

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

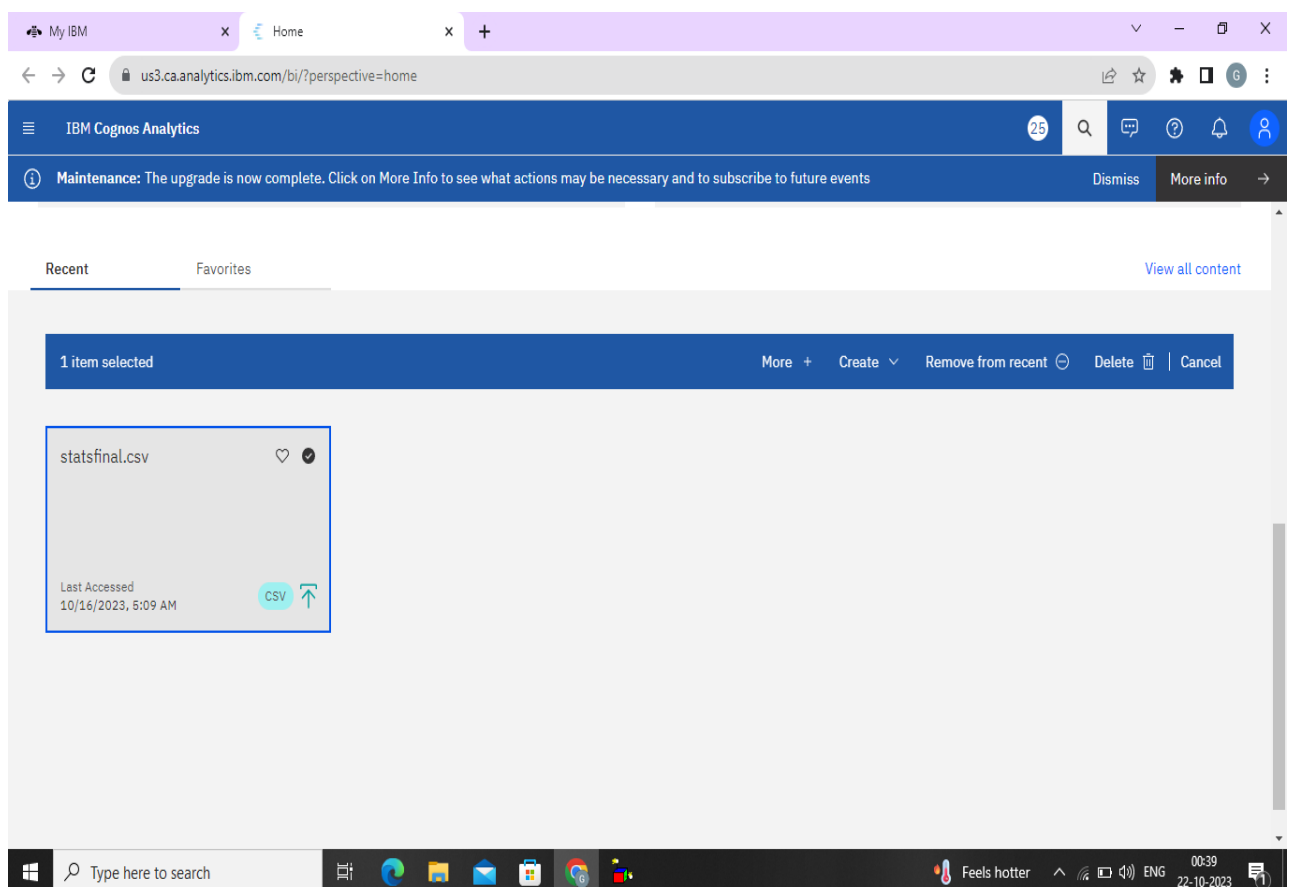
IBM Cognos Analytics:

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, patterns, and opportunities in their data.

