**Website Traffic Analysis**

**Project Overview**

The Website Traffic Analysis project aims to assess and understand the patterns, trends, and user behaviour on a specific website. By collecting and analysing data, the project will provide insights into the site's performance, user engagement, and areas for improvement. The results will inform decision-making, helping to optimize the website's content, design, and marketing strategies for enhanced user experience and increased traffic.

**OBJECTIVE**

Equipped with the right website traffic analysis tools, identify your top site pages, track visitor trends, calculate conversion rates, and ensure your marketing spend translates into an increase in conversions and sales.

**INSTALLATION OF JUPYTER NOTEBOOK**

Command to install jupyter notebook:

****

Output:



**WORKING OF JUPYTER NOTEBOOK:**

**Command to open jupyter notbook:**



**Output:**



**1.DATA COLLECTION**

In Python, you can collect data using various methods and libraries depending on your specific requirements. Determine the source of your data. It could be a file (e.g., CSV, JSON, Excel), a web API, a database, sensors, or user input.

From the given dataset (Kaggle):

Dataset Link: https://www.kaggle.com/datasets/bobnau/daily-website-visitors.

This consists of the CSV dataset for the further analysis.

From the collection of the given data we use the packages to extract and make the analysis with the following:

1. NUMPY
2. PANDAS

To install the packages:  


If the data is stored in files, you can use Python's built-in file handling capabilities or libraries like open() for text-based files or libraries like pandas for more complex data formats like CSV or Excel.

In Python, the can collect data using different methods and libraries tailored to your specific needs. This process typically involves selecting the data source, reading or retrieving data, preprocessing it as required, and potentially storing it for future use.

THE USAGE OF THE PACKAGES



Data collection is a fundamental step in the data analysis and decision-making process. It involves the systematic gathering of information from various sources, including files, databases, web services, sensors, user input, and more. Python provides a rich ecosystem of libraries and tools to facilitate data collection, making it a versatile language for this purpose.

**2. DATA EXTRACTION:**

Packages needed:

To extract command for those modules:

**Input:**



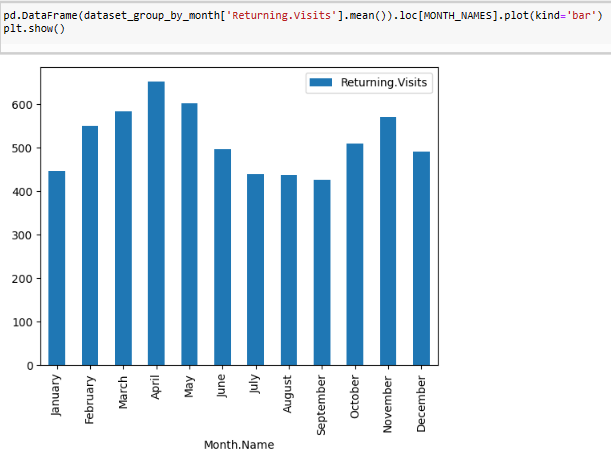
**CODE DESCRIPTION:**

* The code starts by importing the pandas library.
* The code then creates a variable called FILE\_LOCATION and assigns it to the path of a file on your computer.
* Next, the code reads in that CSV file into a DataFrame object using read\_csv().
* The index column is set to Date, which means that this DataFrame will have one row for each day of data.
* The index\_col='Date' parameter specifies that the column with the date should be used as the index.
* The thousands=',' parameter tells pandas to use commas for thousands separators in this column.
* The code opens a file called "P:\ibm\daily-website-visitors.csv" and reads in the data using csv.reader().
* The data is then stored in a list, which is assigned to variable "data".
* Next, numpy is imported as np so that we can use it to analyze the data.
* Finally, the first row of our dataset is analyzed with np.array() and printed out on screen for us to see what's going on with this dataset.
* The code will open the file "P:\ibm\daily-website-visitors.csv" and read the data in as a list of tuples, one for each row of data.

