06_worksheet

November 16, 2022

1 Week 6 worksheet - GB cycling accidents

In the `data` folder of the course materials you should find a `CSV` file called `gb_cycling_accidents.csv` which contains data on bicycle accidents in Great Britain from 1970 to 2018. I retrieved the data set from kaggle, which cites data.world as the original source. Each row holds information about a specific accident, and each column holds information about the accident, such as the date, time of day, day of week, number of vehicles involved, weather conditions, severity, etc. Here is the full explanation of the columns in the data set.

Variable	Definition
Accident_Index	Unique identifier for the accident. This may be thought of as the accident "case number".
Number_of_Vehicles	Number of vehicles that were involved in the accident
Number_of_Casualties	Number of casualties resulting from the accident
Date	Date when the accident happened
Time	Time when the accident happened
Speed_limit	Speed limit on the part of the road where the accident took place
Road_conditions	Road condition (e.g., "frost") at the time and place of the accident
Weather_conditions	Whether condition (e.g., "rain") at time and place of the accident
Day	Day of the week when the accident occurred
Road_type	Type of road (e.g., "Dual carriageway") where the accident happened
Light_conditions	Light conditions (e.g., "Daylight") at time of accident
Gender	Whether the accident victim was Male or Female
Severity	How severe (e.g., "Serious") the accident was
Age_Grp	Age group of the accident victim

Let's explore the frequency of accidents with respect to the different variables.

1.0.1 1. Import `pandas` and read `data/gb_cycling_accidents.csv` into a `DataFrame`

[ ]:

1.0.2 2. How many unique values are in the following columns?

- `Speed_limit`
- `Road_conditions`
- `Weather_conditions`

- `Road_type`
- `Light_conditions`
- `Gender`
- `Severity`
- `Age_Grp`

[ ]:

1.0.3 3. What road conditions were associated with the most and least accidents?

[ ]:

1.0.4 4. What weather conditions were associated with the most and least accidents?

[ ]:

1.0.5 5. What road type was associated with the most and least accidents?

[ ]:

1.0.6 6. What light conditions were associated with the most and least accidents?

[ ]:

1.0.7 7. What speed limit was associated with the most and least accidents?

[ ]:

1.0.8 8. Based on the above, write a single sentence that summarises the conditions in which most accidents appeared to occur.

[ ]:

1.0.9 9. Create a bar chart showing how accidents were distributed by `Age_Grp`

[ ]:

1.0.10 10. Across all accidents, what percentage involved Males, what percentage involved Females, and what percentage involved people identifying as 'Other'? Show the results in a pie chart.

[ ]:

1.0.11 11. What was the highest number of vehicles involved in a single accident?

[ ]:

1.0.12 12. What was the highest number of casualties involved in a single accident?

[ ]:

1.0.13 13. On which day of the week did the accident with `Accident_Index` `201443N027074` occur?

[ ]:

1.0.14 14. Create a separate `DataFrame` for all serious accidents that happened on a Sunday in wet road conditions. How many were there?

[ ]:

1.0.15 15. Create and assign a new `DatetimeIndex` for the `DataFrame` using the `Date` and `Time` columns

[ ]:

1.0.16 16. Add a new column to the `DataFrame` called `long_date`. It should contain the correct dates matching the following format.

- Wednesday 09 February 2012

[ ]:

1.0.17 17. What is the worst day on record in terms of the number of accidents that were reported?

[ ]:

1.0.18 18. Make a bar chart showing total accidents by month of the year

[ ]:

1.0.19 19. Make a line graph showing the total number of accidents that occurred each year from 1979-2018. Have accidents declined overall? In which years did the most and least cycling accidents occur?

[ ]:

1.0.20 20. Repeat the above, but this time with separate lines for `Gender`

[ ]:

1.0.21 21. Repeat the above, but this time with separate lines for `Age_Grp`

[ ]:

1.0.22 22. Repeat the above, but this time with separate subplots for `Severity`

[ ]:

1.0.23 23. Make a bar chart showing the total number of accidents for each hour in the day from 1979-2018

[ ]:

1.0.24 24. As above, but with stacked bars using different colours for each day of the week

[ ]:

1.0.25 25. Make a bar chart showing the year-on-year percentage change for accidents with different coloured bars for each `Severity`

[ ]: