PYTHON

Q.1 :- Write a program that takes a string as input, and counts the frequency of each word in the string, there might be repeated characters in the string. Your task is to find the highest frequency and returns the length of the highest-frequency word.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python(kunalchauhan).py>

Q.2 :- Consider a string to be valid if all characters of the string appear the same number of times. It is also valid if he can remove just one character at the index in the string, and the remaining characters will occur the same number of times. Given a string, determine if it is valid. If so, return YES , otherwise return NO .

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python2(kunalchauhan).py>

Q.3 :- Write a program, which would download the data from the provided link, and then read the data and convert that into properly structured data and return it in Excel format.

Note - Write comments wherever necessary explaining the code written.

Link - <https://raw.githubusercontent.com/Biuni/PokemonGO-Pokedex/master/pokedex.json>

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python3(kunalchauhan).py>

Q.4 :- Write a program to download the data from the link given below and then read the data and convert the into the proper structure and return it as a CSV file.

Link - <https://data.nasa.gov/resource/y77d-th95.json>

Note - Write code comments wherever needed for code understanding.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python4(kunalchauhan).py>

Q.5 :- Write a program to download the data from the given API link and then extract the following data with proper formatting

Link - <http://api.tvmaze.com/singlesearch/shows?q=westworld&embed=episodes>

Note - Write proper code comments wherever needed for the code understanding

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python5(kunalchauhan).py>

Q.6 :- Using the data from Question 3, write code to analyze the data and answer the following questions Note 1.

Draw plots to demonstrate the analysis for the following questions for better visualizations.

2. Write code comments wherever required for code understanding

Insights to be drawn -

● Get all Pokemons whose spawn rate is less than 5%

● Get all Pokemons that have less than 4 weaknesses

● Get all Pokemons that have no multipliers at all

● Get all Pokemons that do not have more than 2 evolutions

● Get all Pokemons whose spawn time is less than 300 seconds.

Note - spawn time format is "05:32”, so assume “minute: second” format and perform the analysis.

● Get all Pokemon who have more than two types of capabilities

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python6(kunalchauhan).py>

Q.7 :- Using the data from Question 4, write code to analyze the data and answer the following questions Note -

1. Draw plots to demonstrate the analysis for the following questions for better visualizations

2. Write code comments wherever required for code understanding

Insights to be drawn -

● Get all the Earth meteorites that fell before the year 2000

● Get all the earth meteorites co-ordinates who fell before the year 1970

● Assuming that the mass of the earth meteorites was in kg, get all those whose mass was more than 10000 kg

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python7(kunalchauhan).py>

Q.8 :- Using the data from Question 5, write code the analyze the data and answer the following questions Note -

1. Draw plots to demonstrate the analysis for the following questions and better visualizations

2. Write code comments wherever required for code understanding

Insights to be drawn -

● Get all the overall ratings for each season and using plots compare the ratings for all the

seasons, like season 1 ratings, season 2, and so on.

● Get all the episode names, whose average rating is more than 8 for every season

● Get all the episode names that aired before May 2019

● Get the episode name from each season with the highest and lowest rating

● Get the summary for the most popular ( ratings ) episode in every season

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python8(kunalchauhan).py>

Q.9 :- Write a program to read the data from the following link, perform data analysis and answer the following questions

Note -

1. Write code comments wherever required for code understanding

Link - https://data.wa.gov/api/views/f6w7-q2d2/rows.csv?accessType=DOWNLOAD

Insights to be drawn -

● Get all the cars and their types that do not qualify for clean alternative fuel vehicle

● Get all TESLA cars with the model year, and model type made in Bothell City.

● Get all the cars that have an electric range of more than 100, and were made after

2015

● Draw plots to show the distribution between city and electric vehicle type

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python9(kunalchauhan).py>

Q.10 :- Write a program to count the number of verbs, nouns, pronouns, and adjectives in a given particular phrase or paragraph, and return their respective count as a dictionary.

Note -

1. Write code comments wherever required for code

2. You have to write at least 2 additional test cases in which your program will run successfully and provide an explanation for the same.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/python10(kunalchauhan).py>

STATISTICS

Q-1. A university wants to understand the relationship between the SAT scores of its

applicants and their college GPA. They collect data on 500 students, including their SAT

scores (out of 1600) and their college GPA (on a 4.0 scale). They find that the correlation

coefficient between SAT scores and college GPA is 0.7. What does this correlation

coefficient indicate about the relationship between SAT scores and college GPA?

