

## Project Development Part – 1

### Product Sales Analysis Phase 3

Date	24 October 2023
Project Name	Product Sales Analysis

#### Phase 3:

- Data Collection
- Prepare the dataset
- Data Cleaning and preprocessing

#### Data Collection:

Data collection is the pivotal first step in constructing a dataset for analysis, machine learning, or any data-driven project. It involves systematically gathering information from diverse sources, which can encompass structured data from databases and spreadsheets, as well as unstructured data from text, images, and more. The process begins with clearly defining project objectives to guide the selection of appropriate data sources. Subsequently, data is extracted, transformed, and cleansed to ensure high quality and accuracy. Attention to data privacy and security is essential to safeguard sensitive information during collection. Detailed documentation of the process is maintained, and if data collection is ongoing, continuous monitoring is established to keep the dataset current and reliable. Effective data collection forms the bedrock upon which robust datasets are built, enabling organizations and individuals to derive meaningful insights from their data.

#### Download the Dataset:

<https://www.kaggle.com/datasets/ksabishek/product-sales-data>

#### Load the Dataset:

**Tool used:** IBM Cognos Analytics

## IBM Cognos Analytics:

IBM Cognos analysis is a powerful tool for extracting actionable insights from data. This platform offers a wide range of analytical features, including data visualization, reporting, and predictive analytics, to help organizations make informed decisions. With Cognos, users can explore data, create interactive dashboards, and generate comprehensive reports that provide a clear understanding of business performance. Its ability to seamlessly integrate with various data sources, along with features for data exploration, makes it a valuable asset for data-driven organizations seeking to uncover trends,

The screenshot displays the IBM Cognos Analytics web application interface. At the top, a blue header bar contains the 'IBM Cognos Analytics' logo, a notification bell with '30' alerts, a search icon, and a user profile icon. Below the header, a dark blue banner features the text 'Get quick answers with the Assistant' and a prompt to 'Ask the Assistant a question in your own words to uncover insights about your data.' A search bar with the placeholder 'Ask a question' is positioned below the banner. To the right, a user profile card for 'Viswa K (f6286c86d05d)' is shown, including the email 'kmviswa73@gmail.com' and a link to 'Profile and settings'. A dropdown menu is open, listing options such as 'My Inbox', 'My schedules and subscriptions', 'My Watch Items', 'Log my session', 'About', 'Manage product subscription', 'Privacy', 'Cookie Preferences', 'IBM Cognos Analytics Mobile', 'Get started', and 'Log out'. Below the banner, a section titled 'Introduction to Cognos Analytics' provides a brief overview of self-service analytics. At the bottom, two light blue boxes offer guidance: the first, with an upload icon, instructs users to 'Upload data and start creating content' by uploading spreadsheets, CSV files, and other types of files; the second, with a folder icon, instructs users to 'Create content from existing data' by locating data sources in the Content view.

IBM Cognos Analytics

30

1/3 Alerts

What's New: To read about what's new in Getting Started, click More

### Get quick answers with the Assistant

Ask the Assistant a question in your own words to uncover insights about your data.

Ask a question

**Viswa K (f6286c86d05d)**  
kmviswa73@gmail.com  
Profile and settings

- My Inbox
- My schedules and subscriptions
- My Watch Items
- Log my session
- About
- Manage product subscription
- Privacy
- Cookie Preferences
- IBM Cognos Analytics Mobile
- Get started
- Log out

#### Introduction to Cognos Analytics

Leverage self-service analytics to make more confident decisions.

