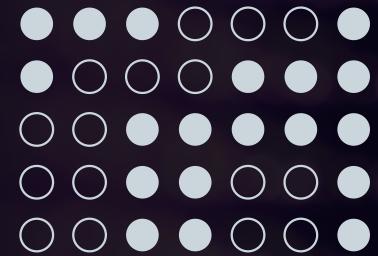




# AI/ML MONTH PROBLEM STATEMENTS



## Loan Default Prediction

The bank claims possession over the collateral against which the loan has been taken if the defaulters do not pay the EMI on time. The grace period given to the people depends upon their annual income and bank balance. If a person has an annual income greater than 100k and a bank balance greater than 5k, the bank will give a grace period of 15 days, else the grace period increases, if and only if the person is unemployed, else the bank charges a fine equivalent to 2.5% of the bank balance per day.

- What is the number of people who will be given a grace period of exactly 15 days?
- Find the maximum fine that can be levied by the bank in the period of 15 days.
- Assuming that all the people falling into the category of defaulters have not paid their EMI in the grace period, what is the total amount being generated by the bank as a part of the fine.
- Use a suitable plot to show the difference in bank balance of the people who are not eligible for a grace period, and those who are.
- Find the difference in the median of the bank balance of those people who are eligible for grace period and the original value.
- Use a suitable graph to show the changes in the annual income of those people who are eligible for grace period and those who are not.

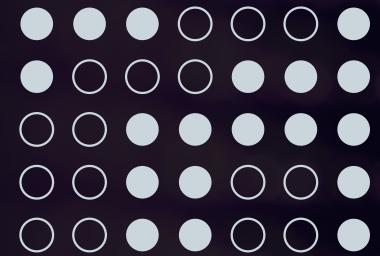
## Stock Price Analysis

Uniqlo has had a growing market since the year 2016. The following dataset shows the variations and changes of their stocks during this period. Perform EDA for the following questions and give a brief explanation about what you understood.

- Plot data for all periods to the low value of the stock and study this graph in relation to change in the open value.
- Analyze the price change within a day using standard deviation.
- Use a suitable plot to show the change in median of open, high, low and close in each quartile as per the volume.
- Given that the difference between high and low is below the average and that of open and close is below the average, find the mean volume and compare it to the original value.
- Use a suitable plot to study the changes in stock trading with respect to time, also plot the changes of difference between open and close to study the fluctuations and relationship between the two.
- Find the percentage change in median for the stock trading if the difference between high and low is below the average and that of open and close is above the average.



# AI/ML MONTH PROBLEM STATEMENTS



## Honey Production

A global concern was raised over the rapid decline in the honeybee population, an integral component to American honey agriculture.

This dataset provides insight into honey production supply and demand in America. Analyze the given dataset for the following questions and list your observations

- Explore quantitative variables and qualitative variables in the dataset.
- Find the relationship between numerical variables using pair plots and correlation plots. Explain what you infer from these plots.
- Look at the overall trend of honey production in the US over the years.
- Are there any patterns that can be observed between total honey production and value of production every year? How has the value of production, which in some sense could be tied to demand, changed every year?
- Observe the Variation in the number of colonies over the years.
- Analyze the Variation of yield per colony over the years and the production trend at state level; and brief out what you observed.
- Analyze what effect the declining production trend has had on the value of production.

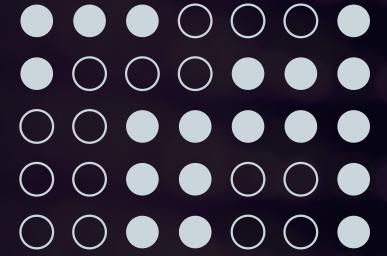
## Superstore

The given dataset is used to analyze and identify trends and patterns in the current retail sales and identify which sector of the market is under loss and which sector is making huge profits. Every sector offers discounts on sales, but do they collect profits as needed on the discounts they offer?

- Box plot for all numeric variable and check for outliers
- Plot pairplot for Region Wise Data Distribution of Parameters: sales, postal code, discount, profit and quantity
- Is there any Impact of Sales through Ship Mode
- Find Segment Wise Sales and its Impact. Plot bar plot and find which segment has highest sales and profit.
- Create bar plot for Sub-Category Wise Order Counts. Which Sub-Category Wise has maximum order and minimum order?
- Analyze the City-Wise Performance of Order Counts.
- Analyze State-Wise Performance of Order Counts.
- Analyze category wise sales performance and sub category wise profit performance. Which shipment class boosts the sales of which category?



# AI/ML MONTH PROBLEM STATEMENTS



## COVID-19 Clinical Trials

The first step towards solving a machine learning problem is to identify any discrepancies in the input data, because if the data is not correct then it will not give the desired output. Do necessary steps to remove any unwanted data, replacing null values etc.

- There are different methods to replace null values in a data. Mention all of them and mention which one you chose and why?
- Explore different columns in your dataset and write your observations about the results. Use all the different plots you've been taught like pie chart, bar graph, histogram etc. and give reasons about why you think this is the best plot for a particular study.
- What conditions are the above studies treating? Use visualization to show this and write your observations.
- Use a map (world map) to show where these studies are taking place. You can refer to the link below:
  - <https://www.geeksforgeeks.org/plotting-world-map-using-pygal-in-python/>
- Analyze what age bracket and gender these studies are considering and use different graphs to show these. Write your observations.