자동차 보험금 데이터 분석

Auto Insurance Data Analysis Project

기초 데이터 분석 및 실습

프로젝트 개요 (Overview)

분석 목표

2024년 상반기 자동차 사고 피해자 데이터를 기반으로 다음을 분석합니다.

- 부상 급수와 보험금 지급액의 상관관계
- 장애 유무에 따른 보험금 차이 검증
- 월별 보험금 지급 추이 파악

데이터 출처 (Data Source)

공공데이터포털(data.go.kr)의 API를 활용하여 데이터를 수집하였습니다.

항목	내용
API 명	자동차사고 피해자 월별 정보 (GetAutoInsVicMonthInfo)
제공 기관	한국교통안전공단
수집 기간	2024년 1월 ~ 2024년 6월
데이터 포맷	JSON -> DataFrame 변환

데이터 구조 (Data Structure)

- **YearMonth:** 접수 년월 (예: 202401)
- **InjuryGrade:** 부상 급수 (1급 ~ 14급)
- **Disability:** 장애 여부 (0: 없음, 1: 있음)
- **Count:** 해당 건수
- **AvgInsurance:** 평균 보험금 지급액 (원)

데이터 특성

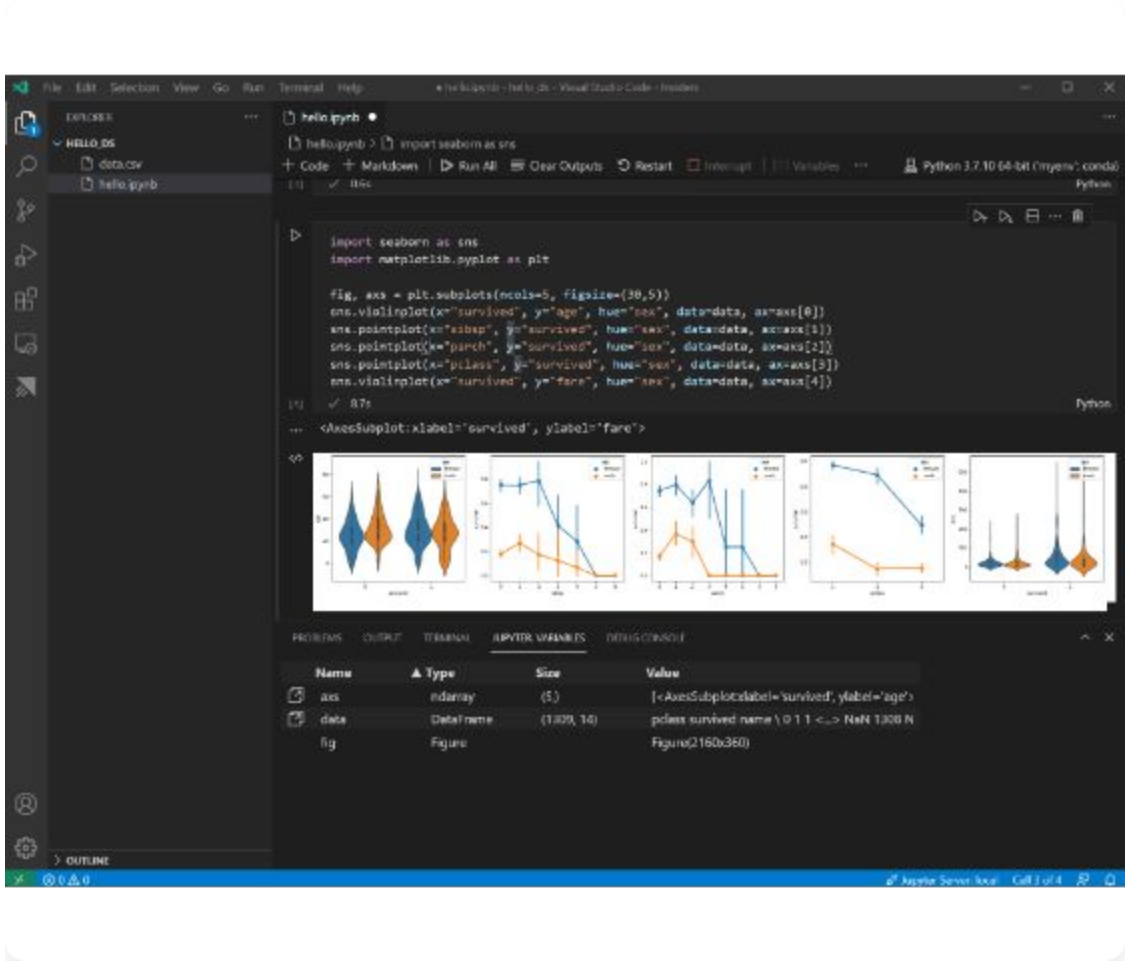
수치형 데이터(보험금, 급수)와 명목형 데이터(장애여부, 년월)가 혼합되어 있어 다양한 통계적 분석이 가능합니다.

