

Customer spending patterns

Question: What is the driving force to what a customer is willing to pay, and how can we use it to our advantage?

Approach: We will do data exploration. What trends exist? Is there any obvious significance in our data? Regression tree vs. logistic regression for a possible pricing model that will allow for predictions in the future.

The Big Bang: An interactive tool that gives a user diagnostics in which will allow a business user to understand trends and model predictions of ranges of values in which a customer is likely to purchase.

Abstract

Central to development of society is transportation, more specifically freight and good transportation. As our society has grown so too has our technology. Currently technology and data is growing at an alarming rate. Petabytes of information exists, and few people know how to utilize it. Databases have become massive and extremely expensive. For multiple reasons: security, confidentiality, and many others; companies are not willing to move their data to high efficiency data storage centers (such as Hadoop). Transportation data is collected at extremely high rates, which is more than a single person can analyze. It has become crucial to squeeze as many pennies out of the data as possible. Currently intermodal (multiple means of transporting goods) companies have flat rate pricing systems, but it is their desire to incorporate a more dynamic pricing model which utilizes statistical methods and data visualization. It is this projects goal to attempt to understand the consumer of intermodal transportation services and develop a deployable interactive tool in which we can interactively observe consumers buying habits.

1 Introduction:

The most essential part of big data is ensuring reproducibility of data products by any (business or otherwise) user. This is a difficult task. Everyone is different, and no two people have the exact same skills or the same understanding of data. However, for any study it becomes crucial to understand the data, how it is stored, and where it is coming from. In this article we will first define some terms that are common to the transportation business. Then, we will explore the data in order to fully understand everything that will be presented in this project. Last, we will analyze the data, create a data product, and will discuss the results.

1.1 Understanding Intermodal Transportation

Transportation business is extremely complicated. Millions of moving and working parts are constantly in motion. Each part can be collected and stored as observable data. Clearly any data from such a complicated industry needs to

have a common language. So, we must first define a few common transportation terms. These words will be used throughout the paper and are essential to understanding our objectives.

1. Intermodal transportation - Intermodal transportation is the movement of freight by more than one type of carrier. Carriers include trains, trucks, barge, etc.
2. Drayage - Location where freight is moved from one transportation type to another: train to truck, truck to barge, or any other transportation combination.
3. Ramp - defined by city, state, and three letter acronym giving location and specific ramp for the transaction.
4. Broker LOB - Stands for "broker type and line of business". There are three main categories: marketing firm, brokerage, intermodal company.
5. Bill to name - Company which the transportation services were purchased from. Often when dealing with brokers they act as an intermediate between two companies in order to find the cheapest means of transportation.

1.2 Understanding Intermodal Data

This data is coming from a door to door intermodal company, meaning that each purchase will consist of transportation where from origin, truck delivers to train, then train delivers to truck, and last truck delivers to destination. Due to the extreme levels of confidentiality in which the data was originally stored, we have eliminated all columns of which could be identified to any individual. We will show a few of the rows.

Even though the column names have been simplified, this is still a complicated topic. We will define a few topics which will be common themes throughout this research.

