

Project Development Phase

Sprint 2

Date	03 November 2022
Team ID	PNT2022TMID41225
Project Name	Global Sales Data Analytics

Sprint 2 tasks:

- Connecting data from IBM cloud to IBM cognos
- Connecting schema and table
- Data visualization
- Prepare the dataset
 - Data Joins
 - Data calculations
 - Data navigation
 - Data formatting
- Creating Dashboard

Connecting data from IBM cloud to IBM cognos:

The screenshot displays the IBM Cognos Analytics with Watson web application. The browser's address bar shows the URL `us1.ca.analytics.ibm.com/bi/?perspective=home`. The page header includes the title "IBM Cognos Analytics with Watson", a search bar, and user profile icons. A blue banner at the top of the main content area contains a maintenance notice: "Maintenance: Scheduled maintenance completed. Click More Info for details and to subscribe to future events". On the left side, a sidebar menu is open, showing "Data server connections". Below this, a table lists the connected data sources:

Name	Modified
Global sales	11/14/2022 5:54 AM
Weather Company	3/31/2022 8:44 PM

The main content area features a large header image with the text "Analytics with" and a diagram illustrating data flow. Below the header, there are three prominent tiles: "data" (with subtext "modules to clean it data from"), "Exploration" (with subtext "Quickly find unbiased answers by identifying trends"), and "Present data" (with subtext "Create sophisticated, multi-page, multi-query").

Connecting schema and table:

The screenshot shows the IBM Cognos Analytics web interface. The left sidebar displays a list of connections, with 'Global sales' selected. The main panel shows the 'Global sales' connection details, including the owner (DHAR ... 92b706), creation/modification dates, and type (Data Server). The 'Connections' tab is active, showing a table of connections:

Name	Modified
Global sales	11/14/2022 7:25 AM

The 'Settings' tab is also visible, showing the 'Authentication method' section with options: 'Connect anonymously' (selected), 'Prompt for the user ID and password', 'Use an external namespace', and 'Use the following signon:'. A 'Test' button shows a 'Success' status. A 'Save' button is at the bottom.

Data visualization:

The screenshot shows a Jupyter Notebook titled 'data_process.ipynb' in Visual Studio Code. The notebook is running on a Jupyter Server (Local). The code in the notebook is as follows:

