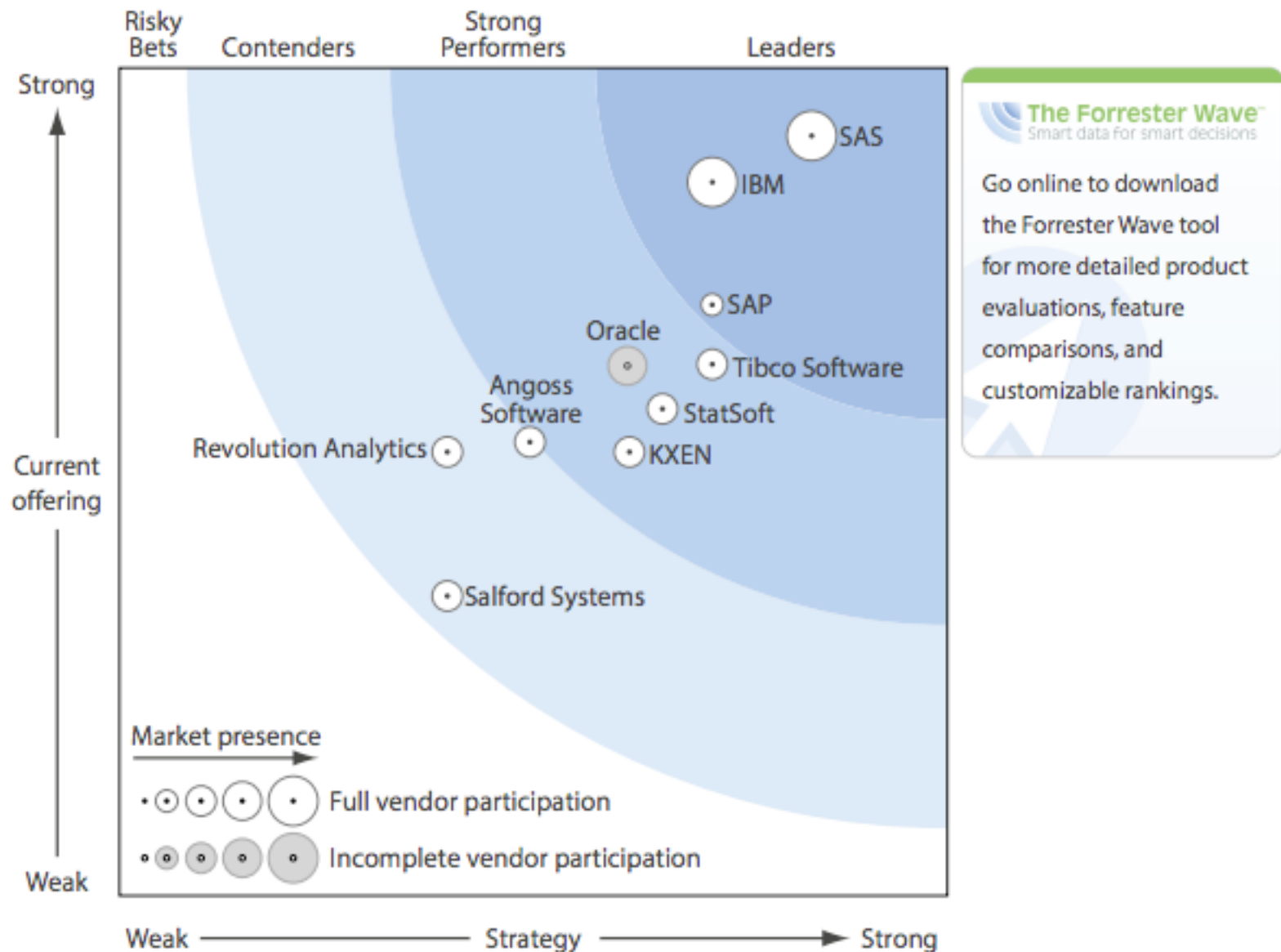# TruQua

Demo: Using Predictive Analytics to Increase Profits at Rosie's Lemonade Stand

Figure 3 Forrester Wave™: Big Data Predictive Analytics Solutions, Q1 '13

Source: Forrester Research, Inc.

Predictive Analytics Library

- Predictive analysis algorithm – seven data mining categories
    - Clustering
    - Classification
    - Association
    - Time Series
    - Preprocessing
    - Social Network Analysis
    - Misc.

