Analyzing Road Traffic Accidents in Switzerland

CAS Applied Data Science 2021
Statistical Inference for Data Science
University of Bern

October 7th, 2021

Maria Mumtaz
Corinna Rutschi

The Road Planning Team of Bern

Maria

Corinna

Data Gathering and Cleaning

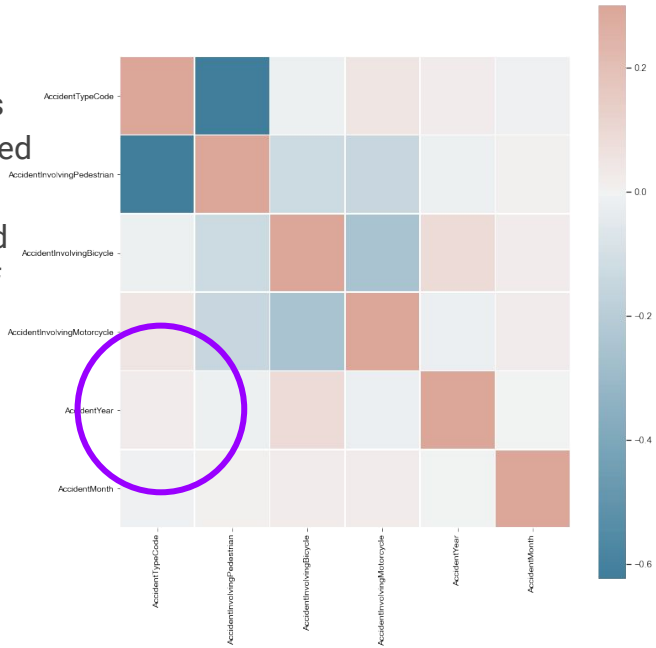
- Loading the dataset from opendata.swiss
- Selecting relevant columns
- Cleaning the data

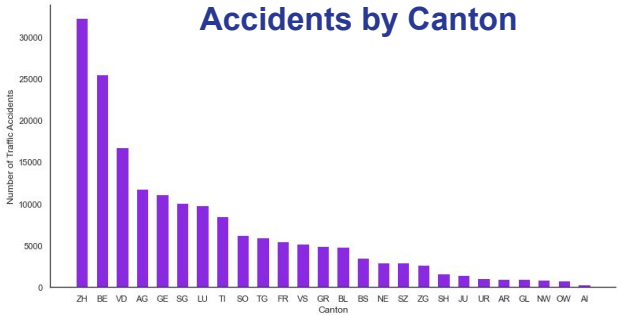
AccidentUID	AccidentType	AccidentType_en	AccidentSeverityCategory	AccidentSeverityCategory_en	AccidentInvolvingPedestrian	AccidentInvolvingBicycle	AccidentInvolvingMotorcycle	RoadType	RoadType_en	CantonCode	MunicipalityCode	AccidentYear	AccidentMonth	AccidentMonth_en	AccidentWeekDay	AccidentWeekDay_en	AccidentHour
41F802C20A6E0430A865E3320A6	at0	Accident with skidding or self-accident	as3	Accident with light injuries	False	True	False	rt433	Minor road	ZH	0261	2011	1	January	aw406	Saturday	1.0
B25356510B0E0430A865E3310B0	at4	Accident when turning-into main road	as3	Accident with light injuries	False	False	True	rt433	Minor road	GE	6621	2011	1	January	aw406	Saturday	1.0
6471BA579094E0430A865E339094	at00	Other	as2	Accident with severe injuries	False	True	False	rt433	Minor road	BE	0371	2011	1	January	aw406	Saturday	2.0
49744917E014E0430A865E33E014	at3	Accident when turning left or right	as3	Accident with light injuries	False	False	False	rt433	Minor road	BS	2701	2011	1	January	aw406	Saturday	2.0
7F3285BB044E0430A865E33B044	at0	Accident with skidding or self-accident	as2	Accident with severe injuries	False	False	False	rt433	Minor road	TI	5203	2011	1	January	aw406	Saturday	2.0

