

# Big Data Analytics: London Crime Data Analysis

Gianmarco Ricciarelli<sup>1</sup>

<sup>1</sup>University of Pisa,  
gianmarcoricciarelli@gmail.com

# Overview

1 Introduction

2 Data Understanding

# Introduction: the analysis' purpose

To discover the patterns among the criminal activities in the London metropolitan area in a distinct window of time.

# Introduction: the dataset(1)

**London Crime Data, 2008-2016:** this dataset, hosted by **Kaggle**, is composed by 13 millions rows describing the London metropolitan area's criminal activities by *Borough*, *Category*, *Month* and *Year* in a window of time that ranges from January 2008 to December 2016.

# Introduction: the dataset(2)

The dataset is composed by 7 variables:

- **lsoa\_code**: code for Lower Super Output Area in Greater London;
- **borough**: common name for London borough;
- **major\_category**: high level categorization of crime;
- **minor\_category**: low level categorization of crime within major category;
- **year**: year of reported counts, 2008 – 2016;
- **month**: month of reported counts, 1 – 12;
- **value**: monthly reported count of categorical crime in given borough;

# Introduction: the dataset(3)

The variables *lsoa\_code*, *borough*, *major\_category*, *minor\_category*, *year* and *month* are **categorical** variables, while *value* is a **discrete numerical** variable.

# Data Understanding: Numeric Variables' Analysis