

‘ Major Port Traffic & Capacity ’

(using IBM Cognos Analytics)

A PROJECT REPORT

Submitted by

Utkarsh Srivastava | Reg.No: 19BCE10383

B.tech | (CSE (core))

at

Vellore Institute of Technology , Bhopal University

in partial fulfillment for the requirement of the award for

TRAINING

in

‘ Data Analytics ’

SMART INTERNZ In Collaboration with **IBM**

May 2022

under the guidance of

Mrs. Uma Rani

ACKNOWLEDGEMENT

First and foremost, I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to **Smart Internz** for so much of his valuable support encouragement for carrying out this work . I would like to thank my guide **Mrs. Uma Rani** ,

for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

Last, but not the least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

PROJECT DESCRIPTION:

The Indian Railways has a capital base of about Rs. 100000 crores and is often referred to as the lifeline of the Indian economy because of its predominance in transportation of bulk freight and long distance passenger traffic. The network criss-crosses the nation, binding it together by ferrying freight and passengers across the length and breadth of the country. As the Indian economy moves into a high growth trajectory the Railways have also stepped-up developmental efforts and are preparing themselves for an even bigger role in the future.

TECHNOLOGY & CONCEPTS involved :

Data Analytics :

Data analytics (DA) is the process of examining data sets in order to find trends and draw conclusions about the information they contain. Increasingly, data analytics is done with the aid of specialized systems and software. Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions. Scientists and researchers also use analytics tools to verify or disprove scientific models, theories and hypotheses. As a term, data analytics predominantly refers to an assortment of applications, from basic business intelligence (BI), reporting and online analytical processing (OLAP) to various forms of advanced analytics.

Data analytics initiatives can help businesses increase revenue, improve operational efficiency, optimize marketing campaigns and bolster customer service efforts. Analytics also enable organizations to respond quickly to emerging market trends and gain a competitive edge over business rivals. The ultimate goal of data analytics, however, is boosting business performance.

Four main types of data analytics

1. Predictive data analytics

Predictive analytics may be the most commonly used category of data analytics. Businesses use predictive analytics to identify trends, correlations, and causation. The category can be further broken down into predictive modeling and statistical modelling.

2. Prescriptive data analytics

Prescriptive analytics is where AI and big data combine to help predict outcomes and identify what actions to take. This category of analytics can be further broken down into optimization and random testing. Using advancements in ML, prescriptive analytics can help answer questions such as “What if we try this?” and “What is the best action?” We can test the correct variables and even suggest new variables that offer a higher chance of generating a positive outcome.

3. Diagnostic data analytics

Diagnostic data analytics is the process of examining data to understand cause and event or why something happened. Techniques such as drill down, data discovery, data mining, and correlations are often employed. Diagnostic data analytics help answer why something occurred.

Like the other categories, it too is broken down into two more specific categories:

- discover and alerts ,
- query and drill downs

Query and drill downs : used to get more detail from a report ,

Discover and alerts : notify of a potential issue before it occurs

4. Descriptive data analytics

Descriptive analytics are the backbone of reporting—it's impossible to have business intelligence (BI) tools and dashboards without it. It addresses basic questions of “how many, when, where, and what.”

PROBLEM STATEMENT:

Because of its dominance in the transportation of bulk freight and long-distance passenger traffic, the Indian Railways is sometimes referred to as the lifeline of the Indian economy, with a capital base of almost Rs. 100000 crores. The network crisscrosses the country, connecting it by ferrying freight and passengers across the country's length and width. As the Indian economy continues to grow at a rapid pace, the railways have increased their development efforts and are preparing for a larger role in the future.

As a data analyst we need to :

- To regain some of the market, it has lost over past decades and regain market share in some commodities and overcome the challenges and to maintain sustainable growth in all its commodities.
- Reducing the congestion on rail corridors and improving port connectivity.
- The development of two Dedicated Freight Corridors across key ports

REQUIREMENTS:

Hardware Requirements :

- Operating system- Windows 7,8,10,11
- Processor- dual core 2.4 GHz (i5 or i7 series Intel processor or equivalent AMD)

- RAM-4GB

Software Requirements :

- an active official mail-id
- IBM Cloud Account

OBJECTIVE:

As a data analyst we need to :

- To regain some of the market, it has lost over past decades and regain market share in some commodities and overcome the challenges and to maintain sustainable growth in all its commodities.
- Reducing the congestion on rail corridors and improving port connectivity.
- The development of two Dedicated Freight Corridors across key ports

PROCEDURE :

To accomplish the objective, we have to follow the activities and tasks listed below:

- creating IBM Cloud Account
- Login to Cognos Analytics
- Working with the Dataset
- Understanding the Dataset
- Loading the Dataset
- Data Prep - Conversion of Data and Calculations
- Do required calculations in Dataset
- Finally , create Data Visualizations Charts
 - ✓ Port-wise Traffic Distribution, Port wise Capacity Distribution
 - ✓ Port-wise Traffic vs Capacity by Line Chart
 - ✓ Port-wise Traffic Projected vs Achieved by Column Chart
 - ✓ Port-wise Traffic Projected vs Achieve by Stacked Column Chart
 - ✓ Port-wise Total Capacity Projects vs Total Capacity Achieve by Line and Bar Chart
 - ✓ Port-wise Traffic Projects vs Total Projected by Area Chart
 - ✓ Port-wise Total Capacity Achieve, Traffic Achieved using Stacked Bar
 - ✓ Filters
 - ✓ Port-wise Total Capacity Achieved using Map
 - ✓ Summary Cards and Visual using Total Capacity vs Actual Capacity Column Chart

Step 1:

Firstly, we need to create an IBM Cloud Account

Harness the power of IBM. Get easy no-charge access to the tools you need to develop the next great thing. Enjoy powerful technical and strategic resources from IBM. Jump right in with access to powerful services and the most prominent open-source computer technologies, or take advantage of hands-on resources that will teach you about data science, artificial intelligence, security and more.

Already registered? Log in Register now

Click >> Log-in

Enter your academic institution issued email to begin

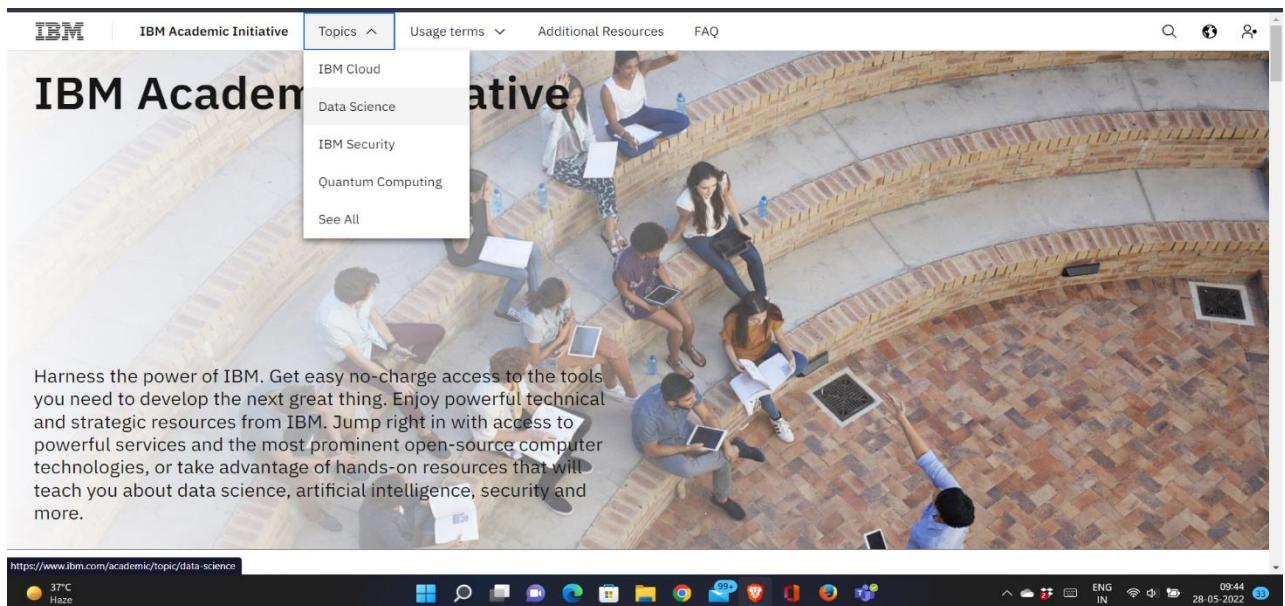
Only the students and faculty of participating academic institutions are eligible to access this website. Please enter your academic institution issued email below to register.

