

INTRODUCTION

From the beginning we wanted to work on a resale platform. So we naturally thought of the leader in online second-hand sales: Vinted. However, the second hand clothing market is very complex because we can find for a single product, many resellers offering this item at very different prices, especially because of the different qualities involved. The price dispersion is in fact not too applicable anymore because the products on sale are not equivalent. We then thought that the luxury market could be a good alternative: These products are often much better maintained because of the price that people have spent to obtain them. However we realized that luxury resale platforms did not have different suppliers since they buy products from individuals before restoring and reselling them. We chose to focus on smartphones sold by CDiscount since smartphones have all the characteristics of luxury goods and CDiscount has several suppliers.

The research question is in which ways price dispersion is affected by the quantity of suppliers and the characteristics of the different models. To address the issue of price dispersion we intend to proceed by scrapping offers of different models of smartphones. We will group together the models that have roughly the same characteristics, otherwise it will be hard to correctly analyze price dispersion. Once this action is done, we will be able to analyze the different prices according to the characteristics of the smartphone : brand, size, storage capacity, etc. We will scrap at regular intervals to be able to identify the characteristics of the different suppliers and which product they sell. We might try to scrape data through a VPN from other areas in the world to see if suppliers have other behavior and characteristics.

References

- <https://doi.org/10.1145/2872427.2883089>
An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace
- <https://www.sciencedirect.com/science/article/pii/S1567422320300454>
What determines online transaction price dispersion? Evidence from the largest online platform in China
- <https://www.mdpi.com/0718-1876/16/2/16>

Price Discount and Price Dispersion in Online Market: Do More Firms Still Lead to More Competition?

- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1625847

Online Price Dispersion: An International Comparison

DATA COLLECTION STRATEGY

The data source we will use is the website of CDiscount : <https://www.cdiscount.com/>

The purpose of the data is to present an offer about a product to a consumer: The consumer searches for a given type of goods in the search bar and the website responds with a list of offers corresponding to the types of goods.

We access the data by scraping it, we will use the Selenium package of python to access the data.

We have tried to scrape the data using BeautifulSoup. At first we were blocked by a 403 Error, it seems that the website did not allow scrapers to fetch the data. We went around this issue by changing the user-agent to Google Chrome. However since the website used Javascript we could not use BeautifulSoup so we finally used Selenium and were able to fetch the data.

We looked at the CDiscount Marketplace API but it didn't allow us to fetch the data we needed.

In our data an observation is an offer for a smartphone made by a supplier.

Dimensions of our data:

- Smartphone model
- Smartphone characteristics
- Price
- Supplier
- Supplier characteristics

Some of the data we collect do not have the Smartphone characteristics dimension so we will remove those data before the analysis.

These are the informations we will extract for each offer:

- 1 pid - (string) The smartphone id
- 3 sid - (string) The unique seller id
- 4 price - (float) The smartphone price listed by a seller
- 5 sid_rating - (float) The "star" rating of a seller ranging from 0 to 5
- 6 sid_rating_cnt - (integer) The number of ratings received by a seller.
- 7 sid_nb_sales - (integer) The number of sales made by a seller the 12 last month
- 8 page - (integer) The page number at which a seller was listed
- 9 pid_rating - (float) The "star" rating of a product ranging from 0 to 5
- 10 pid_rating_cnt - (integer) The number of ratings received by a product
- 11 default_sid - (string) The seller selected by CDiscount as the default
- 12 default_price - (float) The price by the default seller

For every recent model of smartphone we will search on the website the given model and scrape all the offers proposed. We will filter the offers by model (ex: offers for iphone 12, offers for iphone 13, etc.). We expect to have around 10 000 data points in the end.

In the end we will have all the offers for a given model we searched on CDiscount, however we will not have all the models in our data set. It is difficult to know what rate of coverage we will have but we will choose a representative list of models.

Describe your first ideas for the analysis

During preprocessing we will get rid of all duplicates in the data, we will remove all the data that do not have the right dimension (all the data that do not have the smartphone characteristics). We will also try to eliminate all the data that is not relevant (not a smartphone, wrong model, ect.). Those steps should not take too much time.

We want to study the impact of the number of suppliers and the characteristics of a model on price dispersion. We also want to study the price dispersion for suppliers with different characteristics. We will explain those variations in price dispersion by the characteristics and the number of suppliers for a given model of smartphone.

It will be interesting to produce a table with for every model of smartphones, the price dispersion, number of suppliers, highest price, lowest price. We can also group models of smartphones by the same characteristics to evaluate the impact of certain brands or types of characteristics on price dispersion.

We could use regression techniques in order to evaluate the relation between the number of suppliers for a given product and price dispersion associated with this product. We could also evaluate the relation between the price of a product and its price dispersion.

A possible result is that price dispersion is higher when the number of suppliers is important (as seen in the references). We found it interesting because the opposite happens in physical businesses : price dispersion is lower when there are more suppliers for a given product. We might find an interesting relation between the time of the day and price dispersion for all products.