2 Exploration of the data

Exploration of the data is a major component of this project.

```
## Warning in gzfile(file, "rb"): cannot open compressed file 'profitCostMod2.rds',  
probable reason 'No such file or directory'  
## Error in gzfile(file, "rb"): cannot open the connection
```

```
## Error in ggplot(profitCostMod2, aes(ProfitMile, CostMile), alpha
= 0.1): object 'profitCostMod2' not found
```

```
## Error in ifelse(profitCostMod2$Buy == "Load", 1, 0): object 'profitCostMod2'
not found
## Error in eval(expr, envir, enclos): object 'profitCostMod2' not
found
## Error in lapply(list(...), .num_to_date): object 'profitCostMod8'
not found
## Error in eval(expr, envir, enclos): object 'profitCostMod8' not
found
## Error in head(profitCostMod8.1b): object 'profitCostMod8.1b' not
found
## Error in ggplot(profitCostMod8.1b, aes(TRUNC.QREC.CREATE_DT., Buyrate,
: object 'profitCostMod8.1b' not found
```

```
## Error in eval(expr, envir, enclos): object 'profitCostMod8' not
found
## Error in ggplot(profitCostMod3a, aes(Total.Cost, Profit, color =
BROKER_LOB, : object 'profitCostMod3a' not found
## Error in ggplot(profitCostMod3a, aes(Total.Cost, Profit, group =
BROKER_LOB)): object 'profitCostMod3a' not found
## Error in ggsave("Contour_filtered1.png", width = 14, height = 5,
dpi = 200): plot should be a ggplot2 plot
## Error in ggplot(profitCostMod3a, aes(Total.Cost, Profit, group =
BROKER_LOB)): object 'profitCostMod3a' not found
```

```
## Error in eval(expr, envir, enclos): argument is missing, with no
default
```

3 Statistical Analysis:

```
head(profitCostMod8)

## Error in head(profitCostMod8): object 'profitCostMod8' not found

head(profitCostMod8.1b)

## Error in head(profitCostMod8.1b): object 'profitCostMod8.1b' not
found
```

```

head(profitCostMod2)

## Error in head(profitCostMod2): object 'profitCostMod2' not found

Xiangspairs<-profitCostMod2 %>%
  group_by(BILL_TO_NAME, ORIG_RAMP_NAME, DEST_RAMP_NAME) %>%
  summarise(Buy=sum(Buy),Search=sum(Search), Total.Cost=sum(TOTAL_COST),
            Total.Revenue=sum(TOTAL_REVENUE),Profit=sum(COC)) %>%
  filter(Buy==1) %>%
#   filter(Search>20) %>%
#   filter(BROKER_LOB != "#N/A") %>%
#   filter(BROKER_LOB != 0) %>%
#   filter(Profit<30000) %>%
  mutate(Buyrate=Buy/Search)

## Error in eval(expr, envir, enclos): object 'profitCostMod2' not
found

head(Xiangspairs)

## Error in head(Xiangspairs): object 'Xiangspairs' not found

set.seed(2345)
trainIndex<-createDataPartition(Xiangspairs$BILL_TO_NAME, p=.8,
                                list = FALSE,
                                times=1)

## Error in createDataPartition(Xiangspairs$BILL_TO_NAME, p = 0.8, list
= FALSE, : object 'Xiangspairs' not found

XiangsTrain<-Xiangspairs[trainIndex,]

## Error in eval(expr, envir, enclos): object 'Xiangspairs' not found

XiangsTest<-Xiangspairs[-trainIndex,]

## Error in eval(expr, envir, enclos): object 'Xiangspairs' not found

XiangsTrainLabs<-XiangsTrain[,9]

## Error in eval(expr, envir, enclos): object 'XiangsTrain' not found

dim(XiangsTrain)

## Error in eval(expr, envir, enclos): object 'XiangsTrain' not found

dim(XiangsTest)

## Error in eval(expr, envir, enclos): object 'XiangsTest' not found

dim(Xiangspairs)

```

```
## Error in eval(expr, envir, enclos): object 'Xiangspairs' not found
lm.fit1<-lm(Buyrate~., data=XiangsTrain[5:9])
## Error in is.data.frame(data): object 'XiangsTrain' not found
summary(lm.fit1)
## Error in summary(lm.fit1): object 'lm.fit1' not found
lm.fit2<-lm(Buyrate~Search, data=XiangsTrain[5:9])
## Error in is.data.frame(data): object 'XiangsTrain' not found
summary(lm.fit2)
## Error in summary(lm.fit2): object 'lm.fit2' not found
```

Loyal Customers
High Demand Routes
Loyal Customers on High Demand Routes
Different Months?

4 Results

5 Conclusion