

DELIVARIES ANALYSIS

DATA ANALYST

10/17/2021

R Markdown

```
w=read.csv("clipboard","\t",header = TRUE)
xx=w
xx[xx==""]=NA
xxx=na.omit(xx)
xxx %>% mutate_at(vars(Customer.placed.order.datetime,Placed.order.with.restaurant.datetime,Driver.at.restaurant.datetime,Delivered.to.consumer.datetime),~paste("2021-10-",.))->Deliverydata
head(Deliverydata)
```

```
## Customer.placed.order.datetime Placed.order.with.restaurant.datetime
## 1 2021-10- 01 02:52:12 2021-10- 01 03:00:25
## 2 2021-10- 13 15:58:57 2021-10- 13 17:56:45
## 3 2021-10- 17 19:02:37 2021-10- 17 19:12:51
## 4 2021-10- 12 04:01:57 2021-10- 12 04:08:18
## 5 2021-10- 08 23:46:38 2021-10- 08 23:49:32
## 6 2021-10- 12 03:54:30 2021-10- 12 03:57:30
## Driver.at.restaurant.datetime Delivered.to.consumer.datetime Driver.ID
## 1 2021-10- 01 03:08:09 2021-10- 01 03:35:20 279
## 2 2021-10- 13 18:24:39 2021-10- 13 18:57:01 303
## 3 2021-10- 17 19:19:59 2021-10- 17 19:31:09 212
## 4 2021-10- 12 04:13:02 2021-10- 12 04:37:57 352
## 5 2021-10- 08 23:54:29 2021-10- 09 00:37:15 313
## 6 2021-10- 12 04:10:26 2021-10- 12 04:41:00 314
## Restaurant.ID Consumer.ID Delivery.Region Is.ASAP Order.total
## 1 50 6738 Mountain View TRUE 16.33
## 2 96 64746 Palo Alto TRUE 76.14
## 3 190 12484 San Jose TRUE 16.77
## 4 194 13920 San Jose TRUE 25.03
## 5 9 7037 Palo Alto TRUE 51.57
## 6 350 95392 Palo Alto TRUE 16.17
## Amount.of.discount Amount.of.tip Refunded.amount
## 1 0 0.82 0
## 2 0 6.45 0
## 3 6 2.52 0
## 4 0 5.00 0
## 5 0 5.16 0
## 6 6 0.93 0
```

Data manipulation to required levels

```
Deliverydata$Customer.placed.order.datetime=ymd_hms(Deliverydata$Customer.placed.order.datetime)
Deliverydata$Placed.order.with.restaurant.datetime=ymd_hms(Deliverydata$Placed.order.with.restaurant.datetime)
```

```

Deliverydata$Driver.at.restaurant.datetime=ymd_hms(Deliverydata$Driver.at.restaurant.datetime)
Deliverydata$Delivered.to.consumer.datetime=ymd_hms(Deliverydata$Delivered.to.consumer.datetime)

Deliverydata$Delivery.Region=as.factor(Deliverydata$Delivery.Region)
Deliverydata$Is.ASAP=as.factor(Deliverydata$Is.ASAP)
str(Deliverydata)

```

```

## 'data.frame': 13524 obs. of 13 variables:
## $ Customer.placed.order.datetime : POSIXct, format: "2021-10-01 02:52:12" "2021-10-13 15:58:5"
## $ Placed.order.with.restaurant.datetime: POSIXct, format: "2021-10-01 03:00:25" "2021-10-13 17:56:4"
## $ Driver.at.restaurant.datetime : POSIXct, format: "2021-10-01 03:08:09" "2021-10-13 18:24:3"
## $ Delivered.to.consumer.datetime : POSIXct, format: "2021-10-01 03:35:20" "2021-10-13 18:57:0"
## $ Driver.ID : int 279 303 212 352 313 314 154 315 248 160 ...
## $ Restaurant.ID : int 50 96 190 194 9 350 27 107 212 309 ...
## $ Consumer.ID : int 6738 64746 12484 13920 7037 95392 7426 95797 2122 128...
## $ Delivery.Region : Factor w/ 4 levels "Mountain View",...: 1 3 4 4 3 3 3 1 4
## $ Is.ASAP : Factor w/ 2 levels "FALSE","TRUE": 2 2 2 2 2 2 2 2 2 ...
## $ Order.total : num 16.3 76.1 16.8 25 51.6 ...
## $ Amount.of.discount : num 0 0 6 0 0 6 0 6 0 0 ...
## $ Amount.of.tip : num 0.82 6.45 2.52 5 5.16 0.93 5.9 1.36 1.69 2.78 ...
## $ Refunded.amount : num 0 0 0 0 0 0 0 0 0 0 ...
## - attr(*, "na.action")= 'omit' Named int [1:4554] 9 12 14 19 22 23 29 34 35 40 ...
## ..- attr(*, "names")= chr [1:4554] "9" "12" "14" "19" ...