데이터 수집 (Data Collection)

Python requests 라이브러리를 사용하여 API 데이터를 호출하고, pandas를 통해 데이터프레임으로 변환하였습니다.

주요 로직:

202401부터 202406까지 월별 반복문을 통해 데이터를 수집(extend) 후 통합하였습니다.



데이터 전처리 (Preprocessing)

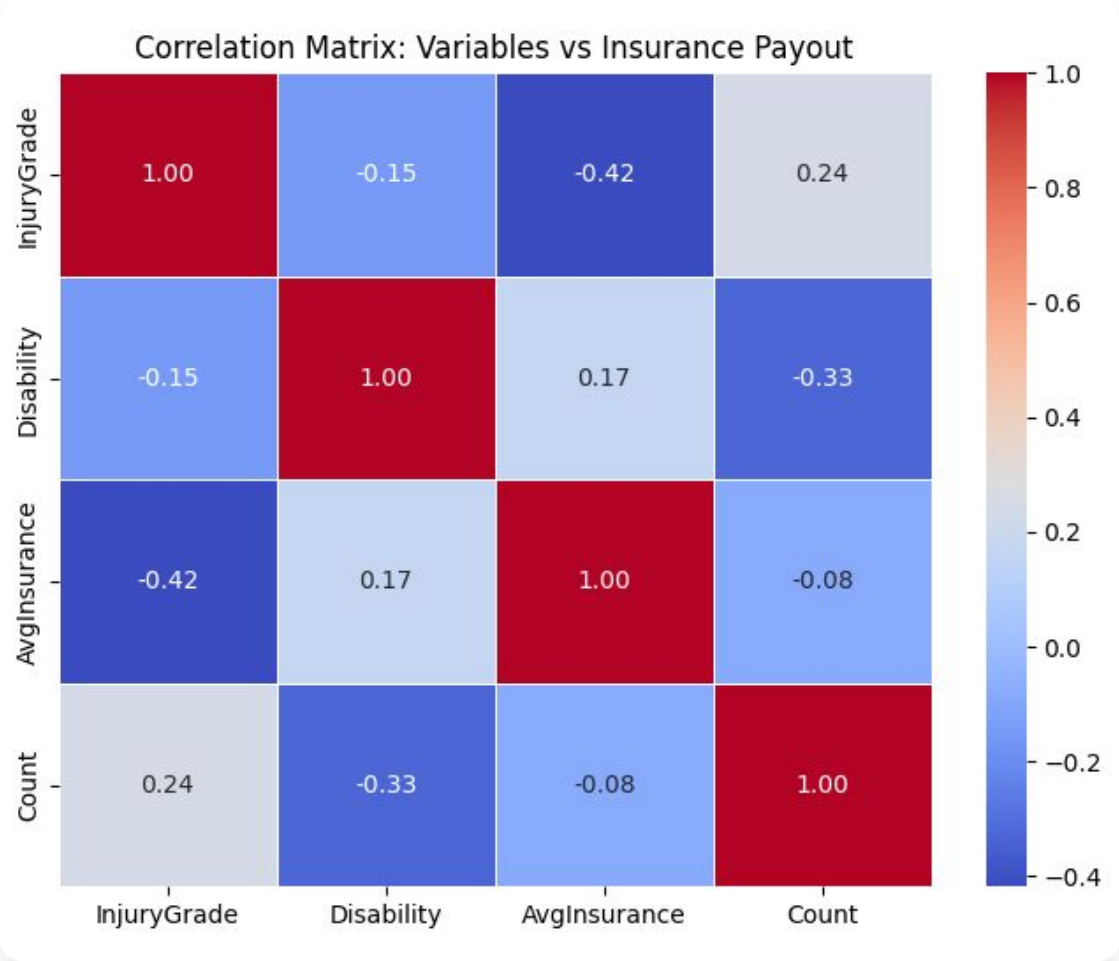
- **컬럼명 변경:** 분석 편의를 위해 영문 컬럼명으로 매핑
- **형 변환:** pd.to_numeric을 사용하여 숫자형 데이터 변환 (Errors='coerce')
- **고유 ID 생성:** 중복 제거를 위해 YearMonth + InjuryGrade + Disability 조합
- **중복 및 결측치 제거:** drop_duplicates 및 dropna 적용