Your academic institution issued email
utkarsh.srivastava2019@vitbhopal.ac.in

Find answers in our [frequently asked questions](#)

Submit

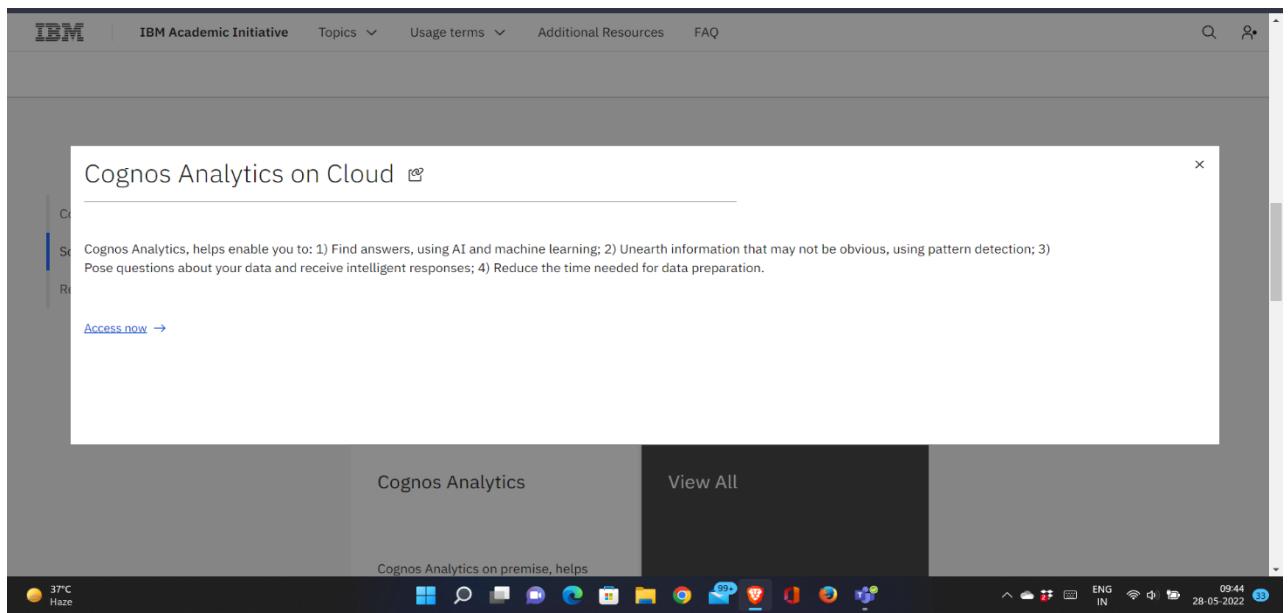
Logged-in via academic mail-id



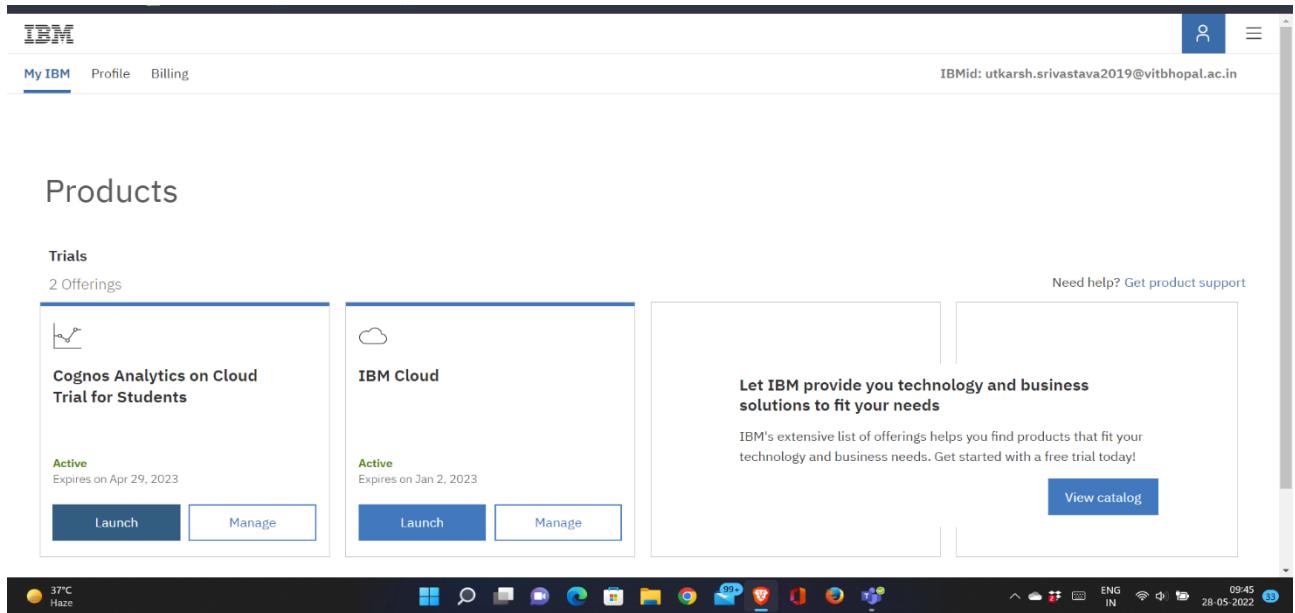
Click >> Data Science

The screenshot shows the 'Software' section of the IBM Academic Initiative website. On the left, a sidebar lists 'Courseware', 'Software', and 'Resources'. The main content area displays four software products in a grid: 'ILOG CPLEX Optimization Studio', 'SPSS Modeler Premium', 'Cognos Analytics on Cloud', and 'Cognos Analytics'. Each product has a brief description and a blue arrow pointing to the right. The 'Cognos Analytics on Cloud' section is highlighted with a grey background. Below the grid, a note says 'Cognos Analytics on premise, helps'. The URL in the address bar is https://www.ibm.com/academic/software. The taskbar at the bottom shows various application icons and system status.

Click >> Software >> Cognos Analytics on Cloud



Further, Login >> Cognos Analytics on Cloud >> Access now (when already account exist)



Screen of IBM Cognos Analytics on Cloud , which offers 2 product as a free trial of 1-year, on academic mail-id(s) :

- ✓ Cognos Analytics on Cloud Trial for Students
- ✓ IBM Cloud

Launch >> Cognos Analytics on Cloud Trial for Students

Working with the Dataset

Before we build a view and do analysis of our data, we must first connect the data to IBM Cognos. Cognos supports connecting to a wide variety of data, stored in a variety of places.

The data can be stored on our computer in a spreadsheet or a text file, or in a big data, relational, or cube (multidimensional) database on a server in our enterprise.