```

checking on delivery delays across the regions

```

Deliverydata %>% select(Delivered.to.consumer.datetime,Placed.order.with.restaurant.datetime,Delivery.Region)
mutate(delay=Delivered.to.consumer.datetime - Placed.order.with.restaurant.datetime) %>% arrange(desc(delay))
head(newdata,10)

```

```

##   Delivered.to.consumer.datetime Placed.order.with.restaurant.datetime
## 1      2021-10-30 20:53:12      2021-10-01 00:23:32
## 2      2021-10-09 00:04:54      2021-10-07 18:04:33
## 3      2021-10-04 18:16:51      2021-10-04 00:18:40
## 4      2021-10-09 18:47:38      2021-10-09 02:48:48
## 5      2021-10-01 19:13:51      2021-10-01 03:54:38
## 6      2021-10-24 18:43:02      2021-10-24 03:23:52
## 7      2021-10-11 19:33:32      2021-10-11 04:22:19
## 8      2021-10-18 18:45:43      2021-10-18 04:11:25
## 9      2021-10-10 18:53:03      2021-10-10 04:33:06
## 10     2021-10-18 02:09:00      2021-10-17 17:25:35
##   Delivery.Region Restaurant.ID      delay
## 1   Mountain View      408 42989.6667 mins
## 2     Palo Alto       19 1800.3500 mins
## 3     Palo Alto       10 1078.1833 mins
## 4     Palo Alto       47  958.8333 mins
## 5   Mountain View       62  919.2167 mins
## 6     Palo Alto       14  919.1667 mins
## 7   Mountain View     239  911.2167 mins
## 8     Palo Alto       10  874.3000 mins
## 9     Palo Alto       10  859.9500 mins
## 10    Palo Alto     324  523.4167 mins

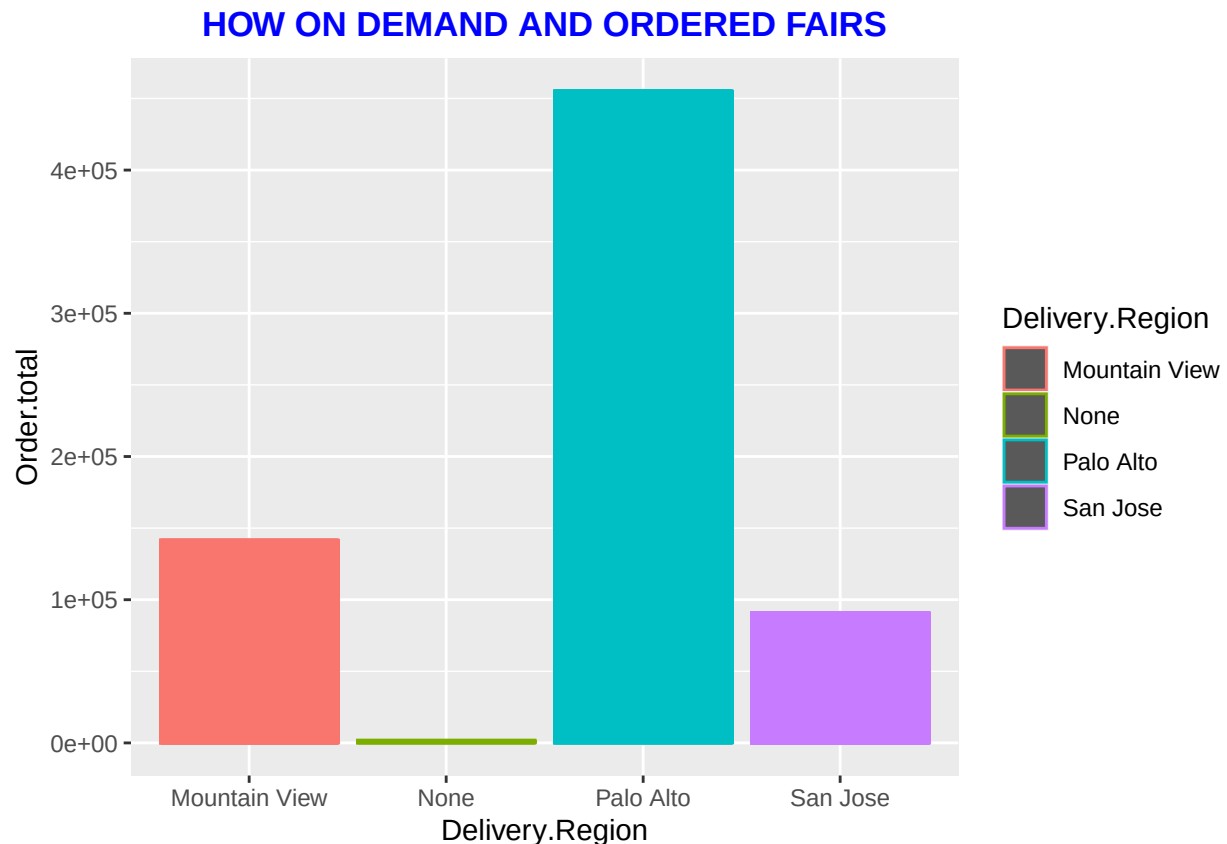
```

```
tail(newdata,5)
```

```
##      Delivered.to.consumer.datetime Placed.order.with.restaurant.datetime
## 13520      2021-10-01 00:30:04      2021-10-31 23:58:13
## 13521      2021-10-01 00:09:00      2021-10-31 23:37:15
## 13522      2021-10-01 00:09:28      2021-10-31 23:43:55
## 13523      2021-10-01 00:05:22      2021-10-31 23:40:31
## 13524      2021-10-01 00:19:21      2021-10-31 23:55:48
##      Delivery.Region Restaurant.ID      delay
## 13520      San Jose      190 -44608.15 mins
## 13521      Palo Alto      4 -44608.25 mins
## 13522      Mountain View      88 -44614.45 mins
## 13523      Mountain View      112 -44615.15 mins
## 13524      Palo Alto      14 -44616.45 mins
```

