Here, you will have the opportunity to be as creative as you want and come up with an idea to leverage the accident data that you found to predict the different accidents' severity. Make sure to provide sufficient justification of why your solution is valid and why would your community or a group of people be interested in your project.

For this week, you will required to submit the following:

A description of the problem and a discussion of the background. (15 marks)

A description of the data and how it will be used to solve the problem. (15 marks)

Project Title \*

**Car accident severity**

Clearly define a problem or an idea of your choice. Remember that data science problems always target an audience and are meant to help a group of stakeholders solve a problem, so make sure that you explicitly describe your audience and why they would care about your problem.

This submission will eventually become your **Introduction/Business Problem** section in your final report. So I recommend that you push the report (having your Introduction/Business Problem section only for now) to your Github repository and submit a link to it.

Car accident are one of the leading causes of death in US, the consequences may range from minor injury/vehicle damage to major personal injury or even death. It is therefore vital for us to understand which are the factors influencing the likelihood of a car accident occurring at certain location, as well as the factors which influence the severity of these car accidents.

Factors likely to influence the likelihood and severity of car accident may include: weather, road condition, timing in the day, and traffic congestion. Additional factors may include driver’s personal condition, e.g. consumption of alcohol, tiredness etc.

The report aims to provide understanding of the factors influencing the frequency and severity of car accidents. These insights may provide reference for city planners and emergency service providers when allocating road design/medical response resources.

Describe the data that you will be using to solve the problem or execute your idea. So make sure that you provide adequate explanation and discussion, with examples, of the data that you will be using.

This submission will eventually become your **Data** section in your final report. So I recommend that you push the report (having your **Data** section) to your Github repository and submit a link to it.

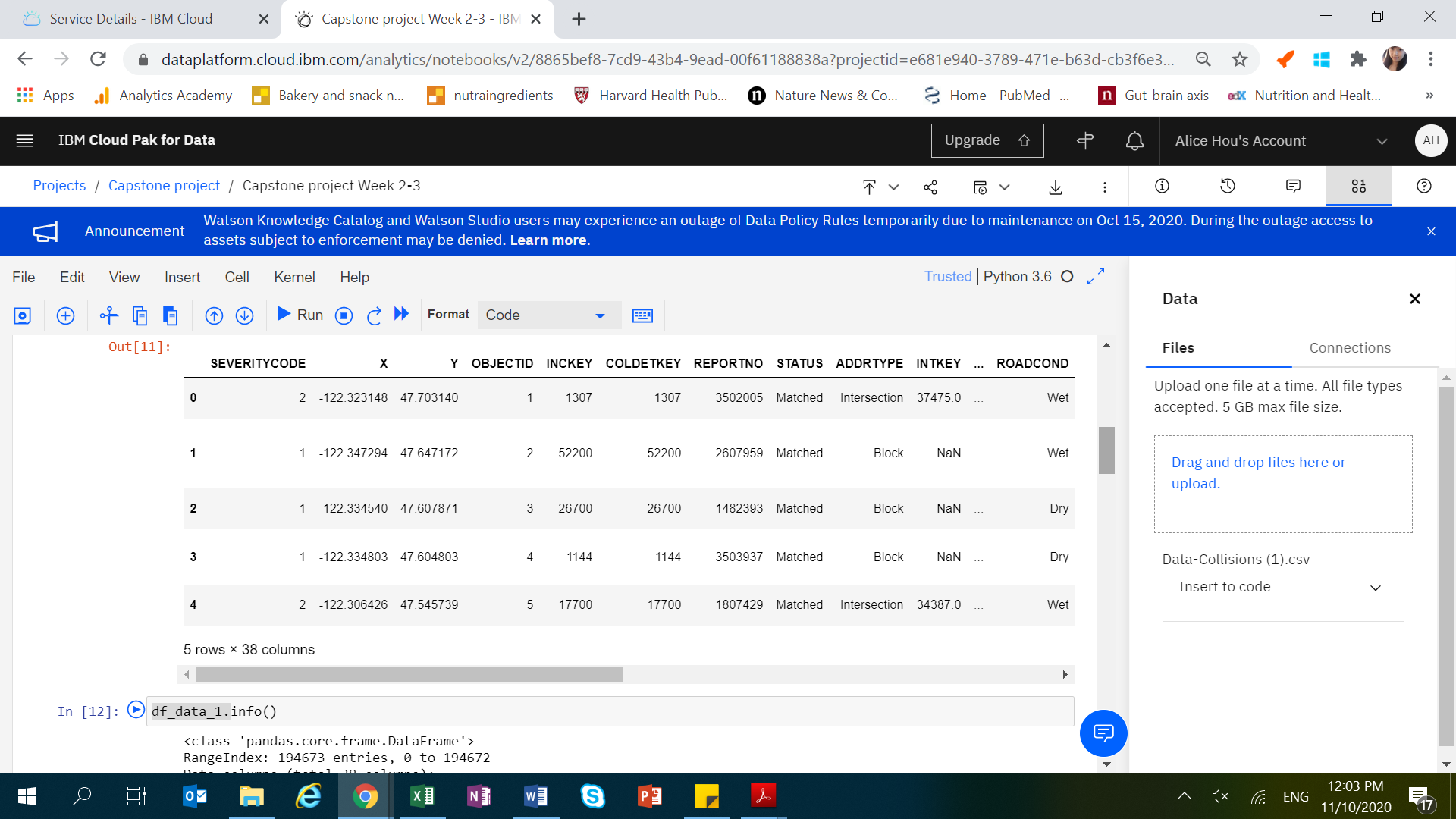
Data description

The data is provided by SDOT Traffic management division which include all car collisions from 2004 to present in Seattle and has been updated weekly. The data has been downloaded and read into Pandas Dataframe, and the content and data types displayed using the HEAD and DTYPES functions.