최종 분석 대상 데이터: 총 619건

Analysis 1
상관관계 분석

부상 급수와 보험금 사이의 관계

상관관계 히트맵 (Correlation)



주요 발견

- **InjuryGrade vs AvgInsurance:** 뚜렷한 음의 상관관계가 관찰됩니다.
- **해석:** 부상 급수는 숫자가 작을수록(1급) 상해 정도가 심각함을 의미합니다.
- 따라서, 급수 숫자가 작아질수록(심각할수록) 보험금 지급액이 증가하는 논리적 결과를 확인했습니다.

Analysis 2
통계적 가설 검정

장애 유무에 따른 보험금 차이 (T-test)

T-test 결과 (T-Test Result)

가설 설정

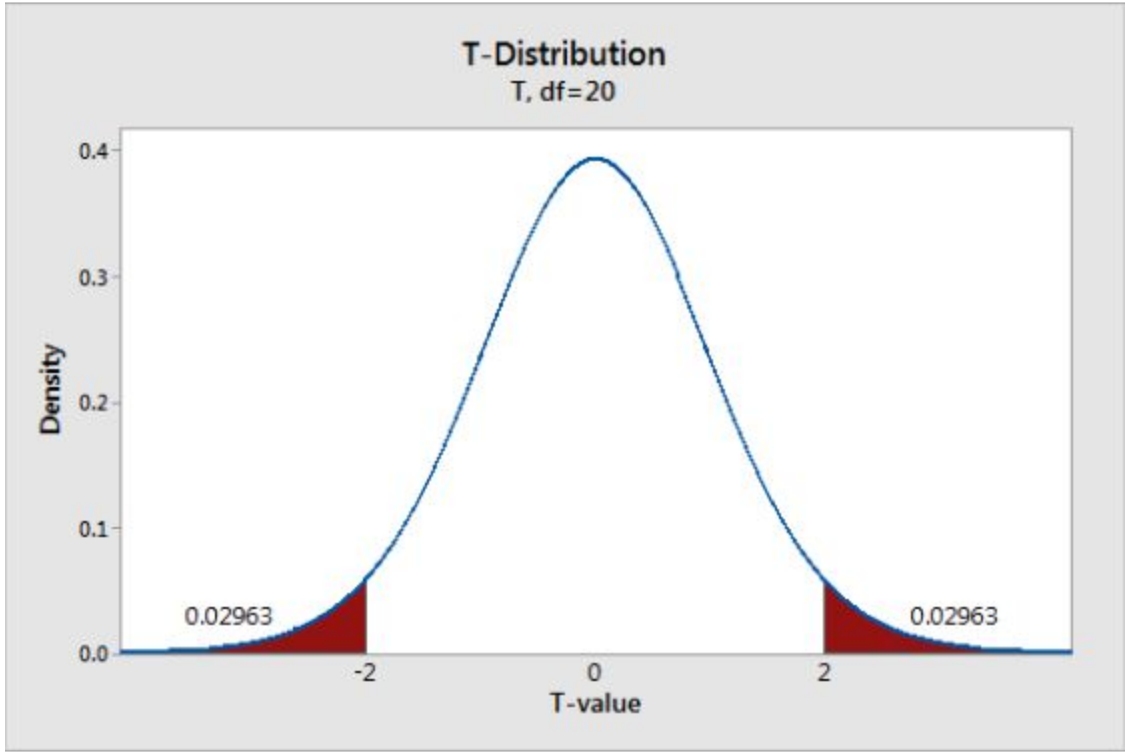
"장애가 있는 그룹의 보험금 평균은 없는 그룹과 유의미하게 다른가?"

