Columnar format:

* Columnar file format is efficient way of storing the dataset examples: parquet, orc
* Row based file formats are csv, tsv etc.

1. Read optimization (query data to be pulled using the given column)
2. Compression improvements (similar entries in the columns make it easier to compress)+
3. Better suited for analytical operations

Apache iceberg

* Improves data processing on large data sets in data lake.