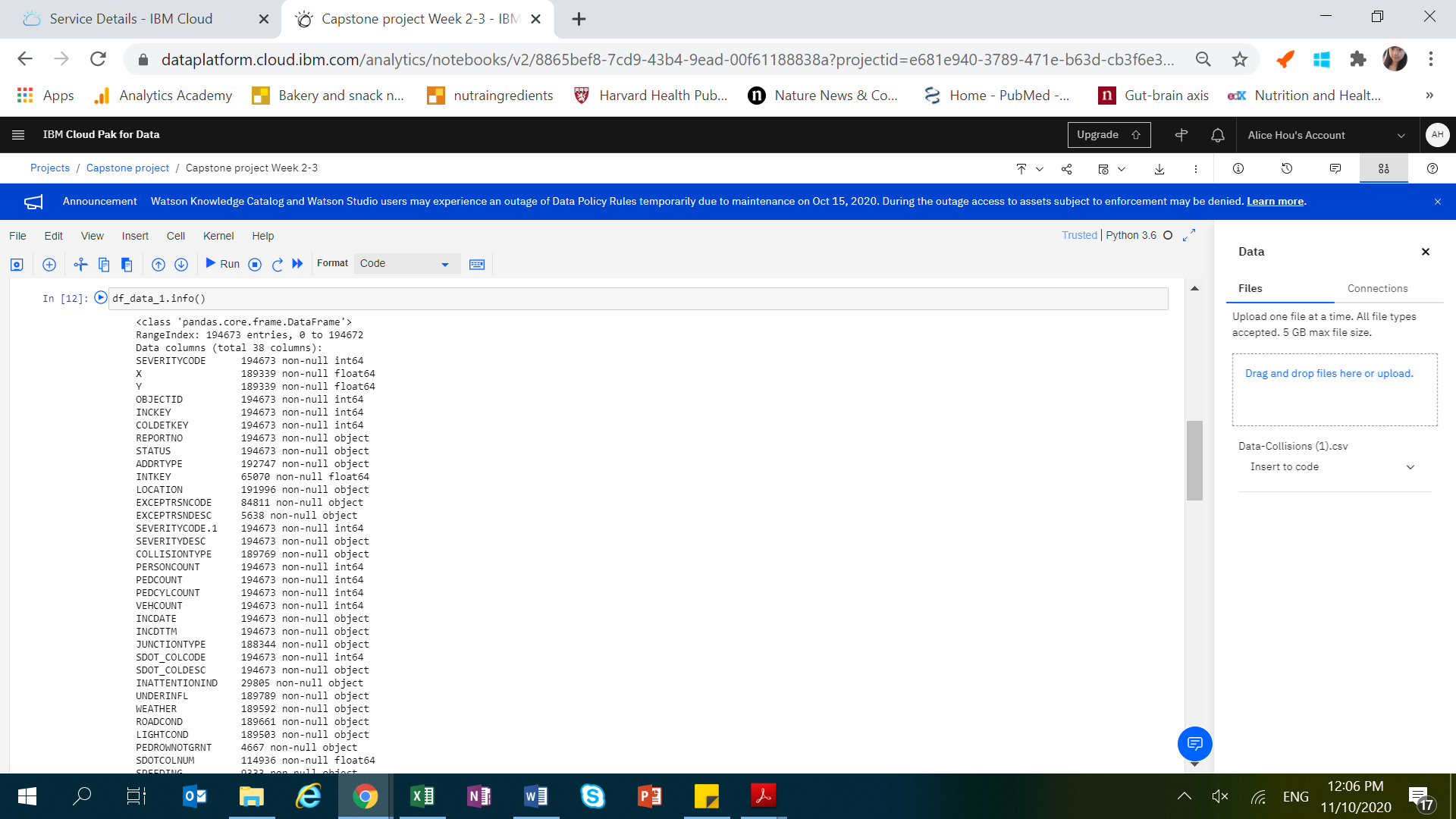
The target variable is SEVERITYCODE which takes the value from 0, 1, 2, 2b or 3. The definition of the severity codes are provided in the metadata and shown below.

|  |  |
| --- | --- |
| Severity Code | Meaning |
| 0 | Unknown |
| 1 | Prop damage |
| 2 | Injury |
| 2b | Serious injury |
| 3 | Fatality |

As shown below, the data consists of 38 columns.



From the data information description , we can see that some variables are categorical (data type as object).



​For the second week, the final deliverables of the project will be:

1. A link to your Notebook on your Github repository, showing your code. (**15 marks**)
2. A full report consisting of all of the following components (**15 marks**):

* Introduction where you discuss the business problem and who would be interested in this project.
* Data where you describe the data that will be used to solve the problem and the source of the data.
* Methodology section which represents the main component of the report where you discuss and describe any exploratory data analysis that you did, any inferential statistical testing that you performed, if any, and what machine learnings were used and why.
* Results section where you discuss the results.
* Discussion section where you discuss any observations you noted and any recommendations you can make based on the results.
* Conclusion section where you conclude the report.

3. Your choice of a presentation or blogpost. (**10 marks**)