



WEB SCRAPING AND DATA ANALYSIS

**CORA
SUPERMARKETS**

ALDO PIETROMATERA

GABRIELE LORENZO

LEONARDO H. C. SANTOS

YURI DE SENE ALVIZI

CONTENT

- 01** Introduction
- 02** Background
- 03** Data collection strategy
- 04** Data description
- 05** Analysis
- 06** Conclusion

INTRODUCTION

01

Motivation

02

Research question

03

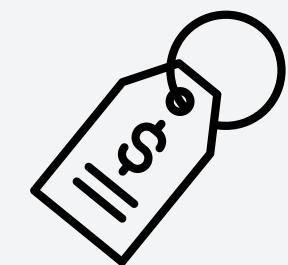
Procedure (collected data and method to analyse it)

MOTIVATION



Consumer Behavior Insights:

- Nutritional quality and pricing of food products
- Valuable information for supermarkets



Market Competition and Pricing Strategy:

- Geographical variation in pricing may provide insights
- Interest to economists and business analysts

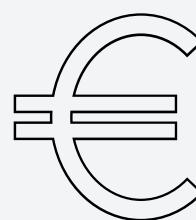
RESEARCH QUESTION



What is the geographical distribution, by region, of:

- the pricing
- the nutritional quality (as measured by Nutri-Score)

for the food products across France.



HOW DID WE PROCEED?

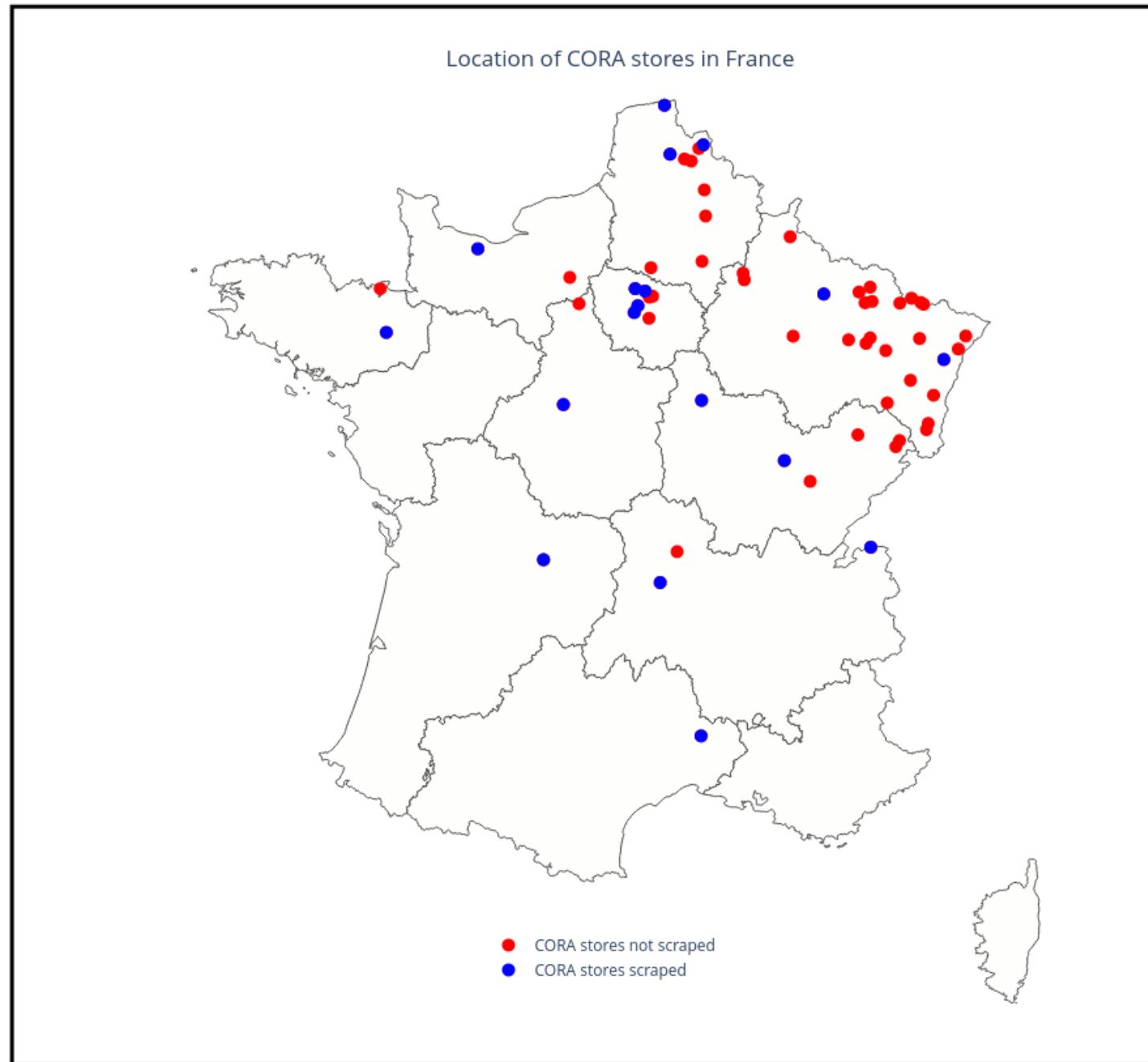


FIGURE 1: Location of CORA stores in France

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

In order to address our research questions, we:

- scraped some selected stores to be as homogeneous as possible
- collected the necessary information
- analyzed the collected information
- compared results with salaries in France
- drew conclusions after comparisons

BACKGROUND

01

Research on the salary/hour per region in France

02

Geographical visualization: median salary/hour per region in France

BACKGROUND

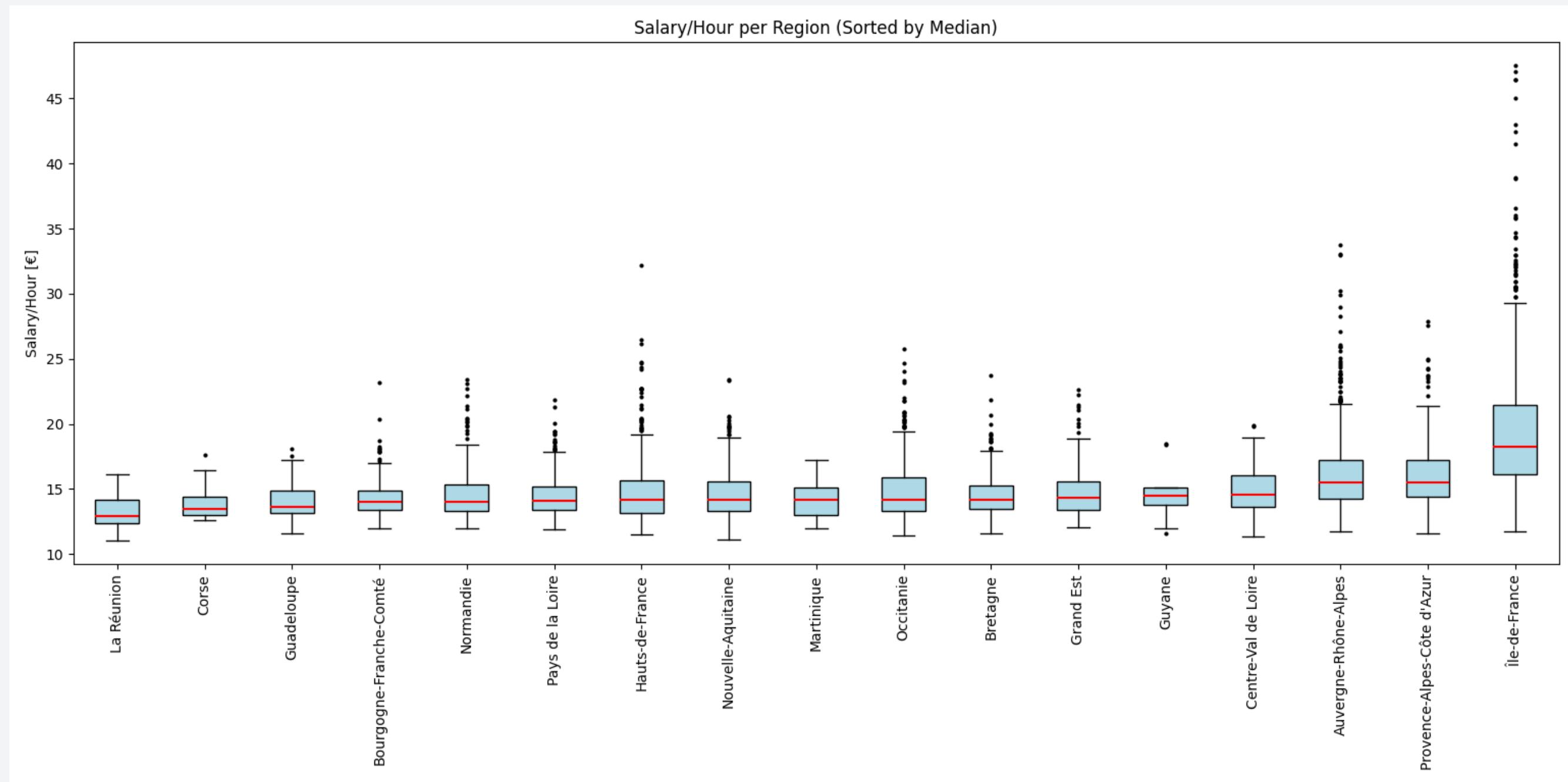
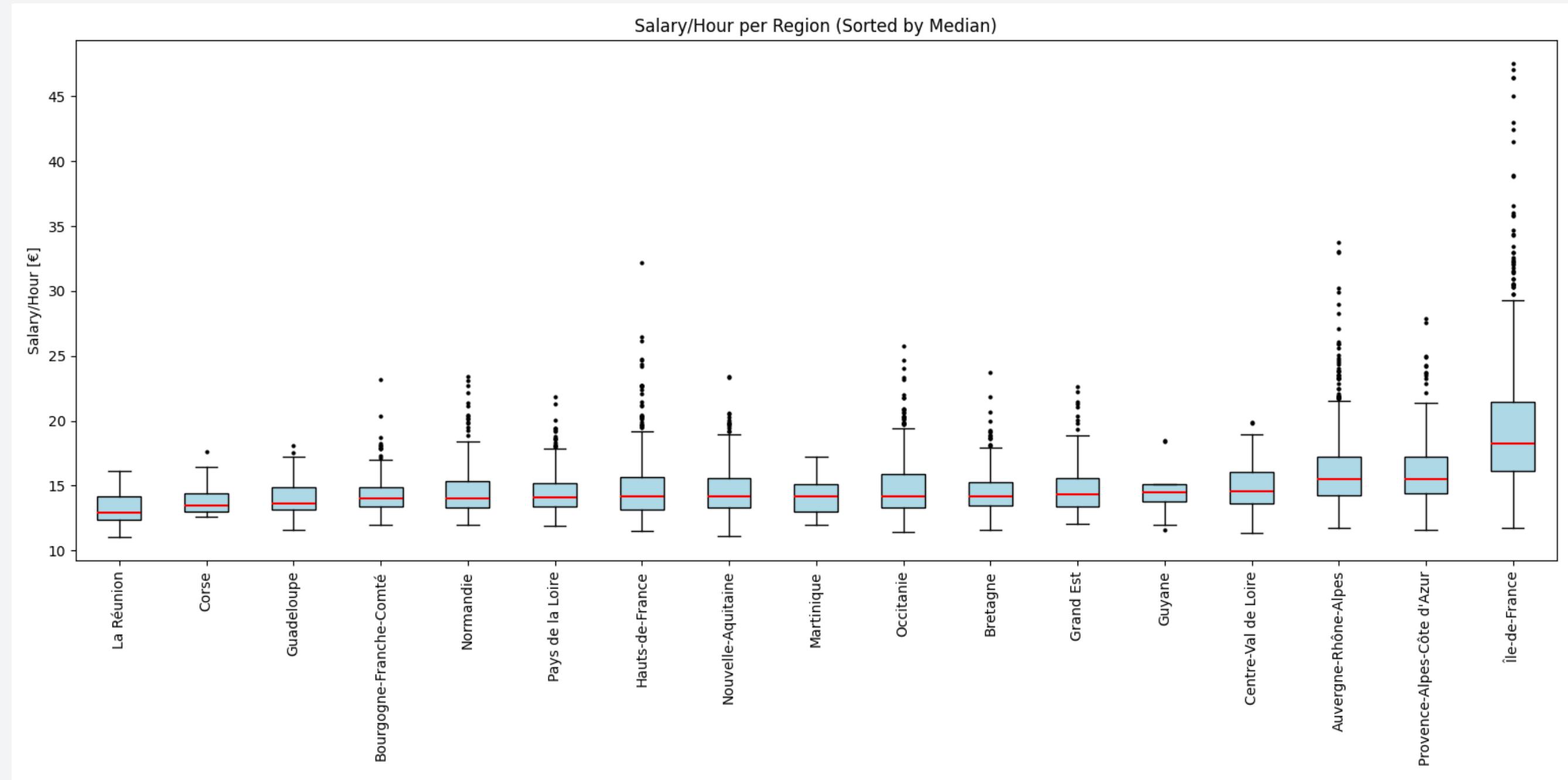


Figure 2: Stability of median hourly salary in all the regions of France, except for Auvergne-Rhône-Alpes, PACA and Île-de-France.

SOURCE: [HTTPS://WWW.INSEE.FR/FR/STATISTIQUES/5359146](https://www.insee.fr/fr/statistiques/5359146)

BACKGROUND



Observing the salary differences identified, we expected a corresponding variation in prices in the supermarkets of the affected regions.

BACKGROUND

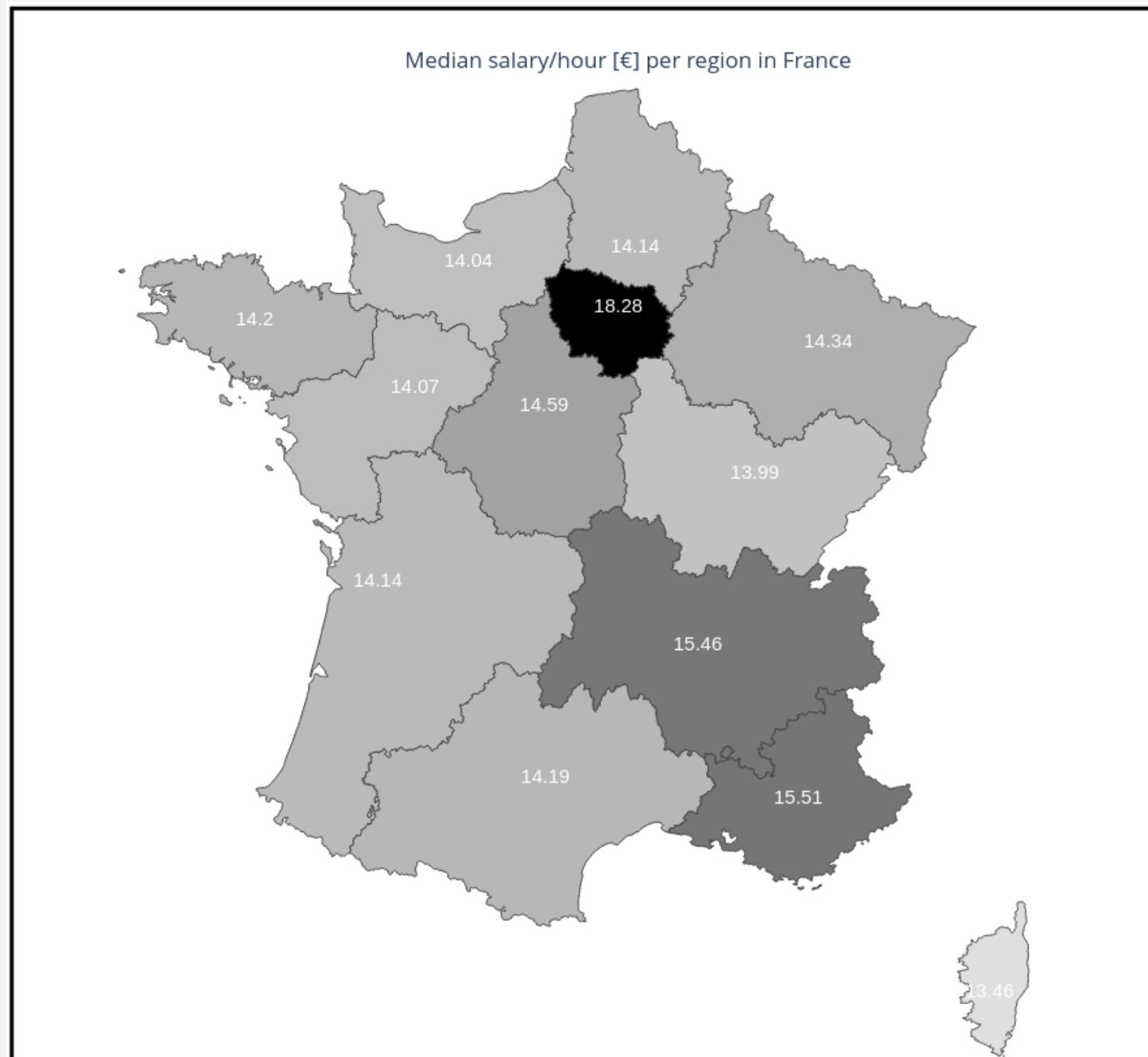


FIGURE 3: Median salary/hour [€] per region in France

SOURCE:

[HTTPS://WWW.INSEE.FR/FR/STATISTIQUES/5359146](https://www.insee.fr/fr/statistiques/5359146)

Observing the salary differences identified, we expected a corresponding variation in prices in the supermarkets of the affected regions.

DATA COLLECTION STRATEGY

How did we collect data ?



BeautifulSoup

DATA COLLECTION STRATEGY

The screenshot shows the Cora website's shopping cart page. At the top, there is a navigation bar with the Cora logo, a search bar, and links for 'Mon magasin Massy', 'Prospectus', 'Mes listes', 'Ma fidélité', 'Mon compte', and a shopping cart icon showing '0,00 €'. Below the navigation, there are tabs for 'Faire mes courses', 'Idées repas', 'Maison et loisirs', 'High-Tech', and 'Mode'. A promotional message encourages users to download the Cora app. The main content area is titled 'Faire mes courses' and features a grid of food categories:

Hypers bons plans et Promos	Les sélections du moment	Alimentation bio et produits écologiques	Régimes alimentaires et nutrition	Fruits et légumes	Viandes, charcuterie et poissons	Traiteur	Produits laitiers, oeufs, fromages et desserts
Pains, viennoiseries et pâtisseries	Surgelés	Epicerie salée	Epicerie sucrée				
Animaux	Le coin des marques						

On the right side of the page, a developer tools window is open, showing the HTML code for the 'Fruits et légumes' category. The code includes a title image with a radius-full border, a label 'Fruits et légumes', and a component ID 'component-2149'.

FIGURE 4: Iterating over categories

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

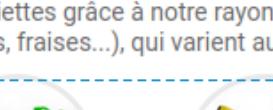
DATA COLLECTION STRATEGY

Accueil > Faire mes courses > Fruits et légumes

Fruits et légumes

Ajoutez de la fraîcheur dans vos assiettes grâce à notre rayon fruits et légumes ! Mettant en avant des producteurs locaux, Cora vous propose toute l'année des fruits et légumes bio (tomates, pommes de terre, pommes, fraises...), qui varient au rythme des saisons. Personnalisez votre panier bio et commandez-le en ligne pour gagner du temps !





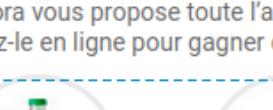
Fruits et légumes bio



Fruits frais



Légumes frais



Fruits et légumes secs



Jus de fruits et légumes frais



Prêt à consommer et fraîche découpe

```
<!-->
 [event]
</picture>
</div>
<div id="component-2897" class="c-title-image_label" data-v-3c66f424="">
  Légumes frais</div>
</a>
</li>
<li class="c-list_item children-categories_list-item" data-v-33b45019="" data-v-652c1147="" data-v-59207ec2="">
  <a class="c-title-image c-children-categories_item c-button c-title-image--radius-full c-title-image--border" data-v-0f661df5="" data-v-3c66f424="" data-v-652c1147="" href="/faire_mes_courses/fruits_et_legumes/fruits_et_legumes_secs-c-177055" tag="button" data-v-33b45019="" aria-labelledby="component-2903" data-v-3c66f424=""> [event] flex
    <div class="c-title-image_image-wrapper" data-v-3c66f424=""> [flex]
      <picture class="c-title-image_image" data-v-4b362a23="" data-v-3c66f424="">
        <!-->
```

FIGURE 5: Iterating over SUB categories

SOURCE: HTTPS://WWW.CORA.FR/

DATA COLLECTION STRATEGY

Accueil > Faire mes courses > Fruits et légumes > Légumes frais

Légumes frais | 153 articles

Découvrez la fraîcheur de nos légumes! Faites le plein de vitamines et de saveurs avec notre large sélection de légumes frais, cultivés avec soin. Régalez-vous avec des légumes qui vous inspirent à cuisiner sainement et à savourer chaque bouchée. Faites confiance à Cora pour des légumes frais et irrésistibles à chaque repas.