In my case, I used spreadsheet or text file for making my analysis

Understanding the Dataset

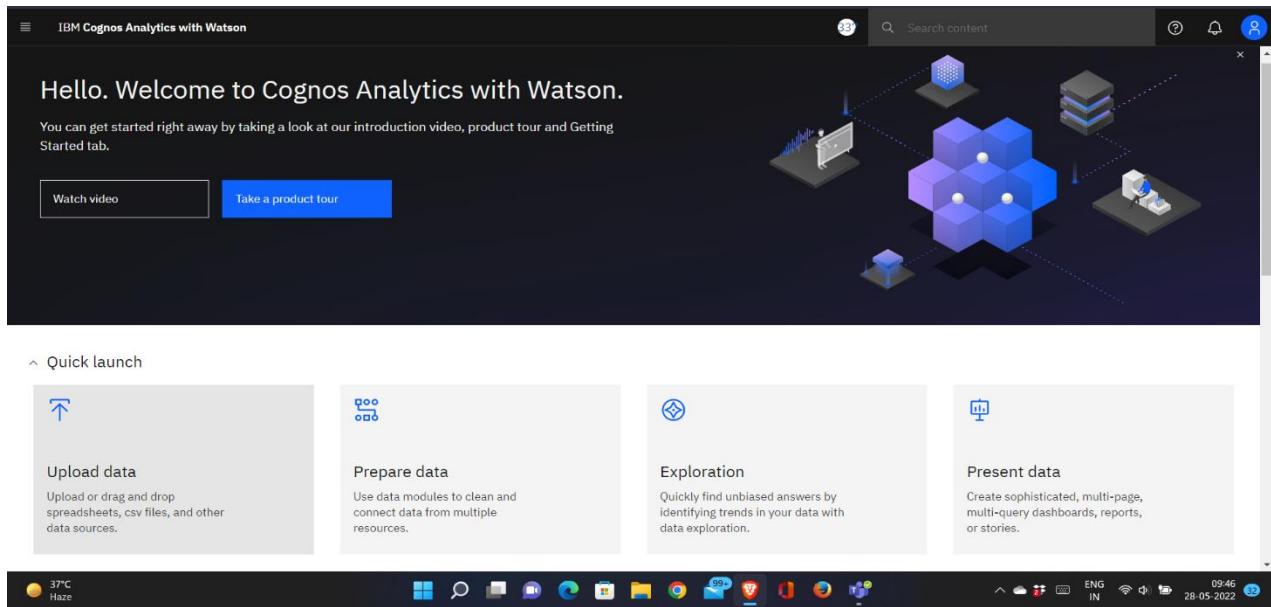
This dataset consists information about used ‘Major Port : Traffic and its Capacity’, this dataset is provided by my guide , which I have uploaded to my drive :

<https://docs.google.com/spreadsheets/d/1QBjS3uWs5WF6cDGkFVjeXNOd3VRUNEC/edit?usp=sharing&ouid=100033017801034948022&rtpof=true&sd=true>

It has 7 columns each column consists information about ‘Major Port : Traffic and its Capacity’ like **Traffic in Eleventh Plan (MT) (2011-12)Proj.** : traffic projected in 11th plan of project , i.e; 2011-12 session , **Traffic in Eleventh Plan (MT) (2011-12) Ach.** : traffic achieved in 11th plan of project , i.e; 2011-12 session , **Traffic in Eleventh Plan (MT) (2011-12) %** : traffic percentage in 11th plan of project , i.e; 2011-12 session , **Total Capacity in Eleventh Plan (MT) (2011-12) Proj.** : total capacity projected in 11th plan of project , i.e; 2011-12 session, **Total Capacity in Eleventh Plan (MT) (2011-12) Ach.** : total capacity achieved in 11th plan of project , i.e; 2011-12 session **Total Capacity in Eleventh Plan (MT) (2011-12) %** : total capacity percentage in 11th plan of project , i.e; 2011-12 session & **Port** : name of the ports .

Step 2:

Loading the Dataset



Click >> Upload Data

Step 3:

Data Prep – Conversion of Data and Calculations

Now, we need prepare dataset as per our convenience , i.e; we need to rename the ‘existing column name’ to ‘simple column name’ , as :

Existing Column Name	Rename to
Traffic in Eleventh Plan (MT) (2011-12)Proj.	Traffic Projection
Traffic in Eleventh Plan (MT) (2011-12) Ach.	Traffic Achievement
Traffic in Eleventh Plan (MT) (2011-12) %	Traffic Percentage
Total Capacity in Eleventh Plan (MT) (2011-12) Proj.	Total Capacity Projection
Total Capacity in Eleventh Plan (MT) (2011-12) Ach.	Total Capacity Achievement
Total Capacity in Eleventh Plan (MT) (2011-12) %	Total Capacity Percentage

IBM Cognos Analytics with Watson

* dataset_mpvc.xlsx

Properties

Grid Relationships Custom tables

Row Id	Port	Traffic in E...11-12)Proj.	Traffic in E...11-12) Ach.	Traffic in ... (2011-12) %	Total Capac...1-12) Proj.	Total Capac...1-12) Ach.	Total C
1	Kolkata	1343	1223	9100	3145	1635	5100
2	Haldia	4450	3101	7000	6340	5070	7900
3	Paradeep	7640	5425	7100	10640	7650	7100
4	Visakhapatnam	8220	6742	8200	10810	7293	6700
5	Ennore	4700	1496	3200	6420	3100	4800
6	Chennai	5750	5571	9700	7230	7972	11000
7	Tuticorin	3172	2810	8900	6398	3334	5200
8	Cochin	3817	2010	5300	5475	4098	7400
9	NMPT	4881	3294	6800	6050	5097	8400
10	Mormugao	4455	3900	8800	6690	4190	6200
11	Mumbai	7105	5618	7900	9191	4453	4800
12	JNPT	6604	6575	10000	9560	6400	6600
13	Kandla	8672	8250	9500	12220	8691	7100

31°C Haze

02:30 28-05-2022

Uploaded Dataset

IBM Cognos Analytics with Watson

* dataset_mpvc.xlsx

Properties

Grid Relationships Custom tables

Row Id	Port	Traffic in E...11-12)Proj.	Traffic in E...11-12) Ach.	Traffic in ... (2011-12) %
1	Kolkata	1343	1223	9100
2	Haldia	4450	3101	7000
3	Paradeep	7640	5425	7100
4	Visakhapatnam	8220	6742	8200
5	Ennore	4700	1496	3200
6	Chennai	5750	5571	9700
7	Tuticorin	3172	2810	8900
8	Cochin	3817	2010	5300
9	NMPT	4881	3294	6800
10	Mormugao	4455	3900	8800
11	Mumbai	7105	5618	7900
12	JNPT	6604	6575	10000
13	Kandla	8672	8250	9500

31°C Haze

02:30 28-05-2022

Existing Column Name : displayed in Screenshot

The screenshot shows the IBM Cognos Analytics with Watson interface. The top navigation bar includes 'IBM Cognos Analytics with Watson', a file menu, and a search bar. The main area is titled 'Data module' and shows a tree view of datasets: 'dataset_mptc.xlsx' and 'dataset.xlsx'. Under 'dataset.xlsx', there are columns: '# Row Id', 'Port', 'Traffic Projection', 'Traffic Achievement', and 'Traffic Percentage'. A preview grid displays 13 rows of data corresponding to these columns. The bottom status bar shows weather information (31°C Haze), system icons, and the date/time (28-05-2022, 02:30).

# Row Id	Port	Traffic Projection	Traffic Achievement	Traffic Percentage
1	Kolkata	1343	1223	9100
2	Haldia	4450	3101	7000
3	Paradeep	7640	5425	7100
4	Visakhapatnam	8220	6742	8200
5	Ennore	4700	1496	3200
6	Chennai	5750	5571	9700
7	Tuticorin	3172	2810	8900
8	Cochin	3817	2010	5300
9	NMPT	4881	3294	6800
10	Mormugao	4455	3900	8800
11	Mumbai	7105	5618	7900
12	JNPT	6604	6575	10000
13	Kandla	8672	8250	9500

New column name : as per tabular image

Step 4 :

Do required calculations in Dataset

we need to do following calculations :

- ❖ Traffic Percentage as
Traffic Achievement / Traffic Projection
- ❖ Total Capacity Percentage as
Total Capacity Achievement / Total Capacity Projection

The screenshot shows the IBM Cognos Analytics with Watson interface. On the left, the 'Data module' pane lists 'dataset_mptc.xlsx' with its contents: 'dataset.xlsx' (Navigation paths, Row Id, Port, Traffic Projection, Traffic Achievement, Traffic in Eleventh Plan (MT) (2011-12) %, Total Capacity Projection, Total Capacity Achievement, Total Capacity in Eleventh Plan (MT) (2011-12) %). A context menu is open over a row in the 'Traffic Achievement' table, with 'Create calculation...' highlighted.

