

Exploratory Data Analysis

- StudentID: 22100809
- Name: 황은지
- 1st Major: 사회복지
- 2nd Major: 데이터 사이언스

본 보고서는 사고 심각도 변수에 영향을 주는 주요 위험 요인을 파악하고, 향후 사고 예방 방안을 제안 하고자 함. Start_Date, Weather_condition, City 등의 변수를 사용하여 날씨조건과 사고 심각도의 관계를 분석하여, 사고 위험도 및 사고 건수가 높아지는 특정 날씨를 분석하고자 함.

1. Data overview

- 총 관측치 : 7,728,394의 데이터, 49개의 열
- 지역 범위: 미국을 기준으로 여러 주/도시를 대상으로 함.
- 주요 변수 설명:

1. 사고 정보

- **ID**: 각 사고를 식별하는 고유 ID (예: A-1, A-2 등)
- **Severity**: 1부터 4까지의 등급으로 나뉨. 숫자가 클수록 더 심각한 사고를 의미
- **Description**: 사고의 간단한 설명, 도로 이름과 사고 원인 등이 포함됨
- **Start_Time / End_Time**: 사고가 발생한 시간과 종료된 시간을 나타내는 타임스탬프
- **Distance(mi)**: 사고가 발생한 거리(마일 단위)로, 주로 짧은 거리를 기록

2. 위치정보

- **Start_Lat / Start_Lng**: 사고가 발생한 위치의 위도와 경도
- **Street / City / County / State**: 사고가 발생한 도로, 도시, 카운티, 주 정보

3. 날씨

- **Weather_Condition**: 사고 발생 시의 날씨 상태 (예: "Light Rain", "Foggy", "Overcast")
- **Temperature(F), Wind_Speed(mph), Visibility(mi)**: 사고 당시 세부적인 날씨 조건
- **Weather_Timestamp**: 사고 발생 시의 날씨 정보를 기록한 시간

4. 도로 및 교통 정보

- **Street / Junction / Traffic_Signal**: 사고가 발생한 도로의 특성을 나타냄
- **Amenity / Bump / Crossing**: 도로 주변의 특징을 나타내는 변수

2. Univariate analysis

아래와 같이 3가지의 변수를 활용함.
특별히 Severity 변수 활용 시 사고 심각도를 3, 4에 해당하는 값만 추출하여 높은 심각도에 영향을 주는 요인을 파악하고자 함.

