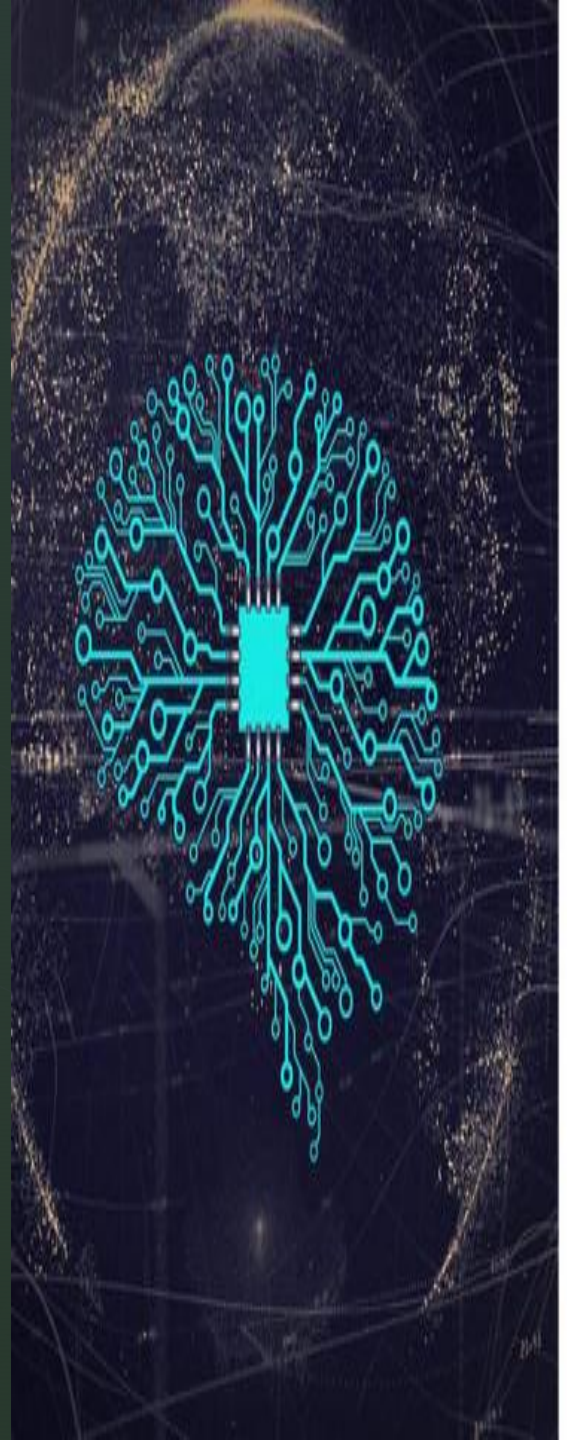
IBM

Basic Presentation

Objective/ key components

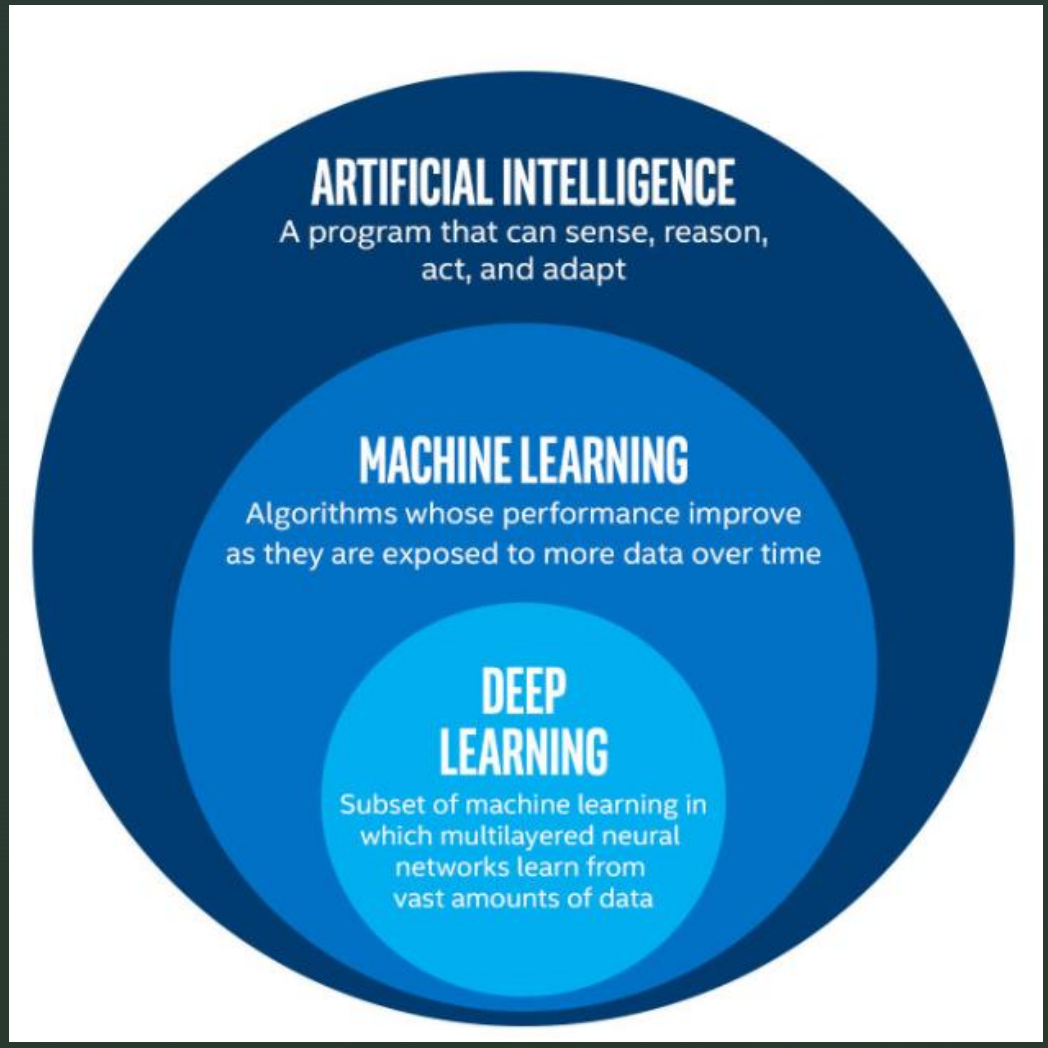
- data collection and data wrangling methodology
- EDA and interactive visual analytics methodology
- predictive analysis methodology
- Plotly Dash dashboard
- predictive analysis



OVERVIEW on ML and AI

Machine learning (ML) is a type of artificial intelligence (AI) focused on building computer systems that learn from data. The broad range of techniques ML encompasses enables software applications to improve their performance over time.