Product tour

Upload data and start creating content

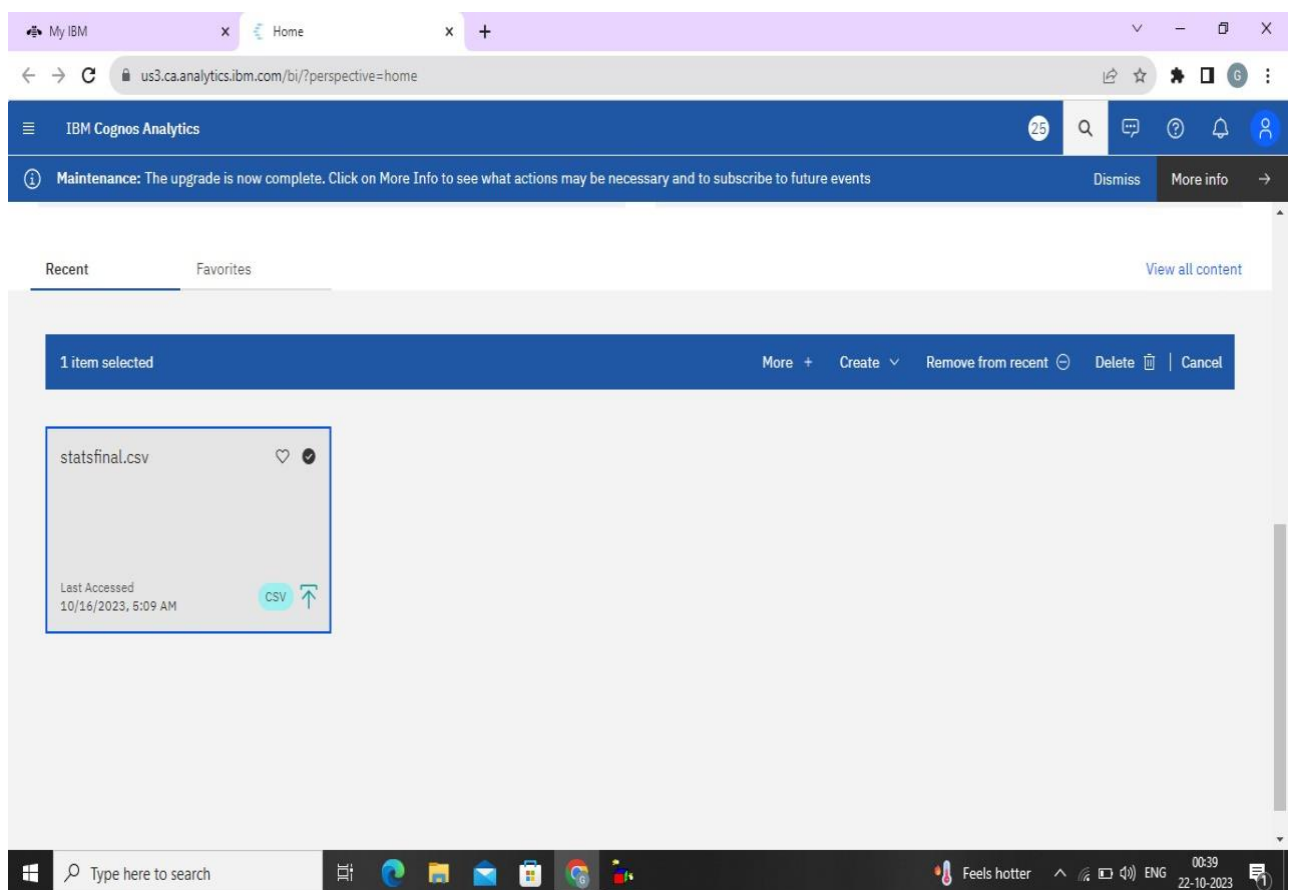
Upload spreadsheets, CSV files, and other types of files, and create content based on these files.

Create content from existing data

Locate data sources in the Content view, and create content based on these sources.

