# Big-Data-Analysis-with-IBM-Cloud-Databases-for-Iris-Dataset

This project uses IBM Cloud Databases to analyze the iris dataset, which has 150 iris flowers with four features.

This project uses IBM Cloud Databases to store and manage the iris dataset, and performs various data analysis tasks to extract valuable insights from the data. The project also includes designing and presenting the analysis results using appropriate visualizations and business intelligence tools.

🔺 Installation and Usage

To run this project, you will need:

    - An IBM Cloud account

    - A PostgreSQL database service on IBM Cloud

    - A database client tool, such as pgAdmin or DBeaver

    - Python 3.7 or higher

    - Jupyter Notebook

    - The following Python libraries: pandas, numpy, matplotlib, sklearn

🔺 Features and Functionality

This project performs various data analysis tasks on the iris dataset using SQL or python. Some of the tasks are:

           - Check the basic information about the data, such as number of rows, columns, data types, missing values, duplicates, etc.

           - Calculate some descriptive statistics for each feature and target variable, such as mean, median, mode, standard deviation, range, etc.

           - Perform some data cleaning and preprocessing tasks, such as handling missing values, outliers, errors, etc.

           - Perform some data transformation tasks, such as scaling, normalization, encoding, etc.

           - Perform some feature engineering tasks, such as creating new features from existing ones, selecting relevant features, etc.

           - Perform some exploratory data analysis (EDA) tasks to understand the distribution and relationship of the data using visual techniques such as histograms, box plots, scatter plots, etc.

           - Perform some predictive analytics tasks to build models that can predict the target variable (species) from the features using techniques such as classification, decision trees, random forests, k-nearest neighbors, support vector machines, etc.

           - Perform some prescriptive analytics tasks to provide recommendations or suggestions based on the analysis results using techniques such as optimization, simulation, scenario analysis, etc.

This project also designs visualizations that can present the analysis results in an understandable and impactful manner using various tools and libraries. Some of the tools and libraries are:

    - Matplotlib: A Python library for creating static 2D plots.

    - Seaborn: A Python library for creating statistical and aesthetic plots.

    - IBM Watson Studio : An integrated environment designed to make it easy to develop, train, manage models, and deploy AI-powered applications.

    - IBM Watson ML :A powerful tool that allows you to explore data, prepare it for training and evaluation.