On the use of R for building a responsible data science workflow in the retail industry

Hinda Haned

HR Analytics group, Ahold Delhaize

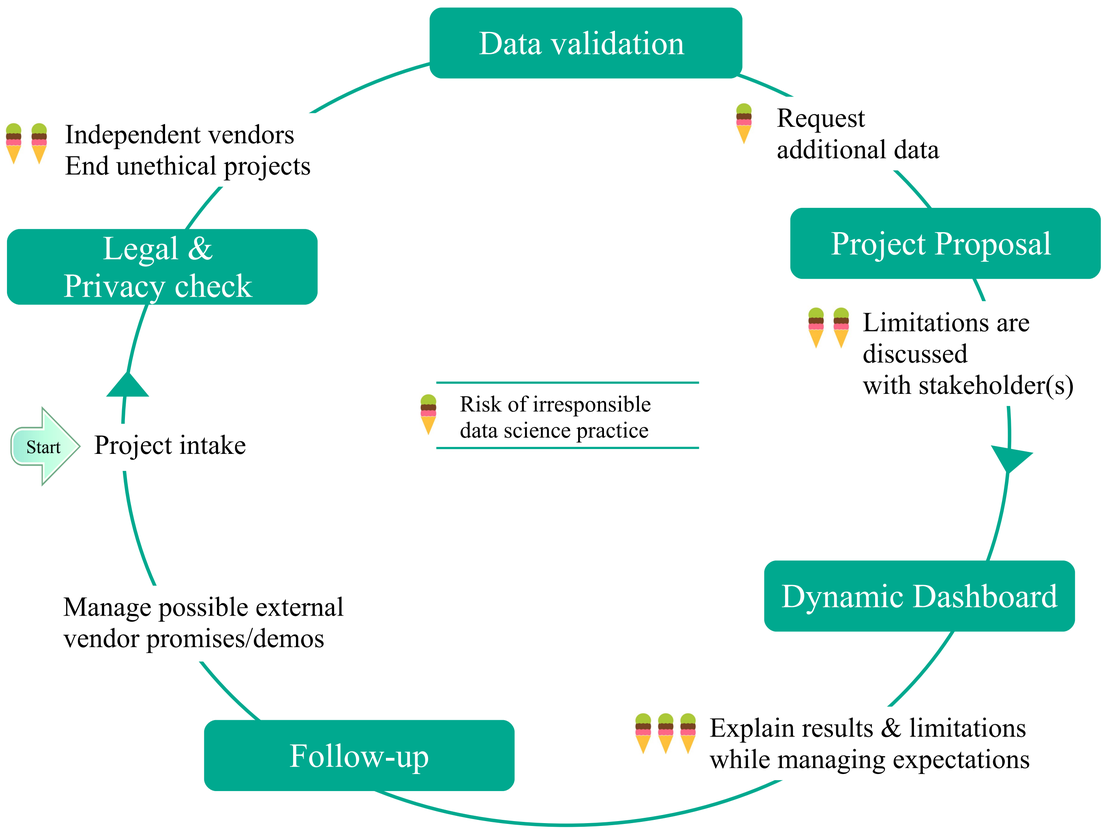
Summary

Analyzing data to answer human resources (HR) related questions raises important questions on the accountability of data scientists:

- How can we balance responsible data science practices in applying machine learning algorithms and (pressure) of utility for business?
- Are there sensitive data sources that should not be collected and used? Where can we (or shall we) draw the line?
- How do we validate our results in a reproducible way?
- How can we communicate results to businesses in a transparent and interpretable way?

In this poster, we illustrate how R is helping us in solving some of these questions in the context of an HR analytics team of an international food retailer with 6,500 stores worldwide and More than 375,000 associates working in Europe, USA and Southeast Asia.

