**IBM Capstone Project**

**"Weather-related car crashes severity prediction"**

## Data

### Dataset

The dataset called "Collisions" was provided by Transportation\_SeattleCityGIS and obtained from the gisdata.seattle.gov website [6], it contain all the collisions provided by SPD in Seattle and recorded by Traffic Record since 2004 and it is update weekly, at this moment it has 221,525 car crashes reports with 40 different attributes like [7]:

* Latitude and Longitude.
* Location (Name or description)
* Severity
* Collision Type
* Environment Conditions, etc.

### Features and Labels

To build a Machine Learning model capable of predict the severity of a collision it will be train based on attributes like:

* The location where the person would transit (longitude, latitude, neighborhood)
* Date (Day and Month)
* Time (HH)
* Weather conditions (Clear, Raining, Overcast, Snowing, etc.)
* Road conditions (Dry, Wet, Slush, etc.)
* Light conditions (Daylight, Dark with streetlights, Dusk, etc.)

And it will have the task to predict between one of four possible severity classes:

* 1 - prop damage
* 2 - injury
* 2b - serious injury
* 3 - fatality

### Data Wrangling

All observations which have an "unknown" severity code, weather, road or light conditions will be drop, a neighborhood attribute will be generate from the coordinates and a geojason file, the date will be separated on day and month attributes and the time transform into a 0-23 integer. Weather, road and light conditions categories will be separate as one-hot encoded attributes, finally the rest of attributes will be drop.

**Bibliography**

[6] **Collisions (dataset source)**<https://data-seattlecitygis.opendata.arcgis.com/datasets/collisions>

[7] **Collisions-All Years (dataset metadata)** <https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Metadata.pdf>