dun & bradstreet

Understanding Global Sales & Employee Models

This document is intended to address the following questions:

- What do the Global Sales & Employee Models predict?
- What are availability rules?
- What is the model development process?
- What is the model performance?
- What are the key attributes used by the models?

I. INTRODUCTION

The Dun & Bradstreet Global Sales & Employee models predict Sales and Employees Total for unlinked businesses in Dun & Bradstreet Data Cloud that do not contain actual sales and employees figures. These models help our customers evaluate the capacity of a business to purchase their product or services based on its size.

Unlinked business entities comprise of businesses that are defined as Single Locations in Dun & Bradstreet Data Cloud. They do not have majority ownership or linkage relationships (e.g. headquarter, parent, branches or subsidiaries) with other entities. They may be minority owned by another business, joint ventures, franchises, dealers or agents. A business considered as Single Location in the Data Cloud does business only at the location captured by the specific D-U-N-S Number®. There are over 165 million Single Locations in the Data Cloud, representing 90% of the entities in the overall population.

Our proprietary DUNSRight® Quality Process ensures the integrity of the information contained in our Data Cloud. DUNSRight is our process for collecting and enhancing information. Our expert team of statisticians and economists lead the development of our Analytics solutions, the fifth and final component of the sequential DUNSRight process, and are responsible for turning our vast commercial database into actionable business insight. The Global Sales & Employee models is the latest in our suite of standard analytical offerings in the sales and marketing space.

The Global Sales & Employee models are highly effective in predicting the size of a business where actual sales and employees figures are not reported. The models utilize advanced modeling techniques and provide a more reliable and accurate prediction of business size. They enhance data completeness and allow customers to engage in large scale target and outreach activities.

The models will help customers with:

- Customer Portfolio Segmentation
- Prospect Identification & Prioritization
- Demand Estimation for Products and Services
- Market Sizing
- Sales Territory Planning
- Identify largest businesses in a market segment or industrial sector
- Understanding the geographic distribution of a business' employees and/or sales

This document explains in detail the Global Sales & Employee models that are applied to Single Locations (unlinked records) in the Data Cloud. Dun & Bradstreet will be updating the underlying models for - linked records in a future release.

II. WHAT DO THE SALES & EMPLOYEE MODELS PREDICT?

The Sales & Employee models provide a statistical valid calculation of:

- Sales, which represents annual sales or revenue of a business in U.S. Dollars.
- Employees Total, which represents total number of employees for a business entity

Dun & Bradstreet also populates a field called Employees Here, which represents number of employees for a business entity at a specific site or location. For Single Locations, Employees Here value will be equal to Employees Total.

III. AVAILABILITY OF THE MODELED GLOBAL SALES & EMPLOYEE VALUES

Modeled Sales & Employee values are available on the majority of the Dun & Bradstreet active universe, except for the following circumstances:

- Business records that already have reported actual sales and employee figures
- Businesses that are considered as "Out of Business"
- Businesses with unclassified industry. These records will not receive modelled Sales figures, but will receive employee figures.

IV. MODEL DEVELOPMENT PROCESS

Dun & Bradstreet has developed a suite of statistical models to provide modeled Sales and Employee values on businesses operating in over 220 countries.

The model development process involved two phases. In phase 1, clusters of countries with similar characteristics within the same continent were identified. These clusters enabled to capture homogeneity of variables across countries, and as a result were modeled together. In phase 2, separate models were created for Sales and Employee values by assessing records with actual reported Sales & Employees values against various Dun & Bradstreet data elements such as Firmographics, Financials, Payment performance and Signals. Appendix A provides list of some of the data elements used by the models. Each of these models in turn are made up of country-specific, cluster of countries or continental type of models.

- *Country-specific:* For mature markets where the number of records with actual Sales & Employees figures are robust, country-specific models were developed.
- *Cluster of Countries:* If the actual collected data did not allow the development of country specific models, then of two or more countries within the same region/continent were clustered and a model developed specific to the cluster.
- *Continental:* For markets where country-specific or cluster of countries models did not provide strong predictive performance, continent specific models using all available actual reported data of that continent was applied.

