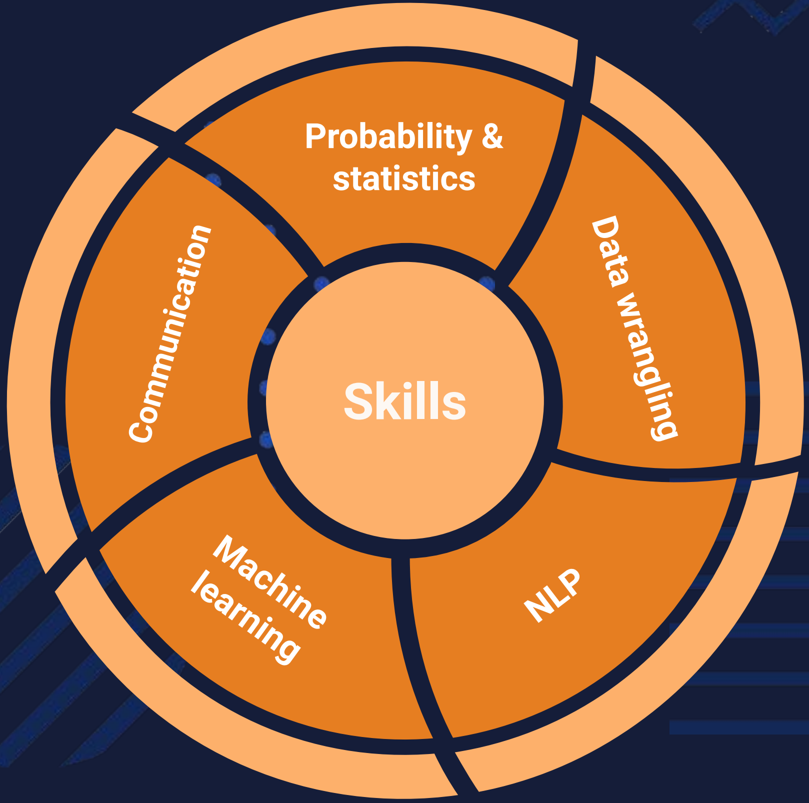
Search Data Scientist Application Case Study

bunnystudio.com

Purpose

The purpose of this case study is to assess the skills of the applicants for the **Search Data Scientist** position.



Context

Bunny Studio is a two-sided creative service marketplace. This means that customers book Audio, Video, Article writing and Image design projects. We refer to these types of services as *Categories*. We connect customers with the best freelancer to do the job, and we refer to these freelancers as *Pros*.

Customers come with an intent to buy from one of the Categories that we offer. The intent of the users is manifested through search queries performed in the search bars across the website. Other data can be used to further understand the customer intent, such as the language of the site, the placement of the search bar, previous search terms, and their shopping behavior in the past.

Objective

Customers input search terms and are presented with the most relevant Samples, according to their perceived intent. Samples are a representation of the Pro’s work in their respective Categories. Users assess the quality and fit of the Sample according to their need, and decide to book a Pro. Users can further narrow their search through the use of filters.

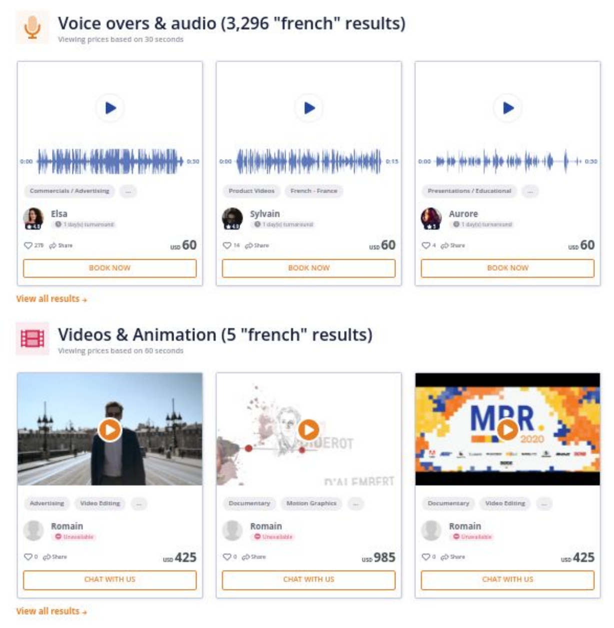
The purpose of this case study is to identify customer intent and present the most relevant results, based on a small sample of our data. You will be given 20 queries and we expect the results from these queries.

Samples

Pros upload samples to the platform so that customers can assess their quality and fit. Samples belong to one Category. Samples can have several tags to facilitate their search.

Once the samples are uploaded, the QC team checks their quality and they can reject or approve the samples. Furthermore, the time to send samples can expire and the number of samples that expire is measured.

Customers can add Samples to their favorite list.



Voice overs & audio (3,296 "french" results)

Viewing prices based on 30 seconds

Commercials / Advertising ... Elsa 4.8 1 day(s) turnaround 279 Share USD 60

Product Videos French - France Sylvain 4.9 1 day(s) turnaround 14 Share USD 60

Presentations / Educational ... Aurore 5 1 day(s) turnaround 4 Share USD 60

Videos & Animation (5 "french" results)

Viewing prices based on 60 seconds

Advertising Video Editing ... Romain Unavailable 0 Share USD 425

Documentary Motion Graphics ... Romain Unavailable 0 Share USD 985

Documentary Video Editing ... Romain Unavailable 0 Share USD 425

Description of fulfillment flow

Users decide to book a Pro. The Pro receives the brief of the project and works on it. Once it's finished, the Pro uploads the work and our Quality Control (QC) team checks the quality of the deliverable. After QC approval, the customer gets the result.

Then, the customer can approve, reject or request revisions. In case of revisions, the Pro is informed about the required changes, implements them, and resubmits their work. The approval cycle is repeated until the customer approves or cancels the project.

Finally, the customer can rate the quality of the work.

Search requirements

To optimize conversions, the following conditions must be met:

- Customers use text search terms, which may include synonyms, plurals, verb conjugations, and other natural language expressions.
- Results should be as relevant as possible, considering the search term and the information about the samples.
- Furthermore, besides being relevant and to minimize the risk of rejected projects, the results should be ordered by the stats of the Pros.
- Result samples are returned in batches of 50 Samples, such that they can be properly shown in the results page.
- To ensure availability of options, the result samples in each of the batches should not be from more than one Pro, or in general it should be avoided that a batch of results has many repeated Pros.
- Pros without projects should not be heavily penalized in the sorting, as they are important too, since they might be new users.

Datasets

You are given three datasets that describe The samples, properties of the samples and stats of the pros.

1. **sample_attributes**: Describes the samples and their attributes.

2. **sample_tags**: Samples have multiple attributes, to facilitate search.

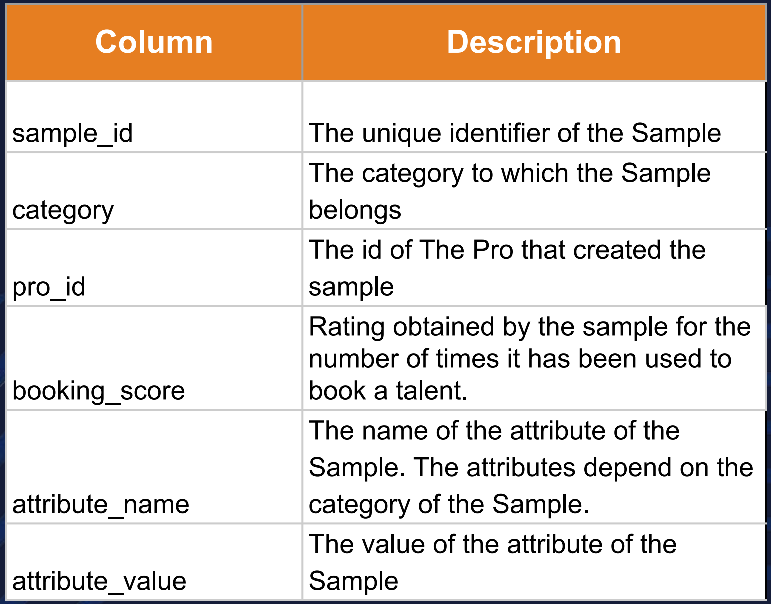
3. **pro_stats**: The stats of the pros, which indicate the projects that they have worked.

sample_attributes

This dataset contains information about the Samples, such as the Pro that created the content, the category to which the Sample belongs, and the attributes of the Sample.

Attributes are specific for each of the Categories. Examples of attributes include:

- Audio: the purpose, language, age and gender of the Pro.
- Image & Video: the purpose, service and language.



Column	Description
sample_id	The unique identifier of the Sample
category	The category to which the Sample belongs
pro_id	The id of The Pro that created the sample
booking_score	Rating obtained by the sample for the number of times it has been used to book a talent.
attribute_name	The name of the attribute of the Sample. The attributes depend on the category of the Sample.
attribute_value	The value of the attribute of the Sample

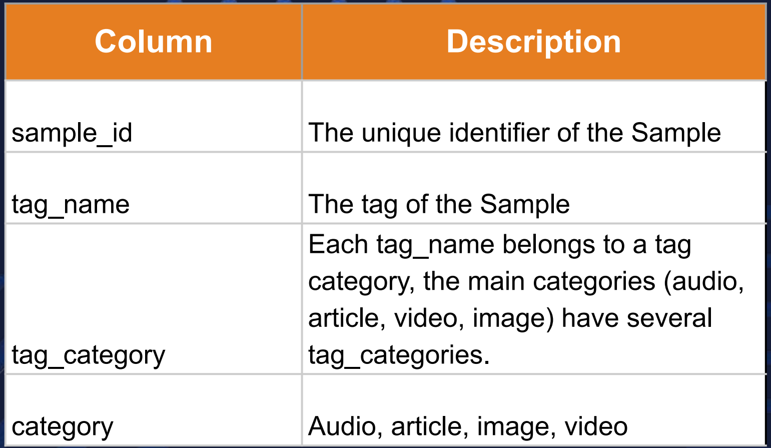
sample_tags

This dataset contains information about the tags of the Samples.

Tags are useful to enrich the data about samples and facilitate search.

Tags are specific for each of the Categories. Examples of tags include:

- Audio: the style, character, impersonation, accent, pitch
- Image & Video: the software, platform and style
- Article: topic, tone, and specialties

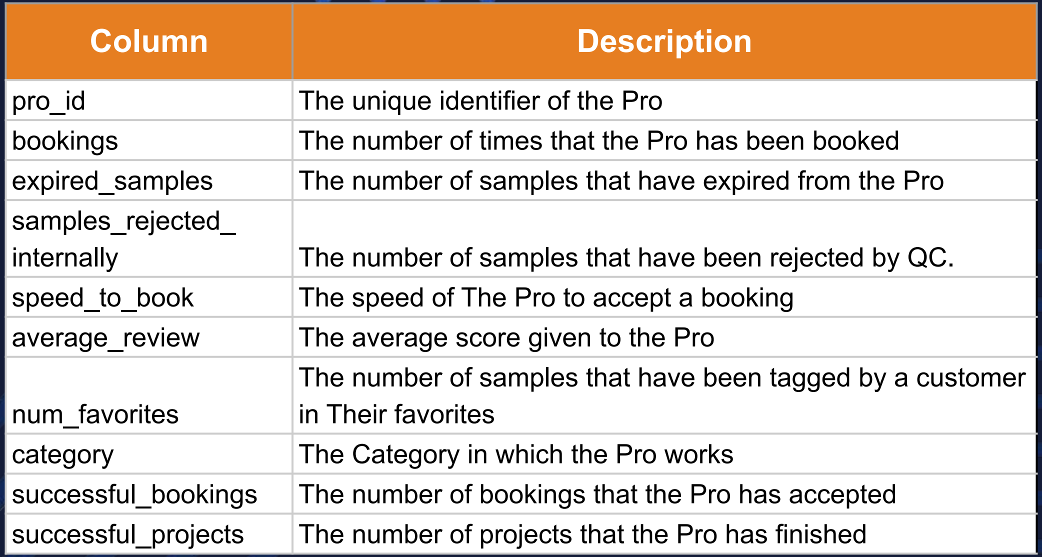


Column	Description
sample_id	The unique identifier of the Sample
tag_name	The tag of the Sample
tag_category	Each tag_name belongs to a tag category, the main categories (audio, article, video, image) have several tag_categories.
category	Audio, article, image, video

pro_stats

This dataset contains The stats of The Pros.

The difference between bookings, successful_bookings, and successful_projects is: a Pro is booked (bookings), a Pro accepts the booking (successful_booking), and the Pro finalizes a project that is accepted by the customer (successful_project).



Column	Description
pro_id	The unique identifier of the Pro
bookings	The number of times that the Pro has been booked
expired_samples	The number of samples that have expired from the Pro
samples_rejected_internally	The number of samples that have been rejected by QC.
speed_to_book	The speed of The Pro to accept a booking
average_review	The average score given to the Pro
num_favorites	The number of samples that have been tagged by a customer in Their favorites
category	The Category in which the Pro works
successful_bookings	The number of bookings that the Pro has accepted
successful_projects	The number of projects that the Pro has finished

Assignment

Produce the results from the 20 search terms found at the end of this document.

Take into account the relevance of the Sample and sort the results using the stats of the Pro. Note that you can create a score of the Pro based on their stats.

Also, make sure to use the search requirements mentioned before.

Search terms

1. Voice over english uk
2. Voice over eng usa
3. Male voice over
4. Female
5. Young adult voice over
6. French blog post
7. Morgan Freeman
8. Child voice for videogame character
9. Elegant man voice for advertising in english
10. Female commercial voice in Spanish
11. Phone system female voice actor
12. Animations 2D characters
13. Icons for Facebook
14. Blog for government
15. Social media post
16. Geology
17. Funny blogpost
18. Ads video
19. Piano music
20. Voice for radio advertisement

Tips and tricks

- You have one week to finish the case. Try to around 8 hours on this assignment.
- Take into account that search terms use natural language and contain synonyms, spelling mistakes, plurals, etc.
- Since customers want to buy from one specific category, you could try to understand which category they want and then present results for that category. Alternatively, you can present results for each category separately.
- Clean the data before using it and document the cleaning steps.
- To report the relevance of your results create a metric for this.
- To sort results, create a measure of the Pros reliability based on their stats. Remember that new Pros (those with no projects) are important too.
- We don't expect every single query to have an exact match. It's about how robust your solution can be.
- If something is not clear, please let us know! Send us an email to diego@bunnystudio.com

bunnystudio.com