shiyan.R

daitu

Tue Jun 21 23:15:23 2016

## 2016年暑期课程设计  
## 问题：Grupo Bimbo Inventory Demand  
## 宾堡集团的库存需求  
## 最大限度地提高销售和最大限度地减少烘焙食品的退回  
## Daitu  
## start:2016.06.21  
## 参考借鉴kaggle上的公开程序  
  
  
setwd("/Users/Daitu/数据分析/kaggle/Grupo Bimbo")  
getwd()

## [1] "/Users/daitu/数据分析/kaggle/Grupo Bimbo"

## 加载包  
  
library(data.table)  
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.2.4

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:data.table':  
##   
## between, last

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(treemap)

## Warning: package 'treemap' was built under R version 3.2.4

## 读取数据####  
## 1:读取训练集  
system.time({  
 traindata <- fread("train.csv",sep=",",header = TRUE)  
})

##   
Read 0.0% of 74180464 rows  
Read 4.4% of 74180464 rows  
Read 8.8% of 74180464 rows  
Read 13.1% of 74180464 rows  
Read 17.5% of 74180464 rows  
Read 21.9% of 74180464 rows  
Read 26.2% of 74180464 rows  
Read 30.6% of 74180464 rows  
Read 35.0% of 74180464 rows  
Read 39.4% of 74180464 rows  
Read 43.8% of 74180464 rows  
Read 48.1% of 74180464 rows  
Read 52.4% of 74180464 rows  
Read 56.7% of 74180464 rows  
Read 61.1% of 74180464 rows  
Read 65.5% of 74180464 rows  
Read 70.0% of 74180464 rows  
Read 74.4% of 74180464 rows  
Read 78.8% of 74180464 rows  
Read 83.2% of 74180464 rows  
Read 87.6% of 74180464 rows  
Read 92.0% of 74180464 rows  
Read 96.2% of 74180464 rows  
Read 74180464 rows and 11 (of 11) columns from 2.980 GB file in 00:00:29

## user system elapsed   
## 26.496 2.241 28.740

# Read 74180464 rows and 11 (of 11) columns from 2.980 GB file in 00:00:29  
# 用户 系统 流逝   
# 26.553 2.077 28.869  
  
## 2:读取客户名单数据  
cliente\_tabla <- fread("cliente\_tabla.csv",sep=",",header = TRUE)  
  
## 3:读取产品名单数据  
producto\_tabla <- fread("producto\_tabla.csv",sep=",",header = TRUE)  
  
## 4:读取城镇和国家（州）数据  
town\_state <- fread("town\_state.csv",sep=",",header = TRUE)

## 数据的描述统计####

## 1:分析数据的周数：Semana  
Semana <- data.frame(table(traindata$Semana))  
colnames(Semana) <- c("Semana","Freq")  
# 3 4 5 6 7 8 9   
# 11165207 11009593 10615397 10191837 10382849 10406868 10408713   
## 条形图