Ais Challenge

Aziz Becheur

Vollständigkeitsprüfung:

Die Heatmap zeigt den Prozentsatz der Vollständigkeit jedes Kunden-Gebäudes im Vergleich zur Referenz

		Cover	age (%) of 'A	ınlagen' per	Customer Bu	uilding		100
2028 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	100
546 -	92.3	61.5	53.8	46.2	84.6	92.3	61.5	
2024 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2032 -	85.7	61.5	58.3	50.0	78.6	85.7	61.5	
2033 -	92.3	66.7	58.3	50.0	84.6	92.3	66.7	
2025 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2029 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	- 80
2022 -	80.0	53.3	46.7	40.0	73.3	80.0	53.3	
2034 -	85.7	66.7	58.3	50.0	84.6	85.7	66.7	
2043 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	
2096 -	85.7	57.1	50.0	42.9	78.6	85.7	57.1	
2038 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2039 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	- 60
2042 -	85.7	57.1	50.0	42.9	78.6	85.7	57.1	00
<u><u><u><u></u></u> <u><u></u> 2035 -</u></u></u>	92.3	66.7	58.3	50.0	84.6	92.3	66.7	
. <u>B</u> 2023 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
ည္တို့ 2036 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	
Seference Building 2035 - 2020 - 2020 - 2104 - 2104 - 2020 - 2104 - 2020 - 2104	85.7	61.5	53.8	46.2	78.6	85.7	61.5	
ž 2104 -	100.0	88.9	77.8	75.0	100.0	100.0	88.9	
2041 -	80.0	53.3	46.7	40.0	73.3	80.0	53.3	- 40
2040 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2017 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2095 -	92.3	80.0	70.0	60.0	91.7	92.3	80.0	
543 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	
2021 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2037 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	- 20
544 -	80.0	53.3	46.7	40.0	73.3	80.0	53.3	
2030 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2026 -	80.0	53.3	46.7	40.0	73.3	80.0	53.3	
2027 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
2031 -	80.0	57.1	50.0	42.9	73.3	80.0	57.1	
545 -	85.7	61.5	53.8	46.2	78.6	85.7	61.5	
	26287	21083	28580 C	30009 ustomer Buildir	22225 ng	15859	18925	- 0

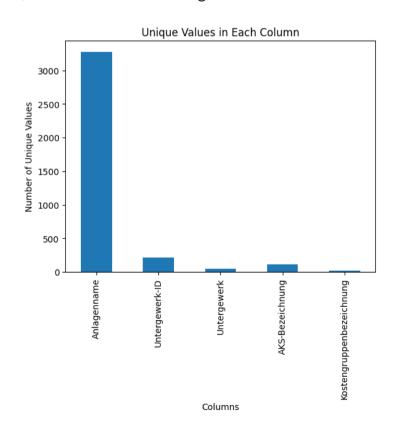
Lösungsweg:

Data Exploration

1. Die Kundendaten und Referenzdaten in einer hierarchischen Struktur organisieren, sodass:

Die Kostengruppenbezeichnung (entsprechende Kostengruppe als ID) einzigartige Unterkategorien von AKS-Bezeichnung (entsprechendes Untergewerk als ID) enthält und jede AKS-Bezeichnung wiederum in mehrere Untergewerk-IDs unterteilt ist.

- 2. In JSON Datei Speichern als Referenz
- -> customer_data/jsons/ reference_data/jsons
- 3. Jedes Gebäude sollte in einer eigenen CSV-Datei gespeichert werden, um sie miteinander vergleichen zu können.



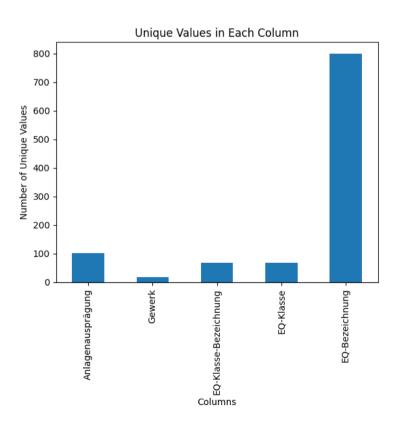
Beispiel von Hierarchie: Lufttechnische Anlagen

- Kompressionskältemaschine

- Kleinkaltwassersatz 109 kW
- Kleinkaltwassersatz 237kW
- VRV Anlage
- Kältemaschine Lüftungsanlagen
- KM Kühlraum Küche
-

— Verteilnetz / Pumpen - Kälteerzeugung

- Kältenetz Dach Deckenkühl..
- Kältenetz Lüftungsgeräte



Lösungsweg:

Data Manipulation

- 4. Jede Mögliche Sequenz von Kategorie (KostenGruppenBezeichnung) + Unterkategorie (AKS-Bezeichnung) + Anlagen werden in einem separaten langen Text gespeichert und zu einem Vektor (Embedding) umwandeln. (Gleich auch für Kundendatei)
- -> 3_create_mapping.ipynb
- 5. **Semantischer Vergleich:** Die Nähe jedes **Embedding** zwischen Kundendatei und Referenz wird als ein Wert berechnet, den wir Cosine Similarity nennen. Diese Cosine Similarity zeigt, wie ähnlich sich zwei Einbettungen sind. (Check outputs/mapping.json)
- -> 3_create_mapping.ipynb
- 6. Nested semantischer Vergleich, sodass wir Übereinstimmungen der Unterkategorien zweier übereinstimmender Kategorien finden.
- -> 3_create_mapping.ipynb
- 7. Der verschachtelte semantische Vergleich hilft uns letztlich dabei, den EP-Katalog der Kundendatei zu erstellen. Jede übereinstimmende Kategorie/Unterkategorie mit den Referenzdaten erhält die entsprechende Referenz-ID zugewiesen.

RLT: 1. Lufttechnische Anlagen (score: 0.64) 2. Baukonstruktionen in Außenanlagen (score: 0.62) 3. Nutzungsspezifische Anlagen (score: 0.61) 4. Abwasser-, Wasser-, Gasanlagen (score: 0.60) 5. Neu Ausstattung (score: 0.59) Elektro: 1. Baukonstruktionen in Außenanlagen (score: 0.63) 2. Neu Ausstattung (score: 0.58) 3. Gebäudeautomation (score: 0.57) 4. Starkstromanlagen (score: 0.55) 5. Fernmelde- und informationstechnische Anlagen (score: 0.53)

```
{'31211':

1. {'432' (score: 0.72)
2. {'439' (score: 0.71)
3. {'431' (score: 0.70)
4. {'433' (score: 0.65)
5. {'434' (score: 0.64)

{'31221':
1. {'439' (score: 0.62)
2. {'431' (score: 0.60)
3. {'432' (score: 0.60)
4. {'434' (score: 0.55)
5. {'433' (score: 0.50)
```

Um die einzelnen Schritte der Lösung nachvollziehen zu können, folgen Sie bitte der Reihenfolge der Jupyter-Notebooks:

- 1. 1_preparation_ref.ipynb
- 2. 2_preparation_data.ipynb
- 3. 3_create_mapping.ipynb
- 4. 4_comparison.ipynb
- 5. 5_zuordnen.ipynb

Eine ausführlichere Beschreibung der einzelnen Schritte finden Sie ebenfalls in den jeweiligen Notebooks.