

# 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 temperature?
2. How the rate of road accident increasing day by day?
3. Is there any realtion between road accident and the temperature?

## Used Data

Average Monthly Temperature Data in Germany:

- This dataset contains the average monthly temperatures for various states in Germany from January 2018 to December 2020.
- Data Type: txt
- License Type: OpenData License

	Year	Month	Date	Brandenburg/Berlin	Brandenburg	Baden-Wuerttemberg	Bayern	Hessen	Mecklenburg-Vorpommern	Niedersachsen
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

```
avg_temp_df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 19 columns):
 #   Column              Non-Null Count  Dtype
---  -
 0   Year                36 non-null    int64
 1   Month               36 non-null    object
 2   Date               36 non-null    object
 3   Brandenburg/Berlin  36 non-null    float64
 4   Brandenburg         36 non-null    float64
 5   Baden-Wuerttemberg  36 non-null    float64
 6   Bayern              36 non-null    float64
 7   Hessen              36 non-null    float64
 8   Mecklenburg-Vorpommern  36 non-null    float64
 9   Niedersachsen       36 non-null    float64
10  Niedersachsen/Hamburg/Bremen  36 non-null    float64
11  Nordrhein-Westfalen  36 non-null    float64
12  Rheinland-Pfalz     36 non-null    float64
13  Schleswig-Holstein  36 non-null    float64
14  Saarland            36 non-null    float64
15  Sachsen              36 non-null    float64
16  Sachsen-Anhalt      36 non-null    float64
17  Thuringen/Sachsen-Anhalt  36 non-null    float64
18  Thuringen           36 non-null    float64
dtypes: float64(16), int64(1), object(2)
```

Fig2:Data Structure

Fig 1: Average temperature data

Road Accident Data in Germany:

- This dataset contains the number of road accidents recorded by the police in various states in Germany from January 2018 to December 2020. This data also shows that the accident in Indoor and outdoor as well.
- Data Type: CSV
- License Type: OpenData License

	Year	Month	Baden-Württemberg	Bayern	Berlin	Brandenburg	Bremen	Hamburg	Hessen	Mecklenburg-Vorpommern	Niedersachsen	Nordrhein-Westfalen	Rheinland-Pfalz	Saarland
0	2018	January	25125	31852	11254	6483	1865	5243	11627	4412	17092	53635	11714	2952
1	2018	February	23267	28969	10427	5428	1796	5013	10766	3638	14940	50510	10621	2511
2	2018	March	26471	32690	11640	6800	1953	5527	11363	4499	16214	54837	11529	2892
3	2018	April	28399	35224	12856	7273	2108	5831	12563	4623	17879	56248	12271	2994
4	2018	May	27337	35379	13075	7784	2223	5901	12723	5424	19160	55686	12540	2928
5	2018	June	28526	35606	12365	7062	2104	5776	12370	4866	17518	55224	12011	2757
6	2018	July	29877	38684	11339	6809	1800	5295	11701	5491	16331	55717	11219	2526
7	2018	August	23031	32697	11643	7009	2215	5882	13445	5197	18304	49458	12964	2876
8	2018	September	27080	35203	12081	6999	2222	5824	12611	4930	17805	55779	12380	2971
9	2018	October	30282	37343	12869	7198	2165	5927	13114	5170	19120	58535	13032	2986
10	2018	November	27698	33833	12860	7098	2092	5929	12719	4755	17952	60557	12498	3005

Fig 3: Road accident data in Germany

```
acc_data_df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 36 entries, 0 to 35
Data columns (total 18 columns):
#   Column              Non-Null Count  Dtype  
---  -
0   Year                 36 non-null    int64  
1   Month                36 non-null    object  
2   Baden-Württemberg    36 non-null    object  
3   Bayern               36 non-null    object  
4   Berlin               36 non-null    object  
5   Brandenburg           36 non-null    object  
6   Bremen               36 non-null    object  
7   Hamburg              36 non-null    object  
8   Hessen               36 non-null    object  
9   Mecklenburg-Vorpommern 36 non-null    object  
10  Niedersachsen         36 non-null    object  
11  Nordrhein-Westfalen   36 non-null    object  
12  Rheinland-Pfalz       36 non-null    object  
13  Saarland              36 non-null    object  
14  Sachsen               36 non-null    object  
15  Sachsen-Anhalt        36 non-null    object  
16  Schleswig-Holstein    36 non-null    object  
17  Thüringen             36 non-null    object  
dtypes: int64(1), object(17)
memory usage: 5.7+ KB
```