검정 결과

T-statistic: 10.0628

P-value: 2.9571e-20 (< 0.05)

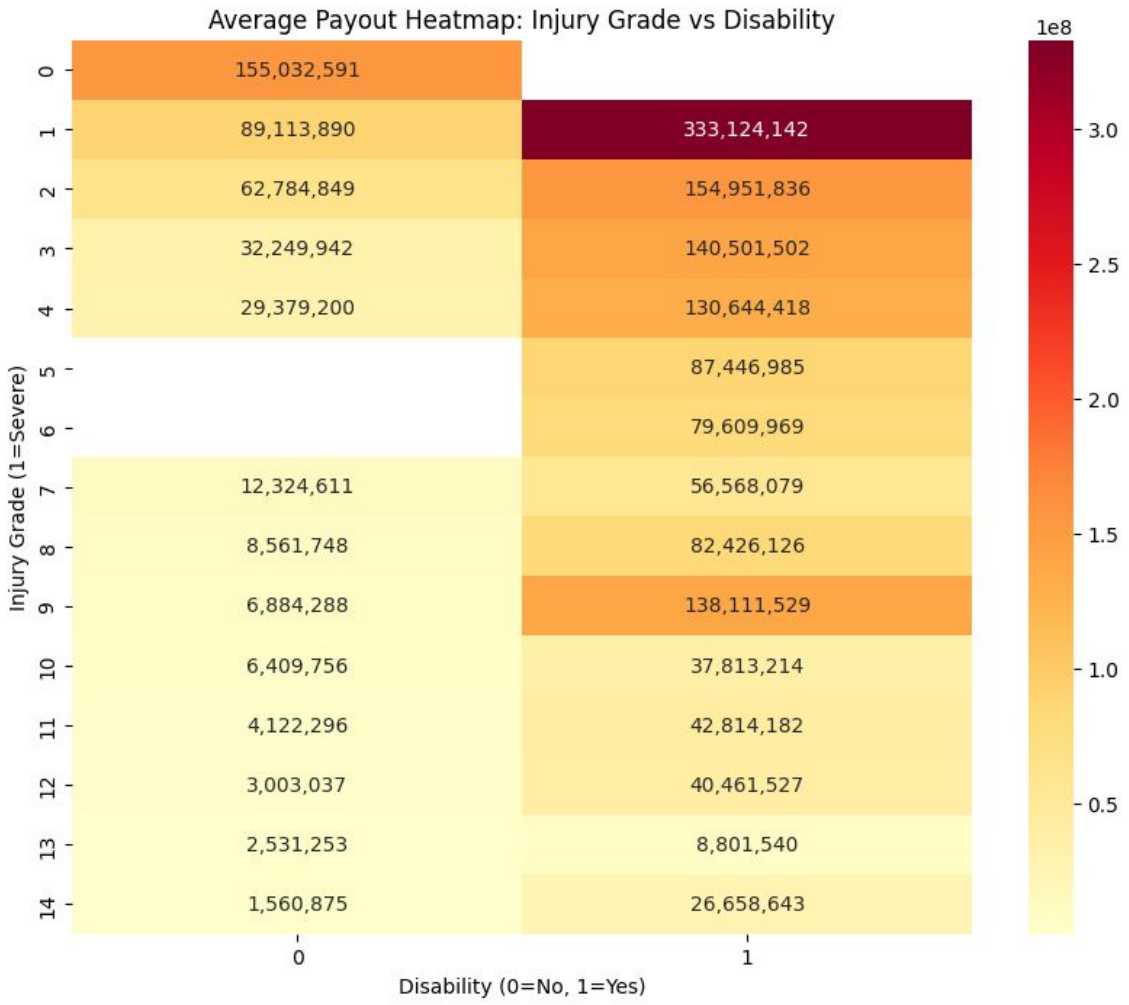
결론: P-value가 0.05보다 매우 작으므로, 두 그룹 간의 평균 보험금 차이는 **통계적으로 매우 유의미**합니다.



