Introduction

- **Background:**
  - Wholesale spendings of customers in Portugal
  - For each customer: spending amounts (continuous, in monetary units) of 6 diverse product categories and 2 categorical features about their region and the channel used
  - Dataset of reasonable small size

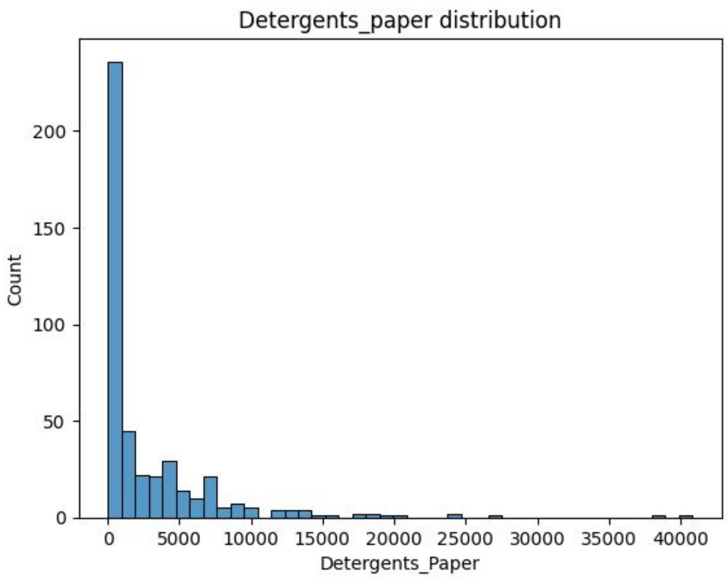
- **Objectives:**
  - Identify anomalous customers building a reproducible outlier detection model
  - Equip the model with explainability capabilities

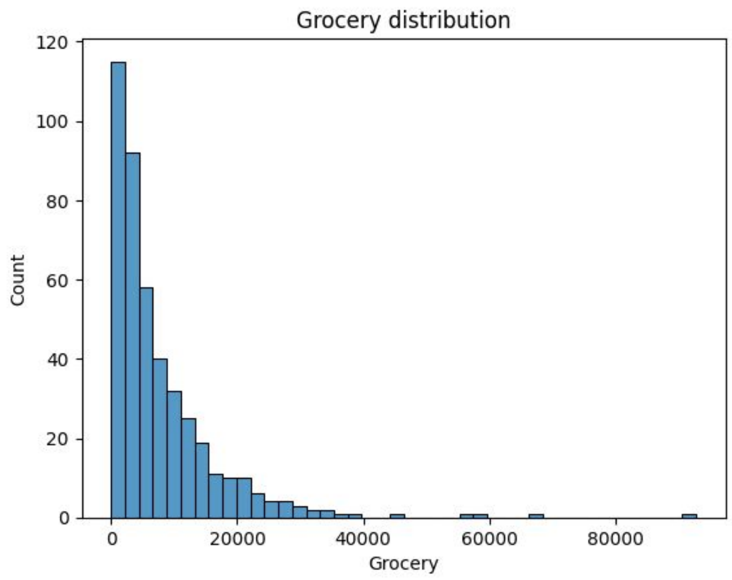
Exploratory data analysis

- **Attributes**:
    - Identified for the 2 categorical attributes, already integer encoded, which region or channel correspond to each encoded value
    - Feature selection

- Checked for **missing** data
- Plotted main descriptive statistics and **attributes' distribution (skewness)**

Exploratory data analysis





Exploratory data analysis

- Applied the following **log transformation** to the data:

$$x \mapsto \log(x + \theta)$$

- **Visual inspection** for the optimal offset
- Correlation matrix

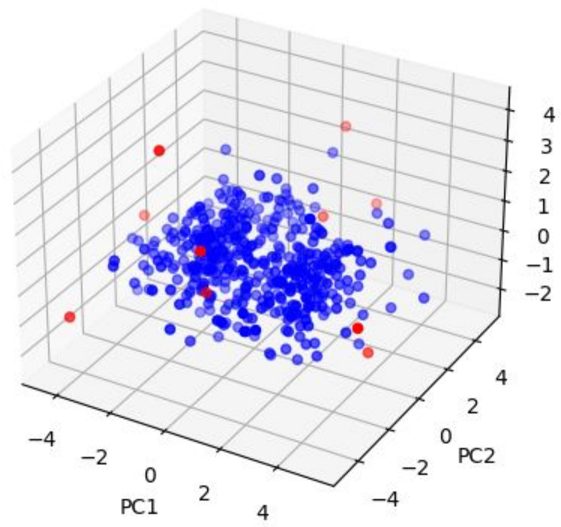
Anomaly detection - 1st method

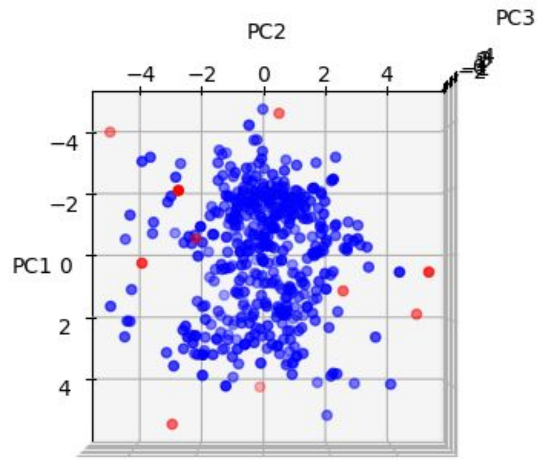
- **Choice of the model**:
  - distance-based leveraging **k-nearest neighbor distance**
- **Implementation**:
  - scikit-learn 2-Nearest Neighbor
  - anomaly score = minimum (squared) distance over the N-1 remaining points (hard min)

$$z_{jk} = \|\boldsymbol{x}_j - \boldsymbol{x}_k\|^2$$
$$y_j = \min_{k \neq j} z_{jk}$$

Anomaly detection - 1st method

- **Extreme-value analysis** and **PCA:** printed and plotted 10 most anomalous records





Anomaly detection - 2nd method

- **Issue** of the 1st model: low reproducibility
- **Solution**: replace hard-min distance with a soft-minimum one

$$y_j = \operatorname{soft\,min}_{k \neq j} \{z_{jk}\}$$

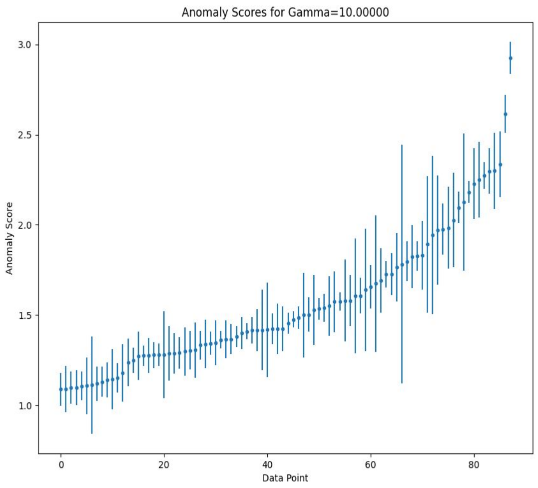
$$\operatorname{soft\,min}_{k \neq j} \{z_{jk}\} = -\frac{1}{\gamma} \log \left(\frac{1}{N-1} \sum_{k \neq j} \exp(-\gamma z_{jk}) \right).$$

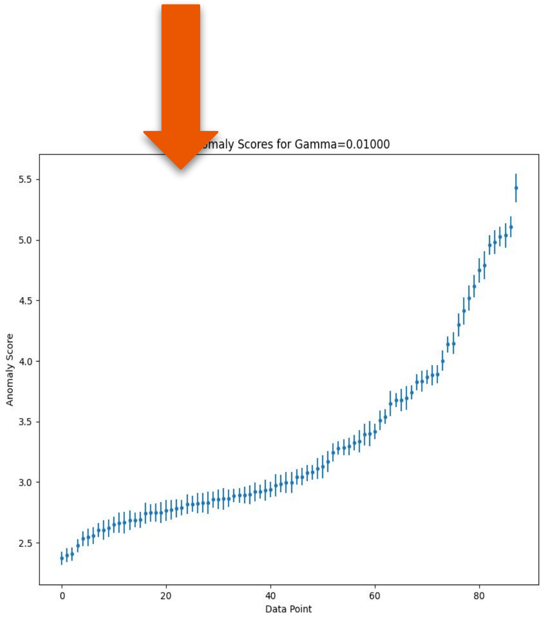
Anomaly detection - 2nd method

- **Hyperparameter optimization**:
  - data set splitted in train set (80%) and test set (20%)
  - hyperparameter gamma tuned in the range np.log (-4, 2, 7)
  - for each value of gamma, bootstrapped train set 100 times, computed the anomaly score for test instances along with mean and variance
  - chosen optimal value through visual inspection and evaluation metric

