FROM: J. Rodriguez
TO: D. Sherman
Subject: Investigating customer buying patterns

Customer Buying Patterns Report

In the report I will answer the following questions and also present some findings and evidences to support it. At the end you can find some charts and reporting I produced with the statistical software.

1. Do customers in different regions spend more per transaction? Which regions spend the most/least?
2. Are there differences in the age of customers between regions? If so, can we predict the age of a customer in a region based on other demographic data? ↓
3. We need to investigate Martin's hypothesis: Is there any correlation between age of a customer and if the transaction was made online or in the store? Do any other factors predict if a customer will buy online or in our stores?
4. Finally, is there a relationship between number of items purchased and amount spent?

Upshot:

The region in which a transaction is made serves as a good indicator for how much will the customer spend. The typical customer in region 2 spends the least and the one from region 3 and 4 spends the most. There are some clear differences regarding age and region: region 4 has the youngest customers aged in average 40 y.o. and region 2 has elderly customers in average over 55 y.o. Region 1 and 3 lay inbetween with similar age distributons around 45 y.o. Region is the best predictor for the customers age, and predicting accurately the age in a given region taking other information in account is in general not possible. Martin's hypothesis does not seem to be supported by the available data: there is no correlation between customers age and their inclination to buy online, exceptuating that the ones aged over 75-80 y.o. make their shopping exclusively online. The principal factor for buying online or in-store is the region, but for some of the regions other useful factors will be the spent amount and the costumer's age too. There is no apparent relationship between the number of items purchased and the amount spent.

Results:

1. Do customers in different regions spend more per transaction? Which regions spend the most/least?

   (see Chart 1.a and Chart 1.b)

The region in which a transaction is made serves as a good indicator for how much will the customer spend.

The typical customer in **region 2 spends the least**, the practical totality of the purchases being **under 500$** with an average around 250$.

The customers from **region 4 spend the most**, almost all of them spending between 200$ and 2700$ around **an average of 1300$**. One of every two customers spend between 700$ and 1800$.

Region 3 presents many similarities with region 4. There almost all its customers spend between 100$ and 2600$, with an average of around 900$. Half of the customers spend 400-1100.

In the region 1 the expense is almost always less than 1750$, starting from about 100$. Half of the customers buy for a value of 400$ to 1000$.

2. Are there differences in the age of customers between regions? If so, can we predict the age of a customer in a region based on other demographic data?

   (*see Chart 2.a and Chart 2.b)*

There are some clear differences regarding age and region. **Region 4 has the youngest customers** aged in average 40 years old, the half of them being between 30 and 50 y.o.

Region 1 and 3 have a very similar customers **age distributon**, with an average around 45 years old and half of the customers being between 35 and 55 y.o.

The **region 2 has elderlier customers**, in average about 55 years old, and the half of them being between 40 and 70 y.o.

Region is the best predictor for the customers age, and **predicting accurately the age in a given region taking other information in account** is in general **not possible**.

3. Is there any correlation between age of a customer and if the transaction was made online or in the store? Do any other factors predict if a customer will buy online or in our stores?

Martin's hypothesis does not seem to be supported by the available data: **there is no correlation between customers age and their inclination to buy online**, exceptuating that the ones aged over 78-80 y.o. make their shopping exclusively online *(see Chart 3.a).*

The principal factor for buying online or in-store **is the region**: customers in region 1 will buy mostly in-store (in 80% of the cases) and the ones in region 2 will do it almost all of them online *(see Chart 3.b).*

For some of the regions **other useful factors** will be the spent amount and the costumer's age too. In region 3, if the amount of the purchase is under 1000$, it will be done in the store in 85% of the cases. If the amount is higher it will be done almost certainly online *(see Chart 3.c).*

In region 4, if the amount of the purchase is over 2000$, it will be done almost always online. If the purchase is below this amount, the decision will depend on the age on the customer: 3 out of 5 will prefer the online option if they are over 50 y.o. but if they are younger, it is more probable (in 60% of cases) that they prefere to buy it in-store *(see Chart 3.d).*

4. Is there a relationship between number of items purchased and amount spent?

   (See Chart 4.a)

There is **no apparent relationship** between the number of items purchased and the amount spent.

