

# Sales transactions

Yarden Toren & Oran Shemesh

March 30, 2016

For this assignment we found dataset of the sales carried out in January 2009. The data is about 3 main products, where and who bought them and how much it costs. In the dataset you can find the following data: transaction date, the product that was purchased, the price of the product, the payment type, the name of the person who bought the product, The city state and country where the purchase was made, the latitude and the longitude of the place where the purchase was made. Payment type: Amex=1 Diners=2 Mastercard=3 Visa=4

```
##      Product  Transaction_date  ProductNum      Price
## Product1:847  Min.      : 1.00    Min.      :1.000  Min.      : 250
## Product2:136  1st Qu.: 7.00    1st Qu.:1.000  1st Qu.: 1200
## Product3: 15  Median   :14.00   Median   :1.000  Median   : 1200
##              Mean     :14.94   Mean     :1.166  Mean     : 1634
##              3rd Qu.:22.75   3rd Qu.:1.000  3rd Qu.: 1200
##              Max.     :31.00   Max.     :3.000  Max.     :13000
##
##      Payment_Type      Latitude      Longitude      Name
## Min.      :1.000    Min.      : -41.47    Min.      : -159.485  Sarah      : 11
## 1st Qu.:3.000    1st Qu.: 35.82    1st Qu.: -87.992  Elizabeth:  9
## Median :4.000    Median : 42.32    Median : -73.731  Lisa       :  9
## Mean     :3.213    Mean     : 39.02    Mean     : -41.338  Nicole     :  8
## 3rd Qu.:4.000    3rd Qu.: 51.05    3rd Qu.:  4.917  Kim        :  7
## Max.     :4.000    Max.     : 64.84    Max.     : 174.767  Jessica    :  6
##                                     (Other)   :948
##
##              City      State      Country
## London          : 19  England: 86  United States :463
## Calgary         : 11  CA      : 66  United Kingdom:100
## Den Haag        :  9  NY      : 41  Canada        : 76
## New York        :  9  TX      : 37  Ireland        : 49
## Vancouver       :  8  VA      : 30  Australia      : 38
## Houston         :  7  FL      : 29  Switzerland    : 36
## (Other)         :935  (Other):709  (Other)        :236
```

