BRIGHT-CS

Data Science Learning Experience

Acknowledgements

This project is funded by a grant from the National Science Foundation to the University of North Carolina Greensboro (NSF Award # 1752436). Dr. Ryoko Yamaguchi is the principal investigator. Dr. Cyntrica Eaton is the primary author, workbook designer, and video producer. Adam Hall served as the editor.

Any opinions, findings, and conclusions or recommendations expressed in this report are those of the author and do not necessarily reflect the view of the National Science Foundation.

Table of Contents

To access the videos that accompany each unit, click on the corresponding icon: 

Preface: A Word to Parents, Caretakers, and Teachers

According to Eric Schmidt, former CEO of Google, in a quote from 2010[1]: "There were 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every 2 days." While the numbers have definitely shifted over time, the underlying concept still holds true: large volumes of data are collected on a daily basis at unprecedented and rapidly increasing rates. From education to entertainment to healthcare and beyond, the goal of leveraging raw data to supercharge sense-making, decision-making, and discovery is critical in nearly every sector. As Atul Butte, Biomedical Researcher and Professor at the University of California-San Francisco said, "Hiding within those mounds of data is the knowledge that could... change the world."

The effort to collect more data at faster rates introduces a myriad of exciting opportunities to innovate in data capture, storage, analysis, and visualization. To capitalize on these opportunities, advance the development of forward-leaning approaches, and harness the power of data, an adept talent pool *must* be nurtured and empowered to take up the mantle. This will only be possible if young people from all walks of life are given an opportunity to grow their interest in the field, expand their knowledge base, build the appropriate skills, and are invited to fully engage in this space.

The **Data Science Learning Experience** module was created to address this need by helping K-12 students, 6th grade and higher, get an introductory look at the power of Data Science, apply base-level concepts, and develop foundational skills. In addition to providing vital exposure to a dynamic field, this module also introduces students to a host of tools and techniques they can use long after they've completed this learning experience. Data Science – which stands at the intersection of Computer Science and Mathematics – provides an organic platform for not only discussing statistics (a key Common Core Standards topic) but applying concepts in statistics to a problem-solving context and using the outcomes to derive meaning. In addition, this module provides a seamless launch pad for expanding students' awareness of a common data analytics tools that is – more often than not – right at their fingertips: Microsoft Excel.

As a parent, caretaker, or teacher, one of your most critical roles is helping young people connect their strengths and interests to viable careers. This module was created to help you in that regard by providing hands-on learning materials that activate guided discovery, help students develop baseline competencies, and allow students to determine whether their interests align with Data Science. In addition, beyond the technical focus of each module, students are also provided opportunities to enhance other skills including communicating ideas and presenting their findings. This module empowers students to begin exploring a career in Data Science via active discovery of the field's professional outlook and academic preparation requirements.

We invite you to review the materials in this series, get an idea for what students will learn, and observe how the materials are presented. As you do, think about young people who could benefit from this module and share as widely as you can. After introducing the materials, also follow up with students to get a feel for their progress and encourage them to share the PowerPoint presentation they will develop as a result of their work. For the students, learning the technical concepts is important but, by having them discuss what they've learned with you, they can gain vital experience communicating new concepts, receive the support necessary to realize their potential, and be encourage to maximize their growth.

Let's work together to ensure that the next generation of Data Scientists and Engineers have access to the learning experiences and materials they need to gain critical exposure to new ideas and fields; discover their capabilities early on; and build the foundation necessary to thrive in a fulfilling, high-demand career.

[1] https://www.forbes.com/sites/ciocentral/2012/04/24/the-web-is-much-bigger-and-smaller-than-you-think/

Preface: A Word to Students

A defining and fascinating aspect of the electronic age is the ability to capture, store, and process vast amounts of data. This capability has led to exciting discoveries and critical insights about the patterns, outliers, and trends that significantly enhance our understanding of the world around us and the people in it. With the proper keys to unlock the insights stored in data, we can better understand critical factors about the past, learn more about our present, and predict key aspects about the future. Yet, to enhance these capabilities, enrich analysis, and truly realize the power of data, we must have the right tools, techniques, and **people** prepared to address these issues.

That's where **you** come in! As our future scientists, engineers, technologists, and professionals, it is critical that you have the exposure necessary to identify careers that align with your aspirations as early as possible. Recognizing careers that fit your interests and gaining awareness of how to properly prepare for those careers can significantly amplify your future success. Early exposure is the key to growing skills, gaining access to the tools necessary to strengthen your innate capabilities, and preparing academically for the career pathway ahead. Yet, in some cases, the learning materials that can assist you in gaining critical career-focused exposure aren't always available – especially in Computer Science. The **Data Science Learning Experience** was created to fill this gap and help you explore base level data analytics and statistics concepts, gain exposure to key ideas, and apply key skills in a real-world context.

A critical factor about Data Science is that its importance spans a wide variety of sectors. As John Elder, a Data Scientist said[2], "Learning from data is virtually universally useful. Master it and you will be welcomed anywhere." This means that you can build your Data Science skills and then apply them in just about any area you choose including medicine, entertainment, sociology, criminal justice, and others. A career as a Data Scientist would allow you to merge your interest in a variety of fields with in-demand technical capabilities and make a significant impact on sense making, decision-making, and discovery as a result.

The **Data Science Learning Experience** provides you with foundational information and support throughout the process that can supercharge your awareness and understanding. But, to take full advantage of the short- and long-term benefits of this rich learning experience, be sure to do the following:

- **Use this workbook as a companion to your learning experience.** The workbook provides an introductory overview of topics, space for notes, exercises that deepen learning, access to additional resources, and opportunities to practice key concepts and skills.
- **Access and review the accompanying YouTube videos.** Clips that support and reinforce the learning process have been developed for select units. Each video is 7 minutes or less and allows you to customize the pace of your learning experience since you can watch, pause, and rewind as much as and whenever you need to.
- **Enlist a friend or a group of friends to go on this journey with you.** Learning Data Science is an engaged activity. Having friends, siblings, or classmates to experience this with you can help to make the learning process that much more enjoyable and help to keep you on track and accountable. Consider asking someone to go through the process with you and then work together to complete the series.

By using the full suite of resources provided here, completing each unit – either by yourself or as a team – and working through the exercises, you will experience overlapping benefits. Completing this journey will teach you how to retrieve and explore data sets, pose and answer interesting questions, apply mathematical concepts and make key discoveries. Over the course of this project, you will become familiar with fundamental statistics and receive guidance on using Microsoft Excel as a data analytic tool. Moreover, you will gain the experience, training, and support you need to expand awareness and start forming an opinion about whether a career in Data Science is right for you.

[2] https://www.sjsu.edu/pdp/programs/msda/

Unit 1: Mapping Out the Road Ahead and Preparing for Growth (Part I)

Welcome to the **Data Science Learning Experience**! This learning module is designed to help you unlock the power of data by building knowledge through hands-on activities. Around the world, every day, large volumes of data are collected at unprecedented and rapidly increasing rates. The need to explore this data for critical insights continues to grow and take on amplified significance. The value of this data and the people who have the skills to make sense of it increases with each passing year. As the world becomes more data-focused and the value of data-driven insight increases, it is important that all pre-college students are provided the base level capability to access, investigate, and draw conclusions from data.

Recognizing the demands of the time, this module will help you build key data skills and learn what it means to be a data professional while also refining your communication skills. The driving goal here is to provide a well-rounded learning experience that can help you enhance essential skills, learn if Data Science interests you, and fill in the blanks about the pathway towards a career in data.