The Country-specific and Cluster of Countries models were based on regression techniques, such as Linear Regression, Quintile Regression, B-Spline Regression, and Multivariate Adaptive Regression Splines. The continental models were based on decision trees modelling technique (CHAID).

A sample of records with actual employees or actual sales as of December 2014 was selected from Dun & Bradstreet Data Cloud - 80% of this sample was used for model development while the remaining 20% was used as a holdout sample for model validation. The models were built using the underlying relationship between sales and employees with industry classification playing a critical role.

Amongst all competing models, the one that delivered the best accuracy results was selected for a specific market(s). Dun & Bradstreet has developed a suite of 38 models. Appendix B includes a list of markets with the corresponding model type.

V. MODEL PERFORMANCE

One way to measure model performance is by examining the association between modeled sales and employees and the actual reported sales and employees' figures. The accuracy of these models vary from market to market, and is dependent on the underlying data depth and quality. Table 1 and 2 capture the above mentioned association between the actual figures and modeled figures.

Table 1. Association between modeled sales range and actual sales range.

SALES BAND	% OF VALIDATION SAMPLE
Exact Band	63.5%
Under by 1 Bannd	11.8%
Under by 2 Bands	5.8%
Under by 3 or more Bands	2.3%
Over by 1 Band	9.6%
Over by 2 Bands	4.5%
Over by 3 or more Bands	2.4%

The following sales range bands were created to measure modeled sales against actual sales: <500K, 500K-1M, 1M-2.5M, 2.5-5M, 5-10M, 10-25M, 25-50M, 50-100M, 100-500M, >500M.

Interpretation: The validation table indicates that there is a strong correlation between modeled sales and actual sales across all levels. 63.5% of all records had the same modeled sales and actual sales range. 11.8% of the records had modeled sales range that was under by one sales band. That is, if actual sales ranged between 1 and 2.5 million, modeled sales ranged between 500K to 1 million. 9.6% had modeled sales ranged that was over by one sales range. That is, if actual sales ranged between 1 and 2.5 million, modeled sales ranged between 2.5 and 5 million.

Table 2. Association between actual employee ranges and modeled employee range

SALES BAND	% OF VALIDATION SAMPLE
Exact Band	76.9%
Under by 1 Bannd	10.2%
Under by 2 Bands	1.6%
Under by 3 or more Bands	0.70%
Over by 1 Band	9.9%
Over by 2 Bands	0.70%

The following sales range bands were created to measure modeled sales against actual employees: 1-5, 6-25, 26-50, 51-100, 101-250, 251-500, 501-1000, 1000+

Interpretation: The validation table indicates that there is a strong correlation between modeled employees and actual employees across all levels. 76.9% of all records had the same modeled employees and actual employees range. 10.2% of the records had modeled employee range that was under by one employee band. For instance, if actual employees ranged between 6 and 25 employees modeled employees was less than 6 employees. 9.9% had modeled employee range that was over by one band. That is, if actual sales ranged between 1 and 2.5 million, modeled sales ranged between 2.5 and 5 million.

These modeled figures should be viewed as estimates based on observed characteristics associated with business with actual sales and employees. They should not be considered as precise assessment of the size of a business when actual figures are not available.

Additional detailed performance results are captured in Appendix C of this document.

dun & bradstreet

Appendix A

List of the Data Elements in the Total Loss Predictor model

Following are some of the key data inputs used by the scoring models. The data inputs used and the weights of the inputs will vary by scorecard.

