📊 Group Project

Project A: Collecting Data on Extinct Animals Due to the Impact of Global Warming

The main objective of this project is to gather data before and after 1900 on the various species that have gone extinct in order to analyze the impact of climate change on these animals and determine the most affected species. To collect data on extinct animals, you need to utilize various sources such as biodiversity databases, academic papers, government reports, and conservation organizations' data.

As a group, complete the following steps:

1. Selection of data collection method. The students will determine the most appropriate data collection method based on the research topic and available resources. This could include observational techniques, transactional data, surveys, experimental data, or web scraping. The method chosen should allow for the collection of reliable and relevant data on extinct animals.

2. Independent research on reliable sources of data. Each student will conduct independent research to locate credible sources of data on animal extinctions.

3. Data organization into a table format. Using the collected information, students will create a table to present data on the distribution and causes of animal extinctions per continent. The table will include columns for the animal's common name, scientific name, date of extinction, continent, and primary reason for extinction. Table 2.10 provides an example.

Common	Scientific	Date of	Contine	Primar Reason Extinctio
Passenger	Ectopistes migratorius	191	North	Overhuntin
Quagg	Equus quagga	188	Afric	Habitat

Table 2.10 Project A Required Data and Example

4. Collaborative data cleaning. Students will work together to clean the collected data, identifying and addressing any missing or duplicate information. They will also ensure that the data is standardized, allowing for accurate analysis and comparison.

5. Data storage determination. In this step, the students will determine the storage size of the collected

2 • Critical Thinking

data, taking into consideration the amount of information gathered from various sources. They will also decide on the most efficient and accessible way to store the

data, ensuring all group members have access to it for further analysis and collaboration.

6.   Source citation. As part of the data collection process, students will document and list all the sources used to gather information on extinct animals. This will ensure proper credit is given to the original sources and allow for further verification of the data.

Project B: Data Collection Using Experimental Technique

Your group project will involve recording temperature changes at a specific location over the course of 24 hours. This project will be completed in subgroups, with each group member responsible for recording data during a specific time segment of the day. Before beginning your research, your group will select a location to observe. (This could be a park, a city street, a backyard, or any other outdoor location. It is important that the location is accessible and allows for consistent monitoring throughout the day.) Each subgroup will be responsible for recording temperature data for a six-hour time period. The first subgroup will record data from midnight to 6 a.m., the second subgroup from 6 a.m. to 12 p.m., the third subgroup from 12 p.m. to 6 p.m., and the fourth subgroup from 6 p.m. to midnight.

It is crucial that each subgroup records data at the exact same time each day in order to maintain consistency. During the data collection process, it is important for each subgroup to maintain the same surrounding conditions. This includes using the same tools, such as a thermometer, and staying in the same location throughout the time period. Any uncontrolled changes or errors should be noted and recorded to ensure accuracy in the data.

Repeat this process for seven days to gather a sufficient amount of data, and once all the data has been collected, calculate the average temperature for each time segment of the day by adding up all the recorded temperatures and dividing by the number of data points.

Once all data has been collected, the group will come together to analyze and graph the temperature changes over the course of 24 hours. This will allow for a visual representation of the temperature patterns throughout the day.

Discuss potential sources of error and ways to improve the data collection process in the future. Then summarize the collected data process, present the project to the rest of the group, and discuss any insights or lessons learned from the data collection process.

Critical Thinking

1.   Megan is interested in opening a gaming center that utilizes virtual reality (VR) technology. She wants to gauge the level of interest in VR gaming among the

residents of her city. She plans to conduct a survey using a random sampling method and then use this data to determine the potential demand for a VR gaming center in her city. What is the nature of the collected data, experimental or observational?

2. A high school student is interested in determining the relationship between the weight of an object and the distance it can travel when launched from a catapult. To collect data for analysis, the student plans to launch objects of varying weight and then measure the corresponding distance traveled. Using this data, the student plans to plot a graph with weight on the x-axis and distance on the y-axis. a. Will this data collection result in observational data or experimental data?

b. How can the student gather the necessary data to establish this relationship?

3. A high school is conducting a survey to understand the opinions of the school's students toward the food options available in the cafeteria. They have divided the student population into four groups based on grade level (9th, 10th, 11th, 12th). A random sample of 60 students is selected from each group and asked to rate the cafeteria food on a scale of 1–10. The results are shown in the following table:

2 • Critical Thinking

Grade	Number of	Average
9t	6	8.
1 t	6	7.
1 t	6	6.
1 t	6	6.

Table 2.11 High School Survey

What sampling technique was used to collect this data?

4. A research study is being conducted to understand the opinions of people on the new city budget. The population consists of residents from different neighborhoods in the city with varying income levels. The following table provides the data of 130 participants randomly selected for the study, categorized by their neighborhood and income level.

Neighborhod	ncome	Number of
Northsid	Low	3
Southsid	Middle	5
Eastsid	High	4

Table 2.12 Neighborhood and Income level Category

What sampling technique was used to collect this data?

5. John is looking to gather data on the popularity of 3D printing among local college students to determine the potential demand for a 3D printing lab on campus. He has decided on using surveys as his method of data collection. What sampling methods would be most effective for obtaining accurate results?

6. A group of researchers were tasked with gathering data on the price of electronics from various online retailers for a project. Suggest an optimal approach for collecting data on electronics prices from different online stores for a research project.

