

# data\_processing

June 5, 2023

## 1 Data Processing - Complying with the EU AI Act

By: Jacintha Walters

This notebook is part of the research on ‘Complying with the EU AI Act’ which focuses on understanding to what extent organizations are prepared for the proposed legislation. To determine this, a questionnaire has been made to gain insight. In this notebook the data gathered from the questionnaire will be processed and prepared for clustering. The data has been anonymized, organization name is only known with the authors of the paper.

```
[111]: import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from sklearn.model_selection import train_test_split

dataset = pd.read_csv("anon_question_data.csv")
print("Data imported.")
```

Data imported.

The questionnaire contains 16 respondents.

### 1.1 Requirements

The goal is to score each respondent (organization) based on their answers. These scores will depict how well an organisation is prepared for the EU AI Act. Based on these scores an analysis can be made on different parameters to determine what influence these parameters have on the extent to which organizations are prepared to conform with the upcoming EU AI Act. Please refer to the paper to gain insights into the full process and argumentation. For this notebook, the following requirements were used: - Data must be made anonymously so that organization names and interviewees cannot be extracted. - The data must be scored using a rule-based AI, this means that: - Data must be encoded into numbers only. - Rules must be written as described in the paper. - A scoring system must be set-up so that each respondent is given a score on each section of the questionnaire along with a reflection score which depicts their self-knowledge. - The influence of the organizations characteristics must be examined, so graphs must be made depicting the relations. - At least one characteristic influence must be proven to be significant using a statistical method.

## 1.2 Dataset

The dataset is downloaded from Qualtrics, with text values turned into numbers and multiple choice questions are one-hot encoded. Some text values are still visible as text and should be processed. The data is made anonymous by removing the name and e-mail columns before importing the data into this notebook. Since the characteristics that are asked are very broad, this still ensures anonymity since it could be a number of organizations.

```
[95]: dataset.dtypes
```

```
[95]: StartDate      object
      EndDate        object
      Status         object
      Progress       object
      Duration (in seconds) object
      ...
      Q145           object
      Q146           object
      Q147           object
      Q148           object
      Q8             object
      Length: 183, dtype: object
```

All values are currently 'objects', these must be changed to integers for clustering. There are also irrelevant columns for clustering.

## 1.3 Dataset Cleaning

First the metadata columns should be removed.

```
[96]: dataset = dataset.drop(['StartDate', 'EndDate', 'Status', 'Progress', 'Duration_
      ↪(in seconds)', 'Finished', 'RecordedDate', 'ResponseId',
      ↪'DistributionChannel', 'UserLanguage'], axis=1)
```

Some questions were answered through the 'other' options where a text can be given. These must also be encoded. The written option is judged, and if worthy of a point encoded as such.

```
[98]: dataset['Q9_1'][3] = 1
      dataset['Q9_2'][3] = 1
```

At one questions organizations were asked to provide an example. This example will be used for the research, but it will be removed for clustering since it is not quantifiable. Questions that are variables interesting for comparison (sector, role etc.) are removed since they are not relevant for the score. Columns with text belonging to the 'other' option is also removed and points are added if the other option is worth points.

```
[100]:
```

```
dataset = dataset.drop(['Q132', 'Q9_5', 'Q109_4_TEXT', 'Q103_4_TEXT',
↳ 'Q83_5_TEXT', 'Q39_5_TEXT', 'Q61_5_TEXT', 'Q42_5_TEXT', 'Q91_6_TEXT',
↳ 'Q128_6_TEXT', 'Q85_5_TEXT', 'Q9_5_TEXT', 'Q48_5_TEXT', 'Q53_5_TEXT',
↳ 'Q5_4', 'Q5_4_TEXT', 'Q35_2_TEXT', 'Q22_NPS_GROUP', 'Q23_NPS_GROUP',
↳ 'Q16_NPS_GROUP', 'Q18_NPS_GROUP', 'Q25_NPS_GROUP'], axis=1)
```

For some questions one respondent used the ‘other...’ option to add ‘management’, this will be encoded by setting the other option to 1. If more options arise this should also be one-hot encoded.