Data Science sits at the intersection of Computer Science and Mathematics. It is not only a hot topic with important implications in modern times, it also provides a natural launch pad for introducing, applying, and making sense of statistical concepts at the K-12 level. In short, Data Science is not just an important and expanding field, but mastering its core concepts will allow you to translate skills learned in the classroom into real-life applications. This ability can empower you to explore topics that interest you in ways that reveal fascinating insights. The world is expected to become more and more data-driven; knowing how to access large and diverse data sets, perform analysis to extract meaning, and leverage that meaning to answer questions is vital for today's students entering tomorrow's workforce.

One of the most intriguing things about Data Science is its wide-ranging relevance and applicability. Any interest you may have in medicine, finance, entertainment, economics, and beyond can be merged with Data Science skills in a way that advances our collective understanding and leads to critical and potentially lucrative discoveries. Data Science is one field where having multiple interests is definitely a plus.

The main purpose of this unit is to provide you with a preview of the journey ahead. As a result, an overview of (1) the ***pathway*** we will take, (2) the ***tools*** we will use, and (3) a way to ***track the growth*** you will experience once you've completed the module. Capturing what you know and think *before* starting the journey and comparing it to what you have learned *after* completing the module is a powerful way to track your growth and learning. We will start that process in this unit.

The Pathway

The **Data Science Learning Experience** pathway is largely shaped by its goals which include introducing Data Science, providing an overview of fundamental concepts in statistics, and fostering hands-on practice in both exploring and making meaning of data. The module also includes links to resources that can further your understanding of Data Science beyond this module and support you in continued discovery.

To solidify your vision of what's ahead, the table below highlights how we will navigate key parts of the learning experience along with the corresponding units.

Key Topic	Learning Goal(s)	Unit(s)
Data Fundamentals	Understanding base aspects about Data Science and Data Analytics	**2 & 3**
Planning and Executing Analysis	Developing a data analysis plan and executing it on a data set	**4**
Data Science As a Career	Exploring the factors key to Data Science as a career	**5**
Next Steps	Presenting your work and looking toward a future in Data Science	**6 & 7**

Table 1: Overview of the steps covered in this learning experience

The Tools

We will use Microsoft Excel to perform analytics because of its accessibility and its capabilities. Excel is usually widely available on personal, school, and library computers, making it widely accessible. It also has built-in features that allow you to perform key analyses. Excel also provides support and has a large user community that that can be helpful to new or learning users. Since having data is critical to using Excel, curated data set have also been developed for this module.

We will also use tools to manage our learning experience and maximize our learning outcomes. These tools include the Need to Know (NTK) Question Tool, which will help keep track of questions that arise as you navigate the module, and the KIMS Tool which you will use to track any new words or concepts encountered. As an important note, KIMS is an acronym for keyword (**K**), important information (**I**), memory clue (**M**), and sentence (**S**). Combined, the NTK and KIMS tools will not only underpin progress in this learning experience, they will also provide a launch pad for further discovery beyond the completion of this module. Identifying new questions to explore and expanding your vocabulary are natural outcomes of learning activities like these and keeping a record of each can have a notable impact beyond this project.

Finally, an important aspect of any discovery process is presenting your findings and sharing your experience with others. As a result, this module includes a special section on preparing for and structuring a Data Science talk. As a result, PowerPoint is a key tool of interest. A comprehensive list of the tools you will use to support you throughout this learning experience are provided below:

- Microsoft Excel
- Curated data sets
- Need to Know (NTK) Question Tool
- KIMS Tool
- Data Analysis Planning Tool
- PowerPoint
- This workbook
- The YouTube videos that accompany Units 2 - 7
- Pen and paper

As you use the tools listed, get acquainted with each and feel free to re-use them in the future as you explore other data sets of your choosing. Understanding how to apply tools is key to your growth and learning.

The Transformation

Growth in your knowledge, understanding, ability, and interests is a key goal of the **Data Science Learning Experience** – having a way to track your progress is important. As you navigate through this module, you will learn new concepts, apply new skills, tap into your curiosity, and build a knowledge-skill base that you can use to turn raw data into new discoveries. In addition to learning more about Data Science, you will either develop or refine your Excel skills, your understanding of statistics, and your experience applying fundamental statistics using a real-world data set. Capturing your reasoning process and tracking your knowledge growth over time is critical to helping you understand and improve your own problem-solving approaches. This, in turn, will help you better use Data Science concepts and tools.

The next page contains a list of 8 questions designed to get a snapshot of your start point, before you start the learning experience. **Answer the questions as best you can** before moving on to Unit 2. Keep in mind that it is okay if you do not know all of the answers at this point. After you complete the learning experience in Unit 8, you will revisit these same questions to see how much you've learned. For now, simply answer the questions that follow, without looking anything up, to capture your starting state of knowledge and understanding. Happy learning!

Track Your Progress: Take a look at the questions below and answer each, as best as you can, based on what you know right now. You will answer the same set of questions at the end, in Unit 8, to see how far you've come over time.

1. **What is data? What are two examples of datum?**

2. **What do you know about Data Science and data analysis?**

3. **In statistics, what is the *mean*?**

4. **In statistics, what is the *median*?**

5. **In statistics, what is the *mode*?**

6. **What aspects of your personality or interests might serve you well as a Data Scientist?**

7. **On a scale from 1 – 10: rate your understanding of Data Science.**

8. **On a scale from 1 – 10: rate your interest in Data Science.**

Congratulations! You have finished Mapping Out the Road Ahead and Preparing for Growth (Part I). Let's move on to Data Science Basics.

Unit 2: Data Science Basics

The core theory of Data Science is this: by having the right data set, asking the right questions, and using the right analysis techniques, we can learn interesting facts, derive meaning from the analysis, and use those insights to make improved, more informed decisions in the future. The realization that data can be used to predict outcomes, draw conclusions, and stimulate informed inquiry has supercharged the demand to collect more data at faster, unprecedented rates. This, in turn, introduces exciting opportunities to innovate in data capture, storage, analysis, and visualization. Yet, to be in the position to capitalize on these opportunities, meet the needs of steadily evolving data requirements, and truly harness the power of data, we must establish a solid platform by building our knowledge from the ground up.

In this unit, we will explore data analysis and lay a foundation for the exploration and discovery that will follow in the units to come. We'll begin by discussing fundamental ideas about data, the problem-solving aspects inherent to data analysis, the importance of curiosity in these endeavors, baseline ideas about statistics, and insight about locating data stores of interest. In the next unit, we will build on this discussion by applying some of the key ideas discussed in an actual analytics tool: Microsoft Excel.

What Is Data?

Given its importance and prevalence in modern-day society, the word data is used *a lot*. In fact, it's used so much, you may feel that you already know its definition. Yet, given the focus of this module, it is important that we have a clear understanding of what *data* is and move forward in our discussions with an agreed-upon definition. With this in mind, take a second to write your definition of the word data in the box below – without looking at any references – just off the top of your head:

Use the space below to write your own definition for data.

Now for our definition: data is simply information. It can take the form of numbers, text, images, and many other formats. One important fact to keep in mind: the word *data* is actually a pluralized version of the word *datum*.

