Codebook for Attrition Dataset Data Overview

## Credentials

This data set can be downloaded from Kaggle data sets.

The direct link to data is:<https://www.kaggle.com/tiredgeek/predict-bo-trial>

## Business goal

This data was collected to predict product backorders.

## Data description

This data set is a data frame of 23 variables over more than a million rows. Each row represents SKU (=item) that is sold by the company .

Need to accurately describe which employees are on the data. For example, randomly selected sample of employees from all 200,000 employees that worked in the company in 2007-2017

Need to describe the type of the company, country, company size, etc. even if the name of the company is not revealed.

When looking at the dataset in R, we can get the following summary –

'data.frame': 1687861 obs. of 23 variables:

$ sku : Factor w/ 1687861 levels "(1687860 rows)",..: 2 3 4 5 6 7 8 9 10 11 ...

$ national\_inv : int 0 2 2 7 8 13 1095 6 140 4 ...

$ lead\_time : int NA 9 NA 8 NA 8 NA 2 NA 8 ...

$ in\_transit\_qty : int 0 0 0 0 0 0 0 0 0 0 ...

$ forecast\_3\_month : int 0 0 0 0 0 0 0 0 15 0 ...

$ forecast\_6\_month : int 0 0 0 0 0 0 0 0 114 0 ...

$ forecast\_9\_month : int 0 0 0 0 0 0 0 0 152 0 ...

$ sales\_1\_month : int 0 0 0 0 0 0 0 0 0 0 ...

$ sales\_3\_month : int 0 0 0 0 0 0 0 0 0 0 ...

$ sales\_6\_month : int 0 0 0 0 0 0 0 0 0 0 ...

$ sales\_9\_month : int 0 0 0 0 4 0 0 0 0 0 ...

$ min\_bank : int 0 0 0 1 2 0 4 0 0 0 ...

$ potential\_issue : Factor w/ 3 levels "","No","Yes": 2 2 2 2 2 2 2 2 2 2 ...

$ pieces\_past\_due : int 0 0 0 0 0 0 0 0 0 0 ...

$ perf\_6\_month\_avg : num -99 0.99 -99 0.1 -99 0.82 -99 0 -99 0.82 ...

$ perf\_12\_month\_avg: num -99 0.99 -99 0.13 -99 0.87 -99 0 -99 0.87 ...

$ local\_bo\_qty : int 0 0 0 0 0 0 0 0 0 0 ...

$ deck\_risk : Factor w/ 3 levels "","No","Yes": 2 2 3 2 3 2 3 3 2 2 ...

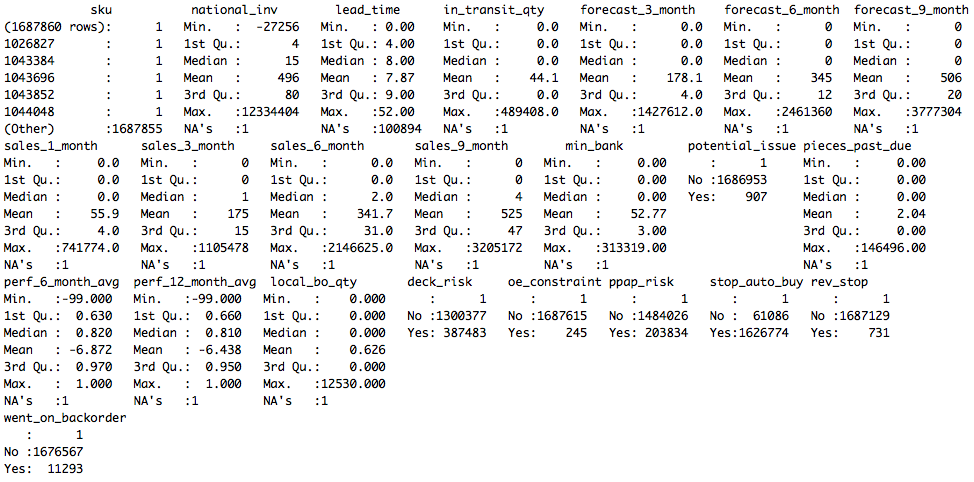
$ oe\_constraint : Factor w/ 3 levels "","No","Yes": 2 2 2 2 2 2 2 2 2 2 ...