Q85\_5\_TEXT = optie management Q91\_6\_TEXT = optie management Q128\_6\_TEXT = management Q42\_5\_TEXT = no department, option added later.

```
[101]: dataset['Q85_5'][10] = 0
dataset['Q42_6'][2] = 1
dataset['Q42_5'][2] = None
dataset['Q48_5'][10] = 0 #answered with 'I don't know, so 0 points
```

The first two rows are dropped with the question and option data. One response is deleted since a significant amount of questions have not been answered.

```
[102]: dataset = dataset.drop(index=[0, 1, 16])
```

Convert each row into floats.

```
[104]: dataset = dataset.apply(pd.to_numeric, downcast='float')
```

```
[105]: dataset.dtypes
```

```
[105]: Q5_1    float32
Q5_2    float32
Q5_3    float32
Q6      float32
Q9_1    float32
...
Q145    float32
Q146    float32
Q147    float32
Q148    float32
Q8      float32
Length: 151, dtype: object
```

NaN’s are replaced with 0 since NaN implies that data is missing, when instead the answer should be seen as a No.

```
[106]: dataset.fillna(0, inplace=True)
```

```
[107]: dataset = dataset.reset_index(drop=True)
```

```
[108]: dataset.head()
```

```
[108]:  Q5_1  Q5_2  Q5_3  Q6  Q9_1  Q9_2  Q9_3  Q9_4  Q7  Q10  ...  Q138  Q140  \
0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  0.0  1.0  2.0  ...  0.0  0.0
1  1.0  0.0  1.0  1.0  1.0  1.0  0.0  0.0  1.0  2.0  ...  0.0  0.0
2  1.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0  3.0  ...  0.0  0.0
3  0.0  0.0  0.0  4.0  0.0  0.0  0.0  0.0  1.0  4.0  ...  0.0  0.0
4  0.0  0.0  0.0  2.0  0.0  0.0  0.0  0.0  1.0  1.0  ...  0.0  0.0

      Q139  Q142  Q143  Q145  Q146  Q147  Q148  Q8
0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
1  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
2  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
3  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
4  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
```

[5 rows x 151 columns]

This finalizes the data processing. The requirements for data processing have been met by changing all answers to numerical values.

## 1.4 Scoring

Each questionnaire will be scored using a rule-based AI model.

```
[109]: rules = [
  {"question": "Q28", "condition": "response == 2", "score": 2, "aspect": "\u2192data", "type": "process", "perfect": "yes"},
  {"question": "Q28", "condition": "response == 1", "score": 1, "aspect": "\u2192data", "type": "process", "perfect": "no"},
  {"question": "Q29", "condition": "response == 1", "score": 2, "aspect": "\u2192data", "type": "process", "perfect": "yes"},
  {"question": "Q29", "condition": "response == 2", "score": 1, "aspect": "\u2192data", "type": "process", "perfect": "no"},
  {"question": "Q80", "condition": "response == 3", "score": 0.5, "aspect": "\u2192data", "type": "scale", "perfect": "no"},
  {"question": "Q80", "condition": "response == 4", "score": 1, "aspect": "\u2192data", "type": "scale", "perfect": "no"},
  {"question": "Q80", "condition": "response == 5", "score": 2, "aspect": "\u2192data", "type": "scale", "perfect": "yes"},
  {"question": "Q32", "condition": "response == 1", "score": 2, "aspect": "\u2192data", "type": "process", "perfect": "yes"},
  {"question": "Q32", "condition": "response == 2", "score": 1, "aspect": "\u2192data", "type": "process", "perfect": "no"},
  {"question": "Q33", "condition": "response == 1", "score": 2, "aspect": "\u2192data", "type": "process", "perfect": "yes"},
  {"question": "Q33", "condition": "response == 2", "score": 1, "aspect": "\u2192data", "type": "process", "perfect": "no"},
```

```

{"question": "Q79", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q79", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q79", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q35", "condition": "response == 2", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q36", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q36", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q37", "condition": "response == 5", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q37", "condition": "response == 3", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q78", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q78", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q78", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q39_1", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q39_2", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q39_3", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q39_4", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q39_5", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q77", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q77", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q77", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q41", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q77", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q77", "condition": "response == 3", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},

