THE CAUSALITY BETWEEN ROAD ACCIDENTS AND WEATHER CONDITIONS

A RELATIONAL DATABASE APPROACH FOR ENHANCED INSIGHTS

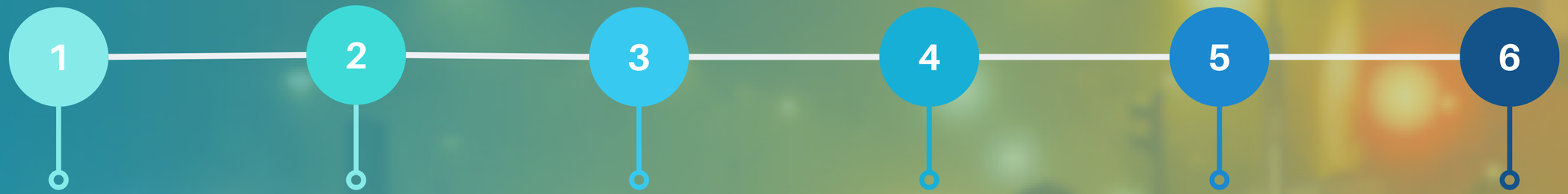
Presented By:

Khajjou Yasmine

Sestito Alfredo

Tadesse Amen Sime

PROJECT OVERVIEW



1. DATA ACQUISITION
2. DATA INTEGRATION AND ENRICHMENT
3. DATABASE STRUCTURE
4. DATA QUALITY
5. CONCLUSION
6. FUTURE WORK

DATA ACQUISITION

Accidents Data

kaggle

- Accident Date: Date of each accident occurrence, providing temporal context crucial for analyzing trends and patterns.
- Light Conditions: Lighting conditions at the time of the accident, influencing visibility and potentially contributing to accident severity.
- Longitude and Latitude: Geographical coordinates pinpointing the location of each accident, essential for spatial analysis and hotspot identification.
- Accident Fatality: Denotes whether the accident resulted in fatalities, a critical metric for assessing accident severity.
- Number of Deceased Individuals: Provides the count of individuals who lost their lives in each accident, aiding in casualty assessment.
- Number of Injured Individuals: Indicates the number of individuals injured in each accident, offering insights into the magnitude of physical harm.
- District: Specifies the district or administrative region where each accident oc- curred, enabling regional analysis and policy targeting.

Narrowed to only Kent and East London.

CSV

DATA ACQUISITION

Weather Dataset

- maxtempC: The maximum temperature recorded during the day, in degrees Cel- sius.
- mintempC: The minimum temperature recorded during the day, in degrees Cel- sius.
- totalSnow: The total amount of snowfall measured in centimeters during the spec- ified time period.
- sunHour: The total hours of sunlight received during the day.
- uvIndex: A measure of the strength of ultraviolet radiation from the sun, indicating the potential risk of harm to exposed skin.
- moon illumination: The percentage of the moon’s surface illuminated by sunlight.
- moonrise: The time at which the moon rises above the horizon.
- moonset: The time at which the moon sets below the horizon.
- sunrise: The time at which the sun rises above the horizon.
- • sunset: The time at which the sun sets below the horizon.
- DewPointC: The temperature at which air becomes saturated with moisture, mea- sured in degrees Celsius.
- FeelsLikeC: The perceived temperature, which factors in humidity and wind con- ditions, expressed in degrees Celsius.
- HeatIndexC: A measure of how hot it feels when relative humidity is factored in with the actual air temperature, in degrees Celsius.

Data Integration and Enrichment

Dataset configuration:

Primary dataset on accidents in Kent region in 2022, augmented with 12 distinct weather datasets for each month.

Integration:

Unified entity created for comprehensive weather conditions corresponding to accident locations.

Refinement:

Removal of irrelevant attributes: "sunHour," "uvIndex," "moonillumination," etc.
Addressing type heterogeneity: Reconstructed "location" attribute to resolve mismatches.
Attribute optimization: Removal of "Weather Conditions" attribute due to redundant data.

