dun & bradstreet

Understanding Dun & Bradstreet Sales & Employee Models

THIS DOCUMENT IS INTENDED TO ADDRESS THE FOLLOWING QUESTIONS:

– What do the Dun & Bradstreet 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?

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

I. INTRODUCTION

The Dun & Bradstreet Sales & Employee models predict Sales and Employees Total for businesses without actual sales and employee figures in Dun & Bradstreet Data Cloud. These models help our customers evaluate the capacity of a business to purchase their product or services based on its size and are applied on Linked and Stand-alone businesses without actual sales and employee values. These models do not apply to locations classified as Branch in the Data Cloud.

The Dun & Bradstreet Data Cloud delivers the world's most comprehensive business data and analytics and provides unparalleled depth and breadth of business information that can be used to help improve our customers' business performance. From the Data Cloud we derive our Live Business Identity, which delivers comprehensive and continually updated view of any company in the Data Cloud. The Sales & Employee models are the latest in our suite of standard analytical offerings in the sales and marketing space.

The 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.

II. WHAT DO THE SALES & EMPLOYEE MODELS PREDICT?

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

- Sales, which represents annual sales or revenue of a business in US Dollars and are based on in-depth information that Dun & Bradstreet has on the business, its industry and peers. For Linked businesses, the Sales value represents the figures of that site as well as all the sites reporting into it. If the sales value is calculated in local currency, it is converted into US dollars using the exchange rate as of the date of when the model was applied. Modeled Sales values are not available on Branch locations.
- Employees Total, which represents total number of people employed by the business and is based on in-depth information that Dun & Bradstreet has on the business, its industry and its peers. The Employee Total field represents the total employee count for a commercial entity; it is not an aggregation of Employees Here of individual entities reporting into a parent within the corporate family tree. For Stand-alone Locations, the Employees Here and Employee Total counts are represented by the same value.

dun & bradstreet

Note: Due to differences in local market practices and regulations, what constitutes as Employee Total varies market to market (eg. some markets count only domestic employees towards Employee Total whereas in other markets it includes employees located globally). The modeled employee figures are expected to reflect the market dynamics of a specific country.

III. AVAILABILITY OF THE MODELED SALES & EMPLOYEE VALUES

MODELED SALES & EMPLOYEE TOTAL VALUES ARE AVAILABLE ON MAJORITY OF THE ACTIVE BUSINESSES IN DUN & BRADSTREET DATA CLOUD. THESE VALUES ARE NOT AVAILABLE IN 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 except those located in the US
- Businesses that are considered as Branch location in Dun & Bradstreet Data Cloud. A branch (or division) is a secondary location of its headquarters. It is not a separate corporation, has no legal responsibility for its debts, even though bills may be paid from the branch location.

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, Linkage and Signals. **Appendix B** 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 two or more countries within the same region/continent were grouped and modelled together.

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. 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 2016 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 50 models. **Appendix C** 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. Higher the association between actual and modeled values, the more the robust the performance of the model. The accuracy of the Sales and Employee models vary from market to market and is driven by underlying data depth and quality.

Tables 1 and 2 show that the models hold a strong association between the actual figures and modeled figures on the validation sample.

Table 1. Association between modeled sales range and actual sales range on Linked and Stand-alone records in Dun & Bradstreet Data Cloud.

SALES BAND	% OF VALIDATION SAMPLE
Exact Band	60.6%
Under by 1 Band	10.5%
Under by 2 Bands	5.1%
Under by 3 or more Bands	1.7%
Over by 1 Band	12.6%
Over by 2 Bands	6.0%
Over by 3 or more Bands	1.6%

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 the Overall Success Rate which correspond to the observed proportion of correctly predicted instances across all bands. 60.6% of all records had the same modeled sales and actual sales range. 10.5% 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. 12.1% 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.

EMPLOYEE BAND	% OF VALIDATION SAMPLE
Exact Band	80%
Under by 1 Band	5.9%
Under by 2 Bands	2.5%
Under by 3 or more Bands	1.8%
Over by 1 Band	7.6%
Over by 2 Bands	1.7%
Over by 3 or more Bands	0.6%

The following sales range bands were created to measure modeled sales against actual employees: 1-10, 11-25, 26-50, 51-100, 101-250, 251-500, 501-1000, 1000+

Interpretation: The validation table indicates that the Overall Success Rate which correspond to the observed proportion of correctly predicted instances across all bands. 80% of all records had the same modeled employees and actual employees range. 5.9% 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 and modeled employees was less than 6 employees. 7.6% had modeled employee range that was over by one band. That is, if actual employees ranged between 6 and 25 employees and modelled employees ranged between 26 and 50 employees.

