SMM635 Mid-Term Project Description

Simone Santoni

Abstract

This document illustrates i. the scope of SMM635's MTP, ii. the dataset to be used to address the data visualization problem, iii. the template that the students shall adopt to present their solution to the data visualization problem.

1 MTP project scope

Carrying out a thorough exploratory data analysis (EDA) is vital for any business analytics project. A well designed and executed EDA can help us spot data quality issues, isolate unexpected relationships, and even generate important insights about the business problem at hand.

Choose the dataset you prefer from sources like Kaggle[1], London Datastore[2], and r/datasets[3] and produce three plots that emphasize:

- Quality issues, intrinsic weaknesses, or idiosyncratic aspects of the chosen dataset
- The novelty and insights emerging from your EDA
- The implications of your EDA for the analysis of a business phenomenon (e.g., possible follow-up data gathering activities or data analysis opportunities)

You can decide to have three plots dealing with the same point (e.g., data quality issues) or dealing with diffents points (a data quality issue, some novelty, and possible business analytics implications).

Any visual form is acceptable.

2 Data

Ideally, the chosen data should deal with a business phenomenon. The adoption of datasets including time-series and geo-referenced information is discouraged. We will deal with these subjects in the second half of the module.

3 Mid-term project submission template

The uploaded materials shall adhere with the following document template:

1. Plot 1 (max 800 words). Describe the fundamental aspects of the adopted visual form and the relevance of your plot in terms of quality issues, novelty, or business analytics implications
2. Plot 2, *ibidem*.
3. Plot 3, *ibidem*.

You are warmly invited to make commit to a few design principles and apply them consistently across the three plots.

[1] https://www.kaggle.com/
[2] https://data.london.gov.uk/
[3] https://www.reddit.com/r/datasets/