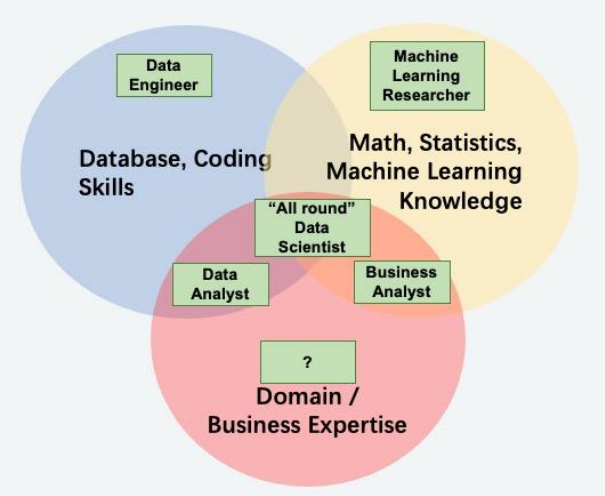
Essay Questions:

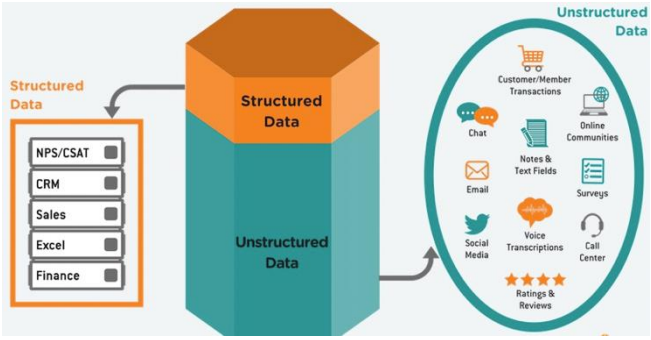
Question 1 (15 points): Provide a sufficient description about the knowledge and skills that one needs to possess to become a successful data analyst.



Mention at least 3 key words from the figure.

Example:
A successful data analyst possesses knowledge and skills coming from 3 domains:
1. Computer programming
2. Math / Stat
3. Field knowledge
He is able to use tools such as programming and mathematical modeling to solve problems if different fields of applications such as in business

Question 2 (15 points): What is structured and unstructured data respectively? Please provide your definitions with examples.



1. Structured data can be processed by machine directly, while unstructured data cannot. For example the data on student grades collected by the Academic Registry are structured data while the content of student emails is unstructured data.

2. Structured data are stored in well designed database such as sales data of a company of the customer relationship management system of a company, while unstructured data could be collected and stored but not in specifically designed database such as phone calls and blog posting on weibo.

Question 3 (16 points): What is the relation and difference between computer science, statistics, and data analytics?

If viewed as a pipeline, data analytics is the bridge that connects statistics and computer science. It focuses on using statistical methods to discover insights from data. Statistics is more traditional and theoretical. Computer science focuses on solving all problem in a computable way, including topics in computability, algorithms, system design, networks, artificial intelligence, software engineering, etc.

(As long as student can describe their understanding in a reasonable way, we'd be lenient with grading.)

Hands-on Data Analytics, Fall2022

Multiple Choices (Only one choice is correct. Each question is worth **6 points**)

1. Where can we use data analytics? ___E (All of the above)

A. Autonomous Driving B. Financial markets prediction
C. Recommender system D. Predicting temperature E. All of the above

2. Which of the following is NOT structured data? ____B (Email files)

A. CSV files B. Email files C. JSON files D. XML files

3. The following data looks like ____B (JSON)

```
{
  "string": "Hi",
  "number": 2.5,
  "boolean": true,
  "null": null,
  "object": { "name": "Kyle", "age": 24 },
  "array": ["Hello", 5, false, null, { "key": "value", "number": 6 }],
  "arrayOfObjects": [
    { "name": "Jerry", "age": 28 },
    { "name": "Sally", "age": 26 }
  ]
}
```

A. CSV
B. JSON
C. XML
D. Excel

4. In ".csv" format, **new records** are separated by which of the following? ____B___

A. Comma B. New Line C. Semicolon D. Tab E. Undefined

5. Consider the following table, what is the corresponding CSV format? _____A_____

2021	Data Science	ZHANG San	Piano, basketball,"laugh"

A. 2021, Data Science, ZHANG San, "Piano, basketball, ""laugh"""
B. 2021, Data Science, ZHANG San, ""Piano, basketball, ""laugh""""
C. 2021, Data Science, ZHANG San, Piano, basketball, "laugh"
D. 2021, Data Science, ZHANG San, "Piano, basketball, "laugh""

6. What is the data type of **Blood Type: {A, B, AB, O}**? _____A_____

A. Categorical B. Ordinal C. Text D. Numerical (discrete) E. Numerical (continuous)

7. What is the data type of **Weekday: {Mon, Tue, Wed, Thu, Fri}**?_____B____

A. Categorical B. Ordinal C. Text D. Numerical (discrete) E. Numerical (continuous)

8. Consider the following data table, which statement is **NOT true?** _____C_____

A. The table has five attributes
B. The table has nine instances
C. "Sex" and "Blood" are ordinal data
D. "Age" and "Height" are numeric data
E. "Drug" is categorical data

9. KNIME uses different colors to represent different status of a node. Which of the following statement is **incorrect**? _____D_____

   A. A red node means the node has not been configured.
   B. A yellow node means the node has been properly configured, but not executed yet.
   C. A Green node means the node has been executed successfully.
   D. When a node is re-configured properly, its color will turn red again. (Assuming it was green previously.)