Our project workflow

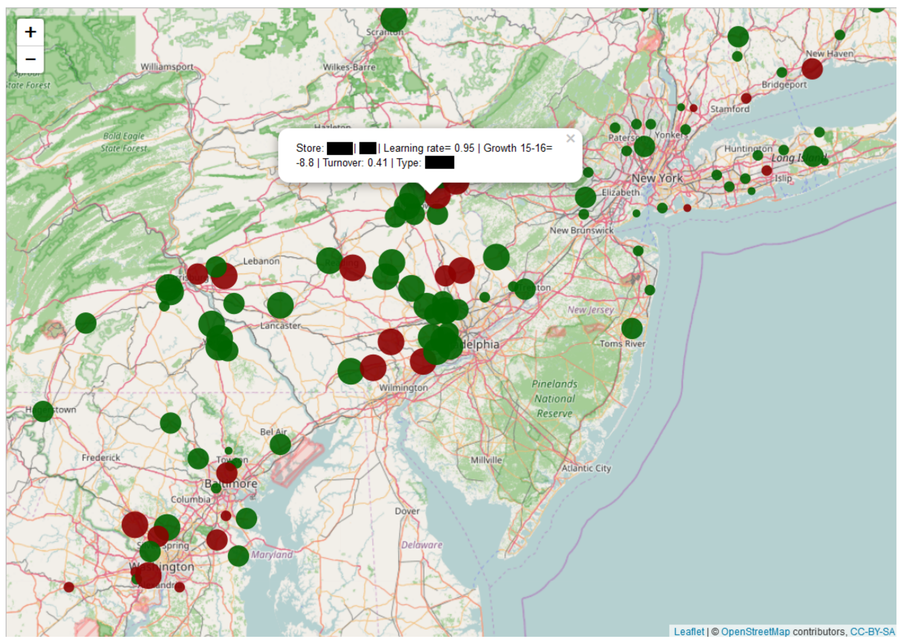


How R is helping us out

Markdown/ Flexdashboard	• Dynamic user-friendly dashboards • Reproducible and automated analyses • Independent code verification and auditing
Shiny apps	• Showcase novel opportunities from open data sources
Specialized packages	• ML packages for forecasting store performance (e.g. XGboost, caret, e1071) • Large scale survey data validation and exploration (e.g. lavaan.survey, semPlot, multilevel) • Natural language processing (e.g. NLP, LDA, qdap)

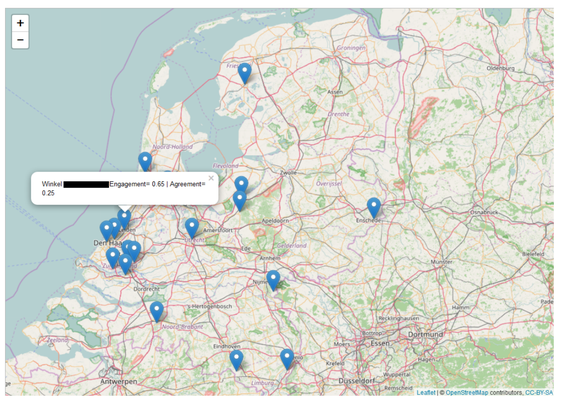
Example 1: Store performance

The visualized map is part of an automated HTML (flex)dashboard that displays information about stores sales growth as a function of learning efforts. We introduced a nuanced measure of performance that includes information about local conditions (e.g. associate turnover, type of store). The users can zoom in and out to view details of each store.



Example 2: Associates feedback

The visualized map below is part of an automated HTML dashboard summarizing our annual survey data. The user can select a store and display survey data along with nuanced measures of performance such as agreement scores. This contributed to a fairer store comparison. Wordclouds summarizing the main suggestions from store associates are also included.



Example 3: Open data

This Shiny app was used in the context of a diversity and inclusion project within the company. It helped illustrate the possibilities of exploring open data on international funding towards gender equality efforts.

`https://hindantation.shinyapps.io/genderequality_2017/`