Here you can see a sample of the data

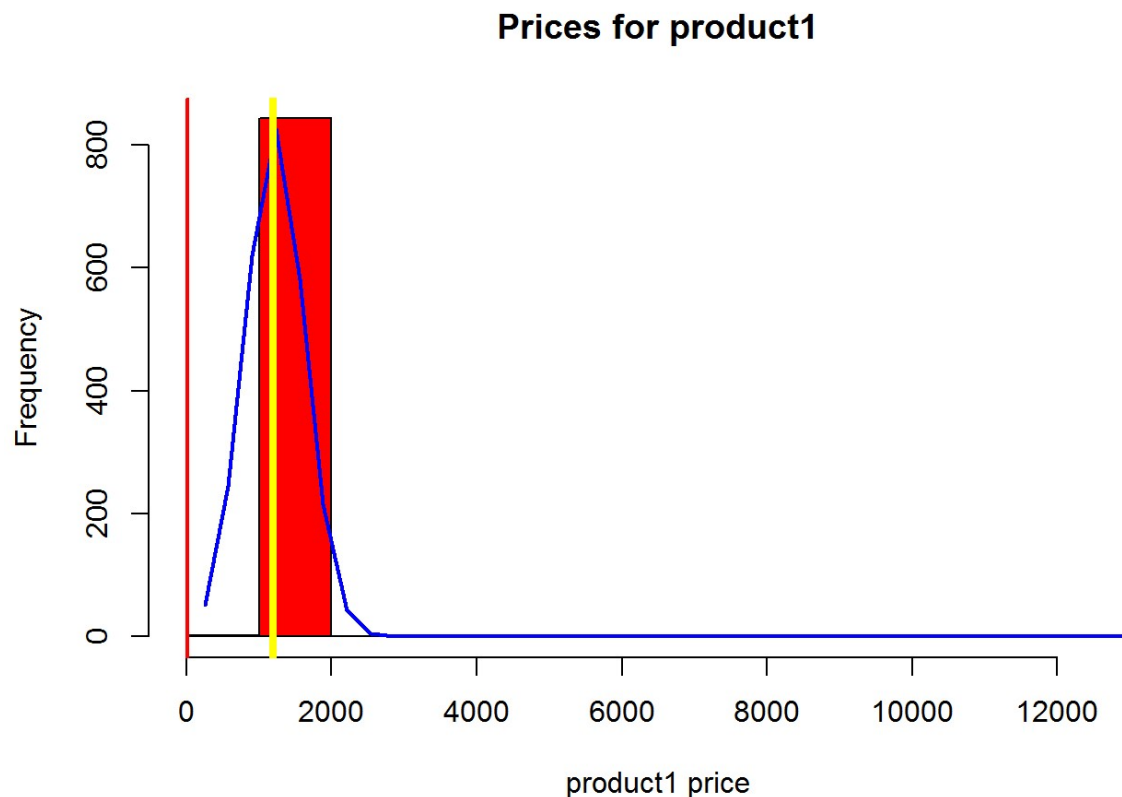
```

##      Product Transaction_date ProductNum Price Payment_Type Latitude
## 1 Product1                1           1  1200             3  51.50000
## 2 Product1                2           1  1200             4  39.19500
## 3 Product1                2           1  1200             3  46.18806
## 4 Product1                3           1  1200             4 -36.13333
## 5 Product1                4           1  1200             4  39.79000
## 6 Product1                4           1  1200             3  40.69361
##      Longitude              Name              City      State
## 1   -1.116667      carolina      Basildon    England
## 2   -94.681940      Betina Parkville              MO
## 3  -123.830000  Federica e Andrea Astoria              OR
## 4   144.750000      Gouya      Echuca Victoria
## 5   -75.238060      LAURENCE Mickleton              NJ
## 6   -89.588890      Fleur Peoria              IL
##      Country
## 1 United Kingdom
## 2 United States
## 3 United States
## 4      Australia
## 5 United States
## 6 United States

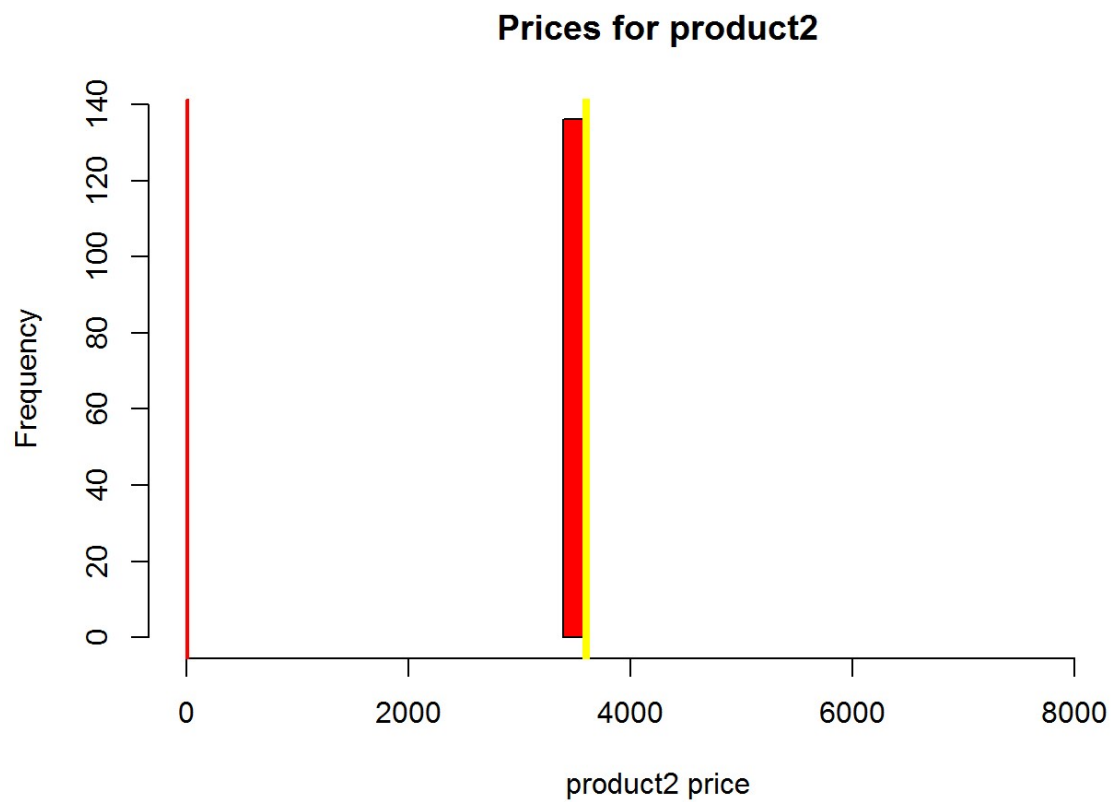
```

## The data analysis

### Frequency of each product price



As we can see, for product1, we have variety of prices.