Answer:- A correlation coefficient of 0.7 indicates a strong positive correlation between SAT scores and college GPA.

A correlation coefficient measures the strength and direction of the linear relationship between two variables. In this case, a correlation coefficient of 0.7 suggests a strong positive linear relationship between SAT scores and college GPA.

A positive correlation means that as SAT scores increase, college GPA also tends to increase. The strength of the correlation is considered strong because the correlation coefficient is close to 1 (the maximum value for a positive correlation).

Therefore, based on the correlation coefficient of 0.7, we can conclude that there is a strong positive relationship between SAT scores and college GPA for the given dataset of 500 students from the university.

Q-2. Consider a dataset containing the heights (in centimeters) of 1000 individuals. The

mean height is 170 cm with a standard deviation of 10 cm. The dataset is approximately

normally distributed, and its skewness is approximately zero. Based on this information,

answer the following questions:

a. What percentage of individuals in the dataset have heights between 160 cm

and 180 cm?

b. If we randomly select 100 individuals from the dataset, what is the probability

that their average height is greater than 175 cm?

c. Assuming the dataset follows a normal distribution, what is the z-score

corresponding to a height of 185 cm?

d. We know that 5% of the dataset has heights below a certain value. What is

the approximate height corresponding to this threshold?

e. Calculate the coefficient of variation (CV) for the dataset.

f. Calculate the skewness of the dataset and interpret the result.

Answer:- (a):-

First, we calculate the z-scores for the heights of 160 cm and 180 cm using the formula:

z = (x - μ) / σ

Where:

x = height value

μ = mean height

σ = standard deviation

For 160 cm:

z1 = (160 - 170) / 10 = -1 = 0.1587

For 180 cm:

z2 = (180 - 170) / 10 = 1 = 0.8413

percentage of individuals in the dataset have heights between 160 cm

and 180 cm = 84.13 - 15.87 = 68.26%

(b):- Using this information, we calculate the z-score for the sample mean height of 175 cm using the formula:

z = (x - μ) / (σ / √n) = (175-170)/(10/10) = 5

Where:

x = sample mean height value

μ = mean height

σ = standard deviation

n = sample size (100 in this case)

(c):- To calculate the z-score corresponding to a height of 185 cm, we use the formula:

z = (x - μ) / σ = (185 - 170)/10 = 1.5

Z score = 0.9332

This calculated z-score represents the number of standard deviations above or below the mean that the height of 185 cm is.

(d):- z score for 5 percentile = -1.645

z = (x - μ) / σ

-1.645 = (x-170)/10

x= 153.55

(e):- The formula to calculate CV is:

CV = (σ / μ) \* 100 = (10/170)\*100 = 5.88

(f):- To calculate the skewness, we can use the skewness formula provided by the third standardized moment

skewness = (3 \* (μ - median)) / σ

But as given in the question that the skewness of the curve is zero so a skewness value close to zero suggests that the dataset is approximately symmetric.

Q-3. Consider the ‘Blood Pressure Before’ and ‘Blood Pressure After’ columns from the

data and calculate the following

https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\_

a. Measure the dispersion in both and interpret the results.

b. Calculate mean and 5% confidence interval and plot it in a graph

c. Calculate the Mean absolute deviation and Standard deviation and interpret

the results.

d. Calculate the correlation coefficient and check the significance of it at 1% level

of significance.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/Statistics3(kunalchauhan).py>

Q-4. A group of 20 friends decide to play a game in which they each write a number between 1 and 20 on a slip of paper and put it into a hat. They then draw one slip of paper at random. What is the probability that the number on the slip of paper is a perfect square (i.e., 1, 4, 9, or 16)?

Answer:- Favorable outcomes: The perfect square numbers between 1 and 20 are 1, 4, 9, and 16. So, there are 4 favorable outcomes.

Total outcomes: Since there are 20 slips of paper with numbers from 1 to 20, there are 20 possible outcomes.

Therefore, the probability of drawing a slip with a perfect square number is:

Probability = Favorable outcomes / Total outcomes = 4 / 20 = 0.2

So, the probability is 0.2 or 20%.

Q-5. A certain city has two taxi companies: Company A has 80% of the taxis and Company B has 20% of the taxis. Company A's taxis have a 95% success rate for picking up passengers on time, while Company B's taxis have a 90% success rate. If a randomly selected taxi is late, what is the probability that it belongs to Company A?

