Project Title

First Author1, Second Author2, Third Author3 & Fourth Author4

1first author student number and e-mail contact

2second author student number and e-mail contact

3third author student number and e-mail contact

4fourth author student number and e-mail contact

Abstract**:** maximum of 250 words, font Times New Roman, size 10, line spacing 1.0

Keywords: maximum three keywords separated by semicolon  
Statement of Contribution: clearly state the contributions of each group member to the project in particular who performed the following: Data collection, Data Cleaning, Data Analysis, Wrote the report, Design the study, Discussion of the Results.

# Introduction

Motor vehicle accidnts can have a high impact on our every day life, not only regarding costs for the public structure and atraffic but most importantly when we consider the persons involved. Sonly in Portugal, in the last years there have been an average of more than 30 thousand ref 1 accidents per year that cause more than 40 thousand wounded and 400 deaths. For this reasons cities around Portugal have an increased interest in understanding not only the causes that lead to accidents but also the main factors that cause the occurrence of wounded passengers and pedestrians. One of those cities is the Portuguese capital, which lunched a competition with the objective to pinpoint the locations where a high number of accidents occur, bur also to infer which of the city/persons characteristics are leading to accidents.

Accident occurrence is complex and normally it has two main influences. Driver characteristics such as skill level, experience, risk taking behaviours (cellphones usage) and age, and external conditions such as quality of the road, traffic and atmospheric conditions number of persons in the vehicle and vehicle type.

Although most of the external factors can be obtained and evaluated easily, some of the drivers characteristics can be difficult to obtain. How to evaluate if a driver is inexperienced or not? And soes he has risk taking behaviour? Another problem is that for example in Portugal you have the option of doing a mutual agreement when an accident occurs, leading to many of the accident not being registered by the police and only by the insurance companies. This for instance can also limit our analyss since only the more violent accidents or the ones were no agreement are reached are registered.

Several studies have been made that

# Data

## Description and Extraction

For this we used several datasets. Most of them were supplied by the LxDataLab and focused on the accidents and city characteristics, however we also tryiend to get information we though pertinent to the accident evaluation.

The LxDataLab data set where:

* Lisbon accidents registered in 2019 by the “Autoridade Nacional de Segurança Rodoviária” (ANSR)

This data set was supplied by the LxDataLab team and had multiple information regarding the accident characteristics and persons involved.

* Lisbon accidents registered in 2019 by the “Regime de Sapadores Bombeiros” (RSB)

This data set was supplied by the LxDataLab team and had only information regarding the time, type and location of the accidents.

* Lisbon street height, slopes, crossings and traffic lights

This data set was supplied by LxDataLab team and had multiple information reagarfing street characteristics

* Lisbon traffic jams in 2019

This data set was supplied by the LxDataLab and had information regarding the city traffic jams measured by waze in the year of 2019

* Portuguese precipitation

This data set was extracted from the IPMA site and contains the precipitation measurements in Portugal since the year of 1950 ref2.

## Transformation

We started by evaluating the accident data consistency for errors and missclassifications. Footnote 1 Has previously stated the RSB data set was very simple, however the ANSR accident presented a high number of descriptive variables regarding the accidents.

We needed to create a single accident data set, however we had some limitations since some of the accidents in the ANSR data set didn’t have geolocation. For this reason we decided to divide the accidents data set in it in three set:

1. Accident data without geolocation and considering only the general accident data (date and type)
2. Accident data with geolocation and considering general accident data (date, type and location)
3. Accident data with only the ANSR descriptive data;

Next, we evaluated the remaining data sets namely the supplied shapefiles (Lisbon street information) and traffic jams data performed the required cleaning.

Finally, when performing geolocation evaluation we needed to aggregate the data by some means. For this purpose we used the python package “h3” developed by Uber which convert each geolocation to an unique hex, allowing the aggregation of our accident, traffic jam and street characteristics.

# Results and Discussion

Basic evaluation

* **Mapa de dados de acidentes em Lisboa (separar por cores os dados de cada ficheiro)**
* **Análise mais global**
* **Acabar com uma análise mais estatística… pensar como fazer isto…**

# Conclusions

# Acknowledgements

Use this section to acknowledge the contribution of a third party for your project, for instance additional mentoring or data acquisition, or if you are using data from your company for instance.

# References

1 pordata: <https://www.pordata.pt/Portugal/Acidentes+de+via%C3%A7%C3%A3o+com+v%C3%ADtimas++feridos+e+mortos+++Continente-326>

2 ipma site para os dados de precipitação

Footnote 1: see cleaning data notebook

Use the sections/sub-sections included in this document as a template for your project report.

The report should have no more than 3500 words, plus Figures, Tables, and max of 30 References.

Figures and tables should be correctly enumerated and have a self-explanatory caption.

The font should be Times New Roman with size 10 and line spacing of 1.5. Text must be justified.

Follow the Physical Review Style for references, <https://cdn.journals.aps.org/files/styleguide-pr.pdf>.

Consider using a reference managing tool such as Mendeley, Zotero, or Read Cube.

References should be done using the APA style and must reference meaningful sources (don’t cite blogs for instance, consider instead citing research articles). Maximum of 15 references allowed.

References made to particular content in the Jupyter Notebook should be done as a footnote[[1]](#footnote-1).

Don’t forget to correctly specify the original data sources, and correctly credit their authors.

Consider sharing your Python materials and all the resources you used for this project in a public GitHub repository.

1. Check Jupyter Notebook XXX [↑](#footnote-ref-1)