R Scripting

- R is an open source programming language and environment for statistical computing.

- R is embedded in SAP HANA SQL code (RLAN procedure)
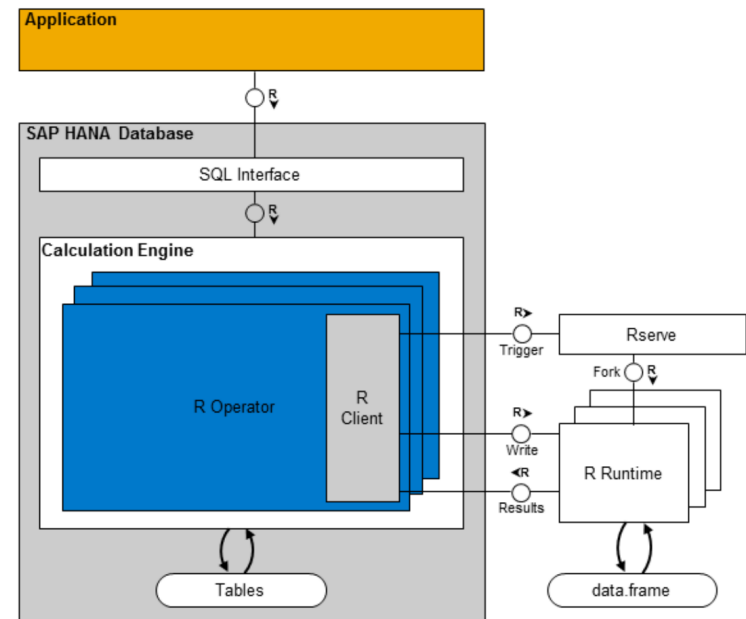
http://prezi.com/otfregblbwmt/lemon/?utm_campaign=share&utm_medium=copy

Predictive Model

Predictive Method

$$\text{SALES}_t = \alpha + \beta_1 \text{AverageTemp}_t + \beta_2 \text{PROMO}_t + \beta_3 \text{WEEKEND}_t + \beta_4 \text{EVENT}_t$$

```
Call:
lm(formula = SalesUnits ~ ActualMaxTemp + Promotion + Event +
    Weekend)

Residuals:
     Min       1Q   Median       3Q      Max
-1.87145 -0.31657 -0.05559  0.35901  2.53100

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)    2.155332   0.757134   2.847   0.0053 **
ActualMaxTemp  0.080491   0.009357   8.602 7.26e-14 ***
Promotion      1.740285   0.186982   9.307 1.89e-15 ***
Event          1.317993   0.209077   6.304 6.63e-09 ***
Weekend        2.480638   0.181068  13.700  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8366 on 107 degrees of freedom
Multiple R-squared:  0.8255,     Adjusted R-squared:  0.819
F-statistic: 126.6 on 4 and 107 DF,  p-value: < 2.2e-16
```

Least Squares

The method of least squares is a standard approach to the approximate solution of over determined systems, i.e., sets of equations in which there are more equations than unknowns. "Least squares" means that the overall solution minimizes the sum of the squares of the errors made in the results of every single equation.

Linear least squares [edit]

Main article: Linear least squares

A regression model is a linear one when the model comprises a linear combination of the parameters, i.e.,

$$f(x, \beta) = \sum_{j=1}^{m} \beta_j \phi_j(x)$$

where the functions, ϕ_j, are functions of x.

Letting

$$X_{ij} = \frac{\partial f(x_i, \boldsymbol{\beta})}{\partial \beta_j} = \phi_j(x_i).$$

we can then see that in that case the least square estimate (or estimator, in the context of a random sample), $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T \boldsymbol{y}.$$

For a derivation of this estimate see Linear least squares (mathematics).

(Source: Wikipedia)

Realtime Predictive v. Predictive

Predictive Method

$$\text{SALES}_t = \alpha + \beta_1 \text{AverageTemp}_t + \beta_2 \text{PROMO}_t + \beta_3 \text{WEEKEND}_t + \beta_4 \text{EVENT}_t$$

Real-Time Method

$$\text{SALES}_t = \alpha + \beta_1 \text{ActualTemp}_t + \beta_2 \text{PROMO}_t + \beta_3 \text{WEEKEND}_t + \beta_4 \text{EVENT}_t + \beta_5 \text{SALES}_{t-1}$$