Légumes de saison	Tomates et concombres	Pommes de terre	Poireaux et navets	Choux, brocolis et artichauts	Carottes, radis et betteraves
Haricots et petits pois	Oignons, ail, échalotes, herbes aromatiques	Poivrons, courgettes et aubergines	Champignons	Salades, endives, épinards, blettes	Avocats et légumes exotiques

```

    > <a class="c-title-image c-children-categories__item c-button c-title-image--radius-full c-title-image--border" data-v-0f661df5="" data-v-3c66f424="" data-v-652c1147="" href="/faire_mes_courses/oignons_ails_échalotes_herbes_aromatiques-c-177070" tag="button" data-v-33b45019="" aria-labelledby="component-3988" event flex></a>
    </li>
    <li class="c-list__item children-categories__list-item" data-v-33b45019="" data-v-652c1147="" data-v-59207ec2="">
        <a class="c-title-image c-children-categories__item c-button c-title-image--radius-full c-title-image--border" data-v-0f661df5="" data-v-3c66f424="" data-v-652c1147="" href="/faire_mes_courses/fruits_et_légumes/légumes_frais/poivrons_courgettes_et_aubergines-c-177071" tag="button" data-v-33b45019="" aria-labelledby="component-3994" event flex></a>
        <div class="c-title-image__image-wrapper" data-v-3c66f424="" flex>
            <picture class="c-title-image__image" data-v-4b362a23="" data-v-3c66f424="">
                <!-->
                
            </picture>
        </div>
        <div id="component-3994" class="c-title-image__label" data-v-3c66f424=""></div>
    </a>
    </li>
    <li class="c-list__item children-categories__list-item" data-v-33b45019="" data-v-652c1147="" data-v-59207ec2="">
        <a class="c-title-image c-children-categories__item c-button c-title-
    
```

FIGURE 6: Iterating over SUB SUB categories

SOURCE: HTTPS://WWW.CORA.FR/

DATA COLLECTION STRATEGY

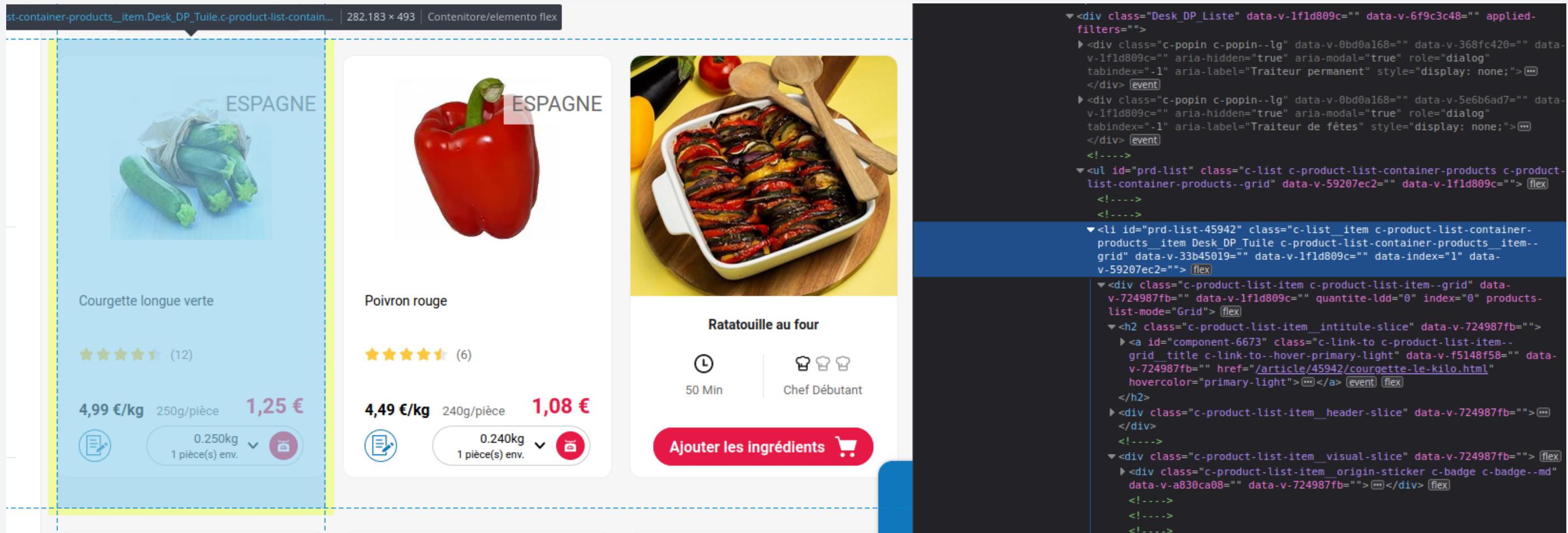


FIGURE 7: Iterating over all the products

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

DATA COLLECTION STRATEGY

The screenshot shows a product page for 'Cora lentilles vertes 500g'. The product image is on the left, followed by the brand name 'CORA' and the product name 'Cora lentilles vertes 500g'. Below the name is a rating of 4,8/5 and a link to 'Lire les 10 avis'. A short description follows: 'Produits cora, le meilleur rapport qualité prix'. To the right is a large price box with '2,30 €' prominently displayed. Below the price are smaller text elements: '4,60€' and 'NUTRI-SCORE A'. At the bottom of the page, there are buttons for 'Ajouter au panier' (Add to cart) and a 'spectus' button. The right side of the image shows the corresponding HTML source code for the price section.

FIGURE 8: Retrieving price for each product

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

The screenshot shows a nutritional information page for the same product. It features a 'Nutri-score' logo and a detailed description of what it represents. Below this is a table titled 'Valeurs Nutritionnelles' with rows for 'Valeur énergétique en kJ' (1449.00 kJ), 'Valeur énergétique en kcal' (343.00 kcal), 'Matières grasses' (1.70 g), and 'Dont acides gras saturés' (0.20 g). The right side of the image shows the corresponding HTML source code for the nutritional table.