Analysis 3
복합 세그먼트 분석

부상급수 X 장애여부 교차 분석

평균 보험금 히트맵 (Pivot Heatmap)



인사이트 도출

- Y축: Injury Grade (1=중상)
- X축: Disability (0=무, 1=유)
- **결과:** 상위 급수(1~3급)이면서 장애가 있는 경우 (Disability=1) 가장 짙은 색(높은 보험금)을 보입니다.
- 단순 부상 급수뿐만 아니라, 장애 후유증 여부가 보상 규모에 큰 영향을 미침을 시각적으로 확인했습니다.

Analysis 4
시계열 추세 분석

2024년 상반기 월별 추이

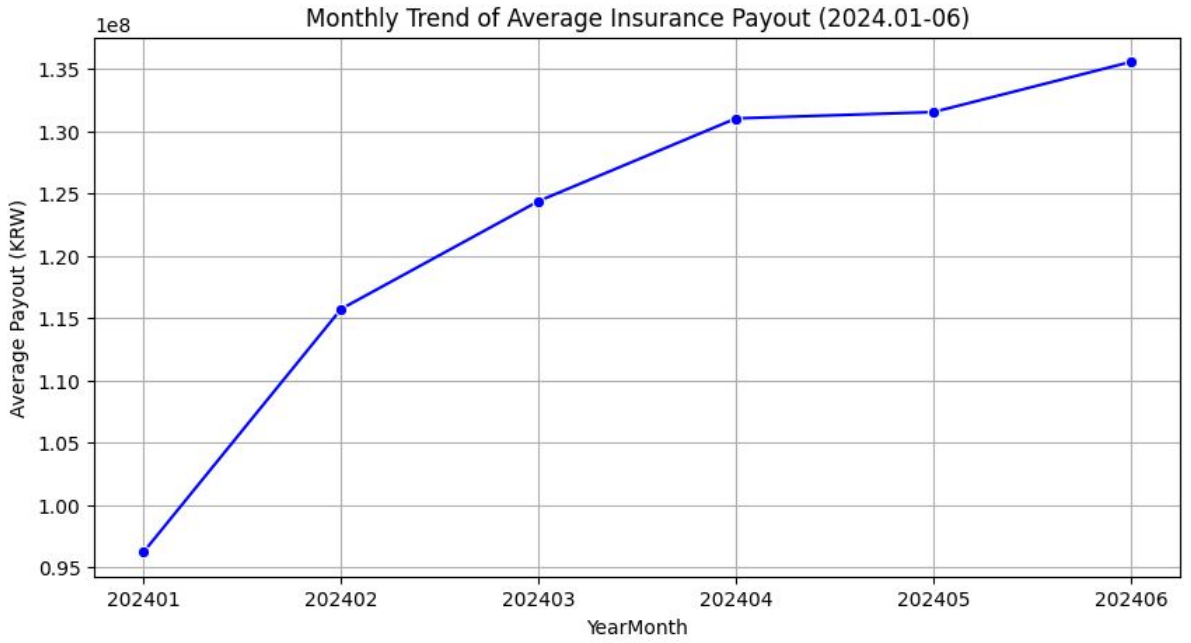
월별 평균 보험금 추이 (Monthly Trend)

변동성 확인

2024년 1월부터 6월까지의 평균 보험금 지급액 변화를 Line Chart로 시각화하였습니다.

- 특정 월에 급격한 변동이 있는지 확인하여, 계절성이나 특정 대형 사고의 영향을 유추해볼 수 있습니다.

- 전반적인 추세(상승/하락)를 통해 보험 재정 흐름을 파악할 수 있는 기초 자료가 됩니다.



최종 결론 (Conclusion)

데이터 분석 요약

- **부상 급수의 영향:** 부상 정도가 심할수록(낮은 급수) 보험금은 뚜렷하게 증가합니다.
- **장애의 영향:** 장애가 동반된 사고는 그렇지 않은 경우보다 통계적으로 유의미하게 높은 보상금을 받습니다. (P-value < 0.05)
- **복합 요인:** 고위험군(상위 급수 + 장애)에 대한 집중적인 관리 및 예산 예측이 필요함을 시사합니다.

본 프로젝트를 통해 공공 데이터를 활용한 데이터 수집, 전처리, 통계 분석 및 시각화의 전 과정을 수행하였습니다.