

Descripción del proyecto

Trabajas como analista para el operador de telecomunicaciones Megaline. La empresa ofrece a sus clientes dos tarifas de prepago, Surf y Ultimate. El departamento comercial quiere saber cuál de los planes genera más ingresos para ajustar el presupuesto de publicidad.

Vas a realizar un análisis preliminar de las tarifas basado en una selección de clientes relativamente pequeña. Tendrás los datos de 500 clientes de Megaline: quiénes son los clientes, de dónde son, qué tarifa usan y la cantidad de llamadas que hicieron y los mensajes de texto que enviaron en 2018. Tu trabajo es analizar el comportamiento de los clientes y determinar qué tarifa de prepago genera más ingresos.

Descripción de las tarifas

Nota: Megaline redondea los segundos a minutos y los megabytes a gigabytes. Para **llamadas**, cada llamada individual se redondea: incluso si la llamada duró solo un segundo, se contará como un minuto. Para **tráfico web**, las sesiones web individuales no se redondean. En vez de esto, el total del mes se redondea hacia arriba. Si alguien usa 1025 megabytes este mes, se le cobrarán 2 gigabytes.

Surf

Pago mensual: 20\$

500 minutos al mes, 50 SMS y 15 GB de datos

Si se exceden los límites del paquete:

1 minuto: 3 centavos

1 SMS: 3 centavos

1 GB de datos: 10\$

Ultimate

1. Pago mensual: 70\$ 2.3000 minutos al mes, 1000 SMS y 30 GB de datos
2. Si se exceden los límites del paquete:
 - 1 minuto: 1 centavo
 - 1 SMS: 1 centavo
 - 1 GB de datos: 7\$

Instrucciones para completar el proyecto

Paso 1. Abre el archivo de datos y estudia la información general

Ruta de archivo:

/datasets/megaline_calls.csv [Descarga el dataset](#)

/datasets/megaline_internet.csv [Descarga el dataset](#)

/datasets/megaline_messages.csv [Descarga el dataset](#)

/datasets/megaline_plans.csv [Descarga el dataset](#)

/datasets/megaline_users.csv [Descarga el dataset](#)

Paso 2. Prepara los datos

- Convierte los datos en los tipos necesarios
- Encuentra y elimina errores en los datos.

Explica qué errores encontraste y cómo los eliminaste. Nota: muchas llamadas tienen duración de 0.0 minutos. Estas pueden ser llamadas perdidas. Tú decides si quieres preprocesar o no estos valores; evalúa cuánto afectaría su ausencia a los resultados de tu análisis.

Para cada usuario, busca:

- El número de llamadas realizadas y minutos utilizados al mes
- La cantidad de los SMS enviados por mes.
- El volumen de datos por mes.
- Los ingresos mensuales de cada usuario (resta el límite del paquete gratuito del número total de llamadas, SMS y datos; multiplica el resultado por el valor de tarifa de llamadas; agrega el cargo mensual según tu tarifa de llamadas)

Paso 3. Analiza los datos

Describe el comportamiento de los clientes. Encuentra los minutos, SMS y volumen de datos que requieren los usuarios de cada tarifa por mes. Calcula

la media, la varianza y la desviación estándar. Traza histogramas. Describe las distribuciones.

Paso 4. Prueba las hipótesis H_0

- El ingreso promedio de los usuarios de las tarifas Ultimate y Surf difiere.
- El ingreso promedio de los usuarios en el área de estados Nueva York-Nueva Jersey es diferente al de los usuarios de otras regiones.

Tú decides qué valor alfa usar.

Explica:

- Cómo formulaste las hipótesis nula y alternativa.
- Qué criterio utilizaste para probar las hipótesis y por qué.

Paso 5. Escribe una conclusión general

Formato: Completa la tarea en un Jupyter Notebook. Escribe tu código en las celdas code y tus explicaciones en las celdas `markdown`, después aplica los formatos y encabezados.

Descripción de los datos

¡No te olvides! Megaline redondea los segundos a minutos y los megabytes a gigabytes. Para **llamadas**, cada llamada individual se redondea: incluso si la llamada duró solo un segundo, se contará como un minuto. Para **tráfico web**, las sesiones web individuales no se redondean. En vez de esto, el total del mes se redondea hacia arriba. Si alguien usa 1025 megabytes este mes, se le cobrarán 2 gigabytes.

La tabla `users` (datos sobre los usuarios):

- *user_id* — identificador único del usuario
- *first_name* — nombre del usuario
- *last_name* — apellido del usuario
- *age* — edad del usuario (en años)
- *reg_date* — fecha de suscripción (dd, mm, aa)
- *churn_date* — la fecha en que el usuario dejó de usar el servicio (si el valor es ausente, la tarifa se estaba usando cuando se recuperaron estos datos)

- *city* — ciudad de residencia del usuario
- *plan* — nombre de la tarifa

La tabla `calls` (datos sobre las llamadas):

- *id* — identificador único de la llamada
- *call_date* — fecha de la llamada
- *duration* — duración de la llamada (en minutos)
- *user_id* — el identificador del usuario que realiza la llamada

La tabla `messages` (datos sobre los SMS):

- *id* — identificador único del SMS
- *message_date* — fecha del SMS
- *user_id* — el identificador del usuario que manda el SMS

La tabla `internet` (datos sobre las sesiones web):

- *id* — identificador único de la sesión
- *mb_used* — el volumen de datos gastados durante la sesión (en megabytes)
- *session_date* — fecha de la sesión web
- *user_id* — identificador del usuario

La tabla `plans` (datos sobre las tarifas):

- *plan_name* — nombre de la tarifa
- *usd_monthly_fee* — pago mensual en dólares estadounidenses
- *minutes_included* — minutos incluidos al mes
- *messages_included* — SMS incluidos al mes
- *mb_per_month_included* — datos incluidos al mes (en megabytes)
- *usd_per_minute* — precio por minuto tras exceder los límites del paquete (por ejemplo, si el paquete incluye 100 minutos el operador cobrará el minuto 101)
- *usd_per_message* — precio por SMS tras exceder los límites del paquete
- *usd_per_gb* — precio por gigabyte de los datos extra tras exceder los límites del paquete (1 GB = 1024 megabytes)

¿Cómo será evaluado mi proyecto?

Hemos recopilado los criterios de evaluación del proyecto. Léelos atentamente antes de empezar a trabajar.

Esto es lo que buscan los revisores de proyecto cuando evalúan tu proyecto:

- Cómo explicas los problemas identificados en los datos
- Cómo preparas los datos para el análisis
- Qué gráficos trazas para las distribuciones
- Cómo interpretas los gráficos resultantes
- Cómo calculas la desviación estándar y varianza
- Si formulas las hipótesis nula y alternativa
- Qué métodos utilizas para probar tus hipótesis
- Si interpretas los resultados de tus pruebas de hipótesis
- Si sigues la estructura del proyecto y mantienes el código ordenado
- Las conclusiones a las que llegas
- Si dejas comentarios en cada paso