Week 1: Introducing Data Analysis

Core Concepts

- **The Goal:** Data and analysis lead to insights, information, and informed decisions.
- **Data Analyst:** Someone who collects, transforms, and organizes data in order to help make informed decisions.
- **Course Outline:** The course consists of eight steps:
    1. Foundations: Data, Data, Everywhere.
    2. Ask Questions to Make Data-Driven Decisions.
    3. Prepare Data for Exploration.
    4. Process Data from Dirty to Clean.
    5. Analyze Data to Answer Questions.
    6. Share Data Through the Art of Visualization.
    7. Data Analysis with R Programming.
    8. Data Analytics Capstone Project: Complete a Case Study.

Business and Data

- **GAP Analysis:** This is the process companies use to compare their current performance with their desired, expected performance.
- **Why Businesses Organize Data:** Organizations need their data structured to improve processes, identify opportunities and trends, launch new products, serve customers, and make thoughtful decisions.
- **Six Steps of the Data Analysis Process:**
    1. Ask questioning and define the problem.
    2. Prepare data by collecting and storing the information.
    3. Process data by cleaning and checking the information.

4. Analyze data to find patterns, relationships, and trends.
5. Share data with the stakeholders.
6. Act on the data and use the analysis results.

Business Analytics vs. Data Science

- **Business Analytics:** The use of math and statistics to derive meaning from data in order to make better business decisions. The main goal is to extract meaningful insights from data to guide organizational decisions. Business analysts participate in tasks such as budgeting, forecasting, and product development.

- **Four Types of Business Analytics:**
  - **Descriptive:** The interpretation of historical data to identify business trends and patterns.
  - **Predictive:** Uses the above data to forecast future outcomes.
  - **Diagnostic:** Can be used to find out the root cause of a business problem.
  - **Prescriptive:** Analysis done to determine which outcome will yield the best result in a given scenario.

- **Data Science:** Focused on turning raw data into meaningful conclusions through using algorithms and statistical models. It involves creating new ways of modeling and understanding the unknown by using raw data. Data scientists focus on data wrangling, programming, and statistical modeling.

- **The Distinction:** While both processes use big data to solve business problems and are vital to today's organizations, they are separate fields consisting of different functions. There are four examples of how organizations use business analytics to their benefit.

- **Data Analysis Definition:** The collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making.

The Data Environment

- **Data Ecosystem:** The various elements that interact with one another in order to produce, manage, store, organize, analyze, and share data.

- **Cloud:** A place to keep data online, rather than on a computer hard drive.
- **Data-Driven Decision Making:** Defined as using facts to guide business strategy.

Data Life Cycles

Core Six-Stage Data Life Cycle

1. **Plan:** Deciding what kind of data is needed, how it will be managed throughout its life cycle, who will be responsible for it, and the optimal outcome.
2. **Capture:** Collecting data from a variety of different sources and bringing it into the organization. When storing customer data in a database, it is essential to ensure data integrity, credibility, and privacy.
3. **Manage:** Caring for and maintaining the data. This includes determining how and where it is stored, the tools used to keep it safe, and actions taken to maintain it properly.
4. **Analyze:** Using data to solve problems, make great decisions, and support business goals.
5. **Archive:** Storing relevant data in a place where it is still available for long-term and future reference, even if it may not be actively used.
6. **Destroy:** Erasing the data to protect it from unintended users after analysis. Companies use special software to remove data from storage and delete any shared copies.

Life Cycles Used by Specific Organizations and Frameworks

- **Google Data Analytics Certificate Process:** 1. Ask (Business Challenge/Objective) 2. Prepare (Data generation, collection, storage) 3. Process (Data cleaning/integrity) 4. Analyze (Exploration, visualization) 5. Share (Communicating results) 6. Act (Putting insights to work).
- **EMC's Data Analysis Life Cycle:** 1. Discovery 2. Pre-processing data 3. Model planning 4. Model building 5. Communicate results 6. Operationalize.
- **SAS Iterative Life Cycle:** 1. Ask 2. Prepare 3. Explore 4. Model 5. Implement 6. Act 7. Evaluate.

- **Project-Based Data Analytic Life Cycle:** 1. Identifying the problem 2. Designing data requirements 3. Preprocessing data 4. Performing data analysis 5. Visualizing data.