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 D** of this document.

APPENDIX A: Detailed explanation of the terminologies used in this document is provided below.

- **Annual sales** value (including all sites reporting into it or below it in their corporate family tree). If the sales value is collected in a local currency, local values are then converted to US dollars using the exchange rate of the date of the financial figures/statement. When the sales figure is not available for the business, D&B models are based on in-depth information that we have on the business, its industry and its peers.
- **Employees Here** is defined as the number of employees at a given physical site of the business.
- **Employees Total** is defined as the total number of people employed by all branches and subsidiaries of a business.
- **Linked businesses** share relationship between different active business entities or specific sites within a corporate family. Linkage occurs in Dun & Bradstreet's Data Cloud when one business location has financial & legal responsibility for another business location. The percentage of financial and legal responsibility determines the type of linkage relationship
- **Stand-alone** businesses are entities which do not have any linkage relationships e.g. headquarter, parent, branches or subsidiaries. It is the only location of that business.
- A **Branch** (or division) is a secondary location of its headquarters. It is not a separate corporation, has no legal responsibility for its debts, even though bills may be paid from the branch location. It will usually have the same legal business name as its headquarters, but can carry out a specific operation related to the headquarters and can even have its own trade style name. It is possible for them to also be located at the same address as the headquarters.

APPENDIX B: 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	
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	
D&B Rating	The D&B 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	
Export Import Indicator	Indicates if the business entity imports, exports, or Both.
Payment Data (Used only in US and UK)	
Number of payment experiences	Number of Trade experiences for a specific Duns
Linkage	Indicates whether the business is a Subsidiary or Headquarter; and if the business is a Global or Domestic Ultimate.

APPENDIX C: List of markets by Scorecard

The below table indicates the type of scorecard applied to individual markets. Scorecard varies by market and based on whether a record is Linked or Stand-alone location.

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

APPENDIX D: Modeled Employee Range

The below table provides the detailed association between modeled and actual sales and employee values.

	PREDICTED								
ACTUAL	1 - 10	11 - 25	26 - 50	51 - 100	101 - 250	251 - 500	501 - 1000	1000+	TOTAL
1 - 10	76.65%	6.36%	1.31%	0.35%	0.09%	0.01%	0.00%	0.00%	84.79%
11 - 25	3.88%	2.15%	0.61%	0.29%	0.06%	0.01%	0.00%	0.00%	7.01%
26 - 50	1.44%	1.12%	0.58%	0.41%	0.10%	0.01%	0.00%	0.00%	3.66%
51 - 100	0.62%	0.56%	0.43%	0.44%	0.14%	0.01%	0.00%	0.00%	2.20%
101 - 250	0.27%	0.28%	0.25%	0.29%	0.16%	0.03%	0.01%	0.00%	1.28%
251 - 500	0.09%	0.11%	0.09%	0.11%	0.10%	0.03%	0.01%	0.00%	0.54%
501 - 1000	0.04%	0.06%	0.04%	0.05%	0.07%	0.03%	0.01%	0.00%	0.31%
1000+	0.03%	0.04%	0.03%	0.03%	0.04%	0.03%	0.01%	0.01%	0.21%
TOTAL	83.02%	10.68%	3.34%	1.98%	0.77%	0.16%	0.04%	0.02%	100.00%

ACCURACY METRICS	GLOBAL	AFRICA	ASIA	EUROPE	MIDDLE EAST	SOUTH AMERICA	NORTH AMERICA
Exact Bands	80%	48%	53%	93%	45%	84%	87%
+/- 1 Bands	93%	83%	85%	97%	80%	95%	97%
+/- 0.1% Error	36%	22%	6%	53%	7%	19%	3%
+/- 10% Error	36%	24%	8%	53%	10%	20%	7%
+/- 25% Error	41%	32%	18%	55%	21%	22%	15%
+/- 50% Error	50%	44%	33%	61%	40%	33%	29%
PPV	56%	60%	54%	59%	63%	63%	59%
TPR	26%	41%	29%	17%	39%	26%	59%
Somers' D	47%	54%	37%	46%	52%	43%	42%
Spearman Corr	49%	61%	41%	48%	59%	45%	46%
Sample size	60,964,118	283,293	17,319,929	36,574,445	324,854	6,461,597	13,364,179