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn import preprocessing
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
```

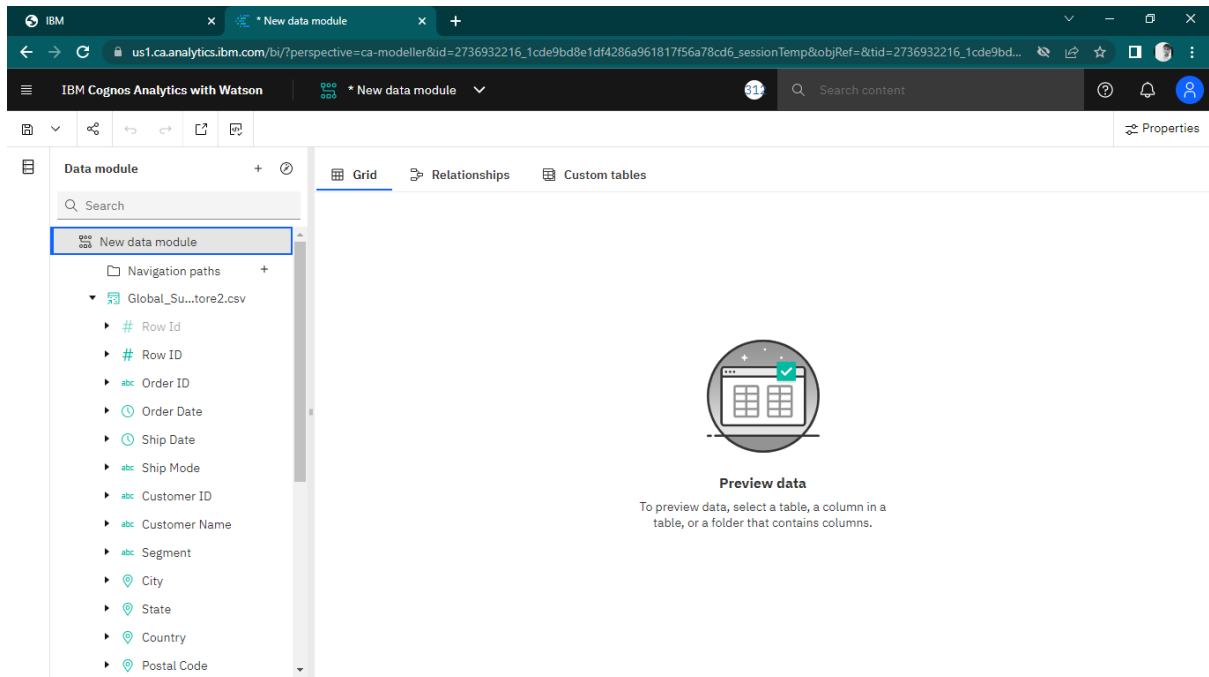
The output of the code shows the 'data' variable, which is a DataFrame with the following columns: Row ID, Order ID, Order Date, Ship Date, Ship Mode, Customer ID, Customer Name, Segment, City, State, Product ID, Category, and Sub-Category.

