Data analysis assignment 2 (3 weeks)

For this second assignment we are going to be using a dataset containing countywide car accident data, which covers data from February 2016 to June 2020. A more detailed description of the dataset and its columns can be found on the following pages. For this assignment imagine an insurance company that deals with car insurances. For them it's crucial to best understand accidents and what causes them. Please help the company by making their collected data of the accidents easily digestible. To do this, create a small dashboard for the management containing a high level overview of the data and some more in-depth analyses. Even though this assignment has similar requirements as the previous assignment, the size of this dataset brings a lot of new challenges with it. For example the dataset is now so big, that you can't open the entire file in Excel. So changing one row in excel before opening the data in the notebook, as could be done for the previous assignment, is no longer possible. Also I'd like you to focus on a few different aspects as compared to last time. For this assignment you may work in groups of 2.

- The goal of this assignment is to deliver a small dashboard based on the provided dataset in Microsoft PowerBI. One of the many traps people often fall for is to make the pages of the dashboard way too busy. This overloads the user of the dashboard with data and makes it difficult to get the correct information. With a lot of data to choose from and a lot of possible data to display you will need to choose what you want to highlight on your dashboard and what to leave out. To be able to make a good decision on this, Exploratory Data Analysis becomes even more important and will probably take a lot longer. Look into the data first and afterwards discuss your findings with your group and decide what you will focus on. Try to tell a story throughout your dashboard. Try to convey your data as concise as possible. You don't have to use every column or show every single category of a column. Your report/dashboard should have a maximum of 4 pages.

- For the people that are going to work with creating a model and optimizing it, clearly state beforehand which results you're expecting and why you think the model will be valuable. It can become quite easy to create a model and optimize the parameters and just call it a day, while the model will not be as useful in practice. Really think about why your model will improve on their decision making. This is also helpful for the people who are just now starting to learn a bit of modeling, since then even if the model didn't turn out to perform well, the basis could still be sound.

- There is more time for this assignment than the previous one, so try to play around with things you think could still use improvement. If you're good at visualizing, try to create a model and if you're good at modeling, maybe spend a bit more time in tidying up your report. This is not to say you should focus on things you don't enjoy doing, but it's a good opportunity to go outside of your comfort zone. At the end of the assignment we will discuss the created dashboards in a meeting.

- Somewhere along these 5 weeks you may ask to show a first draft of your report, to get feedback on whether or not you're working in the right direction. The decision to ask for this meeting is up to the groups themselves and can be done early on, or later in the process.

Data Description

#	Attribute	Description	Nullable
1	ID	This is a unique identifier of the accident record.	No
2	Source	Indicates source of the accident report (i.e. the API which reported the accident.).	No
3	TMC	A traffic accident may have a [Traffic Message Channel (TMC)](#) code which provides more detailed description of the event.	Yes
4	Severity	Shows the severity of the accident, a number between 1 and 4, where 1 indicates the least impact on traffic (i.e., short delay as a result of the accident) and 4 indicates a significant impact on traffic (i.e., long delay).	No
5	Start_Time	Shows start time of the accident in local time zone.	No
6	End_Time	Shows end time of the accident in local time zone. End time here refers to when the impact of accident on traffic flow was dismissed.	No
7	Start_Lat	Shows latitude in GPS coordinate of the start point.	No
8	Start_Lng	Shows longitude in GPS coordinate of the start point.	No
9	End_Lat	Shows latitude in GPS coordinate of the end point.	Yes
10	End_Lng	Shows longitude in GPS coordinate of the end point.	Yes
11	Distance(mi)	The length of the road extent affected by the accident.	No
12	Description	Shows natural language description of the accident.	No
13	Number	Shows the street number in address field.	Yes
14	Street	Shows the street name in address field.	Yes
15	Side	Shows the relative side of the street (Right/Left) in address field.	Yes

#	Attribute	Description	Nullable
16	City	Shows the city in address field.	Yes
17	County	Shows the county in address field.	Yes
18	State	Shows the state in address field.	Yes
19	Zipcode	Shows the zipcode in address field.	Yes
20	Country	Shows the country in address field.	Yes
21	Timezone	Shows timezone based on the location of the accident (eastern, central, etc.).	Yes
22	Airport_Code	Denotes an airport-based weather station which is the closest one to location of the accident.	Yes
23	Weather_Timestamp	Shows the time-stamp of weather observation record (in local time).	Yes
24	Temperature(F)	Shows the temperature (in Fahrenheit).	Yes
25	Wind_Chill(F)	Shows the wind chill (in Fahrenheit).	Yes
26	Humidity(%)	Shows the humidity (in percentage).	Yes
27	Pressure(in)	Shows the air pressure (in inches).	Yes
28	Visibility(mi)	Shows visibility (in miles).	Yes
29	Wind_Direction	Shows wind direction.	Yes
30	Wind_Speed(mph)	Shows wind speed (in miles per hour).	Yes
31	Precipitation(in)	Shows precipitation amount in inches, if there is any.	Yes
32	Weather_Condition	Shows the weather condition (rain, snow, thunderstorm, fog, etc.)	Yes
33	Amenity	A POI annotation which indicates presence of amenity in a nearby location.	No

#	Attribute	Description	Nullable
34	Bump	A POI annotation which indicates presence of speed bump or hump in a nearby location.	No
35	Crossing	A POI annotation which indicates presence of crossing in a nearby location.	No
36	Give_Way	A POI annotation which indicates presence of give_way in a nearby location.	No
37	Junction	A POI annotation which indicates presence of junction in a nearby location.	No
38	No_Exit	A POI annotation which indicates presence of no_exit in a nearby location.	No
39	Railway	A POI annotation which indicates presence of railway in a nearby location.	No
40	Roundabout	A POI annotation which indicates presence of roundabout in a nearby location.	No
41	Station	A POI annotation which indicates presence of station in a nearby location.	No
42	Stop	A POI annotation which indicates presence of stop in a nearby location.	No
43	Traffic_Calming	A POI annotation which indicates presence of traffic_calming in a nearby location.	No
44	Traffic_Signal	A POI annotation which indicates presence of traffic_signal in a nearby location.	No
45	Turning_Loop	A POI annotation which indicates presence of turning_loop in a nearby location.	No
46	Sunrise_Sunset	Shows the period of day (i.e. day or night) based on sunrise/sunset.	Yes
47	Civil_Twilight	Shows the period of day (i.e. day or night) based on civil twilight.	Yes
48	Nautical_Twilight	Shows the period of day (i.e. day or night) based on nautical twilight.	Yes
49	Astronomical_Twilight	Shows the period of day (i.e. day or night) based on astronomical twilight.	Yes