05

실습예제 (생년월일(YYMMDD) 데이터 분석)

ChatGPT 고급 데이터 분석 실습 예제

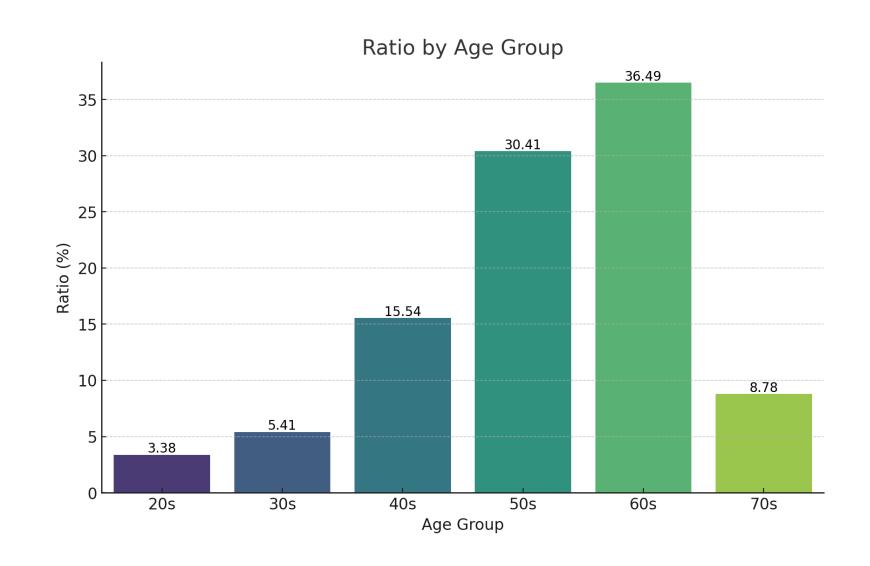
생년월일(YYMMDD) 데이터 분석

생성 AI 기술 발전으로 데이터 분석의 장벽이 낮아지고 있다.

생년월일 예제 데이터

YYMMDD 형식을 가지는 1 4 8 건의 데이터 샘플에서 연령대별(20대 ~ 70대) 비율을 보여주는 예제

Date of Birth	
781224	
561025	
610423	
680823	
630517	
630517	
490525	
671116	
720925	
610525	
731117	
731117	
620625	
620120	
720324	
581003	
581003	
591103	
670219	
570830	
650305	
580425	
610302	
660725	
680310	
680730	
590219	
730410	



엑셀을 사용한다면?

매크로 활용 또는 데이터 가공 이후 시각화

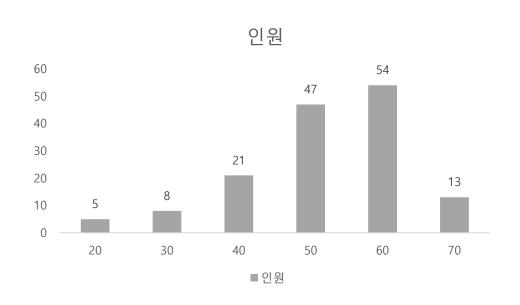
=C	ountByAge(•			
	D	E	F	G	н	J
					연령	인원
	781224	44	40		20	5
	561025	66	60		30	8
	610423	62	60		40	21
	680823	55	50		50	47
	630517	60	60		60	54
	630517	60	60		70	13

```
Function CountByAgeGroup
    (rng As Range, lowerAge As Integer, upperAge As Integer) As Integer
  Dim cell As Range
  Dim count As Integer
  Dim birthYear As Integer
  Dim currentYear As Integer
  Dim age As Integer
  count = 0
  current Year = Year(Now)
  For Each cell In rng
    birthYear = 1900 + Left(cell.Value, 2)
    age = currentYear - birthYear
    If age >= lowerAge And age < upperAge Then
      count = count + 1
    End If
  Next cell
  CountByAgeGroup = count
End Function
```

엑셀을 사용한다면?

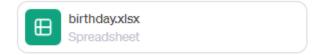
직접적인 데이터 가공 이후 시각화 (FLOOR함수와 COUNTIF함수)

fx > =COUNTIF(F\$6:F\$153,H25)								
D		Е	F	G	н	J		
6702	219	56	50		연령	인원		
5708	330	66	60		20	5		
6503	305	58	50		30	8		
5804	125	65	60		40	23		
6103	302	62	60		50	45		
6607	725	57	50		60	54		
6803	310	55	50		70	13		



CHATGPT를 활용하면?