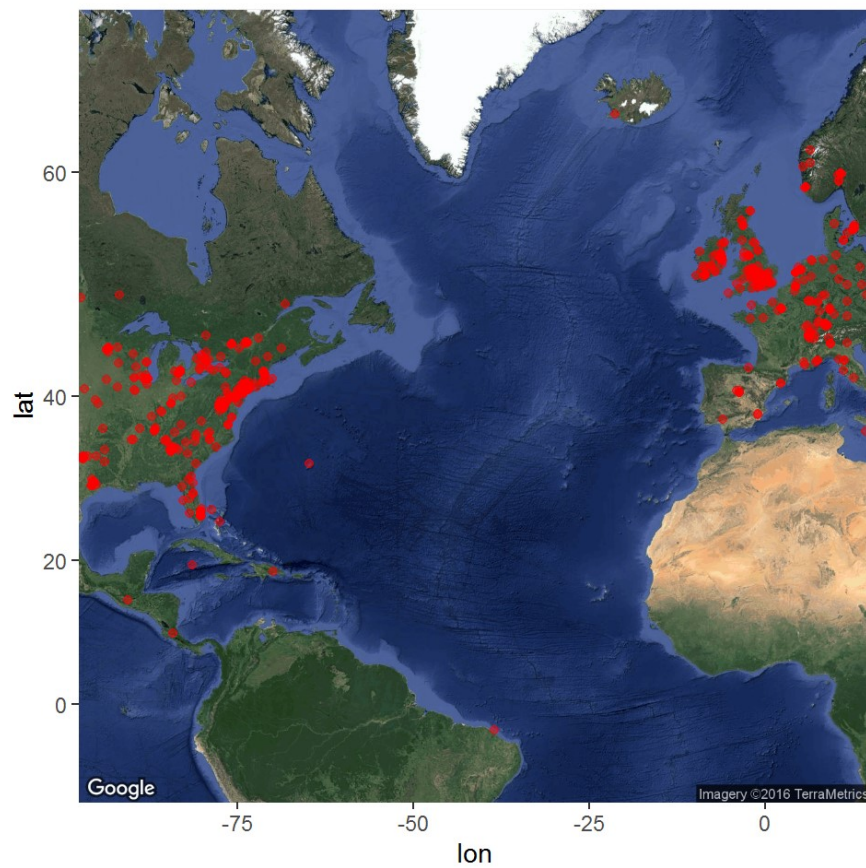


As we can see, for product2, we have only on price



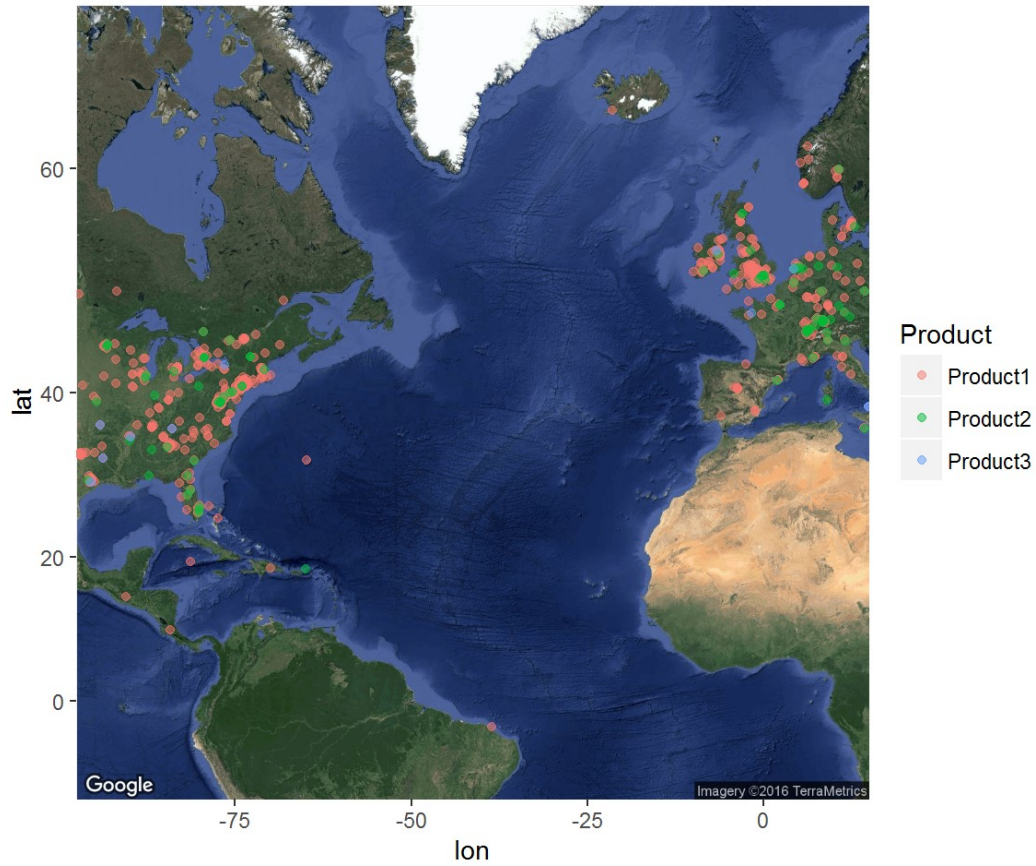
As we can see, for product3, we have only on price

We want to explore more the prices for product1.