- **Big Data Analytics Life Cycle:** 1. Business case evaluation 2. Data identification 3. Data acquisition and filtering 4. Data extraction 5. Data validation and cleaning 6. Data aggregation and representation 7. Data analysis 8. Data visualization 9. Utilization of analysis results.

- **US Fish and Wildlife Service:** 1. Plan 2. Acquire 3. Maintain 4. Access 5. Evaluate 6. Archive.

- U.S. Geological Survey: 1. Plan 2. Acquire 3. Process 4. Analyze 5. Preserve 6. Publish/share.

- **Financial Institution:** 1. Capture 2. Quality 3. Transform 4. Utilize 5. Report 6. Archive 7. Purge.

- **Harvard Business School:** 1. Generation 2. Collection 3. Processing 4. Storage 5. Management 6. Analysis 7. Visualization 8. Interpretation.

Week 2: All About Analytical Thinking

Analytical Skills and Thinking

- **Analytical Skills:** The qualities and characteristics associated with solving problems using facts.

- **Analytical Thinking:** Identifying and defining a problem, and then solving it by using data in an organized, step-by-step manner.

- **Five Essential Aspects of Analytical Skills:**
    1. **Curiosity:** Wanting to learn something.

    2. **Understanding Context:** Context is the condition in which something exists or happens; understanding it is the process of getting insights from that context.

    3. **Having a Technical Mindset:** The ability to break things down into smaller steps or pieces and work with them in an orderly and logical way.

4. **Data Design:** How you organize information. For a data analyst, this could include database activities.

5. **Data Strategy:** The management of people, processes, and tools used in data analysis.

- **Five Key Aspects to Analytical Thinking:**
  1. **Visualization:** The graphical representation of information.

  2. **Strategy:** Having a strategic mindset helps analysts stay focused on track, seeing what they want to achieve with the data and how to get there.

  3. **Problem-Orientation:** Using a problem-oriented approach to identify, describe, and solve problems, keeping the issue top of mind throughout the project.

  4. **Correlation:** The relationship between two or more pieces of data. Correlation does not equal causation; just because data trends together doesn't mean they are related.

  5. **Big-Picture and Detail-Oriented Thinking:** Being able to see both the big picture and the details.

Problem Solving Strategies

- **How a Data Analyst Hunts a Solution:**
  1. **Find Root Cause:** Discover the reason why a problem occurs using the "five whys".

  2. **Find the Gaps:** Data analysts use GAP analysis to find where the gaps in a process are to get to the solution.

  3. **Consider Missing Information:** Ask "what did we not consider before?" to identify missing information or procedures, aiding in better decision-making.

- **Data-Driven Decision-Making:** Using facts to guide business strategy.

How the Data Analysis Process Guides This Program

- **Ask:** Ask effective questions, define the problem, use structured thinking, and communicate with others.

- **Prepare:** Understand how data is generated and collected, identify and use different formats/types/structures, make sure data is unbiased and credible, and organize and protect data.

- **Process:** Create and transform data, maintain data integrity, test data, clean data, and verify and report on cleaning results.

- **Analyze:** Use tools to format and transform data, sort and filter data, identify patterns and draw conclusions, make predictions and recommendations, and make data-driven decisions.

- **Share:** Understand visualization, create effective visuals, bring data to life, use data storytelling, and communicate to help others understand results.

- **Act:** Apply your insights, solve problems, make decisions, and create something new.

Week 3: The wonderful world of Data

Life cycle of data There are six stages to the data life cycle:

Plan: Decide what kind of data is needed, how it will be managed, and who will be responsible for it. This actually happens before starting the project and includes determining the optimal outcome.

Capture: Collect or bring in data from a variety of different sources. When storing customer data in a database, it is important to ensure data integrity, credibility, and privacy.

Manage: Care for and maintain the data. This includes determining how and where it is stored, the tools used to keep it safe and secure, and the actions taken to make sure it's maintained properly.

Analyze: Use the data to solve problems, make decisions, and support business goals.

Archive: Keep relevant data stored for long-term and future reference. This means storing the data in a place where it's still available but may not be used.

Destroy: Remove data from storage and delete any shared copies of the data. This protects the data from unintended users, and companies use special software to erase the whole data.