데이터를 요약해주세요.

2. Unnamed: 1:

- 평균: 56.31
- 표준 편차: 11.43
- 최솟값:24
- 25% 분위수: 50
- 중앙값 (50% 분위수): 57
- 75% 분위수: 64
- 최댓값:79

관찰 사항:

- "Date of Birth (yymmdd)" 열은 날짜 값을 숫자 형식 (yymmdd)으로 나타내고 있습니다. 데이 터의 범위는 상당히 넓으며, 가장 오래된 것은 1944년 4월 25일 (440425)이고, 가장 최근의 것 은 1998년 11월 30일 (981130)입니다.
- "Unnamed: 1" 열은 연령 또는 개인과 관련된 다른 숫자 지표일 수 있습니다. 값의 범위는 24에서 79까지이며, 평균은 약 56.31, 표준 편차는 약 11.43입니다.

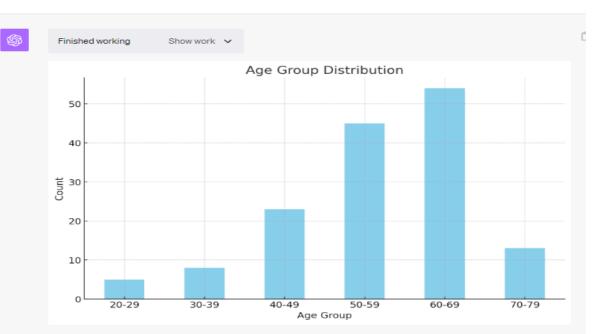
다음 단계:

분석의 구체적인 목표에 따라 다음과 같은 단계를 진행할 수 있습니다:

- 1. "Date of Birth (yymmdd)"를 더 쉬운 분석을 위해 적절한 날짜-시간 형식으로 변환합니다.
- 2. "Unnamed: 1" 열의 이름을 그 목적에 따라 더 설명적인 것으로 변경합니다.
- 3. 누락된 값이나 이상치가 있다면 처리합니다.



연령대별 분포를 막대그래프로 그려주세요.



연령대별 분포

- 20-29:5명
- 30-39:8명
- 40-49: 23명
- 50-59: 45명
- 60-69:54명
- 70-79:13명

위의 막대그래프와 데이터를 보면, 가장 많은 사람들이 60-69 연령대에 속해 있으며, 그 다음으로는 50-59 연령대가 많습니다. 반면, 20-29 연령대와 30-39 연령대에는 상대적으로 적은 수의 사람들이 속해 있습니다.

이러한 연령 분포는 고령층이 많이 포함되어 있음을 나타냅니다, 이는 고령층 중심의 서비스나 제품을 고려할 때 유용한 인사이트가 될 수 있습니다.

다음 분석 단계나 특별한 요구사항이 있으신가요?

CHATGPT를 활용하면?



막대 그래프 위에 인원수를 넣어주세요.





막대그래프 위에 각 연령대별 인원 수가 표시되어 있습니다. 이렇게 하면 그래프를 보면서 쉽게 각 연령대에 얼마나 많은 사람들이 있는지 파악할 수 있습니다.

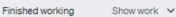
이러한 시각화는 데이터의 분포를 빠르게 이해하고, 특정 연령대에 초점을 맞춘 전략을 구성할 때 도움이 될 수 있습니다.

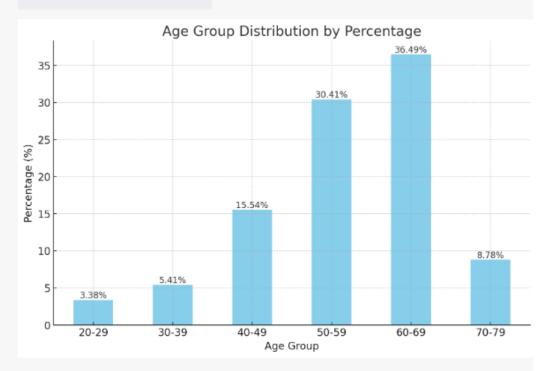
다른 분석이나 질문이 있으신가요?



인원수 대신 비율로 변경해주세요.







