Analysis of Prices

Floriano Peixoto

Here we are going to analyse the prices each competitor gives over time in a data set and discover what it has give us.

The data is described as bellow:

- PROD_ID: Product ID. the values varies between P1 to P9;
- DATE_EXTRACTION: Date and Time of the extraction of the competitors' price, under YYYY-MM-DD HH:MM:SS format;
- COMPETITOR: Competitors' ID (C1 to C6);
- COMPETITOR_PRICE: Competitors' price per product, which can depend on the payment method;
- PAY_TYPE: Payment Method (1=deferred payment, 2=immediate payment)

```
prices <-  read.csv("comp_prices.csv", sep=",", stringsAsFactors = T)

tail(prices)
```

```
##       PROD_ID     DATE_EXTRACTION COMPETITOR COMPETITOR_PRICE PAY_TYPE
## 50109      P5 2015-10-11 20:10:34         C3            819.00        2
## 50110      P5 2015-10-11 20:10:34         C1            819.00        2
## 50111      P5 2015-10-11 20:10:34         C2            853.52        2
## 50112      P5 2015-10-12 08:11:27         C2            853.52        1
## 50113      P5 2015-10-12 08:11:27         C3            819.00        1
## 50114      P5 2015-10-12 08:11:27         C1            819.00        1
```

As the data documentation tells us the competitor is monitored twice a day.

```
prices <- arrange(prices, COMPETITOR, PROD_ID, DATE_EXTRACTION)

tail(prices)
```

```
##       PROD_ID     DATE_EXTRACTION COMPETITOR COMPETITOR_PRICE PAY_TYPE
## 50109      P9 2015-10-14 08:11:39         C6         39999.00        1
## 50110      P9 2015-10-14 08:11:39         C6         39999.00        2
## 50111      P9 2015-10-14 20:10:24         C6           399.99        1
## 50112      P9 2015-10-14 20:10:24         C6           399.99        2
## 50113      P9 2015-10-14 20:11:30         C6           399.99        1
## 50114      P9 2015-10-14 20:11:30         C6           399.99        2
```

Ordering the set by COMPETITOR, PROD_ID and DATE_EXTRACTION we can see that it´s true, we can turn that into a single line displaying it as min and max price.

```
prices_min_max <- mutate(prices, YEAR = lubridate::year(DATE_EXTRACTION),
                           MONTH = lubridate::month(DATE_EXTRACTION),
                           DAY = lubridate::day(DATE_EXTRACTION),
                           PROD_ID = as.factor(PROD_ID)) %>%
                  group_by(COMPETITOR, PROD_ID, YEAR, MONTH, DAY) %>%
                  summarise(MIN_PRICE = min(COMPETITOR_PRICE),
                            MAX_PRICE = max(COMPETITOR_PRICE)) %>%
                  arrange(COMPETITOR, PROD_ID, YEAR, MONTH, DAY)
```

```
tail(prices_min_max)
```

```
## Source: local data frame [6 x 7]
## Groups: COMPETITOR, PROD_ID, YEAR, MONTH [1]
##
##   COMPETITOR PROD_ID  YEAR MONTH   DAY MIN_PRICE MAX_PRICE
##       <fctr>  <fctr> <dbl> <dbl> <int>     <dbl>     <dbl>
## 1         C6      P9  2015    10     9    399.99    399.99
## 2         C6      P9  2015    10    10    399.99    399.99
## 3         C6      P9  2015    10    11    399.99    399.99
## 4         C6      P9  2015    10    12    399.99    399.99
## 5         C6      P9  2015    10    13    399.99    399.99
## 6         C6      P9  2015    10    14    399.99  39999.00
```

