

Sprint 5 - Proyecto

¡Felicidades! Has completado la sección de Análisis estadístico de datos. Es hora de aplicar el conocimiento y las habilidades que has adquirido en un proyecto: un estudio de caso analítico real que completarás por tu cuenta.

Cuando finalices el proyecto, envía tu trabajo al revisor de proyecto para su evaluación. Te dará su opinión en 24 horas. Utiliza los comentarios para realizar cambios y luego envía la nueva versión al revisor.

Es posible que recibas más comentarios sobre esta nueva versión. Esto es totalmente normal. Es común que pases por varios ciclos de feedback y revisiones.

Tu proyecto se considerará completado una vez que el revisor del proyecto lo apruebe.

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 poder 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, así como 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. Más adelante,

encontrarás en las instrucciones del proyecto cuáles son exactamente los aspectos del comportamiento de los clientes que debes analizar.

Determinar qué plan, en promedio, aporta más ingresos es una cuestión que se abordará mediante pruebas estadísticas. Más adelante encontrarás más información al respecto en la sección de instrucciones del proyecto.

Descripción de las tarifas

Nota: Megaline redondea los segundos a minutos y los megabytes a gigabytes. Para las llamadas, cada llamada individual se redondea: incluso si la llamada duró solo un segundo, se contará como un minuto. Para el 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.

A continuación puedes ver una descripción de las tarifas:

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

Pago mensual: \$70.

3000 minutos al mes, 1000 SMS y 30 GB de datos.

Si se exceden los límites del paquete:

1 minuto: 1 centavo.

1 SMS: 1 centavo.

1 GB de datos: \$7.

Diccionario de datos

En este proyecto, trabajarás con cinco tablas diferentes.

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 la que el usuario dejó de usar el servicio (si el valor es ausente, la tarifa se estaba usando cuando fue extraída esta base de 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).

Instrucciones para completar el proyecto

Se te facilita una plantilla del cuaderno. El cuaderno te orientará sobre qué código escribir y te pedirá que expliques tus resultados a lo largo del proceso. Para completar el proyecto, tienes que llenar cada celda de código en la plantilla y editar las celdas Markdown donde la plantilla te pide que expliques tus resultados.

Asegúrate también de incluir una introducción que describa brevemente cuáles son tus objetivos y una conclusión que resuma brevemente tus resultados.

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

A continuación encontrarás las rutas de los archivos que hay que leer, junto con los enlaces para descargarlos si es necesario:

`/datasets/megaline_calls.csv` [Descargar conjunto de datos](#)

`/datasets/megaline_internet.csv` [Descargar conjunto de datos](#)

`/datasets/megaline_messages.csv` [Descargar conjunto de datos](#)

`/datasets/megaline_plans.csv` [Descargar conjunto de datos](#)

`/datasets/megaline_users.csv` [Descargar conjunto de datos](#)

Paso 2. Prepara los datos

Convierte los datos en los tipos necesarios.

Encuentra y elimina errores en los datos. Asegúrate de explicar qué errores encontraste y cómo los eliminaste.

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 por cada usuario. Para ello, necesitas:

Restar el límite del paquete gratuito del número total de llamadas, mensajes de texto y datos.

Multiplicar el resultado por el valor de la tarifa de llamadas.

Añadir la cuota mensual en función del plan de llamadas.

Paso 3. Analiza los datos

Describe el comportamiento de la clientela:

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

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. Además, tienes que explicar:

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 todas las tareas en un Jupyter Notebook. Almacena todo el código en las celdas `code` y las explicaciones de texto en las celdas `markdown`. Añade títulos y el formato adecuado si es necesario.

¿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 la 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.