Chart 1.a ***amount spended by region (Box Plot)***
Coloured box containing second and third quartile, lenght between the whiskers represents approx. 98% of the values. Midbox mark is the median, average is marked by a squared point. [*1.*]

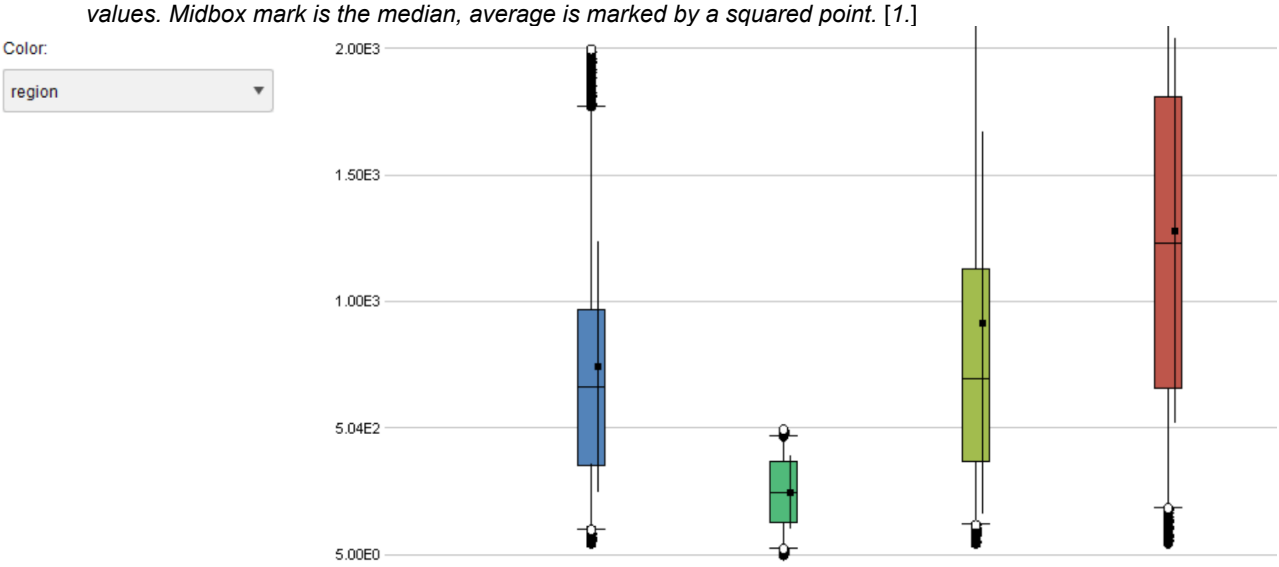

Chart 1.b ***amount spended by region (Distribution)***
Adapted distributions of customer's amount expended by region [*1.*]

