

Mamma-mia ! A pizza visualization, the process book

Our path :

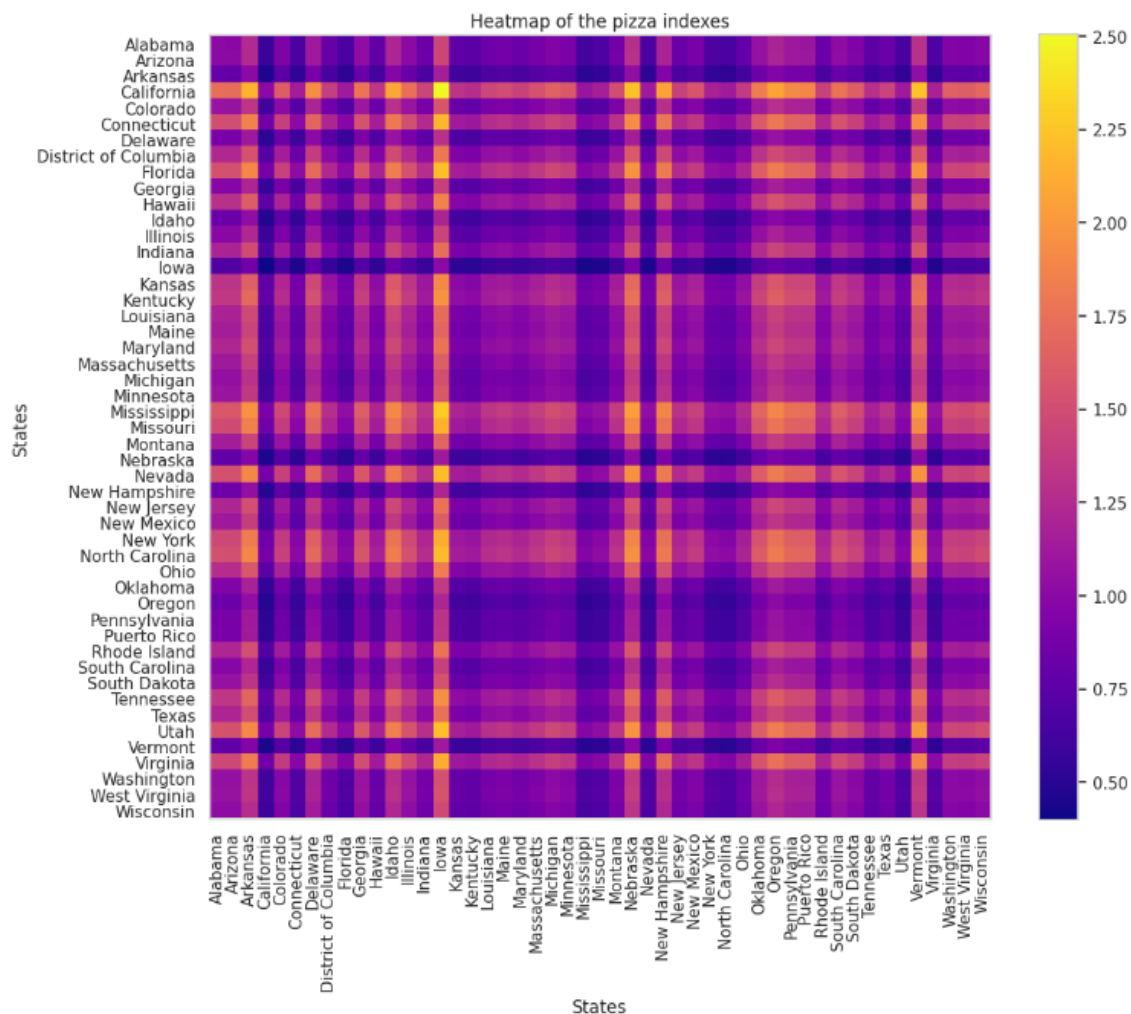
Once we knew that pizza would be our main topic, we knew we wanted to present information in a beautiful and concise manner. We were heavily impressed by the work of a group from last year, **the wine project**, and we liked their website format even more, a simple slide scrolling website that felt fresh and clean. This guided us in the creation of this beautiful website separated with sections.