```

```

{"question": "Q42_1", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q42_2", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q42_3", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q42_4", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q75", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q75", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q75", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q76", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q76", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q76", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q45", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q45", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q46", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q46", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q47", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q47", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q48_1", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q48_2", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q48_3", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q48_4", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q48_5", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q74", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},

```

```

{"question": "Q74", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q74", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q50", "condition": "response == 1", "score": 1, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q50", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q51", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q51", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q73", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q73", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q73", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q53_1", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q53_2", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q53_3", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q53_4", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q53_5", "condition": "response == 1", "score": 0.5, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q71", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q71", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q71", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q55", "condition": "response == 1", "score": 2, "aspect": "data", "type": "process", "perfect": "yes"},
{"question": "Q55", "condition": "response == 2", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q55", "condition": "response == 3", "score": 1, "aspect": "data", "type": "process", "perfect": "no"},
{"question": "Q70", "condition": "response == 3", "score": 0.5, "aspect": "data", "type": "scale", "perfect": "no"},
{"question": "Q70", "condition": "response == 4", "score": 1, "aspect": "data", "type": "scale", "perfect": "no"},

```

```

{"question": "Q70", "condition": "response == 5", "score": 2, "aspect": "data", "type": "scale", "perfect": "yes"},
{"question": "Q58", "condition": "response == 1", "score": 1, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q58", "condition": "response == 2", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q58", "condition": "response == 3", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q69", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q69", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q69", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q68", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q68", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q68", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q61_1", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q61_2", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q61_3", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q61_4", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q62", "condition": "response == 3", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q62", "condition": "response == 1", "score": 2, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q63", "condition": "response == 1", "score": 1, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q67", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q67", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q67", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q66", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q66", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},

```



```

{"question": "Q66", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q82", "condition": "response == 1", "score": 2, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q82", "condition": "response == 3", "score": 2, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q82", "condition": "response == 2", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q82", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q83", "condition": "response == 1", "score": 1, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q84", "condition": "response == 1", "score": 2, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q84", "condition": "response == 2", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q85_1", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q85_2", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q85_3", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q85_4", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q85_5", "condition": "response == 1", "score": 0.5, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q86", "condition": "response == 1", "score": 2, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q87", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q87", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q87", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q88", "condition": "response == 1", "score": 1, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q88", "condition": "response == 2", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q89", "condition": "response == 1", "score": 2, "aspect": "documentation", "type": "process", "perfect": "yes"},
{"question": "Q89", "condition": "response == 2", "score": 2, "aspect": "documentation", "type": "process", "perfect": "no"},
{"question": "Q89", "condition": "response == 2", "score": 1, "aspect": "documentation", "type": "process", "perfect": "no"},

```

```

{"question": "Q90", "condition": "response == 3", "score": 0.5, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q90", "condition": "response == 4", "score": 1, "aspect": "documentation", "type": "scale", "perfect": "no"},
{"question": "Q90", "condition": "response == 5", "score": 2, "aspect": "documentation", "type": "scale", "perfect": "yes"},
{"question": "Q91_5", "condition": "response == 1", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q93", "condition": "response == 1", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q93", "condition": "response == 2", "score": 1, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q94", "condition": "response == 3", "score": 0.5, "aspect": "user", "type": "scale", "perfect": "no"},
{"question": "Q94", "condition": "response == 4", "score": 1, "aspect": "user", "type": "scale", "perfect": "no"},
{"question": "Q94", "condition": "response == 5", "score": 2, "aspect": "user", "type": "scale", "perfect": "yes"},
{"question": "Q95", "condition": "response == 1", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q95", "condition": "response == 2", "score": 2, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q95", "condition": "response == 3", "score": 1, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q96", "condition": "response == 1", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q96", "condition": "response == 2", "score": 1, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q97", "condition": "response == 3", "score": 0.5, "aspect": "user", "type": "scale", "perfect": "no"},
{"question": "Q97", "condition": "response == 4", "score": 1, "aspect": "user", "type": "scale", "perfect": "no"},
{"question": "Q97", "condition": "response == 5", "score": 2, "aspect": "user", "type": "scale", "perfect": "yes"},
{"question": "Q98", "condition": "response == 1", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q98", "condition": "response == 2", "score": 1, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q100", "condition": "response == 1", "score": 1, "aspect": "user", "type": "process", "perfect": "no"},
{"question": "Q100", "condition": "response == 2", "score": 2, "aspect": "user", "type": "process", "perfect": "yes"},
{"question": "Q99", "condition": "response == 3", "score": 0.5, "aspect": "user", "type": "scale", "perfect": "no"},

