Examining the evolving trends and connection between road accidents and the average temperature in Germany.

The link between environmental conditions and road safety has consistently drawn attention. In this report, I explore how the average temperature in Germany from 2018 to 2020 correlates with the number of accidents on motorways inside and outside built-up areas. Uncovering this relationship can offer crucial information for creating measures to improve road safety and lessen the effects of harsh weather.

Question

1. The changes of the temparature?
2. How the rate of road accident increasing day by day?
3. Is there any realtion between road accident and the temparature?

Data Sources

1. Datasource: Federal Statistical Office of Germany

    - Metadata URL: https://www.destatis.de/
    - Data URL: https://www-genesis.destatis.de/genesis/online?language=en&sequenz=statistikTabellen&selectionname=46241#abreadcrumb
    - Data Type: CSV
    - Description: The road accident data in Germany. This data also shows that the accident in Indoor and outdoor as well.
    - License Type: OpenData License
2. Datasource: climate_environment/CDC/

    - Metadata URL: https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/

    - Data URL: https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_01.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_02.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_03.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_04.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_05.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_06.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_07.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_08.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_09.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_10.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_11.txt https://opendata.dwd.de/climate_environment/CDC/regional_averages_DE/monthly/air_temperature_mean/regional_averages_tm_12.txt

    - Data Type: txt
    - License Type: OpenData License
    - Description: This is monthly average air temparature in Germany. In this dataset it also shows that the temparature in state wise.

Data Pipeline

The data pipeline involved several crucial steps to prepare the datasets for analysis. It addressed missing or incorrect entries in both the temperature and accident datasets by using imputation techniques and rigorous data validation. Temporal alignment of the data was crucial for accurately correlating temperature fluctuations with accident frequency.

Technologies Used: Python,Pandas,SQLite,SQLAlchemy

Data Cleaning/Transformation Steps:
Missing temperature data were filled using interpolation methods. Outliers in the accident data were identified and removed to avoid distorted results. Additionally, the data was normalized to compensate for differences in reporting standards.

Problems Encountered: While working on the data engineering process, one of the main challenges was dealing with varying accident reporting standards across different regions. Additionally, ensuring consistent time resolution in both datasets was crucial for establishing a meaningful correlation. These challenges were addressed through meticulous data validation and cleaning procedures

	Year	Month	Date	Brandenburg/Berlin	Brandenburg	Baden-Wuerttemberg	Bayern	Hessen	Mecklenburg-Vorpommern	Niedersachsen	N
0	2018	January	January-2018	3.31	3.31	4.33	3.14	4.10	2.87	3.90	
1	2018	February	February-2018	-1.64	-1.66	-2.09	-2.96	-1.99	-1.47	-1.02	
2	2018	March	March-2018	1.40	1.39	3.19	2.24	2.83	0.76	2.62	
3	2018	April	April-2018	12.92	12.90	12.50	12.52	12.63	10.81	12.02	
4	2018	May	May-2018	17.18	17.16	15.23	15.66	15.87	15.76	16.63	

