

Python Assignment

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 1 Python: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/01.Question 01.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/01.Question%2001.ipynb)

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 2 Python: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/02.Question_02.ipynb

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 3 Python: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/03.Question_03.ipynb

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",
  "reclong": "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 4 Python: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/04.Question 04.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/04.Question%2004.ipynb)

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 5 Python: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/05.Question_05.ipynb

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 6 Python: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/06.Question_06.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 7 Python: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/07.Question_07.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 8 Python: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra Padwal/blob/main/Interview%20Based%20Assignment/Python/08.Question 08.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/08.Question%2008.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 9 Python: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra Padwal/blob/main/Interview%20Based%20Assignment/Python/09.Question 09.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/09.Question%2009.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.

Example Output -

```
dic = {  
    "nouns": "count of nouns",  
    "pronouns": "count of pronouns",  
    "verbs": "count of verbs",  
    "adjectives": "count of adjectives"  
}
```

Answer 10 Python: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra Padwal/blob/main/Interview%20Based%20Assignment/Python/10.Question_10.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Python/10.Question_10.ipynb)

Statistics Assignment

Question 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 relationship between SAT scores and college GPA. The correlation coefficient ranges from -1 to 1, where 1 represents a perfect positive relationship, 0 represents no relationship, and -1 represents a perfect negative relationship. In this case, the correlation coefficient of 0.7 suggests a strong positive association, indicating that higher SAT scores are generally associated with higher college GPAs.

It's important to note that correlation does not imply causation. While the correlation coefficient indicates a relationship between SAT scores and college GPA, it does not establish a cause-and-effect relationship. Other factors, such as study habits, motivation, and personal circumstances, can also influence college GPA and should be considered when interpreting the results.

Question 2 -

Consider a dataset containing the heights (in centimetres) 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:

- What percentage of individuals in the dataset have heights between 160 cm and 180 cm?
- If we randomly select 100 individuals from the dataset, what is the probability that their average height is greater than 175 cm?
- Assuming the dataset follows a normal distribution, what is the z-score corresponding to a height of 185 cm?
- We know that 5% of the dataset has heights below a certain value. What is the approximate height corresponding to this threshold?
- Calculate the coefficient of variation (CV) for the dataset.

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

Answer:

a. To find the percentage of individuals with heights between 160 cm and 180 cm, we can use the standard normal distribution table or calculator. First, we need to convert the heights to z-scores.

The z-score formula is:

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

Where:

x = height value

μ = mean height

σ = standard deviation

For 160 cm:

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

For 180 cm:

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

Now, using the standard normal distribution table, we can find the area under the curve between z_1 and z_2 . The area represents the percentage of individuals with heights between 160 cm and 180 cm.

Looking up the standard normal distribution table, the area between -1 and 1 is approximately 0.6826. This means that approximately 68.26% of individuals in the dataset have heights between 160 cm and 180 cm.

b. To find 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. The mean of the sample means will be the same as the population mean (170 cm), but the standard deviation of the sample means (also known as the standard error) will be the population standard deviation divided by the square root of the sample size.

$$\text{Standard error (SE)} = \sigma / \sqrt{n}$$

Where:

σ = standard deviation of the population (10 cm)

n = sample size (100)

$$SE = 10 / \sqrt{100} = 1$$

Now, we can convert the value 175 cm to a z-score using this formula:

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

$$z = (175 - 170) / 1 = 5$$

Next, we can find the probability of the average height being greater than 175 cm by calculating the area under the standard normal distribution curve to the right of $z = 5$. This can be done using the standard normal distribution table or calculator.

However, since $z = 5$ is quite far from the mean, the probability will be extremely small, close to 0.

c. To find the z-score corresponding to a height of 185 cm, we use the z-score formula:

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

Substituting the values:

$$z = (185 - 170) / 10 = 1.5$$

Therefore, the z-score corresponding to a height of 185 cm is 1.5.

d. To find the approximate height corresponding to the threshold below which 5% of the dataset falls, we need to find the z-score corresponding to the 5th percentile. The 5th percentile corresponds to a z-score of -1.645. We can use the standard normal distribution table or calculator to determine this.

Next, we can convert the z-score back to the height value using the formula:

$$x = \mu + (z * \sigma)$$

$$x = 170 + (-1.645 * 10) \approx 153.55 \text{ cm}$$

Therefore, the approximate height corresponding to the threshold where 5% of the dataset falls below is approximately 153.55 cm.

e. The coefficient of variation (CV) is a measure of relative variability and is calculated by dividing the standard deviation by the mean and multiplying by 100 to express it as a percentage.

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

Substituting the values:

$$CV = (10 / 170) * 100 \approx 5.88$$

Therefore, the coefficient of variation for the dataset is approximately 5.88%.

f. The skewness of a dataset measures the asymmetry of its distribution. Given that the dataset's skewness is approximately zero, it indicates that the dataset is symmetric and has a bell-shaped normal distribution. In other words, the dataset has roughly equal numbers of

individuals above and below the mean, resulting in a balanced distribution without any significant skew to either tail.

Stats: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/02.Question_02.ipynb

Question 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 3 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/03.Question_03.ipynb

Question 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 4 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/04.Question_04.ipynb

Question 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 5 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/05.Question_05.ipynb

Question 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 6 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/06.Question_06.ipynb

Question 7 -

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

$2X + 3 - 8 = 0$ $2Y + X - 5 = 0$ The variance of X = 4 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 :

To calculate the variance of Y, we can rearrange the equation of the regression line for

$$Y: 2Y + X - 5 = 0$$

Solving for Y:

$$2Y = -X + 5$$

$$Y = (-1/2)X + 5/2$$

Since the equation for Y is in the form $Y = aX + b$, where a is the coefficient of X and b is the intercept, we can determine the variance of Y.

a. Variance of Y:

The variance of Y can be calculated using the formula: $\text{Var}(Y) = a^2 * \text{Var}(X)$

Given that the variance of X is 4, we substitute the values:

$$\text{Var}(Y) = (-1/2)^2 * 4$$

$$\text{Var}(Y) = 1 * 4$$

$$\text{Var}(Y) = 4$$

Therefore, the variance of Y is 4.

b. Coefficient of determination of X and Y:

The coefficient of determination, denoted as R^2 , measures the proportion of the variance in the dependent variable (Y) that can be explained by the independent variable (X).

$$R^2 = (\text{Var}(Y) - \text{Var}(\text{residuals})) / \text{Var}(Y)$$

The variance of the residuals can be calculated by subtracting the explained variance from the total variance:

$$\text{Var}(\text{residuals}) = \text{Var}(Y) - \text{Var}(\text{explained})$$

In this case, the variance of the residuals is the same as the variance of Y since the regression line explains all the variance.

$$\text{Var}(\text{residuals}) = \text{Var}(Y) = 4$$

Substituting the values:

$$R^2 = (4 - 4) / 4$$

$$R^2 = 0 / 4$$

$$R^2 = 0$$

Therefore, the coefficient of determination (R^2) between X and Y is 0.

c. The standard error of estimate represents the average distance between the observed values and the predicted values obtained from the regression line.

For the standard error of estimate of X on Y (S_{xy}), we use the formula:

$$S_{xy} = \sqrt{\text{Var}(\text{residuals})}$$

$$S_{xy} = \sqrt{4}$$

$$S_{xy} = 2$$

For the standard error of estimate of Y on X (S_{yx}), we use the same formula:

$$S_{yx} = \sqrt{\text{Var}(\text{residuals})}$$

$$S_{yx} = \sqrt{4}$$

$$S_{yx} = 2$$

Therefore, the standard error of estimate for both X on Y (S_{xy}) and Y on X (S_{yx}) is 2.

Question 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 8 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Statistics/08.Question_08.ipynb

Question 9 -

Given the score of students in multiple exams

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 9 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Statistics/09.Question_09.ipynb

Question 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?

Answer 10 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/10.Question_10.ipynb

Question 11 -

Given the data of a feature contributing to different classes

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 11 Statistics: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/11.Question_11.ipynb

Question 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 12 Statistics: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/12.Question 12.ipynb](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Stats/12.Question%2012.ipynb)

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 1 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/01.Solution_Q1/01.Solution_Q1.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 2 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/02.Solution_Q2/02.Solution_Q2.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 3 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/03.Solution_Q3/03.Solution_Q3.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 4 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/04.Solution_Q4/04.Solution_Q4.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 5 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/05.Solution_Q5/05.Solution_Q5.ipynb

Q-6. Imagine you have a dataset where you have predicted loan Eligibility using any4 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 6 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/06.Solution_Q6/06.Solution_Q6.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 7 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/07.Solution_Q7/07.Solution_Q7.ipynb

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 8 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/08.Solution_Q8/08.Solution_Q8.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 9 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/09.Solution_Q9/09.Solution_Q9.py

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 10 ML: - https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Intermediate%20Question%20Solution/10.Solution_Q10/10.Question%2010.py

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 1 ML: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Advance Question Solution/01.Solution_Q1/Question %201.py](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Advance%20Question%20Solution/01.Solution_Q1/Question%201.py)

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 2 ML: - [https://github.com/kamleshp95/Placement-Assignment-Ramchandra Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Advance Question Solution/02.Solution_Q2/02.Question%202.py](https://github.com/kamleshp95/Placement-Assignment-Ramchandra-Padwal/blob/main/Interview%20Based%20Assignment/Machine%20Learning/Advance%20Question%20Solution/02.Solution_Q2/02.Question%202.py)

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 3 ML: -