2.1 weather condition

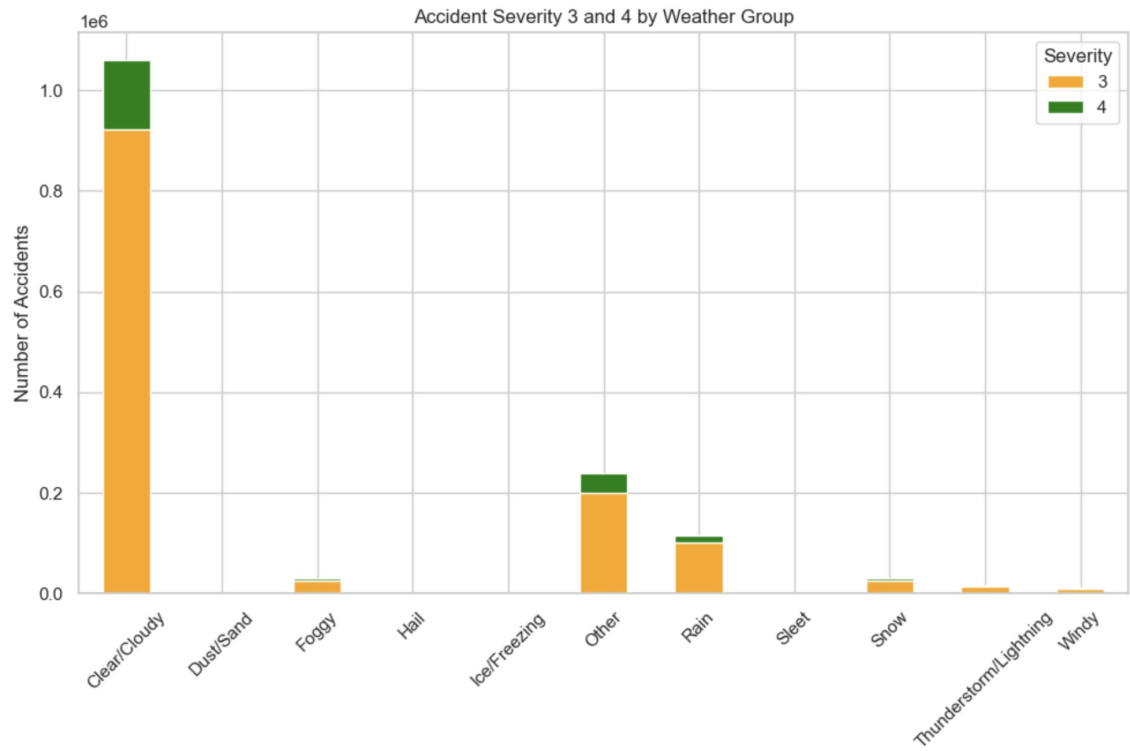


Figure 1. Accident severity 3 and 4 by Weather Group

Weather_Condtion 변수의 값을 총 11개의 날씨 요소로 그룹화하여 Weather_Group 이라는 변수를 새롭게 만듦. Figure 1에서 Clear/Cloudy 날씨일 때, 심각도 3과 4 수준의 사고건수가 가장 높은 것을 볼 수 있음. 순차적으로 Others, Rain, Snow 등이 사고건수가 높다고 파악됨.

- 정확한 수치:

Weather	Count	Severity 3	Severity 4
Clear/Cloudy	1,0611,291	921907	139384
Others	238,110	199608	38502
Rain	11,441	101515	12926
Snow	31,030	25238	5792
Foggy	29,989	25956	4033

Table 1. Count Accident based on Weather

전체적으로 Clear/Cloudy 날씨일 때, 사고 심각도 3 · 4 수준에서 사고가 자주 발생하며, 운전자들은 일상적인 운전 습관에 대한 경각심을 높일 필요가 있음.

2.2 계절별(여름/겨울) 날씨

2.1에서 Clear/Cloudy와 Other 날씨 조건에서 사고가 많이 발생함을 파악할 수 있었음.

2.2는 그룹화한 날씨 변수를 토대로, 사고 발생 날짜 중 여름(6,~8월)과 겨울(11~2월) 계절로 나누어 계절적 특성을 반영한 날씨 조건(Rain, snow 등) 중 발생하는 사고 건수를 살펴보고자 함.

Weather_Group	Count
Clear/Cloudy	1809841
Other	318166
Rain	108210
Thunderstorm/Lightning	48331
Foggy	38920
Windy	11055
Dust/Sand	130
Snow	81
Hail	20
Ice/Freezing	1

Weather_Group	Count
Clear/Cloudy	1959111
Other	562604
Rain	216123
Snow	126191
Foggy	104524
Windy	38393
Thunderstorm/Lightning	5494
Ice/Freezing	4921
Sleet	795
Dust/Sand	212
Hail	70

Table 2 & 3. Summer & Winter Weather frequency Table

- Table 2: 여름(6월, 7월, 8월) 계절적 특성을 반영한 날씨 조건에서 Rain이 가장 높고, Thunderstorm/Lightning, Foggy, Windy 에서 사고가 많이 발생함. 여름철에 비, 뇌우/번개 등이 사고 위험도를 높이는 요인임을 알 수 있음.
- Table 3: 겨울(11월, 12월, 1월, 2월) 계절적 특성을 반영한 날씨 조건에서 Rain이 가장 높고, Snow, Foggy, Windy 순으로 사고가 많이 발생함. 겨울철에 눈과 비가 동시에 발생할 경우 미끄러움과 관련된 사고 위험도가 높아 보임.

2.3 여름/겨울철 사고가 가장 많이 발생한 도시

2.2에서 언급한 여름/겨울에 발생하는 날씨 요건을 반영하고, 가장 많은 영향을 받은 국가가 무엇인지 알아보고자 함.