Fig 4: Data Structure

## Analysis:

### Data Pipeline Overview

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:

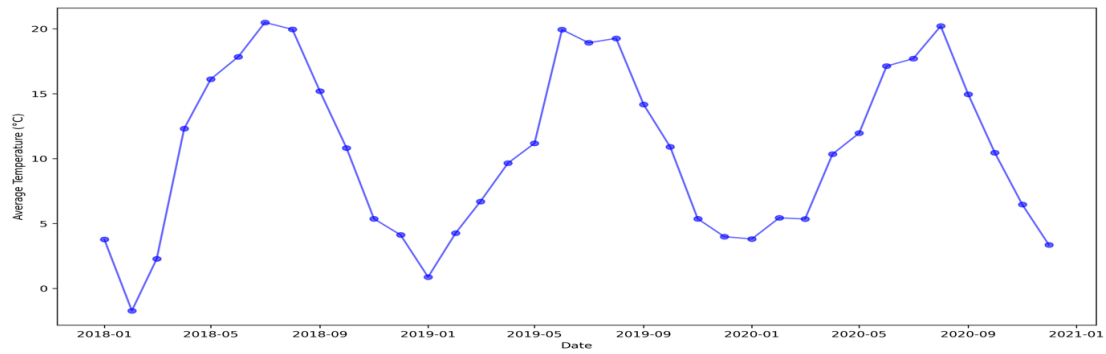
Missing temperature data were filled using interpolation methods. Outliers in the accident data were identified and removed to avoid distorted results. Additionally, the data was normalized to compensate for differences in reporting standards.

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

## Result and Interpretation

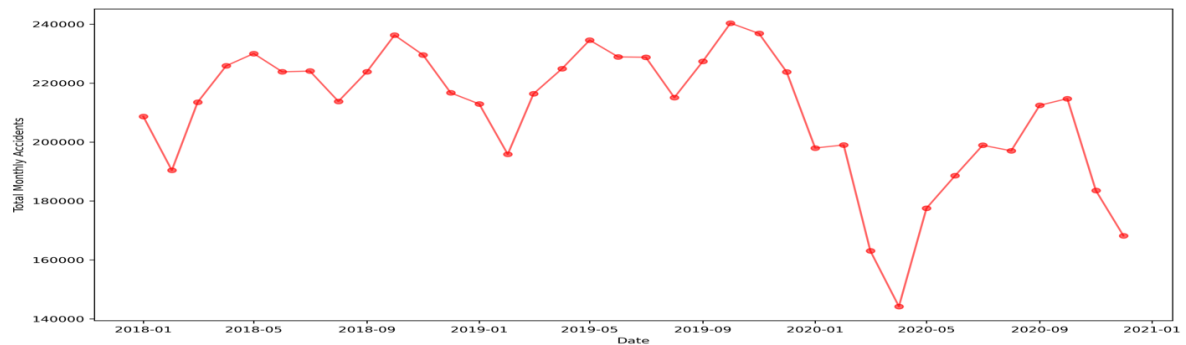
### Visualization :

The (Fig 5) shows average monthly temperatures from January 2018 to December 2020. Peaks occur in the summer months (June to August) and troughs in the winter months (December to February).The temperature pattern is periodic, indicating regular seasonal variations.