Prepare the dataset:

Understanding rows and columns of the data.

a. Data Joins:

Analysing that the dataset module requires joins.

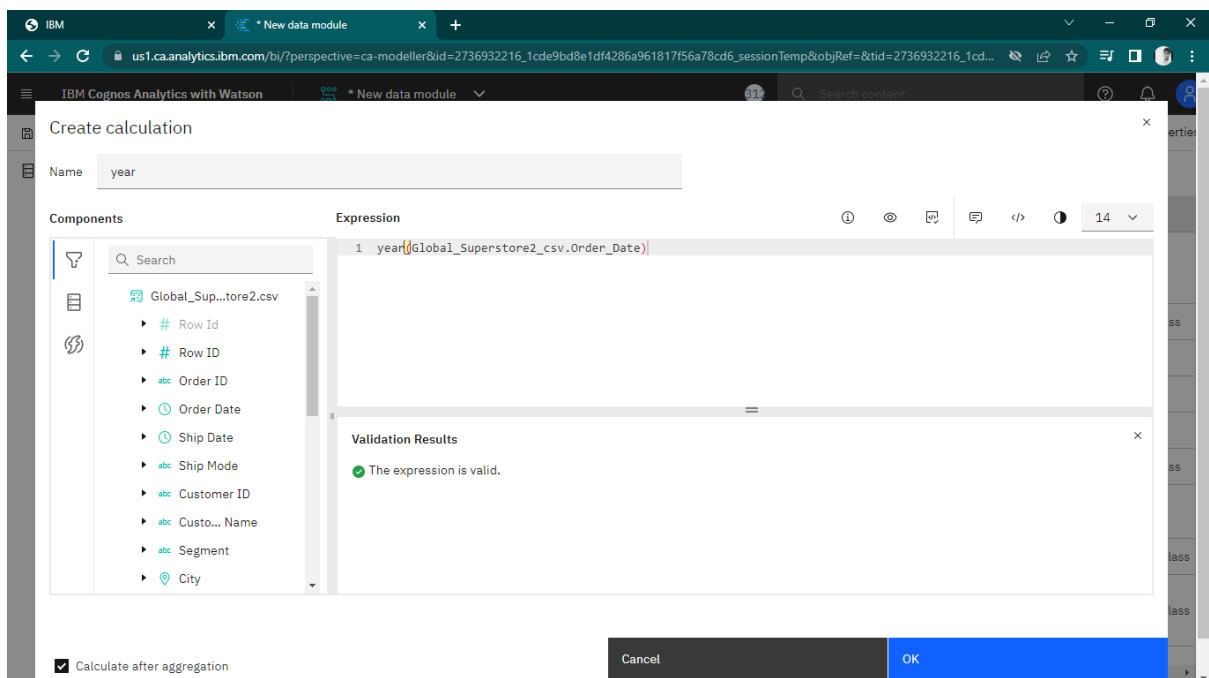


b. Data Calculations:

Creating calculations for extracting year, month and day from order date column.

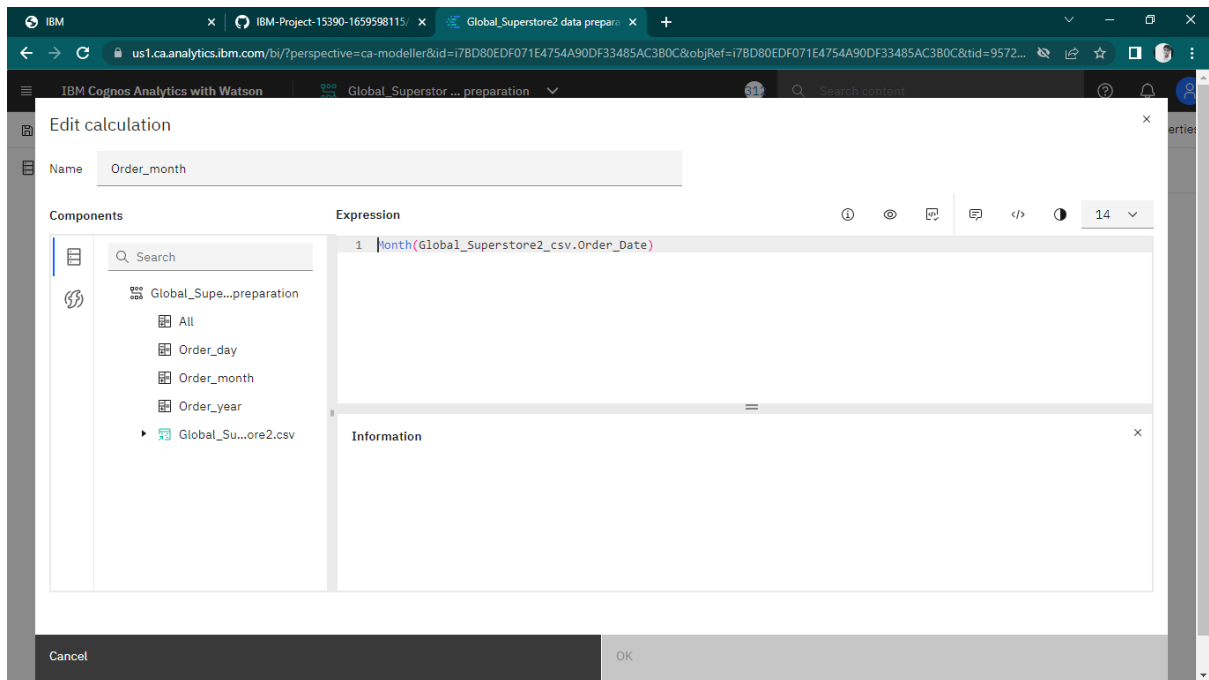
1. Year calculation

Creating calculations for extracting year from order date column.