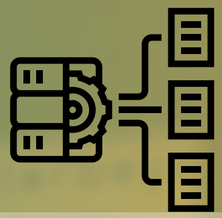
Attribute Optimization:

Removal of redundant attributes like "Weather Conditions" to enhance dataset robustness.

Dataset Integration:

Unified entity created for comprehensive weather conditions.

DATABASE STRUCTURE



Rationale for Relational Database:

Analytical Flexibility: Supports various analytical tasks, including quantitative analysis and data mining.

Focus on Read Operations: Prioritizes efficient data retrieval and processing.

Structured Data: Aligns well with the structured nature of dataset entities.

Database Structure Design:

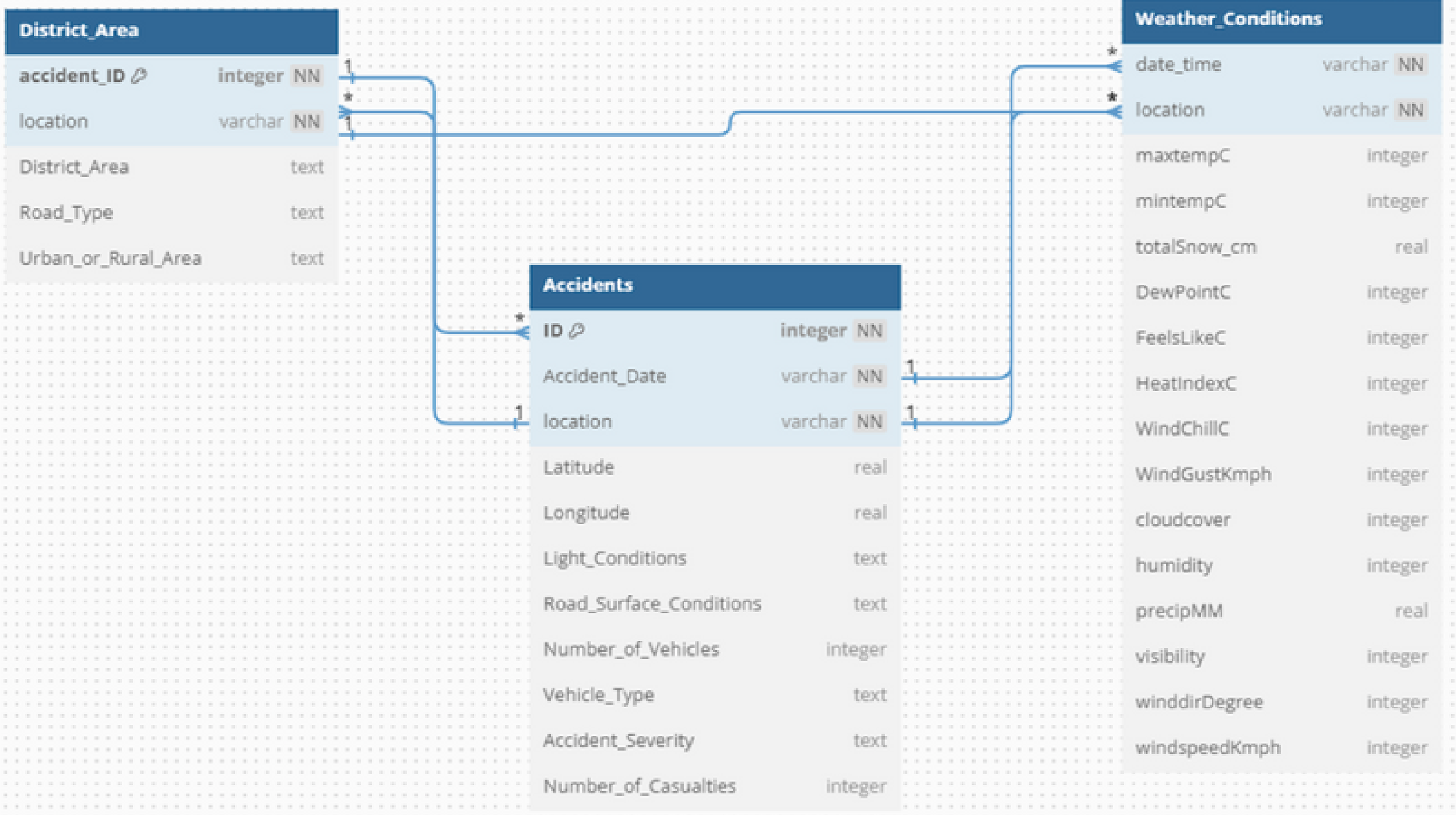
Three distinct entities: Accidents, Weather Conditions, and District Area.