	Traffic Projection	Traffic Achievement	Traffic in ... (2011-12) %	Total Capac... Projection	Total Capac... chievement
1343	1223	9100	3145	1635	
	3101	7000	6340	5070	
	5425	7100	10640	7650	
	6742	8200	10810	7293	
	1496	3200	6420	3100	
	5571	9700	7230	7972	
	2810	8900	6398	3334	
	2010	5300	5475	4098	
	3294	6800	6050	5097	
	3900	8800	6690	4190	
	5618	7900	9191	4453	
	6575	10000	9560	6400	
	8250	9500	12220	8691	

Click >> create calculation

The screenshot shows the 'Create calculation' dialog box. The 'Name' field is set to 'Traffic Percentage'. The 'Components' pane shows 'dataset.xlsx' (Row Id, Port, Traffic Projection, Traffic Achievement, Traffic in Eleventh Plan (MT) (2011-12) %, Total Capacity Projection, Total Capacity Achievement, Total Capacity in Eleventh Plan (MT) (2011-12) %). The 'Expression' field contains the formula 'Traffic in Eleventh Plan (MT) (2011-12) % / Traffic in Eleventh Plan (MT) (2011-12) Proj'. The 'OK' button is visible at the bottom right.

Calculated column : Traffic Percentage

- ❖ Traffic Percentage as
Traffic Achievement / Traffic Projection

Similarly ,

The screenshot shows the 'Create calculation' dialog in IBM Cognos Analytics with Watson. The 'Name' field contains 'Total Capacity Percentage'. The 'Components' pane lists items from 'dataset.xlsx' such as 'Traffic Percentage', '# Row Id', 'Port', 'Traffic Projection', 'Traffic Achievement', 'Total Capacity Projection', 'Total Capacity Achievement', and 'Total Cap...11-12) %'. The 'Expression' pane contains the formula: `1 Total_Capacity_in_Eleventh_Plan_MT_2011_12_Ach / Total_Capacity_in_Eleventh_Plan_MT_2011_12_Proj]`. The 'Information' pane provides details about the `total` function. The bottom right of the dialog has 'Cancel' and 'OK' buttons.

Calculated column : Total Capacity Percentage

- ❖ Total Capacity Percentage as
Total Capacity Achievement / Total Capacity Projection

The screenshot shows the Data module interface in IBM Cognos Analytics with Watson. The left sidebar shows a tree structure with 'dataset_mptc.xlsx' expanded, revealing 'Navigation paths', 'dataset.xlsx', 'Total Capacity Percentage', and 'Traffic Percentage'. The 'Traffic Percentage' node is selected. The main area displays a grid of data with columns: 'Total Capa...Percentage' (highlighted), 'Traffic Percentage', 'Row Id', 'Port', and 'Traffic Projection'. The data rows list various port names and their corresponding values. The bottom right of the screen shows a Windows taskbar with icons for various applications and system status.

New Column(s) added :

Traffic Percentage & Total Capacity Percentage

IBM Cognos Analytics with Watson

* dataset_mptc.xlsx

Properties

Data module

dataset_mptc.xlsx

- Navigation paths
- dataset.xlsx
 - Total Capacity Percentage
 - Traffic Percentage

Row Id, Port, Traffic Projection, Traffic Achievement, Total Capacity Projection, Total Capacity Achievement

Grid Relationships Custom tables

Traffic Percentage

Row Id	Port	Traffic Projection
461	Kolkata	1343
7508	Haldia	4450
0151	Paradeep	7640
4236	Visakhapatnam	8220
70716	Ennore	4700
462	Chennai	5750
4554	Tuticorin	3172
9315	Cochin	3817
9752	NMPT	4881
0478	Mormugao	4455
4845	Mumbai	7105
5067	JNPT	6604
2356	Kandla	8672

Filter... Create calculation... Create data group... Edit calculation... Hide from users Remove Refresh properties... Format data... Clean... Rename Cut Copy Properties

IBM Cognos Analytics with Watson

* dataset_mptc.xlsx

Properties

Data format

Column: Traffic Percentage

Format type:

Unformatted, Percent

Percent

Currency, Date, Date/Time, Time, Time interval, Custom

By default, this data is unformatted.
You can select a different format type and specify its properties.

Advanced options, Reset properties

Cancel, OK

Port, Traffic Projection

Port	Traffic Projection
Kolkata	1343
Haldia	4450
Paradeep	7640
Visakhapatnam	8220
Ennore	4700
Chennai	5750
Tuticorin	3172
Cochin	3817
NMPT	4881
Mormugao	4455
Mumbai	7105
JNPT	6604
Kandla	8672

IBM Cognos Analytics with Watson

* dataset_mptc.xlsx

Properties

Data format

Column: Traffic Percentage

Format type: Percent

Number of decimal places, Use thousands separator, Missing value characters

Default, 0, 1, 2, 3, 4, 5

Advanced options, Reset properties

Cancel, OK

Port, Traffic Projection

Port	Traffic Projection
Kolkata	1343
Haldia	4450
Paradeep	7640
Visakhapatnam	8220
Ennore	4700
Chennai	5750
Tuticorin	3172
Cochin	3817
NMPT	4881
Mormugao	4455
Mumbai	7105
JNPT	6604
Kandla	8672

Since , new columns : Traffic Percentage & Total Capacity Percentage,

values are not in proper % format so , above screenshot shows as:

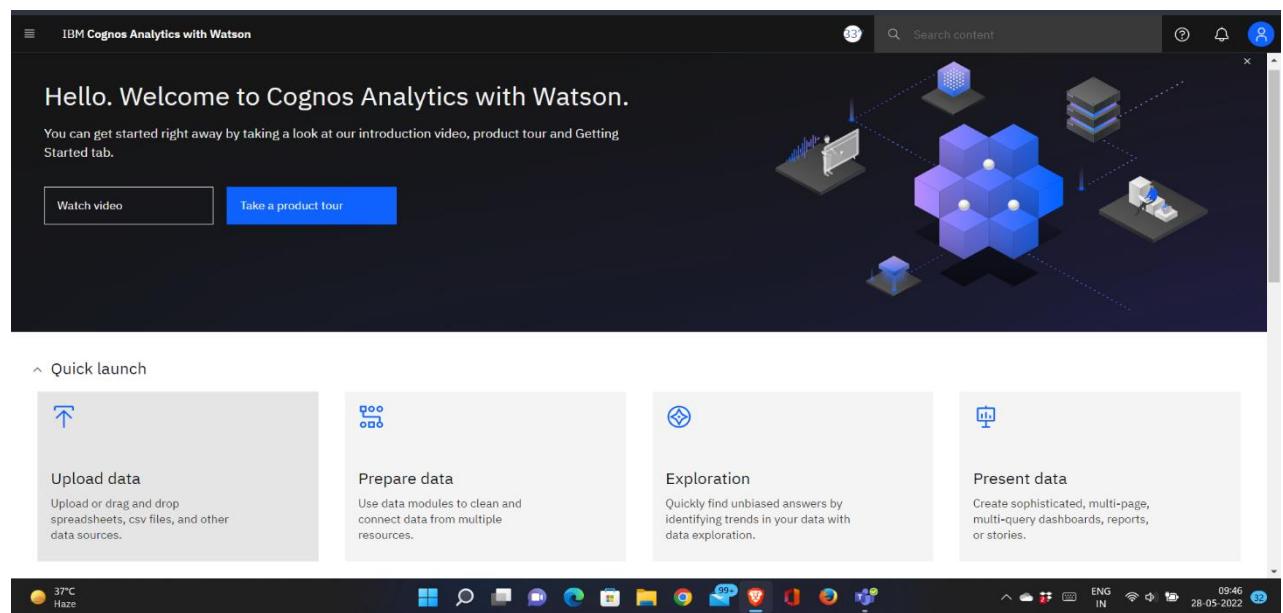
Click >> Format Data >> Format Type >> % >> Number of decimal place >> 2

Final Step :