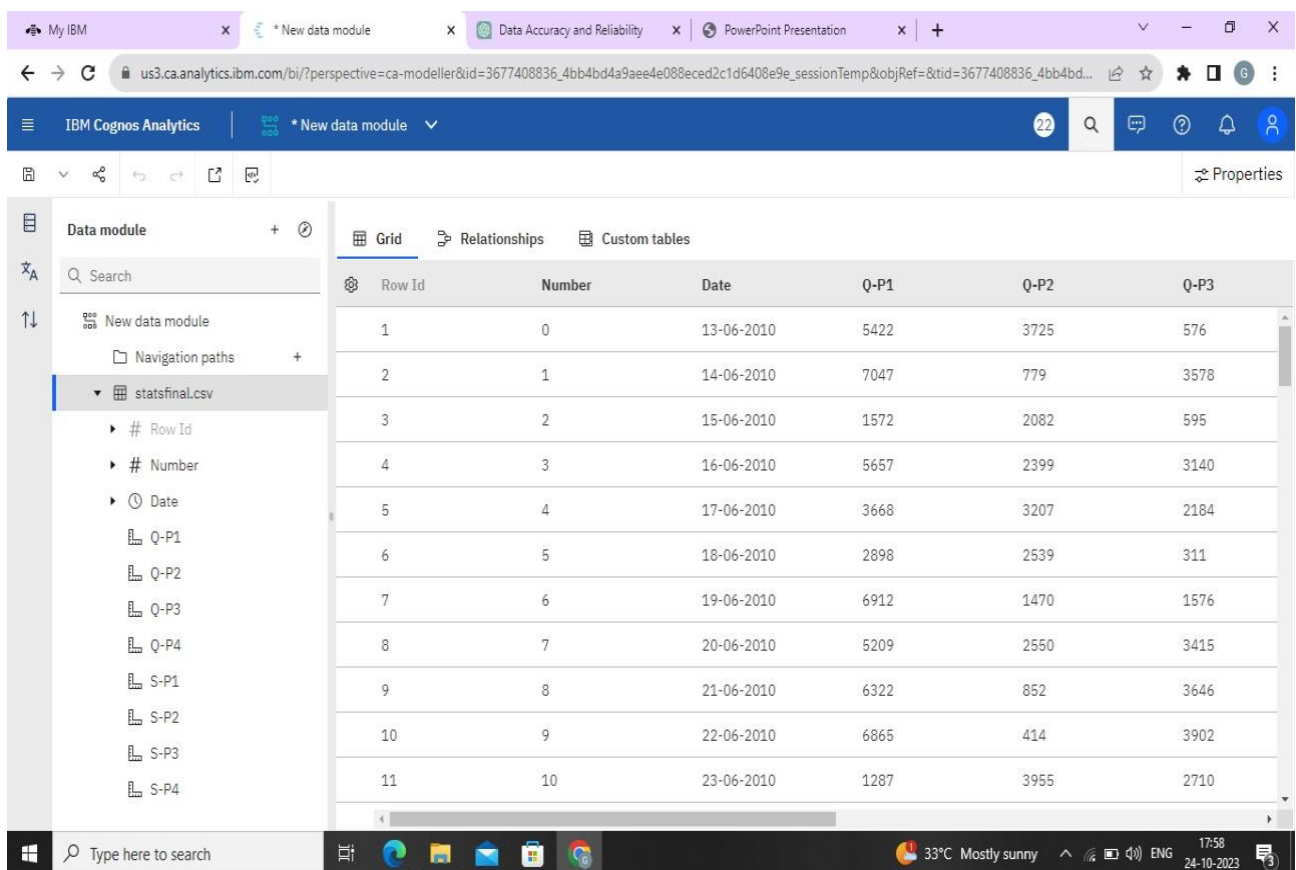
## Load the dataset:

Loading datasets in IBM Cognos analysis is a straightforward yet crucial step for data-driven insights. This process involves importing and connecting data from various sources, including databases, spreadsheets, and cloud services, into the Cognos environment. Users can then prepare, transform, and clean the data to create a unified, structured dataset suitable for analysis. By providing a seamless integration with multiple data sources, Cognos streamlines the process, ensuring that data is readily available for reporting, visualization, and advanced analytics. This efficient dataset loading feature empowers organizations to harness the full potential of their data for informed decisionmaking and actionable insights.



## Data Preparation:

Data preparation in IBM Cognos analysis is a crucial phase where raw data is refined into a usable and valuable asset. This process involves tasks such as cleansing, transformation, and structuring of data to ensure accuracy and consistency. With Cognos, users can easily reshape data, handle missing values, and create calculated fields, making it ready for analysis. Data preparation in Cognos enables organizations to work with cleaner, more meaningful datasets, thus improving the quality and reliability of insights and reports generated within the platform. It's a vital step that empowers users to uncover patterns, trends, and actionable information from their data with confidence.