Accidents: Represents accident dynamics with unique IDs, date, and location.

District Area: Geographical position with coordinates, district, road type, and urban/rural classification.

Weather Conditions: Detailed weather data including temperature, humidity, and precipitation.

One-to-Many paradigm for deeper insights and understanding.



DataBase Schema

Possible Queries

- **The Comparative Analysis of Urban and Rural Area Accidents: :Percentage Distribution Relative to Total Accidents:**

```
SELECT Urban_or_Rural_Area, count(*) as Accident_Count, round((count(*)*100)/
(SELECT count(*) from District_Area), 2) as Percentage
FROM District_Area
GROUP by Urban_or_Rural_Area
;
```

Urban_or_Rural_Area	Accident_Count	Percentage
Rural	2939	55.0
Urban	2317	44.0

Figure 4.2: Output - Table Display of The Query

- District Areas and Causalities Count with High-Speed Wind and Low Temperature:

```
SELECT District, sum(Number_of_Casualties)
FROM Accidents join District_Area on Accidents.location = District_Area.location
where Accidents.location in (SELECT location
FROM Weather_Conditions
WHERE FeelsLikeC <= 5 AND windspeedKmph >= 20)
GROUP by District
ORDER by Number_of_Casualties DESC
;
```

District	sum(Number_of_Casualties)
Tunbridge Wells	46
Suffolk Coastal	2
Wealden	10
Thurrock	24
Tandridge	4
Swale	32
Shepway	1
Sevenoaks	25
Rother	4
Medway	94
Maidstone	66
Lewisham	22
Greenwich	71
Gravesham	28
Gateshead	2
Dartford	25
Canterbury	2
Bromley	45
Bexley	44
Ashford	49

