Data Analytics

Lesson 01. Introduction

Understanding Data Analytics Terminologies

Dr. Hai Tran

hai.tran@sbsuni.edu.vn

Scholar: https://scholar.google.com/citations?user=kHZvlTkAAAAJ&hl=en&oi=ao

Co-Founder: XAI - https://xai.foo/

Saigon Business School

In partnership with

UNIVERSITY CANADA WEST

MACQUARIE University SYDNEY·AUSTRALIA

University of Northampton

AMITY GLOBAL INSTITUTE

Saigon Business School

Course Information

- 4 credits: 48 lecture hours
- Main content of the subject
  - The theoretical foundation of data analytics and a range of data analytic processes and techniques to provide hands-on experience to enhance their skills.
  - Implementing data analytic processes and techniques using a programming language such as Python, or a tool such as Weka, PowerBI, Excel.
- Assessments:
  - Attendance (20%): Activities in class.
  - Assignment 1(40%): Week 1 to Week 3.
  - Assignment 2(40%): Week 4 to Week 6.
- Reference: textbook, slide and Jupiter notebook on Google Colab.

Saigon Business School

Learning Outcomes

- LO1 Discuss the theoretical foundation of data analytics that determine decision making processes in management or business environments
- LO2 Apply a range of descriptive analytic techniques to convert data into actionable insight using a range of statistical techniques
- LO3 Investigate a range of predictive analytic techniques to discover new knowledge for forecasting future events
- LO4 Demonstrate prescriptive analytic methods for finding the best course of action for a situation

Saigon Business School

Learning materials

- Textbook
  - Evans, J. (2016) Business Analytics. 2nd edn. Pearson.
  - Runkler, T. (2016) Data Analytics: Models and Algorithms for Intelligent Data Analysis. 2nd edn. Vieweg+Teubner Verlag.

- Online reference materials
  - archive.ics.uci.edu/ml/
  - powerbi.microsoft.com
  - https://github.com/topics/data-analysis-python
  - https://media.pearsoncmg.com/ph/esm/esm_evans_eba3e_20/tools/eba3e_analytic_solver.html
  - https://data.imf.org/

Saigon Business School

Agenda

- Lesson 1: Understanding Data Analytics Terminologies.
- Lesson 2: Foundation of Business Analytics
- Lesson 3: Visualizing and Exploring data
- Lesson 4: Applying Descriptive Analytic Techniques
- Lesson 5: Data Modeling
- Lesson 6: Predictive Analytics
- Lesson 7: Regression, Classification and Clustering
- Lesson 8: Forecasting Techniques
- Lesson 9: Investigating Predictive Analytic Techniques
- Lesson 10: Introduction to Data Mining
- Lesson 11: Demonstrating Prescriptive Analytic Methods
- Lesson 12: Recap and advanced topics