observation;from the above its clear most delayed delivery was made for Mountain View but its most frequently on palo Alto,the reason could be either distance reason or assignable reason to driver person.

```
Deliverydata %>% ggplot(aes(x=Delivery.Region,y=Order.total,color=Delivery.Region)) + geom_bar(stat = "sum")
labs(title = "HOW ON DEMAND AND ORDERED FAIRS")+theme(plot.title = element_text(hjust = 0.5,face ="bold"))
```



ypur most sales are on oder with palo alto having the greatest sahre on order totals.COnsider looking for new markets for on demands for maximization of sales

```
Deliverydata %>% ggplot(aes(x=Delivery.Region,y=Order.total,color=Is.ASAP)) + geom_bar(stat = "identity")
labs(title = "HOW ON DEMAND AND ORDERED FAIRS")+theme(plot.title = element_text(hjust = 0.5,face ="bold"))
```

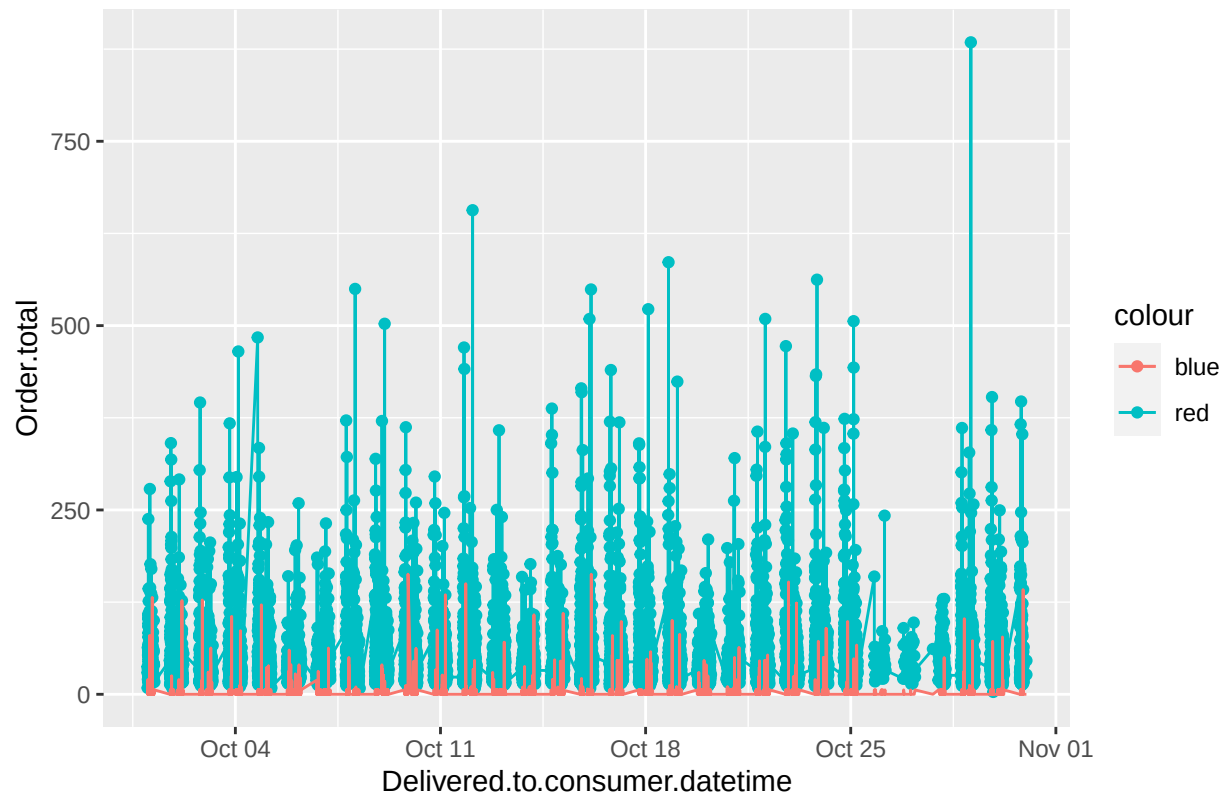
HOW ON DEMAND AND ORDERED FAIRS



few deliveries were made on demand, but Palo Alto having high impromptu demand, consider seeking employer to deal with on demand deliveries for maximised sales

```
Deliverydata %>% ggplot()+geom_line(aes(x=Delivered.to.consumer.datetime,y=Order.total,color="red"))+
  geom_point(aes(x=Delivered.to.consumer.datetime,y=Order.total,color="red")) +
  geom_line(aes(x=Delivered.to.consumer.datetime,y=Amount.of.discount,color="blue"))+
  labs(title = "ORDERS CHANGE WITH TIME")+theme(plot.title = element_text(hjust = 0.5,face = "bold",color = "red"))
```