연령대별 분포 (비율)

• **20-29**: 3.38%

• **30-39**: 5.41%

• **40-49**: 15.54%

• **50-59**: 30.41%

• **60-69**: 36.49%

• **70-79**: 8.78%

막대그래프 위에 각 연령대별로 차지하는 비율이 표시되어 있습니다. 이를 통해 60-69 연령대가 전체의 약 36.49%로 가장 큰 비율을 차지하고 있으며, 그 다음으로는 50-59 연령대가 30.41%로 이어집니다. 반면, 20-29 연령대와 30-39 연령대는 각각 3.38%와 5.41%. 30.89

06

실습예제 (비정형 텍스트 데이터 쇼핑몰 리뷰 텍스트 데이터 분석)

한국어▼



Extract data efficiently with

리스틀리

A free easy-to-use web scraping tool. Quickly solve your data problems today!

크롬확장프로그램 추가

사용법 안내 (영상)

★★★★★ 전 세계 **209,400+** 사용자가 신뢰합니다

실시간으로 추출된 URL을 확인해보세요!

O https://new.land.naver.com/offices?ms... 1 hour ago

O https://besthomewarranties.inlinecrm.c... 1 hour ago

O https://www.google.com/search?sca_e... 1 hour ago





로보락 S8 Pro Ultra 최저 1.567.620원 ₽무료

최저가 사러가기

판매처 216	제품정보	블로그리뷰	쇼핑몰리뷰 35,007	AiTEMS추천	
	사용	성 어플 사용법도 어렵지 않	낳고 보기 편하게 되어있습니다		>

랭킹순 | 최신순

포토&동영상리뷰만 보기

70		

전체 (35,007) 5점 (32,884) 4점 (1,884) 3점 (173) 2점 (12) 1점 (54)

주제전체 만족도 가격 관리 소음 디자인 성능 조작성 사용성 기능 효과 향기 색상 품질 🗸

★★★★★ 5 로보락 공식 스토어 · ioja**** · 24.02.20.

처음 구매하는 로봇청소기 이라 여러제품을 기능,AS 등을 비교하느라 10일정도 고민한것 같아요.기능이 좋고 가격이 저렴…

처음 구매하는 로봇청소기 이라 여러제품을 기능,AS 등을 비교하느라 10일정도 고민한것 같아요.

기능이 좋고 <mark>가격이 저렴해 해외 직구도 고민을 해 봤지만 AS 가 걸리고</mark> 해서 비싸더라도 AS,성능,판매량 등 종합적으로 로보락이 좋더라구요. 결정하고 바로 구매 하였어요.

설치:처음 제품 받고 무게가 엄청 무겁더라구요.

