Exploratory Data Analysis

Wine Reviews Dataset

Authors:

* David Jîtcă (407)
* Iordache Ștefan (411)

Contents

[1. Motivation 2](#_Toc29640176)

# Motivation

We wondered how we could create a predictive model to identify which wines are we drinking without opening the bottles. Blind tasting is a discipline for master sommeliers but even normal people can do it with little to no help from our models.

In order to obtain such a model and some great results it was necessary to understand in depth the dataset that we are going to work with. Some exploratory data analysis was necessary and we tried our best to find which features are the most important and what kind of columns we can drop without noticing any other flavors in our wines from those expected.

Such a project might open a path to advanced systems. For example, using our models we can predict the right price for a bottle of wine before buying it or we can verify if a certain drink have the variety and the taste we would like based on what we are drinking. This model might be used in intelligent vending machines were we would like to buy the wine we desire, based on aromas, color or price.

Without any further introduction we are going to squeeze every bit of information out of the dataset we chose to work with, [Wine Reviews from Winemag](https://www.kaggle.com/zynicide/wine-reviews). It is a rich dataset and we are going to deeply understand it and extract complex features and conclusions.

In the next chapters we are going to search and crunch our computers through the data. Also, we are determined to test a large variety of learning algorithms based on different perspectives, such as Logistic Regression, Support Vector Machines (SVM), Regressors, Random Forests, etc.