This exercise was important because it drives home the importance of having a clear understanding of core terminology associated with a project. Because definitions are typically considered elementary, they are often overlooked as an integral part of the exploration process. In some cases, people skip this step altogether because they believe they already have a clear understanding of what the key definitions are. Yet, unless you are an expert in a given field with deep and expansive knowledge, starting with definitions is actually a great way to familiarize yourself with a field, warm up your exploratory muscle before taking a deep dive, and breaking ground for further discovery.

Now that we have a firm understanding of its definition, the next thing to consider when thinking about data is to identify examples. Consider this: **what data can you capture or take note of in the room you're currently sitting in**. Write down three examples in ***Table 2*** on the next page:

What are some data points you can capture about the room you are sitting in? Write three data points below.	
1	
2	
3	

Table 2: Three data examples

In practice, anything you can see, count, or feel in the room you're in can be considered a data point. This includes factors like the temperature, the number of people present, the number of chairs, the average age of the people in the room, and etc. In alignment with our agreed upon definition, each of these data points provides some *information* about your space and, with the appropriate analysis, can be used to derive meaning.

The Data Science Process: Applied Problem-Solving

Although Data Science itself may be new to you, if you're familiar with the problem-solving process, you have a strong basis on which to build understanding. In problem-solving in general, you start with a question, develop a plan to arrive at a solution by identifying the tools and steps necessary, implement the plan, and – once the solution is achieved – evaluate it to ensure that it adequately addresses the original question.

For Data Science, much like with problem-solving in general, the idea is to start out with a question and find your way to an answer through a series of logical steps. What's critical to this particular approach to problem-solving is the focus on accessing, exploring, and analyzing data to identify solutions.

Navigating through a data set to make sense of it and extract meaning from it is an exploratory process that can provide a lot of insight. For Data Science, the problem-solving process typically takes on the following steps:

1. Formulating questions
2. Finding appropriate data sets
3. Cleaning/formatting the data
4. Analyzing the data
5. Presenting findings

The key take-away here is that Data Science is very much a problem-solving activity and understanding how to successfully navigate from 'formulating questions' to 'presenting findings' will sharpen your ability to solve problems and tackle challenges in the future.

The Power of Creative Curiosity

Questions are at the heart of any process focused on problem-solving and they, more broadly, fuel discovery in a very constructive way. Creative curiosity is a mindset that places a high value on questions, recognizes them as rich indicators of knowledge- or experience-based gaps, and welcomes them as an opportunity to enrich the discovery process in a personal way. Consider that, in some cases, the questions you have will be shared by others; but, in other cases, your questions will be unique to your specific interests and experiences. Creative curiosity recognizes that, in order to be fulfilling and interesting to *you*, discovery processes must be driven by questions that *you* genuinely want to answer.

To support your creative curiosity in this process, two tools have been included in your toolkit: the Need to Know (NTKs) Question Tool and the KIMS Tool. In general, NTKs are important because they can keep you aware and mindful of questions that arise in your process and encourage you to actively work toward identifying solutions. By actively using the NTK tool, you will form a clear understanding of the gaps you need to fill and have a solid launch pad for resolving key questions as you move forward. NTKs are also a great way to capture what you're curious about in a given moment so that you don't lose that train of thought and can revisit it later.

The KIMS Tool, on the other hand, supports creative curiosity in a different but equally valuable way. Consider this: as you complete this learning experience, unit by unit, you will not only identify key questions that will fuel discovery, you will also learn new words and concepts. The KIMS Tool will help you keep track of this newfound information and help to ensure that you expand your vocabulary and knowledge base as a result of this experience. As mentioned before, it is important that you get comfortable using these tools in this context but, also, to consider using them in the future as well. A critical part of learning to learn is building a toolbox of resources that help you focus, help you identify areas in which you need more information, and help you carry your new knowledge forward to use in new ways.

What Is Statistics?

Data Science is a natural platform to discuss statistics, a discipline of mathematics that focuses on the collection, organization, analysis, interpretation, and presentation of data that is used to summarize, characterize, or predict.

In this project, we will focus on five particular kinds of descriptive statistics, each of which is highlighted in ***Table 3***. The goal of this unit is to introduce these topics; we will take a deeper dive into each later in Unit 3.

Statistic	Definition
Mean	central or typical value, also called the average of a set of numbers
Median	the middle value of a sorted set of numbers
Mode	the value that occurs the most in a set of numbers
Maximum	the lowest value in a set of numbers
Minimum	the highest value in a set of numbers

Table 3: Summary of Statistical Values Used in this Module

Finding Sources and Extracting Data

As mentioned earlier in this unit, Data Science is all about having the right data, asking the right questions, and performing relevant analyses. In effect, once we have decided on the specific question(s) we will explore for a Data Science project, accessing the right data means specifically identifying the information necessary to derive a solution and then retrieving that data to preform analysis. In practice, data retrieval will require either *directly* collecting information yourself (e.g., by doing your own study) or *indirectly* accessing data collected by someone else. In most cases, data retrieval is done indirectly and is either focused on identifying a single data source or merging multiple data sets in a logical way.

For this project, you are not expected to collect data. In fact, we will use data sets from government agencies in our work. However, we will discuss ways to think about accessing data sets on your own in Unit 7.

We have covered a lot in this unit in an effort to establish a firm foundation for the units to come. Based on the information provided so far, you probably have some questions about the process. Use the NTKs as a basis for writing out those questions now and, more importantly as you identify answers to your questions, be sure to update the table accordingly.

Congratulations! You have finished Data Science Basics. Let's move on to Data Analysis.

Unit 3: Data Analysis

In the **Data Science Learning Experience**, we will use Microsoft Excel as our data analysis software tool. While there are a host of analytical software packages available to choose from,[3] Excel will support our efforts because it is accessible, powerful, and has the functionality to perform sophisticated analytics. Our goal in this unit is to become more familiar with statistical concepts and learn how to perform the corresponding analysis in Excel.

To get a firm understanding of how to use this tool for analysis, we will start with a sample data set and then explain the questions we will ask and the functions we will use to identify solutions. We will anchor our discussion of how data is stored and processed in Microsoft Excel using ***Figure 1*** below.

	A	B	C	D
1	Student	Birth Date	Grade	Progress Comments
2	Anna	5/15/2010	83	Anna has improved
3	Barry	7/3/2010	79	Help Barry with intergrals
4	Cheri	12/19/2009	55	Invite Cheri to tutoring
5	Donald	3/12/2010	97	Donald is doing great
6	Evan	8/7/2010	95	This is Evan's best grade
7	Fallon	11/27/2009	83	This is Fallon's lowest grade
8	Gerald	7/3/2010	100	Gerald is doing great
9				
10				
11				
12				

Calculus Test Scores - Test #3

Figure 1: Sample data set in Excel - Grades for Test #3 in a Calculus course

Data Set Structure

The data set shown in ***Figure 1*** maintains a record of students' scores for the third test in a Calculus course. Note that the data is arranged in rows and columns. In general, the rows are used to maintain a record for a particular student. The row highlighted in red, as an example, captures the name, birth date, grade, and comments for the student named Donald. As a result, if you are interested in learning more about a specific student, you would primarily focus on the rows. Columns, on the other hand, hold a specific kind of data for the entire set of records. The column highlighted in purple, as an example, maintains all of the birth dates for each student enrolled in the course. As a result, if you wanted to do analysis for a particular data characteristic, you would focus on the columns. For instance, if you wanted to understand what the average grade for the entire class was, you would focus on the third column, labeled 'Grade', in the table. One final note, typically, the first entries in the rows and columns are labels that indicate the data contained; the actual data points are stored where the rows and columns intersect. In the example above, the intersection of the purple column and the red row is the birth date for the student named Donald.

