerton Ave **to** Greenview Ave & Fullerton Ave*

비회원의 주요 루트에 기반한 프로모션

1. **목적지 기반 프로모션:** 관광지, 쇼핑센터, 식당과 같이 자주 방문하는 장소와 협력하여 할인이나 프로모션 제공
2. **이벤트 연계 마케팅:** 루트 근처에서 열리는 문화 행사나 축제와 파트너십을 맺어 특별 초대나 참여 기회 제공
3. **지역 비즈니스 파트너십:** 근처 비즈니스(카페, 식당, 상점)와 파트너십을 맺어 목적지에 도착한 라이더에게 특별 할인이나 서비스 제공, 예를 들어 할인 쿠폰이나 프로모션 코드 제공

이 전략들은 인기 있는 루트를 활용하여 라이더 경험을 향상시키고 프로모션과 파트너십을 통해 서비스 사용을 장려하여 회원 전환을 촉진하는 것을 목표로 합니다.

사용 시간, 인기 정류장, 루트를 기반으로 한 혜택을 제공함으로써 비회원을 회원으로 전환하는 비율을 증가시키는 마케팅 노력에 중점을 둡니다

2022년 12월의 데이터를 기반으로 한 마케팅을 위한 데이터 분석을 발표했습니다.

분석 과정과 코드는 밑에 있는 GitHub 링크를 통해서 확인할 수 있습니다.

또한 2023년 상반기 분석에 대한 자료를

SQL을 사용해 추출한 데이터와 R을 통한 시각화한 코드와 자료를 업로드했습니다.

GitHub URL: <https://github.com/Jun6481/portfolio>