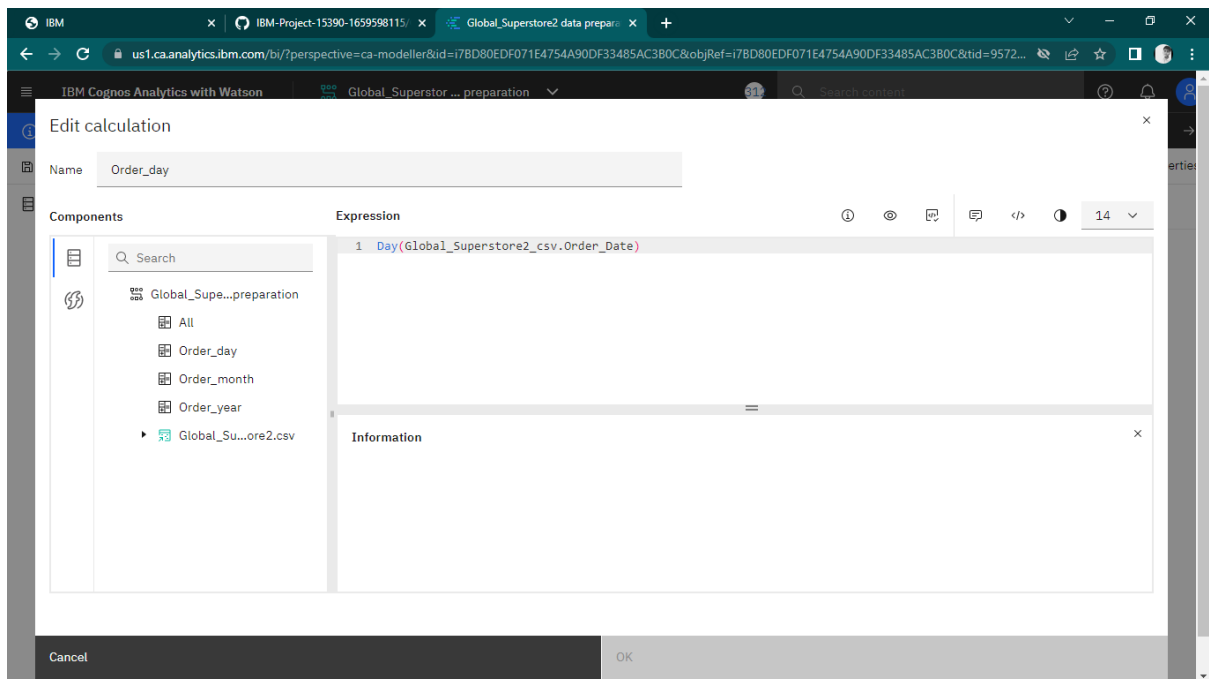
2. Month calculation

Creating calculations for extracting month from order date column.



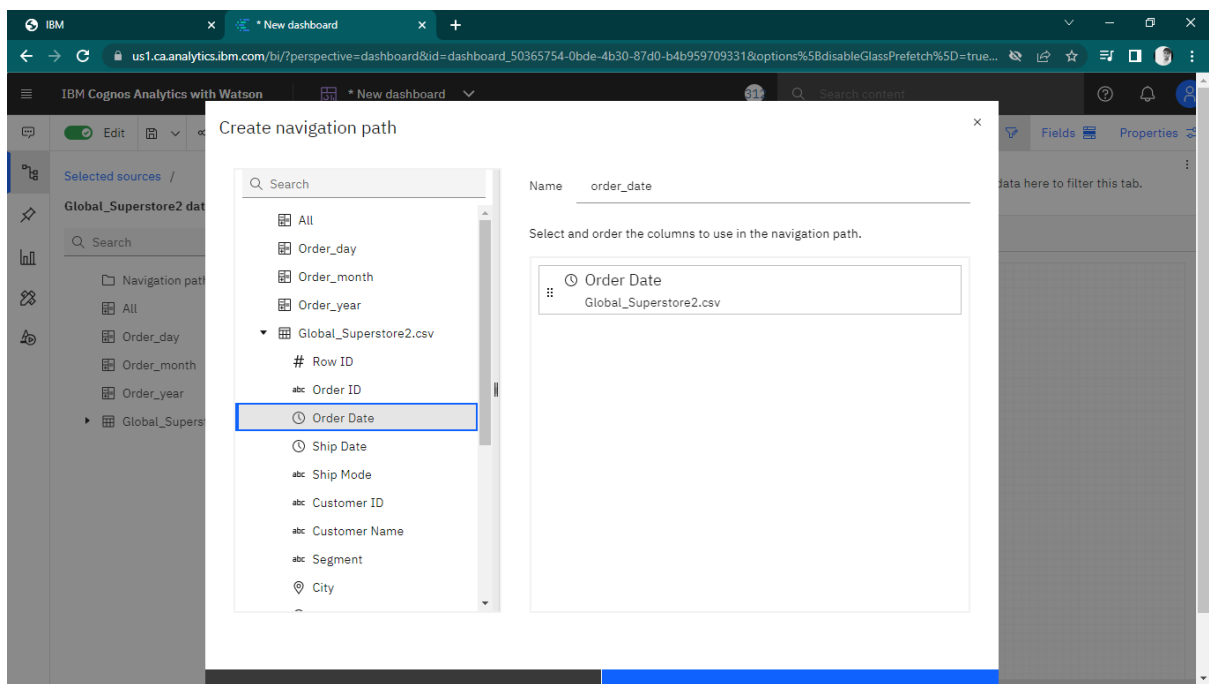
3. Day calculation

Creating calculations for extracting day from order date column.



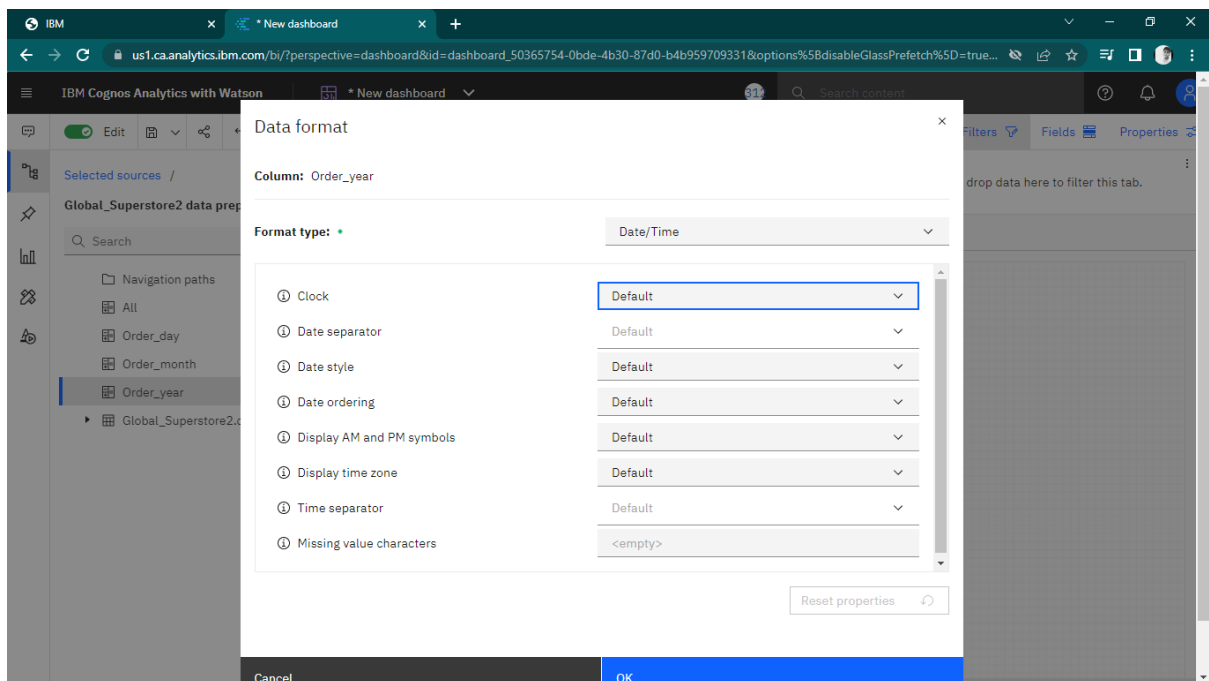
c. Navigation path:

Creating navigation path for order date, order year, order month and order day