Figure 4.3: Output - Table Display of The Query

COMPLETENESS CHECK

Before Integration

KENT_2022_ACCIDENTS

- 92.5% completness
- Duplicate Records: None

DISTRICT AREA

- 94.5% completeness
- Duplicate Records: None

WEATHER CONDITIONS

- 100% completeness
- Duplicate Records: None

After Integration

KENT_2022_ACCIDENTS

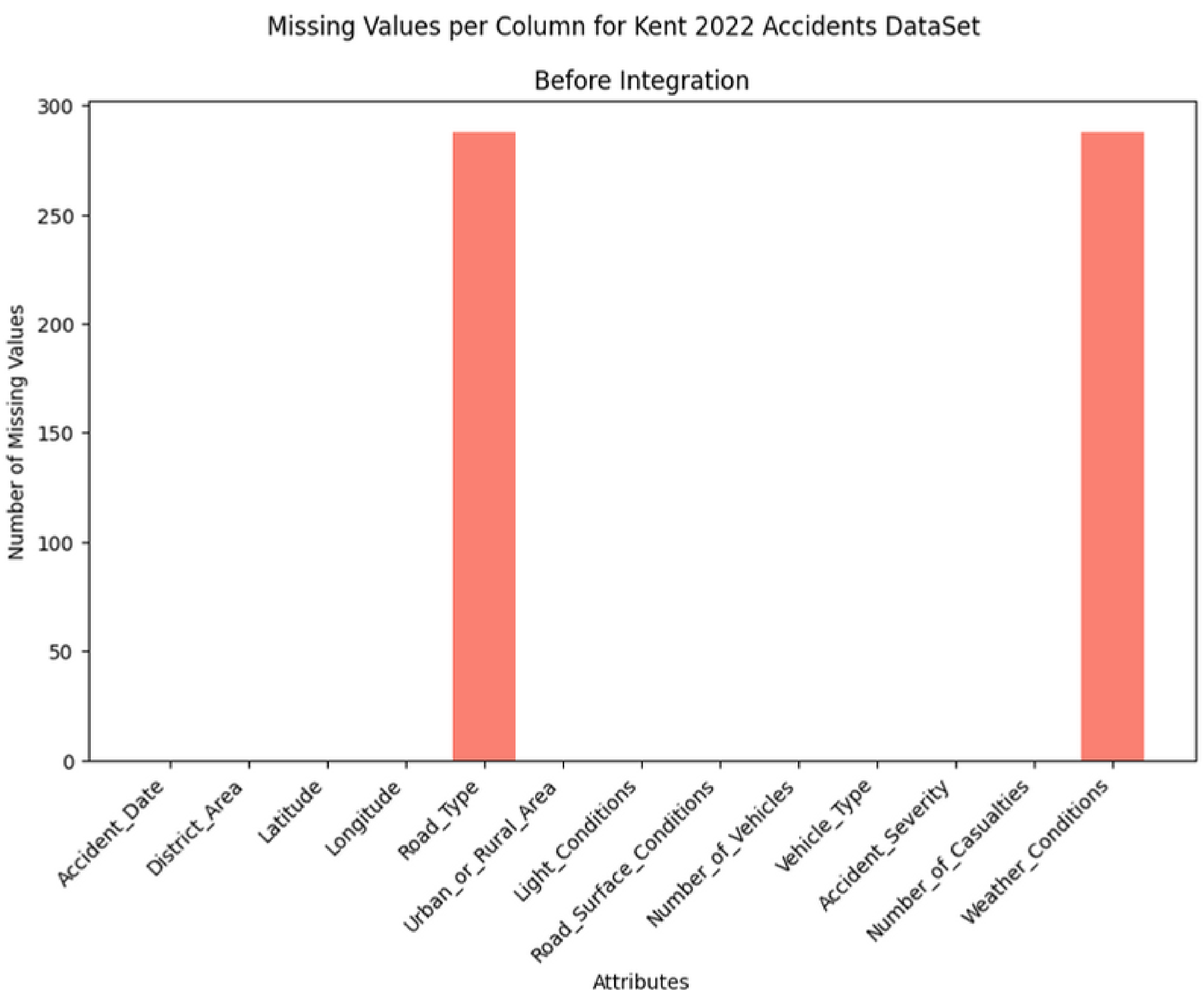
- 100% completeness
- Duplicate Records: None

DISTRICT AREA

- 94.5% completeness
- Duplicate Records: None

WEATHER CONDITIONS

- 100% completeness
- Duplicate Records: None


Missing Values per Column for Kent 2022 Accidents DataSet
Before Integration
Number of Missing Values
300
250
200
150
100
50
0
Accident_Date
District_Area
Latitude
Longitude
Road_Type
Urban_or_Rural_Area
Light_Conditions
Road_Surface_Conditions
Number_of_Vehicles
Vehicle_Type
Accident_Severity
Number_of_Casualties
Weather_Conditions
Attributes

CONSISTENCY CHECK

Before Integration

KENT_2022_ACCIDENTS

- Date Format Check: True
- Latitude Range Check: True
- Longitude Range Check: True

DISTRICT AREA

- Date Format Check: True

WEATHER CONDITIONS

- Date Format Check: True

After Integration

KENT_2022_ACCIDENTS

- Date Format Check: True
- Latitude Range Check: True
- Longitude Range Check: True

