

Date	22 October 2022
Team ID	PNT2022TMID40399
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.

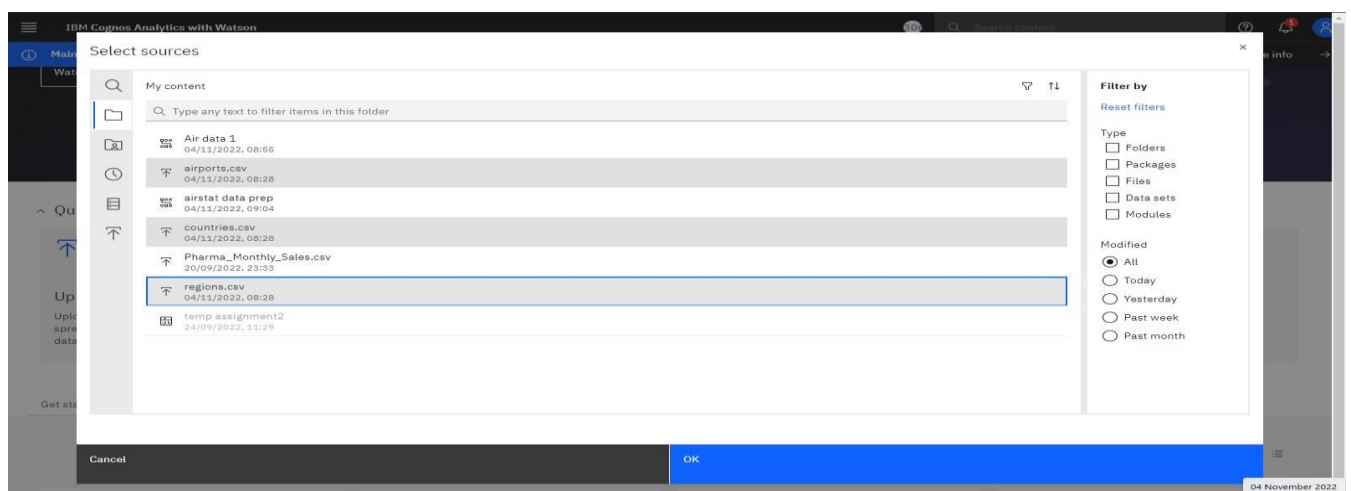
The data can be downloaded from the Links :

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.

The image displays two screenshots of the IBM Cognos Analytics interface, showing data preparation for airports and regions.

Top Screenshot: Airports Data

The interface shows the 'airports.csv' file loaded. The 'Grid' tab is active, displaying a table with the following columns: Row Id, Id, Ident, type, name, latitude_deg, and longitude_deg. The table contains 14 rows of data.

Row Id	Id	Ident	type	name	latitude_deg	longitude_deg
1	6523	00A	heliport	Total RT Heliport	40.07080078125	-74.93360137939453
2	323361	00AA	small_airport	Aero B Ranch Airport	38.704022	-101.473911
3	6524	00AK	small_airport	Lowell Field	59.947733	-151.692524
4	6525	00AL	small_airport	Epps Airpark	34.86479949951172	-86.77030181884766
5	6526	00AR	closed	Newport Hospital & Clinic Heliport	35.6087	-91.254098
6	322127	00AS	small_airport	Fulton Airport	34.9428028	-97.8180194
7	6527	00AZ	small_airport	Cordes Airport	34.305599212646484	-112.16500091552734
8	6528	00CA	small_airport	Goldstone (GTS) Airport	35.35474	-116.885329
9	324424	00CL	small_airport	Williams Ag Airport	39.427188	-121.763427
10	322658	00CN	heliport	Kitchen Creek Helibase Heliport	32.7273736	-116.4597417
11	6529	00CO	closed	Cass Field	40.622202	-104.344002
12	6531	00FA	small_airport	Grass Patch Airport	28.64550018310547	-82.21900177001953
13	6532	00FD	heliport	Ringhaver Heliport	28.848599578857422	-82.34539794921875
14	6533	00FL	small_airport	River Oak Airport	27.230899810791016	-80.96920013427734