```

```

{"question": "Q99", "condition": "response == 4", "score": 1, "aspect": "user", "type": "scale", "perfect": "no"},
{"question": "Q99", "condition": "response == 5", "score": 2, "aspect": "user", "type": "scale", "perfect": "yes"},
{"question": "Q102", "condition": "response == 1", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q102", "condition": "response == 2", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q102", "condition": "response == 3", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q102", "condition": "response == 4", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q103_1", "condition": "response == 1", "score": 0.5, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q103_2", "condition": "response == 1", "score": 0.5, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q103_3", "condition": "response == 1", "score": 0.5, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q106", "condition": "response == 1", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q104", "condition": "response == 3", "score": 0.5, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q104", "condition": "response == 4", "score": 1, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q104", "condition": "response == 5", "score": 2, "aspect": "monitoring", "type": "scale", "perfect": "yes"},
{"question": "Q105", "condition": "response == 1", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q105", "condition": "response == 2", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q108", "condition": "response == 1", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q108", "condition": "response == 2", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q107", "condition": "response == 1", "score": 2, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q107", "condition": "response == 2", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "no"},
{"question": "Q109_1", "condition": "response == 1", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q109_2", "condition": "response == 1", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "yes"},
{"question": "Q109_4", "condition": "response == 1", "score": 1, "aspect": "monitoring", "type": "process", "perfect": "yes"},

```

```

{"question": "Q110", "condition": "response == 3", "score": 0.5, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q110", "condition": "response == 4", "score": 1, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q110", "condition": "response == 5", "score": 2, "aspect": "monitoring", "type": "scale", "perfect": "yes"},
{"question": "Q111", "condition": "response == 3", "score": 0.5, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q111", "condition": "response == 4", "score": 1, "aspect": "monitoring", "type": "scale", "perfect": "no"},
{"question": "Q111", "condition": "response == 5", "score": 2, "aspect": "monitoring", "type": "scale", "perfect": "yes"},
{"question": "Q118", "condition": "response == 1", "score": 2, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q118", "condition": "response == 2", "score": 1, "aspect": "risk", "type": "process", "perfect": "no"},
{"question": "Q119_1", "condition": "response == 1", "score": 0.5, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q119_2", "condition": "response == 1", "score": 0.5, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q119_3", "condition": "response == 1", "score": 0.5, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q119_4", "condition": "response == 1", "score": 0.5, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q120", "condition": "response == 1", "score": 2, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q120", "condition": "response == 3", "score": 1, "aspect": "risk", "type": "process", "perfect": "no"},
{"question": "Q121", "condition": "response == 1", "score": 2, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q121", "condition": "response == 2", "score": 1, "aspect": "risk", "type": "process", "perfect": "no"},
{"question": "Q122", "condition": "response == 3", "score": 0.5, "aspect": "risk", "type": "scale", "perfect": "no"},
{"question": "Q122", "condition": "response == 4", "score": 1, "aspect": "risk", "type": "scale", "perfect": "no"},
{"question": "Q122", "condition": "response == 5", "score": 2, "aspect": "risk", "type": "scale", "perfect": "yes"},
{"question": "Q123", "condition": "response == 1", "score": 2, "aspect": "risk", "type": "process", "perfect": "yes"},
{"question": "Q123", "condition": "response == 2", "score": 1, "aspect": "risk", "type": "process", "perfect": "no"},
{"question": "Q124", "condition": "response == 1", "score": 2, "aspect": "risk", "type": "process", "perfect": "yes"},