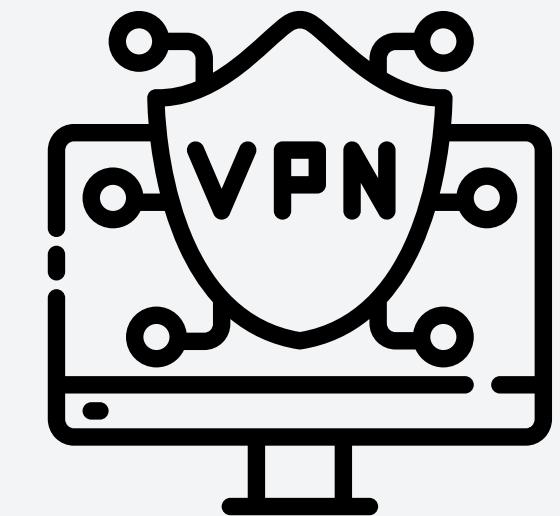
FIGURE 9: Retrieving nutritional values for each product

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

```
</ul>
</div>
<div class="c-nutritional-values c-product-sheet-additional-details__item" data-v-4c71f0df="" data-v-238e6df2="" data-v-78ee5ee4="">
  <h3 class="u-mb-xs u-pt-sm" data-v-4c71f0df="">Valeurs Nutritionnelles</h3>
  <div class="c-nutritional-values__table" data-v-7d841b6d="" data-v-4c71f0df="">
    <div class="c-table_wrapper-outer" data-v-7d841b6d="">
      <div class="c-table_wrapper-inner" data-v-7d841b6d="" responsive="true">
        <table class="c-table c-table--responsive" data-v-7d841b6d="" responsive="true">
          <thead data-v-4c71f0df="" data-v-7d841b6d=""></thead>
          <tbody data-v-4c71f0df="" data-v-7d841b6d="">
            <tr data-v-4c71f0df="" data-v-7d841b6d="">
              <th class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="" scope="row"></th>
              <td class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="">1449.00 kJ</td>
            </tr>
            <tr data-v-4c71f0df="" data-v-7d841b6d="">
              <th class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="" scope="row"></th>
              <td class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="">343.00 kcal</td>
            </tr>
            <tr data-v-4c71f0df="" data-v-7d841b6d="">
              <th class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d=""></th>
              <td class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="">1.70 g</td>
            </tr>
            <tr data-v-4c71f0df="" data-v-7d841b6d="">
              <th class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d=""></th>
              <td class="u-text-center" data-v-4c71f0df="" data-v-7d841b6d="">0.20 g</td>
            </tr>
          </tbody>
        </table>
      </div>
    </div>
  </div>
</div>
```

DATA COLLECTION STRATEGY

- Accessing data by only scraping (No API), using packages and tools presented during the course
- Huge amount of HTTP requests to the website. We got blocked in average once every hour.
→ Solution? VPN and sleep instructions in the code !
- 2 (!!) entire weeks spent scraping
- We tried to cover almost the whole nation scraping at least a store for each region in France
- Preprocessing steps in order to clean data (removed “€” from the price, ...) and use them easily in the analysis step



DATA DESCRIPTION

01

Stores scraped and amount of datapoints collected

02

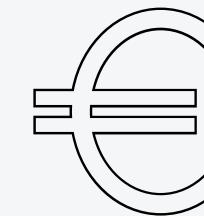
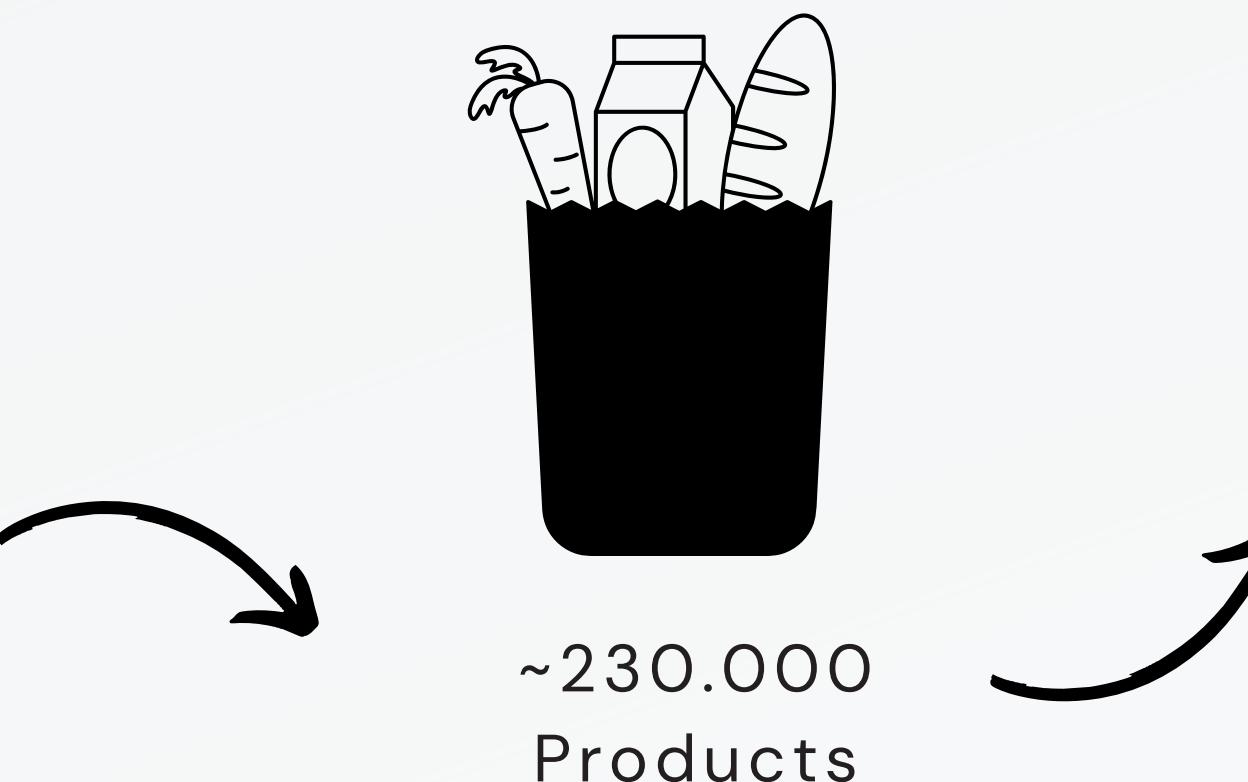
General structure of one datapoint

DATA DESCRIPTION



18 Stores
10 Regions

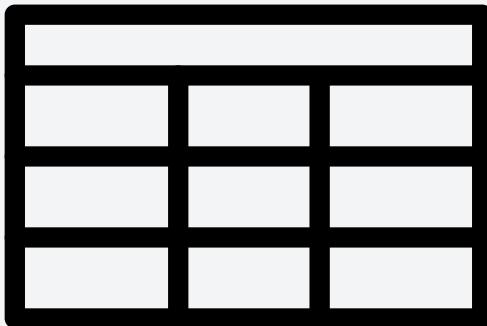
FIGURE 10: CORA stores scraped
SOURCE: [HTTPS://WWW.CORA.FR/](https://WWW.CORA.FR/)



Information retrieved

DATA DESCRIPTION

Our data point has the following features:



- product_id
- category
- subcategory
- sub_sub_category
- product name
- price/unity [€/u]
- price [€]
- nutri-score
- Valeur énergétique [kJ]
- Valeur énergétique [kCal]
- Matières grasses [g]
- Dont acides gras saturés [g]
- Dont acides gras mono-insaturés [g]
- Glucides [g]
- Dont sucres [g]
- Protéines [g]
- Fibres alimentaires [g]
- Sel [g]
- Sodium [mg]
- magasin_id
- name
- adress
- postal code
- city
- latitude
- longitude
- region_name