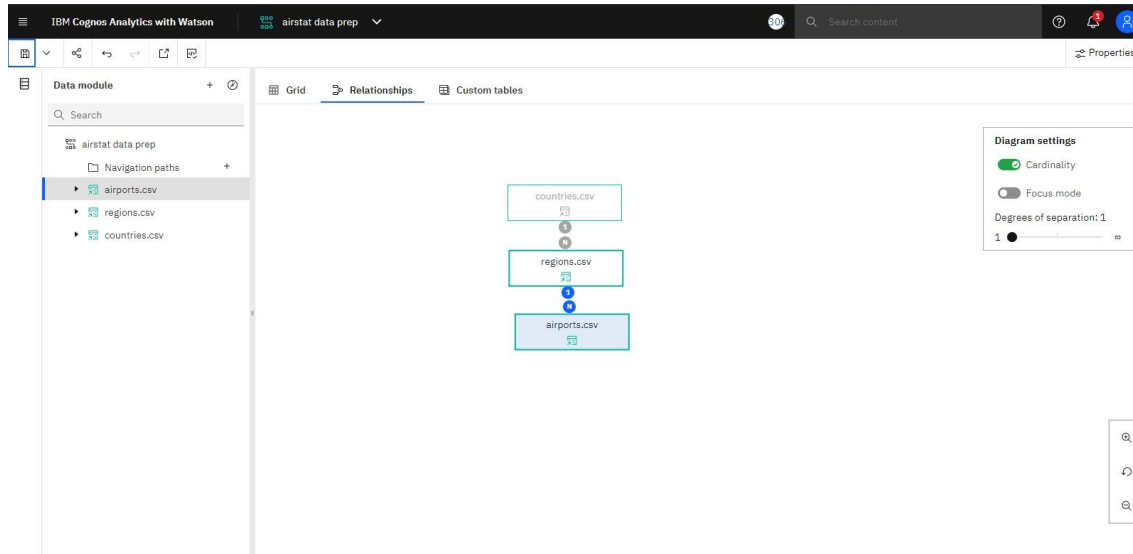
Bottom Screenshot: Regions Data

The interface shows the 'regions.csv' file loaded. The 'Grid' tab is active, displaying a table with the following columns: Row Id, Id, code, local_code, name, continent, and iso_country. The table contains 13 rows of data.

Row Id	Id	code	local_code	name	continent	iso_country
1	302811	AD-02	02	Canillo Parish	EU	AD
2	302812	AD-03	03	Encamp Parish	EU	AD
3	302813	AD-04	04	La Massana Parish	EU	AD
4	302814	AD-05	05	Ordino Parish	EU	AD
5	302815	AD-06	06	Sant Julià de Lòria Parish	EU	AD
6	302816	AD-07	07	Andorra la Vella Parish	EU	AD
7	302817	AD-08	08	Escaldes-Engordany Parish	EU	AD
8	302818	AD-U-A	U-A	(unassigned)	EU	AD
9	302819	AE-AJ	AJ	Ajman Emirate	AS	AE
10	302820	AE-AZ	AZ	Abu Dhabi Emirate	AS	AE
11	302821	AE-DU	DU	Dubai Emirate	AS	AE
12	302822	AE-FU	FU	Fujairah Emirate	AS	AE
13	302823	AE-RK	RK	Ras al-Khaimah Emirate	AS	AE

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	EU	https://en.wikipedia.org/wiki/Andorra	Null	AD		1
1	302672	Andorra	EU	https://en.wikipedia.org/wiki/Andorra	Null	AD		2
1	302672	Andorra	EU	https://en.wikipedia.org/wiki/Andorra	Null	AD		3
1	302672	Andorra	EU	https://en.wikipedia.org/wiki/Andorra	Null	AD		4

Inner join, 1-to-many
No filtering

Matched columns (1)

04 November 2022

Exploration Of Data

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