```

```

    {"question": "Q125", "condition": "response == 2", "score": 2, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q126", "condition": "response == 3", "score": 0.5, "aspect": "
↪risk", "type": "scale", "perfect": "no"},
    {"question": "Q126", "condition": "response == 4", "score": 1, "aspect": "
↪risk", "type": "scale", "perfect": "no"},
    {"question": "Q126", "condition": "response == 5", "score": 2, "aspect": "
↪risk", "type": "scale", "perfect": "yes"},
    {"question": "Q127", "condition": "response == 1", "score": 2, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q127", "condition": "response == 2", "score": 1, "aspect": "
↪risk", "type": "process", "perfect": "no"},
    {"question": "Q128_1", "condition": "response == 1", "score": 0.5, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q128_2", "condition": "response == 1", "score": 0.5, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q128_3", "condition": "response == 1", "score": 0.5, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q128_4", "condition": "response == 1", "score": 0.5, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q128_6", "condition": "response == 1", "score": 0.5, "aspect": "
↪risk", "type": "process", "perfect": "yes"},
    {"question": "Q129", "condition": "response == 3", "score": 0.5, "aspect": "
↪risk", "type": "scale", "perfect": "no"},
    {"question": "Q129", "condition": "response == 4", "score": 1, "aspect": "
↪risk", "type": "scale", "perfect": "no"},
    {"question": "Q129", "condition": "response == 5", "score": 2, "aspect": "
↪risk", "type": "scale", "perfect": "yes"},
]

print("Number of rules:", len(rules))

```

Number of rules: 201

The following code contains the AI-part which uses the rules to automatically score a response.

```

[110]: import math

data_perfect_score = 0
doc_perfect_score = 0
user_perfect_score = 0
monitor_perfect_score = 0
risk_perfect_score = 0
perfect_total_score = 0

process_perfect_score = 0
scale_perfect_score = 0

```

```

# Calculate the maximum achievable points
for rule in rules:
    if rule["perfect"] == "yes":
        if rule["aspect"] == "data":
            data_perfect_score += rule["score"]
        elif rule["aspect"] == "documentation":
            doc_perfect_score += rule["score"]
        elif rule["aspect"] == "user":
            user_perfect_score += rule["score"]
        elif rule["aspect"] == "monitoring":
            monitor_perfect_score += rule["score"]
        elif rule["aspect"] == "risk":
            risk_perfect_score += rule["score"]
        perfect_total_score += rule["score"]
    if rule["type"] == "process":
        process_perfect_score += rule["score"]
    elif rule["type"] == "scale":
        scale_perfect_score += rule["score"]

def score_response(response, rules):
    # Initialize the scores for different aspects
    data_score = 0
    doc_score = 0
    user_score = 0
    monitor_score = 0
    risk_score = 0

    scale_score = 0
    process_score = 0

    # Iterate over the questions for each response
    for question, answer in response.items():
        # Check if the question is in the rules list
        for rule in rules:
            if rule["question"] == question:
                score = rule["score"]
                # Evaluate the condition and add the score if the condition is
                ↪satisfied
                if eval(rule["condition"], {"response": answer}):
                    # Update the corresponding score based on the aspect
                    if rule["aspect"] == "data":
                        data_score += score
                    elif rule["aspect"] == "documentation":
                        doc_score += score
                    elif rule["aspect"] == "user":

```

```

        user_score += score
    elif rule["aspect"] == "monitoring":
        monitor_score += score
    elif rule["aspect"] == "risk":
        risk_score += score
    if rule["type"] == "scale":
        scale_score += score
    elif rule["type"] == "process":
        process_score += score

    # Return the scores for different aspects
    return data_score, doc_score, user_score, monitor_score, risk_score, \
↳scale_score, process_score

data_scored_total = 0
doc_scored_total = 0
user_scored_total = 0
monitor_scored_total = 0
risk_scored_total = 0

# Iterate over the dataset and calculate scores for each response
i = 0
while i < len(dataset.index):
    print("Checking response ", dataset.index[i], "...")
    response = dataset.iloc[i]
    data_score, doc_score, user_score, monitor_score, risk_score, scale_score, \
↳process_score = score_response(response, rules)
    data_scored_total += data_score
    doc_scored_total += doc_score
    user_scored_total += user_score
    monitor_scored_total += monitor_score
    risk_scored_total += risk_score
    print("Score on data and model internals: ", data_score, "/", \
↳data_perfect_score, " - ", math.ceil(100 * data_score / data_perfect_score), \
↳"%")
    print("Score on technical documentation: ", doc_score, "/", \
↳doc_perfect_score, " - ", math.ceil(100 * doc_score / doc_perfect_score), \
↳"%") )
    print("Score on user communication: ", user_score, "/", user_perfect_score, \
↳" - ", math.ceil(100 * user_score / user_perfect_score), "%")
    print("Score on model monitoring: ", monitor_score, "/", \
↳monitor_perfect_score, " - ", math.ceil(100 * monitor_score / \
↳monitor_perfect_score), "%")
    print("Score on risk management: ", risk_score, "/", risk_perfect_score, " \
↳- ", math.ceil(100 * risk_score / risk_perfect_score), "%")