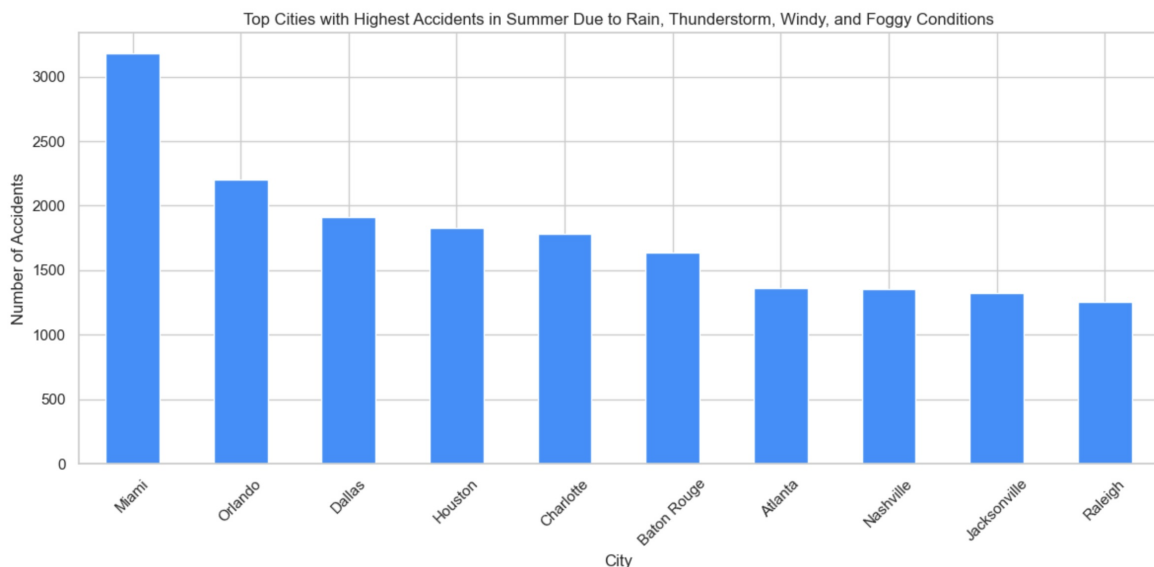


Figure 2. Top cities with Highest Accidents in Summer due to Rain, Thunderstorm/Lightning, Windy and Foggy

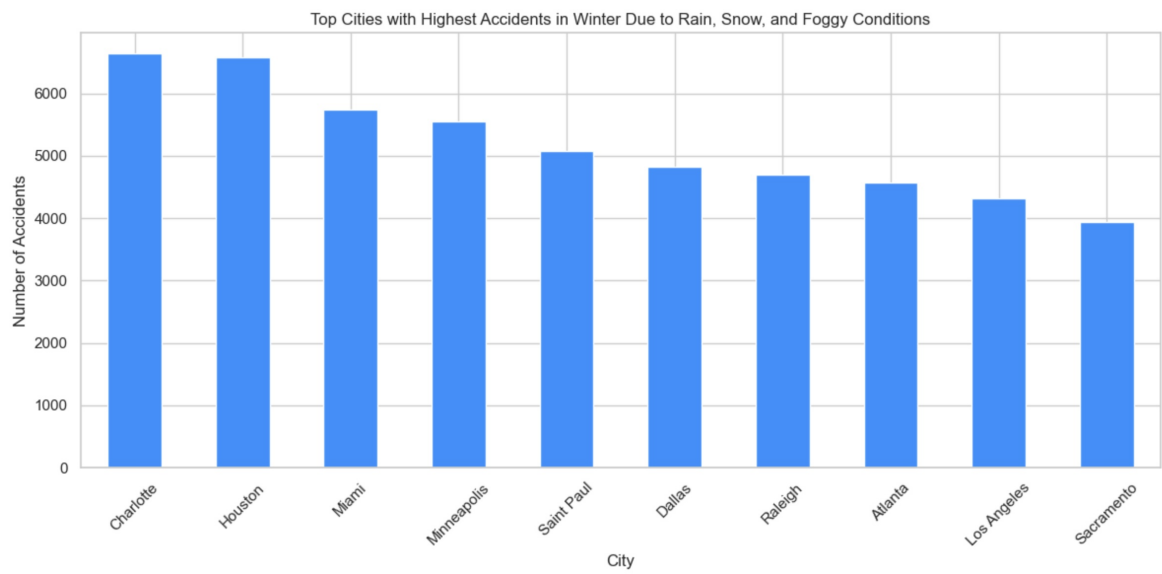


Figure 3. Top cities with Highest Accidents in Winter due to Rain, Snow, Windy and Foggy

- 여름: Miami(3179), Orlando(2203), Dallas(1909), Houston(1827), Charlotte(1779) 등의 국가들이 높은 사고 건수 기록.
- 겨울: Charlotte(6636), Houston(6585), Miami(5747), Minneapolis(5558), Saint Paul(5081) 등의 국가들이 높은 사고 건수를 기록.