Data life cycle following by some organizations:

US fish and wildlife service:

1. Plan, 2. Acquire, 3. Maintain, 4. Access, 5. Evaluate, 6. Archive.

The U.S. geological survey:

1. Plan, 2. Acquire, 3. Process, 4. Analyze, 5. Preserve, 6. Publish/share.

Financial Institution:

1. Capture, 2. Quality, 3. Transform, 4. Utilize, 5. Report, 6. Archive, 7. Purge.

Harvard Business school:

1. Generation, 2. Collection, 3. Processing, 4. Storage, 5. Management, 6. Analysis, 7. Visualization, 8. Interpretation.

Let me know if you'd like me to re-format the full document with this header correctly in place!

Week 4: Set Up Your Toolbox

Data Science Toolbox (DS Toolbox)

- **Spreadsheet:** Microsoft Excel, Google Sheets.
- **SQL:** MySQL, Microsoft SQL Server, BigQuery.
- **Visualization Tools:** Tableau, Looker.

Spreadsheets vs. Databases

Spreadsheets	**Databases**
Software applications	Data stores - accessed using a query language (e.g. SQL)
Structure data in a row and column format	Structure data using rules and relationships
Organize information in cells	Organize information in complex collections

Provide access to a limited amount of data	Provide access to huge amounts of data
Manual data entry	Strict and consistent data entry
Generally one user at a time	Multiple users
Controlled by the user	Controlled by a database management system

Steps to Plan a Data Visualization

Example Scenario: Analyzing data for a clothing distributor updating its website.

- **Step 1: Explore the data for patterns.** Access current sales records and website analytics. Review data to find patterns (e.g., geography and spending amounts among frequent visitors) to explain sales trends.

- **Step 2: Plan your visuals.** Refine the spread-out data. Determine the best visuals for the target audience (e.g., sales-oriented audience needs to see sales over time, sales by location, relationship between sales and website use, and customer growth drivers).

- **Step 3: Create your visuals.** Test different visualization formats and make adjustments. A mix of visuals communicates the best story: line charts for sales over time, maps for location connections, donut charts for customer segments, and bar charts to compare visitors to purchases.

Data Visualization Toolkit

- **Spreadsheets (Microsoft Excel or Google Sheets):** Great for simple visuals like bar

graphs, pie charts, maps, waterfall, and funnel diagrams.

- **Visualization Software (Tableau):** A powerful, popular tool that pulls data from nearly any system to create actionable insights and interactive dashboards. Tableau Public is free and includes resources like how-to videos, practice datasets, and the "Viz of the Day" page for inspiration.
- **Programming Language (R with RStudio):** Data analysts often use R within the RStudio integrated developer environment (IDE) for dashboard-style visualizations. Resources include RStudio Cheatsheets, the Visualize Data Primer, webinars, and videos.
- *Note:* The choice of tool depends on factors like data size and the analysis process used.

Week 5: Endless Career Possibilities

Key Terminology

- **Issue:** A topic or subject to investigate.
- **Question:** Designed to discover information.
- **Problem:** An obstacle or complication that needs to be worked out.
- **Business Task:** The question/problem data analysis answers for a business.
- **Data-Driven Decision-Making:** Using facts to guide business decisions.
- **Fairness:** Ensuring that your analysis doesn't create/reinforce bias.

Decoding the Job Description

The data analyst role shares similarities with other titles, so it is important to read job descriptions carefully as companies often blur the lines between them.

- **Business Analyst:** Analyzes data to help businesses improve processes, products, or services.
- **Data Analytics Consultant:** Analyzes the systems and models for using data.

- **Data Engineer:** Prepares and integrates data from different sources for analytical use.
- **Data Scientist:** Uses expert skills in technology and social science to find trends through data analysis.
- **Data Specialist:** Organizes or converts data for use in databases or software systems.
- **Operations Analyst:** Analyzes data to assess the performance of business operations and workflows.

Comparison of Key Roles

Feature	**Data Analysts**	**Data Scientists**	**Data Specialists**
Problem Solving	Use existing tools and methods to solve problems with existing types of data.	Invent new tools and models, ask open-ended questions, and collect new types of data.	Use in-depth knowledge of databases as a tool to solve problems and manage data.
Analysis	Analyze collected data to help stakeholders make better decisions.	Analyze and interpret complex data to make business predictions.	Organize large volumes of data for use in data analytics or business operations.
Other Relevant Skills	Database queries, Data visualization, Dashboards, Reports, Spreadsheets.	Advanced statistics, Machine learning, Deep learning, Data optimization, Programming.	Data manipulation, Information security, Data models, Scalability of data, Disaster recovery.