설치, 세팅이 어려울까 걱정도 있었는데

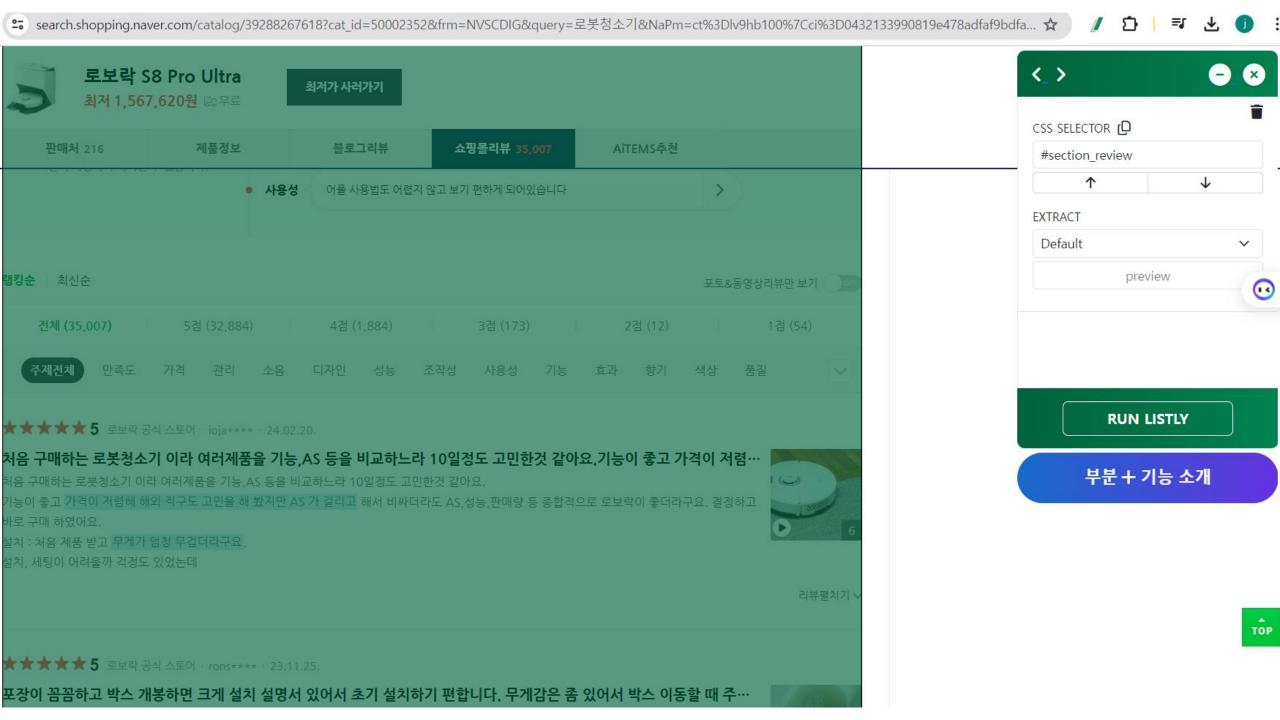


리뷰펼치기 ~

★★★★★ 5 로보락 공식 스토어 · rons**** · 23.11.25.

포장이 꼼꼼하고 박스 개봉하면 크게 설치 설명서 있어서 초기 설치하기 편합니다. 무게감은 좀 있어서 박스 이동할 때 주…







PARTS

Copy Selector

Copy JS Path

② 2024년 4월 21일 9:26 오후

https://search.shopping.naver.com/catalog/39288267618?cat_id=50002352&frm=NVSCDIG&query=%EB%A1%9 C%EB%B4%87%EC%B2%AD%EC%86%8C%EA%B8%B0&NaPm=ct%3Dlv9hb100%7Cci%3D0432133990819e478adfaf9b dfa3f34b9c5f5f09%7Ctr%3Dsls%7Csn%3D95694%7Chk%3D323c3ba8b240f30be2274f57d3934a5e0afa83ac

cjh1446@gmail.com, 님, 프로필 정보를 입력해주세요.

▦ 테이블표 NO

⊙ 쿠키 사용 NO

● 브라우저 설정 접속기기: 데스크탑 | 프록시 | 로그인 | 액션스크립트 | 자동 스크롤 | 자동 저장 | 자동 클릭 | 로딩 대기 | 템플릿

폴더 다운로드

✔ 데이터정렬

하이퍼링크

✓ 컬럼명

ዹ 엑셀

G 구글시트 ▼

% API^{₩|E|}

% Hook^{≌்⊥்}

Enter keyword to search here...

추출할 탭을 선택하세요										
20 개의 카드	4 개의 카드	5 개의 카드	5 개의 카드	5 개의 카드	5 개의 카드					
5 개의카드	5 개의카드	4 개의 카드	4 개의 카드	5 개의 카드	5 개의 카드	5 개의 카드	4 개의 카드	4 개의 카드	5 개의 카드	
4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	4 개의 카드	

파일 홈 삽입 페이지 레이아웃 수식 데이터 검토 보기 도움말

(i) 제한된 보기 주의하세요—인터넷에서 가져온 파일에는 바이러스가 있을 수 있습니다. 편집하지 않는다면 제한된 보기에서 여는 것이 안전합니다.

편집 사용(E)