We then went searching for an impactful way of assembling the different ideas we got while brainstorming; this led us to thinking about the experience we wanted to give to the users. We went for a story that would guide the user through **three main** parts : **At first** we want them to explore the data, discover new pizzas they have never heard of, and play with the interactive map; **then** we wanted to give way to their creative mind, so we built a mini game where they can create their pizza and find a restaurant where their pizza is available. **The final step** is more serious and more mathematical, this time we compare our new pizza index to the living index and let the user puzzle over the possible interpretations.

A first step in this journey to our final result was the acquiring and cleaning of the datasets. From there we performed some modifications and **analysis** on these datasets to use them in the visualization and to draw insights.

Inspired by the very well known and famous Big Mac Index, we have decided to design and develop the most renowned future index addressing social and economic disparities, namely : **the Pizza Index**.

How is it constructed ? It's a ratio of the average prices of one state and of another state : $\text{pizza_index_AB} = \text{mean_price_A} / \text{mean_price_B}$



The heatmap above represents all possible `pizza_index` values between most of the US states (due to the nature of the dataset on the pizza restaurants we only have 49 states).

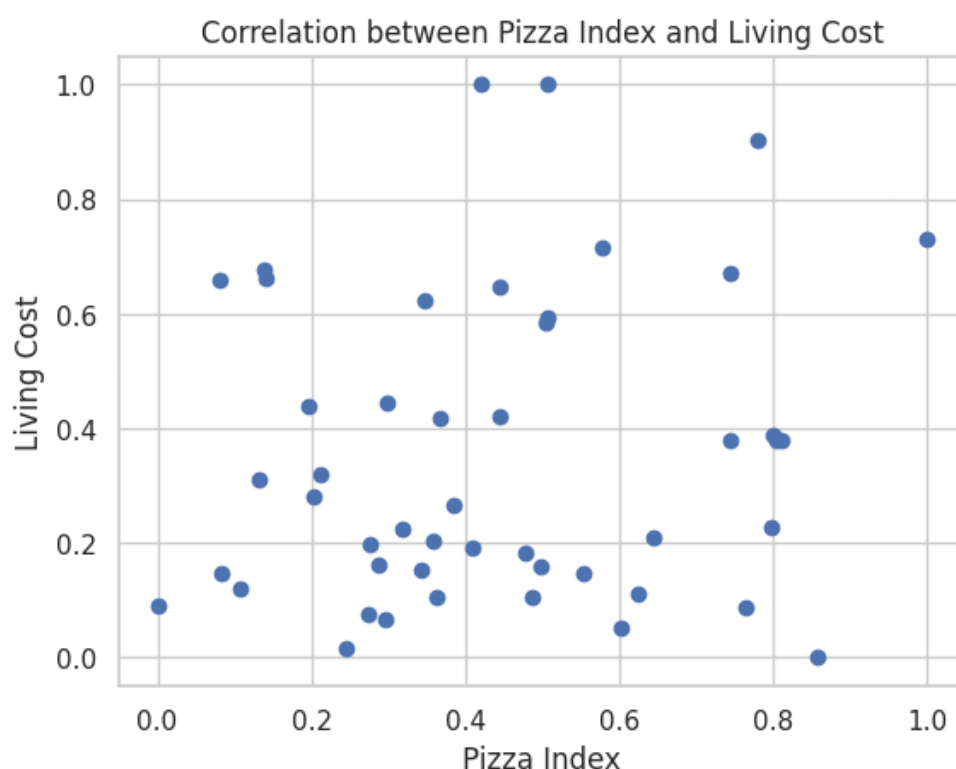
As we can see, we can already spot the states where pizzas are more **expensive** such as California or Mississippi.

In order to generalize the pizza index to all the US states and not only have a value between two states, we had to choose the reference state. As the other datasets covering social and economical indexes were centered around New York, we decided to do the same.

Now we wanted to compare our Pizza Index to the different economic and social indexes we found in a Kaggle Dataset.