각각 여름과 겨울철에 상위 10개에 속한 국가들에 대한 사고 예방 대책 마련 필요. 특별히 Miami의 경우 여름과 겨울 두 계절 모두 상위 3위 안에 속하였기에 운전자들은 안전 운전 주의 필요.

3. Multivariate analysis

3.1 Correlation

단변량 분석을 통해 파악한 주요 변수들이 다변량 분석을 통해 시간대와 계절별로 어떤 상호작용을 보이는지 추가로 분석하였음. 상관관계 분석을 위해 heatmap 시각화를 하였고, 진한 빨간색일 수록 빈도가 높고, 파란색에 가까울 수록 빈도가 낮음.

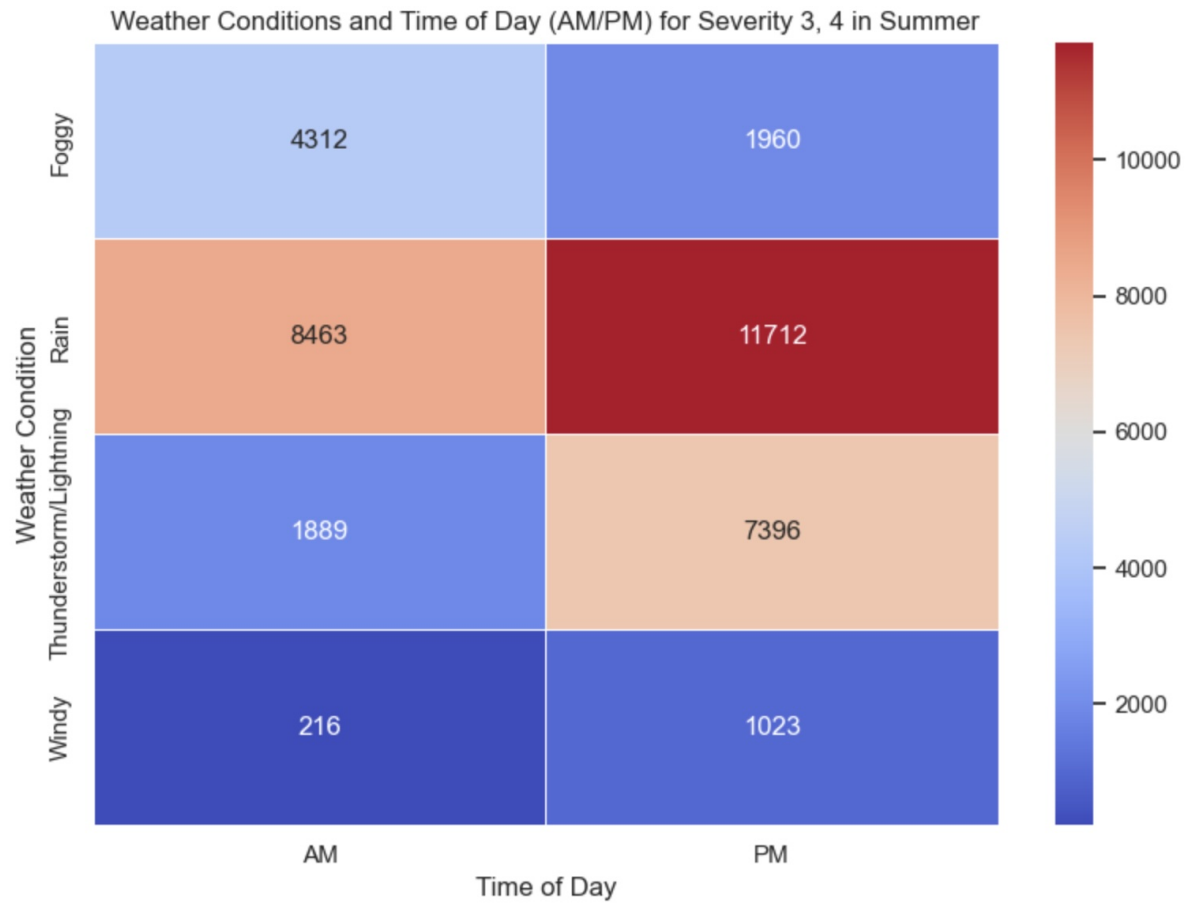


Figure 4. Weather Conditions and Time(AM/PM) for Severity 3,4 in Summer

여름철 날씨 조건별 사고 빈도

- Rain: 전체적으로 사고 빈도수가 가장 높음. 오후 시간대에 사고 빈도가 급증하며, 교통혼잡과 날씨의 복합적 영향이 있어 보임.
- Thunderstorm/Lightning: 번개 동반한 폭우가 오후 시간대에 있어. 사고 빈도 증가에 기여하는 것으로 파악됨.
- Foggy: 오전시간대에 안개가 시야를 제한하여 사고 위험 증가로 예상됨.
- Windy: 다른 날씨 조건보다 영향이 적으나 오후 시간대에 소폭 증가함.

시간대별 사고 빈도

- 전반적으로 오후 시간대에 사고 빈도가 더 높음.
- 출퇴근 시간대 및 교통량 증가가 영향을 미침.

