python

python is a high-level language, used in web dev, data sci, ai/ml, automation, etc. its known for its readability, which means code is easier to write, understand and maintain. its backed by library support, so we dont have to build everything from scratch, there's probably a library that already does what we need.

data types

primitive types
int => 10
float => 10.5
string => "kamlesh"
boolean => true/false
nonetype => none

collection types

```
list => [1,2,3]
    list are an ordered collection of data which is mutable and allows duplicate values and can
    be accessed through indexing.
tuple => (1,2,3)
    tuple are an ordered collection of data which can not be modified thus making them immutable,
    tuple allows duplicates and indexing can be used.
dictionary => {"name":"kamlesh","age":19}
    dictionary are defined by using "{}" brackets and it is an unordered collection which can be
    modified. to access the data from the dicts you have to use keys assigned to the value.
set => {1,2,3}
    set are an unordered collection data type that is mutable but doesn't allows duplicates and
    also can not be accessed through indexing.
```

functions in python
a function is a block of code that runs only when called, and can take inputs and give back an output.

```
syntax:
    def greet(name):
        return "hello", + name
def = keyword to define a function
greet = name of the function
name = parameter (u can set multiple parameters)
return = gives back a result

calling a function:
    greet("kamlesh")
    output: hello, kamlesh

use case:
```

```
    1. reuse the code
    2. organize the logic
    3. avoid repeating the same block

the variables created inside a function cannot be accessed globally.

parameter => placeholder variable in the function definition which appears when we define the
function.

argument => actual value that we pass into the function which is used when we call the function.

example:
    def hello(name, age):
        print(f"hello {name}, you are {age} years old.")

    hello("kamlesh", 19)

here,
 (name, age) is a parameter that is passed in the function.
 hello("kamlesh", 19) they are the arguments that are passed as the actual value we want as the
 output.
```

hypothesis testing is a statistical method used to evaluate claims about a population based on sample data.
null hypothesis => which assumes there is no effect or difference.
alternate hypo => assumes there is a difference in what the claim was made.

p-value
the probability that the result has happened just by random chance.

t-test
compare means of 2 groups
1. one sample t-test => compare group with known value
2. two sample t-test => compare means of two groups

z-test
same as t-test but used when:
1. the sample size is n>30
2. the std dev is given

annova
compare means of 3 or more groups
test scores of group a, b and c

R
r is a programming language used mostly for data analysis, statistics, and visualizations.

data types
numeric => 10, 3.4
character => "kamlesh"
logical => true/false

integer => 5L
factor => categorical data

data frame
a data frame is a 2d table structure made up of rows and columns to store tabular data.

R libraries/packages

1.ggplot2
it is used to create complex and beautiful plots in a simple way. used for data visualization.

2.dplyr
it is used for data manipulation filtering, selecting and grouping the data.

3.caret
it is used to training and evaluating machine learning models in r.

4.tidyr
it is used for cleaning and reshaping the messy data.

5.readr
it is used for reading csv and txt files in r.

python libraries

1.pandas
it is used for data manipulation and analysis; it is used to read data, clean it, filter rows and explore datasets.

2.numpy
numpy is a python library used for handling arrays and doing fast numeric computing.

3.matplotlib
matplotlib is used for visualizing the data in line plots, barchats, piecharts, etc.

4.seaborn
it is also used for visualization built on top of matplotlib for more prettier and complex plots like histograms, heatmaps and boxplots.

5.scikit-learn
it is used to train ml models like decision trees, svm, linear regression, etc.

ml algos

1. liner regression
used for predicting continuous values from the data.
example: predicting house prices based on area.

2. logisitc regression
it is used for classification binary (yes/no, 0/1)
example: will verify the authencity of the user confidentials info

3. decision tree
used for classification and regression
example: loan approval based on income, credit, etc.

4. random forest
used for powerful classification/regression using multiple trees.
rf reduces overfitting and increases accuracy.

5. knn (k nearest neighbors)
used for classification based on similarity
knn looks at nearby data points and picks the most common class.

6.support vector machine
used for classfication with clear margin
ex: email spam detection (spam vs not spam)

7.naive bayes
used for classification using probability
its based on bayes theorem; works well for nlp tasks.

8. k-means clustering
used for unsupervised learning grouping similar data
ex: customer segmentation (grouping buyers based on their behaviors)

9. principal component analysis
used for dimensionality reduction

10. neural networks
used for complex tasks like img recog, speech, etc.

11. xgboost/gradient boosting
used for ml competitions and structured data