Q1. what is the significance of p values?
ans : A statistically significant test result (P ≤ 0.05) means that the test hypothesis is false or should be rejected.
A P value greater than 0.05 means that no effect was observed.
Since the introduction of P value in 1900 by Pearson [1],
the P values are the preferred method to summarize the results of medical articles.
Because the P value is the outcome of a statistical test,
many authors and readers consider it the most important summary of the statistical analyses.

Q2. how to calculate accuracy using the confusion matrix?
ans: Accuracy: It gives you the overall accuracy of the model,
meaning the fraction of the total samples that were correctly classified by the classifier.
To calculate accuracy, use the following formula: (TP+TN)/(TP+TN+FP+FN).

Q6. what is ROC curve?
ans : An ROC curve (receiver operating characteristic curve) is a graph showing the performance of a classification
model at all classification thresholds.
This curve plots two parameters: True Positive Rate. False Positive Rate

Q8 . what is normal distribution?
ans .What Is a Normal Distribution? Normal distribution, also known as the Gaussian distribution,
is a probability distribution that is symmetric about the mean,
showing that data near the mean are more frequent in occurrence than data far from the mean.

A normal distribution has a probability distribution that is centered around the mean.
This means that the distribution has more data around the mean.
The data distribution decreases as you move away from the center.
The resulting curve is symmetrical about the mean and forms a bell-shaped distribution.

Q9. what is cross validation?
ans : Cross-validation is a technique for evaluating ML models by training several ML models on subsets of
the available input data and evaluating them on the complementary subset of the data.
Use cross-validation to detect overfitting, ie, failing to generalize a pattern

There are various types of cross-validation. However,
mentioned above are the 7 most common types - Holdout, K-fold, Stratified k-fold, Rolling, Monte Carlo,
Leave-p-out, and Leave-one-out method.
Although each one of these types has some drawbacks,
they aim to test the accuracy of a model as much as possible

Definition. Cross-Validation is a statistical method of evaluating and comparing learning algorithms by
dividing data into two segments:
one used to learn or train a model and the other used to validate the model

Q10 . which according to you is the most important skill that makes you a good data scientist?

ans :for technical skill
Skill #1- Programming
You need to have knowledge of various programming languages, such as Python, Perl, C/C++,

SQL, and Java, with Python being the most common coding language required in data science roles. These programming languages help data scientists organize unstructured data sets.

Pro tip: To learn more about the Programming languages, click here!

Skill #2- Knowledge of SAS and Other Analytical Tools
An understanding of analytical tools is a helpful data scientist skill for extracting valuable information from an organized data set. SAS, Hadoop, Spark, Hive, Pig, and R are the most popular data analytical tools that data scientists use. Certifications can help you establish your expertise in these analytical tools and help you gain this valuable data scientist skill!

Skill #3- Adept at Working with Unstructured Data
Data scientists should have experience working with unstructured data that comes from different channels and sources. For example, if a data scientist is working on a project to help the marketing team provide insightful research, the professional should be well adept at handling social media as well.

Some of the other skills required are Machine Learning, Artificial intelligence, Deep learning, Probability and Statistics.

for non - technical skill

Skill #1- A Strong Business Acumen
The best way to productively channel technical skills is to have strong business acumen. Without it, an aspiring data scientist may not be able to discern the problems and potential challenges that need to be solved in order for an organization to grow. This is essential for helping the organization you're working for explore new business opportunities.

Skill #2 - Strong Communication Skills
Next on the list of top data scientist skills is communication. Data scientists clearly understand how to extract, understand, and analyze data. However, for you to be successful in your role, and for your organization to benefit from your services, you should be able to successfully communicate your findings with team members who don't have the same professional background as you.

Skill #3 - Great Data Intuition
This is perhaps one of the most significant non-technical data scientist skills. Valuable data insights are not always apparent in large data sets, and a knowledgeable data scientist has intuition and knows when to look beyond the surface for insightful information. This makes data scientists more efficient in their work, and gaining this skill comes from experience and the right training. However, this data scientist skill comes with experience and bootcamps are a great way of polishing it.