Having a clear understanding of how data is structured in Excel is important for analysis overall. As a result, one of the first steps you should take any time you open a new data set is to pay special attention to the labels on its rows and columns. Identifying what data is contained in a data set is an indicator of the kinds of questions you can ask about it. As an example, the data set shown in Figure 1 contains students' names, birth dates, grades, and comments; students' addresses are not included in this set and as a result, we cannot ask questions about it during our analysis. Getting comfortable with this is important as we look to review more advanced data sets in the future.

[3] A few examples of widely used analytical software packages include R, Python, SAS, SPSS, STATA, and Alteryx.

Posing Questions

As we began discussing in Unit 2, for now, we will use the statistical values of *mean*, *median*, *mode*, *maximum*, and *minimum* to conduct analysis. A driving philosophy of this module is that is important for you to start with a question in English and understand which statistical values you need to compute in order to arrive at an answer. To get a general overview of the connection between statistical values and associated questions, the list below makes an explicit link:

- **Mean**: What is the average value for ***M*** in the data set?
- **Median**: What value separates the first half of values in the data set from the second half?
- **Mode**: What value of ***O*** occurs the most in the data set?
- **Maximum**: What is the largest value of ***P*** in the data set?
- **Minimum**: What is the smallest value of ***Q*** in the data set?

Given the values of the data set we're using, the variables ***M***, ***O***, ***P***, and ***Q*** could be used to refer to the names, birth dates, grades, and comments recorded for each student. In the examples that follow, we will round out the discussion of posing questions by transforming each general question into an actual question and then discuss the Microsoft Excel function we will use to identify the answer.

Question #1: What is the average value for *Grade* in the data set?

To calculate the average test score for the third Calculus test, we will need to find the sum of all test scores and divide that number by the total number of test scores. More specifically, we will need to use the following equations (refer to ***Figure 1*** to see where the values came from):

$$\text{mean} = \frac{\text{sum of all } \textbf{\textit{Grades}}}{\text{total number of } \textbf{\textit{Grades}}}$$

$$= \frac{(83 + 79 + 55 + 97 + 95 + 83 + 100)}{7}$$

$$= \frac{592}{7}$$

$$= 84.57$$

Interpretation of the result: The mean value for Test #3 grades is 84.57. Considering that, for this particular course, the grade of B ranges from 80 – 89, students on average earned a B on the third calculus test.

To calculate this value in Microsoft Excel, and more specifically, calculate the mean of *Grade* in the data set shown above, we will use the **AVERAGE** function as follows:

```
=AVERAGE(C2:C8)
```

In English, this function reads, Find the average of the values in Column C from C2 to C8. Note that the grade for Anna is in column C2 and the grade for Gerald is in column C8 so this function captures the full list of grades.

Question #2: What value for *Grade* separates the first and second half of the data set?

To determine the value of *Grade* that separates the first half of the list from the second had, we must first start by sorting the list from the smallest value to the largest value as shown:

55, 79, 83, 83, 95, 97, 100

The value that actually separates the first half of the list from the second half is the value that appears in the middle, as highlighted below:

55, 79, 83, 83, 95, 97, 100

Interpretation of the result: One half of the class made less than 83 and the other half of the class scored higher than 83.

To calculate this value in Microsoft Excel, and more specifically, calculate the median of Grade in the data set shown above, we will use the **MEDIAN** function as follows:

```
=MEDIAN(C2:C8)
```

In English, this function reads: Find the median of the values in Column C from C2 to C8.

Question #3: What value for *Grade* occurs the most in the data set?

To determine the value of *Grade* that appears the most in the data set, it is a good idea to start with a sorted list. This way, all of the values will be grouped together and it will be easier to spot the numbers that appear the most.

55, 79, 83, 83, 95, 97, 100

In looking at the list, all but one of the scores occur once except for 83, which occurs twice.

Interpretation of the result: The grade that occurred the most is 83.

To calculate this value in Microsoft Excel, and more specifically, calculate the mode of Grade in the data set shown above, we will use the **MODE** function as follows:

```
=MODE(C2:C8)
```

In English, this function reads: Find the mode of the values in Column C from C2 to C8.

Question #4: What is the largest value for *Grade* in the data set?

To determine the largest value in a set, we should begin with a list that is sorted from the smallest value to largest value and then, highlight the last value in the list (as shown below):

55, 79, 83, 83, 95, 97, 100

Interpretation of the result: The highest grade on Test #3 is 100

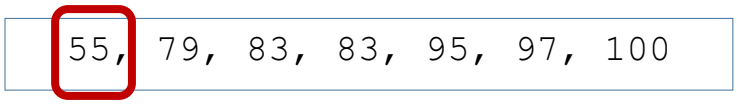
To calculate this value in Microsoft Excel, and more specifically, determine the largest Grade in the data set shown above, we will use the **MAX** function as follows:

```
=MAX(C2:C8)
```

In English, this function reads: Find the maximum value in Column C from C2 to C8.

Question #5: What is the smallest value for *Grade* in the data set?

To determine the smallest value in a set, we should begin with a list that is sorted from the smallest value to largest value and then, highlight the first value in the list (as shown below):



Interpretation of the result: The lowest grade on Test #3 is 55

To calculate this value in Microsoft Excel, and more specifically, determine the largest Grade in the data set shown above, we will use the **MIN** function as follows:

```
=MIN(C2:C8)
```

In English, this function reads: Find the minimum value in Column C from C2 to C8.

This unit provided a great deal of information about how to either calculate or identify the *mean*, *median*, *mode*, *maximum*, and *minimum* values of a data set as well as the functions in Microsoft Excel used to determine each. The knowledge you have gained about how to think about these ideas in English and how to implement them in an analytics tool are a great start for further analysis in the units to come. In the meantime, to get additional practice calculating each statistical value by hand, visit Appendices D & E.

Congratulations! You have finished Data Analysis. Let's move on to Developing and Implementing Your Plan.

Unit 4: Developing and Implementing Your Plan 

In previous units, the goal was to develop a basic understanding of Data Science and assemble a toolkit for data exploration and analysis. In pursuit of that goal, we established a firm definition of data, introduced/reviewed statistics, discussed how to pose data analysis questions, and applied key analysis techniques in Microsoft Excel. All combined, we covered a host of critical topics. Yet, having *awareness* is one thing; successfully *applying* concepts and tools is the next step.

In this unit, your job is to put it all together and use the skills and insights you've gained so far to conduct data analytics on your own. By doing so, you will put what you've learned into practice and begin applying your toolkit to real-world data. Navigating this process will provide invaluable feedback; as you attempt to complete key tasks, you will quickly learn what you firmly know and what you need to reinforce by reviewing prior units or seeking additional information. Keep in mind, needing to review is not a bad thing; on the contrary, it is a great opportunity to revisit the material, fill in the blanks, and ensure growth. All combined, in this unit, the goal is to not only put your new knowledge into practice but also prepare yourself to explore other data sets of interest independently.