Anomaly detection - 2nd method

- **Visual inspection**





Anomaly detection - 2nd method

- Evaluation metric

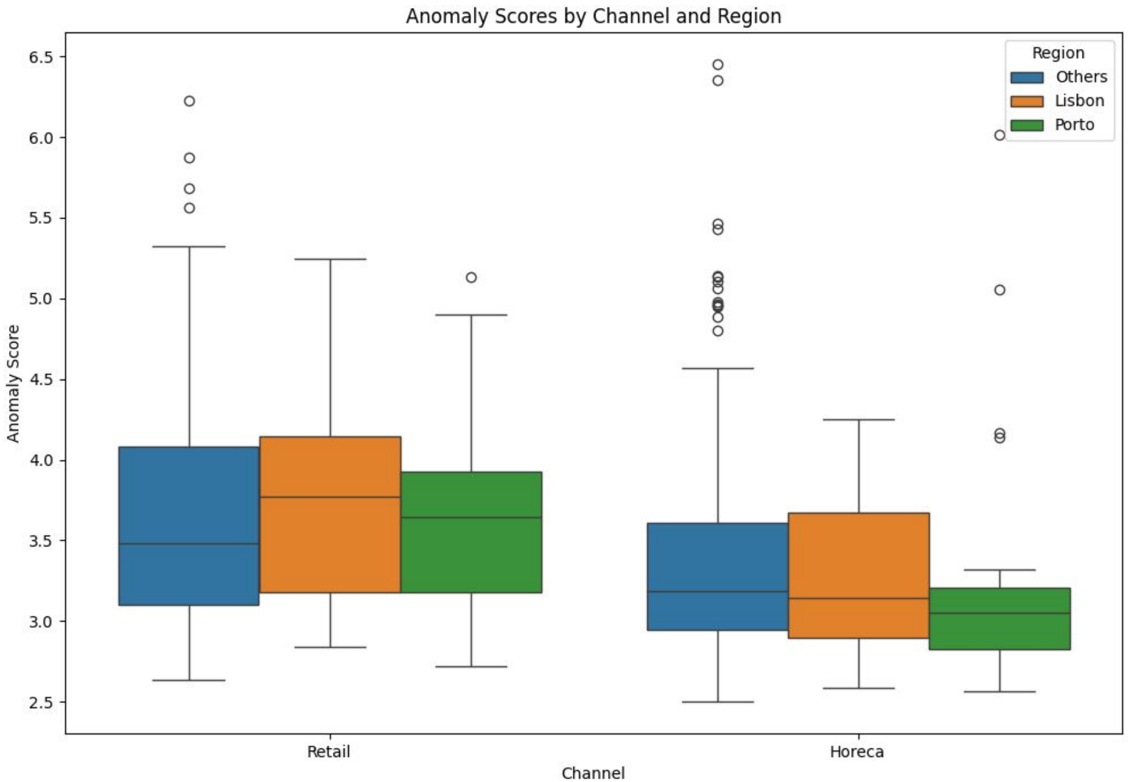
```
# Calculate mean and deviation of each datapoint across bootstraps
mean_scores /= 100
deviation_scores /= 100
deviation_scores -= mean_scores ** 2
deviation_scores = np.sqrt(deviation_scores)

# Normalize the mean and deviation scores
mean_scores_norm = (mean_scores - np.mean(mean_scores)) / np.std(mean_scores)
deviation_scores_norm = (deviation_scores - np.mean(deviation_scores)) / np.std(deviation_scores)

# Calculate metric that benefits low deviation and high difference in mean between normal and anomalous da
outliers = np.where(mean_scores_norm > np.percentile(mean_scores_norm, 97))[0]
non_outliers = np.where(mean_scores_norm <= np.percentile(mean_scores_norm, 97))[0]
metric = np.mean(np.abs(np.mean(outliers) - np.mean(non_outliers))) / np.mean((deviation_scores_norm)**2)
print(np.mean(np.abs(np.mean(outliers) - np.mean(non_outliers))))
print(np.mean((deviation_scores_norm)**2))
print("Gamma: ", gamma, " Metric: ", metric)
```

Explainability

- Where we can find most outliers?



Explainability

- Features' contribution to the overall outlier score?

  'NEON' framework (model → NN → explain via LRP)

Layer 1

Layer 2

$$z_{jk} = \| x_j - x_k \|^2$$

$$\operatorname*{soft\,min}_{k \neq j} \{z_{jk}\} = -\frac{1}{\gamma} \log \left(\frac{1}{N-1} \sum_{k \neq j} \exp(-\gamma z_{jk}) \right).$$

Explainability

- Features' contribution to the overall outlier score?
    - Redistribution of the anomaly score in the intermediate layer via:

Data points' contributions:

$$R_k^j = \frac{e^{-\gamma z_{jk}}}{\sum_{k \neq j} e^{-\gamma z_{jk}}} y_j$$

Features' contributions:

$$R_i^j = \frac{\sum [x_k - x_j]_i^2}{||x_k - x_j||^2} R_k^j$$

Explainability

- Features' contribution to the overall outlier score?
    - **Relevance conservation property**
    - Interactive plots

	Fresh	Milk	Grocery	Frozen	Detergents_Paper	Delicassen
Instance 154	622.000	55.000	137.000	75.000	7.000	8.000
Instance 39	56159.000	555.000	902.000	10002.000	212.000	2916.000
Average	12000.298	5796.266	7951.277	3071.932	2881.493	1524.870
Std	12647.329	7380.377	9503.163	4854.673	4767.854	2820.106