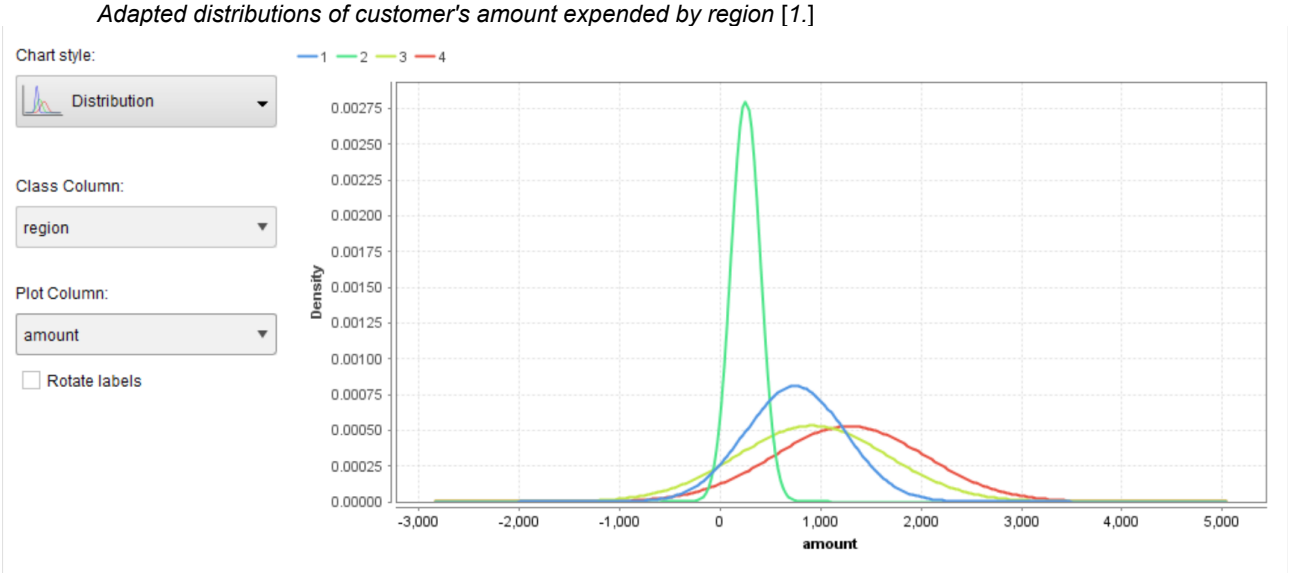

Chart 2.a ***age by region (Box Plot)***
Coloured box containing second and third quartile, lenght between the whiskers represents approx. 98% of the values. Midbox mark is the median, average is marked by a squared point. [2.]

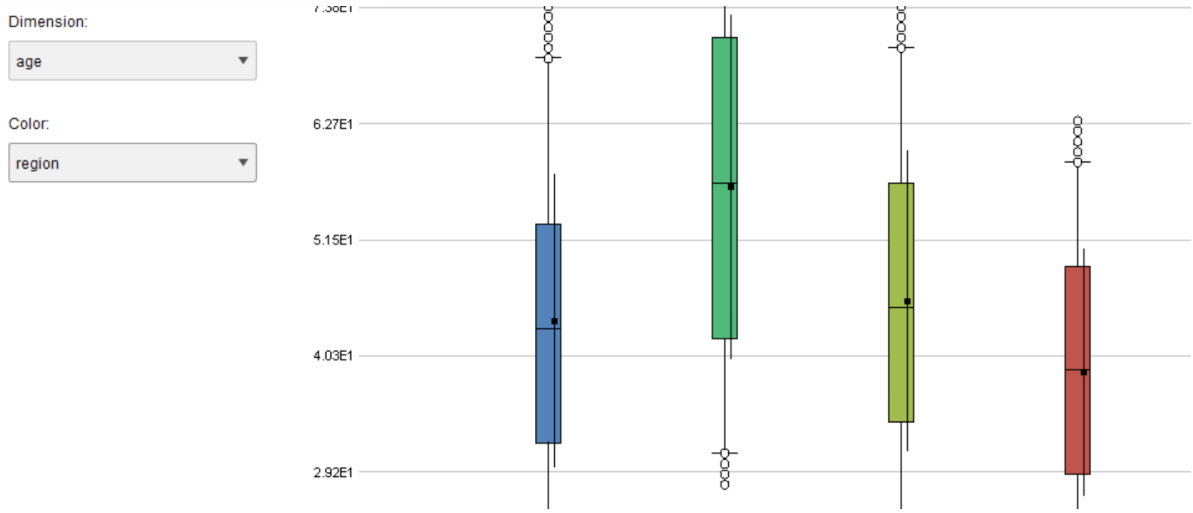

Chart 2.b ***age by region (Distribution)***
Adapted distributions of customer's age by region [2.]