**EXTRACTION WITH VISUALIZATION:**

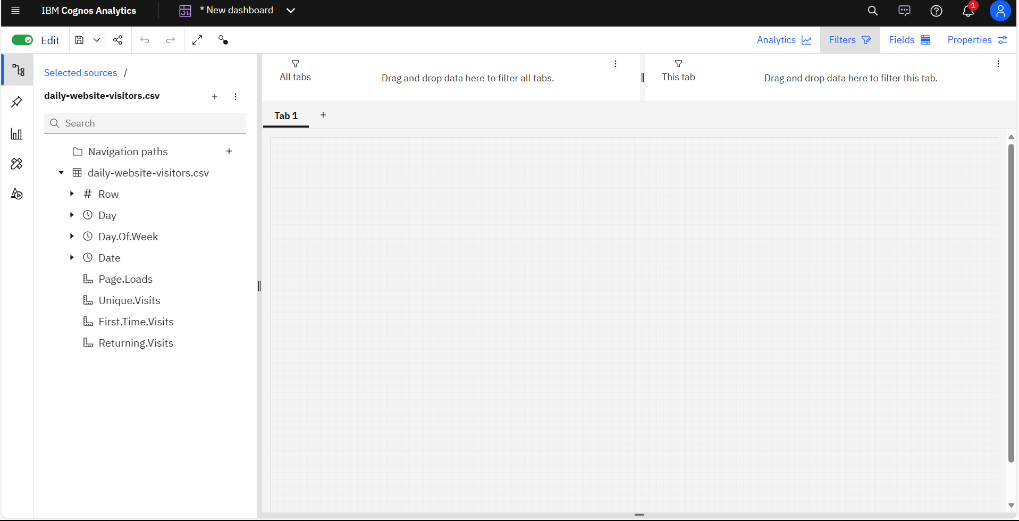
**Extraction is the process of retrieving or pulling data from one or more sources. These sources can be diverse and include databases, spreadsheets, web services, logs, and more.**

**Transformation involves manipulating, cleaning, and structuring the data to make it suitable for the desired use case. This can include operations like filtering, aggregating, joining, and more.**

**COGNOS ANALYTICS IN ACTION**

Cognos Analytics serves as a powerful tool to transform our analytical findings into actionable insights. This section explores the utilization of Cognos Analytics in enhancing the accessibility and applicability of our results.

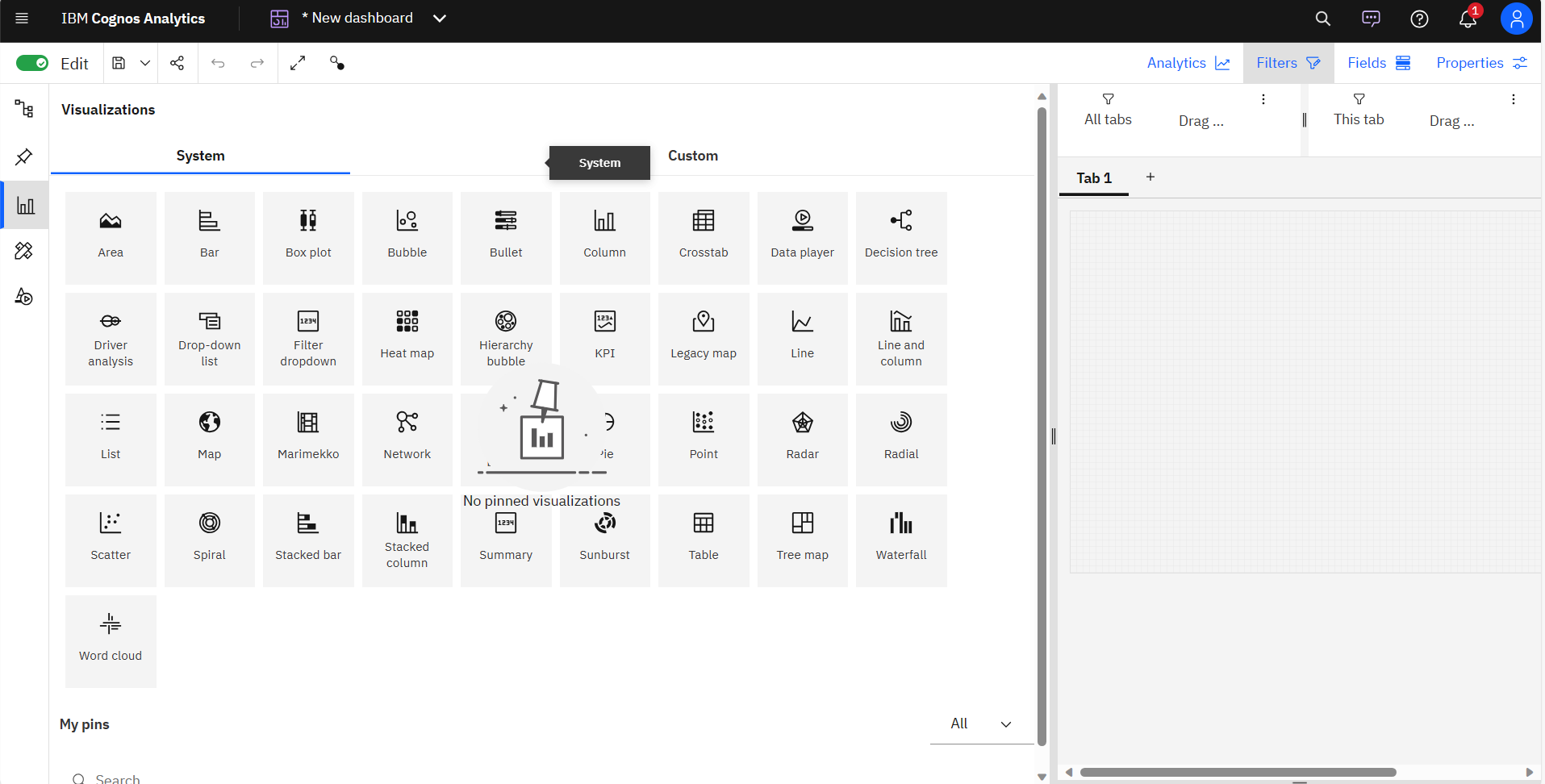
**Dashboard Design:** We delve into the process of designing intuitive dashboards within Cognos Analytics. These dashboards serve as a centralized hub for visualizing key metrics, trends, and predictions derived from our analysis. predictions derived from our analysis.