Answer:- To solve this problem, we can use Bayes' theorem. Let's define the events:

A: Taxi belongs to Company A

B: Taxi is late

We are given the following probabilities:

P(A) = 0.8 (Company A has 80% of the taxis)

P(B|A) = 0.05 (Company A's taxis have a 95% success rate, so the probability of being late is 1 - 0.95 = 0.05)

P(B|not A) = 0.1 (Company B's taxis have a 90% success rate, so the probability of being late is 1 - 0.9 = 0.1)

We need to find P(A|B), which represents the probability that the taxi belongs to Company A given that it is late.

Using Bayes' theorem:

P(A|B) = (P(B|A) \* P(A)) / P(B)

To calculate P(B), we can use the law of total probability:

P(B) = P(B|A) \* P(A) + P(B|not A) \* P(not A)

P(not A) = 1 - P(A) = 1 - 0.8 = 0.2

Now we can substitute the values and calculate P(A|B):

P(B) = P(B|A) \* P(A) + P(B|not A) \* P(not A) = 0.05 \* 0.8 + 0.1 \* 0.2 = 0.04 + 0.02 = 0.06

P(A|B) = (P(B|A) \* P(A)) / P(B) = (0.05 \* 0.8) / 0.06 = 0.04 / 0.06 = 4/6 = 2/3

Therefore, the probability that a randomly selected late taxi belongs to Company A is 2/3 or approximately 0.67.

Q-6. A pharmaceutical company is developing a drug that is supposed to reduce blood

pressure. They conduct a clinical trial with 100 patients and record their blood

pressure before and after taking the drug. The company wants to know if the change

in blood pressure follows a normal distribution.

https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\_

Answer:- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/Statistics6(kunalchauhan).py>

Q-7. The equations of two lines of regression, obtained in a correlation analysis

between variables X and Y are as follows:

and . 2𝑋 + 3 − 8 = 0 2𝑌 + 𝑋 − 5 = 0 The variance of 𝑋 = 4 Find the

a. Variance of Y

b. Coefficient of determination of X and Y

c. Standard error of estimate of X on Y and of Y on X.

Answer:- a. Variance of Y: the coefficient of Y is 2. Therefore, the variance of Y (Var(Y)) can be calculated using the formula:

Var(Y) = b^2 \* Var(X)

Var(Y) = (2^2) \* 4

Var(Y) = 16

So, the variance of Y is 16.

b. Coefficient of determination of X and Y:

For X on Y:

R^2(X|Y) = SSreg(X|Y) / SStotal(X)

From the given equation, 2X + 3 - 8 = 0, we can see that the coefficient of determination for X on Y is 1 since all the variations in X are explained by Y. So, R^2(X|Y) = 1.

For Y on X:

R^2(Y|X) = SSreg(Y|X) / SStotal(Y)

From the given equation, 2Y + X - 5 = 0, we can see that the coefficient of determination for Y on X is 0 since all the variations in Y are not explained by X. So, R^2(Y|X) = 0.

Q-8.The anxiety levels of 10 participants were measured before and after a new therapy.

The scores are not normally distributed. Use the Wilcoxon signed-rank test to test whether

the therapy had a significant effect on anxiety levels. The data is given below: Participant

Before therapy After therapy Difference

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/statistics8(kunalchauhan).py>

Q-9. Given the score of students in multiple exams

Test the hypothesis that the mean scores of all the students are the same. If not, name the

student with the highest score.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/statistics9(kunalchauhan).py>

Q-10. A factory produces light bulbs, and the probability of a bulb being defective is 0.05.

The factory produces a large batch of 500 light bulbs.

a. What is the probability that exactly 20 bulbs are defective?

b. What is the probability that at least 10 bulbs are defective?

c. What is the probability that at max 15 bulbs are defective?

d. On average, how many defective bulbs would you expect in a batch of 500?