d. Data formatting:

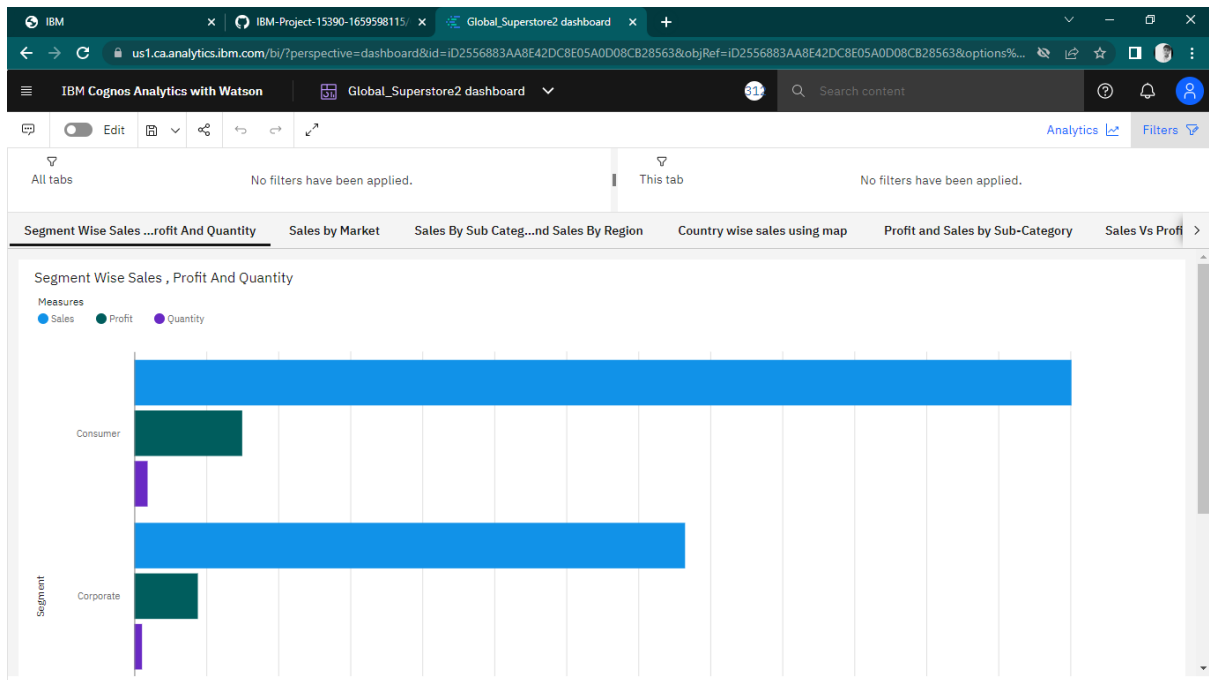
Creating navigation path for order date, order year, order month and order day.



Creating Dashboard:

Segment wise sales, Profit and Quantity:

The following bar chart shows the distribution of Sales, Profit and Quantity by segments.



Sales by Market:

The below Pie visualization represents the Sales distribution by market.

