0.1 Processing and Cleaning the Data

In order to make predictions the authors decided that the PdDistrict would suffice as location information, so the address could be dropped. The description and resolution of the crime and arrest were also unnecessary to make a description of the type of crime. The authors kept the columns dates, category, DayOfWeek, PdDistrict, and the coordinates X and Y. To be able to work with the date the datatype was changed from String to datetime. The authors also assigned for each of the 39 categories a number in order to........