

Issues In Data Sets:

1. No Date column present.

Solution: --

Create Separate Date Column For DEPARTURE / ARRIVAL

2. Departure and scheduled Time column is completely messed up.

Solution: --

Create separate Date/ Time column for DEPARTURE / ARRIVAL

3. No time bucked column present.

Solution: --

Create Time Bucked for **DEPARTURE_TIME / ARRIVAL_TIME**

Extract Hours From **DEPARTURE_TIME / ARRIVAL_TIME** for
DEPARTURE_TIME_BUCKET and **ARRIVAL_TIME_BUCKET** Column.

4. The Flight Data set is too big, nearly 560MB CSV file, I tried for days to upload it to the MySQL table but it won't.

Also faced lost connection to MySQL server during query: Because of null values changed the datatypes to varchar

(Error Code: 1366. Incorrect integer value: " for column 'ARRIVAL_TIME' at row 33)

My sql server time out: Error Code: 2013.

Solution: --

I switched to PostGreSQL.

Other Possible Solution for better analysis:

Data Base Normalisation (2nf) creating a separate data table for cancelled, and delayed flights , but **no flight Id** present in the dataset.

EDA and machine learning implementation can make the analysis better.