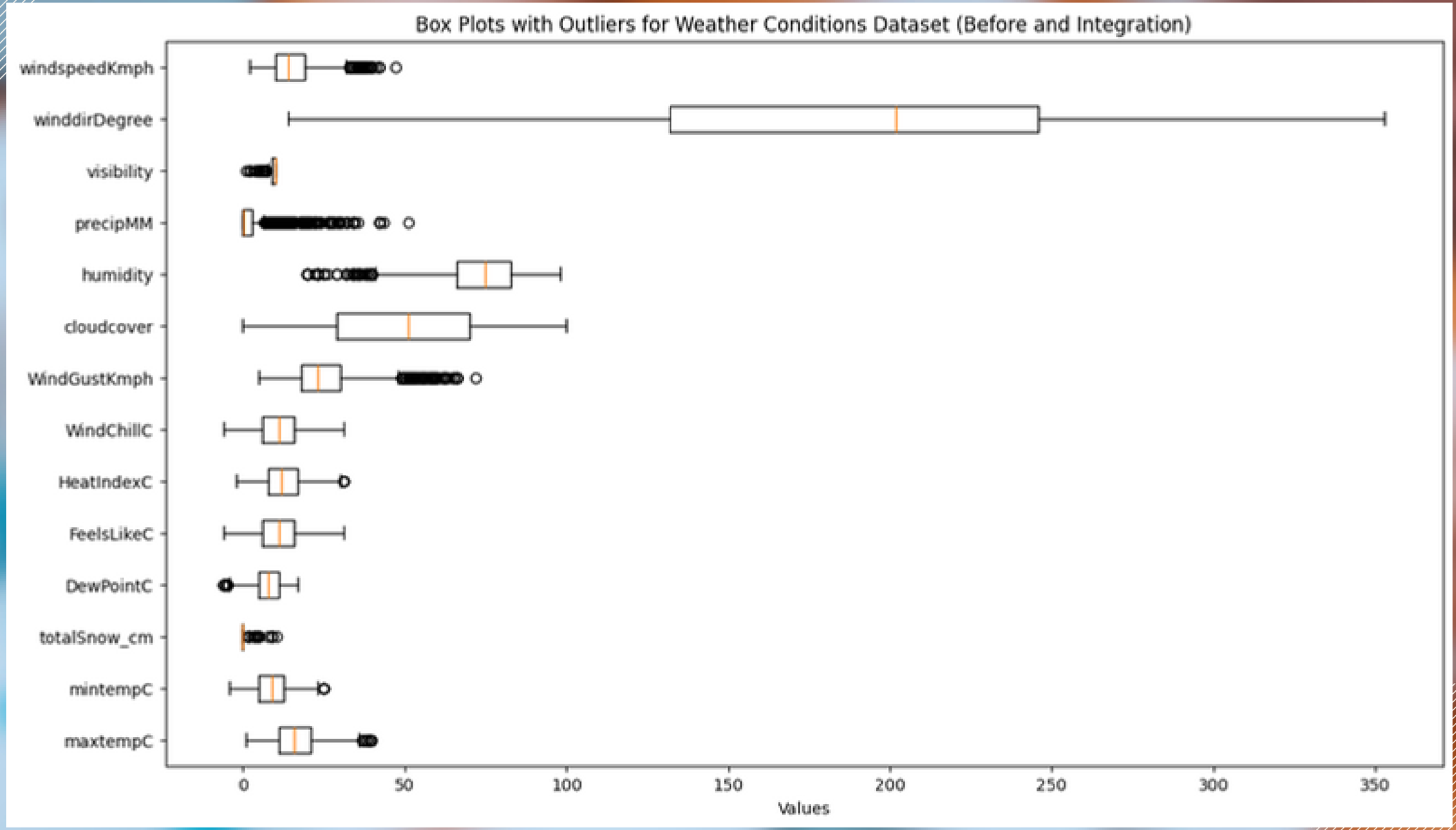
DISTRICT AREA

- Date Format Check: True

WEATHER CONDITIONS

- Date Format Check: True

OUTLINER CHECKS



CONCLUSION

- Successfully implemented a relational one-to-many database integrating accident dataset from Kaggle with meteorological insights via weather API.

- Merged 12 weather datasets into a unified entity for 2022 in Kent, optimizing attributes and addressing type heterogeneity conflicts.

- Rigorous data validation process ensured statistical accuracy leveraging relational database model's flexibility and efficiency.

- Three distinct entities established: Accidents, District Area, and Weather Conditions, laying a solid foundation for comprehensive analysis.

FUTURE WORK

- EXPLORE TEMPORAL AND SPATIAL ANALYSIS
- PREDICTIVE MODELING
- INTEGRATE ADDITIONAL DATA SOURCES, REAL-TIME DATA INTEGRATION,