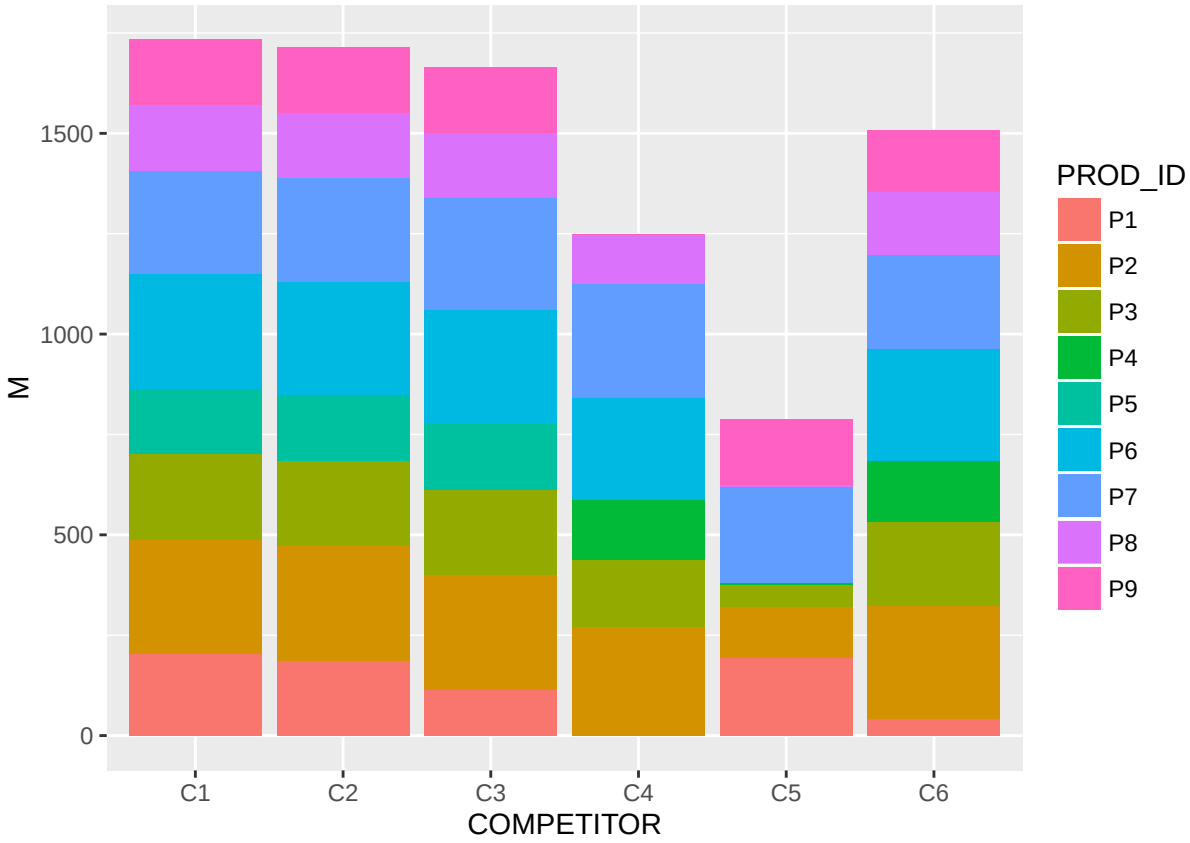
Distribuition of products by competitor

With the data in our hands we can try to discrinate the presence of competitors in the market showing the products it sells.

```
presence_market <- group_by(prices_min_max, COMPETITOR, PROD_ID) %>%
                   summarise(M = n())

tail(presence_market)
```

```
## Source: local data frame [6 x 3]
## Groups: COMPETITOR [1]
##
##   COMPETITOR PROD_ID     M
##       <fctr>  <fctr> <int>
## 1         C6      P3   207
## 2         C6      P4   153
## 3         C6      P6   278
## 4         C6      P7   235
## 5         C6      P8   156
## 6         C6      P9   154
```

```
g <- ggplot(presence_market, aes(y = M, x = COMPETITOR))

g +  geom_bar(stat = "identity", aes(fill = PROD_ID))
```



The plot show us the there are 4 majors competitors C1, C2, C3 and C6, the minor one is C5. It does not seems to have one great competitor, but we have a weaker one in the set. It seems that every major compoetitor sells every product, but the other two has some missing products.

The competitor with lowest price by product

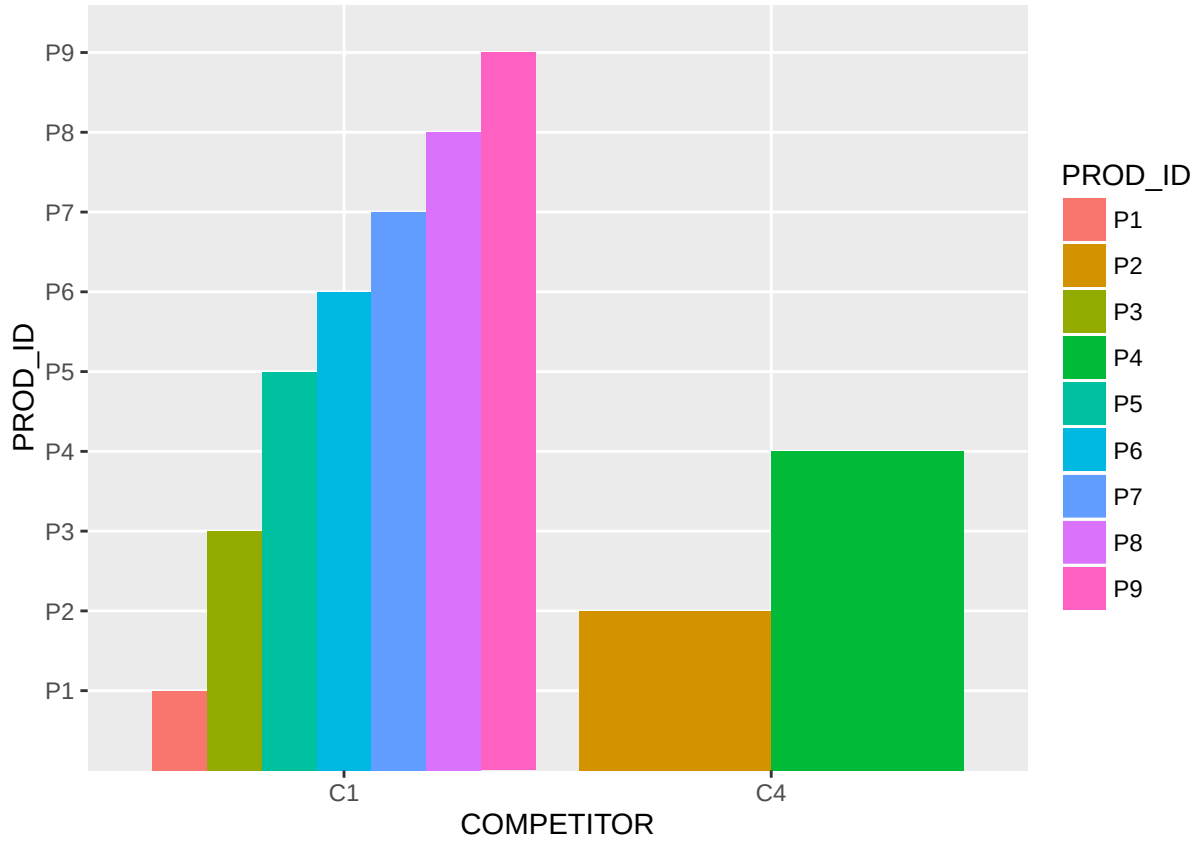
With our data we can see now what are the competitors with the best price by product.

```
competitor_best_price <- group_by(prices_min_max, PROD_ID, COMPETITOR) %>%
                         summarise(PRICE = min(MIN_PRICE)) %>%
                         slice(which.min(PRICE))

competitor_best_price
```

```
## Source: local data frame [9 x 3]
## Groups: PROD_ID [9]
##
##   PROD_ID COMPETITOR   PRICE
##    <fctr>     <fctr>   <dbl>
## 1      P1         C1 1090.00
## 2      P2         C4  506.92
## 3      P3         C1  879.12
## 4      P4         C4  431.10
## 5      P5         C1  674.10
## 6      P6         C1 1225.97
## 7      P7         C1  588.71
```

```
## 8      P8         C1  359.10
## 9      P9         C1  359.10
```

```
g <- ggplot(competitor_best_price, aes(y = PROD_ID, x = COMPETITOR))

g +  geom_bar(stat = "identity", aes(fill = PROD_ID), position = "dodge")
```



The competitor C1 has the best prices for almost all products, loosing to C4 in just two products P2 and P4.