**3.DATA IN COGNOS ANALYTICS:**

Cognos Analytics is a business intelligence and analytics software platform developed by IBM. It is designed to help organizations collect, analyze, and visualize data from various sources to support data-driven decision-making.

IBM Cognos Analytics provides a range of powerful visualization tools and features to help users create compelling and interactive data visualizations within reports and dashboards. These tools enable users to transform data into meaningful insights.



Data Modules simplify data preparation and modeling, making it easier to create visualizations. A user-friendly drag-and-drop interface streamlines the report creation process, while interactive filtering enhances data exploration. Customization options, mapping capabilities, trend analysis, and support for advanced analytics enable users to tailor their visualizations to specific needs. Reports and dashboards are exportable and shareable, ensuring that insights are accessible to stakeholders, and the responsive design ensures a consistent user experience across devices.

**DATA SCIENCE IN PYTHON:**

Data science in Python involves leveraging the Python programming language and its extensive ecosystem of libraries to collect, clean, analyze, and visualize data. Python's popularity in data science is due to its versatility, ease of use, and robust community support. This process encompasses data collection from diverse sources, data cleaning and preprocessing with tools like pandas, exploratory data analysis using libraries like Matplotlib and Seaborn, statistical analysis through SciPy and Statsmodels, and machine learning.

**DATA PREPROCESSION:**

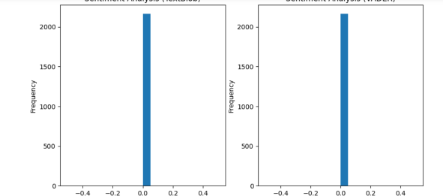
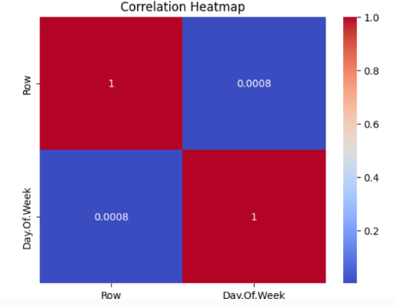
Data preprocessing is a fundamental step in the data analysis pipeline. . The primary purpose of data preprocessing is to improve the quality, consistency, and relevance of the data, making it easier to extract meaningful insights or to train accurate models.