After scaling our Pizza index and comparing it to the different indexes, namely the Living cost index, the Grocery index, the Housing index, the Utilities index, the Transportation index and the Health index, we found, using the mean_squared_error as a metric, that the **Living cost index** was the closest. In the third part of our visualization, we display these indexes by states using a color scale.

But by going deeper in the analysis between the Pizza index and the Living cost index we found that they are **only slightly positively correlated** (as the 0.137 and 0.124 correlation values of the Pearson and Spearman correlation suggests) and that It wasn't statistically significant as we obtained p-values greater than the usual threshold of 0.05.



Then came the implementation of the website, there we applied what we saw in the course, and we mostly used *d3.js*, *leaflet* and a lot of javascript for the content of the website. We also used *interact.js* and *fullpage.js*,

respectively for building the drag and drop minigame and for building the website skeleton of slides. When everything was implemented, our website was already impressive but it clearly lacked text that would create the **story** and **act as a glue** between the three parts and within each part.

With the addition of text our website felt much more like a website which feels like a website you know (HARRY STYLES) – and what was left to do was the yassification of the website : this means playing a lot with the css file ! We will talk more on this in the Design chapter ! ;)

The challenges we encountered on the way :

Concerning the drag and drop part, we constructed a dataset that contains the pizza's information and its list of ingredients. Approximately 43.5% of the pizzas in our dataset had the ingredient information. But in order to execute a query and filter the pizza places that contain the given list of ingredients, we had to clean and normalize the list in the dataframe. For that we designed the following process :

- lemmatization
- removing stop words
- removing trailing spaces
- removing punctuation
- separation to keep simple words
- put them in lower case

Note that after cutting the words on the spaces (for example ['red onion'] becomes ['red', 'onion']), we were left with some ingredients that weren't food items. We then manually created a list of these words in order to remove them :

- ['.', 'fresh', 'top', 'red', 'green', 'black', 'roasted', 'grill', 'italian', 'slice', 'topping', 'extra', 'blend', 'white', '(', ')', 'homemade', 'choice', 'canadian', 'serve', 'smoke', 'hot', 'chop', 'roma', 'include', 'bell', 'traditional', 'house', 'bake', 'classic', 'drizzle', 'large', 'create', '2', 'dressing', 'cover', 'style', 'special', 'sauteed', 'ranch', 'specialty', 'sliced', 'shred', 'hand']

From this, we were able to obtain the most used ingredients in American pizzas and chose to draw the food items ourselves according to that.

On a technical side, we encountered a little challenge because the local testing was unavailable the minute that we fetched the data in .json format because of CORS policies. This forced us to test some parts of the website by deploying the full website on github, which meant latency in debugging.