The competitor with the highest prices

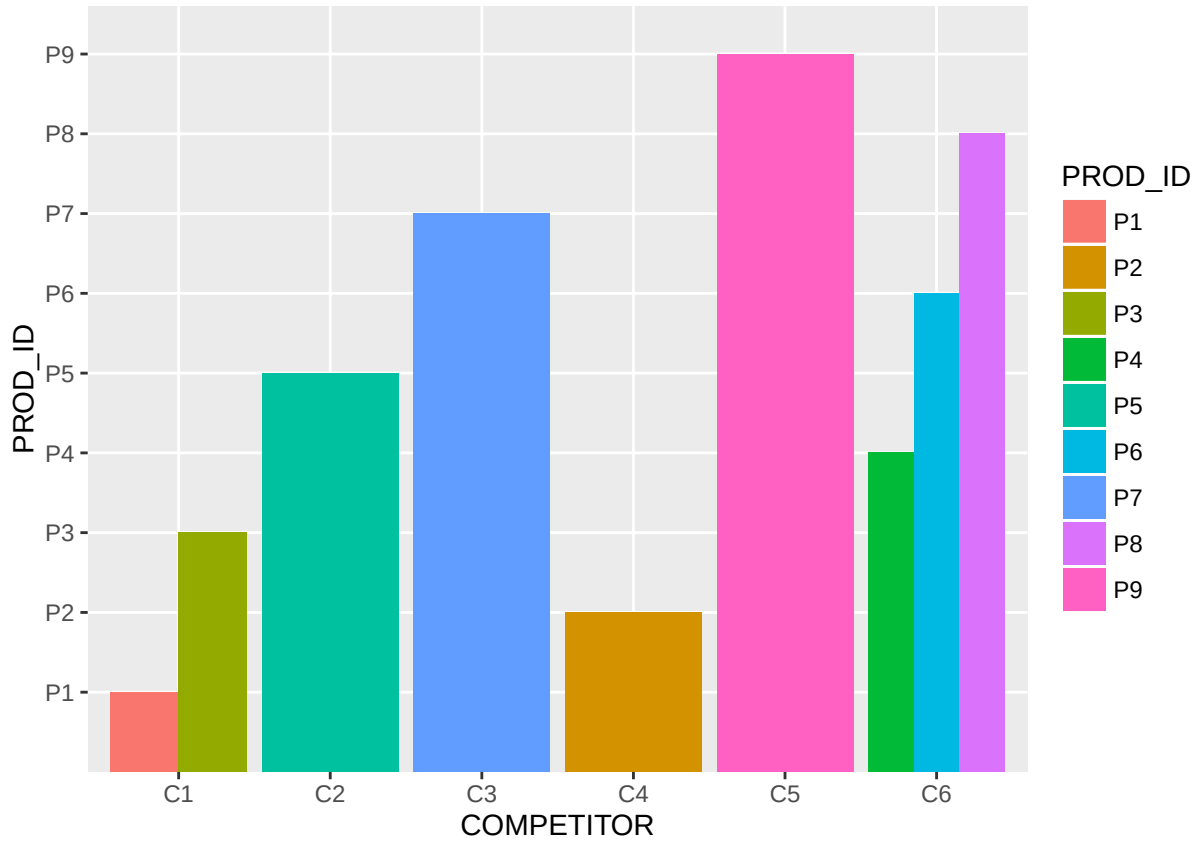
Now let´s see the competitor with the highest prices.

```
competitor_worst_price <- group_by(prices_min_max, PROD_ID, COMPETITOR) %>%
                          summarise(PRICE = max(MAX_PRICE)) %>%
                          slice(which.max(PRICE))

competitor_worst_price
```

```
## Source: local data frame [9 x 3]
## Groups: PROD_ID [9]
##
##   PROD_ID COMPETITOR  PRICE
##    <fctr>     <fctr>  <dbl>
## 1      P1         C1 149900
## 2      P2         C4  79900
## 3      P3         C1 119900
```

```
## 4      P4         C6  49700
## 5      P5         C2  84890
## 6      P6         C6 149900
## 7      P7         C3 104900
## 8      P8         C6  39999
## 9      P9         C5  56900
```

```
g <- ggplot(competitor_worst_price, aes(y = PROD_ID, x = COMPETITOR))

g +  geom_bar(stat = "identity", aes(fill = PROD_ID), position = "dodge")
```



Diferently from the lowest prices, we can see a better distribuition over the competitors.

The lowest and higher prices per competitor by product

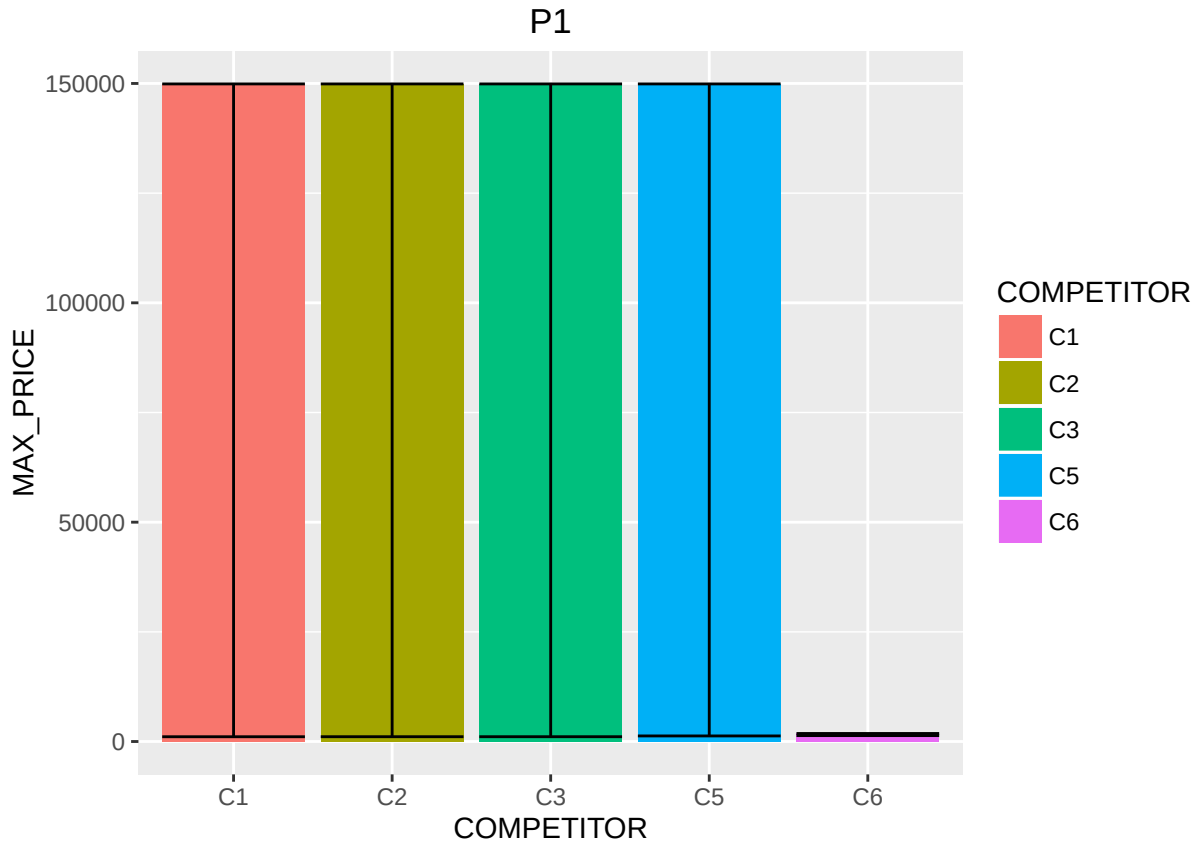
It´s intereting to compare the prices each competitor has, historically, for each product in terms to understand more about their behaviors during the year.

