|  |
| --- |
|  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  | |  |
| 4/1/2015 | |  |
| Walmart Recruiting - Sales in stormy weather | |  |
| Notes | |  |
|  | |  |
| Isaac | |  |
| Authored by: Gino Tesei | |  |
|  | |  |

Exploratory analysis

# weather.csv

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |
| **min** | 1 | 1-gen-12 | -11 | -21 | -16 | -35 | -24 | -15 |
| **max** | 20 | 31-ott-14 | 114 | 88 | 100 | 33 | 77 | 80 |
| **type** | num | num | num, M | num, M | num, M | num, M | num, M | num, M |
|  | **station\_nbr** | **date** | **tmax** | **tmin** | **tavg** | **depart** | **dewpoint** | **wetbulb** |
|  | 1 | 1-gen-12 | 52 | 31 | 42 | M | 36 | 40 |
|  | 2 | 1-gen-12 | 48 | 33 | 41 | 16 | 37 | 39 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 0 | 0 | 406 | 1611 |
| 81 | 35 | 740 | 1949 |
| num,M | num, M | num, - | num, - |
| **heat** | **cool** | **sunrise** | **sunset** |
| 23 | 0 | - | - |
| 24 | 0 | 716 | 1626 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | *emerge che tutte le volte che si ha un + è in FG+* | |  |
|  | *NO* |
| txt | *+FC* |
| **codesum** | **has\_+\_** | **has\_FG+\_** | **has\_+FC** |
| RA FZFG BR | NO | NO | NO |
| RA | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  | *emerge che tutte le volte che si ha un + è in FG+* | |  |
|  | *NO* |
| txt | *+FC* |
| **codesum** | **has\_+\_** | **has\_FG+\_** | **has\_+FC** |
| RA FZFG BR | NO | NO | NO |
| RA | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| *DZ* | *SN* | *SG* | *GS* |
| **has\_DZ** | **has\_SN** | **has\_SG** | **has\_GS** |
| NO | NO | NO | NO |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  | *NO* |  |  |
| *PL* | *IC* | *FG+* | *FG* |
| **has\_PL** | **has\_IC** | **has\_FG+** | **has\_FG** |
| NO | NO | NO | SI |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| *BR* | *UP* | *HZ* | *FU* |
| **has\_BR** | **has\_UP** | **has\_HZ** | **has\_FU** |
| SI | NO | NO | NO |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| *NO* |  | *NO* | *NO* |
| *VA* | *DU* | *DS* | *PO* |
| **has\_VA** | **has\_DU** | **has\_DS** | **has\_PO** |
| NO | NO | NO | NO |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| *NO* |  | *NO* |  |
| *SA* | *SS* | *PY* | *SQ* |
| **has\_SA** | **has\_SS** | **has\_PY** | **has\_SQ** |
| NO | NO | NO | NO |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| *NO* | *NO* |  |  |
| *DR* | *SH* | *FZ* | *MI* |
| **has\_DR** | **has\_SH** | **has\_FZ** | **has\_MI** |
| NO | NO | SI | NO |
| NO | NO | NO | NO |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
| *PR* | *BC* | *BL* | *VC* |
| **has\_PR** | **has\_BC** | **has\_BL** | **has\_VC** |
| NO | NO | NO | NO |
| NO | NO | NO | NO |

|  |  |  |
| --- | --- | --- |
|  |  |  |
| 0 | 0 | 23.72 |
| 16.2 | 7.36 | 30.61 |
| num, M,  T (=< 0,01) | num, M,  T (=< 0,01) | num, M |
| **snowfall** | **preciptotal** | **stnpressure** |
| M | 0.05 | 29.78 |
| 0 | 0.07 | 28.82 |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
| 29.16 | 0 | 1 | 0 |
| 30.8 | 28.4 | 36 | 28.7 |
| num, M | num, M | num, M | num, M |
| **sealevel** | **resultspeed** | **resultdir** | **avgspeed** |
| 29.92 | 3.6 | 20 | 4.6 |
| 29.91 | 9.1 | 23 | 11.3 |

Imputing models

Discarded input variables

* station\_nbr
* date

All predictors treated as numeric (no factors)

**Models**

All.RegModels = c("LinearReg","KNN\_Reg", "PLS\_Reg" , "Ridge\_Reg" , "SVM\_Reg", "Cubist\_Reg")

All.ClassModels = c("SVMClass")

**Performance**

Predicting model #1

Train set (station s, item i) = train data (station s , item i)

Feature selection: …

Model: …