DATA DESCRIPTION

Region Name	Category	Subcategory	Sub Subcategory	Mean Price [€]	A	B	C	D	E	Nutri-Score Null Percentage
Auvergne-Rhône-Alpes	Epicerie salée	Conserves et plats cuisinés	Conserves de légumes	2.38	0.979	0.021	0	0	0	0.27
Auvergne-Rhône-Alpes	Epicerie salée	Conserves et plats cuisinés	Conserves de poisson	2.71	0.102	0.437	0.354	0.083	0.024	0.48
Auvergne-Rhône-Alpes	Epicerie salée	Conserves et plats cuisinés	Plats cuisinés en conserve et bocaux	3.23	0.324	0.612	0.064	0	0	0.49
Auvergne-Rhône-Alpes	Epicerie salée	Conserves et plats cuisinés	Plats cuisinés micro-ondable et déshydratés	2.07	0.106	0.518	0.2	0.176	0	0.23
Auvergne-Rhône-Alpes	Epicerie salée	Conserves et plats cuisinés	Pâtés, terrines, foie gras et conserves de viande	2.95	0	0	0	0.394	0.606	0.67

TABLE 1: Summarized version of the original dataset

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

DATA DESCRIPTION

REGION	count	mean	std	min	25%	50%	75%	max
Auvergne-Rhône-Alpes	765	16.18	2.94	11.74	14.27	15.46	17.21	33.78
Bourgogne-Franche-Comté	201	14.27	1.51	11.94	13.37	14.00	14.87	23.20
Bretagne	427	14.57	1.56	11.60	13.48	14.20	15.28	23.72
Centre-Val de Loire	231	14.86	1.67	11.31	13.59	14.59	16.03	19.90
Corse	033	13.92	1.20	12.57	13.02	13.46	14.44	17.62
Grand Est	479	14.65	1.71	12.03	13.37	14.34	15.60	22.67
Guadeloupe	028	14.09	1.58	11.61	13.11	13.62	14.87	18.11
Guyane	009	14.78	2.40	11.59	13.77	14.50	15.12	18.47
Hauts-de-France	539	14.83	2.48	11.47	13.17	14.14	15.69	32.22
La Réunion	024	13.16	1.26	11.02	12.40	12.92	14.19	16.09
Martinique	027	14.31	1.54	11.95	12.96	14.17	15.14	17.20
Normandie	288	14.67	2.03	11.98	13.30	14.04	15.37	23.41
Nouvelle-Aquitaine	547	14.59	1.82	11.09	13.29	14.14	15.60	23.38
Occitanie	534	14.86	2.22	11.40	13.31	14.19	15.86	25.76
Pays de la Loire	411	14.44	1.51	11.92	13.39	14.07	15.17	21.86
Provence-Alpes-Côte d'Azur	347	16.11	2.53	11.57	14.40	15.51	17.24	27.88
Île-de-France	566	19.76	5.72	11.74	16.09	18.28	21.45	47.52

TABLE 2: Descriptive statistic of mean hourly salary (refer to Figure2 slide 6 for a short explanation of this table)

SOURCE: [HTTPS://WWW.INSEE.FR/FR/STATISTIQUES/5359146](https://www.insee.fr/fr/statistiques/5359146)

ANALYSIS



PRICE ANALYSIS

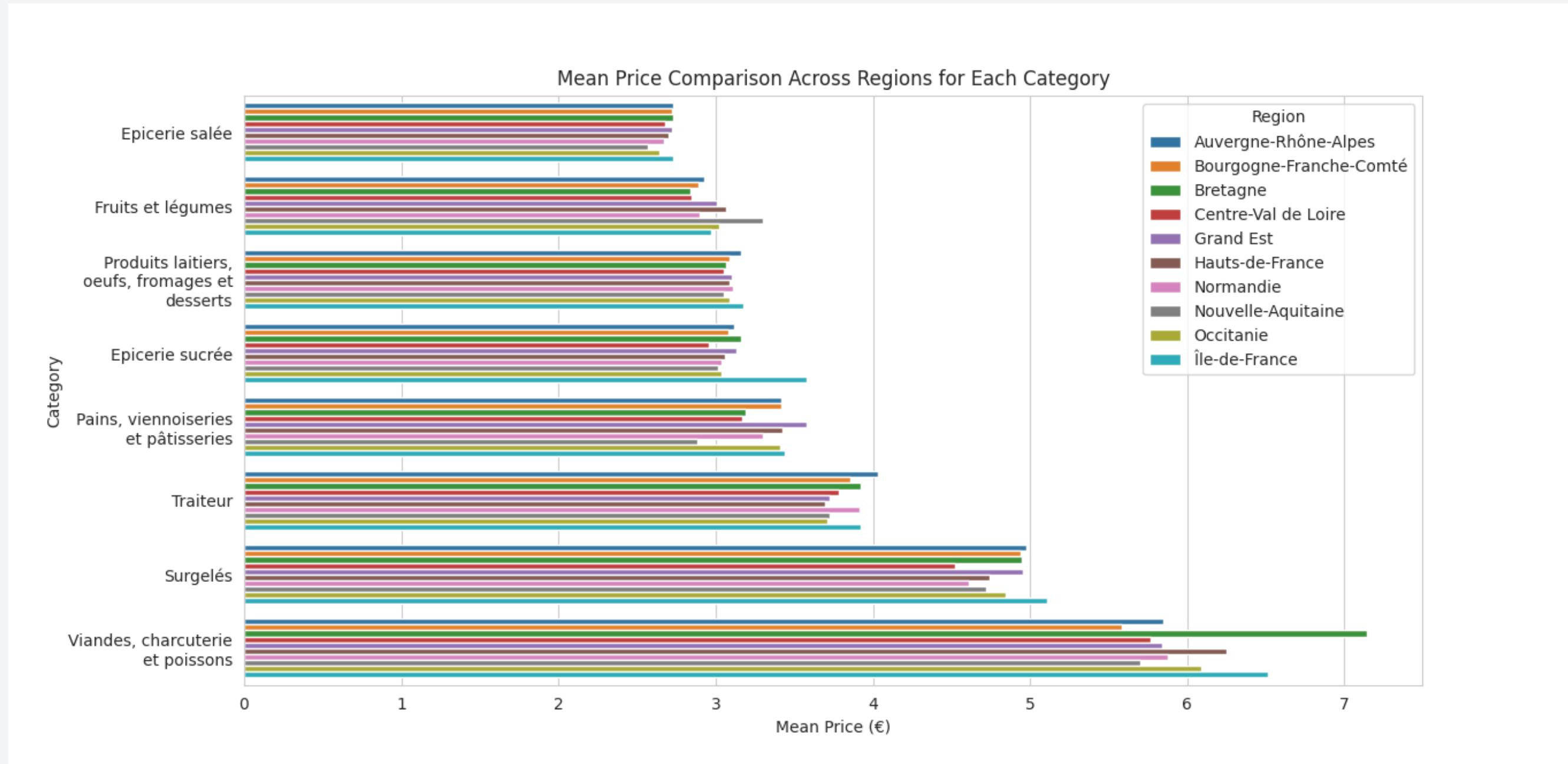


FIGURE 10: Stability of the price by region for each category. The only outlier is the “Viands, charcuterie et poissons” category in Bretagne.