Saigon Business School

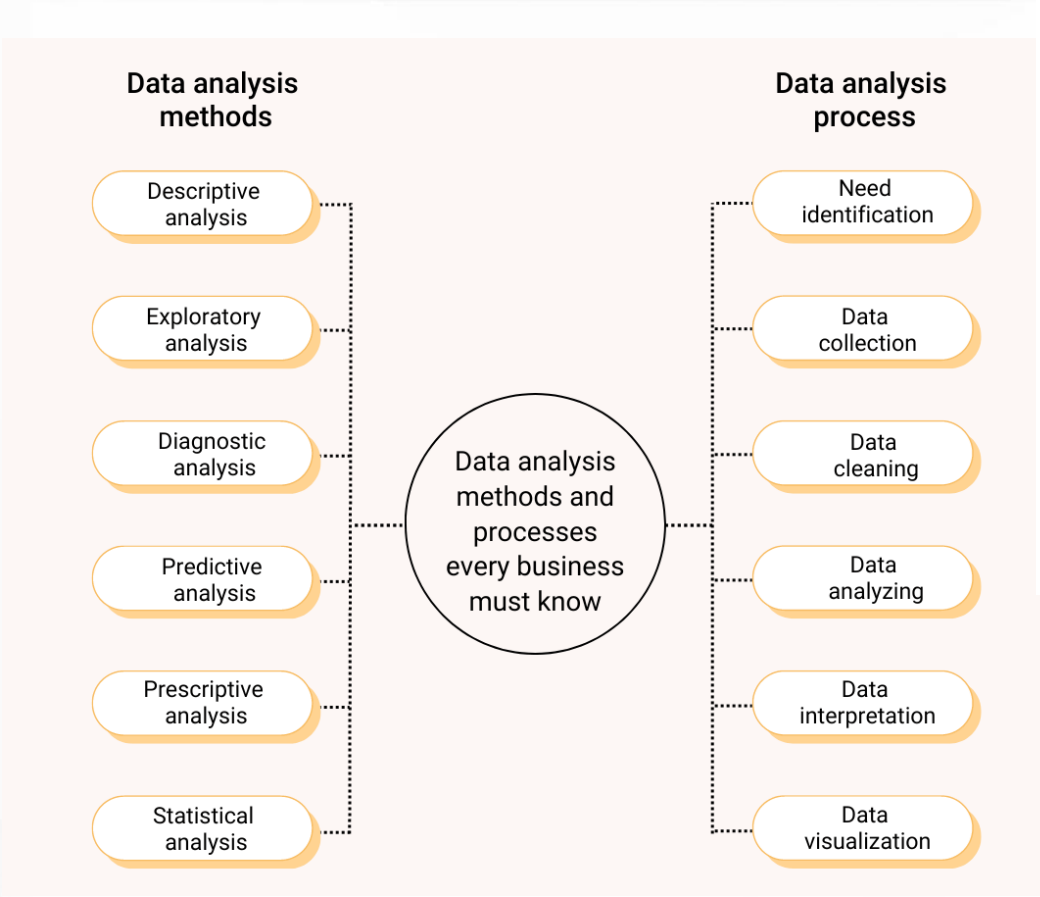
Supplements for Analytic Solver Basic

- 01 Getting Started With Analytic Solver Basic zip
- 02 Using Advanced Regression zip
- 03 Using Forecasting Techniques in Analytic Solver zip
- 04 Using Data Mining in Analytic Solver zip
- 05 Model Analysis in Analytic Solver zip
- 06 Using Monte Carlo Simulation in Analytic Solver zip
- 07 Using Linear Optimization in Analytic Solver zip
- 08 Using Optimization Parameter Analysis in Analytic zip
- 09 Using Integer and Nonlinear Optimization in Analytic Solver zip
- 10 Using Decision Trees in Analytic Solver zip

https://media.pearsoncmg.com/ph/esm/esm_evans_eba3e_20/tools/eba3e_analytic_solver.html

