Task

Nowadays, marketing spending in the banking industry is massive, meaning that it is essential for banks to optimize marketing strategies and improve effectiveness. Understanding customers’ needs lead to more effective marketing plans, smarter product designs, and greater customer satisfaction.

You are working as a data analyst for the Europe Department of ING bank. You have been given the sample of the telemarketing dataset and asked to apply the summary and inferential statistics so that you can make inferences about the larger population based on this sample. You will analyze the telemarketing sample and submit the findings to the marketing team. It will enable the marketing team to develop a more granular understanding of its customer base, predict customers' response to its telemarketing campaign and establish a target customer profile for future marketing plans.

This will not only allow the bank to secure deposits more effectively but also increase customer satisfaction by reducing undesirable advertisements for certain customers.

Note: You will have to use the entire data for solving Q1 - Q11. For the presentation(creative question), you will have to pick ONLY one sample out of 5 and do your statistical analysis.

Question 1

1/1 point (graded)

Story for Q1: We have the sample data from the population of the bank telemarketing dataset. So, the first thing to check is that the sample should be representative of the population. We can do the basic checks through descriptive statistics.

1. Calculate the proportion of customers who have responded to the offer before in the given sample and give the answer for the below question.

Note: The proportion of customers who have responded to the offer before in the main population is 5%

Is the sample representative of the population?

True

False

correct

SELECT Count(last\_select) / (SELECT Count(\*)

FROM bt\_callinfo) \* 100

FROM bt\_callinfo

WHERE last\_select = 'Yes'

The proportion of customers in the sample which has responded to the offer before is 3.5% which comes in the range of the population proportion - threshold percentage (5-2.5). So, we can conclude that the sample is representative of the population.

Question 2

1/1 point (graded)

Story Q2: Average duration of calls is one of the important descriptive statistics. It gives the approximate idea about the change in duration of call when the customer is interested in the product.

2. What is the difference between the mean duration of calls when consumers accepted the offer the last time and when they did not accept the offer the last time?

678.37

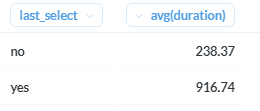
SELECT last\_select,

Avg(duration) as Average\_duration

FROM bt\_callinfo

GROUP BY last\_select;

Above Query will give you the below result

  
If we calculate the difference of average duration then it will be going to be 678.37

Question 3

1/1 point (graded)

Q3 Story: A frequency distribution is an overview of all distinct values in some variable and the number of times they occur. That is, a frequency distribution tells how frequencies are distributed over values. Frequency distributions are mostly used for summarizing categorical variables. So, it is always a great idea to use frequency distribution on the customer categorical variable.

3. Which job type accepted more offers last time?

Blue-collar

SELECT job,

Count(\*) AS count

FROM bt\_callinfo a,

bt\_customerinfo b

WHERE a.id = b.id and a.last\_select='Yes'

group by job

Question 3

1/1 point (graded)

Q3 Story: A frequency distribution is an overview of all distinct values in some variable and the number of times they occur. That is, a frequency distribution tells how frequencies are distributed over values. Frequency distributions are mostly used for summarizing categorical variables. So, it is always a great idea to use frequency distribution on the customer categorical variable.

3. Which job type accepted more offers last time?

Blue-collar

SELECT job,

Count(\*) AS count

FROM bt\_callinfo a,

bt\_customerinfo b

WHERE a.id = b.id and a.last\_select='Yes'

group by job

Question 4

2/2 points (graded)

Story Q4: Understanding the customer is one of the important characteristics of targeting marketing. Once we calculate the average, we should always calculate the standard deviation of the same attribute. It provides the range for targeting the customer.

4. Calculate the Average and Standard Deviation of blue-collar customer age and then calculate the maximum and minimum of range using Average + SD and Average - SD respectively?

Max Range

47.68

30.28

SELECT Stddev(age) AS standard\_deviation,

Avg(age) AS Average\_Age

FROM bt\_customerinfo

WHERE job = 'blue-collar'

Maximum Range = Average + Standard Deviation = 38.98 + 8.7 = 47.68

Minimum Range = Average - Standard Deviation = 38.98 - 8.7 = 30.28

Question 5

1/1 point (graded)

Q5 Story: The demand of the product may change if the person has been defaulted before. So, let's take a look at the distribution of customers with respect to the defaulter status of Customer.