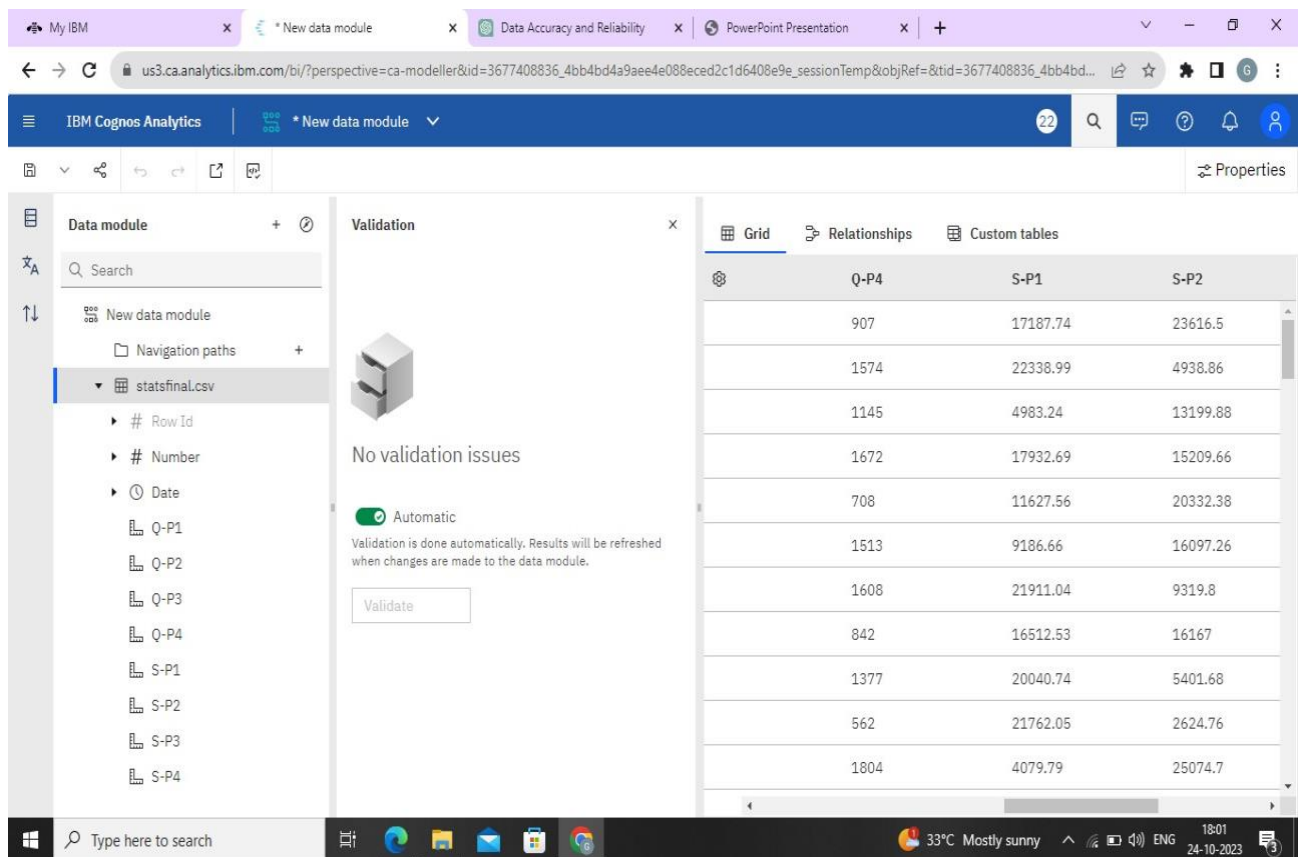


The screenshot displays the IBM Cognos Analytics web interface. The top navigation bar includes the IBM logo and several open tabs: 'My IBM', '\* New data module', 'Data Accuracy and Reliability', and 'PowerPoint Presentation'. The main header shows 'IBM Cognos Analytics' and a search bar. On the left, a sidebar lists the 'Data module' and a search bar. Below it, a tree view shows the file structure, including 'statsfinal.csv' and various calculated fields like '# Row Id', '# Number', 'Date', 'Q-P1', 'Q-P2', 'Q-P3', 'Q-P4', 'S-P1', 'S-P2', 'S-P3', and 'S-P4'. The main content area displays a table with 11 rows and 7 columns. The columns are 'Row Id', 'Number', 'Date', 'Q-P1', 'Q-P2', and 'Q-P3'. The table contains numerical data for each row, representing a time series from June 13, 2010, to June 23, 2010.

Row Id	Number	Date	Q-P1	Q-P2	Q-P3
1	0	13-06-2010	5422	3725	576
2	1	14-06-2010	7047	779	3578
3	2	15-06-2010	1572	2082	595
4	3	16-06-2010	5657	2399	3140
5	4	17-06-2010	3668	3207	2184
6	5	18-06-2010	2898	2539	311
7	6	19-06-2010	6912	1470	1576
8	7	20-06-2010	5209	2550	3415
9	8	21-06-2010	6322	852	3646
10	9	22-06-2010	6865	414	3902
11	10	23-06-2010	1287	3955	2710

## Data Cleaning:

Data cleaning in IBM Cognos analysis is the process of refining and enhancing the quality of your data. It involves identifying and rectifying errors, inconsistencies, and inaccuracies within your datasets. With Cognos, users can detect and address issues such as duplicate records, missing values, outliers, and data discrepancies. By employing various data cleaning techniques, Cognos ensures that your data is accurate, consistent, and reliable. Clean data is the foundation for producing trustworthy reports and insights, ultimately enabling more informed decision-making and analysis within the Cognos platform.