Saigon Business School

Data Analytic Tools



- MS Excel
- SQL
- R
- SAS
- Tableau

Basic data analytics tools

- SPSS
- MATLAB
- Python
- Hadoop
- Spark
- Hive

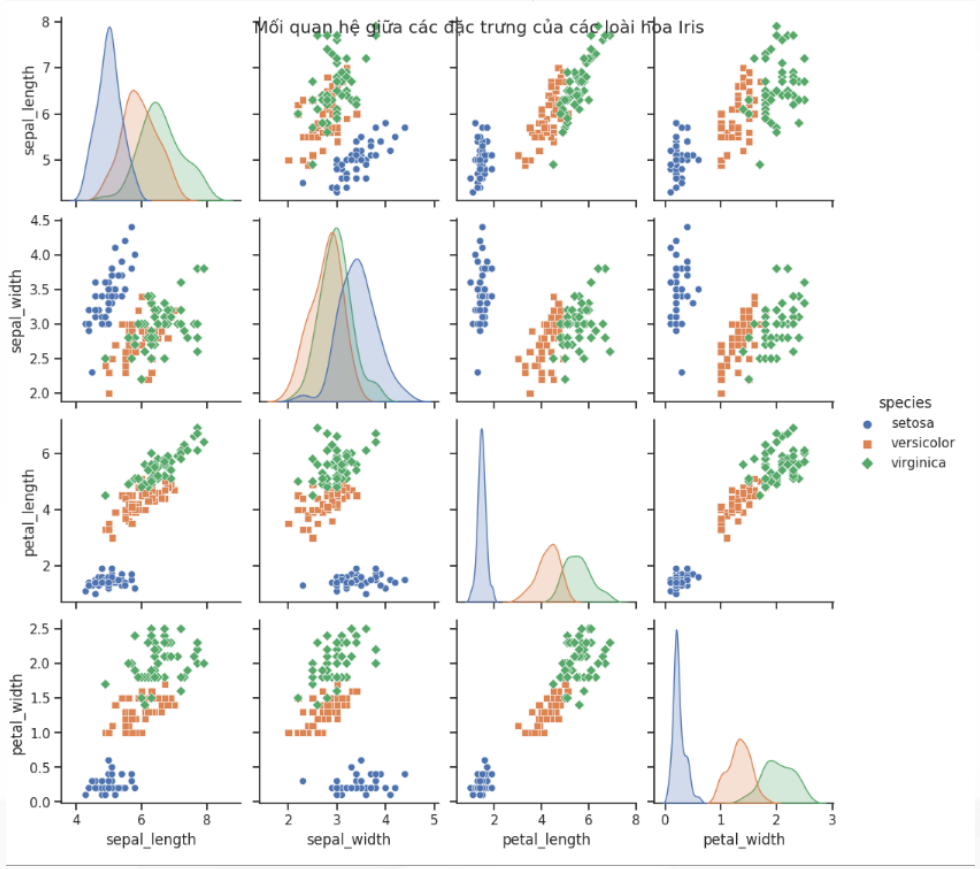
Advanced data analytics tools

Saigon Business School

Introduction – Python Data Analytics

- Example: Python script for Data Analytics

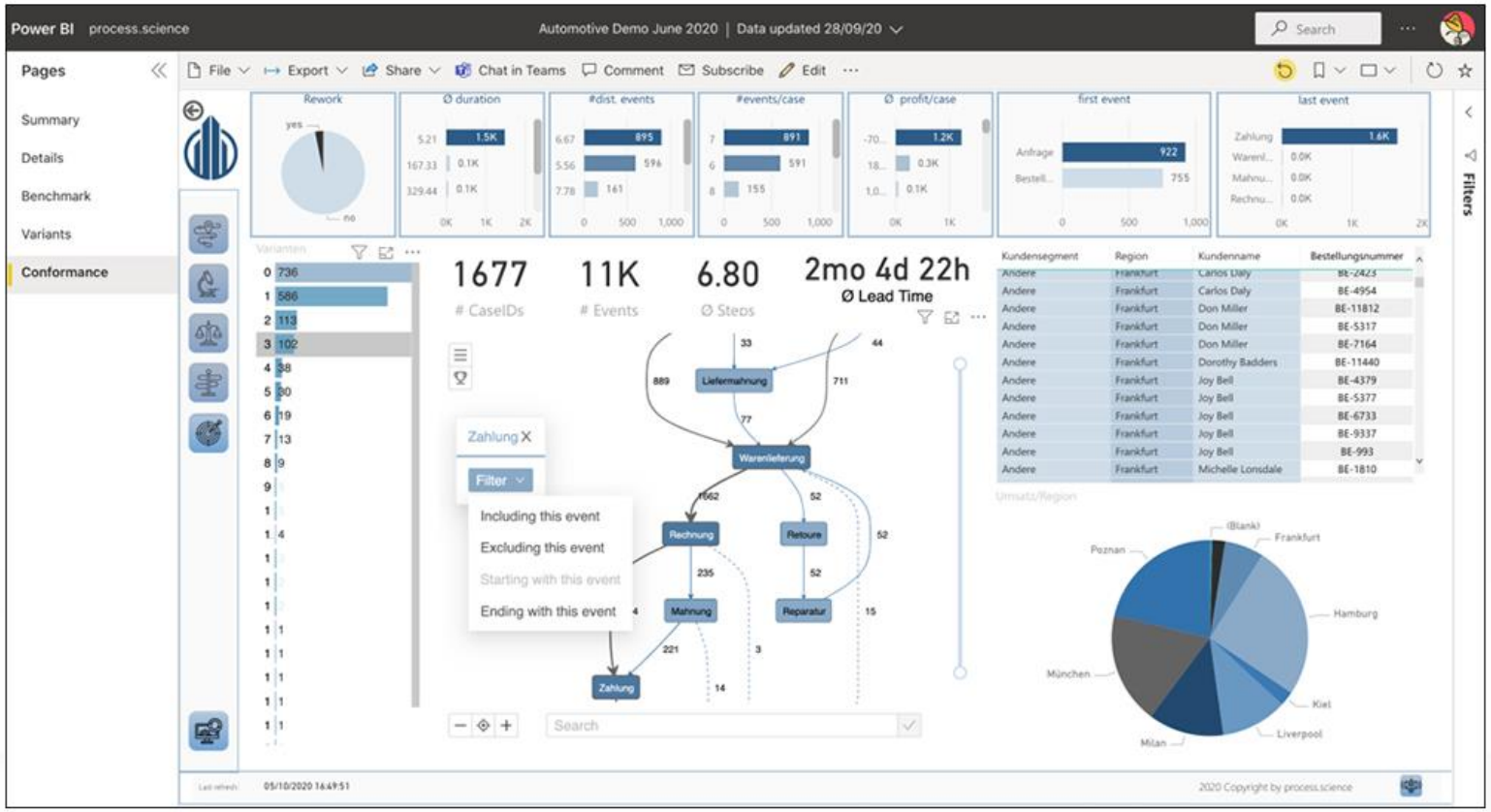
```
import seaborn as sns
import matplotlib.pyplot as plt

# Load Iris dataset from Seaborn
iris = sns.load_dataset("iris")

# Create pairplot diagram and consider
distribution and feature relationship
sns.set(style="ticks")
sns.pairplot(iris, hue="species",
markers=["o", "s", "D"])

# Set title
plt.suptitle("Mối quan hệ giữa các đặc trưng
của các loài hoa Iris")

# Show diagram
plt.show()
```



