



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)**TOPICS**[What are the columns mean?](#)

general

4

[Clarification about features](#)

general

6

[summer-products-and-sales-performance-in-e-commerce-on-wish](#)

general

4

What are the columns mean?

@zarifbin · 3 years ago

Hi @jfreex could you please explain about what that mean following columns:
-merchantratingscount
-meandiscount vs averagediscount: why some of the records are negative? Totally what is different between mean and average?
-meanproductratingscount
-urgencytextrate

Dataset: Computed insight - Success of active sellers.csv

Thank you...

 [Like](#)

@jfreex · 3 years ago · Edited

Hi @zarifbin , thanks for your interest.

I invite you to take a look at the Q&A on the Kaggle's page of this dataset [here](#).

SQL queries in this dataset

Regarding the file [Computed insight - Success of active sellers.csv](#).

This file is actually not a raw data source. It is **computed through a SQL query of**

[summer_products_with_rating_and_performance_2020_08.csv](#). The query that generates the file you mentioned is the one named [Success of Active Sellers](#). If you run the query, you can get the exact same result as the computed file.

Due to an issue on data.world, any description on columns get removed when the SQL query is updated...

So please look at the SQL queries and try to understand things from there.

Regarding the columns you asked for:

More about (Rating, Count, In the last 12 months, Success rate, Urgency rate) ...



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

Due to an issue on data.world, any description on columns get removed when the SQL query is updated...
So please look at the SQL queries and try to understand things from there.

Regarding the columns you asked for:

- MerchantRatingsCount is the total number of ratings received by a merchant (a seller).
- regarding discounts, please have a look at [this discussion](#).
- MeanProductRatingscount just as it sounds. Since the SQL query does an aggregate of products by their sellers, I used some functions to process the aggregates. Here, for each seller, I took the mean of all the product ratings count of products sold by that particular merchant.
- urgencytextrate : the % of appearance of an urgency banner on that seller's products.

Hope I answered your questions.

Regards

[Like](#)

@zarifbin • 3 years ago

Thank you for your additional explanation...

[Like](#)

@jelizabeth • last year

Hello @jfreeex, I was looking for the data for August 2020 for the Wish platform. Is there any chance of accessing it? TIA

[Like](#)[Add a comment...](#)

Markdown is supported [Preview](#)

[Post](#)



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)**TOPICS**[What are the columns mean?](#)

4

[Clarification about features](#)

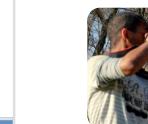
6

[summer-products-and-sales-performance-in-e-commerce-on-wish](#)

4



Clarification about features



@jfreex • 4 years ago • Edited

Hi everyone! Thanks for your interest in this dataset.

I want my datasets to be as clear and straightforward to use as possible. In that regard I'd like to share with you some questions that have been asked about this dataset.

Other questions / remarks / explanations / code examples can be found on [the Kaggle page of this dataset](#).

Some of the questions are related to:

- `has_urgency_banner` and the fact that some some duplicates appear with and without an `urgency_text`

Some people got confused about:

- the `price` and `retail_price` columns.
 - The `retail_price` column should have been named something like "reference price" because `retail_price` is NOT the price at which the product is proposed to clients. The pair `price` and `retail_price` should be used to calculate the discount offered to the client, or how much cheaper is a product compared to the competitors' similar products.

There have been a surge in [submitted notebooks](#) recently and I'm planning to review them and give feedback.

These might in turn prove to be a source of inspiration for your own work.

Hope you're enjoying your time figuring out the secrets to bestseller products... or anything else you take interest in!