MODEL INPUTS	DESCRIPTION
Actual Employees (used only in sales model)	Employee figure collected from a business principle or official source(s)
Actual Sales	Actual sales figures normally collected from financial statements
DEMOGRAPHICS	**DESCRIPTION**
Business Age	Number of years since the current ownership or management assumed control of the business or the years since the business was established if no control change has taken place
Number of Principles	Indicates number of principles associated with the business
Legal Structure	Indicates whether a business entity is a Corporation, Joint Venture, Partnership, Proprietorship, and others)
Industry Classification	Assignment system used to categorize business establishments based upon the type of business activity done by that business at that location
Country	Location where the business is located.
FINANCIALS	**DESCRIPTION**
D&B Rating	The Dun & Bradstreet Rating provides a quick and clear indication of the credit-worthiness of an organization. The first part indicates the size of the company and is referred to as the Financial Strength Indicator or the Rating Classification. This component indicates the size of the company and is based on net worth, issued capital or number of employees. The second part estimates overall credit standing of the company or predicts the likelihood of failure. In some markets it is known as the Risk Indicator, while in others it is known as the Code Condition or the Composite Credit Appraisal. In markets with Failure Score, it is derived from it.
Net Worth	Location where the business is located
Profit and Loss	Indicates whether business has generated income or incurred loss for a specific time period.
SIGNALS	**DESCRIPTION**
Export Import Indicator	Indicates if the business entity imports, exports, or Both.
PAYMENT DATA (USED ONLY IN US AND UK)	**DESCRIPTION**
Number of payment experiences	Number of Trade experiences for a specific D-U-N-S Number

dun & bradstreet

Appendix B

List of markets by Scorecard

SALES MODEL DEVELOPMENT		
CONTINENT	MODEL TYPE	COUNTRY NAME
Asia Pacific	• Country-specific • Cluster of Countries • Cluster of Countries • Continential	• Japan • China • Australia, New Zealand • Others
North America	• Country-specific • Continental	• USA* • Canada
Latin American	• Cluster of Countries • Cluster of Countries • Continental	• Colombia • Others • Brazil, Mexico
Europe	• Cluster of Countries • Country-specific • Cluster of Countries • Country-specific • Cluster of Countries • Cluster of Countries • Continential • Country-specific	• Belgium, Poland, Denmark, Norwary, Ireland, Sweden, Netherlands • Germany • Finland, Italy, Spain • France • Hungary, Czech Republic, Portugal • Slovakia • Others • UK
Middle East	• Continential	• Applicable to all countries
Africa	• Continential	• Applicable to all countries

*Represents US Sales and Employee model developed in 2015

SALES MODEL DEVELOPMENT		
CONTINENT	MODEL TYPE	COUNTRY NAME
Asia Pacific	• Country-specific • Country-specific • Cluster of Countries • Cluster of Countries • Cluster of Countries • Continential	• Japan • India • Thailand • Hong Kong, Korea • Singapore • Others • China, Taiwan, Australia, New Zealand
North America	• Country-specific • Continental	• USA* • Canada
Latin American	• Continental	• Applicable to all countries
Europe	• Cluster of Countries • Cluster of Countries • Country-specific • Cluster of Countries • Cluster of Countries • Cluster of Countries • Continential • Country-specific	• Netherlands, Bulgaria • Switzerland, Czech Republic • Germany • Norway, Belgium, Spain • France, Sweden • Hungary, Poland, Portugal • Others • UK
Middle East	• Continential	• Applicable to all countries
Africa	• Continential	• Applicable to all countries

*Represents US Sales and Employee model developed in 2015

dun & bradstreet

Appendix C

The following table captures actual versus predicted Sales for all sales range bands in table.

ACTUAL SALES RANGE	<500k	500k-1m	1m-2.5m	2.5m-5m	5m-10m	10m-25m	25m-50m	50m-100m	>100m
<500k	56.45%	8.49%	4.75%	1.04%	0.47%	0.16%	0.03%	0.01%	0.00%
500k-1m	5.44%	2.87%	1.88%	0.44%	0.24%	0.11%	0.01%	0.01%	0.00%
1m-2.5m	3.11%	2.18%	2.89%	0.87%	0.35%	0.19%	0.03%	0.01%	0.00%
2.5m-5m	0.97%	0.64%	1.23%	0.70%	0.32%	0.17%	0.04%	0.01%	0.00%
5m-10m	0.42%	0.25%	0.50%	0.44$	0.33%	0.19%	0.05%	0.02%	0.01%
10m-25m	0.20%	0.10%	0.21%	0.22%	0.22%	0.19%	0.06%	0.02%	0.01%
25m-50m	0.04%	0.02%	0.04%	0.04%	0.05%	0.07%	0.03%	0.01%	0.01%
50m-100m	0.01%	0.01%	0.01%	0.01%	0.02%	0.03%	0.02%	0.01%	0.01%
>100m	0.00%	0.00%	0.00%	0.00%	0.00%	0.01%	0.01%	0.01%	0.00%

The following table captures actual versus predicted employees for all employee ranges bands in table.

