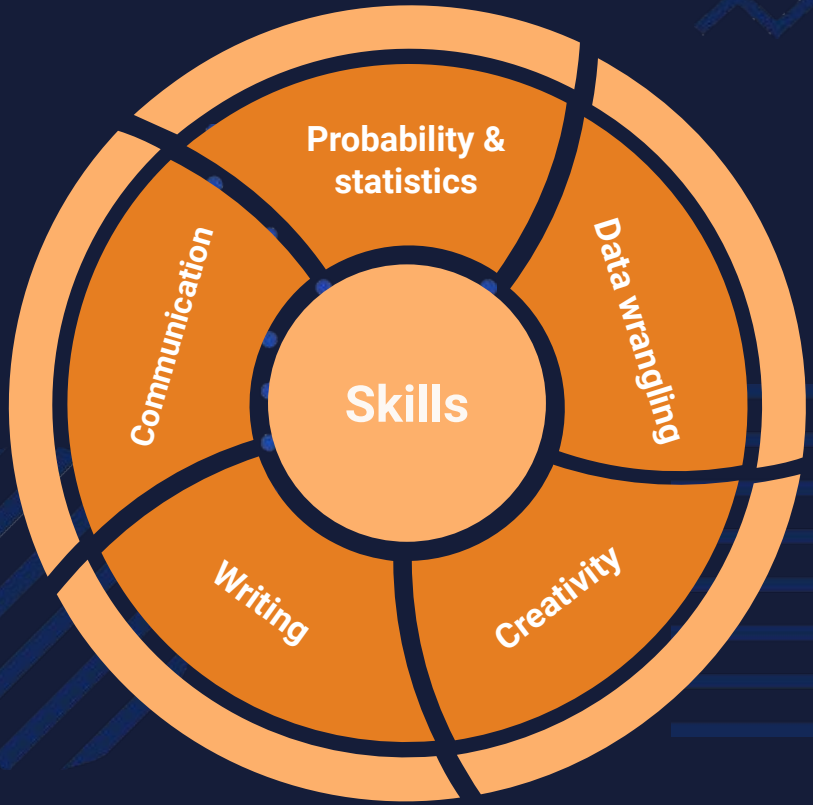
Pricing Data Scientist Application Case Study

Purpose

The purpose of this case study is to assess the skills of the applicants for the **Pricing 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 those broad groups of services as categories. Within those groups, we offer tens of different types of services. For example, within the Audio category, we offer regular audio ads, 3D audio ads, voice-overs, audiobooks, among others. Those specific services are known within Bunny Studio as subcategories.

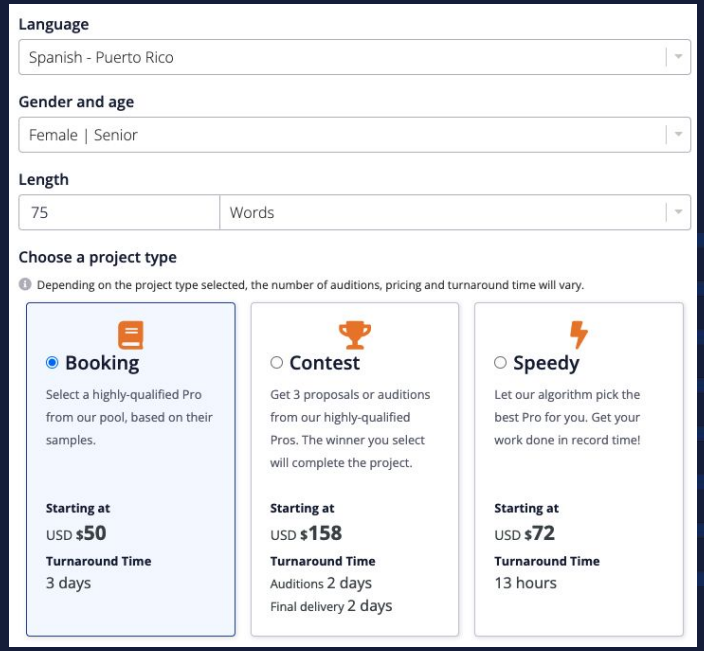
We connect customers with the best creatives to do the job, and we refer to these freelancers as Pros.

Customers come to our website to explore all of our services and, eventually, to buy a project from one of the subcategories that we offer.

About pricing

Regarding pricing, we offer instant and predictable pricing. This means that each time a client comes to the website with the intention of quoting or buying a product, they should be able to find the price by themselves without additional help.