```

```

    print("TOTAL SCORE: ", data_score + doc_score + user_score + monitor_score +
    ↪ risk_score, "/", perfect_total_score, " - ", math.ceil(100 * (data_score +
    ↪ doc_score + user_score + monitor_score + risk_score) / perfect_total_score),
    ↪ "%")

    print("")
    print("Points for process questions:", process_score, "/",
    ↪ process_perfect_score, "-", math.ceil(100 * process_score /
    ↪ process_perfect_score), "%")
    print("Points for self reflection questions:", scale_score, "/",
    ↪ scale_perfect_score, "-", math.ceil(100 * scale_score /
    ↪ scale_perfect_score), "%")
    print("Reflection score:", round((process_score/process_perfect_score) /
    ↪ (scale_score/scale_perfect_score),1))
    print("-----")
    i += 1

#calculate average score per aspect over all respondents
print("Average score per aspect over all respondents...")
print("Score on data and model internals:", data_scored_total, "/",
    ↪ data_perfect_score * len(dataset.index), " - ", math.ceil(100 *
    ↪ data_scored_total / (data_perfect_score * len(dataset.index))), "%")
print("Score on technical documentation: ", doc_scored_total, "/",
    ↪ doc_perfect_score * len(dataset.index), " - ", math.ceil(100 *
    ↪ doc_scored_total / (doc_perfect_score * len(dataset.index))), "%")
print("Score on user communication:", user_scored_total, "/",
    ↪ user_perfect_score * len(dataset.index), " - ", math.ceil(100 *
    ↪ user_scored_total / (user_perfect_score * len(dataset.index))), "%")
print("Score on model monitoring:", monitor_scored_total, "/",
    ↪ monitor_perfect_score * len(dataset.index), " - ", math.ceil(100 *
    ↪ monitor_scored_total / (monitor_perfect_score * len(dataset.index))), "%")
print("Score on risk management:", risk_scored_total, "/", risk_perfect_score *
    ↪ len(dataset.index), " - ", math.ceil(100 * risk_scored_total /
    ↪ (risk_perfect_score * len(dataset.index))), "%")
total_scored = data_scored_total + doc_scored_total + user_scored_total +
    ↪ monitor_scored_total + risk_scored_total
possible_scored = (data_perfect_score + doc_perfect_score + user_perfect_score +
    ↪ monitor_perfect_score + risk_perfect_score) * len(dataset.index)
print("Average total score:", total_scored, "/", possible_scored, '-', math.
    ↪ ceil(100 * total_scored / (possible_scored)), "%")

```

Checking response 0 ...

```

Score on data and model internals: 29.0 / 56.5 - 52 %
Score on technical documentation: 11.0 / 30.5 - 37 %
Score on user communication: 13.5 / 18 - 75 %
Score on model monitoring: 16.5 / 20.5 - 81 %
Score on risk management: 17.0 / 24.5 - 70 %
TOTAL SCORE: 87.0 / 150.0 - 58 %

```



Points for process questions: 56.5 / 100.0 - 57 %  
Points for self reflection questions: 30.5 / 50 - 61 %  
Reflection score: 0.9  
-----

Checking response 1 ...  
Score on data and model internals: 30.0 / 56.5 - 54 %  
Score on technical documentation: 12.5 / 30.5 - 41 %  
Score on user communication: 9 / 18 - 50 %  
Score on model monitoring: 13.5 / 20.5 - 66 %  
Score on risk management: 11.5 / 24.5 - 47 %  
TOTAL SCORE: 76.5 / 150.0 - 51 %

