Recap:

Dataframe reading

**Handling bad records/corrupted data :**

Mode :

1. Permissive - mismatched records are replaced by null

2. Dropmalformed ->it ignores all bad records and

allow only the good records

3. Failfast ---> throws exceptions at any single error in the source file

=====================================

**Handling null values 👍**

Na.drop or dropna

Na.fill or fillna

Na.replace or replace

Team 1 --Retail - Vishnu ,Sahil,Angela, Riya

https://github.com/akgeoinsys/Shell2023/tree/main/Sales360

team 2 -- Carbon emission - Pragati, ,Baskar, Arpit, Manisha,

https://github.com/akgeoinsys/Shell2023/tree/main/EVCarbonEmission

team 3 --covid 19- > Nirmeetha, Rohit, Sandeep, Nikhil

Batch Processing

https://github.com/akgeoinsys/ltimindtree2023/tree/main/casestudies/datasets/healthcare/worldometer\_data.csv

Spark Streaming Datasets: Latest

https://github.com/akgeoinsys/ltimindtree2023/tree/main/casestudies/datasets/healthcare/countr

y\_wise\_latest.csv

team 4 -- Worldbank ->loan => Gaurav , Shrasti ,Thirumalesh,,Mansi

Batch Processing 👍

https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-Of-Loans-Historical-Data/zucq-nrc3/about\_data

Spark Streaming Datasets: Latest

https://finances.worldbank.org/Loans-and-Credits/IBRD-Statement-of-Loans-Latest-Available-Snapshot/sfv5-tf7p/about\_data

team 5 --> Airlines : Manideep ,Likitha,,Dinesh

https://github.com/akgeoinsys/ltimindtree2023/blob/main/casestudies/datasets/airlines.csv

Airports:

https://github.com/akgeoinsys/ltimindtree2023/blob/main/casestudies/datasets/airports.csv

Streaming datasets :

Flights: (historical)

https://github.com/akgeoinsys/ltimindtree2023/blob/main/casestudies/datasets/flights\_sample.cs

v

Flights: (latest)

https://github.com/akgeoinsys/ltimindtree2023/blob/main/casestudies/datasets/flights\_latest.csv