# Statistics Project Statement:

## Task 1: Answer the following Questions

1. Question: Please refer the table below to answer below questions:

|  |  |  |  |
| --- | --- | --- | --- |
| Planned to purchase Product A | Actually, placed and order for Product A - Yes | Actually, placed and order for Product A - No | Total |
| Yes | 400 | 100 | 500 |
| No | 200 | 1300 | 1500 |
| Total | 600 | 1400 | 200 |

* 1. Refer to the above table and find the joint probability of the people who planned to purchase and actually placed an order.
  2. Refer to the above table and find the joint probability of the people who planned to purchase and actually placed an order, given that people planned to purchase.

1. Question: An electrical manufacturing company conducts quality checks at specified periods on the products it manufactures. Historically, the failure rate for the manufactured item is 5%. Suppose a random sample of 10 manufactured items is selected. Answer the following questions.
   1. Probability that none of the items are defective.
   2. Probability that exactly one of the items is defective.
   3. Probability that two or fewer of the items are defective.
   4. Probability that three or more of the items are defective.
2. Question: A car salesman sells on an average 3 cars per week.
   1. Probability that in a given week he will sell some cars.
   2. Probability that in a given week he will sell 2 or more but less than 5 cars.
   3. Plot the poison distribution function for cumulative probability of cars sold per-week vs number of cars sold per-week.
3. Question: Accuracy in understanding orders for a speech based bot at a restaurant is important for the Company X which has designed, marketed and launched the product for a contactless delivery due to the COVID-19 pandemic. Recognition accuracy that measures the percentage of orders that are taken correctly is 86.8%. Suppose that you place order with the bot and two friends of yours independently place orders with the same bot. Answer the following questions.
   1. What is the probability that all three orders will be recognised correctly?
   2. What is the probability that none of the three orders will be recognised correctly?
   3. What is the probability that at least two of the three orders will be recognised correctly?
4. Question: A group of 300 professionals sat for a competitive exam. The results show the information of marks obtained by them have a mean of 60 and a standard deviation of 12. The pattern of marks follows a normal distribution. Answer the following questions.
   1. What is the percentage of students who score more than 80.
   2. What is the percentage of students who score less than 50.
   3. What should be the distinction mark if the highest 10% of students are to be awarded distinction?
5. Question: Explain 1 real life industry scenario [other than the ones mentioned above] where you can use the concepts learnt in this module of Applied statistics to get a data driven business solution.

## Task 2: Complete the Below Project Objective

**DOMAIN*:*** Sports

**CONTEXT*:*** Company X manages the men's top professional basketball division of the American league system. The dataset contains information on all the teams that have participated in all the past tournaments. It has data about how many baskets each team scored, conceded, how many times they came within the first 2 positions, how many tournaments they have qualified, their best position in the past, etc.

**DATA DESCRIPTION*:*** Basketball.csv - The data set contains information on all the teams so far participated in all the past tournaments.

ATTRIBUTE INFORMATION:

1. Team: Team’s name
2. Tournament: Number of played tournaments.
3. Score: Team’s score so far.
4. PlayedGames: Games played by the team so far.
5. WonGames: Games won by the team so far.
6. DrawnGames: Games drawn by the team so far.
7. LostGames: Games lost by the team so far.
8. BasketScored: Basket scored by the team so far.
9. BasketGiven: Basket scored against the team so far.
10. TournamentChampion: How many times the team was a champion of the tournaments so far.
11. Runner-up: How many times the team was a runners-up of the tournaments so far.
12. TeamLaunch: Year the team was launched on professional basketball.
13. HighestPositionHeld: Highest position held by the team amongst all the tournaments played.

**PROJECT OBJECTIVE*:*** Company’s management wants to invest on proposal on managing some of the best teams in the league. The analytics department has been assigned with a task of creating a report on the performance shown by the teams. Some of the older teams are already in contract with competitors. Hence Company X wants to understand which teams they can approach which will be a deal win for them.

## Task 3: Complete the Below Project Objective

**DOMAIN*:*** Startup ecosystem

**CONTEXT*:*** Company X is a EU online publisher focusing on the startups industry. The company specifically reports on the business related to technology news, analysis of emerging trends and profiling of new tech businesses and products. Their event i.e. Startup Battlefield is the world’s pre-eminent startup competition. Startup Battlefield features 15-30 top early stage startups pitching top judges in front of a vast live audience, present in person and online.

**DATA DESCRIPTION*:*** CompanyX\_EU.csv - Each row in the dataset is a Start-up company and the columns describe the company.

ATTRIBUTE INFORMATION:

1. Startup: Name of the company
2. Product: Actual product
3. Funding: Funds raised by the company in USD
4. Event: The event the company participated in
5. Result: Described by Contestant, Finalist, Audience choice, Winner or Runner up
6. OperatingState: Current status of the company, Operating ,Closed, Acquired or IPO

\*Dataset has been downloaded from the internet. All the credit for the dataset goes to the original creator of the data.

**PROJECT OBJECTIVE*:*** Analyse the data of the various companies from the given dataset and perform the tasks that are specified in the below steps. Draw insights from the various attributes that are present in the dataset, plot distributions, state hypotheses and draw conclusions from the dataset.