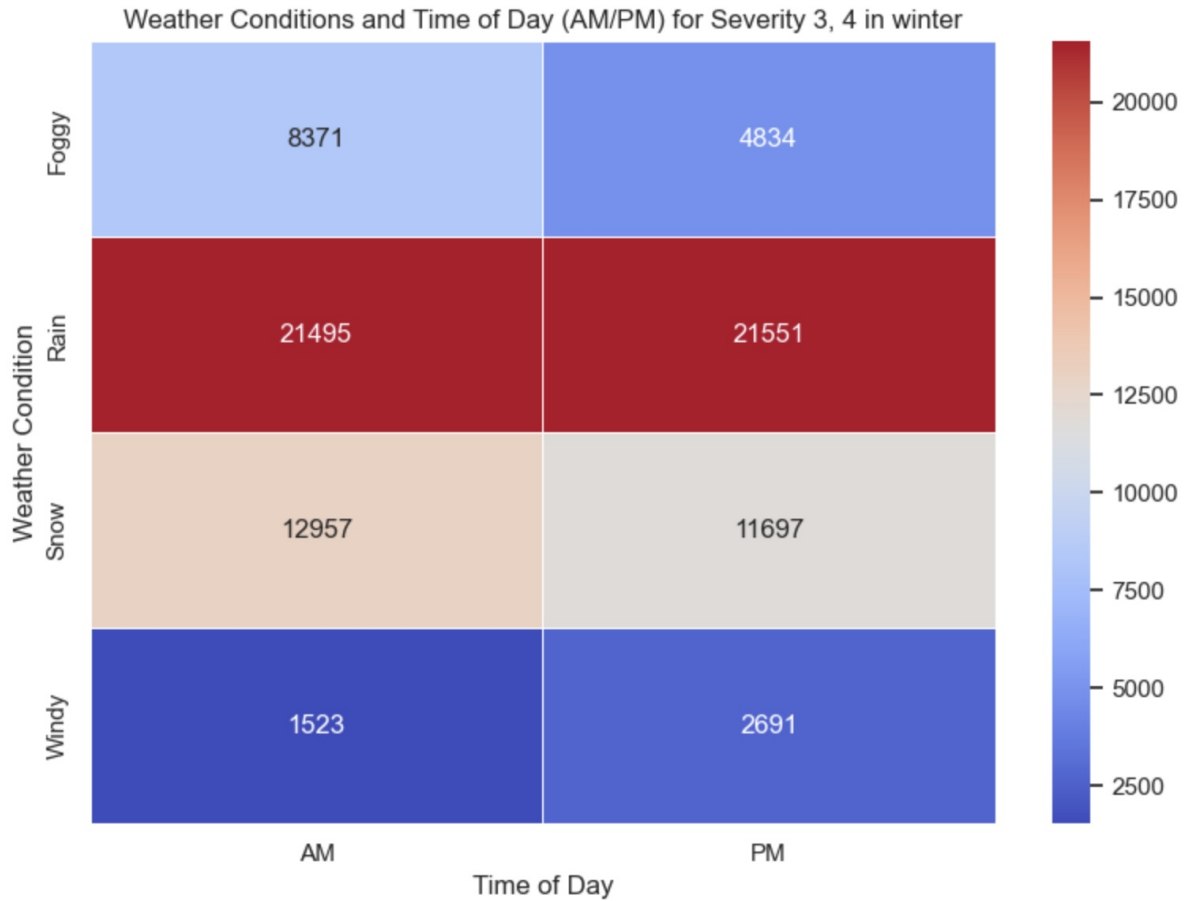


Figure 5. Weather Conditions and Time(AM/PM) for Severity 3,4 in Winter

겨울철 날씨 조건별 사고 빈도

- Rain: 전체적으로 사고 빈도수가 가장 높음. 오전, 오후 둘 다 사고 건수가 높아 하루 종일 교통사고 위험도가 높음.
- Snow: 사고 건수가 오전, 오후 큰 차이는 없지만 오전이 더 높음. 눈이 내리는 아침 시간대에 사고 위험이 더 높음.
- Foggy: 아침 시간대에 더 많은 사고가 발생함. 시야 확보 문제로 사고가 빈번하게 일어날 수 있음.
- Windy: 오후 시간대에 사고가 더 많이 발생함. 오후에 바람이 강해지며 사고의 위험도가 높아질 수 있음.

4. Suggestion

4.1 계절별 사고 안전 대책

- 여름/겨울에 따라 달라지는 기상 조건, 사고 발생 패턴 분석을 토대로 맞춤형 교통 안전 대책 수립 필요함.
- 여름철 대책: 비, 천둥/뇌우 등으로 인한 교통사고 사건 발생 수가 높다는 결과가 나옴. 사고 방지를 위해 도로 배수 시스템 강화, 운전 경고판 미 설치 장소 파악 및 설치 등 실시함. 천둥번개 시 차량 운행 속도를 낮추는 표지판을 다량 설치하는 방안을 고려해야 함.
- 겨울철 대책: 눈/비 등으로 인해 교통사고 발생 건수가 높다는 결과가 나옴. 사고 방지를 위해 도로 미끄러움 방지를 위한 체계적 제설작업 실시 필요함.

4.2 특정 계절에 취약한 지역 집중 관리

