The Groupon Impact Report is a web service designed to provide Groupon’s merchant partners with in-depth analytics pertaining to the performance of their Groupon campaigns. A suite of metrics presented in an easy-to-understand website helps merchant partners gain insights into many aspects of their businesses, including customer acquisition, customer retention, and return on investment (ROI). Launched in January 2013, the Impact Report currently serves over 80,000 of Groupon’s North American merchant partners.

The engine performing the data analysis for the Impact Report is implemented in R. We chose to develop in R because of 1) the ease and speed of development for data scientists, 2) ability to interface with SQL databases and mash-ups with different data sources, and most importantly 3) excellent support of statistical and machine learning tools. The production R-engine queries a database built from a variety of sources for the necessary data. After computing the metrics, the results are written to a database dedicated to serving the metrics via a REST API used by internal clients. The R-engine processes a large amount of data, computing and updating nearly 24 million metrics on a daily basis. Since metric computation for each merchant is separate and independent, parallelization can be achieved by running multiple concurrent calculations. This makes the system easily scalable.

The foundation of the Impact Report is Groupon’s merchant data. Purchase and coupon redemption data provide insights into customer demographics and acquisition. Post-redemption surveys measure customer satisfaction and loyalty. In-store transaction data track customer behavior and provide an estimate of the overall financial impact. Leveraging this powerful data set, the Impact Report presents merchants with a comprehensive view of their Groupon campaigns and beyond. Another major component of the Impact Report is a model that predicts metrics for merchants with missing or incomplete input data. Regression models are trained for a number of metrics using as input historical data and merchant features such as business category and merchant quality. This predictive model makes it possible to expand coverage of the Impact Report to nearly 100% of Groupon’s North American merchant partners.

The future of the Impact Report lies in real-time applications. It is desirable to update metrics as soon as new data arrive. Another interesting application is to allow users to interact with the data, for example, by adjusting inputs specific to their businesses. Implementing R in the backend of a real-time web application will present unique challenges, especially in the areas of performance and web integration.