Thanks — this gives us a clear picture of linie.csv.

**✅ Key Takeaways from linie.csv:**

**📂 Basic Stats:**

* **Rows**: 431 railway line segments
* **Delimiter**: ; correctly parsed
* **Columns**: 10 (but 3 are garbage metadata columns)

**✅ Useful Columns:**

| **Column Name** | **Description** |
| --- | --- |
| Line | Likely a unique numeric ID per segment |
| Line.1 | Human-readable name of the segment |
| START OPK | Start station/op point name |
| END OPK | End station/op point name |
| KM START / KM END | Railway kilometers (position on track) |
| geo\_point\_2d | Single coordinate, possibly mid-point |
| Content-Type... | **Irrelevant garbage**, should be dropped |

**🔥 Most Important Field:**

The column with this value:

{"coordinates": [[8.7171, 47.4921], [8.7093, 47.4898]], "type": "LineString"}

...is embedded under a misleading column name like:

"Content-Type: text/plain; charset=utf-8\n.2"

It’s clearly a **LineString geometry** field — likely copied from linie\_mit\_polygon.csv. So this dataset is a **simpler version** of linie\_mit\_polygon.csv.

🧾 Columns:

- Line

- Line.1

- START OPK

- END OPK

- KM START

- KM END

- Content-Type: text/plain; charset=utf-8

- Content-Type: text/plain; charset=utf-8

.1

- Content-Type: text/plain; charset=utf-8

.2

- geo\_point\_2d

2025-05-12 15:41:00,132 [INFO]

🔍 Sample rows:

Line Line.1 START OPK END OPK KM START KM END Content-Type: text/plain; charset=utf-8\n Content-Type: text/plain; charset=utf-8\n.1 Content-Type: text/plain; charset=utf-8\n.2 geo\_point\_2d

9794 W'thur West Unterquerung - W'thur Toss Winterthur West Unterquerung Winterthur Toss 51.033 51.718 0 1000000 {"coordinates": [[8.71716739222, 47.4921217692], [8.70930994344, 47.4898063841]], "type": "LineString"} 47.49096407665, 8.71323866783

9793 W'thur Eichliacker - W'thur Wylandbrucke Winterthur Eichliacker (Vzw) Winterthur Wylandbrucke (Vzw) 54.583 55.495 0 1000000 {"coordinates": [[8.71370075634, 47.4881074605], [8.71959354602, 47.4952673938]], "type": "LineString"} 47.49168742715, 8.71664715118

9792 Tossmuhle Steigbach-W'thur Eichliacker Tossmuhle Steigbach (Vzw) Winterthur Eichliacker (Vzw) 53.415 54.583 0 1000000 {"coordinates": [[8.70653172432, 47.4788163018], [8.71370075634, 47.4881074605]], "type": "LineString"} 47.48346188115, 8.710116240329999

9791 Wallisellen Furt - Dietlikon Sud-West Wallisellen Furt (Vzw) Dietlikon Sud-West (Abzw) 9.575 10.826 0 1000000 {"coordinates": [[8.60725918287, 47.4108311022], [8.61841308539, 47.4182759904]], "type": "LineString"} 47.4145535463, 8.61283613413

9790 Wallisellen Herti - Wallisellen Belair Zurich Aubrugg (Vzw) Wallisellen Belair (Vzw) 6.841 7.800 0 1000000 {"coordinates": [[8.57180041381, 47.4131270279], [8.58448632752, 47.4133546872]], "type": "LineString"} 47.41324085755, 8.578143370665

2025-05-12 15:41:00,136 [INFO]

📊 Dataset Info:

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 431 entries, 0 to 430

Data columns (total 10 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 Line 431 non-null int64

1 Line.1 431 non-null object

2 START OPK 431 non-null object

3 END OPK 431 non-null object

4 KM START 431 non-null float64

5 KM END 431 non-null float64

6 Content-Type: text/plain; charset=utf-8

431 non-null int64

7 Content-Type: text/plain; charset=utf-8

.1 431 non-null int64

8 Content-Type: text/plain; charset=utf-8

.2 431 non-null object

9 geo\_point\_2d 431 non-null object

dtypes: float64(2), int64(3), object(5)

memory usage: 33.8+ KB

None

2025-05-12 15:41:00,142 [INFO]

🧼 Null value summary:

Line 0

Line.1 0

START OPK 0

END OPK 0

KM START 0

KM END 0

Content-Type: text/plain; charset=utf-8\n 0

Content-Type: text/plain; charset=utf-8\n.1 0

Content-Type: text/plain; charset=utf-8\n.2 0

geo\_point\_2d 0

dtype: int64

2025-05-12 15:41:00,143 [INFO]

📦 Cardinality preview (top 5 columns):

- Line: 431 unique

- Line.1: 431 unique

- START OPK: 356 unique

- END OPK: 370 unique

- KM START: 323 unique

(progress\_env) PS D:\PhD\prog\_report\_2025\_June\_project>