Problem Statement: Predict if opportunity is won or lost, and what factors affect winning the opportunity the most

Ho – Opportunity is lost

Ha – Opportunity isn’t lost

Limitations – factors that affect deals being closed may not be included in the data set, though the data has been included in Watson Analytics

Data

<https://www.ibm.com/communities/analytics/watson-analytics-blog/guide-to-sample-datasets/>

Data is under WA\_Fn\_UseC\_Sales Win\_Loss.csv

**Columns**

**Opportunity Number** – (Categorical, though numerically expressed) Unique number corresponding to sales lead

**Supplies subgroup** – (CAT) subcategory of product under main supplies group – further defines what the lead sells

**Supplies Group** – (CAT) category which sales opportunity belongs to, e.g. whether they

**Region** - (CAT) geographical location of lead in USA

**Route To Market** (CAT) - the mode of sales that the company takes to get products to market – defines the relationship the owner of the dataset has with each opportunity

**Elapsed Days In Sales Stage** – (numeric) number of days lead has been in sales funnel

**Opportunity Result** (TARGET COLUMN) – Yes or No, regarding if opportunity was closed or not

**Sales Stage Change Count** (numerical) -

**Total Days Identified Through Closing** (NUM) – number of days opportunity has spent in closing stage of sales funnel

**Total Days Identified Through Qualified** (NUM) - number of days opportunity has spent as a qualified lead.

**Opportunity Amount USD** (NUM) – potential revenue generated from opportunity

**Client Size By Revenue** (CAT, though numerical) – numerical key denoting size of client by total revenue generated by client.

**Client Size By Employee Count** - (CAT, though numerical in expression) numerical key denoting size of client by total number of employees.

**Revenue From Client Past Two Years** (CAT, though numerical in expression) numerical key denoting amount of revenue from client in the past two years

**Competitor Type** (CAT) – whether competitors for opportunity are known to the owner of the data

**Ratio Days Identified To Total Days** (NUM) – percentage of days client spent in ‘identified’ stage to the total number of days spent as a lead (from start of funnel until deal is closed or lost)

**Ratio Days Validated To Total Days** (NUM) – percentage of days client spent in ‘validated’ stage to the total number of days spent as a lead (from start of funnel until deal is closed or lost)

**Ratio Days Qualified To Total Days** (NUM) – percentage of days client spent in ‘qualified’ stage to the total number of days spent as a lead (from start of funnel until deal is closed or lost)

**Deal Size Category** (CAT, though numerical in expression) - numerical key denoting size of client by total number of employees. The larger the number, the larger the size of the deal (in terms of revenue

78026 rows, 19 columns