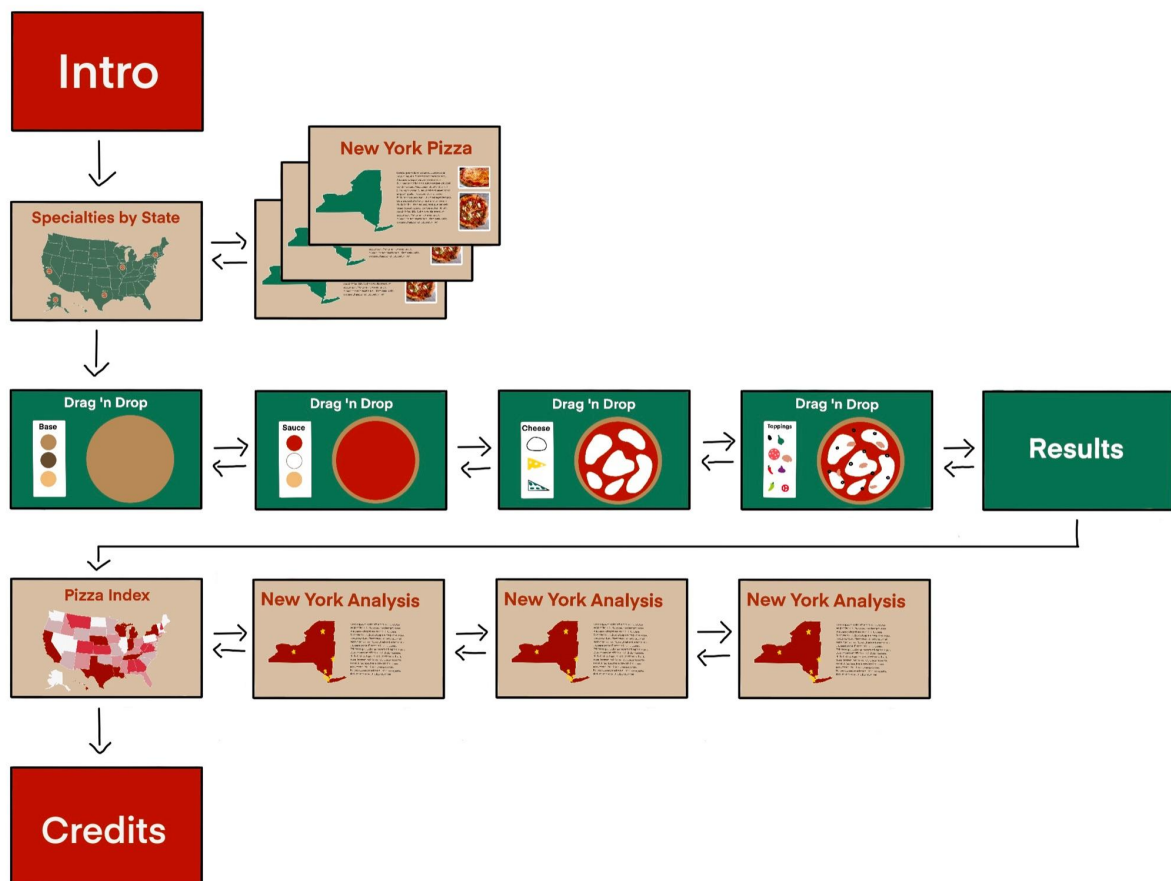
Sketches / Change of plans :

From our original ideas, we simplified the drag and drop a little bit, because creating different visuals for each combination felt a bit overkill, so the ingredients are simply displayed on top of the pizza. A main change to this part was that rather than simply showing the results and their price, we display the pizza restaurants' locations where you can find pizzas matching your choices on an interactive map.

We also wanted to showcase the dataset to the user so we added the possibility to click on the states in the map of the specialities to display more information on them such as the number of restaurants per city in the state and we added a color scale that displays which states have the most pizza restaurants.

Additionally, we preferred to show the pizza index as a global index in all the USA as opposed to in each state as originally planned, to give more meaning and to compare it with the living cost of each state.

For comparison, here is the first draft of our design :



And here is the actual state : [website](#)

Design :

We figured while brainstorming that we wanted our website to have no more than 3 **colors** and we wanted those colors to be reminiscent of pizza so we went for a brownish yellow (for the dough), a deep green (for the basil) and a light red (for the tomato sauce). We of course picked colors that are not too bright so that the website feels beautiful and professional.

As already discussed in the Path chapter, our website design is ruled by this slide layout that is very straightforward and that partitions the different parts very well. This partition is further accentuated by the change of background color when you change between parts.

Another important design choice was the layout on each slide, since this slide skeleton created a constraint for the content that we could put on them. When the moment came to add text we felt we had two options : **either** putting explanations in filler slides between each slide, **or** putting the explanations directly on the slide next to the visualizations, at the risk of putting too much information on the slide. We went for the second option because we found it harder for the user to focus on information if there is no accompanying visualization with it. This made us create a layout where the visualization takes 2/3 of the screen and the explicative text/tips are at the right side.

Peer assessment :

We consider that we were able to split the workload very well and that there was very good teamwork throughout the whole project. We helped each other a lot through all steps of the process.

- Drawings and Maps : Emmanuelle
- Basic analysis of the datasets : Everybody
- Statistical analysis for Pizza index: Etienne
- Specialities : Emmanuelle and Xavier
- Drag and drop functionality: Xavier and Etienne
- Video voice : Emmanuelle
- Process book : Everybody

We invite you to experiment the visualization by yourself on our [github](#) !

Buon appetito ! 🇮🇹 🍕 🍷