Looking forward to reading from you

Jeffrey

Like



Jeffrey Mvutu Mabilama

Summer Products and Sales Performance in E-commerce on Wish

 Comment

783

[Launch workspace](#) |

10

Overview

Discussion

Activity Settings

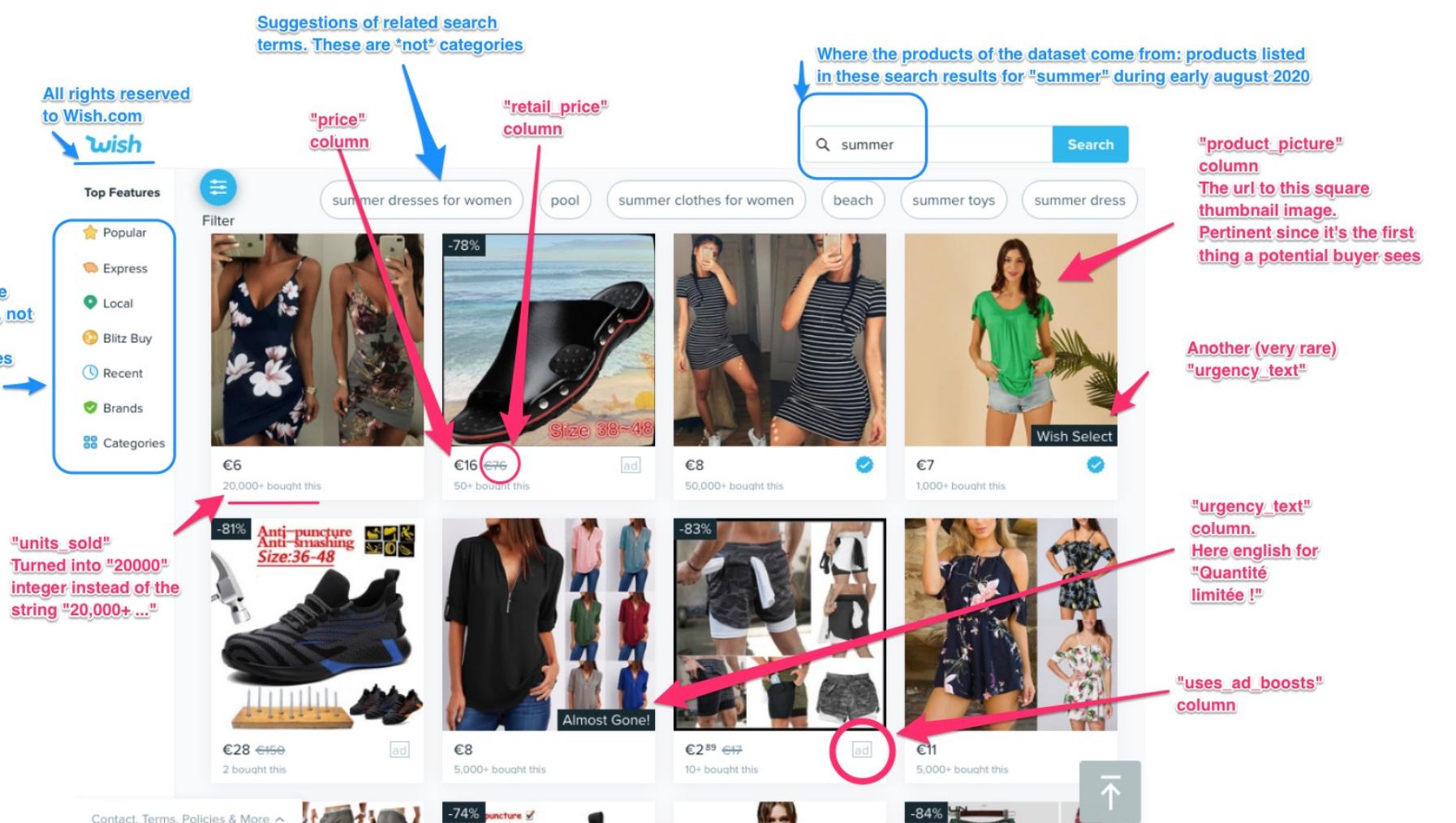
Jeffrey

 Like



@jfreetx • 4 years ago

Here is an annotated screenshot to better explain columns and aspects of the dataset



All in all, we see that the website actually does not have the concept of categories. Rather, it uses tags (keywords/compound keywords) sellers can set. When the keyword is searched for (like "summer"), products related to that tag appear.

 Like



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)[783](#)[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

@yousoufou · 3 years ago

Hello, @jfreex can you please explain from where did you get the following columns (i can't find them in the website) :
Product variation inventory
Countries shipped to
Origin country
Badges

[Like](#)

@jfreex · 3 years ago · Edited

Hi @yousoufou ,
There are simple and more advanced collection methods. Also, Wish is always evolving and fast, so it may not be possible to find some of the columns as of today.

Simple ones:

- Badges could be found on the product badge (they are rare so it's rare to come across them).
- Origin Country is an info about the seller, so found on the seller's page

For other things, you would need to have more advanced or specific knowledge to see "unseen" data. The basics to learn starts with HTML & CSS. Some infos are present but not visible (with CSS display: none or similar things). Then there are knowledge about adaptive websites, which adapt the content based on the user's location/country/language/currency/...

Advanced: to get those with a crawler you need to experiment and have custom configurations. They're mostly for "advanced", since you might need more or less tricks to build a crawler.

- Countries shipped to [advanced]. Just know that based on your location country or language settings, some products pages will be different or not available. If you know about cookies , IP addresses and such, then you would know what I mean. You can get translations this way (for instance by having different cookies with different user settings), or know if a product is not available within a country (if the product page is not showing with a certain IP, or whatnot).
- Product variation inventory [advanced]. It is something I learned more about by creating a Wish seller account.



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)[783](#)[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

- Product variation inventory [advanced]. It is something I learned more about by creating a Wish seller account. Playing around helps figuring out how to interpret the data on other seller's products.

I started working on the scraping program more than 1.5 year ago. Some specific details are currently out of my mind. Moreover I cannot login in Wish right now with several browsers (Chrome, Safari) so maybe the site is broken :/ Can't browse it to search or experiment.

As a takeaway, you need to use several sources and aggregated them. For other things, experimenting will help. And for other advanced things, you might need tricks and specific knowledge.

E-commerce websites have a lot of constraints when they need to be worldwide: their clients speak different languages, have different currencies, and so on... Knowing their constraints and difficulties would help you to know how to do advanced things.

[Like](#)

@jfreex • 3 years ago • Edited

By the way, were you looking for data revolving around a specific **theme or keyword**?

I have data from other periods in time and for other keywords/search terms, even though I only published a dataset about summer 2020.

If some people are interested, I also have **data for other product categories or other search terms**. I would have some processing to do before publishing them, but I can always add that to my todo-list if there is interest.

Some keywords I collected data about during 2019 or 2020:

- Christmas (collected in winter 2019 and around spring 2020)
- gift (collected during 2019)
- games (2019)
- other keywords ... (collected during 2019 and 2020)

Anyone interested to have data about other themes or something? Please like or comment so that I can know more of your interest.

[Like](#)



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)**TOPICS**[What are the columns mean?](#)

4

[Clarification about features](#)

6

[summer-products-and-sales-performance-in-e-commerce-on-wish](#)

4

[general](#)

@faithfullyryan • 4 years ago

hi may i know what's the difference between original & data? I'm new here. Can I use tableau to build visualization for business performance dashboard? Some of the title names have weird or alphabet in it. Do I just ignore it?

[Like](#)

@jfreex • 4 years ago

Hi Ryan,
thanks for your interest and for reaching out.

Difference between `title` and `title_orig`

what's the difference between original & data?

I think you meant `title` and `title_orig`. If not, please correct me.

The data was collected with *french* as preferred language settings. That's why there are non-ascii latin characters (« é » and « à » and such).

- The `title` column contains what would be displayed to a user with *french* settings. Some products can benefit having a title modified to suit the audiences of another country, so sellers can provide a translation. This column is different to `title_orig` if there is such translated/more targeted title.
- The `title_orig` column represents the title specified by the product's seller as the default language. That's why it is mostly in english.



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)CREATE ORIGIN If there is such a translated/more targeted title.

- The `title_orig` column represents the title specified by the product's seller as the default language. That's why it is mostly in english.

Some sellers want to provide a translation for their product's infos. So when `title` and `title_orig` do not match, it's generally because a translation has been provided for some other languages.

So yeah you can just ignore the `title` column if you are not interested in studying translations. **However**, we could suppose that a product might sell better if it has a translation in the client's preferred language.

So we could craft a boolean feature that compares `title` and `title_orig` and says whether a translation was provided. Maybe it would help validate/invalidate the hypothesis I mentioned. But it's up to you to ignore that hypothesis or not.

I'll put it in my todo-list to either update the dataset or add a mapping with that new feature.

Infos about other columns

I also provided explanations about each column. Since you're new, you may appreciate looking at them to get a lot more sense about the data. Here is [how to display column descriptions](#).

I tend to describe columns and features as much as possible in my datasets, so you're sure to have a lot of questions answered by reading them.

Using tableau

Since it is a CSV file, you can easily visualise it using Tableau. This [data.world article](#) shows you how to open a dataset in Tableau.

If it proves too difficult, you can also use another app "integration" with the "Open in third party app" button.

Among others, I have personally used the *Google Data Studio* integration to create insights for [this fashion sellers segmentation](#) project. The graph is basic since it was my first time using Google Data Studio, but that shows you a glimpse of what you can do quickly with it.

[Like](#)



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

@faithfullyryan • 4 years ago • Edited

thank u so answering back. I just started the Analytic Module in my school so it is quite overwhelming & confusing. the question states like this.

In this Group-based Assignment (GBA), you are required to explore the Internet to search for another website (i.e., other than Kaggle) that provides publicly available datasets.

You are encouraged to use professionally prepared graphical and tabular illustrations in your answers. Graphs and tables are not included in word counts.

(a) Describe the newly identified website (that hosts data) to someone who is new to data analytics. Identify one (1) dataset that is suitable for data visualization. State the software (e.g., MS Excel or Tableau) that you will use. Use up to 250 words for your answer. (Note: Though not mandatory, you are encouraged to use Tableau as there is a high chance that we may use it for the End-of-Course Assignment).

(b) Identify one (1) potential problem that can be addressed using visualization of the data identified in Part (a). Discuss why data visualization is recommended in this case. Use up to 200 words for your answer. (Note: The identified problem here can either be a business-related or non-business-related problem, and both are perfectly acceptable for this assignment.)
(20 marks)

(c) Use up to 200 words to describe the data identified in Part (a). Your description should provide details to explain how the data can be used to address the problem identified in Part (b). Does the dataset fit the definition of big data? Why or why not? (Note: For this assignment, it does not matter whether the dataset chosen in Part (a) is "big" or "small"; but you are expected to understand the challenges and opportunities that come with analyzing big data.)
(20 marks)

(d) Perform any necessary preparation of the data, and then produce one or more graphical charts that give useful information in your identified context. Explain how the results address the problem mentioned in Part (b).

I'm trying to an easy business dataset so that i can answer the question easily. i am not sure whether I can use the dataset from here to answer the question. Also, thanks for your help in advance because my due date is coming & trying to rush for this!

[Like](#)



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

@jfreex • 4 years ago • Edited

Regarding (b), I generally tend to see business related questions. You can find inspiration from my other datasets. I tend to put an **inspiration** section to help people get ideas they might try to solve using the given dataset.

Regarding products related datasets, this one also has products and you might find inspiration from it: [NewChic](#).

The kind of business-related questions I think people can try to solve using this dataset are:

- what are the top product categories that sell best ?
businesses want to optimise their stock, so they want to stock more of categories that will have more sales. No one would stock 10K units of a product that is only for a small niche, it would be a loss of money
- Do the rating of a seller matter ?
When a price of a product is sufficiently low, some buyers may tend to not care that much about the rating of a seller since the cost is low enough. One may want to try to see if there is a breakpoint regarding the price and that tendency. Or one might want to play around this question.

There are also non-business related questions.

(c) this one is clearly not big data. Historically there were 3 common V's defining big data (*Volume, Variety, Velocity*). More V's have been added since then so you should check with those to understand the challenges that come with big data.

(d) I uploaded a map of most represented tags in this dataset. It helps a little to show you more of what you can do.
Preparing the dataset will depend on the orientation you choose in (b). Depending on what you want to visualise.

Other thoughts

In order to get new ideas you may start to do some visualisations with simple questions. This way I believe it could help you get a better grasp of the dataset.

For instance, what columns would you use to :

- display the discount policies sellers do on the wish platforms ? (hint: `price` (what the user will pay) and `retail_price` (a comparison price for the same type of products from competitors, or the price before a discount the seller offers))
- does the units sold correlate with the discount offered ? (This might require that you prepare the data by adding a computed



Jeffrey Mvutu Mabilama >

Summer Products and Sales Performance in E-commerce on Wish

[Comment](#)

783

[Launch workspace](#)[Overview](#)[Discussion](#)[Activity](#)[Settings](#)

In order to get new ideas you may start to do some visualisations with simple questions. This way I believe it could help you get a better grasp of the dataset.

For instance, what columns would you use to :

- display the discount policies sellers do on the wish platforms ? (hint: `price` (what the user will pay) and `retail_price` (a comparison price for the same type of products from competitors, or the price before a discount the seller offers))
- does the units sold correlate with the discount offered ? (This might require that you prepare the data by adding a computed `discount` column.)
- does using ad boost propel the sales a lot ?
- does the shipping price have an impact on the number of sales ? If yes, on which conditions ?
(You know that people prefer free shipping, and they may be repelled if the shipping fees are disproportional or seem non legit).
- does the ratings of a product (good/bad/neutral) affect the number of sales ? [I don't recommend tackling this one if you're new to data analysis and short on time]

All in all, there are a lot of things you can do to get a better grasp of the dataset if you are new to data analysis.

If you are already confortable with it, then you may know what to pick in what i mentioned above.

I hope it helps.

Best regards,

Jeffrey

Like



Add a comment...



Markdown is supported [Preview](#)

Post