

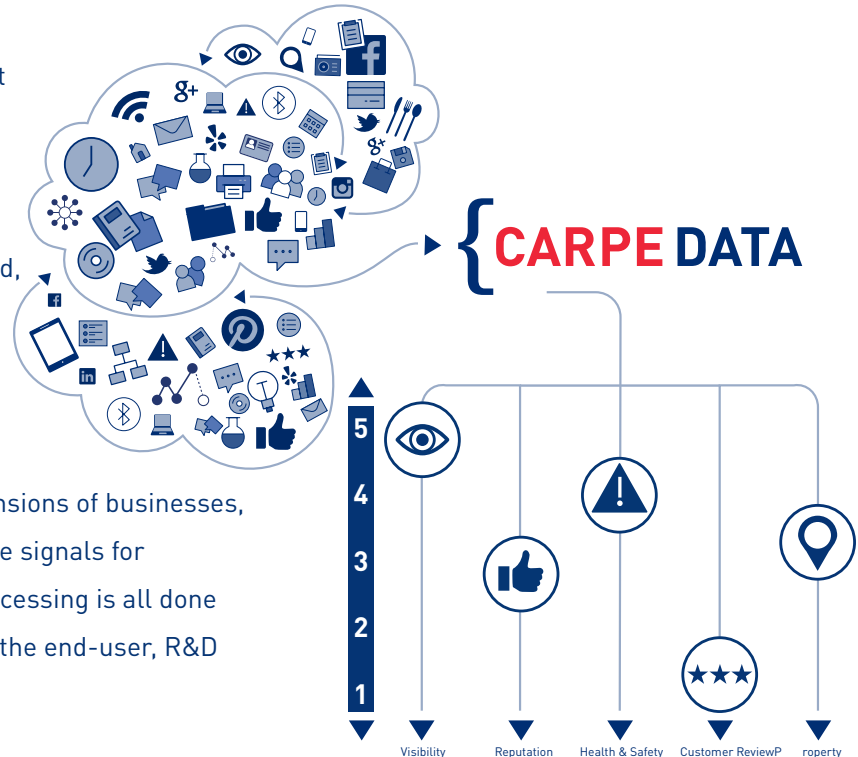
{ **NEXT GENERATION** INDEXES



Harnessing Alternative & Emerging Data

Add a more predictive dimension of data to your risk assessment profile.

Carpe Data Indexes are distilled views of non-traditional data points such as customer sentiment and on-line visibility for businesses from our massive inventory of data elements. Primarily derived from online content, the Indexes clean and normalize highly informative but often unstructured, biased and inconsistent raw data into clear, consumable insurance data points.



The Indexes are designed to capture various dimensions of businesses, blending longer-term, stable signals with real-time signals for dynamic, rapid reflection of risk exposure. The processing is all done systematically and in advance, so upon delivery to the end-user, R&D can start immediately.

Why Indexes?

Big Data is Unwieldy

Everyone wants to leverage Big Data but as the volume and variety grows, the need to streamline its processing and storage for reporting, modeling and consuming for real time decisioning becomes more critical.

Raw, On-line data needs refinement

Data from on line sources ranging from reviews to As the volume and variety of online alternative and emerging data grows, many challenges arise with analyzing this data:

Carpe Data Indexes solve these challenges by providing the missing link between raw data and inputs ready for models. We distill our massive inventory of data into highly explanatory and predictive sources of risk information.

