1) Describe a data you worked on recently?

Ans:

My recent project is "Effective Data Visualization". The goal for the project is to design a story telling dashboard to highlight the present market using a food supply chain in tableau. I built 8 visualizations which highlight the top sales by month-to-month end and year-to-year end. Using various kinds, the beverages like apple pie, lemon pie, donuts, etc... and successfully submitted my project to my team for a review. In the first review, the team members said that the work is technically good, but pointed out the explanatory section to improve better. After their review, as per the aim of project. I found my mistake and started adding the explanation section of each chart and I added some interesting quick filters to view better. Now, I got a positive feedback from my team and, they pointed some changes to fix such as adding a data sources to the file and their interlinking connections using data blending and joining. I feel great by seeing the project every time and got to know the importance of the feedback from the team or others.

I believe that this ability is something could bring to my work at chewy. I would collect feedback from the customers, classify, analyze, and create solutions to improve their satisfaction with pet food services at chewy.

2) You are given a ten-piece box of chocolate truffles. You know based on the label that six of the pieces have an orange cream filling and four of the pieces have a coconut filling. If you were to eat four pieces in a row, what is the probability that the first two pieces you eat have an orange cream filling and the last two have a coconut filling?

Ans:

Here, we have ten pieces of chocolate truffle in total and 6 pieces of orange filling, 4 pieces of

Coconut filling.

Now. Taking a first piece as an orange filling means picking up 6 out of 10 it is (6/10),

We left with 9 pieces, contains 5 are orange filling, it means taking orange filling from 9 is

5 out of 9 it is (5/9)

In the last but one case we are having 8 pieces in total, in that 4 are coconut filling.

Picking up coconut means 4 out of 8 it is (4/8).

The last case taking a coconut from remaining 7 pieces gives us 3 out of 7 it is (3/7).

To find the probability of eating 1st two pieces as orange and next two pieces as coconut, we

Must multiply the four cases explained in the above.

Probability = (6/10)*(5/9)*(4/8)*(3/7) = 360/5040= (1/14)= 0.071428 (Ans)

##Follow-up question: If you were given an identical box of chocolates and again eat four pieces in a row, what is the probability that exactly two contain coconut filling?

Ans:

We are taking a below combinations to be simple.

Combination 1: Cococunt, Cococunt, Cococunt, Cococunt

Combination 2: Cococunt, Cococunt, Cococunt, Orange

Combination 3: Cococunt, Cococunt, Orange, Orange

:

.

Combination 14: Orange, Orange, Orange, Orange

There are 14 combinations here and, now, 2 Coconut pieces from 4 Coconut pieces. Therefore, it is 2C4 combinations.

Here 2C4 = 6 combinations.

The total probability to eat exactly two coconut chocolates is=

$$\frac{no.of\ combinations\ used\ to\ take\ exactly\ two\ coconut\ chocolates}{Total\ no.of\ combinations\ used\ to\ take\ all\ chocolates} = 2C4/14 = 0.428$$

3) Given the table users:

Table "users"

```
+-------------+-----------+
| Column      | Type      |
+-------------+-----------+
| Id          | integer   |
| Username    | character |
| Email       | character |
| City        | character |
| State       | character |
| Zip         | integer   |
| active      | boolean   |
+-------------+-----------+
```

construct a query to find the top 5 states with the highest number of active users. Include the number for each state in the query result. Example result:

```
+------------+------------------+
| state      | num_active_users |
+------------+------------------+
| New Mexico | 502              |
| Alabama    | 495              |
| California | 300              |
| Maine      | 201              |
| Texas      | 189              |
+------------+------------------+
```

Ans:

select state, sum(active) as total_active_users from users group by state

order by total_active_users DESC top 5;

4) Define a function first_unique that takes a string as input and returns the first non-repeated (unique) character in the input string. If there are no unique characters return None. Note: Your code should be in Python.

Ans:

```
def first_unique(string):
  l = len(string)          # intializing l as a length of string
        # checking all the characters in 'l'
        for p in range l:
        # # checking unique characters before and after the character.
          if string[p] in string[0:p] and string[p] in string[p+1:]:
                  return None   # retruns None if there is no unique character
  print string[p] # prints the similar cahracters if they found
  return string[p] # Returns the similar characters if they found in the string.
```

5) What are underfitting and overfitting in the context of Machine Learning? How might you balance them?

Ans:

We come across over fitting when the algorithm is not tuned well, results the high performance of training data. While generalizing the new data it fails and results the low performance to predict.

Coming to the under fitting, the algorithm fails to generalize the train and new data. The performance results will be lower.

To overcome these issues, we must select the limited number of features, using the cross validation by 3:1 or 4:1 ratio and do not over tweak the parameters. By doing this can help us to avoid noise and irrelevant.

The other method to overcome over & under fitting is by using L1 and L2 regularizations, in which L1 (Lasso Regression) helps from the under fitting and L2 (Ridge Regression) helps from the over fitting issues.

Finally, best way is to use cross validation to divide the training data by fixing the value of "**λ"** and train it. Repeat these for all the remaining subsets by changing different values of "**λ",** then select that minimizes the loss function.

6) If you were to start your data analyst position today, what would be your goals a year from now?

Ans:

I am applying for a Tableau Developer position in chewy. So, I able to learn much knowledge about pet food supply reports and analysis. I would feel excited about being a part of a crew that helps the pet food supplies. I would use this opportunity to work with a team and know the nuts and bolts of every single database scheme in chewy, especially in Florida in the first quarter of a year. Then after, I start collaborating the planning's and designs for the business growth by deploying the past data using Tableau. In a year, I develop an automated data visualization for month-to-month and year-to-year deviations which helps the company to save the time to predict the sales of the products by monthly & yearly and makes the company to top 10 levels in the competitive world.