$ ppap\_risk : Factor w/ 3 levels "","No","Yes": 2 2 2 2 2 2 2 3 2 2 ...

$ stop\_auto\_buy : Factor w/ 3 levels "","No","Yes": 3 3 3 3 3 3 3 3 3 3 ...

$ rev\_stop : Factor w/ 3 levels "","No","Yes": 2 2 2 2 2 2 2 2 2 2 ...

$ went\_on\_backorder: Factor w/ 3 levels "","No","Yes": 2 2 2 2 2 2 2 2 2 2 ...



## Variables description

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Variable Name | Description | ID | Type | Possible values |
| 1 | sku | Random ID for the product | ID | Text | random numeric |
| 2 | national\_inv | Current inventory level for the part | Current State | Numeric |  |
| 3 | lead\_time | Transit time for product (if available) | Current State | Numeric |  |
| 4 | in\_transit\_qty | Amount of product in transit from source | Current State | Numeric |  |
| 5 | forecast\_3\_month | Forecast sales for the next 3 months | Forecast | Numeric |  |
| 6 | forecast\_6\_month | Forecast sales for the next 6 months | Forecast | Numeric |  |
| 7 | forecast\_9\_month | Forecast sales for the next 9 months | Forecast | Numeric |  |
| 8 | sales\_1\_month | Sales quantity for the prior 1 month time period | Actual Sales | Numeric |  |
| 9 | sales\_3\_month | Sales quantity for the prior 3 month time period | Actual Sales | Numeric |  |
| 10 | sales\_6\_month | Sales quantity for the prior 6 month time period | Actual Sales | Numeric |  |
| 11 | sales\_9\_month | Sales quantity for the prior 9 month time period | Actual Sales | Numeric |  |
| 12 | min\_bank | Minimum recommended amount to stock | Current State | Numeric |  |
| 13 | potential\_issue | Source issue for part identified | Source variables | Boolean | Yes/No |
| 14 | pieces\_past\_due | Parts overdue from source | Source variables | Numeric |  |
| 15 | perf\_6\_month\_avg | Source performance for prior 6 month period | Source variables | Numeric |  |
| 16 | perf\_12\_month\_avg | Source performance for prior 12 month period | Source variables | Numeric |  |
| 17 | local\_bo\_qty | Amount of stock orders overdue | Source variables | Numeric |  |
| 18 | deck\_risk | Part risk flag | Risk Variables | Boolean | Yes/No |
| 19 | oe\_constraint | Part risk flag | Risk Variables | Boolean | Yes/No |
| 20 | ppap\_risk | Part risk flag | Risk Variables | Boolean | Yes/No |
| 21 | stop\_auto\_buy | Part risk flag | Risk Variables | Boolean | Yes/No |
| 22 | rev\_stop | Part risk flag | Risk Variables | Boolean | Yes/No |
| 23 | went\_on\_backorder | Product actually went on backorder. This is the target value. | Fact | Boolean | Yes/No |

## Business questions

1. Does the sales forecast predict the actual sales and what is the level of accuracy?
2. Which SKUs are in high demand when comparing
3. out of the BO items, which SKU are due to part overdue from source
4. Percentage of backordered items

## Who needs to review the business question

Sales Operations and Inventory Manager

## Related links

<http://projekter.aau.dk/projekter/files/262657498/master_thesis.pdf>

<http://www.business-science.io/business/2017/10/16/sales_backorder_prediction.html>

<https://www.kaggle.com/tiredgeek/predict-bo-trial>

<https://rpubs.com/benjaminlott/BackorderMLR>