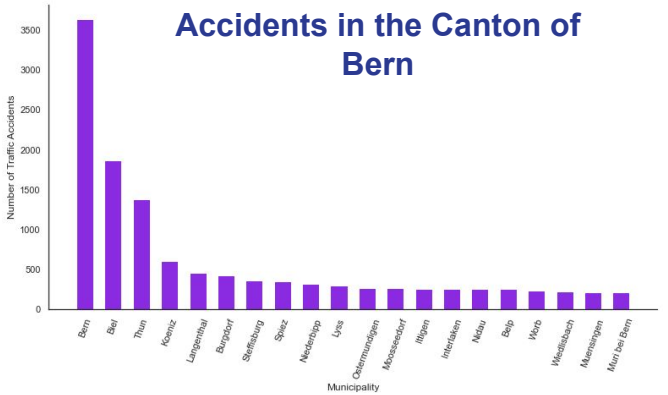
Exploratory Data Analysis [1/3]

Observations

- Accident type and accident times do not seem to be highly correlated
- Accidents involving pedestrians and Accident types are correlated
- Zurich has the highest number of accidents





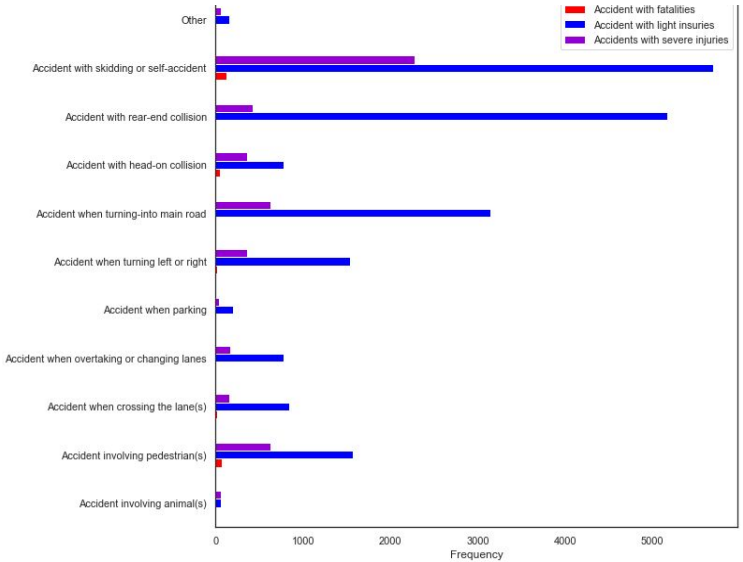


Exploratory Data Analysis [2/3]

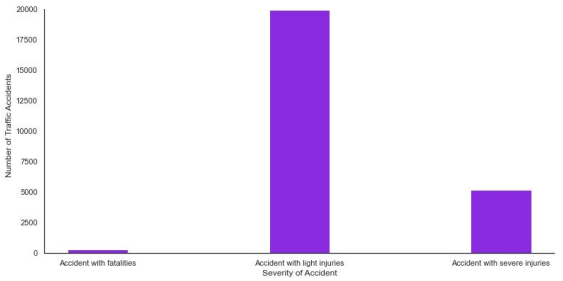
Observations

- Most accidents happen at 5 pm
- Accidents with skidding or self-accidents are more fatal
- Most accidents happen by skidding or rear-end collision
- Accidents involving pedestrians are more fatal

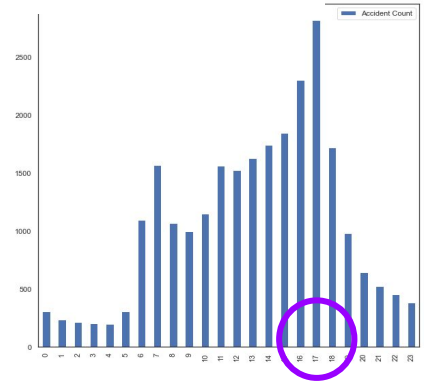
Accident Type in Relation to Severity for BE



Accidents in Bern by Severity



Accident Time



Exploratory Data Analysis [3/3]

Observations

- Most accidents happen on Fridays
- Number of accidents with respect to different categories and time of the year

	AccidentWeekDay_en	Accident Count
0	Friday	4226
6	Wednesday	3916
4	Thursday	3850
5	Tuesday	3741
1	Monday	3675
2	Saturday	3423
3	Sunday	2604

AccidentYear	AccidentInvolvingPedestrian	AccidentInvolvingBicycle	AccidentInvolvingMotorcycle
2011	257	597	504
2012	261	591	442
2013	254	590	407
2014	251	656	422
2015	242	662	419
2016	234	644	374
2017	217	685	437
2018	228	815	413
2019	225	709	373
2020	190	764	393