Saigon Business School

Introduction – PowerBI Data Analytics

- Example: PowerBI for Data Analytics



Saigon Business School

Introduction – Excel Data Analytics

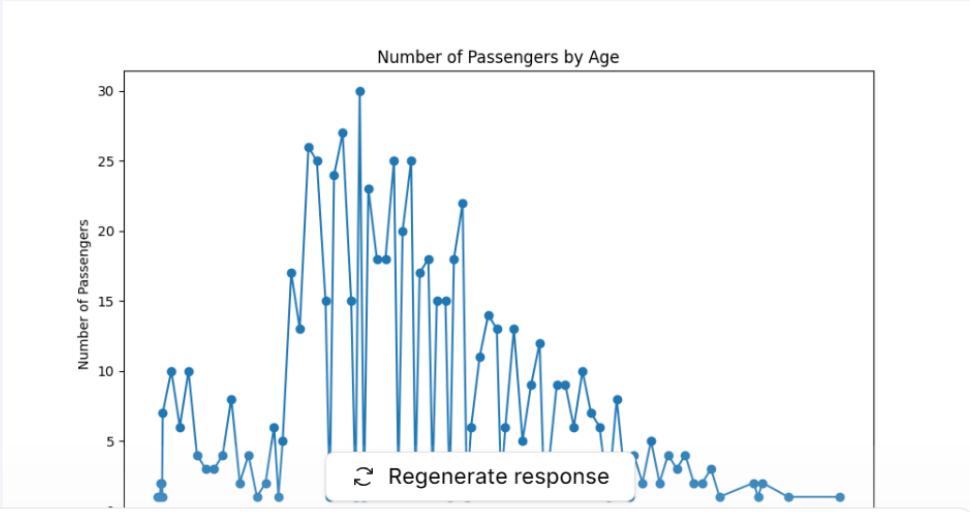
- Example: AI Excie for Excel Data Analytics

chat.excie.org

Excie

Excie

Here is the line chart illustrating the number of passengers by age:

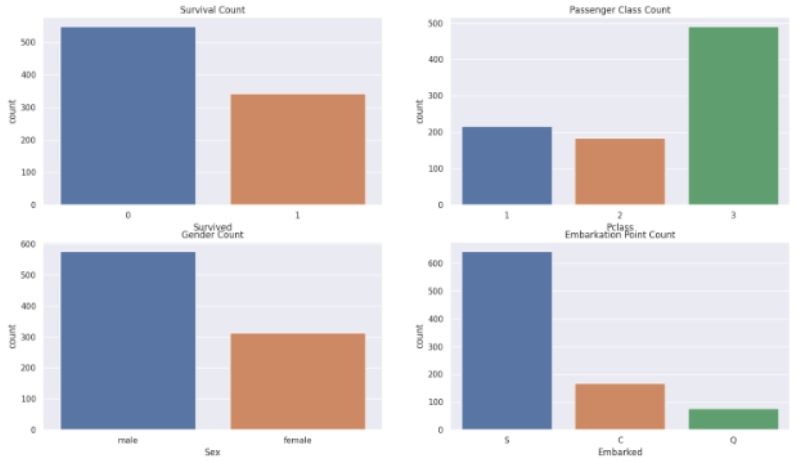


Regenerate response

To get the best answer, please ask in English.

Excie

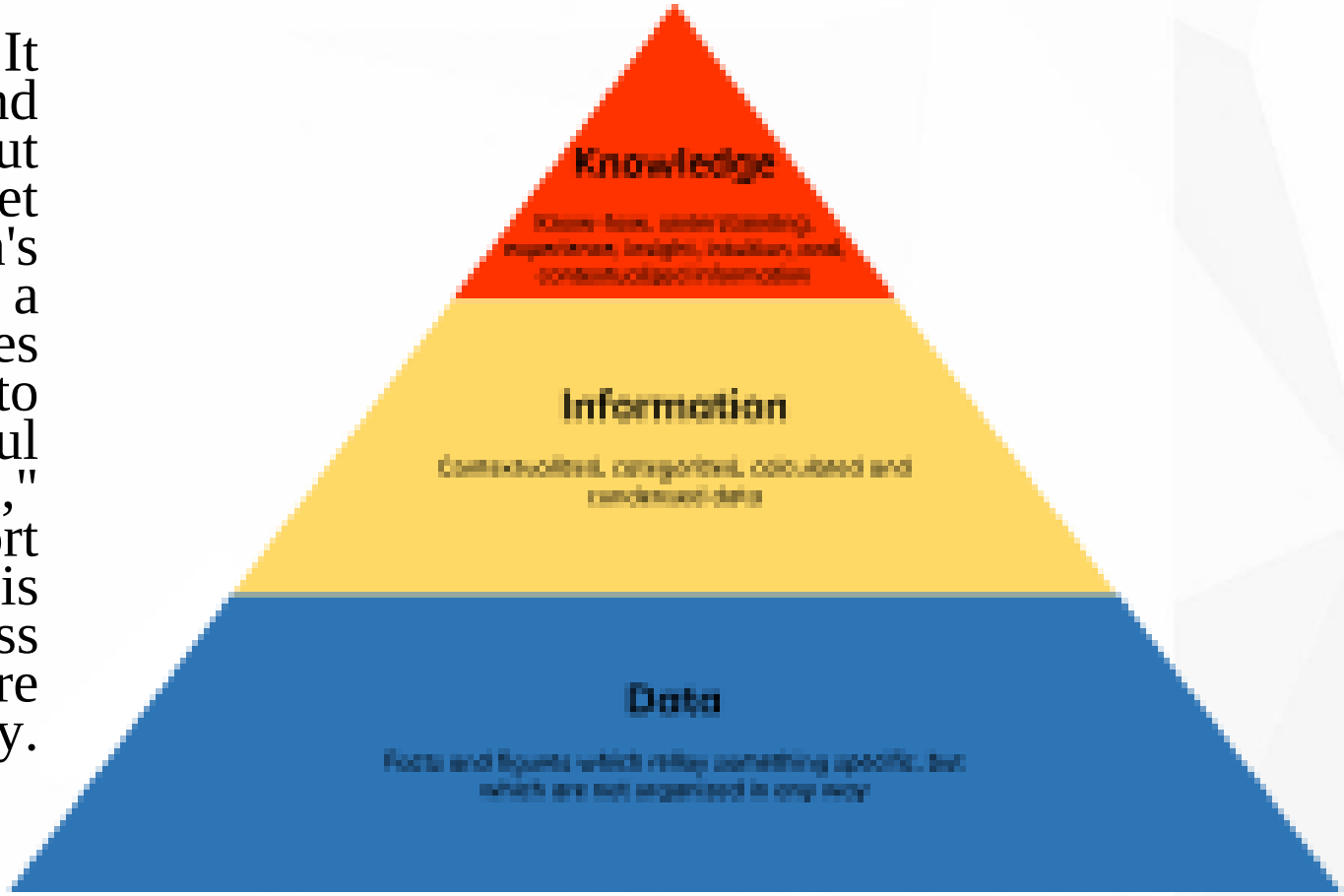
- Survival Count
- Passenger Class Count
- Gender Count
- Embarkation Point Count



Saigon Business School

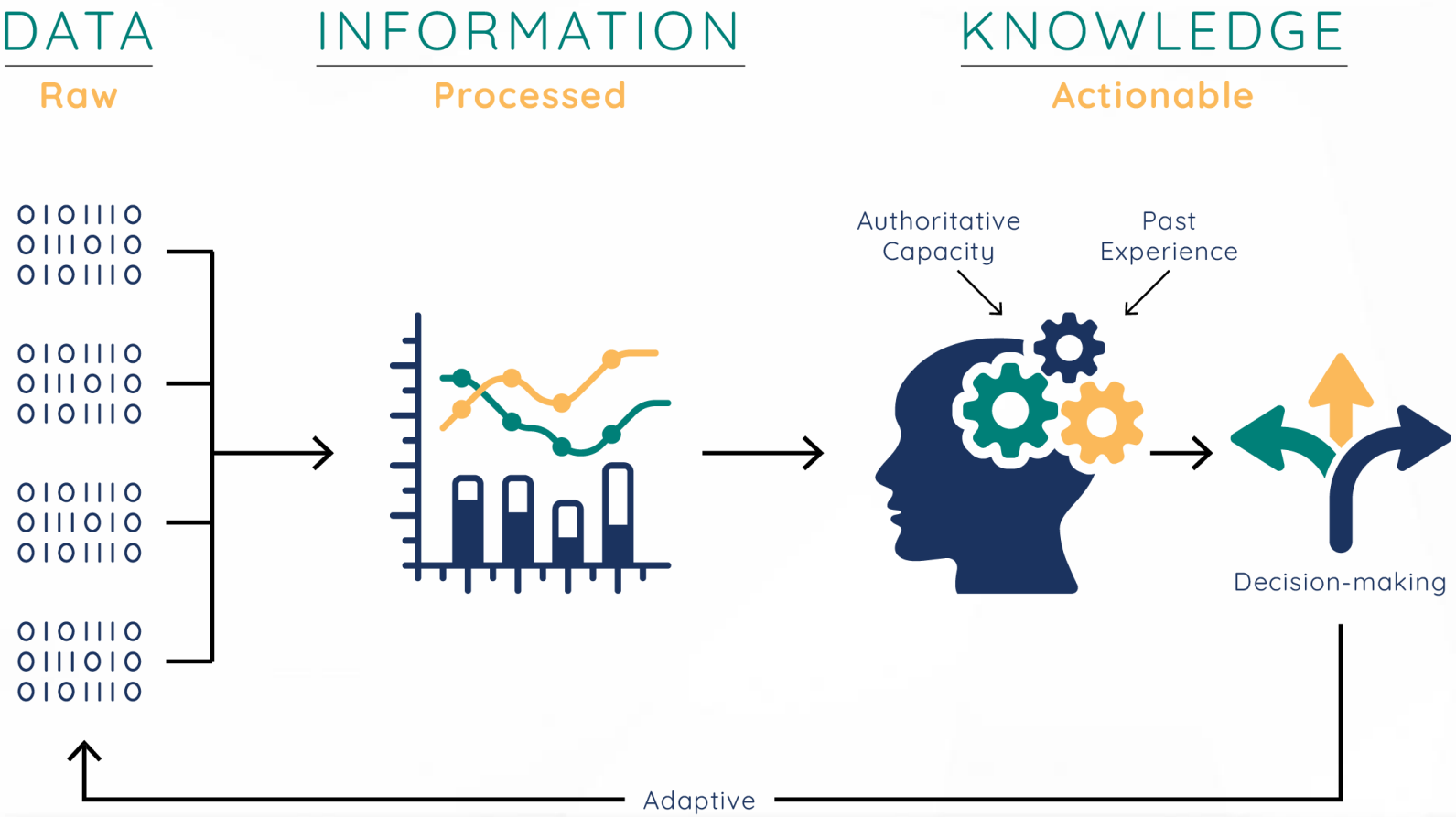
Understanding Data Analytics Terminologies

- There is data all over the place. It would be challenging to comprehend the meaning of this data without understanding how to interpret it, let alone use it to boost an organization's productivity. Data analytics is a collection of procedures that uses various statistical methods to transform data into meaningful information. The term "data analytics," which refers to all decision support and issue solving methodologies, is relatively new. The terms "business analytics" and "data analytics" are sometimes used interchangeably.