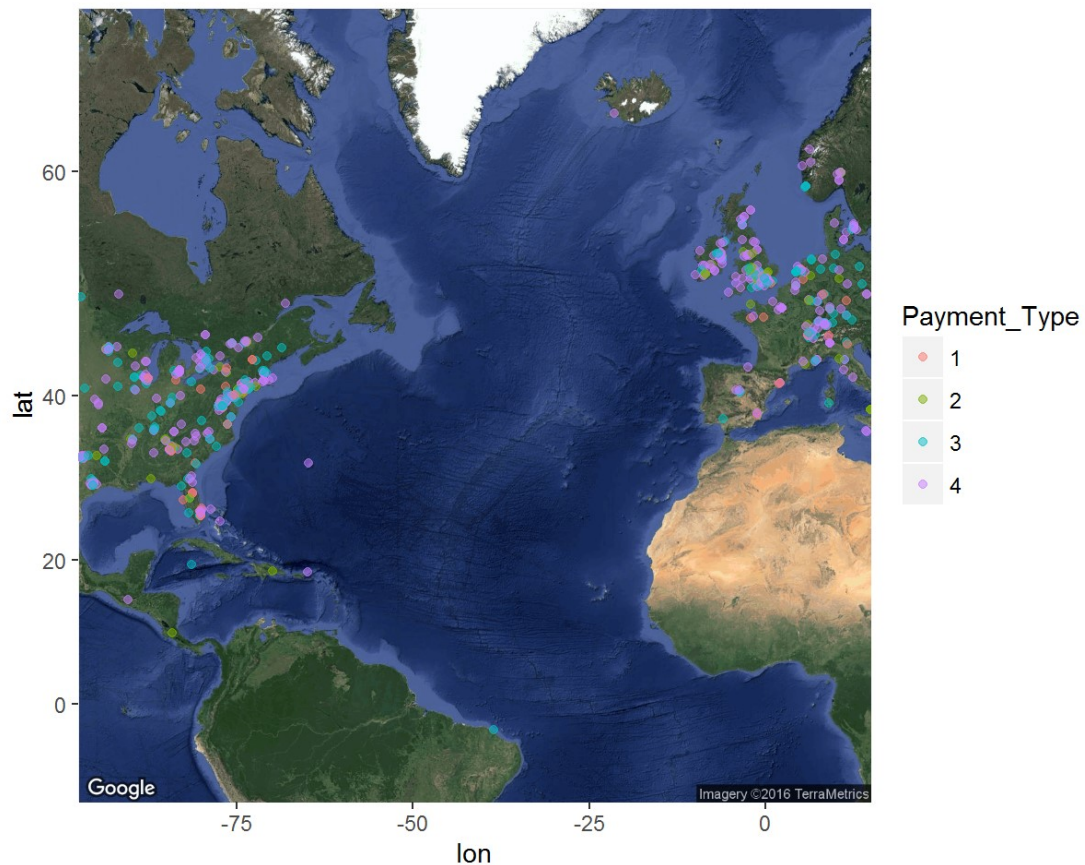
We can see in the map the way that the prices distribut across the world But every location as more than one price. so we can assume that the diffrent prices is not per place

