(Says linespacing is 1.15, but that’s an invalid size.)

**Used Car Characteristics Exploratory Analysis**

**Problem statement:** Buying a new car is a big challenge for a lot of people. There are 2 crucial tasks everyone must do before deciding the vehicle:

* consider their budget
* consider the car

Lots of people can’t afford to buy a brand new vehicle, as that’s something that’s considered a luxury today. That’s why the used auto market has seen major upsurge in popularity. The term „used” sometimes creates a bad taste in your mouth, because it comes with the belief that if someone already used it, then it must be

* dirty
* damaged
* „why would they get rid of something that’s in good condition, there must be something sneaky”

But these are just assumptions we make, and most of the cars don’t have any substantial problems, other than some already driven kilometres. Some businesses that resell cars even have regulations to only accept vehicles in good condition or they can also buy it for cheap and fix it themselves. However this doesn’t mean that we should not analyze our choices before making a purchase. In many cases, we can find correlation between the status of the car and sometimes even causation, regarding the distance driven, previous accidents and year of construction of the model. These are characteristics that we would like to explore in our research. Our goal is to do explorative analysis, data mining from publicly available websites (or even datasets) and clustering analysis on used cars. We want to find underlying patterns in the data and show how different characteristics result in (price) differences.(?)

**Success criteria:** We consider success when we can interpret the results clearly and in high confidence. We want to minimize errors from incorrect or missing data, be able to visualize patterns with different techniques (models, scatterplots, histograms etc…) , identify correlation of data and be able to recognize problematic examples/outliers.