How Indexes Work

Businesses are measured in each Carpe Data Index to show its normative value, enabling it to be compared to:

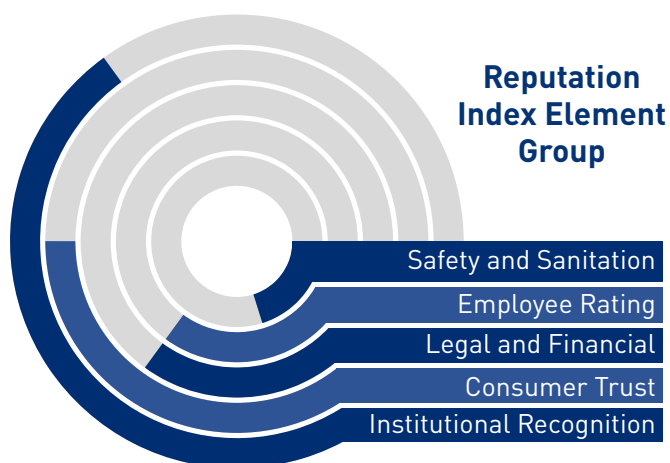
- Businesses within industry segment and location cluster
- The segment and location cluster average
- The measure of the same business over time

Key Concepts

Index Values

Index values are measured to the tenths place and range from 1.0 through 5.0, where 5.0 means the highest value attainable for that dimension, given its segment and subpopulation.

The value itself does not imply the level of risk. For example, the lowest visibility value does not mean highest risk, but it could. The lowest health & safety value, on the other hand, very well could mean highest risk. The interpretation will depend on the dimension and book of business, among other possible considerations.



Required Fill Rate and No Hit

Not every business is found by our search algorithm, and even if it is found, there may be insufficient data to confidently report an index value. Each index has an associated required fill threshold or required percentage of data elements that need to be available to return an Index value.

Example: Customer Review Index for a Restaurant

ACME, Inc. has customer reviews of “★”, “★”, “★★”, “★★★★★”, “★★★★★” or an average of 2.8

We compute the average review for each business in the segment and then create a 1-5 scale:

- The bottom 20% of restaurants have average review < 2.1
- The top 20% of restaurants have average review ≥ 4.2

ACME, Inc. has average review = 2.8, which maps to index value = 2

Data Element Groups

Index is composed of broad types of risk within a given dimension represented by that specific index. Those groups need to be articulated and represented with actual data.

Element groups capture this notion, and an example is the groups shown in the Reputation Index to the right.

The data elements within each Element Group are specific pieces of information used in index calculation. Our data arrives at all levels of structure including text, images, and numeric values. Pre-processing is implemented by a custom set of workflows considering data type, range of values, and directionality. For example, the number of visits to a website over some time horizon could take any non-negative value, while a review could be 1-5 stars. Normalization is applied to handle these issues.

Reason Codes

If an index value is less than a threshold value of 3.0, any gone or more reason codes will be returned with the score. These codes will indicate which data element groups were below the threshold, for transparency purposes.

Versioning

Each version of an index is stored and available. Clients can receive one or more versions of an index, for example if they require version 1.1 in production but want to test a newer version 1.2. Carpe Data tracks reasons behind all version changes, both minor and major. Minor changes (reflected by a change in the tenths place of a version number as in 1.1 to 1.2) could include recalibrating to newer data. Major changes (reflected by a change in the ones place of a version number as in 2.0 to 3.0) could include a methodology change.

Clustering

A restaurant in a large city may be reviewed differently than a restaurant in a small town. Our Indexes are constructed to control for business subpopulations within segment. Extensive analysis determined businesses should be clustered by relevant data elements, as opposed to just geography. Two locations separated by 2000 miles may have more similar attributes than two locations separated by 20 miles. The concept of geography is used by indexes, and geography is reported to customers (e.g., United States), but internally there is a finer analysis of the business location using clustering.

Applications

Features in a model

The indexes can be inputs to various insurance model (e.g., probability of claim, expected loss, loss ratio, pricing, etc). We recommend including multiple indexes together, as they may measure different sources of risk. A business with a low customer review value and high visibility value could be a larger risk than a business with the same customer review value but low visibility. A restaurant with one hundred 1-star reviews is likely to be worse than a restaurant with two 1-star reviews.

Monitoring trends

Monitoring index values in your book over time can help show changes in business attributes. This can produce a red flag on an existing insured or help define a target business.

