

# SPRINT II

Date	3 November 2022
Team ID	PNT2022TMID22723
Project Name	Airlines Data Analytics for Aviation Industry
Maximum Marks	8 Marks

## Working With The Dataset

### Understanding The Dataset

A dataset is a collection of data within a database.

Typically, datasets take on a tabular format consisting of rows and columns. Each column represents a specific variable, while each row corresponds to a specific value. Some datasets consisting of unstructured data are non-tabular, meaning they don't fit the traditional row column format.

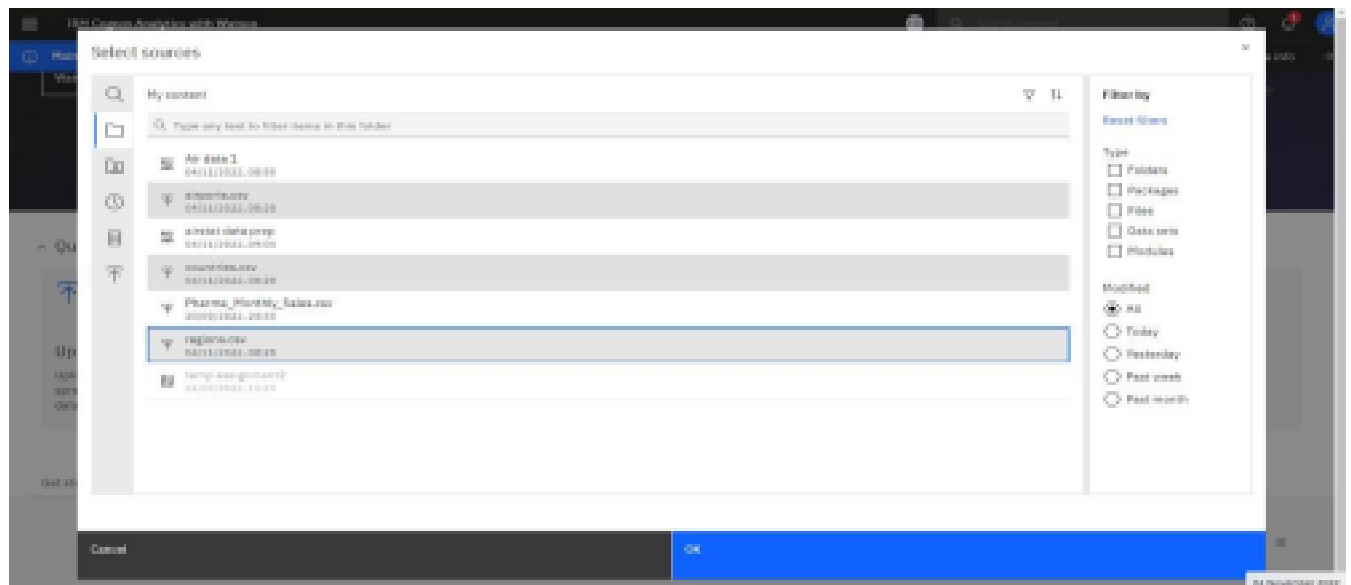
The data can be downloaded from the Links :

1. [AirStats data on airports around the world](#)
2. [Circum - Airport Performance Reports](#)
3. [Resources Coverage data](#)

## Loading of Dataset

### What is meant by load data?

Data Loading is defined as **copying data from one electronic file or database into another**. Data loading implies converting from one format into another; for example, from one type of production database into a decision support database from a different vendor.



# Data Preparation

## Data Preparation.

- Validate all the tables - airports, countries, regions
- Create calculated field - Continent Name using the codes.

Row ID	ICAO	IATA	Type	Name	latitude_deg	longitude_deg
1	H001	0001	airport	Type 00 airport	40.670000000000	-74.630000000000
2	H00001	0004	small_airport	Barro Colorado Airport	9.704000	-82.477000
3	H004	0004	small_airport	Lowell Field	19.947000	-90.180004
4	H005	0004	small_airport	Eggn Airport	14.864700449000	-86.770000000000
5	H009	0004	closed	Newport Hospital & Clinic Helipad	15.408000	-91.214000
6	000177	0006	small_airport	Fulton Airport	16.900000	-87.860000
7	H021	0002	small_airport	Corinto Airport	11.388000000000	-82.180000000000
8	H026	0002	small_airport	San Antonio (ST) Airport	16.360000	-92.380000
9	000334	0002	small_airport	Williams Egg Airport	19.427000	-92.760000
10	000666	0002	airport	Robert Creek Helipad	18.727000	-92.480000
11	H029	0003	closed	San Jose	40.600000	-74.630000
12	H001	0004	small_airport	San Jose Airport	10.400000000000	-84.200000000000
13	H001	0010	airport	Kingman Airport	35.890000000000	-93.240000000000
14	H001	0011	small_airport	San Jose Airport	10.400000000000	-84.200000000000

Row ID	ICAO	IATA	Name	continent	ICAO_code
1	000001	A0-00	Central America	EU	00
2	000002	A0-00	South America	EU	00
3	000003	A0-00	La Paz America	EU	00
4	000004	A0-00	Older America	EU	00
5	000005	A0-00	San Luis de Lima	EU	00
6	000006	A0-00	Andorra Valley	EU	00
7	000007	A0-00	San Juan de los Rios	EU	00
8	000008	A0-04	San Jose	EU	00
9	000009	A0-00	San Jose	EU	00
10	000010	A0-00	San Jose	EU	00
11	000011	A0-00	San Jose	EU	00
12	000012	A0-00	San Jose	EU	00
13	000013	A0-00	San Jose	EU	00
14	000014	A0-00	San Jose	EU	00

# Joining Of Tables

Joining of Tables Airports, Countries and Regions with the related columns.



Edit relationship

Table 1: countries.csv

Table 2: regions.csv

Match selected columns

Row Id	id	name	continent	wikipedia_link	keywords	code	iso_country	Row Id
1	302672	Andorra	ES	https://en.wikipedia.org/wiki/Andorra	NuS	#D		1
1	302672	Andorra	ES	https://en.wikipedia.org/wiki/Andorra	NuS	#D		2
1	302672	Andorra	ES	https://en.wikipedia.org/wiki/Andorra	NuS	#D		3
1	302672	Andorra	ES	https://en.wikipedia.org/wiki/Andorra	NuS	#D		4

Show join, 1-to-many No filtering

Matched columns

04 November 2020

# Exploration Of Data

Exploration of Data to understand the descriptive statistics of the data.