## Frequency of each product



We can see that each product appear in more than one location and each location as more than one product.

## Frequency of each payment type



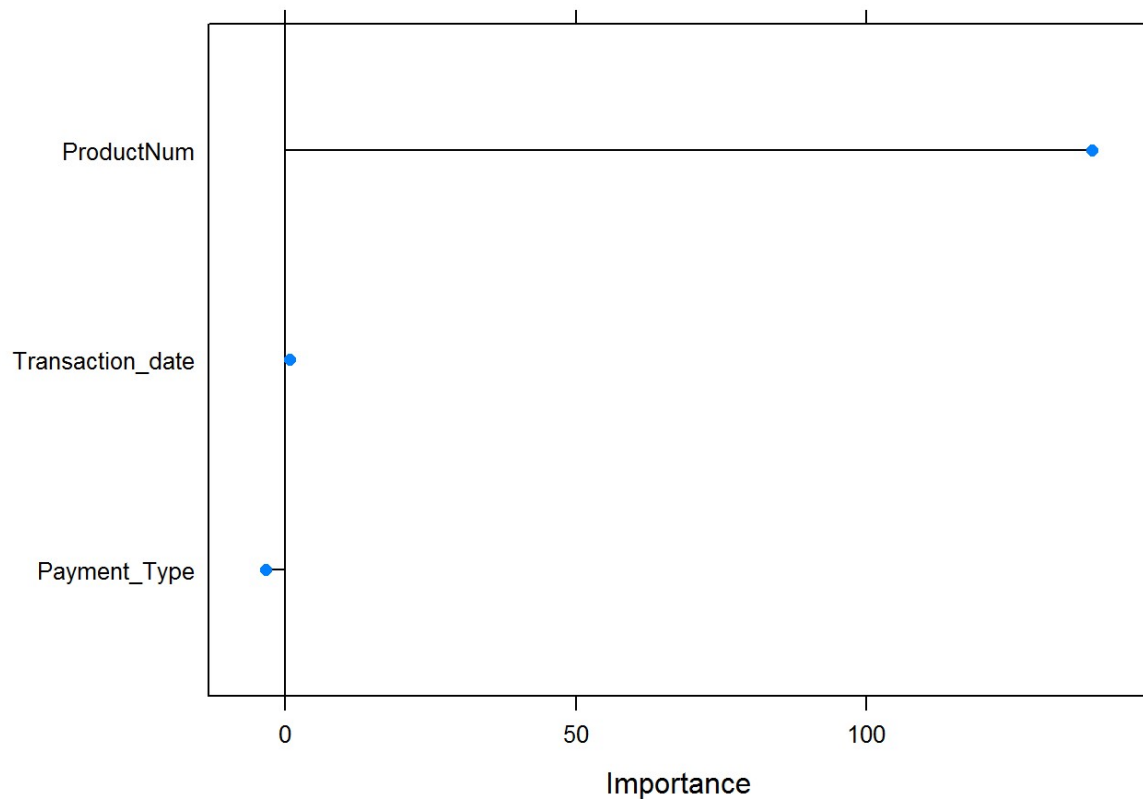
We can see that each location uses more than one payment type. But more people use payment type 3 and 4 (Mastercard and Visa).