Create Data Visualizations Charts

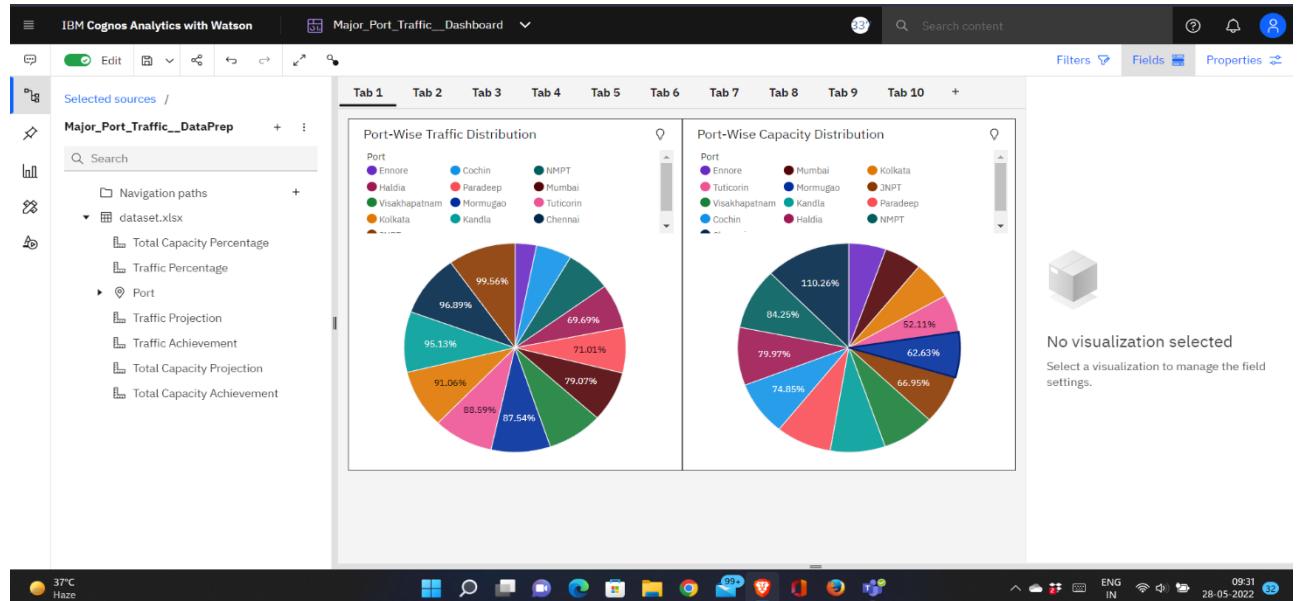
Data Visualization

Data visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data.

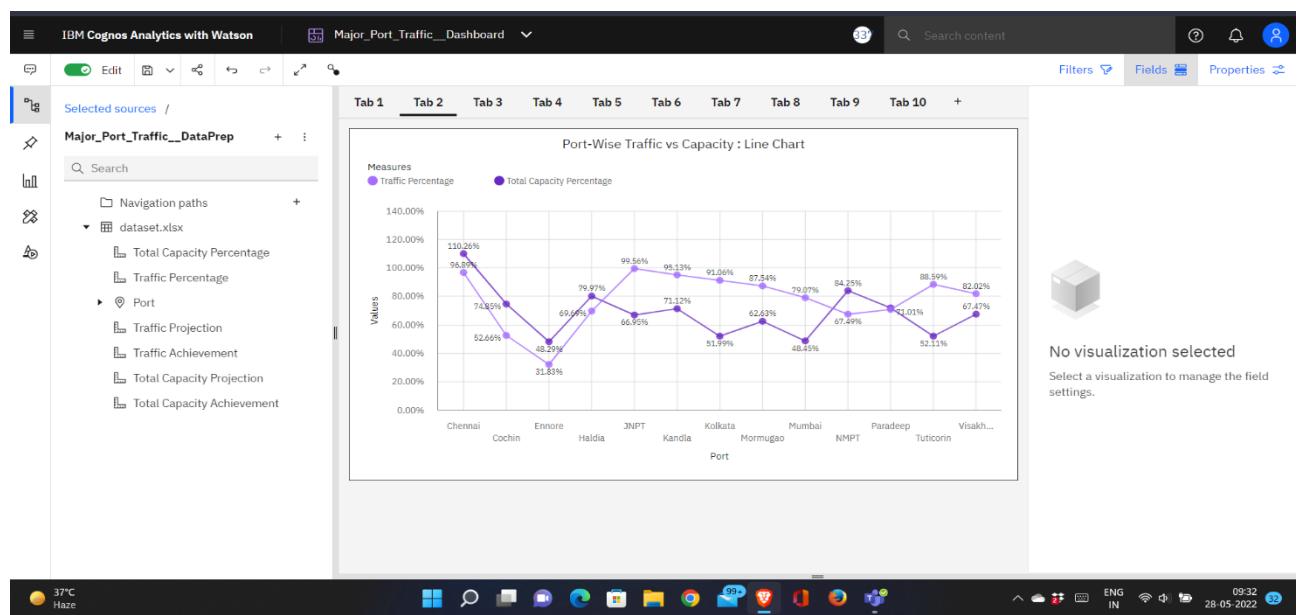


Click >> Present Data >> Dashboard

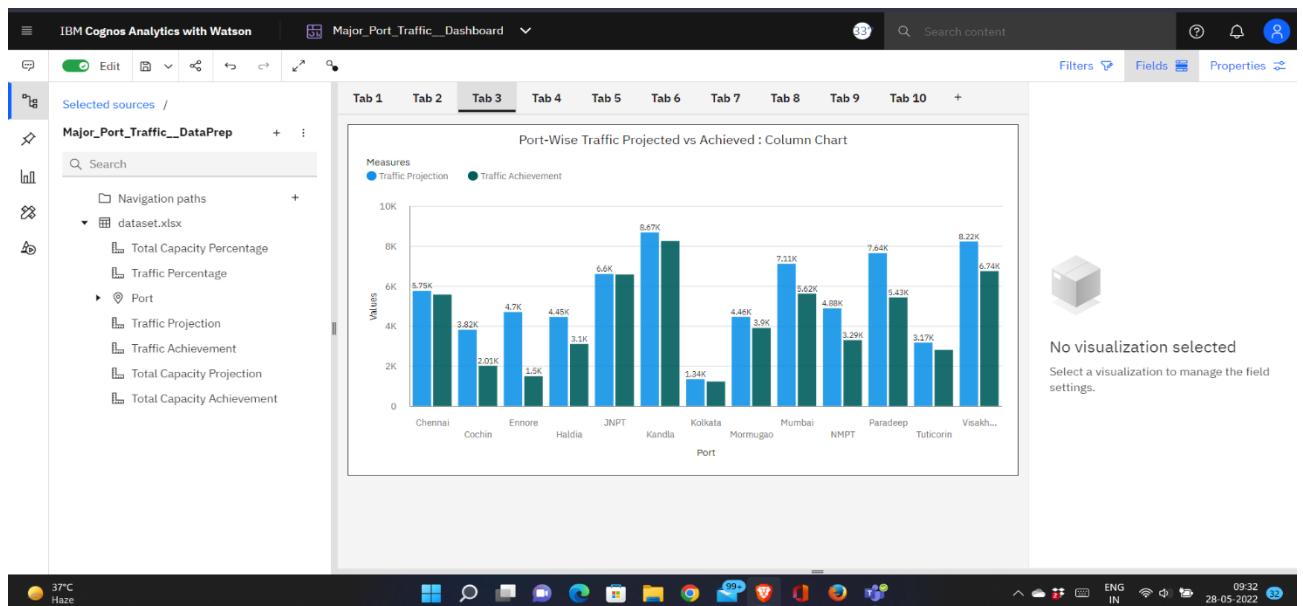
Now, lets create various graphs and charts to highlight the insights and visualizations , for our Data_Prep .



