Pyspark formulas:

Drop syntax:returns new dataframe with dropped column

invalid\_records\_df\_drop=invalid\_records\_df.drop("data")

<https://www.geeksforgeeks.org/defining-dataframe-schema-with-structfield-and-structtype/>

defining schema to split with column

https://mungingdata.com/apache-spark/dataframe-schema-structfield-structtype/