1.What is data science?

Data science combines math and statistics, specialized programming, advanced analytics, artificial intelligence (AI), and machine learning with specific subject matter expertise to uncover actionable insights hidden in an organization’s data. These insights can be used to guide decision making and strategic planning.

2.Importance of data science?

In data science, statistics is at the core of sophisticated machine learning algorithms, capturing and translating data patterns into actionable evidence. Data scientists use statistics to gather, review, analyze, and draw conclusions from data, as well as apply quantified mathematical models to appropriate variables

3.What is data engineering?

Data engineering is the practice of designing and building systems for collecting, storing, and analyzing data at scale. It is a broad field with applications in just about every industry. Organizations have the ability to collect massive amounts of data, and they need the right people and technology to ensure it is in a highly usable state by the time it reaches data scientists and analysts.

4.What is data visualization?

Data visualization is the graphical representation of information and data. By using v[isual elements like charts, graphs, and maps](https://www.tableau.com/data-insights/reference-library/visual-analytics), data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. Additionally, it provides an excellent way for employees or business owners to present data to non-technical audiences without confusion.

5.What is data cleaning?

Data cleaning is the process of preparing data for analysis by removing or modifying data that is incorrect, incomplete, irrelevant, duplicated, or improperly formatted.

6.What is python and why we use it?

Python is a computer programming language often used to build websites and software, automate tasks, and conduct data analysis. Python is a general-purpose language, meaning it can be used to create a variety of different programs and isn't specialized for any specific problems.

7.What is data in statics?

 Data are measurements or observations that are collected as a source of information. There are a variety of different types of data, and different ways to represent data.

8.What is statistics?

Statistics is a branch of applied mathematics that involves the collection, description, analysis, and inference of conclusions from quantitative data.

9.What is mathematics?

**Mathematics** is an area of [knowledge](https://en.wikipedia.org/wiki/Knowledge) that includes the topics of numbers, formulas and related structures, shapes and the spaces in which they are contained, and quantities and their changes.

10.What is computer science?

Computer science is the study of computers, including computational theory, hardware and software design, algorithms and the way humans interact with technology.

11.What is programming language?

A programming language is a computer language used by programmers to communicate with computers.

Assignment – 3

1.What is open source?

**Open-source software** (**OSS**) is [computer software](https://en.wikipedia.org/wiki/Software) that is released under a [license](https://en.wikipedia.org/wiki/Open-source_license) in which the [copyright](https://en.wikipedia.org/wiki/Copyright) holder grants users the rights to use, study, change, and [distribute the software](https://en.wikipedia.org/wiki/Software_distribution) and its [source code](https://en.wikipedia.org/wiki/Source_code) to anyone and for any purpose.[[1]](https://en.wikipedia.org/wiki/Open-source_software#cite_note-1)[[2]](https://en.wikipedia.org/wiki/Open-source_software#cite_note-2) [Open-source](https://en.wikipedia.org/wiki/Open_source) software may be developed in a collaborative, public manner.

2. What is paid tool?

A paid tool is that apply charges according to storage or time

3. What is anaconda and why we use it?

Anaconda is an open-source distribution for python and R. It is used for data science, machine learning, deep learning, etc.

Anaconda Python is a free, open-source platform that allows you to write and execute code in the programming language Python

4.What is difference between version in python?

The two major versions of Python that were introduced to the world were Python 2 and Python 3. Although they are both just different versions of the same programming language, surprisingly, there are striking differences between the two

|  |  |  |
| --- | --- | --- |
| Storage of Strings | In Python 2, strings are stored as ASCII by default. | In Python 3, strings are stored as UNICODE by default. |

5.What is difference between .exe file and .setup file?

.setup file:

a file needed to set up or install a computer program.

.exe file:

An executable file (EXE file) is a computer file that contains an encoded sequence of [instructions](https://www.techtarget.com/whatis/definition/instruction) that the system can execute directly when the user clicks the file icon. Executable files commonly have an EXE [file extension](https://www.techtarget.com/whatis/definition/file-format), but there are hundreds of other executable file formats.

6. What is difference between anaconda navigator and anaconda prompt?

Anaconda works on your command line interface such as Anaconda Prompt on Windows and terminal on macOS and Linux. Navigator is a desktop graphical user interface that allows you to launch applications and easily manage anaconda packages, environments, and channels without using command-line commands.

7. What is difference between jupyter and spider tool?

Jupitor is used for development purpose

Spider is used for deployment purpose

Jupyter is an interactive Python notebook where you can run code, visualize data and include text all in one document, while Spyder is an IDE specifically for scientific programming in Python. Consider Jupyter if you work on data-driven projects where you need to easily present data to a non-technical audience.

An alternative search strategy tool for qualitative/mixed methods research is outlined: SPIDER (Sample, Phenomenon of Interest, Design, Evaluation, Research type).

8.What is difference between development,testing and deployment?

Development, testing, acceptance and production (DTAP) is a phased approach to software testing and deployment. The four letters in DTAP denote the following common steps: Development: The program or component is developed on a development system. This development environment might have no testing capabilities.

Deployment testing refers to test installation (now a days with Uninstall) process for developed software. It is also known as Installation testing or Implementation testing. The testing process may involve full, partial or upgrades install/uninstall processes.17-Dec-2012

9.Why we use jupyter terminal?

JupyterLab terminals provide full support for system shells (bash, tsch, etc.) on Mac/Linux and PowerShell on Windows. You can run anything in your system shell with a terminal, including programs such as vim or emacs.

**1. Programming Practice**

## 2. Collaborating Across Projects and Tools

## 3. Data Organization and Cleaning

## 4. Data Visualization and Sharing

## 5. Teaching Data Science Skills

10.What is files running and clusters in jupyter?

With jupyter Notebook cluster, you can run notebook on the local machine and connect to the notebook on the cluster by setting the appropriate port number