Problem 3.2.7

Investigate with Data

Introduction

People with computational thinking skills see opportunities to use computational power to solve problems and address needs. Consider something you would like to know. A computational thinker wonders what data are available on the Internet, or how one might crowd-source the collection of that data. A computational thinker recognizes opportunities to use computational power to explore, analyze, and visualize the data.

What do you want to know more about? What tools do you naturally turn to?

Materials

- Computer with Canopy distribution of *Python*® programming language and access to Internet

Resources

Problem 3.2.7 Rubric

Procedure

Part I: Asking Questions

1. Form pairs as directed by your teacher. Meet or greet each other to practice professional skills. Set team norms.
2. Develop a cluster of questions in an area of investigation. To develop these questions, record the results of three brainstorming activities.
   - Brainstorm areas in which you might be interested in creating new insight or knowledge. You can include broad and more specific areas in your brainstorming. For example, you might list music, live music, and/or a genre of music as areas for investigation. Recall the guidelines for brainstorming: use a tag line to keep a record, go for quantity, never criticize ideas during brainstorming, and "piling on" is welcome.
   - Brainstorm fields of data that might be obtainable for each of two or more of the areas from the previous brainstorm. For example, for orchestral music, there might be data listing composer, piece, year, conductor, orchestra, date of performance, location, and so on. Keep a record of this second brainstorm.
   - Brainstorm questions that ask how one field of data is related to another field of data. For example, how does the location in which an orchestral music piece was composed relate to the location in which it has been performed? Has the "export" of music changed over

time?
- Select 3-5 questions in a single area of investigation. Describe the area of investigation and why you chose it.

Part II: Finding or Collecting Data

3. Seek one or more sets of data relevant to your investigation. The combined data set should contain multiple attributes and be rich enough to provide meaningful insight in your area of investigation. You can proceed with the next step while continuing to search for additional data.
   - Identify each data set and its source. Use an APA citation with a URL and the date of retrieval.
   - Describe each data set.

Part III: Exploring and Analyzing Data

4. Use any tools you wish to seek patterns in your data. Consider working with subsets of the data by sorting, clustering, and filtering the data. Comment any code you produce. Use visualization and analysis as appropriate, recording and explaining your work as you progress. Keep records of your methods, results, and discoveries in a project notebook. Keep version-controlled copies of your code and derived data.

Part IV: Communicating Your Methods and Discoveries

5. Create a written presentation with the following components.
   - Present the questions you selected with rationale.
   - Provide citations and descriptions of the data you used.
   - Explain the information or knowledge you gained from your analysis of the data.
   - Provide visualizations, tables, or other resources you created. Title each such resource as a numbered Table or Figure and include one or more sentences referring to each Table or Figure in your explanation of the knowledge you gained.
   - Describe the methods that were used to analyze the data. Enough detail should be given to allow replication. Discuss the tools you used as well as the manipulations performed with your data. You might separately address the activities of identifying data, merging or cleaning data, filtering, clustering and classifying data, transforming data, and visualizing data. This written description must also address the scalability of your techniques to data sets several orders of magnitude larger.

Conclusion Questions

1. What new questions were raised by your investigation? Describe how you would proceed if you were to continue your investigation.
2. To what extent computation was necessary to analyze your data and answer your questions.

3. Reflect on your collaboration.
   - How did you share or divide the work?
   - What was the most significant piece of information that you contributed?
   - What was the most significant piece of information contributed by your partner?
   - What question or feedback did you ask or offer that gave your partner the most significant help in reviewing and revising his or her work?
   - What question or feedback did your partner ask or offer that gave you the most significant help reviewing or revising your work?

4. You created visualizations of data for exploration in Part III and for communication in Part IV. Compare and contrast these two creative tasks.

5. Practice Opportunity for the *Create* Performance Task

   Respond to two or more of the following prompts as directed by your instructor.

   - "Create a 50-59 second video in which you demonstrate the running of at least one significant feature of your program." (Adapted from College Board Create Performance Task Part 1.)

   - "Identify the purpose of your program and explain what the video illustrates. (Approximately 150 words)" (Adapted from Create Performance Task Part 2a.)
   - "Describe the incremental and iterative development process you used, focusing on two distinct points in that process. Describe the difficulties and/or opportunities you encountered and how they were resolved or incorporated." (Adapted from Create Performance Task Part 2b.)
   - "Describe how each algorithm within your algorithm functions independently, as well as in combination with others, to form a new algorithm that helps to achieve the intended purpose of the program. (Approximately 200 words)" (Adapted from College Board Create Performance Task Part 2c.)
   - Explain how abstraction helped manage the complexity of your program. (Approximately 200 words) (Adapted from College Board *Create* Performance Task Part 2d.)
   - Paste your previous four responses into a new document and clearly label them 2a through 2d. Use screenshots to capture your entire program code, and paste the image(s) into the document, labeling the images 2e. Draw an oval around the segment of code corresponding to the algorithm of algorithms you described in deliverable (c). Draw a rectangle around the segment of code corresponding to the abstraction you described in deliverable (d). Save the document as a PDF. (Adapted from College Board *Create* Performance Task Part 2e.)