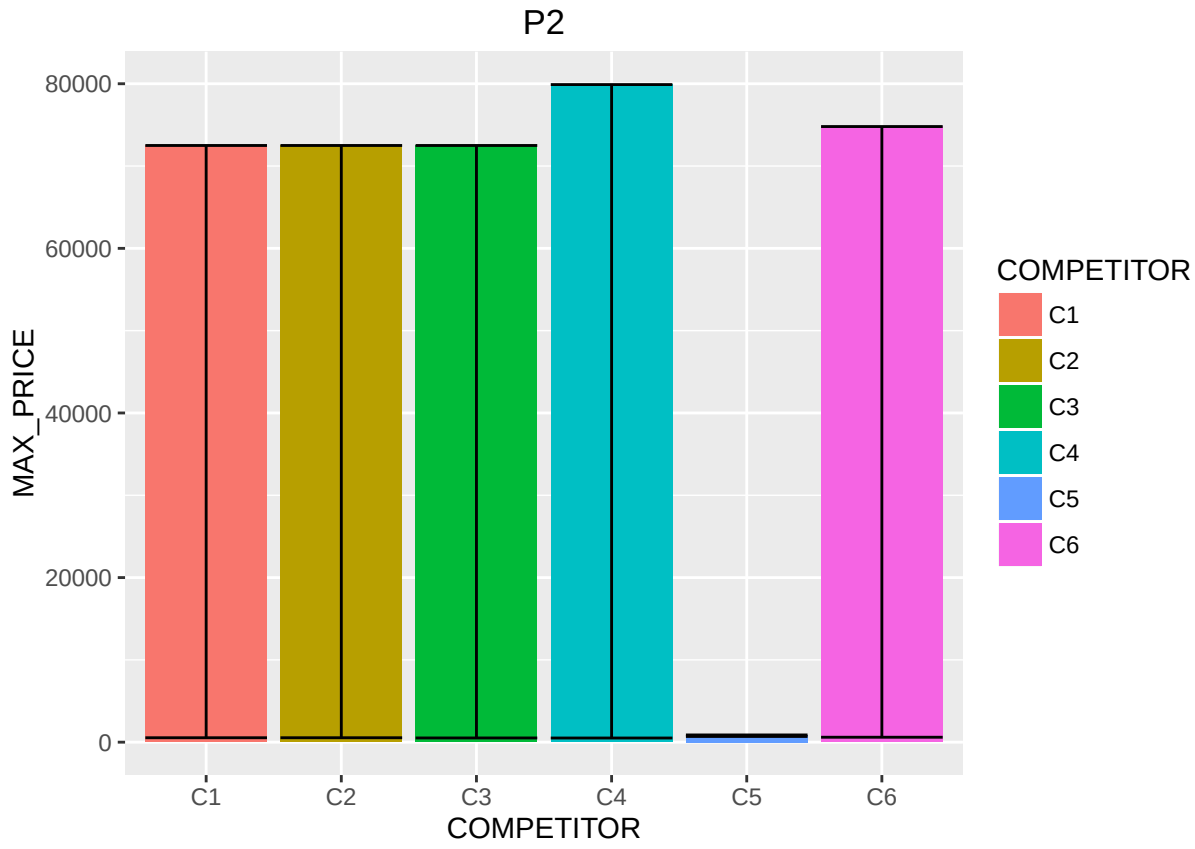
```
comp_best_worst_pric_prod <- group_by(prices_min_max, COMPETITOR, PROD_ID) %>%
                             summarise(MIN_PRICE = min(MIN_PRICE),
                                       MAX_PRICE = max(MAX_PRICE)) %>%
                             arrange(COMPETITOR, PROD_ID)

comp_best_worst_pric_prod
```