The screenshot shows the IBM Cognos Analytics web interface. On the left, a navigation pane lists the data module 'statsfinal.csv' and its columns: Row Id, Number, Date, and a list of products (Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, S-P4). The main area is divided into two sections: 'Validation' and 'Grid'. The 'Validation' section shows a message 'No validation issues' and a 'Validate' button. The 'Grid' section displays a table with three columns: Q-P4, S-P1, and S-P2. The table contains 10 rows of data.

Q-P4	S-P1	S-P2
907	17187.74	23616.5
1574	22338.99	4938.86
1145	4983.24	13199.88
1672	17932.69	15209.66
708	11627.56	20332.38
1513	9186.66	16097.26
1608	21911.04	9319.8
842	16512.53	16167
1377	20040.74	5401.68
562	21762.05	2624.76
1804	4079.79	25074.7

## Product Sales Analysis data:

### About Dataset

REC corp LTD. is small-scaled business venture established in India.

- They have been selling FOUR PRODUCTS for OVER TEN YEARS.
- The products are P1, P2, P3 and P4.

- They have collected data from their retail centers and organized it into a small csv file.

The excel file contains about 8 numerical parameters:

- Q1- Total unit sales of product 1 • Q2- Total unit sales of product 2
- Q3- Total unit sales of product 3 • Q4- Total unit sales of product 4
- S1- Total revenue from product 1 • S2- Total revenue from product 2
- S3- Total revenue from product 3
- S4- Total revenue from product 4

The screenshot shows the IBM Cognos Analytics interface. On the left, a 'Data module' sidebar lists fields from a file named 'statsfinal.csv': Row Id, Number, Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, and S-P4. The main area displays a 'Grid' view of the data. The grid has 11 rows and 6 columns. The columns are labeled 'Row Id', 'Number', 'Date', 'Q-P1', 'Q-P2', and 'Q-P3'. The data shows a sequence of unit sales for three products (Q-P1, Q-P2, Q-P3) over a period of 11 days in June 2010.

Row Id	Number	Date	Q-P1	Q-P2	Q-P3
1	0	13-06-2010	5422	3725	576
2	1	14-06-2010	7047	779	3578
3	2	15-06-2010	1572	2082	595
4	3	16-06-2010	5657	2399	3140
5	4	17-06-2010	3668	3207	2184
6	5	18-06-2010	2898	2539	311
7	6	19-06-2010	6912	1470	1576
8	7	20-06-2010	5209	2550	3415
9	8	21-06-2010	6322	852	3646
10	9	22-06-2010	6865	414	3902
11	10	23-06-2010	1287	3955	2710

My IBM

New data module

Data Accuracy and Reliability

PowerPoint Presentation

us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=3677408836\_4bb4bd4a9ae4e088eced2c1d6408e9e\_sessionTemp&objRef=&tid=3677408836\_4bb4bd...

IBM Cognos Analytics

New data module

22

Q

?

?

?

?

Properties

Grid

Relationships

Custom tables

	Q-P4	S-P1	S-P2	S-P3	S-P4
	950	6178.33	12540.52	22541.78	6773.5
	976	9265.91	18855.16	4249.28	6958.88
	1535	6691.87	11025.26	5685.58	10944.55
	1088	10077.43	23115.64	21723.36	7757.44
	486	21090.01	1616.7	3002.68	3465.18
	1189	12705.36	13719.76	10617.78	8477.57
	1493	22031.5	14112.84	31620.28	10645.09
	1777	4070.28	8159.58	17734.24	12670.01
	1850	3008.33	3487	24736.88	13190.5
	384	17666.41	15862.68	23669.14	2737.92
	1209	22326.31	4520.42	2883.44	8620.17

Data module

+

?

Search

New data module

Navigation paths

statsfinal.csv

# Row Id

# Number

Date

Q-P1

Q-P2

Q-P3

Q-P4

S-P1

S-P2

S-P3

S-P4

Type here to search

33°C Mostly sunny

18:02

24-10-2023