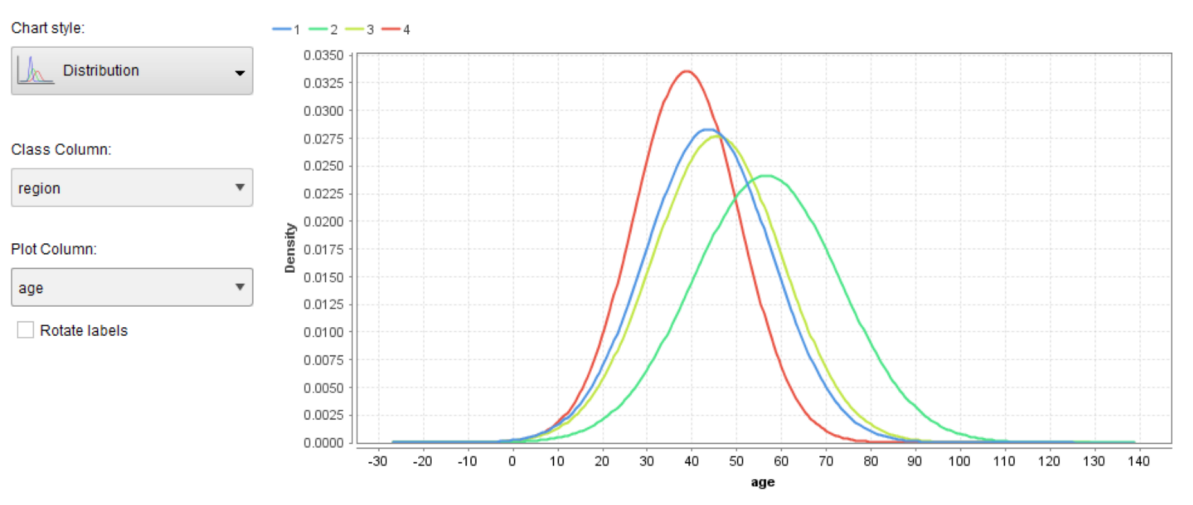

*Chart 3.a **in-store vs age (Scatter Plot)***
There is in general no correlation between buying online (in-store=false) and the age of the customer [3.]

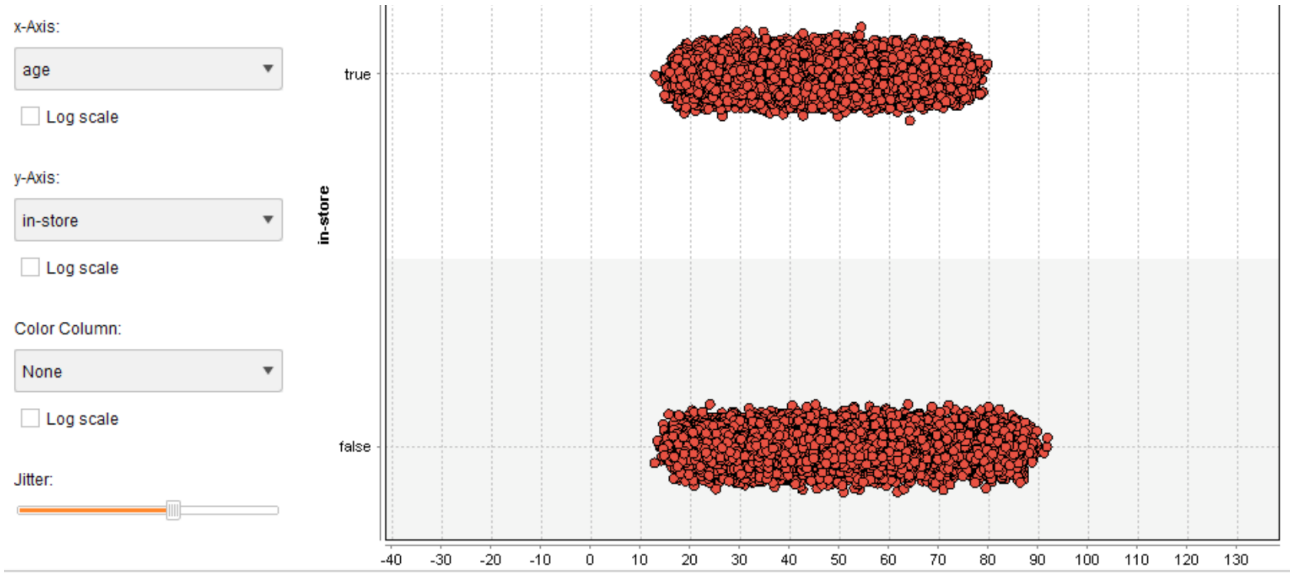

*Chart 3.b **Inclination to buy in-store by region (Decision Tree 2L)***
Probability that the customer is buying in-store (alternative is online) for a given the region [3.]