5. How many blue-collar customers have signed up for the product last time and defaulted a loan?

3

SELECT b.loan\_default,

Count(\*) AS count

FROM bt\_callinfo a,

bt\_customerinfo b

WHERE a.id = b.id

AND last\_select = 'Yes'

AND b.job = 'blue-collar'

GROUP BY b.loan\_default;

Question 6

1/1 point (graded)

Inferential Question

Q6 Story:The conversion rate is the number of conversions divided by the total number of visitors. Generally, A good conversion rate is above 10%. We should check the conversion rate for the telemarketing data so the marketing team can be aware of the success rate of the previous campaign.

6. What is the probability that a random customer will end up signing up for a product?

0.035

SELECT ( (SELECT Count(\*)

FROM bt\_callinfo

WHERE last\_select = 'Yes') / Count(\*) ) AS probability

FROM bt\_callinfo

There is a 3.5% chance that a random customer will end up signing for a product. This is low and needs to be increased.

Question 7

1/1 point (graded)

Q7 Story: Marketing team must quantify the campaign's commitment, and to do so, they must have reasonable estimates of call time.

7. What is the probability that a random call will last for less than 200 seconds?

0.53

correct 0.53

\(\)

SELECT ( (SELECT Count(\*)

FROM bt\_callinfo

WHERE duration < 200) / Count(\*) ) AS probability

FROM bt\_callinfo

There is 53% chance that any random call will last less than 200 seconds.

Questions 8 and 9

1/1 point (graded)

Q8 and 9 Story: Inferencing sample data by calculating the probability on various attributes , the marketing team can be more sure about the customer and can build the next targeted campaign.

8. What is the probability that a random customer will have both a housing loan and a personal loan?

0.126

correct 0.12

\(\)

SELECT ( (SELECT Count(\*)

FROM bt\_customerinfo

WHERE housing\_loan = 'Yes') / Count(\*) ) AS Housing\_probability,

( (SELECT Count(\*)

FROM bt\_customerinfo

WHERE personal\_loan = 'Yes') / Count(\*) ) AS personal\_probability

,

( ( (SELECT Count(\*)

FROM bt\_customerinfo

WHERE Housing\_loan = 'Yes') / Count(\*) ) \* (

(SELECT Count(\*)

FROM bt\_customerinfo

WHERE personal\_loan = 'Yes') / Count(\*) ) ) AS

both\_loan\_probability

FROM bt\_customerinfo

Probability of Housing loan - 0.84

Probability of personal loan - 0.15

Combined probability = 0.84 \* 0.15 = 0.12

Question 9

1/1 point (graded)

9. What is the probability that a random customer will have a personal loan given that he already has a housing loan?

0.0205948705770601

SELECT (SELECT Count(\*)

FROM bt\_customerinfo

WHERE housing\_loan = 'Yes') AS

Housing\_loan,

(SELECT Count(\*)

FROM bt\_customerinfo

WHERE personal\_loan = 'Yes'

AND housing\_loan = 'Yes') AS

personal\_loan\_with\_housing\_loan,

(SELECT Count(\*)

FROM bt\_customerinfo

WHERE personal\_loan = 'Yes'

AND housing\_loan = 'Yes') / (SELECT Count(\*)

FROM bt\_customerinfo

WHERE housing\_loan = 'Yes') AS

conditional\_probability

FROM bt\_customerinfo

LIMIT 1

Count of customer having housing loan - 8422

Count of Customer having Personal Loan - 1317

Conditional Probability = 1317/8422 = 0.16

Question 10 and 11

1/1 point (graded)

Q10 and Q11 Story: When we are analyzing a telemarketing dataset, we may get different statistics, so it is always a good idea to check the likelihood of that shift being due to randomness.

10. Consider the marketing team has launched a new campaign for the product and 5 people have signed up out of 50 random people. How likely is this outcome due to random chance?

0.022395001

The success rate in the sample is 3.5% and in this question, the success rate is 10% so we have to check whether there is a change in the success rate in the new campaign or we are seeing this number only because of the random chance variation. So, we calculate that likelihood by using binomial distribution because the target variable is binary(Yes/No).

BINOM.DIST(5,50,0.035,FALSE) = 0.022

So, There is only a 2.2% chance that 5 people are signing up out of 50 random people which is very low. It means that the success rate has been actually increased in the new campaign.

Hint

Show answer

SubmitSome problems have options such as save, reset, hints, or show answer. These options follow the Submit button.

Answers are displayed within the problemReview

Question 11

1/1 point (graded)

11. What is the probability that if we choose 15 customers, at least 10 of will be younger than 40?

0.07617090451

probability of customers being < 40 in the dataset = 5419/10000 = 0.54

P of at least 10 customers is p(10) + p(11) + p(12) + p(13) + p(14) + p(15)

= 1-p(<=9)

= 1-binom.dist(9,15,0.54,true) = 0.23

Creative Question

Bookmark this page

You have to prepare the presentation for all below questions

Table bt\_callinfo has been randomly divided into 5 different samples, called 5 small samples. The number in column sample\_no represents the 5 different samples. You have to select any one sample out of 5 samples present in the bt\_callinfo table and answer the below questions. Also, please reach out to the support team in case of any confusion or questions.

1) Solve at least 2 questions from this project using any small sample(Choose any sample i.e 1 to 5 with sample\_no column) and explain in the presentation if you find any difference in the statistics between the 10000 sample (bt\_callinfo - original dataset) and small sample (Approx 2000 sample) which is created from bt\_callinfo?

2) What is the probability of 12 customers rejecting the offer and then 1 customer accepting the offer?

3) Assume the marketing team launches a new campaign and 60 new customers accepted the offer in May. Is the rise in the number of accepted offers due to the latest campaign or due to random chance? Calculate the likelihood that the new campaign will result in more than 60 customers accepting the offer during May.

4) What are the chances/probability that if you pick any random customer, then the customer's age is less than or equal to 28 years old?

-- 2) What is the probability of 12 customers rejecting the offer and then 1 customer accepting the offer?

ANSWERS

1) Solve at least 2 questions from this project using any small sample(Choose any sample i.e 1 to 5 with sample\_no column) and explain in the presentation if you find any difference in the statistics between the 10000 sample (bt\_callinfo - original dataset) and small sample (Approx 2000 sample) which is created from bt\_callinfo?

View sheet ⇒ <https://docs.google.com/spreadsheets/d/1qHgR6Dci4TCQ7qIlRNBJ0o_aDSBdINeD_yQk5ShPuCc/edit?usp=sharing>

2) What is the probability of 12 customers rejecting the offer and then 1 customer accepting the offer?

SELECT \* FROM bt\_callinfo

-- NUMBER OF REJECTED / ACCEPTED

SELECT COUNT(Last\_Select) AS NO\_REJECTED

FROM bt\_callinfo

WHERE Last\_Select='no'

SELECT COUNT(Last\_Select) AS NO\_ACCEPTING

FROM bt\_callinfo

WHERE Last\_Select='yes'

-- PROBABILITY

SELECT 12\*100/COUNT(Last\_Select) AS PROB\_REJECTED

FROM bt\_callinfo

WHERE Last\_Select='no'

SELECT 1\*100/COUNT(Last\_Select) AS PROB\_ACCEPTED

FROM bt\_callinfo

WHERE Last\_Select='yes'

-- probability of 12 customers rejecting the offer and then 1 customer accepting the offer

SELECT Last\_Select, 12\*100/COUNT(Last\_Select) AS PROB\_REJECTED, 1\*100/COUNT(Last\_Select) AS PROB\_ACCEPTED

FROM bt\_callinfo

GROUP BY Last\_Select

| Last\_Select | PROB\_REJECTED | PROB\_ACCEPTED |  |  |
| --- | --- | --- | --- | --- |
| no | 0.1244 | 0.0104 |  |  |
| yes | 3.4286 | 0.2857 |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
| -- probability of 12 customers rejecting the offer and then 1 customer accepting the offer | | | | |
|  |  |  |  |  |
| 0.03554108 |  |  |  |  |

3) Assume the marketing team launches a new campaign and 60 new customers accepted the offer in May.

Is the rise in the number of accepted offers due to the latest campaign or due to random chance?

Calculate the likelihood that the new campaign will result in more than 60 customers accepting the offer during May.

SELECT \* FROM bt\_callinfo

SELECT 60\*100/COUNT(Last\_Select) AS 60\_NEW\_CUSTOMER

FROM bt\_callinfo

WHERE month='May' AND Last\_Select='YES'

4) What are the chances/probability that if you pick any random customer, then the customer's age is less than or equal to 28 years old?

SELECT \* FROM bt\_callinfo

SELECT 1\*100/COUNT(day) AS PROB\_CUST\_28 FROM bt\_callinfo

WHERE day<=28

PROBALITY OF CUSTOMER’S AGE LESS THEN OR EQUAL TO 28

0.011.