In the exercises that will follow, you will develop a data exploration plan, execute the plan, and then reflect on your effort. Establishing a strong plan is the first step because it will help to provide structure to your process. From a general perspective, starting with a plan is an important step because it provides you an opportunity to brainstorm and think critically about the path forward, gather or earmark key tools, identify potential challenges, work out any glaring problems in advance, have a more formal way to track your progress, and help to ensure that you have well-supported discovery. To address a common misunderstanding, planning ahead does not reduce your opportunity to be creative. Instead, planning ahead can allow you to set the cornerstones of your approach early on and, thus, structure the exploration process in a way that shifts the focus away from deciding on a path and allowing more space for creative ideas to surface.

The Overall Approach

As you plan your way forward, a great place to start is the basic structure of the problem-solving process for Data Science, as discussed in Unit 2. Use ***Table 4*** below to list those steps. Try to recall them from memory first.

Write the five steps of the Data Science Process below:	
1	
2	
3	
4	
5	

Table 4: The 5 Data Science Steps

As an important note, whether your first step is formulating questions or finding appropriate data sets is heavily dependent on the particulars of your project. In this learning experience, we will start by identifying our data set from a curated list first and then formulate our questions.

Finding Appropriate Data Sets

For this learning experience, the first step we will exercise in the Data Science process is selecting a data set. To get process started, review the list below – which contains U.S.-based data sets in athletics, colleges and universities, population, and health – and select the data set that aligns with your interests.

Data	Source	Link
Equity in Athletics	U.S. Department of Education – Office of Postsecondary Education	https://ope.ed.gov/athletics/#/datafile/list
College Information	National Center for Education Statistics	https://nces.ed.gov/ipeds/use-the-data
U.S. Census Data	U.S. Census Bureau	https://data.census.gov/cedsci/
COVID-19 Data	Centers for Disease Control and Prevention	https://data.cdc.gov/browse?tags=covid-19

Table 5: Data Set Choices

Once you select your data set, write its name below (using the title in the **Data** column of ***Table 5***) along with a short explanation for why you chose it in ***Table 6***. It's a good idea to keep track of this information now, while you're in the moment. This table will be a helpful reference when you start preparing the presentation to document your experience.

Data Set Selection and Justification	
What data set did you choose?	
Why did you choose that data set?	

Table 6: Data Set Selection and Justification

Formulating Questions

Once you have selected your data set, the next thing to focus on is the list of questions that are going to shape your data exploration process. As mentioned before, it is important to start that process by looking at the data set first, learning more about its structure, and understanding what information it captures. Specifically, open the data set you selected and take a look at the row and column labels. Next, decide what data values you want to know more about using the five questions we defined in Unit 3 as a guide. As an example, if you opened the data set and noted that one of the columns was labeled **length**, you would know that the corresponding column captured the length for each entity represented by a row. From there, you could decide that you wanted to know the largest **length** in the set and would apply the `MAX` function in Microsoft Excel to derive that information. With this example in mind, explore the data set you selected and use that knowledge as a launch point to indicate the questions you will ask in ***Table 7*** along with either the Microsoft Excel function (`MEAN`, `MEDIAN`, `MODE`, etc.) or the approach you will use to extract an answer.

Questions	#	Question	Function/Approach
	1		
	2		
	3		
	4		
	5		

Table 7: Data Set Exploration Plan - Questions & Functions/Approach

Understanding what is contained in a data set and how it is arranged is critical to your analysis because it provides, as you've just experienced first-hand, the basis for formulating the questions you will answer. However, it can also provide an indicator you might need to locate additional data in order to do certain kinds of analysis. It is not unheard of to merge several different data sets to answer questions of interests. For now, however, just work with the available data.

Cleaning/Formatting Data

Collecting and storing data is not a fault-free process and, as a result, some data cells might contain noise (or garbage), incorrect entries, or duplicated rows; as a result, you may need to perform data clean-up prior to analysis. To assist with eliminating duplicate rows, navigate the following path in Microsoft Excel: click the **Data** tab and select **Remove Duplicates** in the **Data Tools** panel. Use this process on your data set to get a sense of what it does and how to use it.

Analyzing Data

Once you have formulated your questions and identified the function/approach you will use, the next step is to implement your plan and record the results. Leveraging the planning you did in the 'Formulating Questions' section, record the value or values you derive in ***Table 8*** below for interpretation.

Answers	#	Answer
	1	
	2	
	3	
	4	
	5	

Table 8: Data Set Exploration Plan – Answers Derived

Now that you've done the work and recorded the results, use ***Table 9*** to interpret the answers recorded in ***Table 8***. Use ***Tale 10*** to share general observations. Were the values what you expected? Did you learn anything new by looking at the results? Again, it is important to record those thoughts now, in the moment, so that you can share them in your presentation.

Interpretation	#	Interpretation
	1	
	2	
	3	
	4	
	5	

Table 9:Data Set Exploration Plan - Interpretation

Observations	

Table 10:Data Set Exploration Plan - Observation

The Wrap-Up

One of the best ways to tell if you have learned something is to apply the knowledge. This unit was all about planning the way forward, executing, and simultaneously preparing for the next step – documenting your findings. We will jump into preparing your presentation in Unit 6. In the meantime, we will explore Data Science as a career.

Congratulations! You have finished Developing and Implementing Your Plan. Let's move on to Data Science as a Career.

Unit 5: Data Science as a Career

Throughout the previous units, you have been developing and exercising skills that can be applied in the near-term, in your day-to-day life, exploring topics that are of interest to you but also, in your life as a student, by reinforcing the knowledge of statistics you will need in your courses. Yet, the benefits don't end there. The skills and tools you have been learning and applying throughout this module have helped you to build foundational awareness of what Data Scientists do professionally. As a result, if you love exploring data sets and discovering interesting facts, you may want to consider a rewarding and impactful future as a Data Scientist. This unit will guide your exploration of Data Science as a career and provide insight into how to prepare academically to be a Data Science professional.

A Career in Data Science

One of the most beneficial things you can do for yourself is gain insight on the wide variety of career options available to you; this will not only point you in the right direction for your career of choice but allow you to begin preparing well in advance. One tool that can be helpful in this process is the Occupational Information Network (O*NET), a tool supported by the U.S. Department of Labor/Employment and Training Administration (USDOL/ETA) that provides key critical insights and details about a wide range of careers[4]. In addition to providing exploration-oriented tools, like the Interest Profiler which helps you link your interests directly to a set of viable careers, O*NET also allows you to search for details on specific careers and gather critical information about each. O*NET is an invaluable resource for exploring careers and learning about key occupational factors including, annual salary, working conditions, required skills, and job growth projections. We will use O*NET heavily in this phase of the learning experience, but, now that you know about it, feel free to use it as often as you would like to explore any other career of interest.