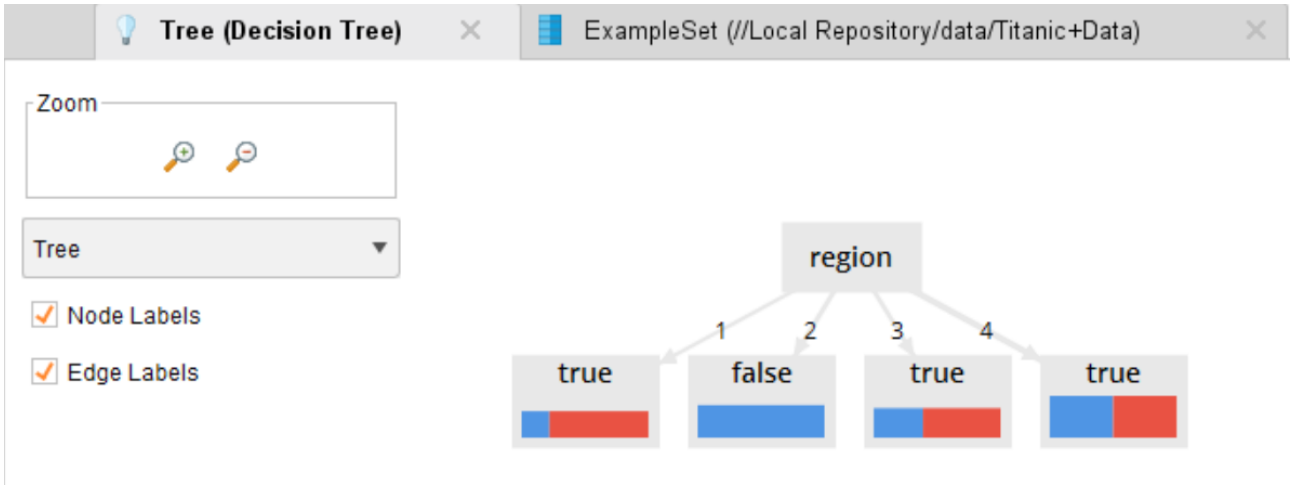

Chart 3.c **Inclination to buy in-store for region 3, amount and age (Decision Tree 5L)**
Probability that the customer is buying in-store (alternative is online) for region 3, given amount and age [3.]

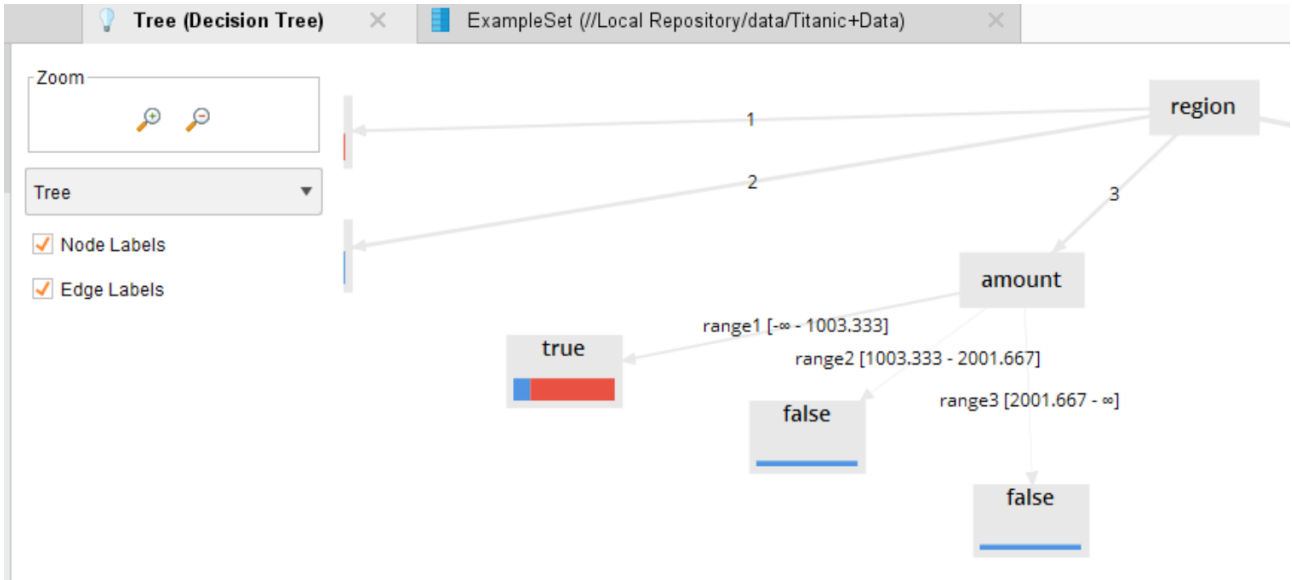

Chart 3.d **Inclination to buy in-store for region 4, amount and age (Decision Tree 5L)**
Probability that the customer is buying in-store (alternative is online) for region 4, given amount and age [3.]

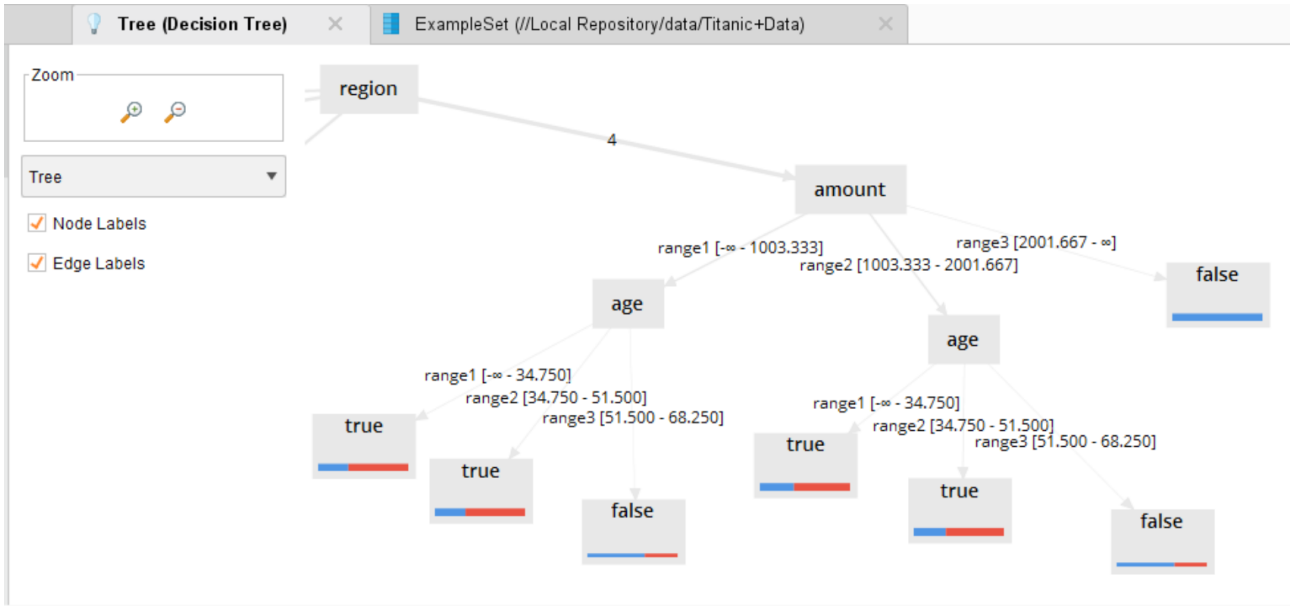

Chart 4.a ***amount vs items (Scatter Plot)***
There is no apparent correlation between the amount spent and number of items bought [4.]