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/statistics10(kunalchauhan).py>

Q-11. Given the data of a feature contributing to different classes

https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp

=share\_

a. Check whether the distribution of all the classes are the same or not.

b. Check for the equality of variance/

c. Which amount LDA and QDA would perform better on this data for

classification and why.

d. Check the equality of mean for between all the classes.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/statistics11(kunalchauhan).py>

Q-12. A pharmaceutical company develops a new drug and wants to compare its

effectiveness against a standard drug for treating a particular condition. They conduct a

study with two groups: Group A receives the new drug, and Group B receives the standard

drug. The company measures the improvement in a specific symptom for both groups after

a 4-week treatment period.

a. The company collects data from 30 patients in each group and calculates the

mean improvement score and the standard deviation of improvement for each

group. The mean improvement score for Group A is 2.5 with a standard

deviation of 0.8, while the mean improvement score for Group B is 2.2 with a

standard deviation of 0.6. Conduct a t-test to determine if there is a significant

difference in the mean improvement scores between the two groups. Use a

significance level of 0.05.

b. Based on the t-test results, state whether the null hypothesis should be

rejected or not. Provide a conclusion in the context of the study.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/statistics12(kunalchauhan).py>

Machine Learning

INTERMEDIATE QUESTIONS :

Q-1. Imagine you have a dataset where you have different Instagram features like username , Caption , Hashtag , Followers , Time\_Since\_posted , and likes , now your task is to predict the number of likes and Time Since posted and the rest of the features are your input features. Now you have to build a model which can predict the number of likes and Time Since posted.

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML1(KunalChauhan).ipynb>

Q-2. Imagine you have a dataset where you have different features like Age ,

Gender , Height , Weight , BMI , and Blood Pressure and you have to classify the people into

different classes like Normal , Overweight , Obesity , Underweight , and Extreme Obesity by using any 4 different classification algorithms. Now you have to build a model which

can classify people into different classes.

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML2(KunalChauhan).ipynb>

Q-3. Imagine you have a dataset where you have different categories of data, Now

you need to find the most similar data to the given data by using any 4 different

similarity algorithms. Now you have to build a model which can find the most similar

data to the given data.

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML3(KunalChauhan).ipynb>

Q-4. Imagine you working as a sale manager now you need to predict the Revenue

and whether that particular revenue is on the weekend or not and find the

Informational\_Duration using the Ensemble learning algorithm

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML4(KunalChauhan).ipynb>

Q-5. Uber is a taxi service provider as we know, we need to predict the high

booking area using an Unsupervised algorithm and price for the location using a

supervised algorithm and use some map function to display the data

Dataset This is the Dataset You can use this dataset for this question.

Answer :-

Q-6. Imagine you have a dataset where you have predicted loan Eligibility using any

4 different classification algorithms. Now you have to build a model which can

predict loan Eligibility and you need to find the accuracy of the model and built-in

docker and use some library to display that in frontend

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML6(KunalChauhan).ipynb>

Q-7. Imagine you have a dataset where you need to predict the Genres of Music

Using an Unsupervised algorithm and you need to find the accuracy of the model, built-in

docker, and use some library to display that in frontend

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML7(kunalchauhan).py>

Q-8. Quora question pair similarity, you need to find the Similarity between two

questions by mapping the words in the questions using TF-IDF, and using a supervised

Algorithm you need to find the similarity between the questions.

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/ML8(KunalChauhan).ipynb>

Q-9. A cyber security agent wants to check the Microsoft Malware so need he came

to you as a Machine learning Engineering with Data, You need to find the Malware

using a supervised algorithm and you need to find the accuracy of the model.

Dataset This is the Dataset You can use this dataset for this question.

Answer :-

Q-10. An Ad- Agency analyzed a dataset of online ads and used a machine learning

model to predict whether a user would click on an ad or not.

Dataset This is the Dataset You can use this dataset for this question.

Answer :-

Advance QUESTIONS :

Q-1. A Social Media Influencer collected data on Facebook friend requests and used

a supervised algorithm to predict whether a user would accept a friend request or

not. Dataset This is the Dataset You can use this dataset for this question. Note : Use

only Dask and Use MLflow

Answer :-

Q-2. A chemist had two chemical flasks labeled 0 and 1 which consist of two

different chemicals. He extracted 3 features from these chemicals in order to

distinguish between them, you provided the results derived by the chemicals and

your task is to create a model that will label chemical 0 or 1 given its three features

and built-in docker and use some library to display that in frontend.