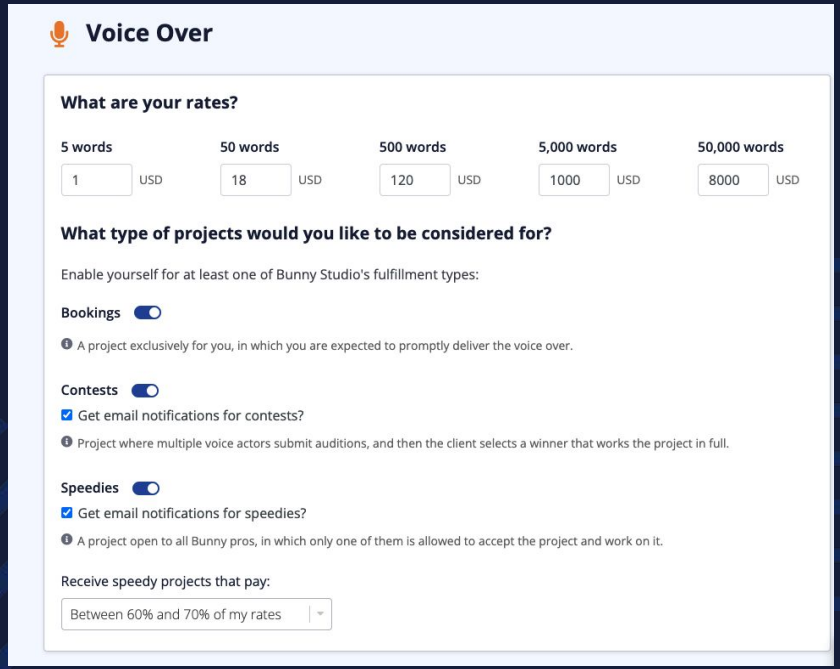
Regarding our cost structure, our main costs are the rates that Pros charge us for completing a project. We refer to the payments to freelancers as “rewards”.



About pricing

Pros register their rates in a UI for different project lengths within the subcategory/subcategories in which they are signed up. Length refers to the number of units of the project, for example the number of words in a voice over.

(See an example for a voice-over Pro in the image in the left)



Objective

The purpose of this case study is to go through the process of pricing a project within Bunny Studio in subcategories, with numerous rows of data and in new subcategories with no data.

Pricing requirements

To guarantee upfront, predictable pricing, the following requirements must be met:

- Model takes a combination of pricing attributes (language, complexity, etc) and project lengths and returns a price for any length > 0.
- Pros receive the reward they expect
- Bunny Studio keeps at least the 15% of the project price
- Fulfillment types also play a role in the pricing, however we will omit that.

Assignment 1

Define an algorithm that allows to price a voice-over project in all of the languages present in the dataset.

Take into account that each language should have differentiated pricing and that you should be able to have a price for any project between 1 and 50,000 words.

Datasets

You will find in the “data” folder a single .csv file.

1. **Pros_rates_voice.csv**: A dataset containing information regarding Pros rates collected from the UI previously presented.

pros_rates_voice.csv

- In this dataset you will find some basic information regarding the rates of the Pros for some voice over lengths measured in number of words.

Column	Description
user_id	The unique identifier of the Pro
subcategory	The subcategory to which the rates belong to
trusted	An identifier of how trusted is the pro. 0=Regular Pros 1=Very trustable Pros 2=Banned Pros
Status_submission	The status of their most recent application test. (labels are self-explanatory)
language_id	An identifier of the language of the Pro
seed_data	A JSON object with the rates. The data is defined in the following units: units (words), time (seconds), cost (rate in local currency of the Pro), costUSD (rate in USD)

Expected deliverable

A GitHub repo or a zip file with the .ipynb file and an .html export. The notebook should contain the following:

- Any data cleaning, data wrangling and EDA you consider necessary
- All the steps required for your solution
- A table with the suggested prices for projects of 30, 75, 150 and 500 words in all of the languages present in the dataset.
- In max. 300 words explain what you did in plain English. Assume that this will be read by a non-technical stakeholder.

Assignment 2

Define the pricing model for a new subcategory

Now let's assume that the Head of Product decides to release a new "love letters" subcategory. Because of the novelty, we lack of any sort of data within our databases regarding love letter writers.

How would you approach the lack of data?
What other pricing approaches would you use to define a launching price?

Report in max. 300 words how would you proceed in non-technical language. **Please report this in the same notebook from the previous assignment.**

Tips and tricks

- You have one week to finish the case. 4 to 8 hours of work should be enough.
- Clean the data before using it and document the cleaning steps.
- Consider the effect that atypical data points could have in your analysis.
- Consider the effect that not having many observations could have on your results.
- We know that this is a very open-ended case. Keep in mind that there are no right or wrong questions and we are trying to understand the way you approach to a complex problem.
- If something is not clear, please let us know! Send us an email to diego@bunnystudio.com

bunnystudio.com