Let's dive in! Start by accessing O*NET using the following link: https://www.onetcenter.org/

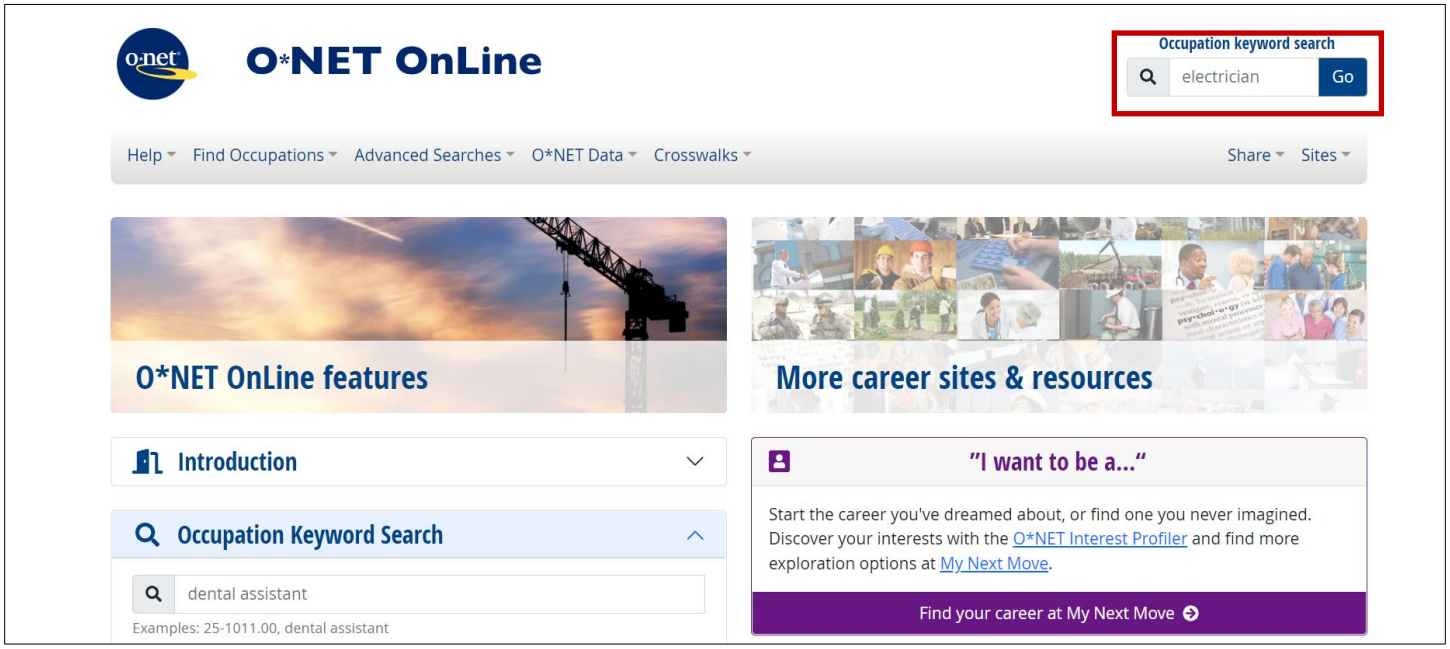


*Figure 2: O*NET homepage with the search box highlighted in red*

[4] https://www.onetcenter.org/overview.html

After arriving at the site, type the following query into the search box (shown in red in ***Figure 2***):

data scientist

As a result of the search query, the first return you will see is the occupation **Data Scientists**, code 15-2051.00. After clicking the link and navigating to the corresponding web page, you will see a host of important information regarding typical work activities, wages and employment trends, and related occupations.

Absorb as much as you can about Data Science as a career using the information provided by O*NET. Then, in ***Table 11***, record the requested data.

Data Science as a Career Record information about the wages, duties, and skills for Computer and Information Research Scientists below.	
Median Wages	
Projected Growth	
Key Tasks	

Table 11: Data Science as a Career

Having navigated this learning experience, picked up critical skills, and gathered information about a career, take some time to capture your thoughts about Data Science as a career in the space below.

Use the space below to jot your thoughts about Data Science as a career.

Identifying a College to Prepare for a Career as a Data Scientist

One of the most traditional ways to launch a career path in Data Science is to attend a college or university and begin the journey by earning an undergraduate degree in Computer Science, Electrical Engineering, or Mathematics but there are also opportunities to earn certificates in Data Science.

This leads us to our next question: how should you start thinking about the educational process for becoming a Data Scientist? One way is to begin identifying a potential institution of choice by finding colleges and universities with strong Data Science programs. To discover institutions that fit the bill, conduct a Google Search on the top colleges and universities in Data Science using the following query – be sure to type it *exactly* as shown:

+“data science” +top + colleges

In your search results, look for a return from U.S. News & World Report labeled, **Best Undergraduate Data Science Programs – USNews.com**:

1. Click the link and then record the top 10 institutions in ***Table 12*** (below).
2. Next, take a look at your list and identify at least one institution that would be of interest to you as a college/university to attend. Place an asterisk beside that institution.
3. Finally, explore the institution you placed the asterisk beside a little further. In particular, identify where the institution is located in the U.S. and find the homepage for the institution online. Copy/past the link to the institution's home page into ***Table 12*** as well.

By completing this exercise, you will not only get an idea about the institutions with high-caliber Data Science programs, you will begin taking the steps necessary to learn about a future college or university of choice. Although you may change your mind about the college/university you want to attend several times between now and when you actually start working on your undergraduate degree, getting the ball rolling on this process as early as possible can be very beneficial in the long run.

Top U.S.-based Colleges and Universities in Data Science Record U.S. News & World Report's top 10 Data Science Institutions below, identify an institution of interest on the list, and record key information about that institution below.		
	Top Ten Institutions	
	Institution of Interest	
	Location (City, State)	
	Homepage Link	

Table 12: College and Universities to Consider in Data Science

Congratulations! You have finished Data Science as a Career. Let's move on to Presenting Your Work.

Unit 6: Presenting Your Work

One of the cornerstones of discovery, especially after intense data analysis, is presenting your findings. This is an especially important endeavor in this project because it will allow you to share your experience with others, heighten awareness of the knowledge you discovered as a result of your work, and enlighten your audience on how your skill and knowledge toolkit grew via the **Data Science Learning Experience**. Most importantly, any opportunity you have to strengthen your ability and confidence as a speaker is a plus.

The goal of this unit is to help you develop a presentation that encapsulates your experience and findings. The idea is to not only encourage you to communicate your results – which is important – but to also share the interesting things you learned about the data set and to think critically about the results that you achieved before sharing them with others.

Each heading below indicates a slide number, a potential title, and the information you should discuss. If you follow the prescribed approach below, you will build a 9-slide deck that documents your discoveries and tells the story of your journey through the **Data Science Learning Experience.** Most importantly, your presentation will capture what you learned, detail the decisions you made in your data analysis, discuss the outcomes you observed, communicate the key take-away messages you want the audience to walk away with, and allow you to lead conversations that heighten others' awareness of Data Science. The unit also provides insight on how to engage your audience.

Slide 1: Introductory Title

Your first slide should, at the very least, list the title of your talk. In general, this slide includes the title of the talk, the presenters' name, and the presenters' affiliation, e.g. the school attended or company of employment. Since you will likely be presenting to your friends, classmates, teachers, parents, or caretakers, your name and affiliation are not requirements since they will likely know both.

As far as how you should deliver this slide, when you share the title of your talk, you should also thank your audience member(s) for agreeing to listen to your presentation. Also, this is a good time to let your audience know how to engage with you during the presentation. For instance, if you would like for them to hold questions until the end or if you are comfortable with them stopping you during your presentation to ask a question, let that be known at the very beginning to avoid confusion and ensure a smooth presentation.

Slide 2: Data Science Overview

Use this slide to introduce Data Science, based on what you learned about it during this learning experience, and list the 5-step process you took during your data exploration process.

When delivering this slide, make it clear to the audience that the 5-step process is an overview of the steps you took and that you will be going in-depth with each step during the remainder of the presentation.

Slide 3: Data Set Selected and Tools Used

Use this slide to indicate the options you had for data selection – namely data sets on U.S.-based athletics, colleges and universities, population, and health (page 16) – specify which one you chose, and indicate why. As far as tools, indicate that you used Microsoft Excel to do your analysis.