ACTUAL EMPLOYEE RANGE	1 to 5	6 to 25	26 to 100	101 to 250	251 to 500	501 to 1000	1000+
1 to 5	67.21%	8.65%	0.63%	0.03%	0.01%	0.00%	0.00%
6 to 25	7.05%	8.68%	1.12%	0.06%	0.01%	0.00%	0.00%
26 to 100	0.92%	2.87%	0.94%	0.09%	0.01%	0.00%	0.00%
101 to 250	0.14%	0.55%	0.28%	0.06%	0.01%	0.00%	0.00%
251 to 500	0.05%	0.20%	0.11%	0.03%	0.01%	0.00%	0.00%
501 to 1000	0.02%	0.10%	0.05%	0.01%	0.01%	0.00%	0.00%
1000+	0.01%	0.05%	0.02%	0.01%	0.00%	0.00%	0.00%

Additional model accuracy measures are outlined in the below table.

SALES	EUROPE	ASIA	L.AMERICA	N.AMERICA	MIDDLE EAST	AFRICE
Exact Bands	60%	30%	32%	63%	40%	39%
+/- 1 Bands	84%	67%	63%	89%	80%	82%
+/- 0% Errors	1%	1%	1%	1%	1%	2%
+/- 10% Errors	7%	7%	6%	9%	10%	12%
+/- 25% Errors	18%	17%	14%	21%	23%	27%
+/- 50% Errors	37%	36%	31%	42%	46%	52%
Positive Predicted Value	66%	79%	77%	78%	85%	84%
True Positive Rate	42%	50%	49%	59%	64%	72%
Somers' D	55%	50%	53%	53%	63%	69%
Spearman Correlation	67%	64%	67%	71%	80%	84%

EMPLOYEES	EUROPE	ASIA	L.AMERICA	N.AMERICA	MIDDLE EAST	AFRICE
Exact Bands	84%	43%	69%	79%	79%	79%
+/- 1 Bands	94%	80%	81%	94%	91%	91%
+/- 0% Errors	44%	10%	31%	1%	28%	28%
+/- 10% Errors	44%	11%	31%	5%	28%	28%
+/- 25% Errors	47%	23%	32%	13%	29%	29%
+/- 50% Errors	57%	44%	44%	27%	38%	38%
Positive Predicted Value	76%	63%	54%	61%	64%	64%
True Positive Rate	13%	25%	13%	54%	15%	15%
Somers' D	40%	34%	19%	45%	35%	50%
Spearman Correlation	50%	45%	28%	53%	56%	66%

- Diagonal and +/- 1 Bands which shows the magnitude of shift in predefined bands for modeled values
- Positive Predicted Value (PPV) which represent how "Precise" the model is, providing the extent to which the prediction reflects the true values. True Positive Rate (TPR) which assess the "Sensitivity" of the model, providing the proportion of positives that are correctly identified as such. Record Level Accuracy
- Error Rate (Error Rate = (Predicted – Actual) / Actual) which measures the deviation of the modelled values away from the actual reported figures

Ranking Capability at band level

- Somers' D
- Spearman Correlation

dun & bradstreet

ABOUT DUN & BRADSTREET

Dun & Bradstreet, a leading global provider of business decisioning data and analytics, enables companies around the world to improve their business performance. Dun & Bradstreet's Data Cloud fuels solutions and delivers insights that empower customers to accelerate revenue, lower cost, mitigate risk, and transform their businesses. Since 1841, companies of every size have relied on Dun & Bradstreet to help them manage risk and reveal opportunity. Twitter: @DunBradstreet