	AccidentYear	AccidentMonth_en	Accident Count
0	2011	April	232
1	2011	August	303
2	2011	December	208
3	2011	February	167
4	2011	January	160
...	...	...	...
115	2020	March	131
116	2020	May	222
117	2020	November	186
118	2020	October	223
119	2020	September	245

Hypothesis Testing

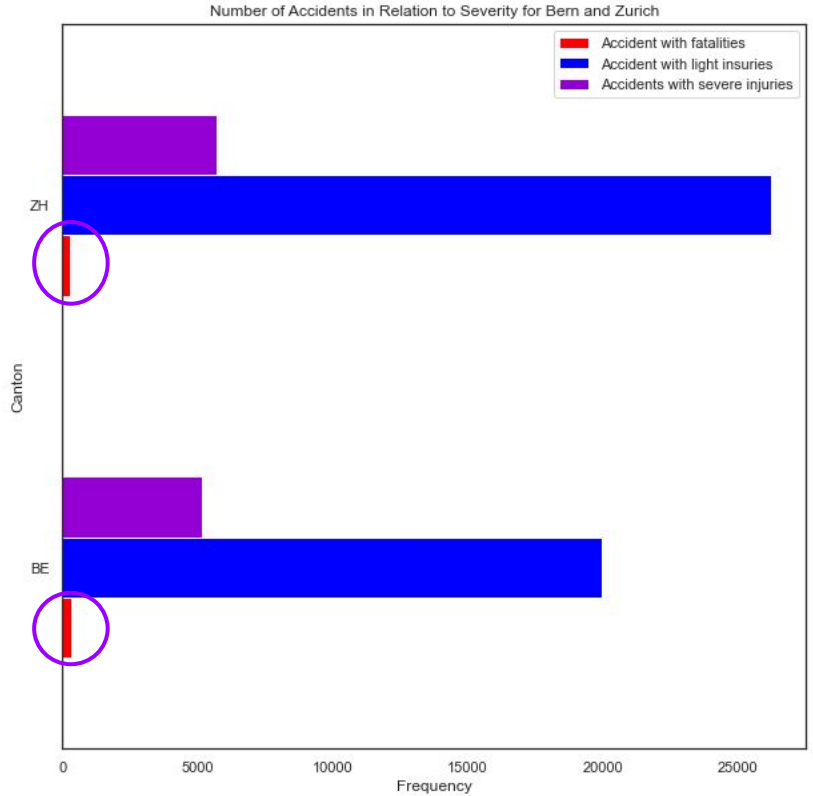
H1: Accidents in ZH are more severe than in BE

Observations

- There are more severe accidents in ZH
- But there are more accidents with fatalities in BE than in ZH

ZH	AccidentSeverityCategory_en	Accident Count
0	Accident with fatalities	286
1	Accident with light injuries	26205
2	Accident with severe injuries	5682

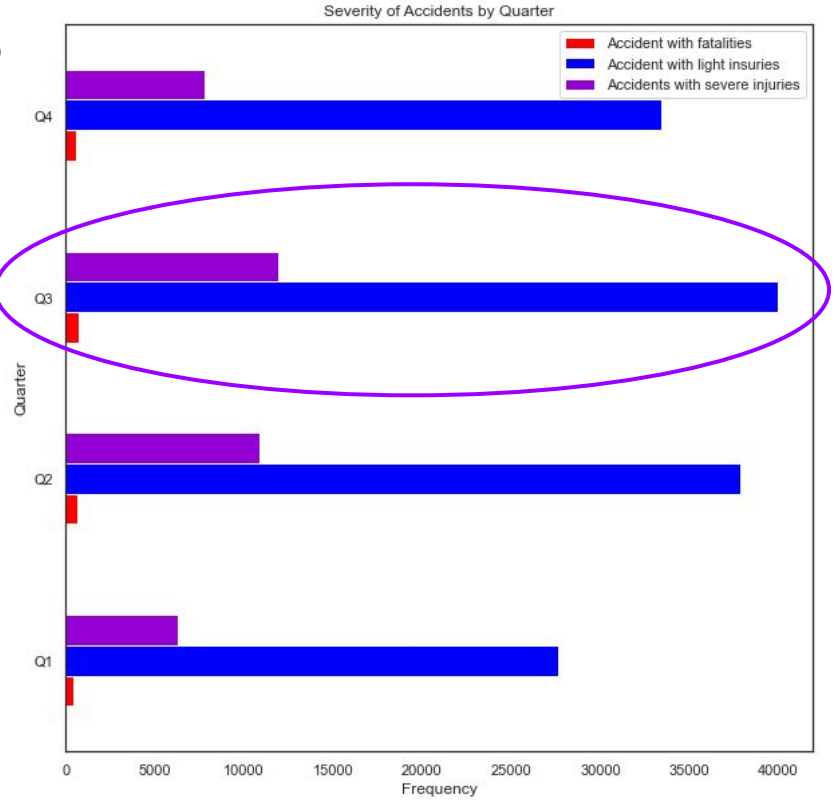
BE	AccidentSeverityCategory_en	Accident Count
0	Accident with fatalities	327
1	Accident with light injuries	19941
2	Accident with severe injuries	5167



