Task1:

**design pattern** is a general repeatable solution to a commonly occurring problem in software design. A design pattern isn't a finished design that can be transformed directly into code. It is a description or template for how to solve a problem that can be used in many different situations.

Benefits of **design pattern**s

1. Foresee and rectify future problems easily.
2. Helps in maintaining binary compatibility with subsequent releases.
3. Just by following [SOLID Principles](http://www.vishalchovatiya.com/single-responsibility-principle-in-cpp-solid-as-a-rock/) helps greatly in agile or adaptive software development.
4. The solution facilitates the development of highly [cohesive](https://en.wikipedia.org/wiki/Cohesion_(computer_science)) modules with minimal coupling. Thus, increasing extensibility and reusability.
5. There are some patterns like Facade, Proxy, etc which encapsulates the complexity in itself to provide easy and intuitive interface to the client. Thus, making the overall system easier to understand and reduce learning curve.
6. Design Patterns make communication between designers and developers more crystal and precise. A developer can immediately picture the high-level design in their heads when they refer to the name of the pattern used to solve a particular issue when discussing software design.

Task2:

**Statistical tests**

Z-test

Your sample size is greater than 30. Otherwise, use a t test.

Data points should be independent from each other. In other words, one data point isn’t related or doesn’t affect another data point.

Your data should be normally distributed. However, for large sample sizes (over 30) this doesn’t always matter.

Your data should be randomly selected from a population, where each item has an equal chance of being selected.

Sample sizes should be equal if at all possible.

T-test

A t-test is used when the population parameters (mean and standard deviation) are not known.

Paired T-Test

A paired t-test is used when we are interested in the difference between two variables for the same subject.

Independent T-test

You want to know if two groups are different on your variable of interest

Your variable of interest is continuous

You have two and only two groups

You have independent samples

You have a normal variable of interest

One sample t-test

The one-sample t-test is used when we want to know whether our sample comes from a particular population but we do not have full population information available to us. For instance, we may want to know if a particular sample of college students is similar to or different from college students in general. The one-sample t-test is used only for tests of the sample mean. Thus, our hypothesis tests whether the average of our sample (M) suggests that our students come from a population with a know mean (m) or whether it comes from a different population.

ANOVA Test

use ANOVA to help you understand how your different groups respond, with a null hypothesis for the test that the means of the different groups are equal. If there is a statistically significant result, then it means that the two populations are unequal (or different).

Non parametric statistical test

Non parametric tests are used when your data isn’t normal. Therefore the key is to figure out if you have normally distributed data. For example, you could look at the distribution of your data. If your data is approximately normal, then you can use parametric statistical tests.

Chi-square test

A chi-square test is a statistical test used to compare observed results with expected results. The purpose of this test is to determine if a difference between observed data and expected data is due to chance, or if it is due to a relationship between the variables you are studying. Therefore, a chi-square test is an excellent choice to help us better understand and interpret the relationship between our two categorical variables.

Task3:

**NFT**

A non-fungible token (NFT) is a unique and irreplaceable unit of data stored in a digital record (called a blockchain). NFTs can be used to represent easily reproducible items such as photos, videos, audio, and other types of digital files as unique items (similar to a Certificate of Authenticity), using blockchain technology to create a public and verified proof of ownership. Copying the original file is not limited to the owner of the NFT, and it can be copied and shared like any file. These tokens are distinguished from other cryptocurrencies such as Bitcoin and Ethereum in that they are not exchangeable

Task4:

**Type of data**

Quantitative data

Quantitative data seems to be the easiest to explain. It answers key questions such as “how many, “how much” and “how often”.

Quantitative data can be expressed as a number or can be quantified. Simply put, it can be measured by numerical variables.

Quantitative data can be represented by a wide variety of statistical types of graphs and charts such as line, bar graph, scatter plot.

Qualitative data

Qualitative data can answer questions such as “how this has happened” or and “why this has happened”.

Qualitative data can’t be expressed as a number and can’t be measured. Qualitative data consist of words, pictures, and symbols, not numbers.

Qualitative data is also called categorical data because the information can be sorted by category, not by number.

Nominal data

Nominal data is used just for labeling variables, without any type of quantitative value.

The nominal data just name a thing without applying it to order. Actually, the nominal data could just be called “labels.”

Ordinal data

Ordinal data is data which is placed into some kind of order by their position on a scale. Ordinal data may indicate superiority.

However, you cannot do arithmetic with ordinal numbers because they only show sequence.

We can also assign numbers to ordinal data to show their relative position. But we cannot do math with those numbers. For example: “first, second, third…etc.”

Discrete data

Discrete data is a count that involves only integers. The discrete values cannot be subdivided into parts.

For example, the number of children in a class is discrete data. You can count whole individuals. You can’t count 1.5 kids.

It has a limited number of possible values e.g. days of the month.

Continuous data

Continuous data is information that could be meaningfully divided into finer levels. It can be measured on a scale or continuum and can have almost any numeric value.

For example, you can measure your height at very precise scales — meters, centimeters, millimeters and etc.

You can record continuous data at so many different measurements – width, temperature, time, and etc.

The continuous variables can take any value between two numbers. For example, between 68 and 90 inches.

Task6:

**Uses of probability in AI**

Sampling – Dealing with non-deterministic processes

Pattern recognition

Training – use in Maximum likelihood estimation

Developing specific algorithms

Hyperparameter optimization

Model evaluation

Task7:

Our World in Data

Worldometer – real time world statistics

World Bank Open Data | Data

The World Economic Forum

National Statistical Agencies of Other Countries

Task8:

**Predictive analytics**

Predictive analytics is the use of data, statistical algorithms, and machine learning techniques to identify the likelihood of future outcomes based on historical data. The goal is to go beyond knowing what has happened to providing a best assessment of what will happen in the future.