In delivering this slide, keep in mind that the discussion of the data set you chose is an opportunity to infuse aspects of your personality and interests into your talk. Maximize this opportunity to frame the story of your experience by sharing how the interests and curiosities you came to the module with influenced your choices.

Slide 4: Data Preparation

Use this slide to introduce the idea that data sets can be messy and incomplete.

In your delivery, discuss the **Remove Duplicates** functionality of Microsoft Excel and indicate how you used it to prepare your data for analysis and the outcomes of your usage? For instance, were there any duplicates found in your data? If so, how many? Do you have any thoughts regarding why these duplicates may have occurred?

Slide 5: Questions Asked and Solution Approach

Use this slide to list the questions you chose to answer about your data set and why you chose them. Also, share the statistical methods you used to arrive at the answers to your questions. In short, this slide should answer the following questions:

- What did you observe about the data contained when you opened the data set?
- What questions did you start out with and how did you plan to answer them?

In your delivery, you should use as many visual aides as necessary. For instance, this might be a good time in your presentation to show your audience a screenshot of your data set or to even show the data set live in Microsoft Excel while exploring its rows and columns in real-time. You may also want to screen shot and display the table you used to keep track of the questions you answered and the functions/approaches you used (***Table 7***).

Slide 6: Data Summary

Use this slide to share the answers you derived after applying the prescribed function/approach. Also discuss your interpretation of the results and whether the results you derived sparked any additional questions. In short, this slide should answer the following questions:

- What were the answers to the questions you posed?
- How did the statistical values you calculated help to answer those questions?
- What did you learn about the data based on the answers you collected?
- What if any new questions (specifically about the data set) did the answers you discovered raise for you?

This slide focuses squarely on the results of your analysis. In your delivery of this slide, be sure to make clear connections between the questions asked, the answers derived, and your interpretation of each. Consider including a screen shot of the tables you used to store the answers you derived and your interpretations (***Table 8 and Table 9***).

Slide 7: Data Science as a Profession

In addition to learning key Data Science skills, you also learned about the role of Data Scientists as a result of your work. Use this slide to discuss your findings. In particular, this slide should include answers to the following questions:

- What is the median wage for Data Scientists?
- What is the projected growth of the field?
- What key tasks are Data Scientists responsible for?

In your delivery of this slide, be sure to indicate what your thoughts are of Data Science as a profession and career. Based on what you've learned, are you considering Data Science as your career path? Share this insight with your audience and be sure to indicate why or why not.

Slide 8: Lessons Learned and Project Experience Take-Aways

Use this slide to provide your audience with a list of key take-aways from your presentation. Use this space/time to share your answers to one or all of the questions below.

- What challenges did you encounter in completing this project?
- What did you find interesting about the data analysis process?
- What do you like about Data Science?
- What surprised you about the analysis process and the answers you discovered?

As for delivery, this is one of the most important slides in your deck. Try to keep your discussion succinct and focus on the most memorable aspects of your project.

Slide 9: Questions

Thank your audience for their time and attention and ask them if they have any questions. Don't be afraid of this segment of your talk. Typically, audience questions lead to further inquiry and discovery. If an audience member asks you a question that you do not know the answer to, tell them that you will add it to your NTKs and get back with them if you find an answer. Use audience questions as an opportunity to grow and a reason to continue exploring further.

Finding your Audience

Learning how to present your work is one of the most critical aspects of the **Data Science Learning Experience**. Being able to talk about the new things you learned is a clear indicator that you have absorbed the material. In addition, getting as many opportunities as possible to practice giving talks is crucial. You will become less intimidated by the process if you feel overwhelmed at first. You can learn about your strengths and work to build on them. At the same time, you can identify your weaknesses and work to make them strengths.

To identify the audience for your talk, start by inviting your parents, siblings, friends, and teachers. In addition, try to schedule more than one presentation; this will allow you to practice in front of one audience, get initial feedback, update your talk and then present again. The more you repeat this cycle, the more practice you will get, and the more you can improve.

Congratulations! You have finished Presenting Your Work. Let's move on to Diving In and Next Steps.

Unit 7: Diving In and Next Steps

A natural start point in understanding and exploring Data Science is actively performing analysis on a wide variety of data sets and using a range of tools, techniques, and functions in pursuit of data-driven discovery and insight. The **Data Science Learning Experience** provides an introduction to the process and initial exposure to help students decide if they want to dive further into the field. In addition, the learning experience is built on the idea that understanding statistical concepts and knowing how to maneuver in Excel have value beyond this module and provide a launch pad to grow and develop skills that will be beneficial in a variety of contexts.

Now that you've been exposed to key concepts and the role of a Data Scientist, you may be interested in continuing to build your skillset. This unit was developed to provide suggestions for moving in the right direction and going further.

One of the most important messages to drive home is to continue to grow and excel in your mathematics courses. Having a strong mathematical foundation is critical in data analysis. Now that you recognize it, it is incumbent on you to do your best in these courses and make the effort to ensure that you fully understand the material. Mathematics is not necessarily easy for everyone. However, now that you know this might be a career path of interest to you, make the effort to excel. As we saw in Unit 4, having *awareness* is one thing; being able to successfully *apply* concepts and tools is the next step. Make sure that you are taking every opportunity to apply the math concepts you are learning and, if you are having trouble grasping topics, don't hesitate to reach out to your teachers or request support from a tutor.

The next thing to consider is how to continue exploring data. Clips on YouTube, like Statistical Data Analysis are a great place to start. As for identifying data sets that you can use to perform analytics in a variety of contexts, consider the following list which covers the areas of justice, health, science, education, political science, and economics:

Bureau of Justice Statistics
Centers for Disease Control and Prevention (CDC)
Data.gov
National Aeronautics and Space Administration (NASA)
National Center for Education Statistics
National Center for Health Statistics (NCHS)
National Oceanic and Atmospheric Administration (NOAA)
US Census Bureau
US Department of Education - Data Express
The World Bank

One very important note: As you explore these and other data sets, continue to share your results with your family members, friends, peers, and teachers. As you move forward, and continue to apply and grow your skills, be sure to maximize the opportunity to grow your communication skills and get increasingly comfortable delivering presentations. In doing so, you will not only spread awareness about your work but you may also inspire someone else to consider Data Science as an area of interest.

Here's the key take-away: you've done a lot to build critical skills – continue to learn as much as you can! The resources listed above are a start but there are many others to consider and choose from. The key is to continue to fuel your thirst for knowledge and watch your abilities grow beyond your wildest imagination.

Congratulations! You have finished the Diving In and Next Step. Let's move on to Reflecting on the Road Traveled and Measuring Growth (Part II).

Unit 8: Reflecting on the Road Traveled and Measuring Growth (Part II)

Data Science is an emergent field with growing impact and a high demand for talent. Given the ever-increasing need for data-driven awareness and a data-adept workforce, the **Data Science Learning Experience** was designed to provide the exposure needed to spark your interests, develop working knowledge of key concepts, build base-level skills, and establish a platform to expand awareness. Providing adequate exposure is critical to helping you identify career paths of interest and we hope this module has enlightened you about the fascinating opportunities available in Data Science. This experience was designed to give you a strong base and, hopefully, the confidence you need to take on and successfully navigate other Data Science opportunities and replicate this process with another topic of your choosing in the near future.