The screenshot displays the IBM Cognos Analytics web application running in a browser. The browser's address bar shows the URL `us3.ca.analytics.ibm.com/bj/?perspective=home`. The application's header is dark blue with the 'IBM Cognos Analytics' logo on the left and user profile information on the right. A maintenance banner at the top states: 'Maintenance: The upgrade is now complete. Click on More Info to see what actions may be necessary and to subscribe to future events'. The main content area features a large dark blue section titled 'Get quick answers with the Assistant', which includes a text input field labeled 'Ask a question' and a blue speech bubble icon. Below this is a light blue section titled 'Introduction to Cognos Analytics' with the text 'Leverage self-service analytics to make more confident decisions.' and two buttons: 'Get started' and 'Watch videos'. On the right side, a user profile dropdown menu is open, showing the user's name 'hari haran (fe25b7d9d88...)', email 'todayuse2345@gmail.com', and a 'Profile and settings' link. The menu also lists 'My Inbox', 'My schedules and subscriptions', 'My Watch Items', 'Log my session', 'About', 'Manage product subscription', 'Privacy', 'Cookie Preferences', 'IBM Cognos Analytics Mobile', and 'Log out'. The Windows taskbar at the bottom shows the search bar, several application icons, and system status information including '32°C Mostly cloudy' and the date '22-10-2023'.

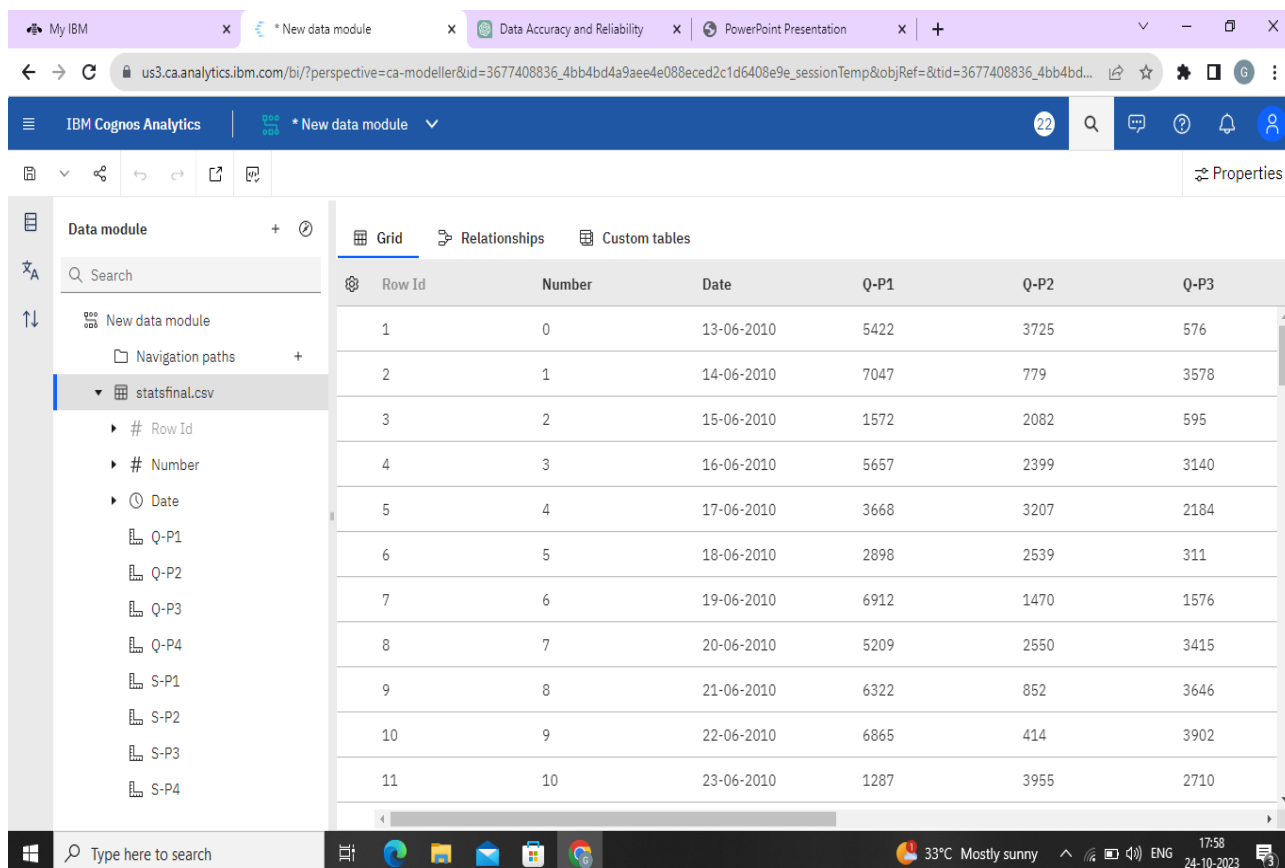
Load the dataset:

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 decision-making 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 main area shows a data table with the following columns: Row Id, Number, Date, Q-P1, Q-P2, and Q-P3. The table contains 11 rows of data. The left sidebar shows the 'Data module' and 'statsfinal.csv' file. The top navigation bar includes 'IBM Cognos Analytics' and 'New data module'.

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 displays the IBM Cognos Analytics web application interface. The browser tabs at the top include 'My IBM', 'New data module', 'Data Accuracy and Reliability', and 'PowerPoint Presentation'. The URL bar shows a path to a Cognos perspective. The application header includes the 'IBM Cognos Analytics' logo and a search bar. The main workspace is divided into three panels:

- Data module:** A sidebar on the left showing a tree view of data modules. The 'statsfinal.csv' module is selected, revealing a list of fields: Row Id, Number, Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, and S-P4.
- Validation:** A central panel displaying 'No validation issues'. It features a toggle for 'Automatic' validation, which is currently turned on. A 'Validate' button is visible at the bottom of the panel.
- Grid:** A table on the right showing data from the 'statsfinal.csv' module. The table has four columns: 'Q-P4', 'S-P1', and 'S-P2'. The data is organized into rows, with the first row containing values 907, 17187.74, and 23616.5, respectively.

The bottom of the screen shows a Windows taskbar with various application icons and a system tray displaying the temperature (33°C), weather (Mostly sunny), and time (18:01, 24-10-2023).

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 web interface. The browser tabs include 'My IBM', 'New data module', 'Data Accuracy and Reliability', and 'PowerPoint Presentation'. The URL is 'us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=3677408836_4bb4bd4a9aee4e088eced2c1d6408e9e_sessionTemp&objRef=&tid=3677408836_4bb4bd...'. The interface has a top navigation bar with 'IBM Cognos Analytics' and a 'New data module' dropdown. Below this is a toolbar with icons for search, navigation, and properties. The main content area is divided into a left sidebar and a central grid. The sidebar shows a 'Data module' tree with 'statsfinal.csv' selected, listing fields: Row Id, Number, Date, Q-P1, Q-P2, Q-P3, Q-P4, S-P1, S-P2, S-P3, and S-P4. The central grid displays a table with 11 rows and 7 columns: Row Id, Number, Date, Q-P1, Q-P2, Q-P3, and Q-P4. The data is as follows:

Row Id	Number	Date	Q-P1	Q-P2	Q-P3	Q-P4
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	

The bottom of the screen shows a Windows taskbar with the search bar, taskbar icons, and system tray showing 33°C, Mostly sunny, and the date 24-10-2023.

My IBM

* New data module

Data Accuracy and Reliability

PowerPoint Presentation

us3.ca.analytics.ibm.com/bi/?perspective=ca-modeller&id=3677408836_4bb4bd4a9aee4e088eced2c1d6408e9e_sessionTemp&objRef=&tid=3677408836_4bb4bd...

IBM Cognos Analytics

New data module

22

Properties

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

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

Type here to search

33°C Mostly sunny 18:02 24-10-2023