H2: Accidents that happen in Q1 and Q4 are more severe than accidents that happen in Q2 and Q3

Observations

- Most severe accidents happened in Q3



H3: Average number of fatal accidents per Canton is equal to average number of fatal accidents in Bern

H0: Average number of fatal accidents per canton is not equal to that of Bern

Results

- Comparing columns 'AccidentTypeCode = Accident with fatalities ' for two dataframes (one for Switzerland and the other one for Bern)
- Two sample tTest
- Since the p-value is greater than 0.05, **we cannot reject the null hypothesis**
- We can conclude that the average number of fatal accidents per canton is most likely not equal to that of Bern, and Bern is among the 'not-so-safe' cantons of Switzerland

```
In [31]: stats.ttest_ind(df_hypothesis['AccidentTypeCode'],df_hypothesis_bern['AccidentTypeCode'], equal_var = False, alternative='two-sid

Out[31]: Ttest_indResult(statistic=0.3455304860978335, pvalue=0.7298681233798292)
```

Predictions for the Future and Outlook

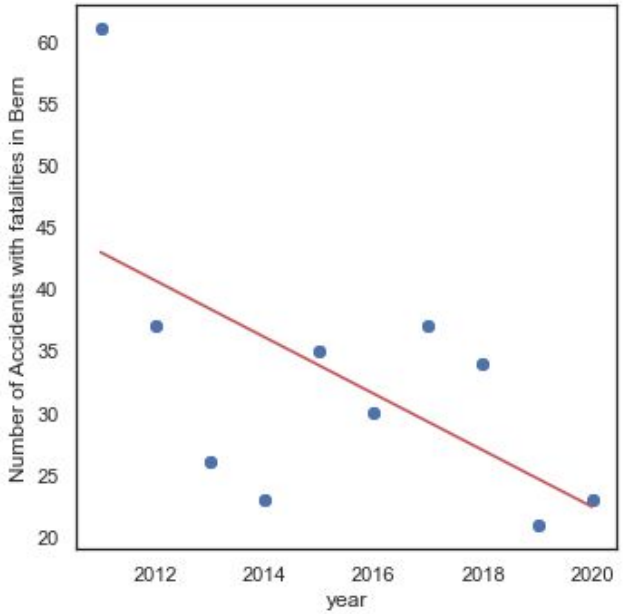
Predicting number of fatal accidents

Observations

- The number of accidents from 2011 to 2020 seem to decrease
- By calculating the gradient and Intercept, we would be able to predict the number of (fatal) accidents for future years

	AccidentYear	AccidentSeverityCategory_en	Count
0	2011	Accident with fatalities	61
3	2012	Accident with fatalities	37
6	2013	Accident with fatalities	26
9	2014	Accident with fatalities	23
12	2015	Accident with fatalities	35
15	2016	Accident with fatalities	30
18	2017	Accident with fatalities	37
21	2018	Accident with fatalities	34
24	2019	Accident with fatalities	21
27	2020	Accident with fatalities	23