7. As the pandemic continues to surge, the US Centers for Disease Control and Prevention (CDC) releases daily data on COVID-19 cases and deaths in the United States, but the CDC does not report on Saturdays and Sundays, due to delays in receiving reports from states and counties. Then the Saturday and Sunday cases are added to the Monday cases. This results in a sudden increase in reported numbers of Monday cases, causing confusion and concern among the public. (See the table provided.) The CDC explains that the spike is due to the inclusion of delayed or improperly recorded data from the previous week. This causes speculation and panic among the public, but the CDC reassures them that there is no hidden outbreak happening.

Dat	Da	New
10/18/202	Monda	311
10/19/202	Tuesda	484
10/20/202	Wednesda	394
10/21/202	Thursda	482
10/22/202	Frida	435
10/23/202	Saturda	0
10/24/202	Sunda	0
10/25/202	Monda	857
10/26/202	Tuesda	446
10/27/202	Wednesda	532
10/28/202	Thursda	501
10/29/202	Frida	471
10/30/202	Saturda	0
10/31/202	Sunda	0
1 /1/202	Monda	1041
1 /2/202	Tuesda	509
1 /3/202	Wednesda	688
1 /4/202	Thursda	540
1 /5/202	Frida	675
1 /6/202	Saturda	0
1 /7/202	Sunda	0

Table 2.13 Sample of COVID-19 Data Cases Within 23 Days (source: https://data.cdc.gov/Case

Surveillance)

1 /8/202	Monda	1006
1 /9/202	Tuesda	529

Table 2.13 Sample of COVID-19 Data Cases Within

23 Days (source: https://data.cdc.gov/CaseSurveillance)

a. What simple strategies can the data scientists utilize to address the issue of missing data in their analysis of COVID-19 data from the CDC?

b. How can data scientists incorporate advanced methods to tackle the issue of missing data in their examination of COVID-19 cases and deaths reported by the CDC?

8. What distinguishes data standardization from validation when gathering data for a data science project?

9. Paul was trying to analyze the results of an experiment on the effects of music on plant growth, but he kept facing unexpected data. He was using automated instrumentation to collect growth data on the plants at periodic time intervals. He found that some of the data points were much higher or lower than anticipated, and he suspected that there was noise (error) present in the data. He wanted to investigate the main source of the error in the data so that he could take corrective actions and thus minimize the impact of the error(s) on the experiment results. What is the most likely source of error that Paul might encounter in this experiment: human error, instrumentation error, or sampling error? Consider the various sources of noise (error).

10. The government is choosing the best hospital in the United States based on data such as how many patients they have, their mortality rates, and overall quality of care. They want to collect this data accurately and fairly to improve the country's health care.

a. What is the best way to collect this data? Can we use web scraping to gather data from hospital websites? Can we also use social media data collection to gather patient reviews and feedback about the hospitals?

b. How can we ensure the data is unbiased and accurate? Can we use a mix of these methods to create a comprehensive analysis?

11. Sarah, a data scientist, was in charge of hiring a new sales representative for a big company. She had received a large number of applications. To make the hiring process more efficient, she decided to streamline the process by using text processing to analyze the candidates' applications to identify their sentiment, tone, and relevance. This would help Sarah gain a better understanding of their

communication skills, problem-solving abilities, and overall fit for the sales representative role. She compiled two lists of keywords, one for successful candidates and one for rejected candidates. After generating summaries of the keywords from each application, Sarah was ready to make informed and efficient hiring decisions. Identify the best candidate based on the list of keywords associated with each candidate, as reviewed in the figure shown. Your answer should be supported with strong reasoning and evidence.

2 • References

Candidates	Customer-Centric	Team Player	Collaborative	Sales Experience	Leadership	Multitasking	Goal-Oriented	Team-Oriented	Effective	Inflexible	Uncooperative	Indecisive	Negative	Lacking	Limited
John	✓	✓	✓	✓	✓										
Sarah		✓	✓	✓			✓		✓						
Lin			✓	✓										✓	✓
Aysha	✓				✓					✓	✓	✓	✓	✓	
Alex		✓		✓			✓		✓	✓					
Peter	✓	✓	✓	✓						✓	✓		✓	✓	
Jamal	✓	✓		✓		✓	✓	✓							
Brian	✓	✓	✓	✓	✓	✓	✓					✓		✓	
Samantha	✓	✓	✓	✓	✓		✓	✓	✓						
Miguel					✓	✓	✓	✓	✓	✓	✓	✓			

Figure 2.6 Data Collection Based on Text Processing

12. How can a data analyst for a large airline company efficiently collect and organize a significant amount of data pertaining to flight routes, passenger demographics, and operational metrics in a manner that ensures reliable and up-to-date records, complies with industry standards, and adheres to ethical guidelines and privacy regulations? The data will be utilized to identify potential areas for improvement, such as flight delays and high passenger traffic, so that the company can implement effective strategies to address these issues.

13. A chain of restaurants that serve fast food needs to improve their menu by providing healthier options to their customers. They want to track the nutrition information of their dishes, including burgers, fries, and milkshakes, to make it easier for the managers to compare their calorie, fat, and sugar content. What data management tools can restaurants utilize to track nutritional data to make healthier options readily available and enhance their menu?

14.The NFL team's talent acquisition specialist is on a mission to find the most talented high school football player to join their team. The specialist is responsible for gathering extensive information on potential players, such as their stats, performance history, and overall potential. The objective is to pinpoint the standout athlete who will bring a significant advantage to the NFL team. To achieve this, the specialist utilizes cloud computing to sift through vast amounts of data and identify the ideal high school player who can make a valuable contribution to the team's success in the upcoming season. Why is cloud computing the perfect tool for collecting and analyzing data on high school football players?