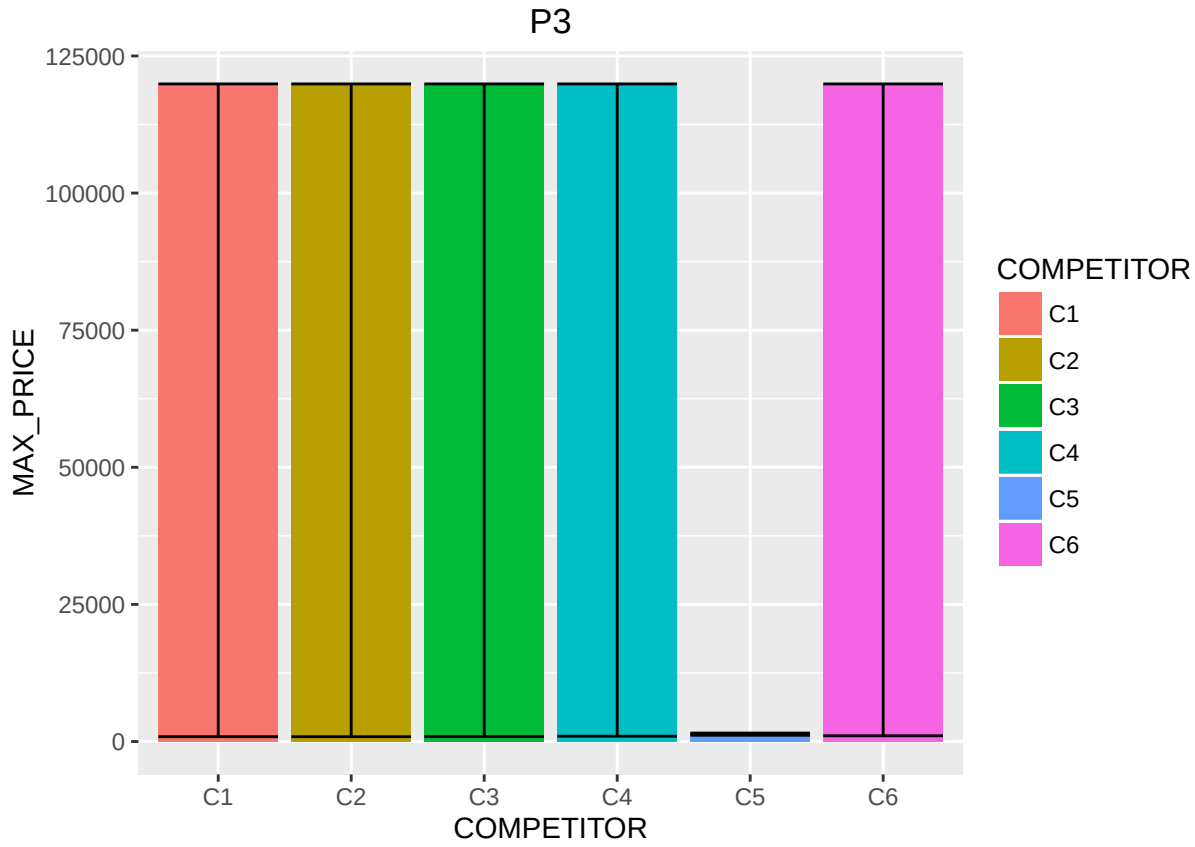
```
## Source: local data frame [46 x 4]
## Groups: COMPETITOR [6]
```

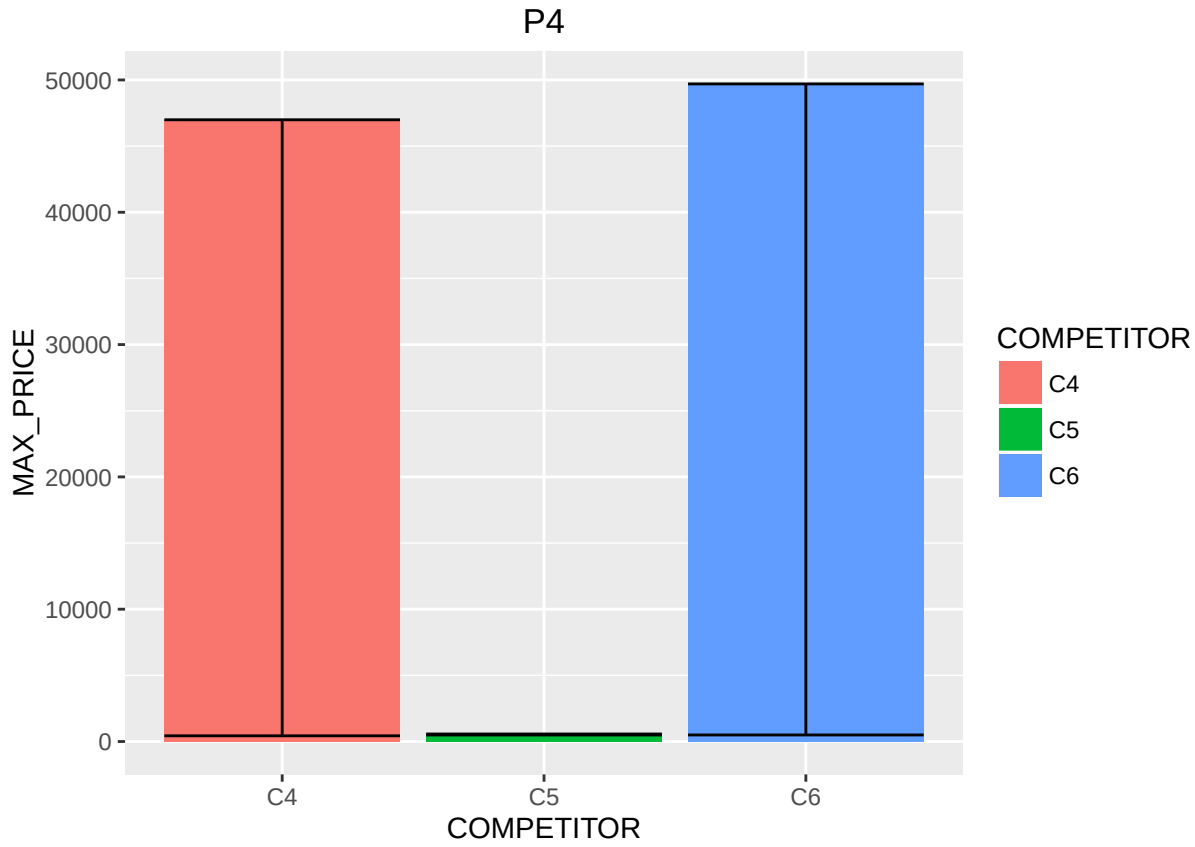
```
##
##    COMPETITOR PROD_ID MIN_PRICE MAX_PRICE
##        <fctr>  <fctr>     <dbl>     <dbl>
## 1          C1      P1   1090.00    149900
## 2          C1      P2    539.10     72498
## 3          C1      P3    879.12    119900
## 4          C1      P5    674.10     81900
## 5          C1      P6   1225.97    148000
## 6          C1      P7    588.71     84090
## 7          C1      P8    359.10     39300
## 8          C1      P9    359.10     39300
## 9          C2      P1   1099.00    149900
## 10         C2      P2    539.10     72498
## # ... with 36 more rows
```