Machine learning algorithms are trained to find relationships and patterns in data. They use historical data as input to make predictions, classify information, cluster data points, reduce dimensionality and so on…

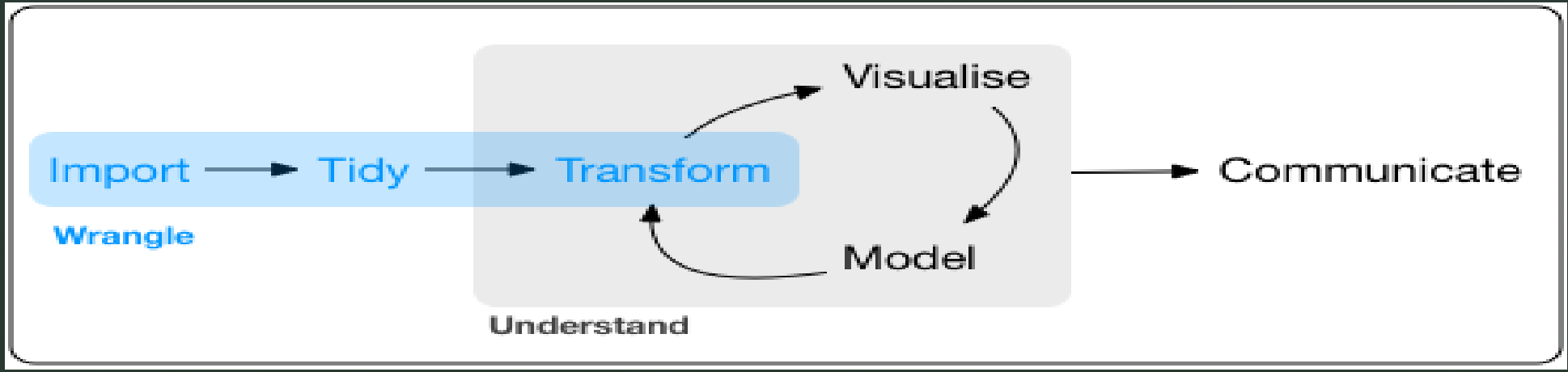


Data collection and Data wrangling methodology

A data wrangling process, also known as a data munging process, consists of reorganizing, transforming and mapping data from one "raw" form into another in order to make it more usable and valuable for a variety of downstream uses including analytics.

Steps are present in the data wrangling process:

- discovery
- structuring
- Cleaning
- enriching
- validating
- publishing



EDA and interactive visual analytics methodology

Exploratory data analysis (EDA) is a task of analyzing data using simple tools from statistics, simple plotting tools.

EDA techniques may include :

- calculating summary statistics
- visualizing data distributions
- identifying outliers
- exploring relationships between variables
- performing hypothesis testing.

This process helps gain insights into the data, identify patterns, and inform further analysis or decision-making.

Data Exploration using numerical analysis

Data Exploration using visual analysis

Data Exploration using TensorFlow validator

Data Exploration using Lux

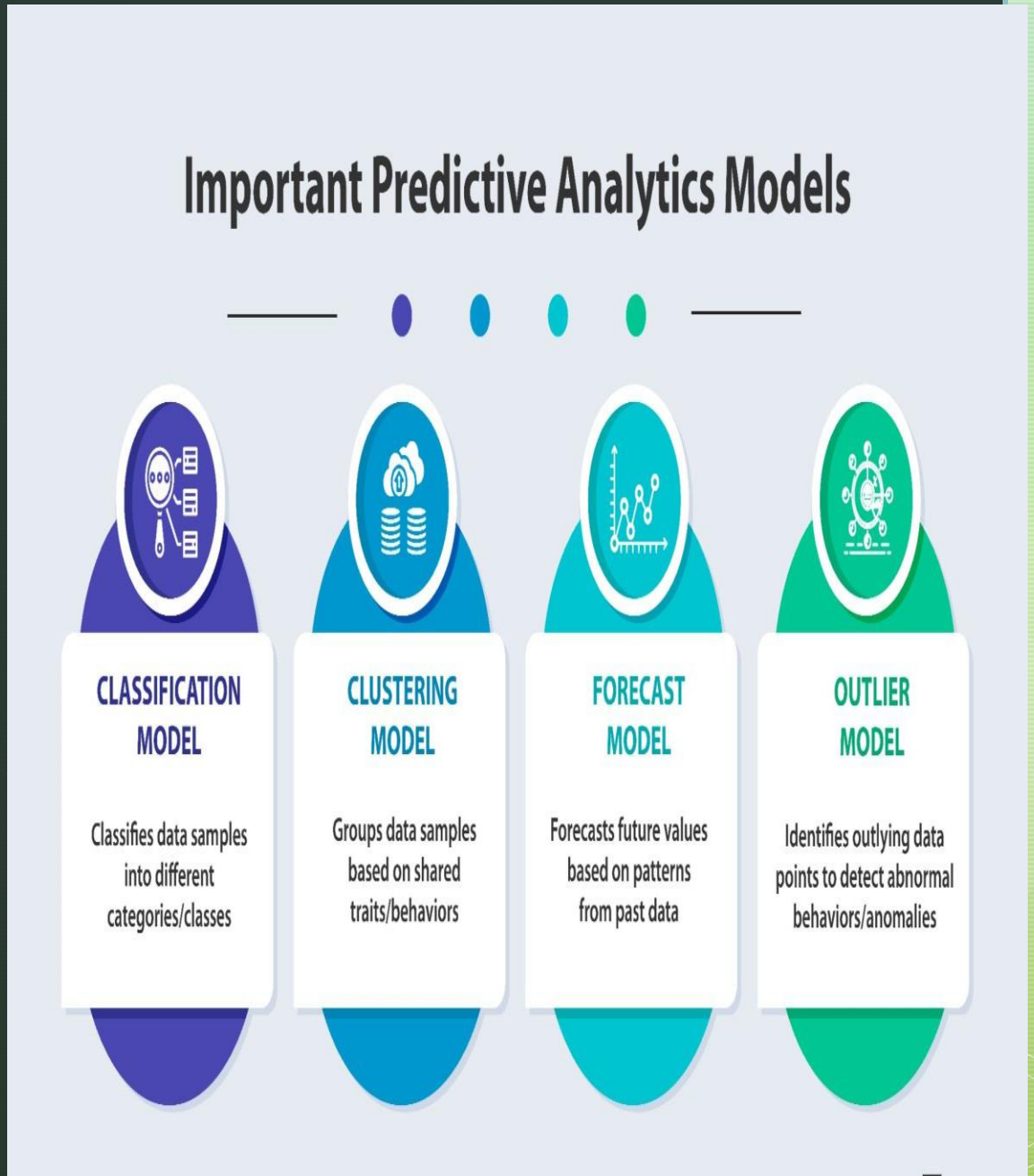
Predictive analysis methodology

Predictive analytics is a type of advanced data analytics that uses historical data to predict future outcomes. It combines statistical modeling, data mining, and machine learning techniques to analyze the data.