As a result of this learning experience, you have explored a data set, asked key questions about it, conducted analysis, interpreted the results, and presented your findings. The more practice you get with each step, the more comfortable you will be exploring other data sets in the future, and the more advanced your toolkit of applicable techniques will become. In addition, the tools and skills you have used throughout this experience can be used in classroom assignments and, hopefully, provide you with good habits for data exploration and discovery.

In all, this experience was designed to build your knowledge base and capabilities in ways that open new doors and provide long lasting effects in how you think about data, problem-solve, explore careers, and communicate your findings. Continue to seek out additional tools and learning opportunities to understand the landscape, grow critical skills, and understand how they might benefit you in the field of Data Science.

The process of taking advantage of opportunities like the **Data Science Learning Experience** and using it as a stepping stone to continue to grow and expand your skill set can help you grow as an independent learner. Problem-solving can prepare you to take control of your own learning and explore your curiosities independently. As you continue to journey throughout your middle and high school years, you will likely find that the skills and topics discussed in this learning experience will be quite helpful in a variety of contexts, not just your math and science classes. For example, being able to present well is universal. Keep the momentum going and make sure that you are putting the skills you picked up here to good, continued use. The key is to use the skills you've already built as they apply and to learn new skills as the need surfaces.

There was a lot to consider in this learning experience and we hope that you were able to gain a lot of new knowledge. To track your progress, we have included a tool you can use to see how your knowledge base has grown since you started the **Data Science Learning Experience**. On the next page, you will find the same set of questions you answered in Unit 1. The answers you provided before you completed the units that followed capture where you were. To see how far you've come, complete the 8 questions on the next page and compare your latest answers to your original responses. The results of that comparison will not only provide insight into how you've advanced your understanding but will also shed light on areas that may need additional review. Celebrate the growth but take note of opportunities to improve and revisit previous units as necessary!

In closing, thank you for taking the journey! We're excited for your continued growth and hope that your time exploring the units, learning key concepts, and doing activities was enriching and that a new door has opened for you. We'd love to hear about your experience and how, from your perspective, it can be improved. Be sure to reach out to us at the link below and share your thoughts: https://forms.gle/t5dYQzLthehLT3aJ8

Track Your Progress: Take a look at the questions below and answer, as best as you can, based on what you've learned. Once you complete the set, compare your responses here with the answers you provided in Unit 1 to see how far you've come over time.

1. **What is data? What are two examples of datum?**

2. **What do you know about Data Science and data analysis?**

3. **In statistics, what is the *mean*?**

4. **In statistics, what is the *median*?**

5. **In statistics, what is the *mode*?**

6. **What aspects of your personality or interests might serve you well as a Data Scientist?**

7. **On a scale from 1 – 10: rate your <u>understanding</u> of Data Science.**

8. **On a scale from 1 – 10: rate your <u>interest</u> in Data Science.**

Citations

[1] B. Upbin, “The Web Is Much Bigger (And Smaller) Than You Think”, *Forbes*. [Online]. Available: https://www.forbes.com/sites/ciocentral/2012/04/24/the-web-is-much-bigger-and-smaller-than-you-think/ [Accessed: August 9, 2022].

[2] San José State University – College of Professional and Global Education, “Data Analytics (MS)” [Online]. Available: https://www.sjsu.edu/pdp/programs/msda/ [Accessed: August 18, 2022].

[3] O*NET Resource Center, “About O*NET” [Online]. Available: https://www.onetcenter.org/overview.html [Accessed: August 11, 2022].

Appendix A: Need to Know Questions

#	Question	Potential Process for Retrieving Answer(s)	Resolved?
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			

Appendix B: KIMS[5,6]

KIMS is the acronym for a strategy that supports the exploration of words and can help you build your vocabulary knowledge across any content area. The acronym stands for keyword (K), important information (I), memory clue (M), and sentence (S). To assist with the KIMS process, a graphic organizer has been provided below with each column labeled accordingly.

To implement the KIMS approach when encountering a term you don't know, apply the following steps:

1. Add the new term to the ***Keyword*** column. [**ex**. corpus]
2. Identify or locate the definition for the term and record it in the ***Important Information*** column. [**ex**. A large dataset that can be used to train a machine.]
3. Think about a memory clue that could help you remember the keyword's definition? Write or draw this ***Memory Clue*** in the third column. [**ex**. A corporation is a group of people working toward the same purpose, so a corpus is a dataset being used for the same thing.]
4. Finally, develop a sentence that correctly uses the word. Write this ***Sentence*** in the last column and circle the keyword in the sentence. [**ex**. I collected every picture of dogs I could find to create a corpus for teaching the AI what a dog is.]

#	Keyword	Important Information	Memory Clue	Sentence
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				

[5] https://learn.k20center.ou.edu/strategy/780

[6] Vibas, K. (1970, January 10) K.I.M "A Highly Effective Strategy to Build Vocabulary Across Any Content Area", http://www.awalkinthechalk.com/2016/12/kim-highly-effective-and-simple.html

Appendix C: Data Analysis Planner

The Data Science process requires you to identify appropriate data sets, ask appropriate questions, clean the data set if necessary, analyze the data and present your findings. This planner was designed to help you structure that process by providing space for you to make decisions and record key outcomes. To plan for data exploration and keep track of your findings, consider the following:

- **Data Source**: What data set will you use during exploration and, upon retrieval, where did you find it?
- **Selection Justification**: Why did you choose this data set?
- **Questions**: What questions do you plan to pose about the data? How will you go about extracting an answer – namely, which function or approach will you use?
- **Answers**: What were the answers to the questions you posed? Record the answers in the space provided below.
- **Observations**: What are your thoughts about the data you collected? Were there any answers that were unexpected? Were there answers you were able to accurately predict in advance?

Overall, these questions are designed to put you in a position where exploration of a new data set is within reach. You have the basic tools and this planner helps you to develop the approach needed to explore on your own.

Key Consideration	**Response**		
Data Source (Link)			
Selection Justification			
Questions	#	Question	Function/Approach
	1		
	2		
	3		
	4		
	5		
Answers	#	Answer	
	1		
	2		
	3		
	4		
	5		
Observations			

Appendix D: Mean, Median, Mode+ in Action – Part 1!

Understanding what the mean, median, and mode are and calculating them along with the maximum and minimum of a data set are a key learning outcome for this module. In this practice set, you will calculate each for the list of numbers shown below, explain your steps in English, and indicate the requested value. Try to use your notes as sparingly as possible.

8, 13, 12, 4, 1, 20, 12

1. What is the **mean**, or average, value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

2. What is the **median** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

Use the space above to write out the sorted list

3. What is the **mode** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

4. What is the **maximum** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

5. What is the **minimum** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

Appendix E: Mean, Median, Mode+ in Action – Part 2!

As discussed in Unit 4, having *awareness* is one thing; being able to successfully *apply* concepts and tools is the next level. The exercise that follows presents a list of numbers and requests that you compute the mean, median, mode, maximum, and minimum by hand.

13, 15, 13, 13, 4, 7, 3, 9, 22

Answer the following questions using the data set above:

6. What is the **mean**, or average, value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

7. What is the **median** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

Use the space above to write out the sorted list

8. What is the **mode** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

9. What is the **maximum** value?
   - Steps in English:
   - Computation of the value:
   - Interpretation of the value:

10. What is the **minimum** value?
    - Steps in English:
    - Computation of the value:
    - Interpretation of the value: