Title: Introduction to Analytics Date:

Topic: Section 4: Additional Topics/Points Continued from:

Key differences between Careers & Roles	Notes
	• **Data Scientist:** Role involves extensive data analysis, model building, and interpretation to derive actionable insights. Utilizes a variety of statistical and machine learning techniques. • **Data Engineer:** Role focuses on designing, constructing, and maintaining the systems and architectures for collecting and storing data. As well as, database management, ETL processes, and ensuring data accessibility. • **Decision Scientist/Analyst:** Role specializes in applying statistical and analytical methods to make data-driven decisions. Often collaborates with business leaders to provide insights for strategic decisions. • **Data Analyst:** Role is primarily responsible for interpreting and analyzing data using statistical techniques. Which provides actionable insights to support decision-making. • **Business Intelligence (BI) Analyst:** Role focuses on analyzing business data to provide insights and support strategic decision-making. As well as, designing, developing and updating BI reports, dashboards, and data visualizations. • **Machine Learning Engineer:** Role concentrates on designing, implementing, and maintaining machine learning models and systems. As well as, extensive knowledge of machine learning algorithms, model training, and deployment.

Summary

Title: Introduction to Analytics Date:

Topic: Section 4: Additional Topics/Points Continued from: Section 4.1

Each role does	Notes
	• **Data Scientist:** Involves Data Collection and Cleaning: Collaborates with Data Engineers to ensure data quality and completeness; Analysis: Leads the analysis phase, employing statistical methods and machine learning models; Interpretation and Presentation: Interprets results, communicates findings, and presents actionable insights. • **Data Engineer:** Involves Data Collection: Plays a crucial role in cleaning and preprocessing data for analysis. Focuses on designing and implementing systems for efficient data collection; Processes: Ensures data is stored efficiently, manages databases, and supports ETL processes. • **Decision Scientist:** Involves Interpretation: Applies statistical methods to interpret data and derive meaningful insights; Presentation: Collaborates with stakeholders to ensure data-driven decisions align with business goals. • **Data Analyst:** Analysis: Are involved in the analysis phase, employing statistical techniques; Interpretation and Presentation: Provides actionable insights and presents findings to support decision-making. • **Business Intelligence (BI) Analyst:** Involves Data Collection and Cleaning: Gathers and ensures the quality of business data for analysis; Analysis: Utilizes BI tools for reporting, dashboard creation, and business performance analysis. • **Machine Learning Engineer:** Involves Analysis: Designs and implements machine learning models for data analysis. Involved in algorithm selection, training, and evaluation; Interpretation and Presentation: Interprets model results, implements machine learning algorithms, trains models, and deploys them for practical use.

Summary

Title: **Introduction to Analytics** Date:

Topic: **Section 4: Additional Topics/Points** Continued from: **Section 4.2**

	Notes
Data Analytics techniques	Includes various statistical and machine learning methods for extracting insights from data.
Regression	Regression is a predictive modeling technique. **Linear regression** predicts numeric outcomes, while **logistic regression** predicts binary outcomes.
Clustering	A technique that **groups similar data points into clusters**, identifying patterns or structures in data.
K-means	A specific **clustering algorithm** that partitions data into K clusters based on similarity.
P-values	A **statistical measure** used to **determine** the **significance** of **observed effects**, commonly **used** in **hypothesis testing**.
PCA	**Principal Component Analysis**, a technique for reducing the dimensionality of data while retaining important information.
Naïve Bayes analysis	A **classification technique** based on the concept of **probability** (Bayes' theorem) and **assigns class labels to instances** based on the possibility of belonging to a particular class.
Oversampling	A **method to balance class distribution** in datasets, especially in machine learning, by duplicating or creating synthetic instances of the minority class.
Customer Life Cycle	Describes the **stages a customer goes through**, from awareness and acquisition to retention and loyalty.

Summary

Title: Introduction to Analytics Date:

Topic: Section 4: Additional Topics/Points Continued from: Section 4.3

	Notes
Subscription models	**Business models** where **customers** pay a **recurring fee** for continued access to a product or service.
Roles of software applications	Refers to the specific functions and responsibilities of software applications in supporting data-related tasks.
Statistical models	Are **mathematical representations** of relationships and patterns in data. These models are used in statistics to describe and analyze the relationships between variables, make predictions, and infer underlying patterns from observed data. Statistical models can be simple or complex, depending on the nature of the data and the questions being addressed.
Imputation	The **process of replacing missing or incomplete data** with estimated values.
Hadoop	A **framework** for distributed storage and processing of large datasets.
Data Wrangler	An **interactive tool** for data cleaning and transformation.

Summary

Title: Introduction to Analytics Date:

Topic: Section 4: Additional Topics/Points Continued from: Section 4.4

	Notes
D3.js	A **JavaScript library** for creating interactive and dynamic data visualizations in web browsers.
Type errors	Statistical errors where **Type I** involves a false positive and **Type II** involves a false negative.
Quantitative vs Qualitative data	**Quantitative** involves numerical data, while ***qualitative*** involves non-numeric data, often categorical or descriptive.
SG&A expenses	**S**elling **G**eneral **and A**dministrative expenses.
Veracity	The accuracy, reliability, and quality of the data collected and analyzed.
Correlation Coefficient	**Measures** the extent of **correlation** between **two variables**, ranging from -1 to 1.
Process mining	Aims to provide insights into actual executions rather than how the executions should be.
ARIMA model	**Autoregressive Integrated Moving Average** model, used for time series analysis after removing trends or seasonality.

Summary