Saigon Business School

Understanding Data Analytics Terminologies



Saigon Business School

Data Analytics Basics

- Definition of Data Analytics
  - Data analytics is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision making.
  - Data analytics is often confused with data analysis. While these are related terms, they aren't exactly the same. In fact, data analysis is a subcategory of data analytics that deals specifically with extracting meaning from data. Data analytics, as a whole, includes processes beyond analysis, including data science (using data to theorize and forecast) and data engineering (building data systems).
- Importance of Data Analytics

Saigon Business School

Data Analytics Basics

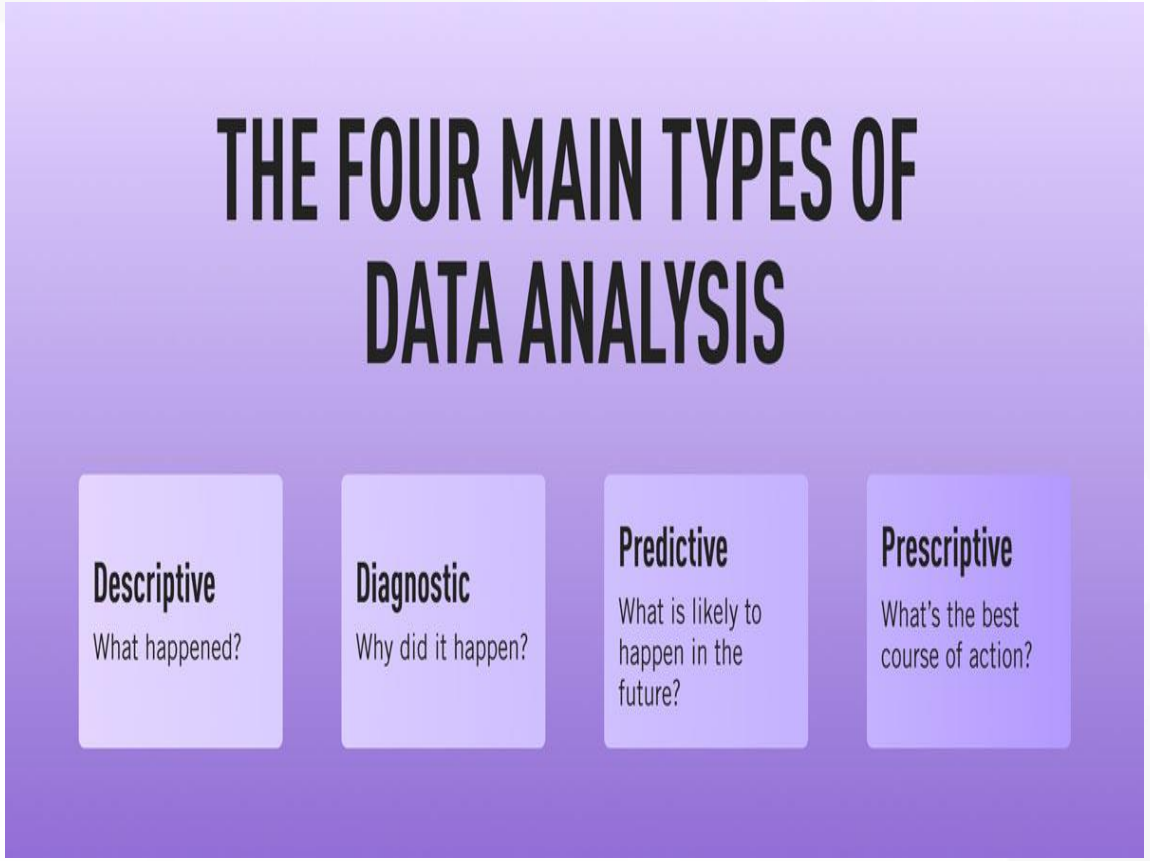
- Importance of Data Analytics

Finding this reference by yourself ?????

