

Algorithms I Used



Apriori Algorithm

Apriori(T, ϵ)

$L_1 \leftarrow \{\text{large 1 - itemsets}\}$

$k \leftarrow 2$

while $L_{k-1} \neq \emptyset$

$C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \notin a\} - \{c \mid \{s \mid s \subseteq c \wedge |s| = k - 1\} \not\subseteq L_{k-1}\}$

for transactions $t \in T$

$C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}$

for candidates $c \in C_t$

$count[c] \leftarrow count[c] + 1$

$L_k \leftarrow \{c \mid c \in C_k \wedge count[c] \geq \epsilon\}$

$k \leftarrow k + 1$

return $\bigcup_k L_k$

Gini Index for
decision tree
grading

$$I_G(f) = \sum_{i=1}^m f_i(1 - f_i) = \sum_{i=1}^m (f_i - f_i^2) = \sum_{i=1}^m f_i - \sum_{i=1}^m f_i^2 = 1 - \sum_{i=1}^m f_i^2 = \sum_{i \neq k} f_i f_k$$

Algorithms I Used



Euclidean Distance

$$d(\mathbf{p}, \mathbf{q}) = d(\mathbf{q}, \mathbf{p}) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \cdots + (q_n - p_n)^2}$$
$$= \sqrt{\sum_{i=1}^n (q_i - p_i)^2}.$$

Probability Function

$$P(X \leq m) = P(X \geq m) = \int_{-\infty}^m f(x) dx = \frac{1}{2}.$$

Mean Function

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

Algorithms I Used



Variable Generation

```
#Gave conditions where the value of SWAM would equal 1
SWAM$SMALL[grepl("S",SWAM$SWAM) | grepl("O",SWAM$SWAM)] <- 1
SWAM$WOMAN[grepl("W",SWAM$SWAM)] <- 1
SWAM$MINORITY[grepl("M",SWAM$SWAM)] <- 1
SWAM$MICRO[grepl("O",SWAM$SWAM)] <- 1
SWAM$SDV[grepl("SDV",SWAM$SWAM)] <- 1

#Added attribute to allow for finding the amount of days in the fy that the order was published
BuyReq$DatesFromFYStart <- as.numeric(as.Date(as.character(BuyReq$PUBLISH_DATE),format="%m/%d/%Y")-as.Date(paste0(substr(
as.character(BuyReq$PUBLISH_DATE),nchar(as.character(BuyReq$PUBLISH_DATE))-3,nchar(as.character(BuyReq$PUBLISH_DATE))),"-07-01")))
%% 365

#made attribute flags to show what type of entity we are focusing on... School/Locality/Agency/ Other... other flag is used for
future purposes for now
BuyReq$IsSchool<-0;
BuyReq$IsSchool[grepl(SchoolIndicator,BuyReq$CLIENTNAME)] <-1;
BuyReq$IsLocality<-0;
BuyReq$IsLocality[grepl(LocalityIndicator,BuyReq$CLIENTNAME)] <-1;
BuyReq$IsOtherType<-0;
BuyReq$IsOtherType[grepl(OtherIndicator,BuyReq$CLIENTNAME)] <-1;
BuyReq$IsAgency <-0;
BuyReq$IsAgency[BuyReq$IsSchool==0 & BuyReq$IsLocality==0 & BuyReq$IsOtherType==0] <-1;

#When importing data from csv, a i.. sometimes appears, this code removes that... a better sql export would fix it too, utf-8-BOM
names(VenRes) <- gsub("i..", "", names(VenRes))
names(BuyReqLines) <- gsub("i..", "", names(BuyReqLines))
names(VenResLines) <- gsub("i..", "", names(VenResLines))
names(BuyReq) <- gsub("i..", "", names(BuyReq))
names(BuyReqLines)[names(BuyReqLines)=="QUANTITY"] <- "QUANTITYrequested"
```

Data Is Power



FORUM 2015

Presented by
Subhash Jaini

Agenda



- Introduce myself
- Discuss the goal of analytics in eVA
- Questions



Introduction



- Subhash Jaini
- Heavy background in restaurant management- passion for food
- Master Gardener
- New dad on May 7th, 2015
- Prior to eVA, Lead JavaScript Developer
- Evangelist for Data Science
- VCU's Executive Masters Decision Analytics
- Logi Report Support/ Discover things

