**WORKING WITH DATASET**

**Understanding The Dataset**

|  |  |
| --- | --- |
| Date | 07 November 2022 |
| Team ID | PNT2022TMID43705 |
| Project Name | Project – AIRLINES DATA ANALYTICS FOR AVAITION INDUSTRY |
| Maximum Marks | 2 Marks |

The main dataset has been sourced from [Kaggle.](https://www.kaggle.com/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018) Additional data with Aiport metadata has also been fetched from Partow [(The Global Airport Database).](https://www.partow.net/miscellaneous/airportdatabase/) Only relevant information will be provided for the latter due to low importance in the report.

The main dataset i.e. [Airline Cancellation/Delay](https://www.kaggle.com/yuanyuwendymu/airline-delay-and-cancellation-data-2009-2018) was fetched in the form of multiple CSV files each representing data for each year. Overall, it was nearly 7 GB in size with nearly 68 million rows, comprising of the following fields/columns:

