

### Question 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.

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

Example input - string = "write write write all the number from from from 1 to 100"

Example output - 5

Explanation - From the given string we can note that the most frequent words are "write" and "from" and the maximum value of both the values is "write" and its corresponding length is 5

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/assignment.py>

### Question 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**.

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

Example input 1 - s = "abc". This is a valid string because frequencies are { "a": 1, "b": 1, "c": 1 }

Example output 1- YES

Example input 2 - s "abcc". This string is not valid as we can remove only 1 occurrence of "c". That leaves character frequencies of { "a": 1, "b": 1, "c": 2 }

Example output 2 - NO

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/q2.py>

### Question 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>

**Data Attributes** - **id**: Identification Number - **int** **num**: Number of the

•

Pokémon in the official Pokédex - **int** **name**: Pokémon name -

•

**string** **img**: URL to an image of this Pokémon - **string** **type**:

•

Pokémon type - **string** **height**: Pokémon height - **float**

•

**weight**: Pokémon weight - **float** **candy**: type of candy used to evolve Pokémon or given

•

when transferred - **string** **candy\_count**: the amount of candies required to evolve - **int**

•

**egg**: Number of kilometers to travel to hatch the egg - **float** **spawn\_chance**:

•

Percentage of spawn chance **(NEW)** - **float** **avg\_spawns**: Number of this pokemon on 10.000 spawns **(NEW)** - **int**

•

**spawn\_time**: Spawns most active at the time on this field. Spawn times are the same for all time zones and are expressed in local time. **(NEW)** - "minutes: seconds" **multipliers**:

Multiplier of Combat Power (CP) for calculating the CP after evolution See below - list of **int**

**weakness**: Types of

•

Pokémon this Pokémon is weak to - list of strings **next\_evolution**: Number and Name of successive evolutions of Pokémon - list of dict **prev\_evolution**: Number and Name of previous evolutions of Pokémon - - list of dict

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/q3.py>

#### Question 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.

**Sample Data** -

```
{
  "name": "Tomakovka",
  "id": "24019",
  "nametype": "Valid",
  "recclass": "LL6",
  "mass": "600",
  "fall": "Fell",
  "year": "1905-01-01T00:00:00.000",
  "reclat": "47.850000",
  "recclong": "34.766670",
  "geolocation": {
    "type": "Point",
    "coordinates": [
      34.76667,
      47.85
    ]
  }
}
```

#### Excepted Output Data Attributes

- Name of Earth Meteorite - string id - ID of Earth
- Meteorite - int nametype - string recclass - string
- mass - Mass of Earth Meteorite - float year - Year at which Earth
- Meteorite was hit - datetime format reclat - float recclong - float
- point coordinates - list of int

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/q4.py>

#### Question 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

**Sample Data** -

```
{
  "id": 2326658,
  "url": "https://www.tvmaze.com/episodes/2326658/westworld-4x05-zhuangzi",
  "name": "Zhuangzi",
  "season": 4,
  "number": 5,
  "type": "regular",
  "airdate": "2022-07-24",
  "airtime": "21:00",
  "airstamp": "2022-07-25T01:00:00+00:00",
  "runtime": 60,
  "rating": {
    "average": 7.8
  },
  "image": {
    "medium": "https://static.tvmaze.com/uploads/images/medium_landscape/416/1042460.jpg",
    "original": "https://static.tvmaze.com/uploads/images/original_untouched/416/1042460.jpg"
  },
  "summary": "<p>God is bored.</p>",
  "_links": {
    "self": {
      "href": "https://api.tvmaze.com/episodes/2326658"
    },
    "show": {
      "href": "https://api.tvmaze.com/shows/1371"
    }
  }
}
```

#### Excepted Output Data Attributes -

- 
- id - int url - string
- 
- name - string season
- 
- int number - int
- 
- type - string airdate -
- 
- date format airtime -
- 
- 12-hour time format
- 
- runtime - float
- 
- average rating - float
- 
- summary - string
- 
- without html tags
- 
- medium image link - string
- 
- Original image link - string

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/q5.py>

#### Question 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/tushar99deep/Assignments/blob/main/Question%206%20EDA.ipynb>

### Question 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 10000kg

**Answer:** <https://github.com/tushar99deep/Assignments/blob/main/question%207.ipynb>

### Question 8 -

Using the data from **Question 5**, write code to 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/tushar99deep/Assignments/blob/main/question%208.ipynb>

### Question 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/tushar99deep/Assignments/blob/main/new\\_assignment/question%209.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/question%209.ipynb)

### Question 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/tushar99deep/Assignments/blob/main/new\\_assignment/question\\_10.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/question_10.ipynb)

## 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

The correlation coefficient is 0.7, it indicates a strong positive linear relationship between SAT scores and college GPA. Here's the mathematical interpretation:

**Positive Relationship:** The positive sign indicates that as SAT scores increase, college GPA tends to increase as well. Higher SAT scores are associated with higher college GPAs.

**Strength of Relationship:** The magnitude of 0.7 suggests a strong relationship. The closer the correlation coefficient is to 1, the stronger the linear relationship between the variables.

**Linear Relationship:** The correlation coefficient measures the strength of the linear relationship. A value of 0.7 indicates that there is a strong linear association between SAT scores and college GPA. It means that as SAT scores increase, the college GPA tends to increase in a consistent manner.

\*\*\*\*\*

**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. To calculate the percentage of individuals with heights between 160 cm and 180 cm, we need to find the z-scores for these heights and then use the standard normal distribution table or a statistical calculator.

The z-score formula is given by:

$$z = (x - \mu) / \sigma$$

where:

- x is the value (height) we want to convert to a z-score
- $\mu$  is the mean of the dataset
- $\sigma$  is the standard deviation of the dataset

For 160 cm:

$$z_1 = (160 - 170) / 10 = -1$$

For 180 cm:

$$z_2 = (180 - 170) / 10 = 1$$

Using the standard normal distribution table or calculator, we can find the percentage of individuals between -1 and 1 z-scores, which represents the percentage of individuals with heights between 160 cm and 180 cm.

b. To calculate the probability that the average height of 100 randomly selected individuals is greater than 175 cm, we need to consider the sampling distribution of the sample mean.

Since the dataset is normally distributed and the sample size is large ( $n = 100$ ), we can use the Central Limit Theorem. The distribution of the sample mean will be approximately normal with the same mean as the population mean (170 cm) and a standard deviation equal to the population standard deviation divided by the square root of the sample size ( $10 \text{ cm} / \sqrt{100}$ ).

Using this information, we can find the z-score for a sample mean of 175 cm and calculate the probability using the standard normal distribution table or calculator.

c. To find the z-score corresponding to a height of 185 cm, we can use the formula mentioned in part (a). The z-score formula is given by:

$$z = (x - \mu) / \sigma$$

where:

- x is the value (height) we want to convert to a z-score
- $\mu$  is the mean of the dataset
- $\sigma$  is the standard deviation of the dataset

For 185 cm:

$$z = (185 - 170) / 10$$

d. To find the approximate height corresponding to a threshold of 5% of the dataset, we can use the inverse of the standard normal distribution (also known as the percent-point function or quantile function). We need to find the z-score corresponding to a cumulative probability of 0.05 and then convert it back to the height using the z-score formula.

e. The coefficient of variation (CV) is a measure of relative variability and is calculated as the ratio of the standard deviation to the mean. The formula for CV is:

$$CV = (\sigma / \mu) * 100$$

where:

- CV is the coefficient of variation
- $\sigma$  is the standard deviation of the dataset
- $\mu$  is the mean of the dataset

f. The skewness of a dataset measures the asymmetry of the distribution. A skewness of approximately zero indicates that the dataset is approximately symmetric. Skewness can be calculated using the following formula:

$$\text{skewness} = (\sum((X - \bar{X})^3) / (n * \sigma^3))$$

where:

- $X$  is the value in the dataset
- $\bar{X}$  is the mean of the dataset
- $\sigma$  is the standard deviation of the dataset
- $n$  is the number of observations in the dataset

Interpreting the skewness result depends on the magnitude and direction of the value. A skewness close to zero suggests a symmetric distribution, while positive or negative skewness indicates a right-skewed or left-skewed distribution, respectively.

\*\*\*\*\*

**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\\_](https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_)

- Measure the dispersion in both and interpret the results.
- Calculate mean and 5% confidence interval and plot it in a graph
- Calculate the Mean absolute deviation and Standard deviation and interpret the results.
- Calculate the correlation coefficient and check the significance of it at 1% level of significance.

**Answer** [https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statsq3.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statsq3.ipynb)

\*\*\*\*\*

**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)?

To calculate the probability that the number on the slip of paper is a perfect square, we need to determine the total number of favorable outcomes (numbers that are perfect squares) and the total number of possible outcomes (all numbers from 1 to 20).

Total number of favorable outcomes:

There are 4 perfect squares between 1 and 20, which are 1, 4, 9, and 16.

Total number of possible outcomes:

There are 20 numbers from 1 to 20.

Probability = Number of favorable outcomes / Number of possible outcomes

Probability =  $4 / 20$

Probability = 0.2 or 20%

Therefore, the probability that the number on the slip of paper is a perfect square is 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?

To find the probability that a randomly selected late taxi belongs to Company A, we can use Bayes' theorem.

Let's define the events:

A: Taxi belongs to Company A

B: Taxi is late

We are given:

$P(A) = 0.8$  (probability that a taxi belongs to Company A)

$P(B|A) = 0.05$  (probability that a taxi from Company A is late)

$P(B|\text{not } A) = 0.10$  (probability that a taxi from Company B is late)

We want to find:

$P(A|B)$  (probability that a late taxi belongs to Company A)

Bayes' theorem states:

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

We can calculate  $P(B)$  using the law of total probability:

$$P(B) = P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A)$$

$$P(\text{not } A) = 1 - P(A) = 1 - 0.8 = 0.2 \text{ (probability that a taxi does not belong to Company A)}$$

Now, let's calculate  $P(B)$ :

$$\begin{aligned} P(B) &= P(B|A) * P(A) + P(B|\text{not } A) * P(\text{not } A) \\ &= 0.05 * 0.8 + 0.10 * 0.2 \\ &= 0.04 + 0.02 \\ &= 0.06 \end{aligned}$$

Finally, we can calculate  $P(A|B)$ :

$$\begin{aligned} P(A|B) &= (P(B|A) * P(A)) / P(B) \\ &= (0.05 * 0.8) / 0.06 \\ &= 0.04 / 0.06 \\ &= 0.6667 \end{aligned}$$

Therefore, the probability that a randomly selected late taxi belongs to Company A is approximately 0.6667 or 66.67%.

\*\*\*\*\*8

**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\\_](https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_)

**Answer**

[https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statsq6.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statsq6.ipynb)

**Q-7.** The equations of two lines of regression, obtained in a correlation analysis between variables X and Y are as follows:

and  $2 + 3 - 8 = 0$   $2 + - 5 = 0$  The variance of Find the

- Variance of Y
- Coefficient of determination of C and Y
- Standard error of estimate of X on Y and of Y on X.

**Answer** [https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statq7.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statq7.ipynb)

**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



| Participant | Before therapy | After therapy | Difference |
|-------------|----------------|---------------|------------|
| 1           | 10             | 7             | -3         |
| 2           | 8              | 6             | -2         |
| 3           | 12             | 10            | -2         |
| 4           | 15             | 12            | -3         |
| 5           | 6              | 5             | -1         |
| 6           | 9              | 8             | -1         |
| 7           | 11             | 9             | -2         |
| 8           | 7              | 6             | -1         |
| 9           | 14             | 12            | -2         |
| 10          | 10             | 8             | -2         |

**Answer**

[https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statsq8.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statsq8.ipynb)

\*\*\*\*\*

**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.

| Name    | Exam 1 | Exam 2 | Final Exam |
|---------|--------|--------|------------|
| Karan   | 85     | 90     | 92         |
| Deepa   | 70     | 80     | 85         |
| Karthik | 90     | 85     | 88         |
| Chandan | 75     | 70     | 75         |
| Jeevan  | 95     | 92     | 96         |

**Answer**

[https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statsq9.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statsq9.ipynb)

\*\*\*\*\*

**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.

- What is the probability that exactly 20 bulbs are defective?
- What is the probability that at least 10 bulbs are defective?
- What is the probability that at max 15 bulbs are defective?
- On average, how many defective bulbs would you expect in a batch of 500?

**Solution**

a. To calculate the probability that exactly 20 bulbs are defective, we can use the binomial probability formula. The probability of success (defective bulb) is 0.05, and the number of trials (light bulbs) is 500. We want to find the probability of getting exactly 20 successes.

$$P(X = 20) = (500 \text{ choose } 20) * (0.05)^{20} * (1 - 0.05)^{(500 - 20)}$$

Using a calculator or statistical software, we can calculate this probability.

b. To calculate the probability that at least 10 bulbs are defective, we can calculate the probability of getting 10, 11, 12, ..., 500 defective bulbs and sum them up.

$$P(X \geq 10) = P(X = 10) + P(X = 11) + \dots + P(X = 500)$$

Again, using a calculator or statistical software, we can calculate this probability.

c. To calculate the probability that at most 15 bulbs are defective, we can calculate the probability of getting 0, 1, 2, ..., 15 defective bulbs and sum them up.

$$P(X \leq 15) = P(X = 0) + P(X = 1) + \dots + P(X = 15)$$

Using a calculator or statistical software, we can calculate this probability.

d. On average, the expected number of defective bulbs can be calculated using the formula for the mean of a binomial distribution.

$$E(X) = n * p$$

Where n is the number of trials (500) and p is the probability of success (0.05).

$$E(X) = 500 * 0.05$$

Therefore, on average, we would expect 25 defective bulbs in a batch of 500.

\*\*\*\*\*

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

classes [https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share\\_](https://drive.google.com/file/d/1mCjtYHiX--mMUjicuaP2gH3k-SnFxt8Y/view?usp=share_)

- Check whether the distribution of all the classes are the same or not.
- Check for the equality of variance/
- Which amount LDA and QDA would perform better on this data for classification and why.
- Check the equality of mean for between all the classes.

**Answer**

[https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/statsq11.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/statsq11.ipynb)

\*\*\*\*\*

**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.

- 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.
- 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

# Machine learning

**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/tushar99deep/Assignments/blob/main/new\\_assignment/mlq1.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq1.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/tushar99deep/Assignments/blob/main/new\\_assignment/mlq2.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq2.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/tushar99deep/Assignments/blob/main/new\\_assignment/mlq3.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq3.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/tushar99deep/Assignments/blob/main/new\\_assignment/mlq4.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq4.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**

[https://github.com/tushar99deep/Assignments/blob/main/new\\_assignment/mlq5.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq5.ipynb)

**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.

*Data set is private in kaggle*

*No answer*

**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/tushar99deep/Assignments/blob/main/new\\_assignment/mlq7.ipynb](https://github.com/tushar99deep/Assignments/blob/main/new_assignment/mlq7.ipynb)