- 계절별로 특정 도시들이 날씨에 따른 교통사고 건수가 높다는 결과를 도출함. 사고 방지를 위해 취약 지역에 대한 집중 관리 필요함.

- 여름철 취약 도시: 여름철 비로 인한 사고 발생률이 높은 Miami의 경우, 폭우 예보 시 경고 시스템을 도입하고, 비가 많이 오는 도로 구간에 대한 도로 관리 강화 대책을 마련해야 함. 6월 전에 배수 시스템, 방수 시스템 등에 대한 의무적 점검 및 보고 체계 확립 필요함.
- 겨울철 취약 도시: 겨울에 가장 사고 건수가 높은 Charlotte의 경우, 폭설·비 발생 시, 날씨 방송 및라디오 송출 확대하고, 11월 전에 제설 장비 확충 해야 함. 사고 대비를 위한 개인 자동차 정비 도구 안내 메뉴얼 배포 고려함.

4.3 시간대별 대응

- 시간대(오전/오후)에 따라 사고 발생 위험이 높아지는 경우가 있음을 도출함. 기상 상황 데이터 기반으로 맞춤형 안전 대책 수립 필요함.
- AM 시간대: 눈과 안개 등으로 아침 시간대 사고 발생률이 높음. 이에 대비하여 주요 도로 시야 확보 시스템(안개등, 도로 시야 보조 장치 등) 개선해야 함. 도로 상황을 실시간으로 전파하는 방송을 확대해야 함.
- PM 시간대: 비와 바람 등으로 인해 오후 시간대 사고 에 대비하여, 오후 운전 경고발송 시스템 (네비게이션 안내, 라디오 방송 등) 강화해야 함.
- 교통량 분산: 아침, 오후 등의 교통 혼잡도를 줄이기 위해 교차로 신호 체계를 시간대별로 조정 필요함. 기상 조건에 따른 교통량 분산 체계 확립 필요함.