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

PRICE ANALYSIS

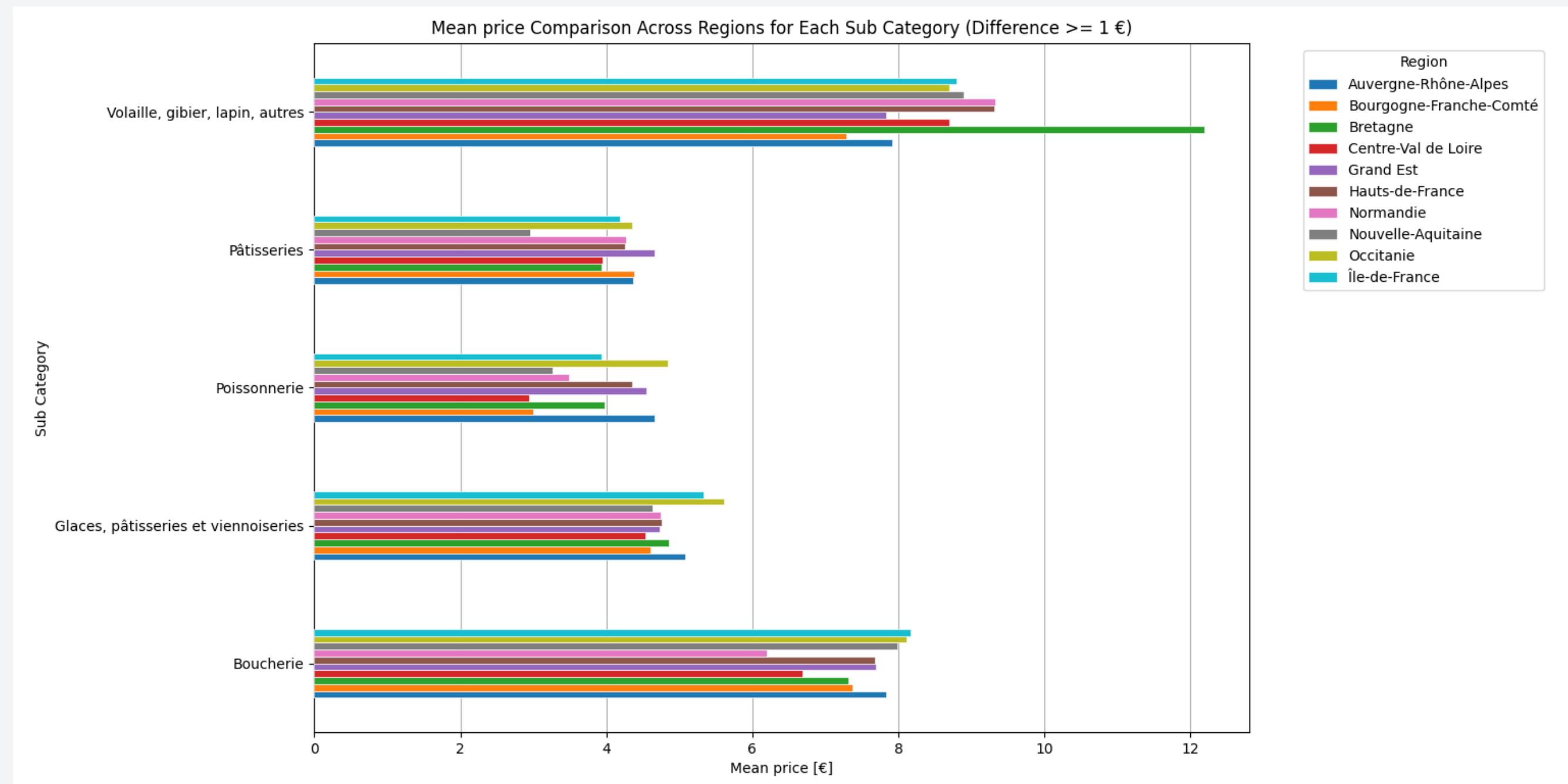


FIGURE 11: Some sub categories where we found a discrepancy of mean price. Despite this, the general trend remains the same.

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

NUTRI-SCORE ANALYSIS

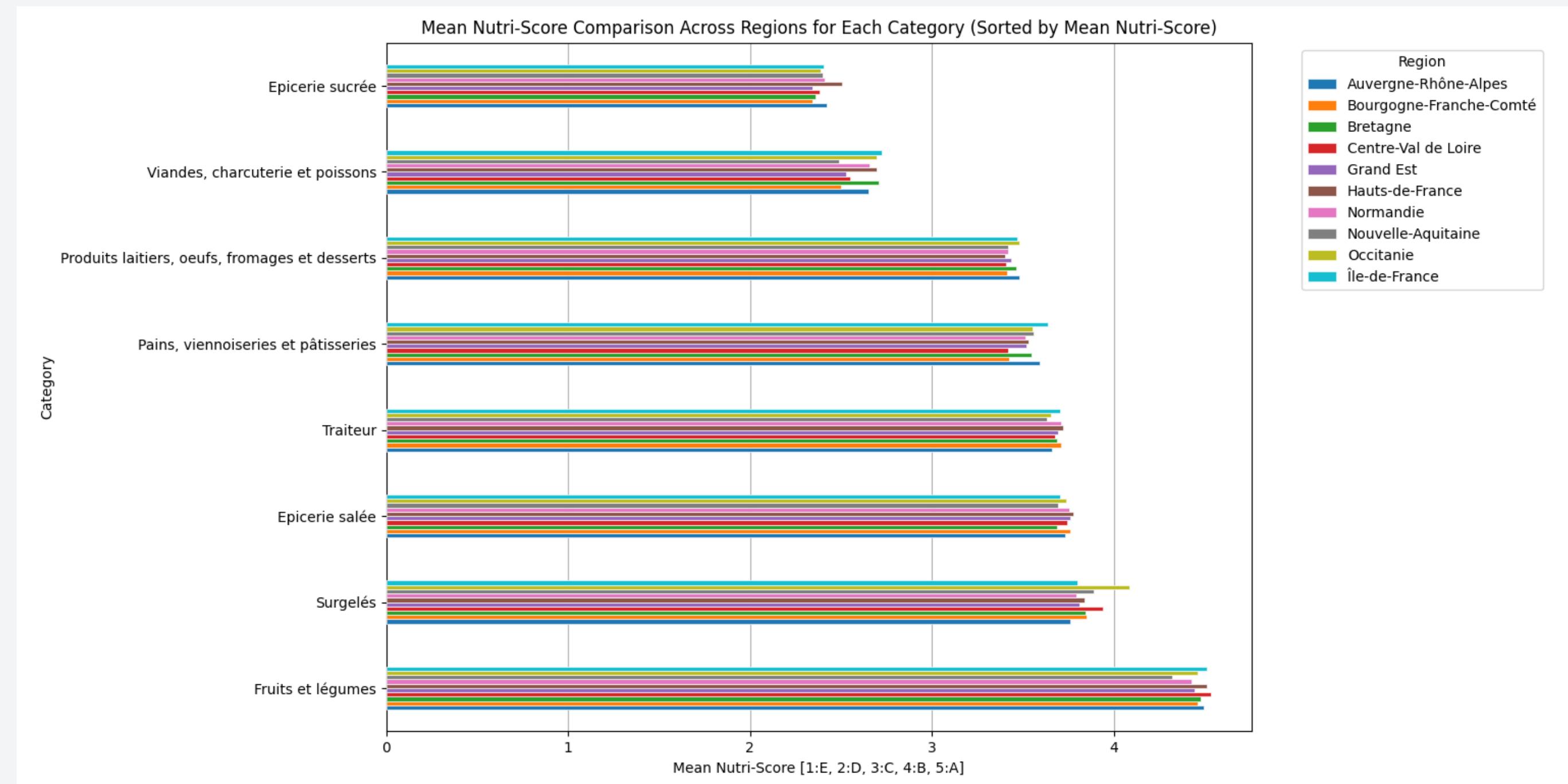


FIGURE 12: Stability of the Nutri-Score by region for each category.