**Fig 5: Average Monthly Temperatures**

The (Fig 6) shows total monthly accidents from January 2018 to December 2020. Over the years, there are consistent ups and downs in the number of accidents. There is a significant drop in the number of accidents around early 2020, likely influenced by external factors such as COVID-19 lockdowns



**Fig 6: Monthly Road accident data in Germany**

### Correlation Analysis:

There is a clear correlation between higher temperatures and increased accidents, especially noticeable during the summer months when both variables peak. It means that summertime might see higher traffic volumes or more unsafe driving.

Around the beginning of 2020, there is a noticeable decrease in accidents, which is according to the start of the COVID-19 epidemic and the lockdown measures that followed, which significantly decreased traffic levels.

Both temperature and accident data shows clear seasonal patterns. Temperature peaks during the summer months (June-August) and drops during the winter (December-February). A slightly similar pattern can also be seen in accidents, although with less noticeable spikes and drops.

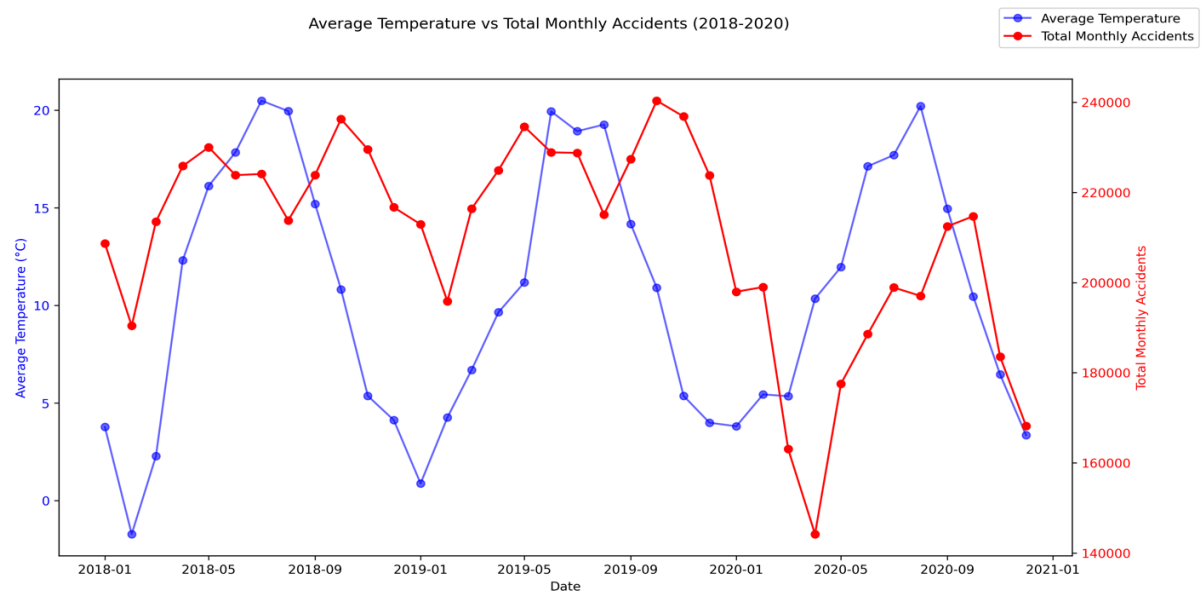


Fig 7: Average Temperature vs Total Monthly Accidents (2018 - 2020)

## Conclusion:

The analysis indicates a significant correlation between high temperatures and traffic accidents. As temperatures increase, the number of accidents correspondingly increased. This correlation indicates a potential link between weather and road safety, suggesting that rising temperatures may pose unique challenges that increase accident rates.

## Limitation:

Although the relationship is clear, it is important to acknowledge the limitations of this study. Other factors, such as road maintenance, traffic volume, and driver behavior, can contribute to inappropriate accident rates regardless of temperature. Furthermore, the data are limited to a specific time period, and the generalization of these findings to different time periods or geographic areas should be done with caution.

## Future Work:

Future research could investigate the specific mechanisms by which temperature affects road safety. Furthermore, deeper study that takes into account additional geographical and environmental factors may offer a more comprehensive view of the connection between accidents and climate.