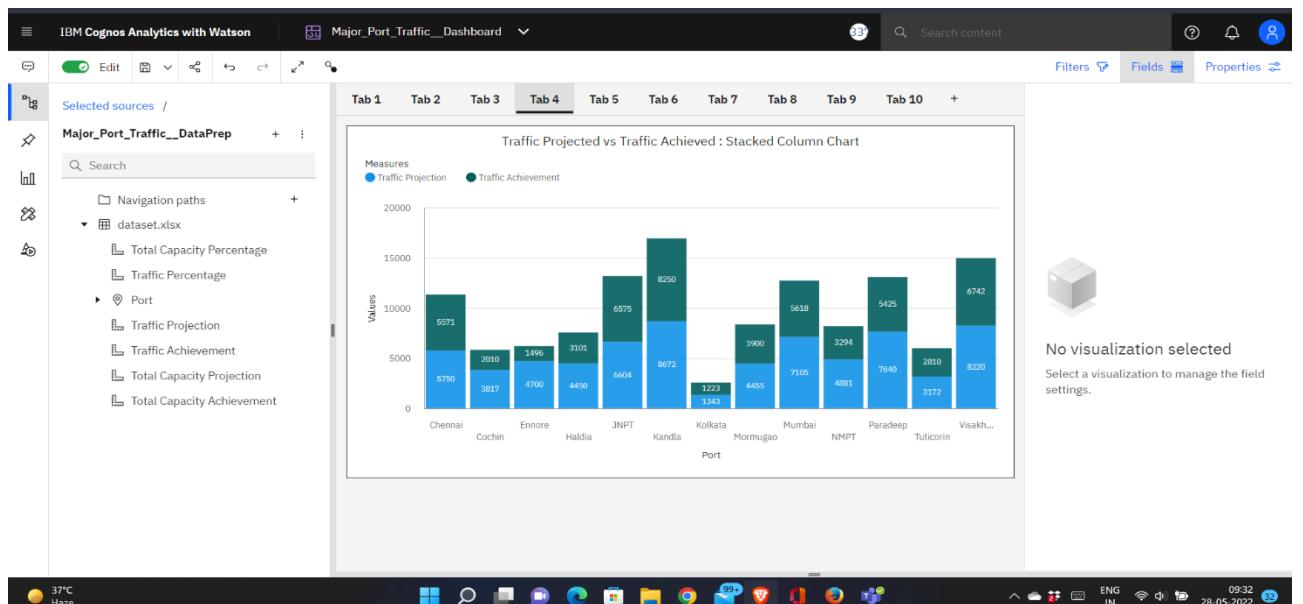
Port-wise Traffic Distribution, Port wise Capacity Distribution



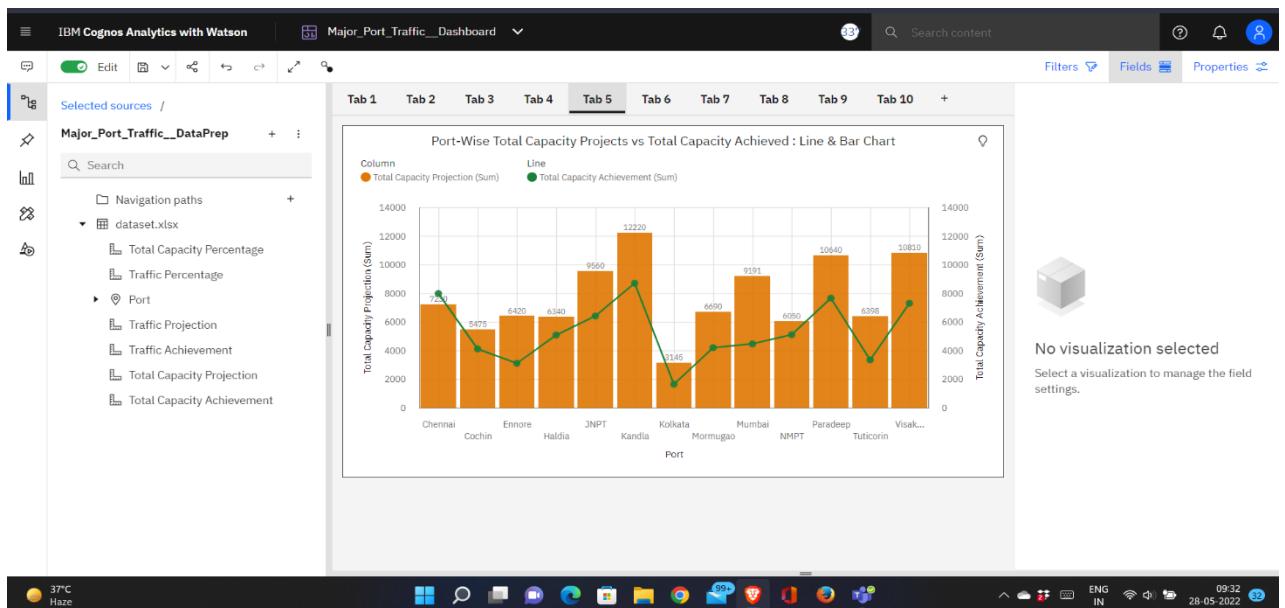
Port-wise Traffic vs Capacity by 'Line Chart'



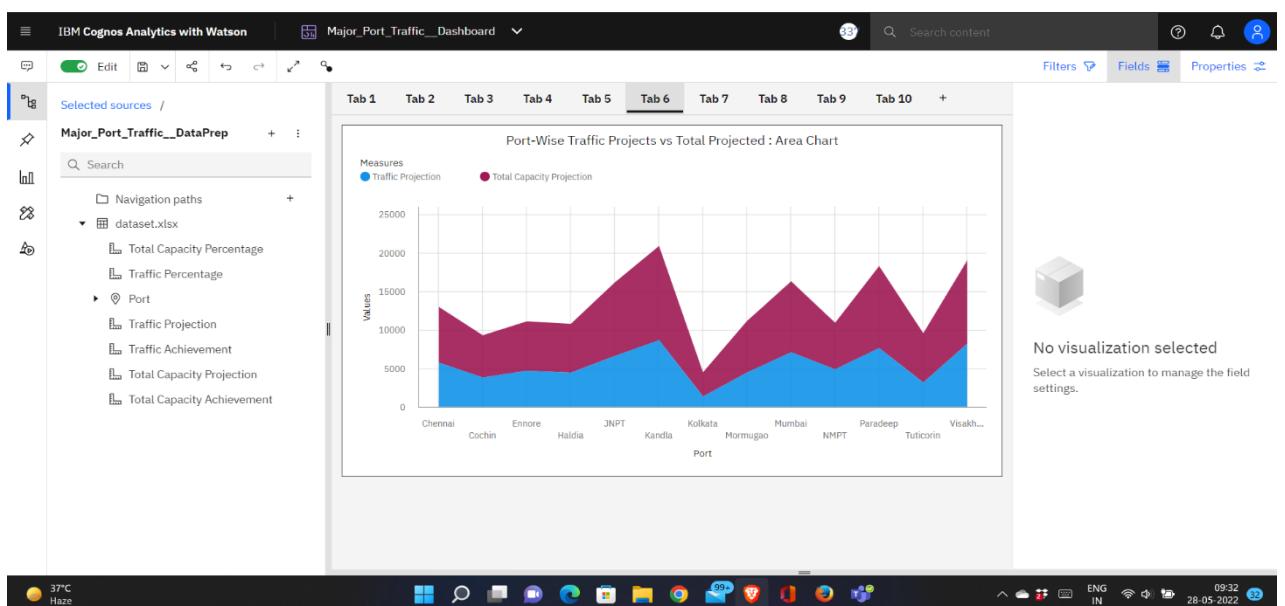
Port-wise Traffic Projected vs Achieved by ‘Column Chart’