SOURCE: [HTTPS://WWW.CORA.FR/](https://www.cora.fr/)

NUTRI-SCORE ANALYSIS

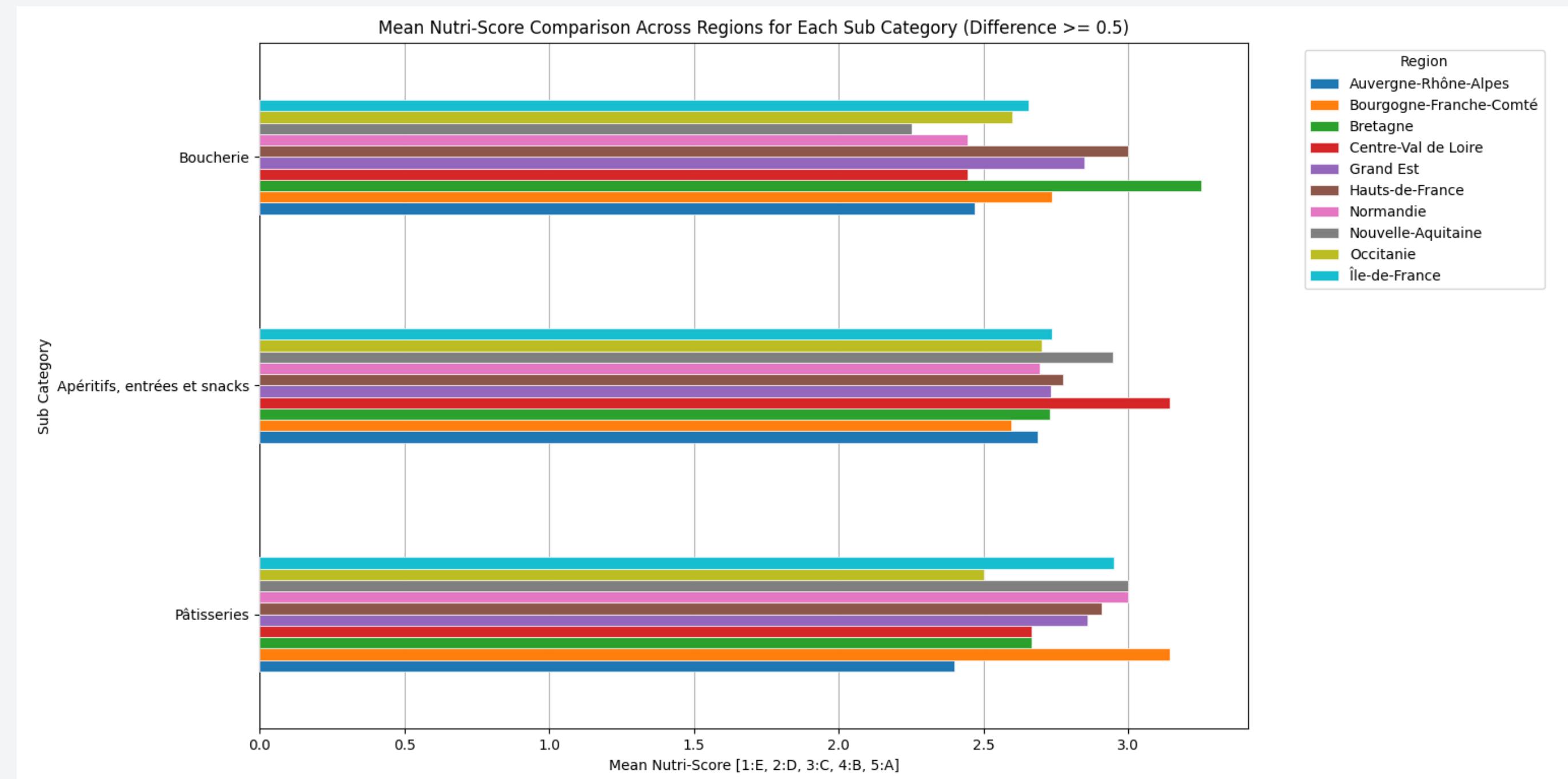


FIGURE 13: Some sub categories where we found a discrepancy of nutri-score. Despite this, the general trend remains the same.

CONCLUSION

Main Findings:

- Distribution of pricing and Nutri-Scores across France doesn't change across regions

Potential Shortcomings:

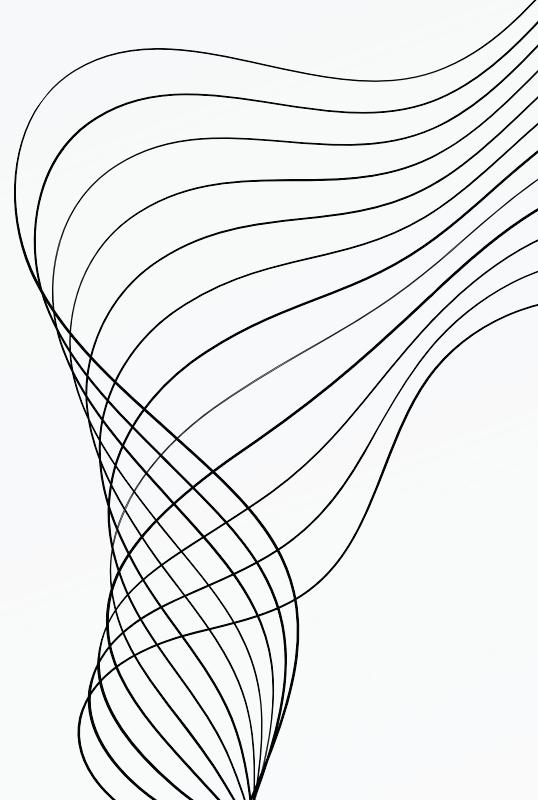
- Challenges in data collection (e.g., website access, VPN use)
- Limitations in data representatives

Future Research Avenues:

- Expanding data scope to more stores/regions
- Incorporating socio-economic factors
- Methodological improvements in data scraping



Thanks for your attention!



ALDO PIETROMATERA
GABRIELE LORENZO
LEONARDO H. C. SANTOS
YURI DE SENE ALVIZI