**EXPLORATORY DATA ANALYSIS:**

Exploratory Data Analysis (EDA) is a preliminary step in data analysis that involves the systematic examination and initial understanding of your dataset. The primary goals of EDA are to identify patterns, detect anomalies, and gain insights into the data.

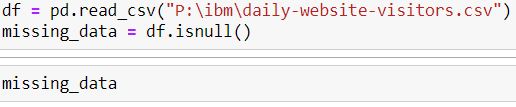
**NATURAL LANGUAGE PROCESSING:**

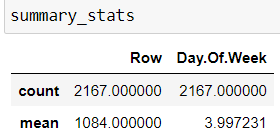
Natural Language Processing (NLP) in Python is a set of techniques and libraries used for the analysis, manipulation, and understanding of human language data. NLP is widely used in data analysis for extracting insights from unstructured text data, and it's particularly valuable in areas like sentiment analysis, etc.

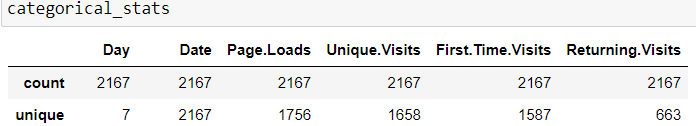


**4.DATA EXPLORATION AND VISUALIZATION**

Data exploration involves the initial phase of understanding and characterizing a dataset, including computing summary statistics, visualizing data distributions, and identifying relationships between variables. It also entails addressing data quality issues such as missing values and duplicates. On the other hand, data visualization is the practice of representing data through graphical elements like bar charts, scatter plots, line charts, and heatmaps, making it easier to perceive patterns, trends, and outliers. Effective data visualization aids in conveying insights and is an essential tool for communicating findings to stakeholders, enabling data analysts and scientists to extract meaningful information and make informed decisions based on the data's visual representations.



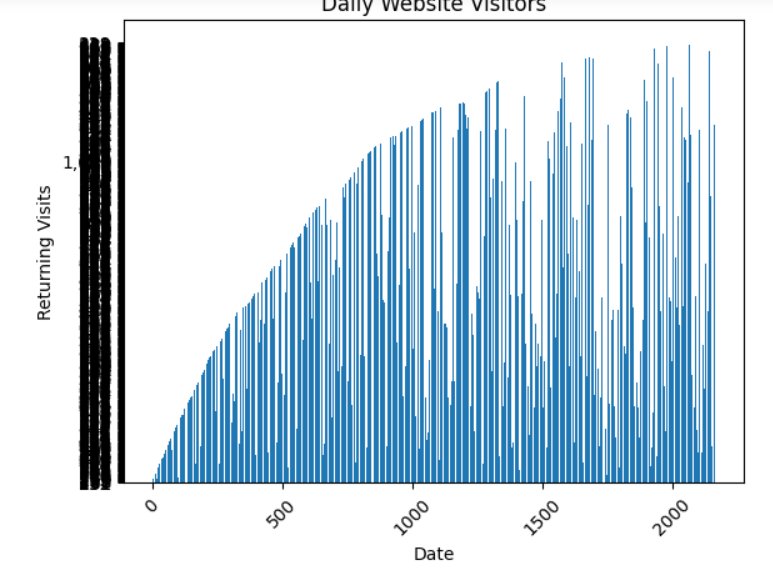
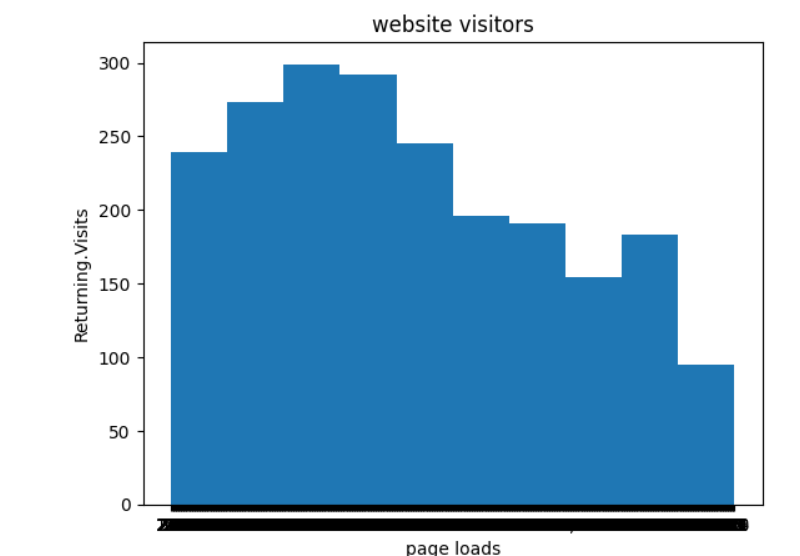