Number of fatal Accidents in Bern



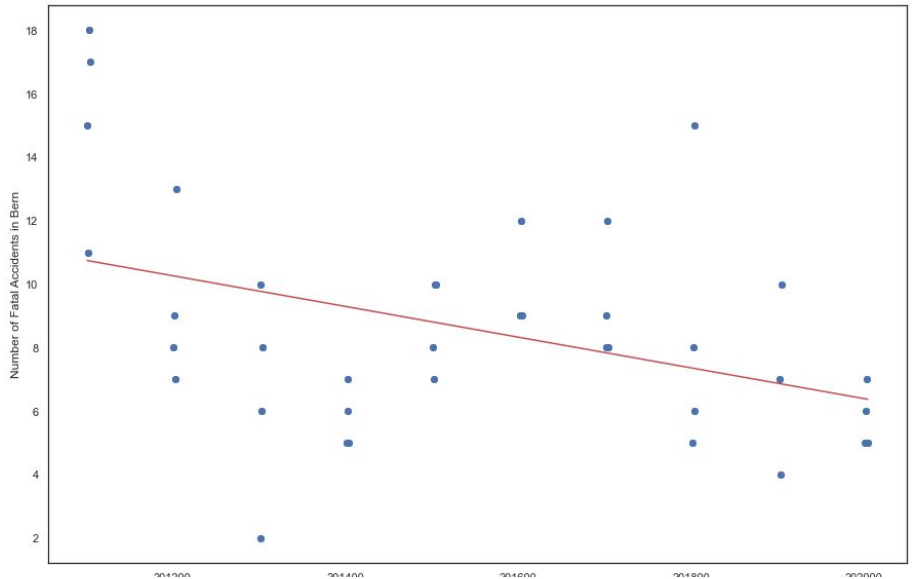

```
In [35]: #predicting number of accidents in Bern in 2022
         result= gradient*2022+intercept
         result

Out[35]: 17.848484848484077
```

Linear Regression 1: Relationship between quarters of the year and number of fatal accidents

Observations

- Number of accidents per quarter and year (to have more data points)

	QuaterAndYear	Count
0	201101	15
1	201102	11
2	201103	18
3	201104	17
4	201201	8
5	201202	9
6	201203	7
7	201204	13
8	201301	2
9	201302	10




```
In [39]: #predicting number of accidents in Bern in first quarter of 2024
         result= gradient*202401+intercept
         result

Out[39]: 4.444638995345258
```

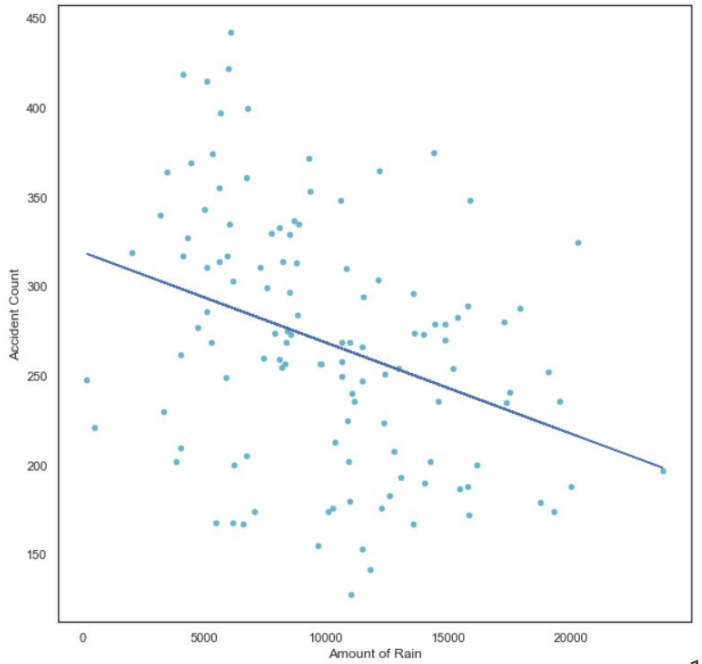
Linear Regression 2: Relationship between amount of rain and the number of accidents per month and year

Observations

- There seems to be a negative relationship between number of accidents and amount of rain
- More rain indicates less accidents
- The p-value is, however, very small

```
Out[55]: 0.00011956385258052757
```

	YearMonth	Accident Count	Amount of Rain
0	2011-01	167	6592.30
1	2011-02	168	5483.28
2	2011-03	202	3829.82
3	2011-04	319	2027.43
4	2011-05	317	5934.17
...	...	...	...
115	2020-08	365	12164.85
116	2020-09	397	5662.16
117	2020-10	273	13963.46
118	2020-11	230	3328.81
119	2020-12	193	13060.54



Findings and Conclusions

- Most accidents happen on Fridays and at around 5pm
- Most accidents are caused by skidding
- Bern has high number of fatal accidents compared to other cantons
- Most accidents happen at the second Quarter of the year
- There seems negative relation of rain on number of accidents

Learnings

- If there are few data points, it is possible to add more data
- Visualizations are easy to read, but with minor variations, aspects could be overlooked
- More can always be done. The important thing is to pay attention to what is relevant
- Colab as a collaboration tool?