## Explore importance between attribute

### By price:

```
## note: only 2 unique complexity parameters in default grid. Truncating the  
grid to 2 .
```

```
## rf variable importance  
##  
## Overall  
## ProductNum      138.7481  
## Transaction_date  0.7408  
## Payment_Type     -3.3036
```

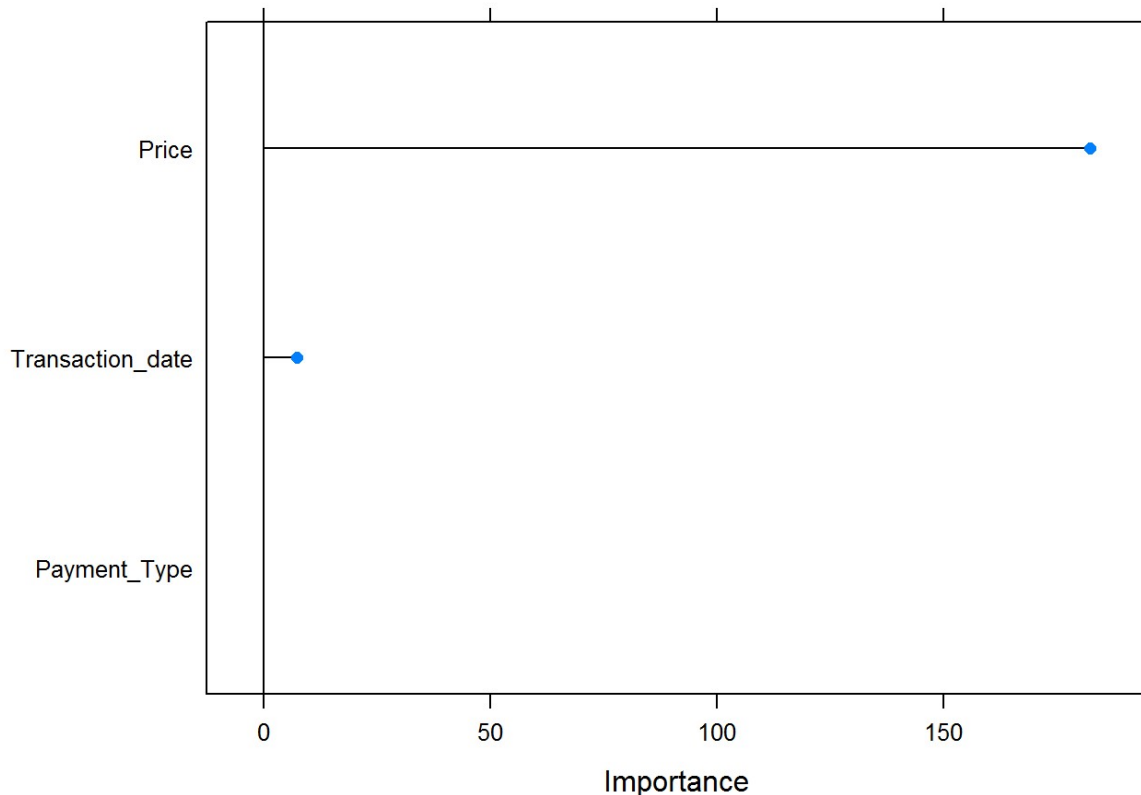


It shows that the product is the most important attribute. payment type attribute is the least important.

## By product:

```
## note: only 2 unique complexity parameters in default grid. Truncating the  
grid to 2 .
```

```
## rf variable importance  
##  
## Overall  
## Price 182.197  
## Transaction_date 7.311  
## Payment_Type 0.000
```



It shows that the price is the most important attribute. payment type attribute is the least important.

## Explore correlation between attribute

```
##           Transaction_date  ProductNum      Price  Payment_Type
## Transaction_date           1.00000000  0.02305024  0.036803494  0.048271160
## ProductNum                0.02305024  1.00000000  0.936085156 -0.018053681
## Price                     0.03680349  0.93608516  1.000000000 -0.008848857
## Payment_Type              0.04827116 -0.01805368 -0.008848857  1.000000000
```

It shows that product and price are highly correlated

## Summary, Conclusions:

In this research we saw that for some products everyone will pay the same price but there is some products that the payment is different between each person.

We couldn't find that for each location has a different price

The price of the product and the date that we bought the product are important when we look at the product.

Product and price is the highly correlated.

## Recommendations:

I think that credit card companies can use this data to target locations who don't use in their credit



card. visa can make a campaign in those locations to get more customers.

If we want to sell a specific product, and we want to get higher payment, We can find out where people buy similar products and what is the price that they are willing to pay for it.