Diagonal and +/- 1 Bands: shows the magnitude of shift in predefined bands for modeled values

Positive Predicted Value (PPV): represents 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) measures the capture rates while allowing +/-X% deviation of the modelled values away from the actual reported figures

Ranking Capability at band level is measured using the combination of the Somers' D and Spearman Correlation

PREDICTED SALES RANGE											
ACTUAL	< 500K	500 - 1M	1M - 2.5M	2.5M - 5M	5M- 10M	10m - 25m	25m - 50m	50m - 100m	100M - 500M	>500M	TOTAL
< 500k	21.97%	7.98%	6.80%	3.40%	0.91%	0.50%	0.06%	0.17%	0.39%	0.12%	42.31%
500k-1m	5.49%	4.52%	3.11%	0.83%	0.23%	0.14%	0.02%	0.02%	0.02%	0.05%	14.42%
1m-2.5m	2.87%	4.07%	5.66%	1.76%	0.70%	0.41%	0.15%	0.09%	0.03%	0.02%	15.75%
2.5m-5m	0.77%	1.12%	2.99%	2.03%	0.85%	0.46%	0.11%	0.08%	0.09%	0.00%	8.50%
5m-10m	0.33%	0.44%	1.23%	1.65%	1.18%	0.85%	0.39%	0.09%	0.06%	0.02%	6.25%
10m-25m	0.15%	0.09%	0.47%	0.68%	0.88%	0.99%	0.91%	0.27%	0.09%	0.05%	4.58%
25m-50m	0.11%	0.00%	0.06%	0.15%	0.23%	0.53%	0.53%	0.29%	0.41%	0.08%	2.38%
50m-100m	0.06%	0.02%	0.02%	0.02%	0.12%	0.20%	0.39%	0.15%	0.77%	0.14%	1.88%
100m-500m	0.15%	0.02%	0.03%	0.06%	0.11%	0.24%	0.33%	0.24%	1.31%	0.29%	2.78%
>500M	0.06%	0.00%	0.00%	0.03%	0.06%	0.11%	0.17%	0.03%	0.38%	0.32%	1.15%
TOTAL	31.97%	18.25%	20.36%	10.62%	5.27%	4.42%	3.07%	1.43%	3.55%	1.06%	100.00%

ACCURACY METRICS	GLOBAL	AFRICA	ASIA	EUROPE	MIDDLE EAST	SOUTH AMERICA	NORTH AMERICA
Exact Bands	62.12%	38.66%	32.12%	66.13%	36.24%	76.56%	67.68%
+/- 1 Bands	84.13%	72.11%	65.88%	86.79%	74.27%	90.32%	89.26%
+/- 0.1% Error	0.47%	0.79%	0.35%	0.50%	0.70%	0.37%	0.75%
+/- 10% Error	5.28%	6.51%	3.92%	5.62%	6.01%	4.16%	8.37%
+/- 25% Error	13.30%	16.62%	9.93%	14.13%	13.88%	10.57%	20.85%
+/- 50% Error	26.89%	34.48%	20.98%	28.38%	28.92%	21.56%	41.18%
PPV	63.35%	62.70%	67.52%	60.15%	84.39%	74.88%	81.16%
TPR	50.43%	65.60%	47.28%	52.73%	66.75%	49.28%	57.55%
Somers' D	53.00%	49.60%	38.61%	53.60%	70.32%	68.79%	60.00%
Spearman Corr	58.16%	57.54%	45.95%	58.17%	82.21%	72.90%	71.00%
Sample size	12,114,397	6,590	1,679,551	9,577,463	4,128	846,665	635,687

Diagonal and +/- 1 Bands: shows the magnitude of shift in predefined bands for modeled values

Positive Predicted Value (PPV): represents 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) measures the capture rates while allowing +/-X% deviation of the modelled values away from the actual reported figures

Ranking Capability at band level is measured using the combination of the Somers' D and Spearman Correlation