Data exploration involves getting to know your data by examining its key characteristics and relationships. This step is crucial for data preparation and quality assurance. In parallel, data visualization leverages graphical representations to convey insights effectively. It helps in revealing patterns, trends, and outliers in the data, making complex information more accessible and understandable, and is a valuable tool for decision-making and data-driven storytelling.

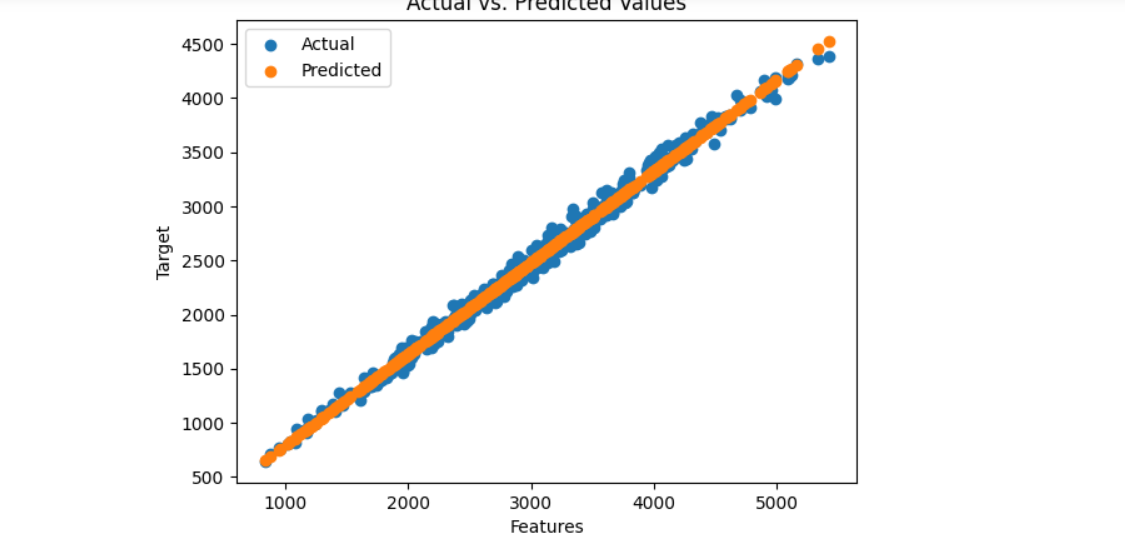
**VISUALIZATION:**

Data visualization is a powerful technique for transforming complex data into easily digestible and meaningful insights. It involves the use of graphical representations, such as charts, graphs, and maps, to convey information visually. By encoding data into visual elements, patterns, trends, and relationships within the data become more apparent, enabling individuals to make data-driven decisions and understand complex datasets at a glance. Effective visualization is not only aesthetically pleasing but also enhances data storytelling, making it accessible to a wide range of audiences, from data analysts to decision-makers. Whether it's depicting sales trends, geographic distributions, or demographic patterns, data visualization plays a vital role in distilling large volumes of information into clear and actionable messages, making it an indispensable tool in fields like data science, business intelligence, and scientific research.