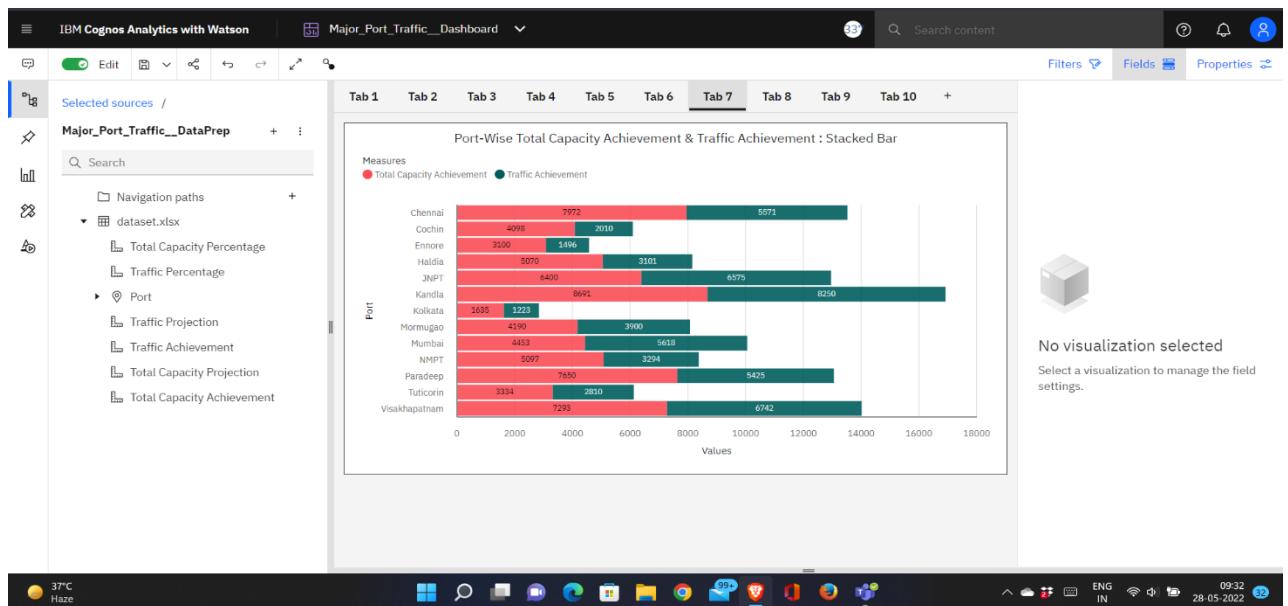
Port-wise Traffic Projected vs Achieve by ‘Stacked Column Chart’



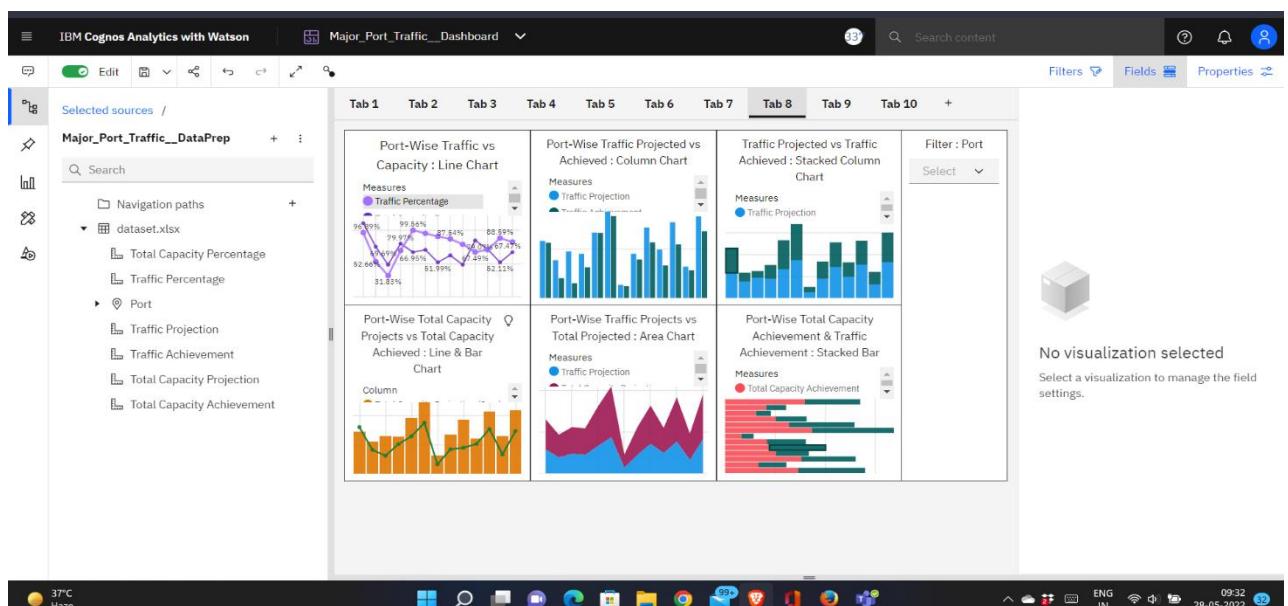
Port-wise Total Capacity Projects vs Total Capacity Achieve by ‘Line and Bar Chart’



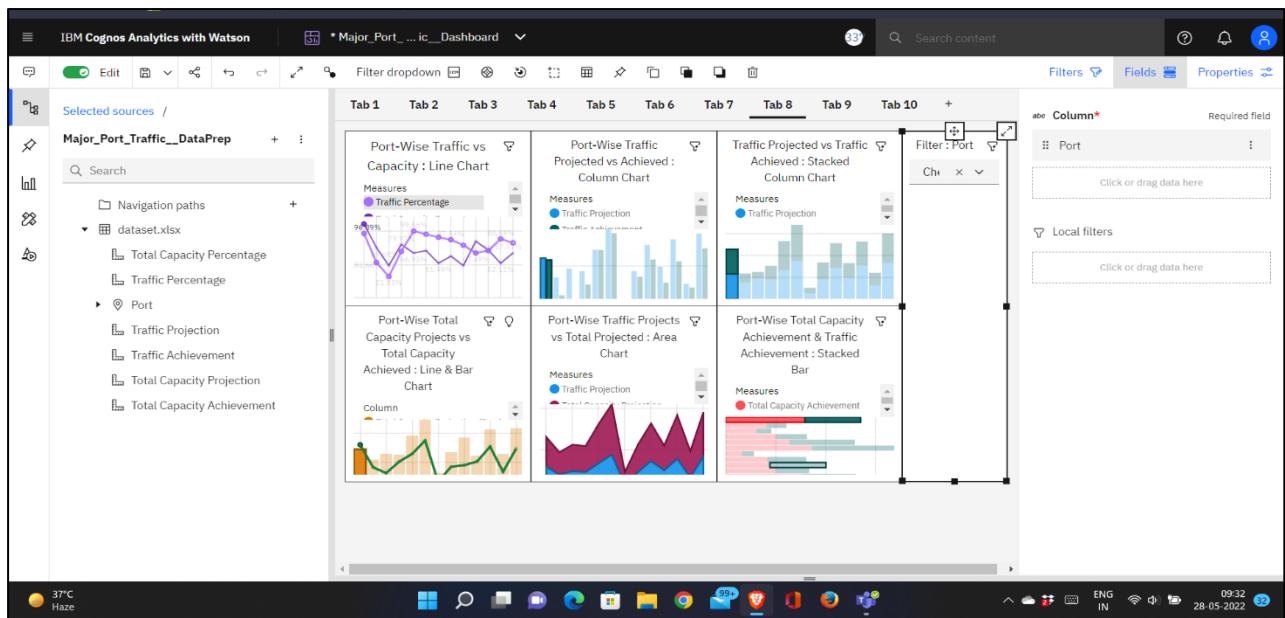
Port-wise Traffic Projects vs Total Projected by ‘Area Chart’



Port-wise Total Capacity Achieve, Traffic Achieved using ‘Stacked Bar’