Saigon Business School

Data Types

- Population
  - Definition: The entire set of data or individuals under consideration.
- Sample
  - Definition: A subset of the population used for analysis.
- Categorical Data
  - Definition: Data that falls into distinct categories or groups.
- Nominal Data
  - Definition: Categorical data with no inherent order.
- Ordinal Data
  - Definition: Categorical data with a specific order or ranking.
- Continuous Data
  - Definition: Data that can take any value within a range.
- Discrete Data
  - Definition: Data with specific, distinct values.



https://careerfoundry.com/en/blog/data-analytics/different-types-of-data-analysis/

Saigon Business School

Types of Data Analytics

- Descriptive Data Analytics
  - Definition: Summarizing and exploring data to understand its characteristics.
- Predictive Data Analytics
  - Definition: Using historical data to make predictions about future outcomes.
- Prescriptive Data Analytics
  - Definition: Recommending actions based on data to achieve specific goals.

Saigon Business School

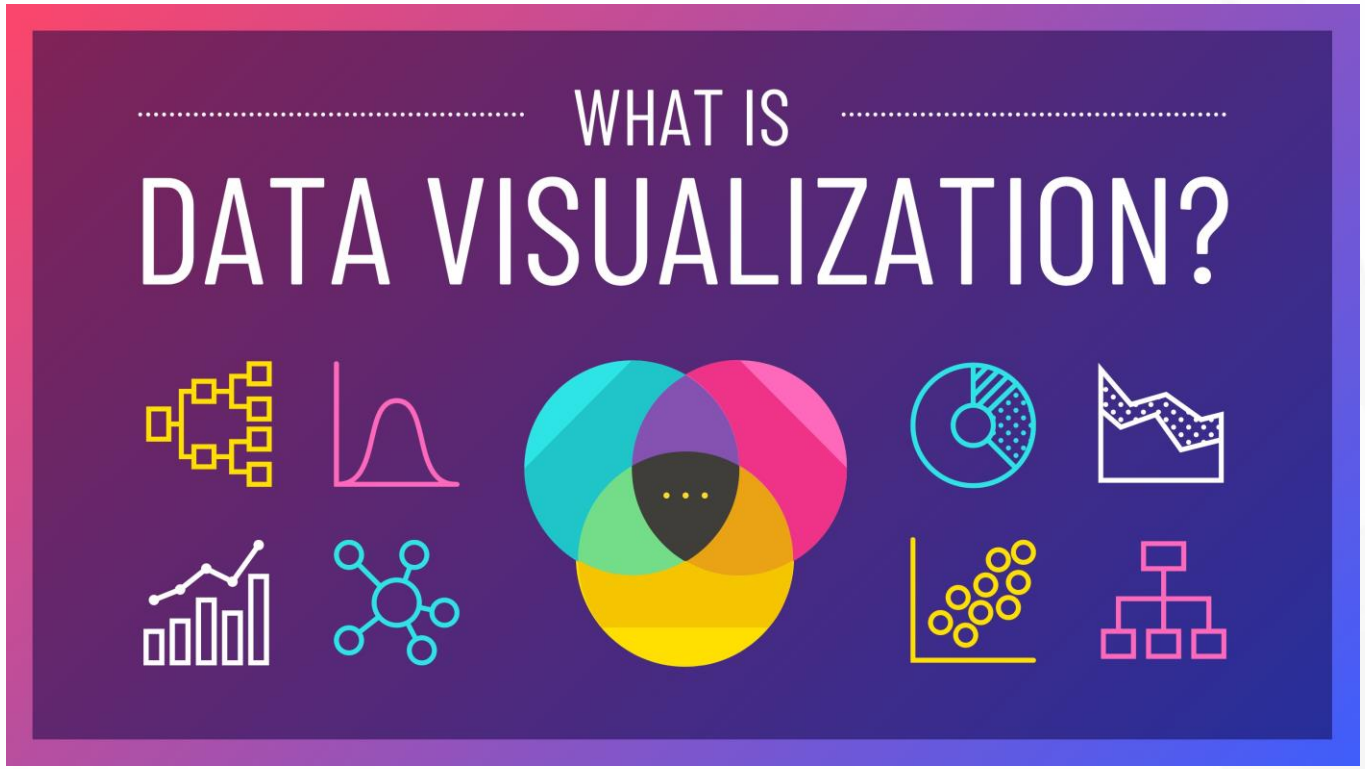
Exploratory Data Analysis (EDA)

- EDA Introduction
  - Definition: The process of analyzing data sets to summarize their main characteristics and uncover patterns.
- Variable Identification
  - Definition: Identifying and understanding the variables in the dataset.
- Univariate Analysis
  - Definition: Analyzing one variable at a time to understand its distribution.
- Bivariate Analysis
  - Definition: Analyzing the relationship between two variables.
- Missing Values Treatment
  - Definition: Strategies for handling missing or incomplete data.