The goal is to provide the most accurate prediction of what will happen in the future.

Predictive analytics can be used to:

- Find patterns in data to identify risks and opportunities
- Forecast future workforce needs and skills requirements
- Analyze employee data to identify factors that contribute to high turnover rates
- Clean and improve data used for estimates



EDA with visualization

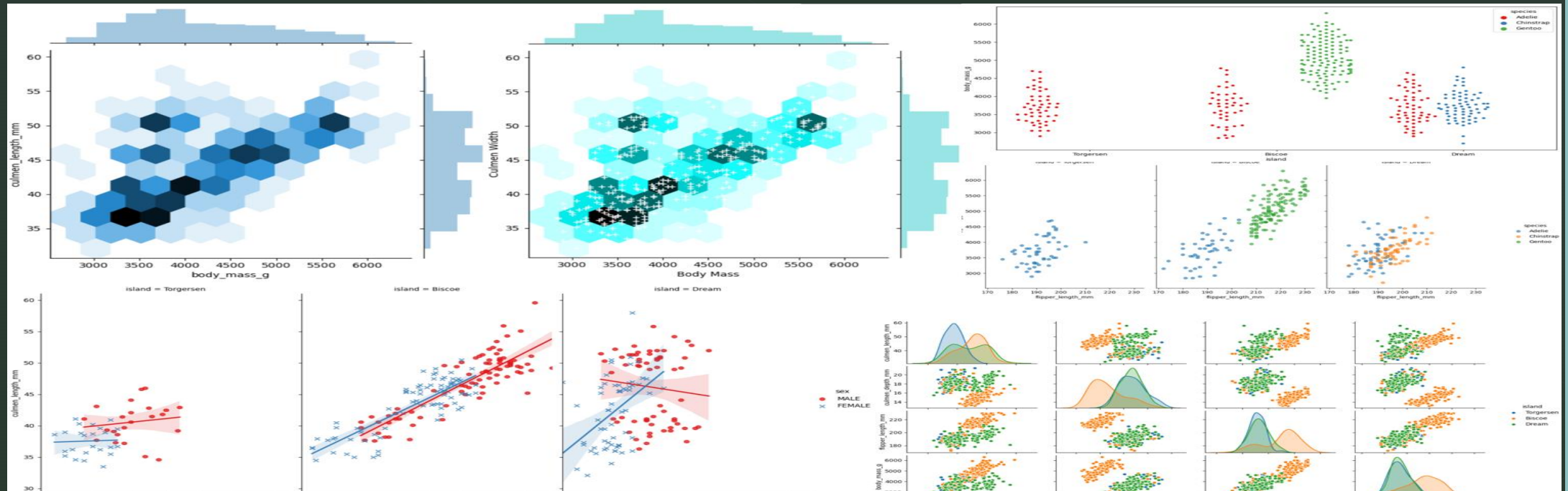
Exploratory Data Analysis (EDA) is a data science process that involves analyzing and summarizing datasets to uncover patterns, trends, and insights. Data visualization is a key component of EDA, and it involves the graphical representation of data using plots, charts, and graphs.

Data visualization helps data analysts:

- **Identify patterns, trends, and relationships within the data**
- **Gain insights into the data's distribution, relationships between variables, and potential outliers**

Some examples of data visualizations include:

Histograms, Box plots, Scatter plots, Line plots, Heatmaps, Bar charts.



EDA with SQL

Exploratory Data Analysis (EDA) in Python using SQL and Seaborn (SNS) Exploratory Data Analysis (EDA) is **a method of analyzing datasets to summarize their main characteristics**. EDA is often used to analyze KPIs and for companies adopting data-driven decision making.

We can use EDA with SQL. For example, you can use the UPDATE statement of SQL to substitute a value.

Conclusion

- Completed the required data collection and data wrangling methodology
- Completed the required EDA and interactive visual analytics methodology
- Completed the required predictive analysis methodology
- Completed the required EDA with visualization
- Completed the required EDA with SQL
- Completed the required interactive map with Folium
- Completed the required Plotly Dash dashboard
- Completed the required predictive analysis