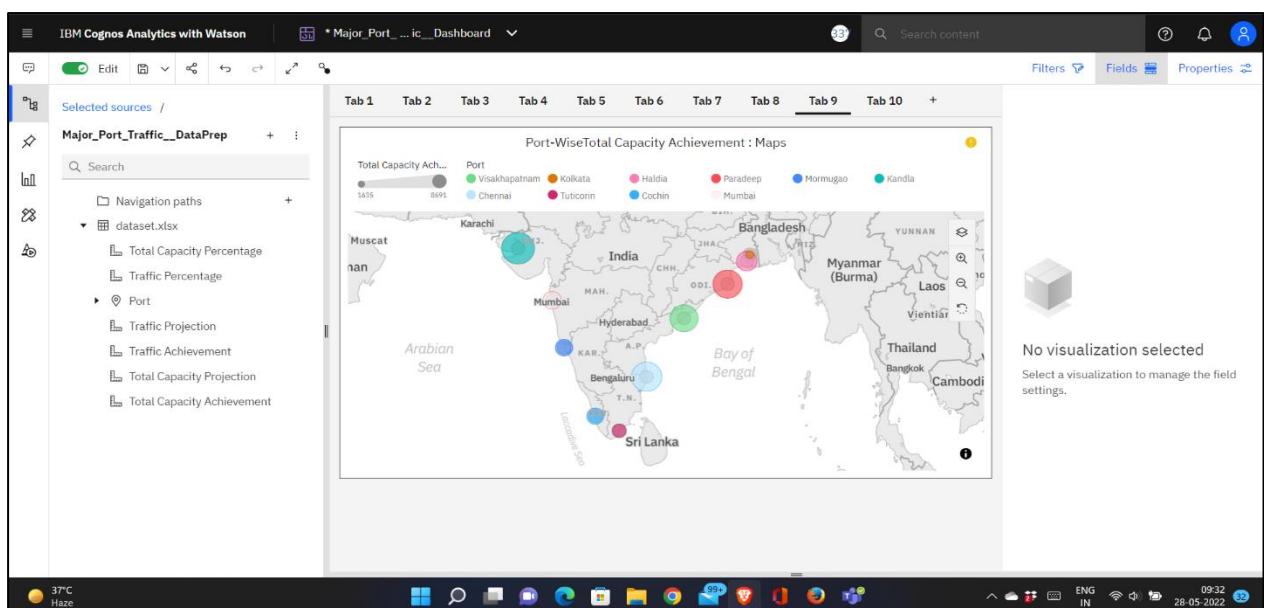
(Screenshot 1.0)



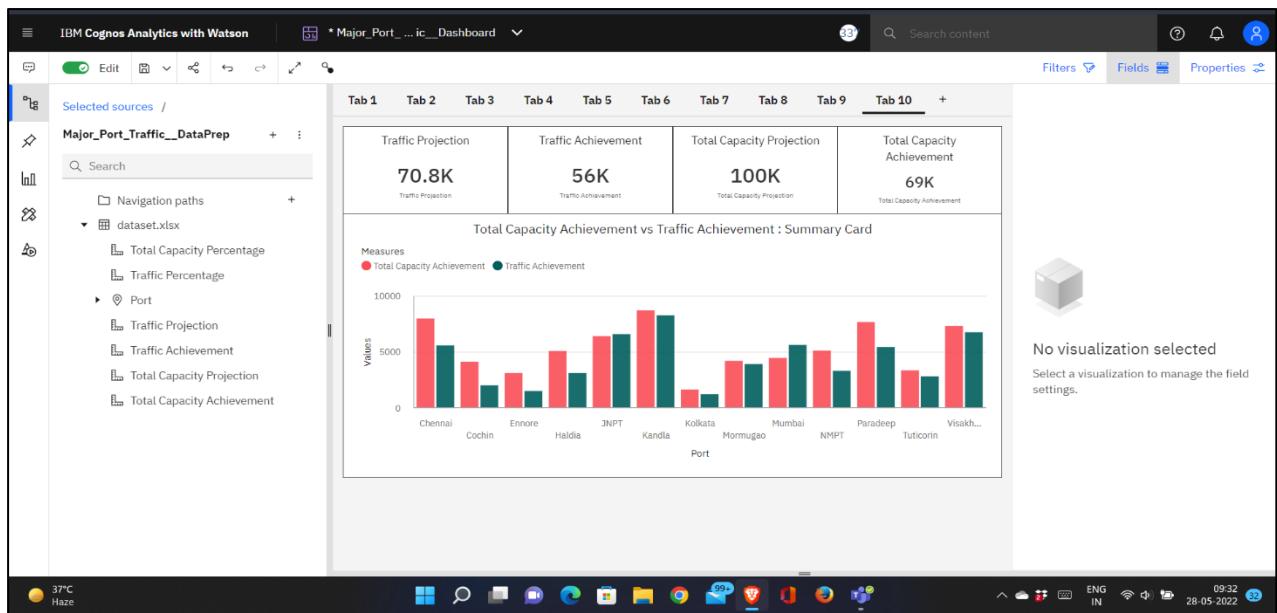
(Screenshot 1.1)

Like here from filters if we selected any one port name , say Chennai :

Then all visualizations would be specifically highlight and show the insights and visuals of data related to Port : Chennai

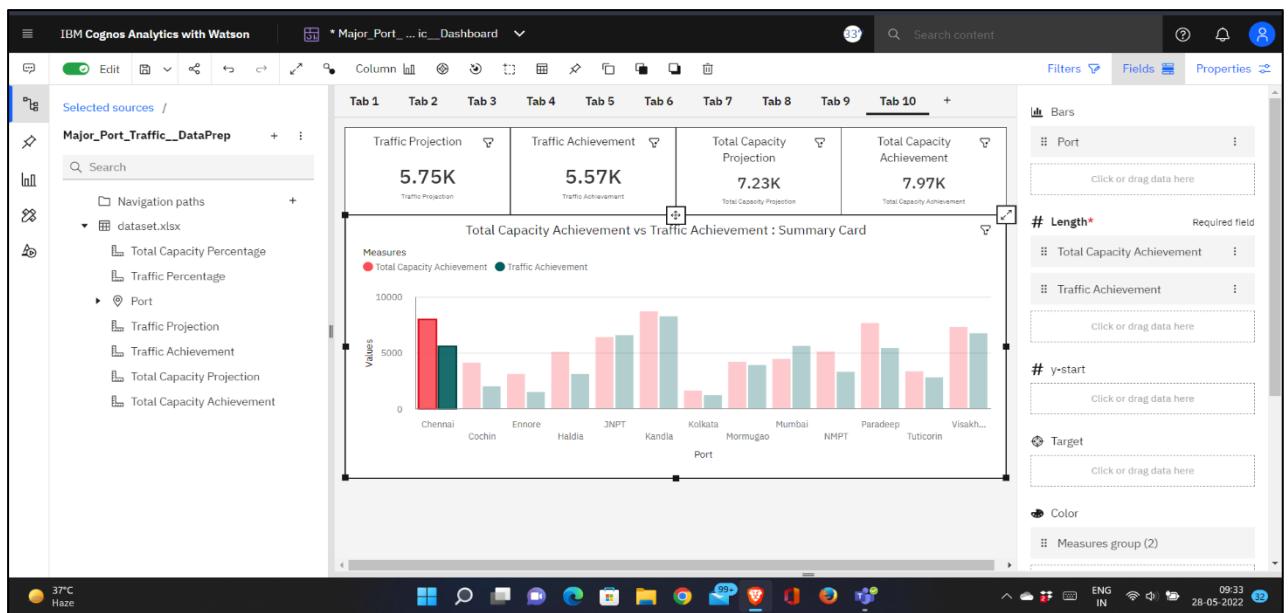


Port-wise Total Capacity Achieved using 'Map'



‘Summary Cards’ and Visual using Total Capacity vs Actual Capacity Column Chart

(Screenshot 1.0)



(Screenshot 1.1)

Here , like if we click any bar so , summary chart of Data_Prep , also changes values in for that specific port , like Chennai

CONCLUSION

as
learning / outcome from (training + project) :

I have learned fundamental concepts and I can work on IBM Cognos Analytics ,
also :

- Understanding big/small industrial dataset ,
- Making data preparation report of real dataset(s) ,
- Calculations in dataset for getting meaningful insights & visuals for appropriate predictive result for data(s) ,
- Have learned a broad understanding of plotting different graphs ,
- Learned to create meaningful dashboard