**Coursera - Applied Data Science Capstone**

Ana Cristina Bittar de Oliveira

Capstone Project – Week 1

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

How to prevent or, at least, reduce the severity of road accidents is a very important question. A better understanding of the probable causes of collisions could be a good approach in order to suggest actions to be taken or improvements to be made in traffic engineering systems willing to prevent them.

The present work will analyze data related to collisions in the city of Seattle. It would be out of scope to generalize this model for sites other than those included on the data to be analyzed, although intuition indicates that, for similar traffic systems, the results could be extended.

The main potential consequences of a car collision are:

* Injuries
* Traffic jams
* Related costs (government, insurance companies, individuals, *et al*)

Seattle Traffic Management Division releases weekly updates about all collisions recorded by Traffic Records since 2004. The main goal is to analyze that data to identify patterns and make predictions about risk and severity of the accidents. With this information an alert system could be developed to provide guidance on what to do in short, medium, or large term.

Some examples of actions to be taken could be:

* deviate traffic in pre-defined times or weather conditions (preventive)
* improve traffic signs
* reduce speed in specific roads in determined weather conditions
* reduce speed in specific roads (permanently)
* make structural changes on traffic
* educative campaigns and advertisement
* recommendations to improve the data collection
* other

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

This project will use the dataset shared in the Capstone.

The dataset provides information about collisions occurred in the city of Seattle since 2004 and is updated weekly. It was also provided very detailed metadata for better understanding.

Some aspects that should be observed are:

* There’s very detailed geographic information, but no demographic about the conductors
* The data went back 15 years, but traffic volume increases rapidly, so there should be any weighting to ensure the significance of most recent occurrences
* Analytic Approach
* Data Requirements
* Data Preparation
* Modelling
* Evaluation

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.

Some aspects that should be observed are:

* There’s very detailed geographic information, but no demographic about the conductors
* The data went back 15 years, but traffic volume
* 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**)