Points for process questions: 54.5 / 100.0 - 55 %  
Points for self reflection questions: 22.0 / 50 - 44 %  
Reflection score: 1.2  
-----

Checking response 2 ...  
Score on data and model internals: 17.0 / 56.5 - 31 %  
Score on technical documentation: 6.0 / 30.5 - 20 %  
Score on user communication: 5 / 18 - 28 %  
Score on model monitoring: 8.5 / 20.5 - 42 %  
Score on risk management: 13.5 / 24.5 - 56 %  
TOTAL SCORE: 50.0 / 150.0 - 34 %

Points for process questions: 34.5 / 100.0 - 35 %  
Points for self reflection questions: 15.5 / 50 - 31 %  
Reflection score: 1.1  
-----

Checking response 3 ...  
Score on data and model internals: 42.0 / 56.5 - 75 %  
Score on technical documentation: 9.5 / 30.5 - 32 %  
Score on user communication: 11.5 / 18 - 64 %  
Score on model monitoring: 10.5 / 20.5 - 52 %  
Score on risk management: 15.0 / 24.5 - 62 %  
TOTAL SCORE: 88.5 / 150.0 - 59 %

Points for process questions: 59.0 / 100.0 - 59 %  
Points for self reflection questions: 29.5 / 50 - 59 %  
Reflection score: 1.0  
-----

Checking response 4 ...  
Score on data and model internals: 42.0 / 56.5 - 75 %  
Score on technical documentation: 23.0 / 30.5 - 76 %  
Score on user communication: 13 / 18 - 73 %  
Score on model monitoring: 15.0 / 20.5 - 74 %  
Score on risk management: 20.5 / 24.5 - 84 %  
TOTAL SCORE: 113.5 / 150.0 - 76 %

Points for process questions: 72.5 / 100.0 - 73 %  
Points for self reflection questions: 41 / 50 - 82 %  
Reflection score: 0.9  
-----

Checking response 5 ...  
Score on data and model internals: 28.5 / 56.5 - 51 %  
Score on technical documentation: 7.5 / 30.5 - 25 %  
Score on user communication: 5 / 18 - 28 %  
Score on model monitoring: 11.5 / 20.5 - 57 %  
Score on risk management: 10.0 / 24.5 - 41 %  
TOTAL SCORE: 62.5 / 150.0 - 42 %

Points for process questions: 45.5 / 100.0 - 46 %  
Points for self reflection questions: 17.0 / 50 - 34 %  
Reflection score: 1.3  
-----

Checking response 6 ...  
Score on data and model internals: 33.0 / 56.5 - 59 %  
Score on technical documentation: 23.5 / 30.5 - 78 %  
Score on user communication: 11 / 18 - 62 %  
Score on model monitoring: 15.0 / 20.5 - 74 %  
Score on risk management: 17.0 / 24.5 - 70 %  
TOTAL SCORE: 99.5 / 150.0 - 67 %

Points for process questions: 65.0 / 100.0 - 65 %  
Points for self reflection questions: 34.5 / 50 - 69 %  
Reflection score: 0.9  
-----

Checking response 7 ...  
Score on data and model internals: 20.0 / 56.5 - 36 %  
Score on technical documentation: 4.5 / 30.5 - 15 %  
Score on user communication: 4.5 / 18 - 25 %  
Score on model monitoring: 7.0 / 20.5 - 35 %  
Score on risk management: 2 / 24.5 - 9 %  
TOTAL SCORE: 38.0 / 150.0 - 26 %

Points for process questions: 24.5 / 100.0 - 25 %  
Points for self reflection questions: 13.5 / 50 - 27 %  
Reflection score: 0.9  
-----

Checking response 8 ...  
Score on data and model internals: 36.0 / 56.5 - 64 %  
Score on technical documentation: 3.0 / 30.5 - 10 %  
Score on user communication: 9.0 / 18 - 50 %  
Score on model monitoring: 18.5 / 20.5 - 91 %  
Score on risk management: 19.0 / 24.5 - 78 %  
TOTAL SCORE: 85.5 / 150.0 - 57 %