Note: *This last set of deliverables is adapted from the official College Board Create Performance Task but they do not duplicate the content of College Board Task or Rubric. The task provided here contains elements that are different than the College Board Performance Task and Rubric. Please reference official College Board materials.*

Problem 3.2.7 Investigate With Data Rubric

Evidence	Criterion	**4**	3	2	1
5.a.i	Questions	**Insightful** questions seek connections among **many significant** attributes	**Good** questions seek connections among **some significant** attributes	**Good** questions seek connections among **limited** attributes	Questions **do not** connect multiple attributes
5.a.ii 5.b.ii	Data	Data set(s) support a **rich** investigation	Data set(s) support a **reasonable** investigation	Data set(s) support a **limited** investigation	Data **does not** support the area of investigation
5.b.iii Project Notebook Files	Discovery Process	**Strategic** and **purposeful** use of tools Uses **extensive** documentation in a project notebook as work progresses Uses **extensive** version control to maintain **well commented** code and data analysis as work progresses	**Sometimes strategic** and **appropriate** use of tools Uses **adequate** documentation in a project notebook as work progresses Uses **adequate** version control to maintain **commented** code and data analysis as work progresses	**Somewhat appropriate** use of tools Uses **limited** documentation in a project notebook as work progresses Maintains **limited** intermediate code and analysis as work progresses, with **limited** comments	Use of tools **lacks strategy or purpose** **Does not maintain** documentation in a project notebook as work progresses **Does not** maintain intermediate versions of code or analysis. Code and analysis **lack** comments
5.b.iii	Description of Methods	Prose **clearly** and **thoroughly** explains how the data were analyzed	Prose **adequately** explains how the data were analyzed with **minor ambiguities** or missing details	Prose explains how the data were analyzed with **significant ambiguities** or missing details	Prose **predominantly is missing** necessary details about how data were analyzed
5.a.iii 5.a.iv	Knowledge Produced	**Meaningful** artifacts **clearly** communicate answers to **all** questions The report **clearly** explains how the analysis contributes **insight** to the area of investigation	**Meaningful** artifacts communicate answers to **most** questions, with **minor** areas for improvement The report **clearly** explains how the analysis contributes **information** to the area of investigation	Artifacts communicate answers to **some** questions with **some inconsistent or inappropriate** analysis The report explains how the analysis contributes **information** to the area of investigation	Artifacts are inconsistent with data or analsysis The report **confuses or misleads** and **weakly connects** data to the area of investigation
		Quality writing **exemplifies all** of the seven traits:	Adequate writing **exemplifies most** of the seven traits:	Writing **demonstrates some** of the seven traits: • convey	Writing **needs significant improvement in most** regards:

Standard	Criterion	(Level 4)	(Level 3)	(Level 2)	(Level 1)
5a 5b C2	Communication	• convey meaningful ideas • well organized • use scientific voice • use precise words • use smoothly constructed sentences • follow conventions such as grammar • present neatly Visualizations are **correct** and **significantly enhance** the answers to the questions posed	• convey meaningful ideas • well organized • use scientific voice • use precise words • use smoothly constructed sentences • follow conventions such as grammar • present neatly Visualizations are **mostly correct** and **contribute to** answering the questions posed	meaningful ideas • well organized • use scientific voice • use precise words • use smoothly constructed sentences • follow conventions such as grammar • present neatly Visualizations are **mostly correct** and make a **limited contribution** to answering the questions posed	• convey meaningful ideas • well organized • use scientific voice • use precise words • use smoothly constructed sentences • follow conventions such as grammar • present neatly Visualizations are **incorrect** or weakly connect the data and the investigation
C2	Collaboration	• Provides **helpful original** input to others • **Promotes** positive, productive, and respectful team dynamic • **Encourages** and incorporates input from others • **Promotes** equitable workload	• Provides **adequate original** input to others • **Maintains** positive, productive, and respectful team dynamic • Positively **incorporates** input from others • **Maintains** equitable workload	• **Significant** but **limited** input • **Usually maintains** positive, productive, and respectful team dynamic • **Receives** input from others • Shares workload **somewhat** equitably	• **Limited** input • Is **not promoting** positive, respectful, or productive team dynamic • **Discourages** or is unresponsive to input from others • **Does not** promote equitable workload
5.b.i 5.b.iii	Scalability	The entire method could **easily** scale to use a much larger data set	**With minor improvements in efficiency**, the method could scale to use a much larger data set	**With significant revision**, the method could scale to use a much larger data set	The method could **not** scale to use a much larger data set without a new approach

Other comments: