Case study 2

Sharleen Kong

2022-04-10

Case study 2

```
customer <- read.csv("casestudy.csv")

# Total revenue for current year
customer_rev <- customer %>%
group_by(year) %>%
summarize(total_revenue = sum(net_revenue))

customer_rev
```

```
## # A tibble: 3 x 2
##    year total_revenue
##   <int>         <dbl>
## 1  2015     29036749.
## 2  2016     25730944.
## 3  2017     31417495.
```

```
# also shows revenue from existing customers current/prior year
```

```
# New Customer Revenue e.g. new customers not present in previous year only
customer_new <- customer %>%
group_by(customer_email) %>%
mutate(first_sale=min(year))%>%
filter(year == first_sale) %>%
ungroup()

new_rev <- customer_new %>%
group_by(year) %>%
summarize(total_revenue = sum(net_revenue))
new_rev
```

```
## # A tibble: 3 x 2
##    year total_revenue
##   <int>         <dbl>
## 1  2015     29036749.
## 2  2016     18245491.
## 3  2017     28676608.
```

```r
# 2015 data does not make sense here since we do not have data for 2014
```

```r
# 2015-2016 growth
growth15_16 <- customer_rev$total_revenue[which(customer_rev$year == 2016)] - customer_rev$total_revenue
growth15_16
```

```
## [1] -3305806
```

```r
# 2016-2017 growth
growth16_17 <- customer_rev$total_revenue[which(customer_rev$year == 2017)] - customer_rev$total_revenue
growth16_17
```

```
## [1] 5686551
```

```r
# revenue lost from attrition
#2015 - 2016
growth15_16/customer_rev$total_revenue[which(customer_rev$year == 2015)]
```

```
## [1] -0.113849
```

```r
#2016-2017
growth16_17/customer_rev$total_revenue[which(customer_rev$year == 2016)]
```

```
## [1] 0.2210005
```

```r
# Total customer current/prior year
customer_count <- customer %>%
group_by(year) %>%
summarize(count = n())
customer_count
```

```
## # A tibble: 3 x 2
##    year  count
##   <int>  <int>
## 1  2015 231294
## 2  2016 204646
## 3  2017 249987
```

```r
# New customers
count_new <- customer_new %>%
group_by(year) %>%
summarize(count = n())
count_new
```

```
## # A tibble: 3 x 2
##    year  count
##   <int>  <int>
## 1  2015 231294
## 2  2016 145062
## 3  2017 228262
```

```
# 2015 not make sense since we do not have info for 2014.
```

```
#lost customers in 2016
customer_lost16 <- customer %>%
group_by(customer_email) %>%
mutate(last_sale = max(year))%>%
  filter(last_sale == 2015) %>%
ungroup()%>%
  group_by(year)%>%
summarize(count = n())
customer_lost16
```

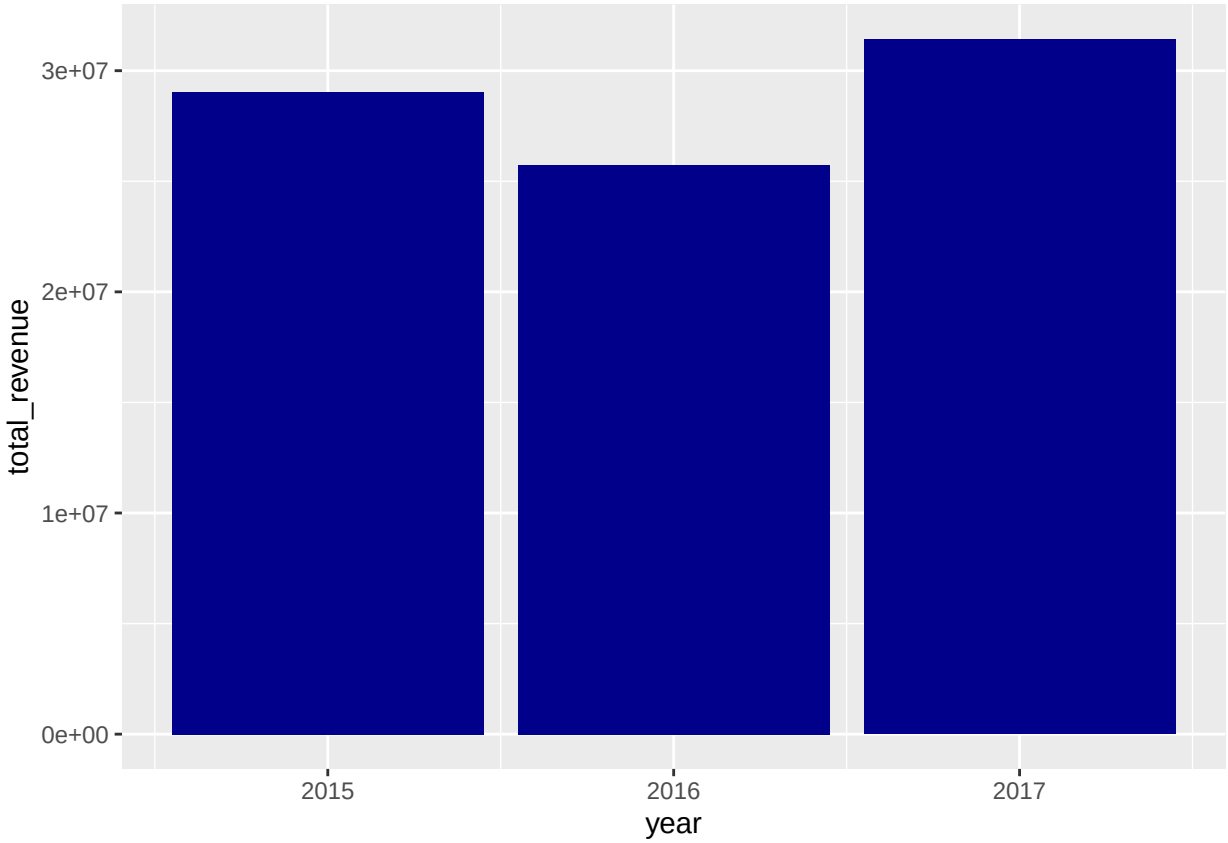
```
## # A tibble: 1 x 2
##    year  count
##   <int>  <int>
## 1  2015 170944
```

```
#lost customers in 2017
customer_lost17 <- customer %>%
group_by(customer_email) %>%
mutate(last_sale = max(year))%>%
filter(last_sale == 2016) %>%
  ungroup()%>%
  group_by(year)%>%
  filter(year == 2016)%>%
summarize(count = n())
customer_lost17
```

```
## # A tibble: 1 x 2
##    year  count
##   <int>  <int>
## 1  2016 183687
```

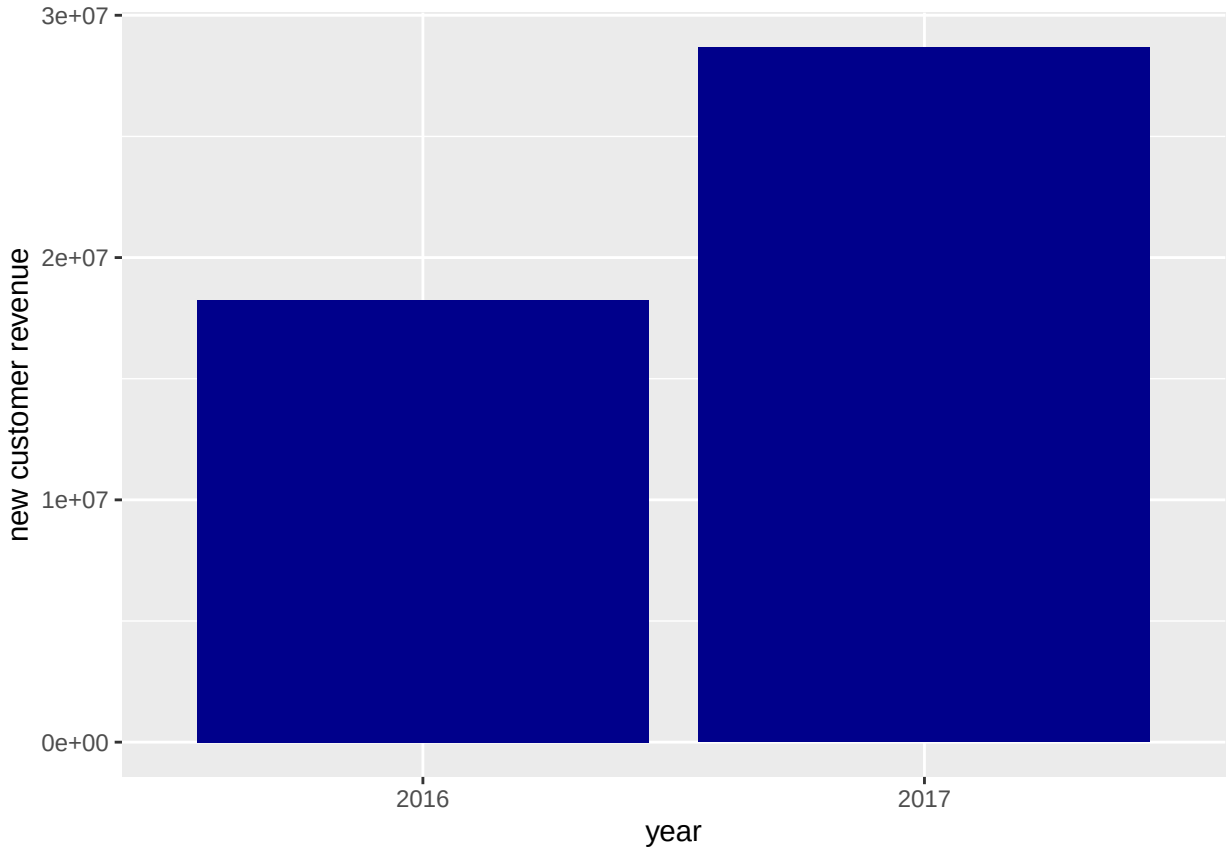
plots

```
ggplot(customer_rev,aes(x = year,y=total_revenue))+geom_bar(stat = "identity",fill="darkblue")
```



2017 has the highest revenue while 2016 has the lowest revenue.

```
new_rev$year <- as.factor(new_rev$year)
ggplot(new_rev[-1,],aes(x = year,y=total_revenue))+geom_bar(stat = "identity",fill="darkblue")+ylab("ne
```



There is a large jump of new customer revenue from 2016 to 2017.