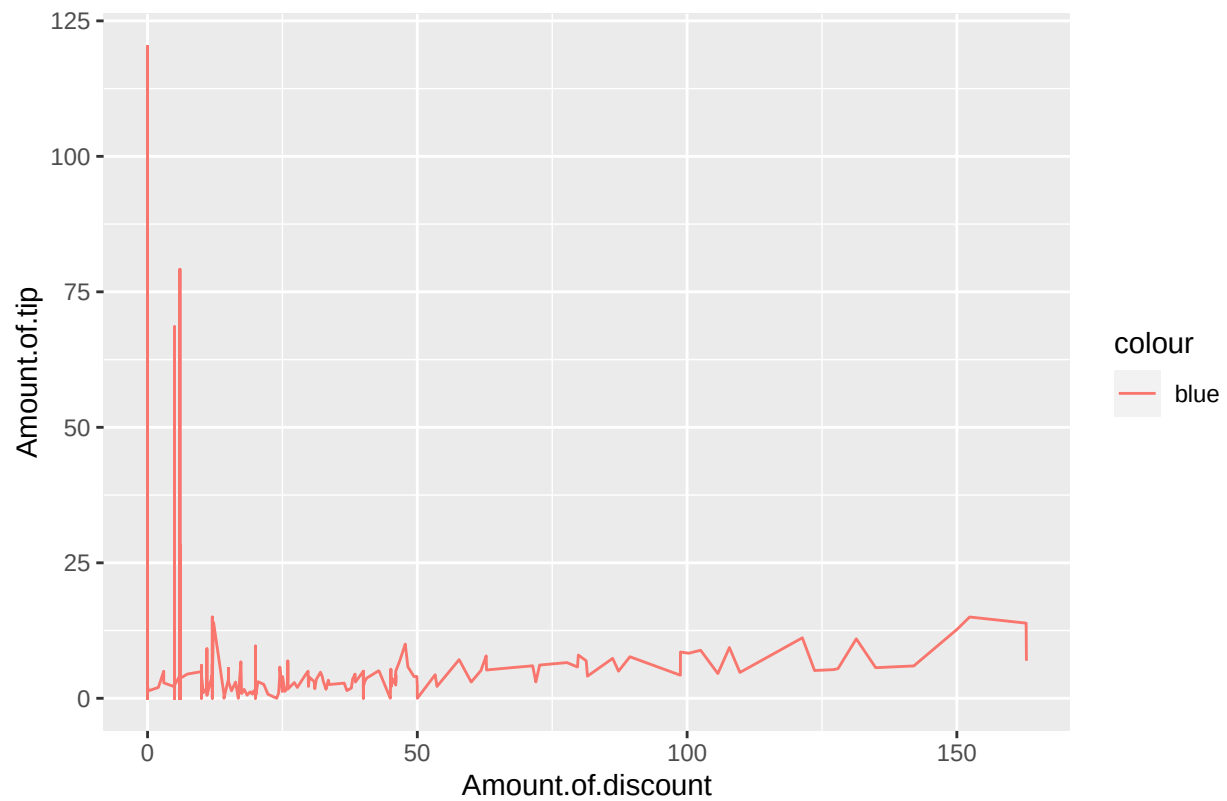
ORDERS CHANGE WITH TIME



there is positive relation between total orders and discount offered on delivery, consider standardized discounts for maximized sales

```
ggplot(Deliverydata, aes(x=Amount.of.discount, y=Amount.of.tip, color="blue")) + geom_line() +
  labs(title = "DISCOUBTS CHANGE IN RELATION TO TIPS") + theme(plot.title = element_text(hjust = 0.5, face
```

DISCOUBTS CHANGE IN RELATION TO TIPS

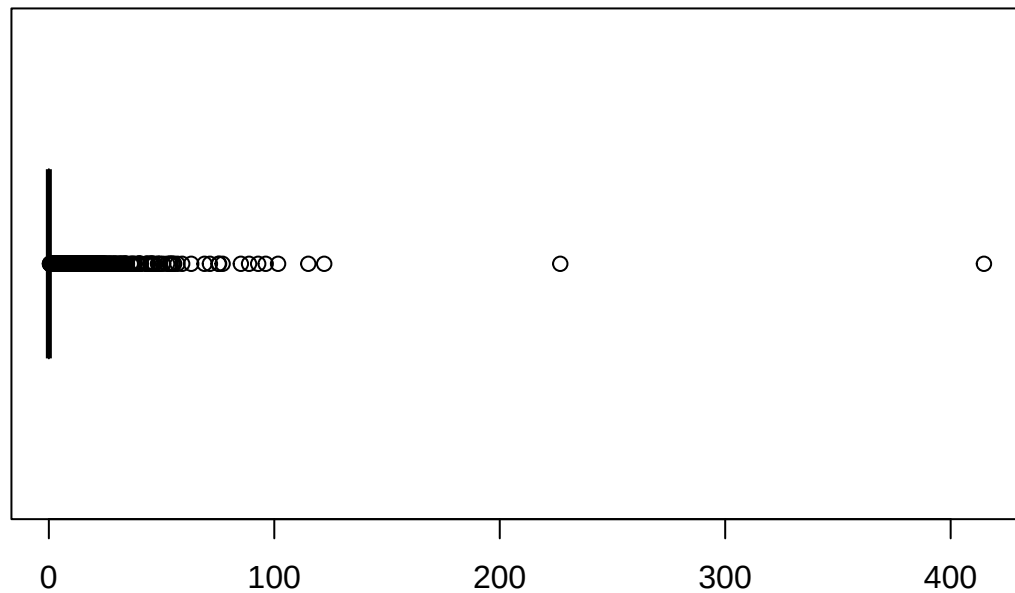


positive association of tip and discount is significant when discount exceeds 140, others consider keeping it at 20, for it registers upward delivery signifying high sales

```
mean(Deliverydata$Refunded.amount)
```

```
## [1] 0.5622101
```

```
boxplot(Deliverydata$Refunded.amount, horizontal = TRUE)
```



there are rampant evidence of outlier(extreme refunding,with highest regesterd being 430+)

```
Deliverydata %>% select(Delivery.Region,Order.total) %>% group_by(Delivery.Region) %>% summarise(count=
marked
```

```
## # A tibble: 4 x 2
##   Delivery.Region count
##   <fct>          <int>
## 1 Mountain View    2763
## 2 None              15
## 3 Palo Alto       8633
## 4 San Jose        2113
```

palo Alto region offers you greatest marked for your product consider prioritizing it and offering giveaways to san jose for improved sales.