Hypothesis Testing Project

Project Deliverables

You will be required to provide the following deliverables.

- A python notebook with your solution.

Instructions

Background Information

You work as a data scientist for the telecom operator Megatelco. The company offers its clients two prepaid plans, Surf and Ultimate. The commercial department wants to know which of the plans brings in more revenue in order to adjust the advertising budget.

You are going to carry out a preliminary analysis of the plans based on a relatively small client selection. You'll have the data on 500 Megatelco clients: who the clients are, where they're from, which plan they use, and the number of calls they made and text messages they sent in 2018. Your job is to analyze clients' behavior and determine which prepaid plan brings in more revenue.

Description of the plans

Megatelco rounds seconds up to minutes, and megabytes to gigabytes. For calls, each individual call is rounded up: even if the call lasted just one second, it will be counted as one minute. For web traffic, individual web sessions are not rounded up. Instead, the total for the month is rounded up. If someone uses 1025 megabytes this month, they will be charged for 2 gigabytes.

Surf

1. Monthly charge: $20
2. 500 monthly minutes, 50 texts, and 15 GB of data
3. After exceeding the package limits:
   - 1 minute: 3 cents
   - 1 text message: 3 cents
   - 1 GB of data: $10

Ultimate

1. Monthly charge: $70
2. 3000 monthly minutes, 1000 text messages, and 30 GB of data
3. After exceeding the package limits:
   ○ 1 minute: 1 cent
   ○ 1 text message: 1 cent
   ○ 1 GB of data: $7

Instructions on completing the project

Step 1. Open the data file and study the general information

● Download Link: https://bit.ly/3lrTn9z

Step 2. Prepare the data

● Convert the data to the necessary types
● Find and eliminate errors in the data
  ■ Explain what errors you found and how you removed them.

For each user, find:

● The number of calls made and minutes used per month
● The number of text messages sent per month
● The volume of data per month
● The monthly revenue from each user (subtract the free package limit from the total number of calls, text messages, and data; multiply the result by the calling plan value; add the monthly charge depending on the calling plan)

Step 3. Analyze the data

● Describe the customers' behavior. Find the minutes, texts, and volume of data the users of each plan require per month. Calculate the mean, variance, and standard deviation. Plot histograms. Describe the distributions.

Step 4. Test the hypotheses

● The average revenue from users of Ultimate and Surf calling plans differs.
● The average revenue from users in NY-NJ area is different from that of the users from other regions.

You decide what alpha value to use.

Explain:

● How you formulated the null and alternative hypotheses.

- What criterion you used to test the hypotheses and why.

Step 5. Write an overall conclusion

- Format: Complete the task in Google Colab. Put the programming code in code cells and text explanations in markdown cells, then apply formatting and headings.

Description of the data

The users table (data on users):

- *user_id* — unique user identifier
- *first_name* — user's name
- *last_name* — user's last name
- *age* — user's age (years)
- *reg_date* — subscription date (dd, mm, yy)
- *churn_date* — the date the user stopped using the service (if the value is missing, the calling plan was being used when this database was extracted)
- *city* — user's city of residence
- *plan* — calling plan name

The calls table (data on calls):

- *id* — unique call identifier
- *call_date* — call date
- *duration* — call duration (in minutes)
- *user_id* — the identifier of the user making the call

The messages table (data on texts):

- *id* — unique text message identifier
- *message_date* — text message date
- *user_id* — the identifier of the user sending the text

   The internet table (data on web sessions):

- *id* — unique session identifier
- *mb_used* — the volume of data spent during the session (in megabytes)
- *session_date* — web session date
- *user_id* — user identifier

The plans table (data on the plans):

- *plan_name* — calling plan name
- *usd_monthly_fee* — monthly charge in US dollars

- *minutes_included* — monthly minute allowance
- *messages_included* — monthly text allowance
- *mb_per_month_included* — data volume allowance (in megabytes)
- *usd_per_minute* — price per minute after exceeding the package limits (e.g., if the package includes 100 minutes, the 101st minute will be charged)
- *usd_per_message* — price per text after exceeding the package limits
- *usd_per_gb* — price per extra gigabyte of data after exceeding the package limits (1 GB = 1024 megabytes)