Fig: Average temperature data

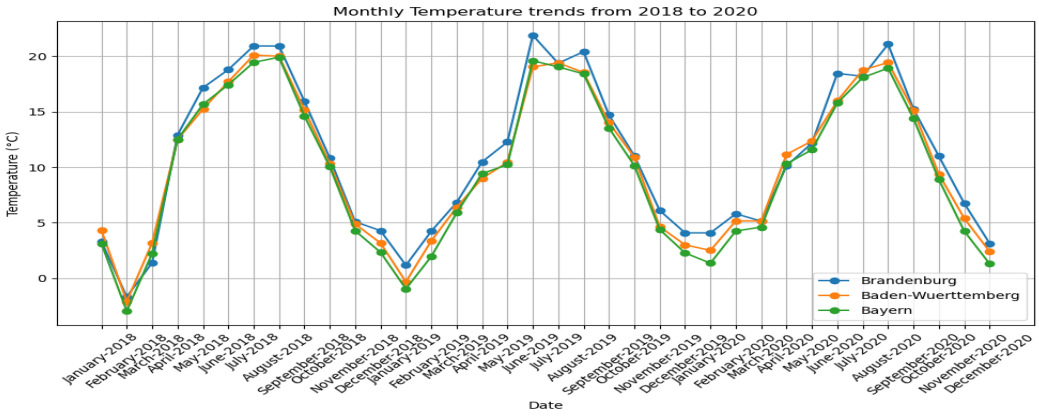

Fig: Monthly Temperature trends from 2018 to 2020

Hamburg	Hessen	Mecklenburg-Vorpommern	Niedersachsen	Nordrhein-Westfalen	Rheinland-Pfalz	Saarland	Sachsen	Sachsen-Anhalt	Schleswig-Holstein	Thüringen
5243	11627	4412	17092	53635	11714	2952	8447	5551	6879	4558
5013	10766	3638	14940	50510	10621	2511	6961	4855	6675	4023
5527	11363	4499	16214	54837	11529	2892	8950	6363	7106	4644
5831	12563	4623	17879	56248	12271	2994	9412	6347	7019	4792
5901	12723	5424	19160	55686	12540	2928	9582	6510	8575	5194

Fig: Road accident data in Germany

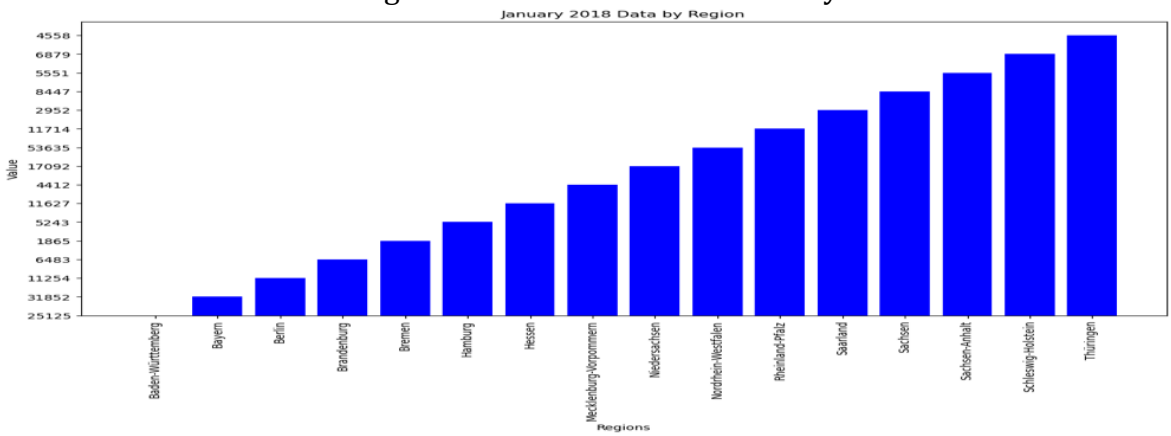

Fig: January 2018 Data by Region

Result and Limitations

The plot clearly shows a typical temperature progression with increases from January to July, followed by a decrease towards September. This indicates normal seasonal variations.There are notable differences in temperature trends between the regions. Baden-Wuerttemberg and Bayern generally experienced higher peaks and lower troughs than Brandenburg, indicating a slightly wider range of temperatures.

The data collected across various regions in Germany for the year 2018 displays distinct seasonal temperature variations, highlighting differences in climatic conditions across regions.The data demonstrates a typical progression of temperatures increasing from January through July, followed by a subsequent decrease towards September.

SQL allows for efficient querying, manipulation, and management of structured data in relational databases.

Resolution: Monthly averages are used, which can mask daily or weekly variations that might be critical for more detailed climate analysis or impact assessments. Region Specificity: The data aggregates temperatures by large regions, which may not accurately reflect local microclimates. For instance, urban areas might have higher temperatures due to the heat island effect