Saigon Business School

Data Visualization

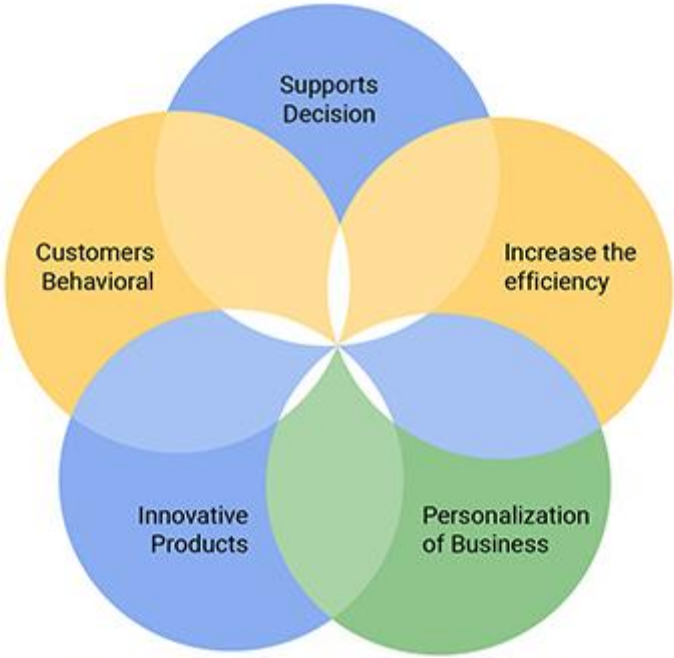
- Data Visualization Introduction
  - Definition: Presenting data in visual forms to facilitate understanding and insights.
- Types of Data Visualization
  - Graphs, Charts, and Plots
    - Bar Charts
    - Line Charts
    - Pie Charts
    - Scatter Plots
    - Histograms
    - Box Plots
    - Heatmaps
- Importance of Data Visualization
  - Enhances data interpretation
  - Communicates insights effectively



Saigon Business School

Benefits of Data Analytics

- Informed Decision-Making
- Improved Business Performance
- Enhanced Competitive Advantage



Top 5 Benefits of Data Analytics in 2023

https://ppcexpo.com/blog/benefits-of-data-analytics

Saigon Business School

Challenges of Data Analytics

- Data Privacy and Security
- Data Quality
- Skill and Resource Constraints

https://intellifysolutions.com/blog/top-5-data-analytics-challenges/

Saigon Business School

Conclusion and Questions

- Overview of data analytics terminologies

- Open questions

Saigon Business School

Assignment 1 (Week 1 – Week 3)

- **Title:** Data Visualization Project
- **Objective:** This project aims to improve your data visualization abilities and show that you can use suitable visual representations to communicate insights from data that are relevant.
- **Task:** You have to choose a dataset (it can be from any discipline; for example, business, medicine, social sciences, etc.) and write an extensive report on data visualization.
- **Deliverables**: A well-documented report in PDF format, all code and scripts used for data preprocessing and visualization and Presentation slides in PowerPoint or PDF format.
- **Assessment Criteria:** based on SBS Assessment Criteria.

Saigon Business School

Assignment 1 (Week 1 – Week 3)

- **Task Details:**

Data Selection and Data Preprocessing.

Data Visualization: Create a variety of data visualizations using tools like Python (with libraries like Matplotlib, Seaborn, …) or any other visualization software you are comfortable with. The types of visualizations may include, but are not limited to:

- ✓ Scatter plots
- ✓ Line charts
- ✓ Bar charts
- ✓ Histograms
- ✓ Box plots
- ✓ Heatmaps
- ✓ Pie charts

Report Document: Write a narrative report that explains your findings and the significance of what you have found and include your visuals. Your story should be understandable, succinct, and perceptive.

Presentation: Write a three- to five-minute presentation outlining the main conclusions and important visuals from your research. This is what you are going to show the class over the last week.

Saigon Business School

Assignment 2 (Week 4 – Week 6)

- **Title:** Data Analytics Project
- **Objective:** This project aims to enhance your data analytics and data mining abilities by using a variety of data mining approaches to extract useful information from a real-world dataset.
- **Task:** It is necessary for you to choose a dataset from the dataset list and carry out a thorough data mining investigation.
- **Deliverables**: A well-documented report in PDF format, all code and scripts used for data preprocessing and visualization and Presentation slides in PowerPoint or PDF format.
- **Assessment Criteria:** based on SBS Assessment Criteria.

Saigon Business School

Assignment 2 (Week 4 – Week 6)

- **Task Details:**

Data Selection and Data Preprocessing.

Utilize data mining approaches in your work. Among the potential methods are, but are not restricted to:

- ✓ Association Rule Mining
- ✓ Classification
- ✓ Clustering
- ✓ Regression
- ✓ Time Series Analysis

Utilize data analytic techniques in your work. Among the potential methods are, but are not restricted to:

- ✓ descriptive analytic techniques
- ✓ predictive analytic techniques
- ✓ Prescriptive analytic techniques

Model Evaluation: Make use of the relevant metrics to assess how well your selected data mining approaches function. Examine and contrast the outcomes.

Interpretation of Results: Describe the learnings from your data mining study and how you may use them to solve a real-world issue or situation.

Report Document: Write a narrative report that explains your findings and the significance of what you have found and include your visuals. Your story should be understandable, succinct, and perceptive.

Presentation: Write a three- to five-minute presentation outlining the main conclusions and important visuals from your research. This is what you are going to show the class over the last week.

Saigon Business School

Thank you