Note : Use only pyspark

Dataset This is the Dataset You can use this dataset for this question.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/MLADVANCE2(KunalChauhan).ipynb>

Q- 3. A company wants to predict the sales of its product based on the money spent

on different platforms for marketing. They want you to figure out how they can

spend money on marketing in the future in such a way that they can increase their

profit as much as possible built-in docker and use some library to display that in

frontend Dataset This is the Dataset You can use this dataset for this question. Note:

Use only Dask

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/MLADVANCE3(KunalChauhan).ipynb>

Q-4. Take any 3 questions and deploy them to AWS using GitHub Actions and

show a demo link

Q-5. Take any 3 questions and deploy them to AWS using Circle-CI and show a demo link

DEEP LEARNING

Question 1 -

Implement 3 different CNN architectures with a comparison table for the MNSIT

dataset using the Tensorflow library.

Note -

1. The model parameters for each architecture should not be more than 8000

parameters

2. Code comments should be given for proper code understanding.

3. The minimum accuracy for each accuracy should be at least 96%

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/DL1(KunalChauhan).ipynb>

Question 2 -

Implement 5 different CNN architectures with a comparison table for CIFAR 10

dataset using the PyTorch library

Note -

1. The model parameters for each architecture should not be more than 10000

parameters

2 Code comments should be given for proper code understanding

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/dl2(kunalchauhan).py>

Question 3 -

Train a Pure CNN with less than 10000 trainable parameters using the MNIST

Dataset having minimum validation accuracy of 99.40%

Note -

1. Code comments should be given for proper code understanding.

2. Implement in both PyTorch and Tensorflow respectively

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/dl3(kunalchauhan).py>

Question 4 -

Design an end-to-end solution with diagrams for object detection use cases

leveraging AWS cloud services and open-source tech

Note -

1. You need to use both AWS cloud services and open-source tech to design the

entire solution

2. The pipeline should consist of a data pipeline, ml pipeline, deployment pipeline,

and inference pipeline.

3. In the data pipeline, you would be designing how to get the data from external or

existing sources and tech used for the same

4. In the ml pipeline, you would be designing how to train the model, and what all

algorithms, techniques, etc. would you be using. Again, tech used for the same 5.

Since this is a deep learning project, the use of GPUs, and how effectively are you

using them to optimize for cost and training time should also be taken into

consideration.

6. In the deployment pipeline, you would be designing how effectively and

efficiently you are deploying the model in the cloud,

7. In the inference pipeline, consider the cost of inference and its optimization

related to computing resources and handling external traffic

8. You can use any tool to design the architecture

9. Do mention the pros and cons of your architecture and how much further it can

be optimized and its tradeoffs.

10. Do include a retraining approach as well.

11. Try to include managed AWS resources for deep learning like AWS Textract,

AWS Sagemaker, etc., and not just general-purpose compute resources like S3,

EC2, etc. Try to mix the best of both services

Answer :-

Question 5 -

In Question 4, you have designed the architecture for an object detection use case

leveraging AWS Cloud, similarly, here you will be designing for Document

Classification use case leveraging Azure Cloud services.

Note -

1. Most of the points are the same as in Question 4, just cloud services will

change

NLP

Q-1. Take any YouTube videos link and your task is to extract the comments from

that videos and store it in a csv file and then you need define what is most

demanding topic in that videos comment section.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/nlp1(kunalchauhan).py>

Q-2. Take any pdf and your task is to extract the text from that pdf and store it in a

csv file and then you need to find the most repeated word in that pdf.

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/nlp2(kunalchauhan).py>

Q-3. from question 2, As you got the CSV and now you need perform key word

extraction from that csv file and do the Topic modeling

Answer :-

Q-4. Take any text file and now your task is to Text Summarization without using

hugging transformer library

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/nlp4(kunalchauhan).py>

Q-5. Now you need build your own language detection with the fast Text model

by Facebook and

Answer :- <https://github.com/Kuna1Chauhan/Assignment_Kunal_Chauhan/blob/main/nlp5(kunalchauhan).py>

Q-6. Generate research papers titles using Bert model and containerize the

application and push to public docker hub

Answer :-

Q-7. Now you need to build your own chatbot using the seq2seq model of

Amazon website by scrape the website and containerize the application and push

to public docker hub

Answer :-

Q-8. Take a any own dataset and build a knowledge bot using Llama model.

Answer :-

Q-9. Using wisher you need transcribe any audio file and then you need to convert

that audio file into text file and now convert that text file into audio file of different

language.

Answer :-

Q-10. Build a whole End- End api and deploy it on Heroku /railways so the task is

that you need build a Auto-Correction of text using NLP

Note: only Jupyter notebook is not allowed from 5th question