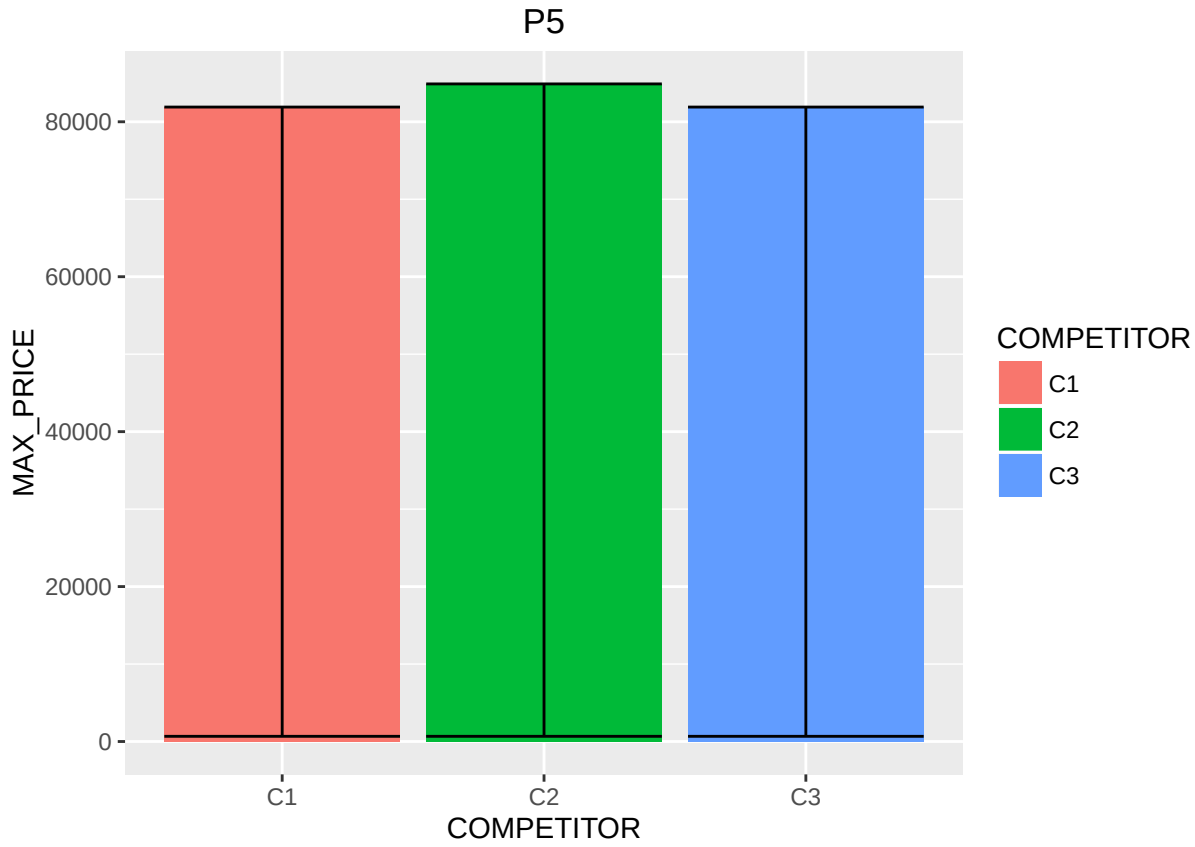
```
for(product in levels(comp_best_worst_pric_prod$PROD_ID)){

  g <- ggplot(filter(comp_best_worst_pric_prod, PROD_ID == product), aes(y = MAX_PRICE, x = COMPETITOR,

  g <- g +  geom_bar(stat = "identity") +
  geom_errorbar()   +
  ggtitle(product)


  print(g)
}
```

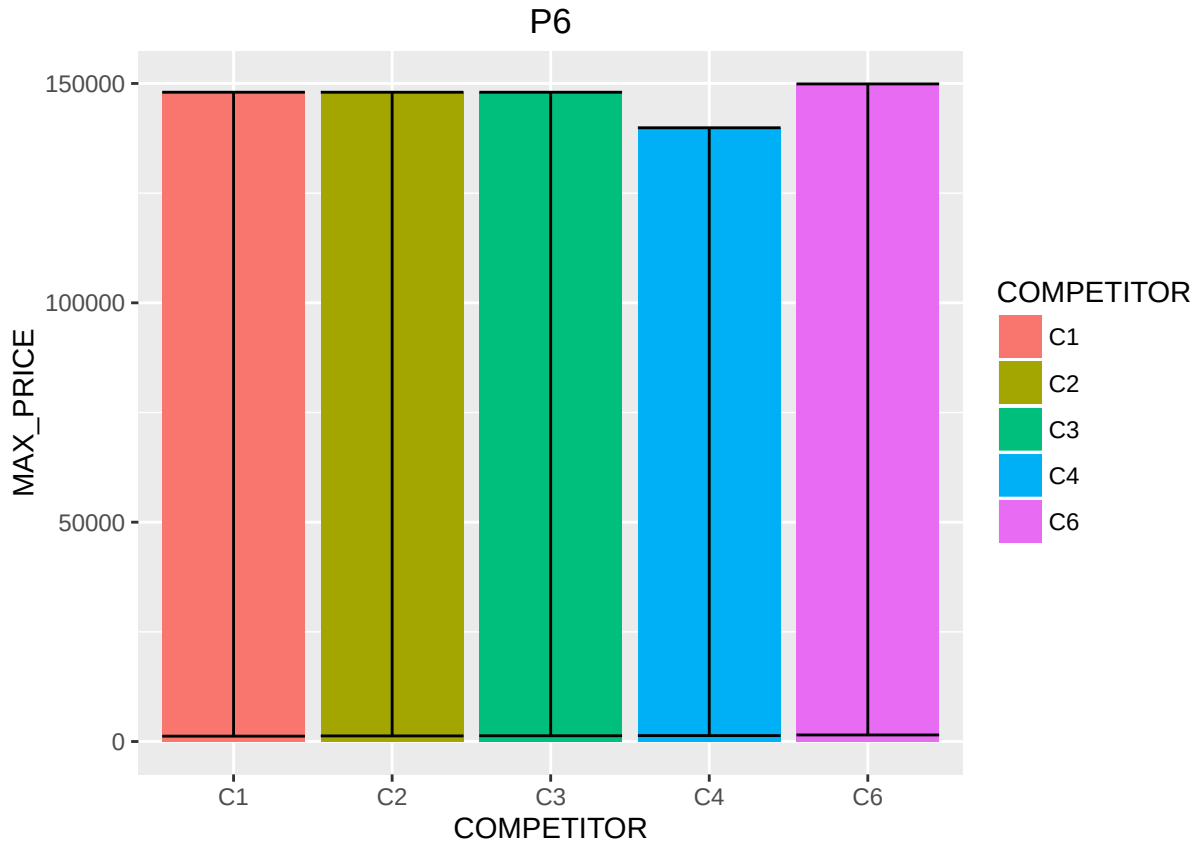

P1
150000
100000
50000
0
MAX_PRICE
C1
C2
C3
C5
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C5
C6

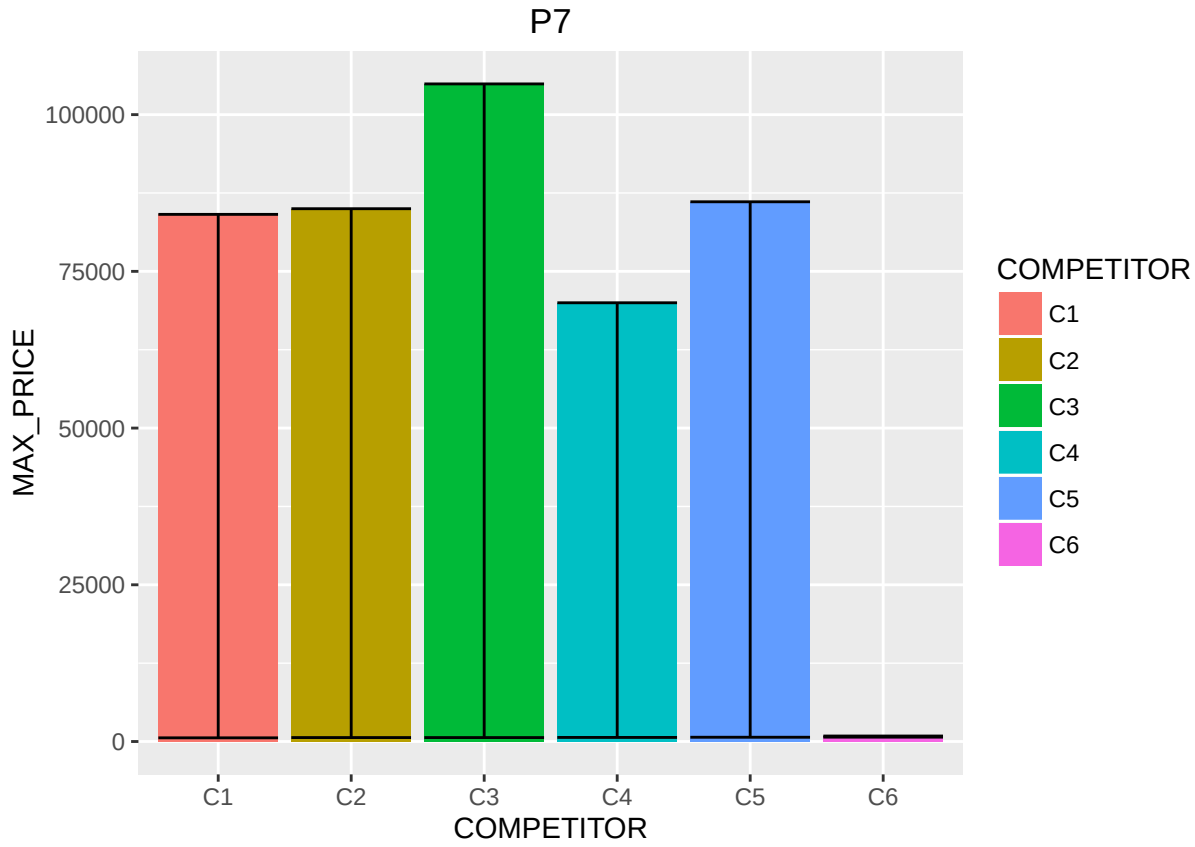

P2
80000
60000
40000
20000
0
MAX_PRICE
C1
C2
C3
C4
C5
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C4
C5
C6

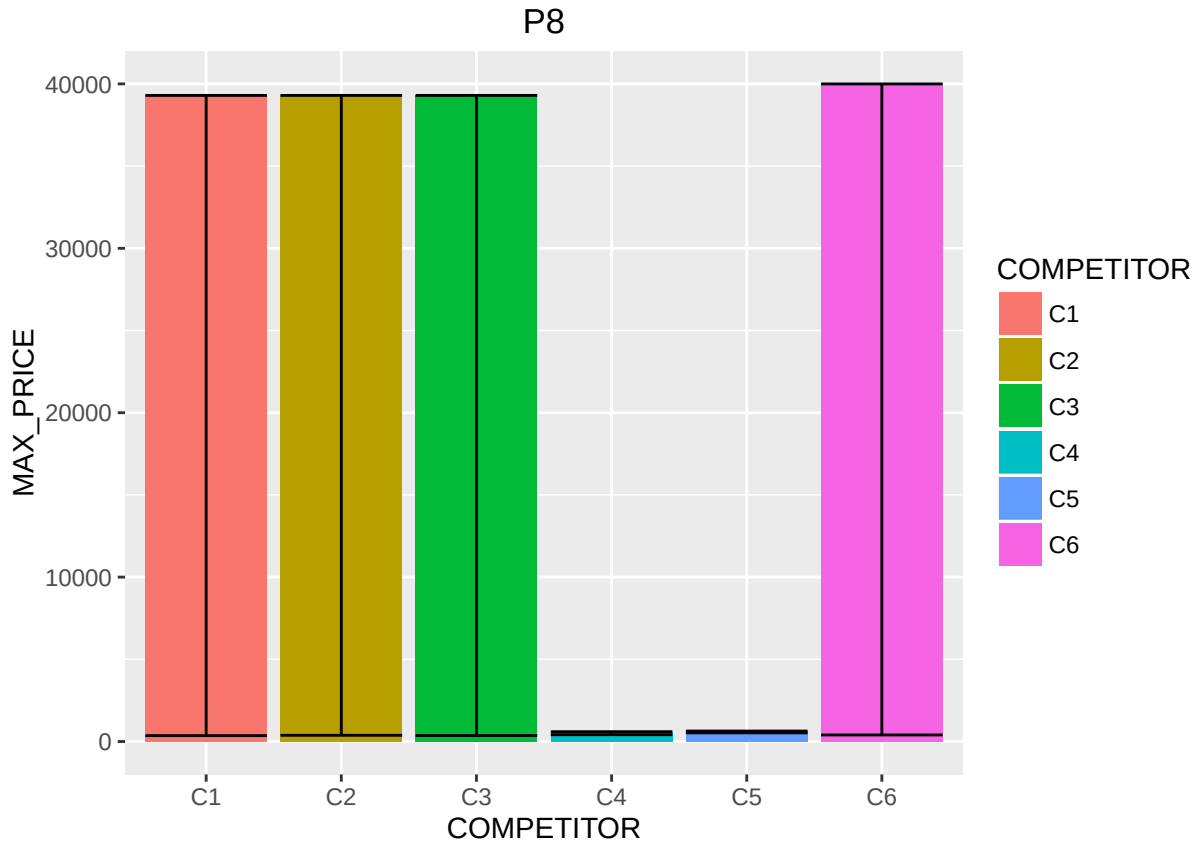

P3
125000
100000
75000
50000
25000
0
MAX_PRICE
C1
C2
C3
C4
C5
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C4
C5
C6

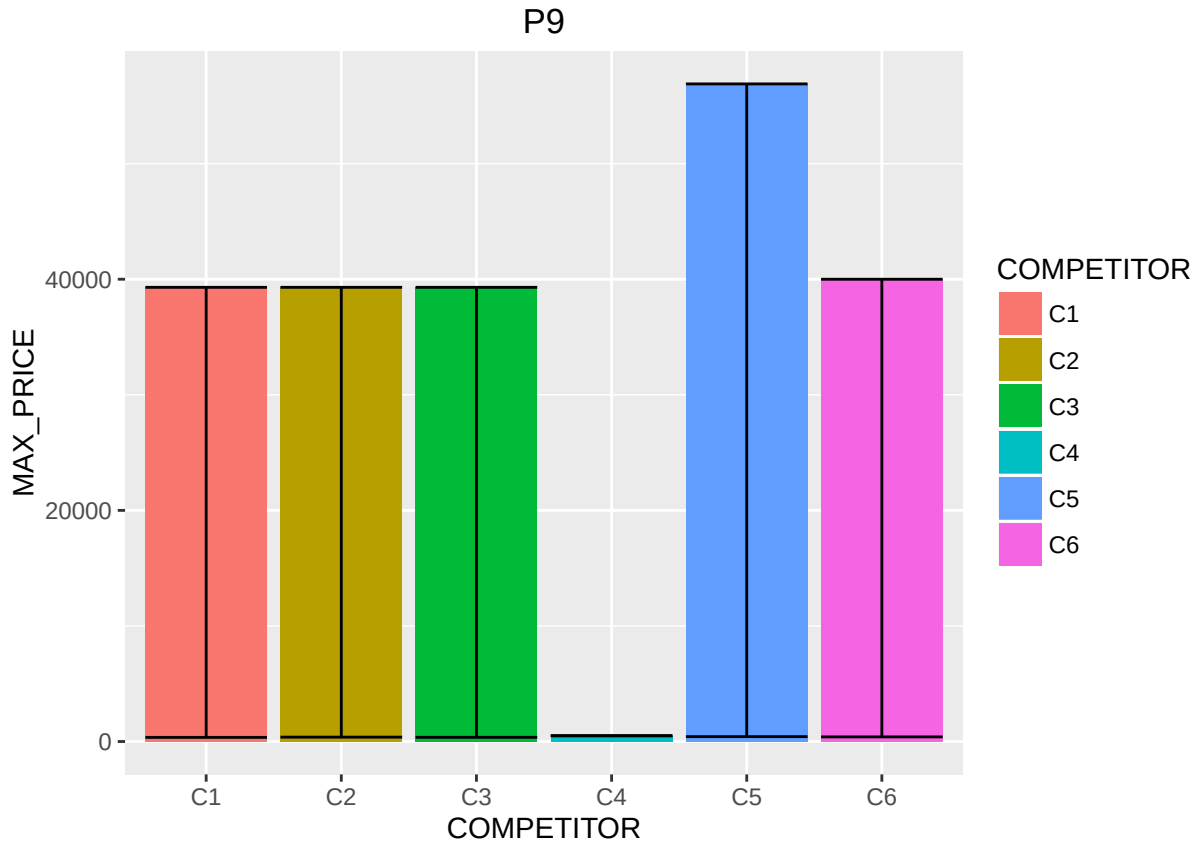

P4
50000
40000
30000
20000
10000
0
MAX_PRICE
C4
C5
C6
COMPETITOR
COMPETITOR
C4
C5
C6


P5
MAX_PRICE
80000
60000
40000
20000
0
C1
C2
C3
COMPETITOR
COMPETITOR
C1
C2
C3


P6
150000
100000
50000
0
MAX_PRICE
C1
C2
C3
C4
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C4
C6


P7
100000
75000
50000
25000
0
MAX_PRICE
C1
C2
C3
C4
C5
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C4
C5
C6


P8
40000
30000
20000
10000
0
MAX_PRICE
C1
C2
C3
C4
C5
C6
COMPETITOR
COMPETITOR
C1
C2
C3
C4
C5
C6



As we can see some competitors have similar prices for some products, but others have a really lower costs. And as we could see before not all of them sell all products, so some of them has higher presence in the market.