THE ULTIMATE GUIDE TO DATA MANIPULATION WITH R AND PYTHON

Zakaria Al Azhar, bigdatahabits.com

2017-11-13

Contents

Chapter 1

Klm

Chapter 2

Introduction

Before you can really mine your data for insights you need to clean it up. Even though it's always good practice to create a clean, well-structured data set, sometimes it's not always possible. Data sets can come in all shapes and sizes (some good, some not so good!), especially when you're getting it from the web.

Data manipulation refers to a set of skills of changing data in an effort to make it easier to read or be more organized, with the eventual goal to get and present insight.

In this book we cover the most used data manipulation formulas, and show how to implement the formulas with R and Python. The formulas are classified in the following sections:

- **Column formulas**
  - *Binning*: grouping data in intervals
  - *Convert and Replace*: casting datatypes, converting, replacing and column renaming.
  - *Filter*: Inclusion, Exclusion, Selecting, and Searching.
  - *Split and Combine*: Splitting, Aggregating, Combining, Merging, Joining, and Appending
  - *Transform*: Converting, Comparing, Resorting, Lagging, Missing Value, Normalizing/Denormalizing, One To Many, Many to One, String manipulation, Subsetting.
- **Row formulas**
  - *Filter*: Row filter, and Data Splitting
  - *Transform*: Concatenate, Group by, Ungroup, Partitioning, Pivoting, Unpivioting, Sampling, and Sorting.

Chapter 3

Prerequisites

3.1   Basic knowledge

3.2   Dataset

3.3   Python and R packages

Chapter 4

Data Exploration

4.1 Data Structure

4.2 Data Summary

Chapter 5

Column Formulas

5.1 Data Binning

5.2 Convert & Replace

5.2.1 Category to Number

Chapter 6

Final Words and this is strange

Chapter 7

Placeholder

Bibliography