Points for process questions: 60.0 / 100.0 - 60 %  
Points for self reflection questions: 25.5 / 50 - 51 %  
Reflection score: 1.2  
-----

Checking response 9 ...  
Score on data and model internals: 22.0 / 56.5 - 39 %  
Score on technical documentation: 24.0 / 30.5 - 79 %  
Score on user communication: 7 / 18 - 39 %  
Score on model monitoring: 9.0 / 20.5 - 44 %  
Score on risk management: 12.0 / 24.5 - 49 %  
TOTAL SCORE: 74.0 / 150.0 - 50 %

Points for process questions: 45.0 / 100.0 - 45 %  
Points for self reflection questions: 29.0 / 50 - 58 %  
Reflection score: 0.8  
-----

Checking response 10 ...  
Score on data and model internals: 36.5 / 56.5 - 65 %  
Score on technical documentation: 13.5 / 30.5 - 45 %  
Score on user communication: 15 / 18 - 84 %  
Score on model monitoring: 14.5 / 20.5 - 71 %  
Score on risk management: 17.0 / 24.5 - 70 %  
TOTAL SCORE: 96.5 / 150.0 - 65 %

Points for process questions: 65.0 / 100.0 - 65 %  
Points for self reflection questions: 31.5 / 50 - 63 %  
Reflection score: 1.0  
-----

Checking response 11 ...  
Score on data and model internals: 36.5 / 56.5 - 65 %  
Score on technical documentation: 27.0 / 30.5 - 89 %  
Score on user communication: 18 / 18 - 100 %  
Score on model monitoring: 19.0 / 20.5 - 93 %  
Score on risk management: 20.5 / 24.5 - 84 %  
TOTAL SCORE: 121.0 / 150.0 - 81 %

Points for process questions: 71.0 / 100.0 - 71 %  
Points for self reflection questions: 50 / 50 - 100 %  
Reflection score: 0.7  
-----

Checking response 12 ...  
Score on data and model internals: 43.5 / 56.5 - 77 %  
Score on technical documentation: 16.0 / 30.5 - 53 %  
Score on user communication: 16 / 18 - 89 %  
Score on model monitoring: 16.5 / 20.5 - 81 %  
Score on risk management: 20.0 / 24.5 - 82 %  
TOTAL SCORE: 112.0 / 150.0 - 75 %

Points for process questions: 68.5 / 100.0 - 69 %  
Points for self reflection questions: 43.5 / 50 - 87 %  
Reflection score: 0.8  
-----

Checking response 13 ...  
Score on data and model internals: 41.0 / 56.5 - 73 %  
Score on technical documentation: 14.0 / 30.5 - 46 %  
Score on user communication: 12 / 18 - 67 %  
Score on model monitoring: 3 / 20.5 - 15 %  
Score on risk management: 12.5 / 24.5 - 52 %  
TOTAL SCORE: 82.5 / 150.0 - 55 %

Points for process questions: 55.0 / 100.0 - 55 %  
Points for self reflection questions: 27.5 / 50 - 55 %  
Reflection score: 1.0  
-----

Checking response 14 ...  
Score on data and model internals: 32.5 / 56.5 - 58 %  
Score on technical documentation: 17.5 / 30.5 - 58 %  
Score on user communication: 11 / 18 - 62 %  
Score on model monitoring: 12.5 / 20.5 - 61 %  
Score on risk management: 19.0 / 24.5 - 78 %  
TOTAL SCORE: 92.5 / 150.0 - 62 %

Points for process questions: 60.0 / 100.0 - 60 %  
Points for self reflection questions: 32.5 / 50 - 65 %  
Reflection score: 0.9  
-----

Average score per aspect over all respondents...  
Score on data and model internals: 489.5 / 847.5 - 58 %  
Score on technical documentation: 212.5 / 457.5 - 47 %  
Score on user communication: 160.5 / 270 - 60 %  
Score on model monitoring: 190.5 / 307.5 - 62 %  
Score on risk management: 226.5 / 367.5 - 62 %  
Average total score: 1279.5 / 2250.0 - 57 %