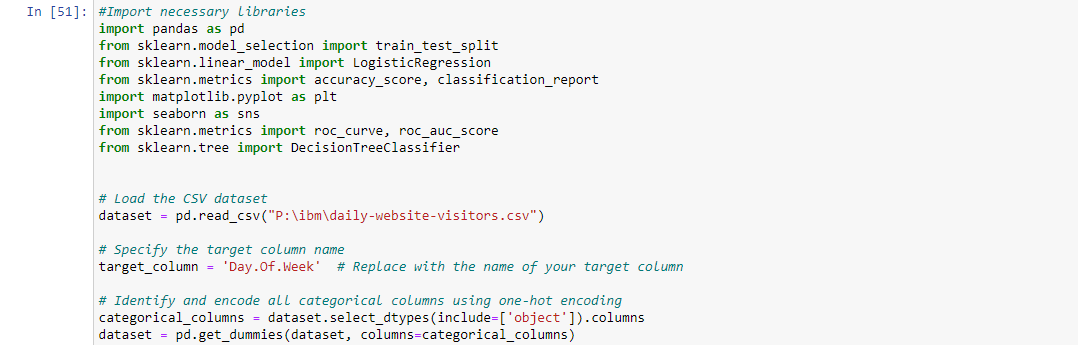
**REGRESSION:**

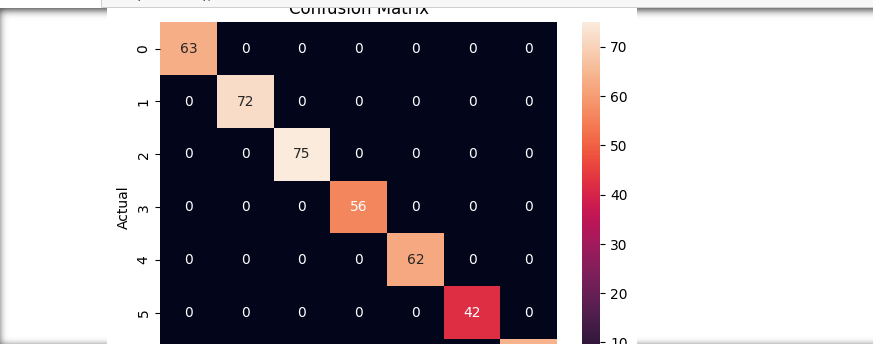
Regression analysis is a statistical method used to understand and model the relationship between one or more independent variables and a dependent variable. It aims to quantify how changes in the independent variables affect the dependent variable. In simple linear regression, there is one independent variable, while multiple linear regression deals with more than one.



**5.SUPERVISED LEARNING CLASSIFICATION:**

Supervised learning classification is a machine learning technique where the algorithm learns to assign predefined labels or categories to input data based on patterns and relationships in the training dataset. It's used to build models that can classify new, unseen data into one of the known classes. Classification algorithms, such as decision trees, support vector machines, and neural networks, are trained on labeled data, enabling them to make predictions and automate tasks like spam detection, image recognition, and sentiment analysis. This method is a fundamental tool in data science and artificial intelligence for solving a wide range of real-world problems where categorizing data is crucial.