A1	A1 · X · / fx LABEL-1										
	Α	В	С	D	E	F	G	Н			
1	LABEL-1	LABEL-2	LABEL-3	LABEL-4	LABEL-5	LABEL-6	LABEL-7	LABEL-8	LABEL-9		
2	평점	평점5	로보락 공식 스토어	ioja****	24.02.20.	처음 구매하는 로봇청소	처음 구매하는 로봇청소	https://phinf.pstatic.net/i	동영상		
3	평점	평점5	로보락 공식 스토어	rons****	23.11.25.	포장이 꼼꼼하고 박스 개	포장이 꼼꼼하고 박스 개				
4	평점	평점5	로보락 공식 스토어	jong****	23.04.16.	로봇청소기가 똑똑하면	로봇청소기가 똑똑하면	https://phinf.pstatic.net/i	동영상		
5	평점	평점5	로보락 공식 스토어	cksg****	23.04.30.	-구매사유-저는 원래 청	-구매사유-저는 원래 청	https://phinf.pstatic.net/i	동영상		
6	평점	평점5	로보락 공식 스토어	gnc0****	23.05.03.	집사람이 직업상 어깨를	집사람이 직업상 어깨를	https://phinf.pstatic.net/i	동영상		
7	평점	평점5	로보락 공식 스토어	pink****	23.08.31.	회사 사람들이 로봇 청소	회사 사람들이 로봇 청소	https://phinf.pstatic.net/i	동영상		
8	평점	평점5	로보락 공식 스토어	bell****	23.07.09.	로봇청소기는 처음 구매	로봇청소기는 처음 구매	https://phinf.pstatic.net/i	동영상		
9	평점	평점5	G마켓	l*y****	23.05.15.	청소 도우미 아주머니 작	안녕하세요. 우선 이전부				
10	평점	평점5	G마켓	W*W****	23.05.24.	저희 집 콩쥐 리뷰입니	도착하기도 전에 이름부				
11	평점	평점5	로보락 공식 스토어	ysr7****	23.06.12.	맞벌이 부부라 청소하는	맞벌이 부부라 청소하는	https://phinf.pstatic.net/i	동영상		
12	평점	평점5	로보락 공식 스토어	dytp****	23.05.06.	3년 간 고민하다 드디어	3년 간 고민하다 드디어				
13	평점	평점5	G마켓	k*y****	23.05.26.	드디어 구매한 로보락 S8	몇 달 간 구매하려고 기				
14	평점	평점5	G마켓	a*r****	23.11.14.	막내가 생겼습니다.	배송은 너무 깜짝 놀랄만				
15	평점	평점4	로보락 공식 스토어	esth****	23.07.02.	남편과 지인이 추천해서	남편과 지인이 추천해서				
16	평점	평점5	로보락 공식 스토어	jiwo****	23.07.09.	먼저 총평을 하자면 너무	먼저 총평을 하자면 너무	https://phinf.pstatic.net/i	동영상		
17	평점	평점5	로보락 공식 스토어	alsw****	23.12.05.	전세집에 살고 있는 신혼	전세집에 살고 있는 신혼	https://phinf.pstatic.net/i	동영상		
18	평점	평점5	로보락 공식 스토어	jds9****	23.04.15.	로봇청소기 사려고 한달	로봇청소기 사려고 한달	https://phinf.pstatic.net/i	동영상		
19	평점	평점4	로보락 공식 스토어	dear****	23.12.03.	항상 기대해왔던 로봇청	항상 기대해왔던 로봇청	https://phinf.pstatic.net/i	동영상		
20	평점	평점5	로보락 공식 스토어	thfk****	24.01.10.	로봇 청소기를 처음 써보	로봇 청소기를 처음 써보				
21	평점	평점5	로보락 공식 스토어	sims****	23.04.18.	LG 로봇청소기 사용하다	LG 로봇청소기 사용하다				
22											

07

실습예제 (자전거 대여 수요예측 kaggle 데이터)







Home

 $oldsymbol{\Phi}$ Competitions

□ Datasets

M Models

<> Code

Discussions

★ Learn

More

Q Search

Competitions

Q bike sharing demand

Results



Bike Sharing Demand

Forecast use of a city bikeshare system Playground \cdot 3242 Teams \cdot 9 years ago