“Statistics Lie”



- eVA is VERY business oriented
- Logically driven, many meetings and debates for logic/data used/ business case
- Statistics/ Code/ Data is used to make the process replicable impersonally

Analytics Goals: Here

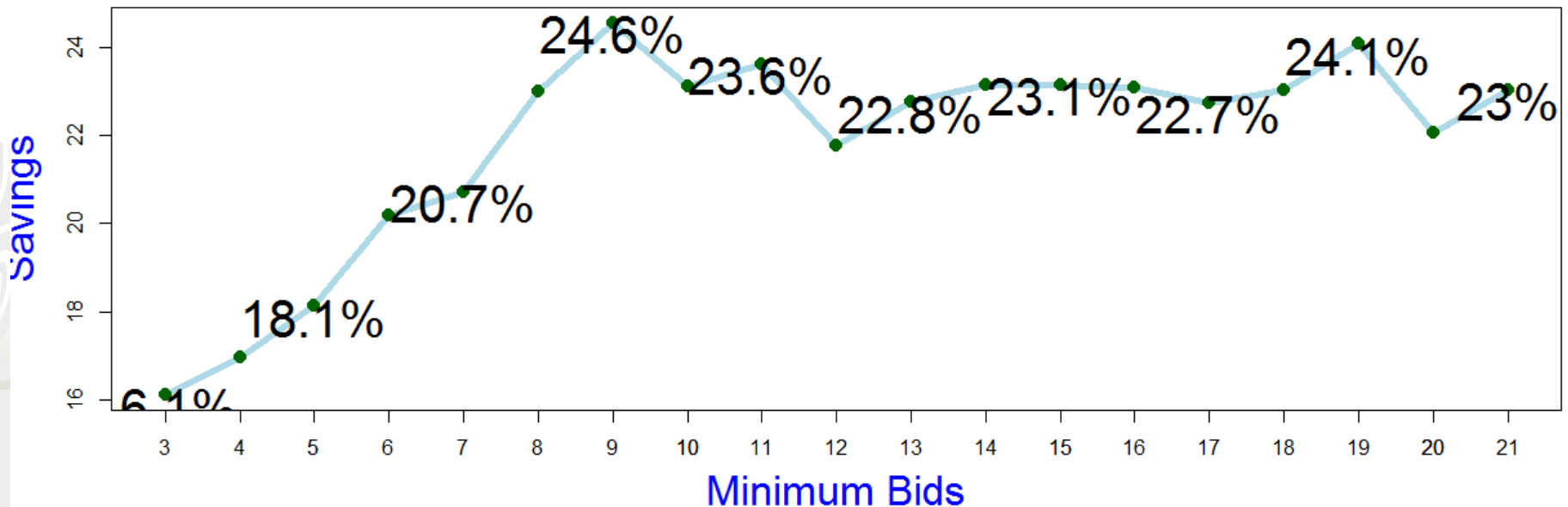


Through competition, we found proven cost savings. As more folks bid, the cost savings increase.

Strategy of Vendors

Here	Relevant
Trusted	Valuable

Savings as Bids Increase



Analytics Goals: Relevant



Through heavier data engineering tactics, we were able to identify an item and run them through scenario analysis. Still experimental, but moving forward

Here	Relevant
Trusted	Valuable

Mandatory Contract vs Non
Contract vs Non
Sole Source vs No
Self Registered vs Non



Analytics Goals: Trusted

Through PO data, we are able to identify an item and group it with other items based on a grouping. Groupings are based on orders, req, buyer AND day, BSO AND day. Still experimental, but neat stuff.

Here	Relevant
Trusted	Valuable



Analytics Goals: Valuable



- Personal Enjoyment
- QQ Savings-Cost to Build
- QQ Savings-Cost to Maintain
- QQ Savings-Savings to be had

Here	Relevant
Trusted	Valuable

Question Time



- Presentation: “How can professional buyers help in this data initiative?”
- Subhash’s Turn
- Any Questions?