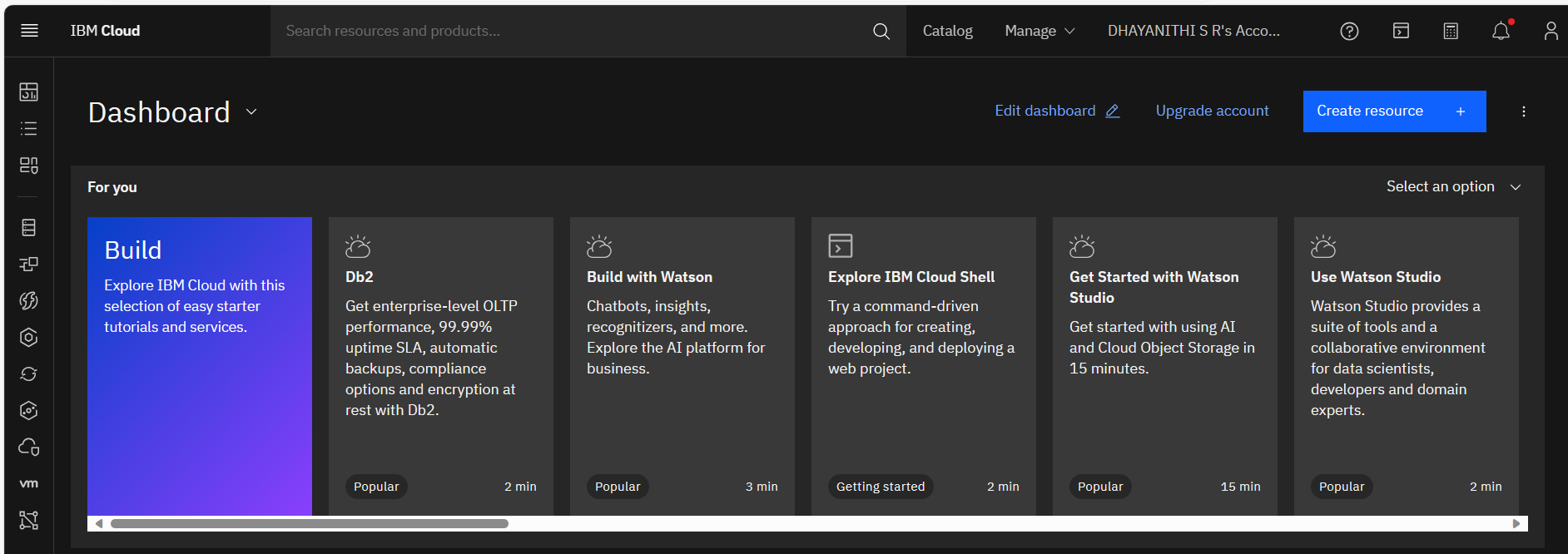


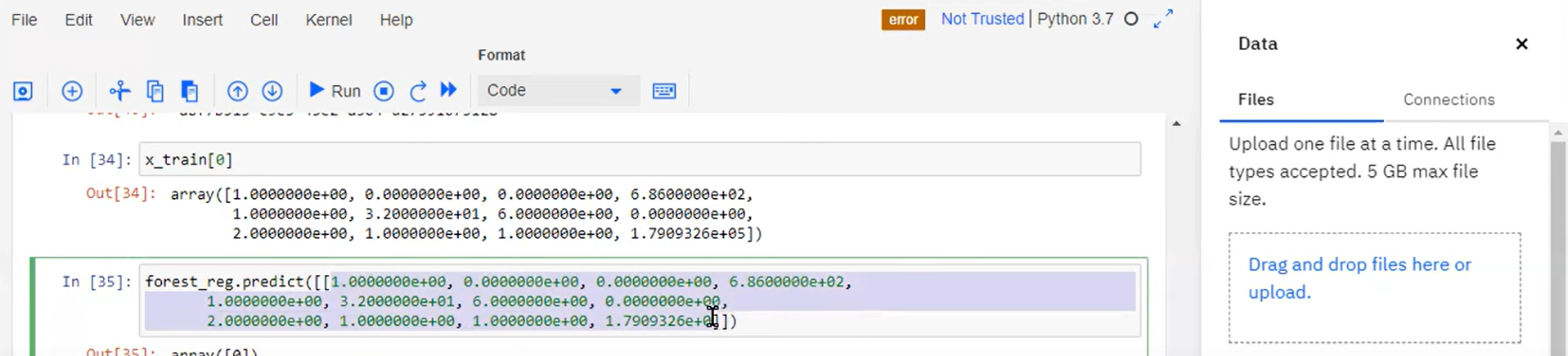


It relies on labeled training data to establish patterns and relationships for accurate predictions.

**DEPLOYMENT IN CLOUD:**

Building and deploying AI models in IBM Cloud involves a streamlined process that leverages IBM's AI and cloud infrastructure often with the help of IBM Watson Studio or AutoAI tools for data preparation, model training, and evaluation. Once the model is ready, IBM Cloud provides a platform for deploying and hosting your AI application or service. This deployment can range from web applications to REST APIs, ensuring that your AI capabilities are accessible over the internet. IBM Cloud services like Watson Machine Learning and Watson cloud hosting within IBM Cloud, you can create, deploy, and manage your AI applications in a seamless and cost-effective manner, while also taking advantage of IBM's robust AI ecosystem and infrastructure.





In IBM Cloud, building and deploying AI models is a streamlined process. You develop your AI model using IBM Watson Studio or similar tools, and once it's ready, IBM Cloud provides a platform for easy deployment, whether it's a web app or an API service. IBM Cloud's AI services help manage, scale, and monitor your AI solutions, making it a convenient and efficient solution for creating and hosting AI applications.

CONCLUSION:

This seamless integration allows organizations to harness the benefits of artificial intelligence in a cost-effective and efficient manner. With features like data preparation, model training, and deployment, IBM Cloud simplifies the entire AI lifecycle, making it accessible to a broad range of businesses and industries