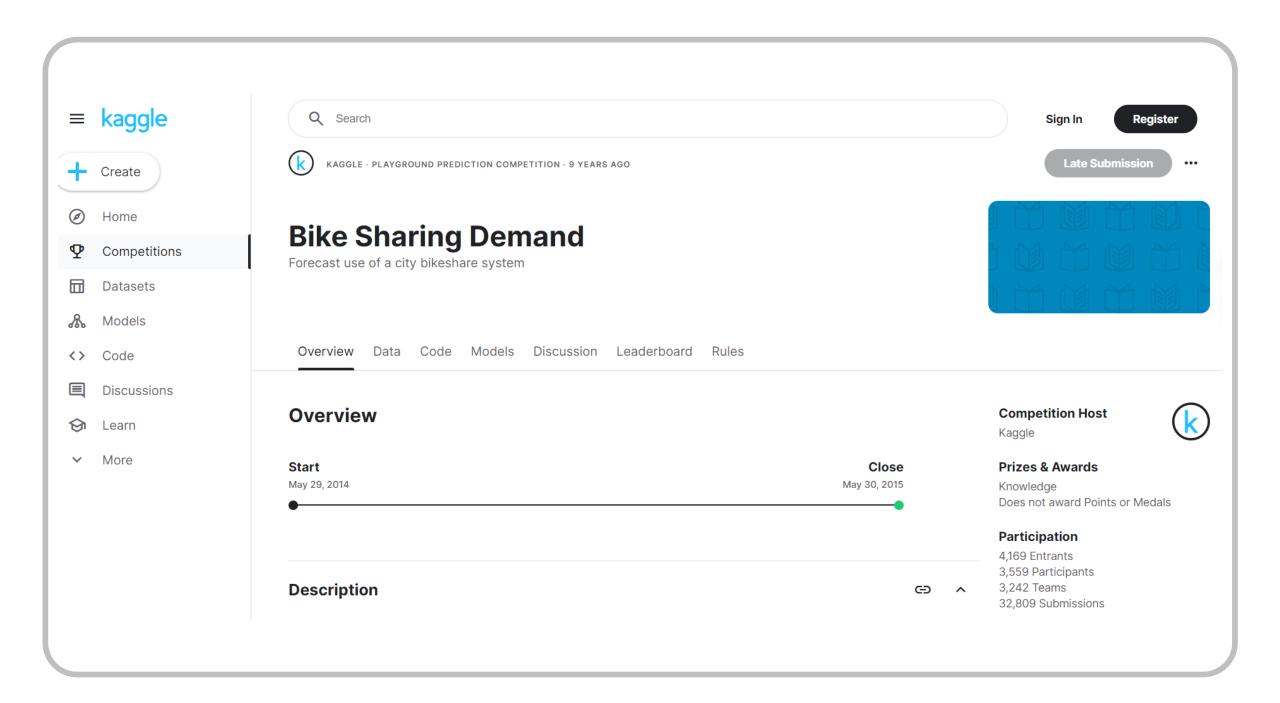
INF131 - 2019 Challenge

Build a regressor to predict bike sharing demand Community \cdot 69 Teams \cdot 4 years ago



Bike Sharing Hourly Demand Forecasting

Forcast bike sharing system hourly demand Community \cdot 44 Teams \cdot Invitation Only \cdot 4 years ago

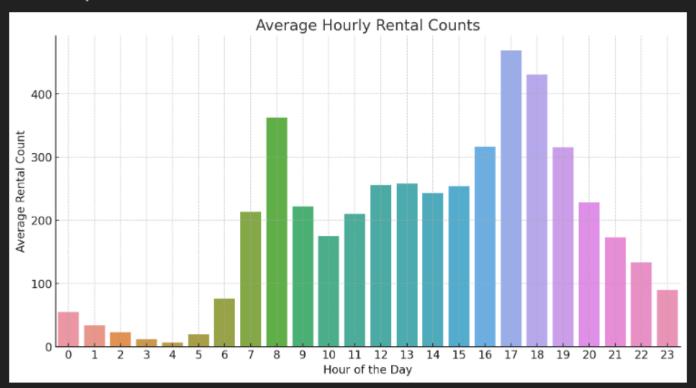


```
In [1]: import pandas as pd
        import numpy as np
        import matplotlib as mpl
        import matplotlib.pyplot as plt
        import seaborn as sns
        from scipy import stats
        # 노트북 안에 그래프를 그리기 위해
        %matplotlib inline
       # 그래프에서 격자로 숫자 범위가 \chi눈에 잘 띄도록 ggplot 스타일을 사용
        plt.style.use('ggplot')
        # 그래프에서 마이너스 폰트 깨지는 문제에 대한 대처
       mpl.rcParams['axes.unicode_minus'] = False
```



ChatGPT

시간별로 총 대여량('count')을 시각화하여, 하루 중 각 시간대별 대여 패턴을 파악해 보겠습니다. 이를 위해 'datetime' 컬럼을 datetime 타입으로 변환하고, 시간 컬럼을 추출한 후, 시간별로 대여량의 평균을 계산하여 시각화하겠습니다.



위 그래프는 하루 중 각 시간대별 평균 대여량을 보여줍니다. 몇 가지 특징적인 패턴을 볼 수 있습니다: