**Introducción**

¡Bienvenido al proyecto del Módulo 2! Aplicarás todas las habilidades y conocimientos adquiridos a lo largo del módulo, así como las novedades que aprenderás durante las próximas dos semanas. No se espera que lo sepas todo desde el principio del proyecto; aprenderás sobre los siguientes temas y los aplicarás a tu proyecto durante las próximas dos semanas:

* EDA y limpieza de datos
* Métricas de rendimiento
* Prueba de hipótesis
* Evaluación del experimento
* Cuadro

**Configuración**

En este proyecto trabajarán en parejas. Al igual que con los proyectos y miniproyectos, deberán crear un [**tablero Kanban**](https://trello.com/) para la gestión del proyecto y así decidir cómo distribuir mejor el trabajo.

**Día típico**

En este proyecto, trabajarás en él por las tardes con tu compañero. Podrás dedicarle el día completo los dos últimos días antes de la presentación del viernes de la semana 6.

**Resumen del proyecto**

**Contexto**

Eres un analista de datos recién incorporado al equipo de Experiencia del Cliente (CX) de Vanguard, la empresa estadounidense de gestión de inversiones. Tu primera tarea te ha lanzado de lleno. Antes de tu llegada, el equipo lanzó un emocionante experimento digital, y ahora esperan con ansias descubrir los resultados y necesitan tu ayuda.

**El desafío digital**

El mundo digital está evolucionando, al igual que los clientes de Vanguard. Vanguard creía que una interfaz de usuario (IU) más intuitiva y moderna, junto con indicaciones contextuales oportunas (señales, mensajes, sugerencias o instrucciones proporcionadas a los usuarios directamente en el contexto de su tarea o acción actual), podría simplificar el proceso en línea para los clientes. La pregunta clave era: **¿Estos cambios animarían a más clientes a completar el proceso?**

**El experimento realizado**

Nota: Revisa qué son las pruebas A/B [**aquí**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/self_guided_lessons/self_guided_ab_testing.md) si aún no sabes cómo funcionan:

* [**Pruebas AB**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/self_guided_lessons/self_guided_ab_testing.md)

El equipo puso en marcha una prueba A/B del 15/3/2017 al 20/6/2017.

* **Grupo de control** : Los clientes interactuaron con el proceso en línea tradicional de Vanguard.
* **Grupo de prueba** : Los clientes experimentaron la nueva interfaz digital mejorada.

Ambos grupos navegaron a través de una secuencia de proceso idéntica: una página inicial, tres pasos subsiguientes y, finalmente, una página de confirmación que señalaba la finalización del proceso.

El objetivo es ver si el nuevo diseño conduce a una mejor experiencia de usuario y mayores tasas de finalización del proceso.

**Sus herramientas de datos**

Hay [**tres conjuntos de datos**](https://github.com/data-bootcamp-v4/lessons/tree/main/5_6_eda_inf_stats_tableau/project/files_for_project) que guiarán tu viaje:

1. **Perfiles de clientes ( [df\_final\_demo](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/files_for_project/df_final_demo.txt" \t "_blank) )** : datos demográficos como edad, género y detalles de la cuenta de nuestros clientes.
2. **Huellas Digitales (df\_final\_web\_data)** : Un registro detallado de las interacciones de los clientes en línea, dividido en dos partes: [**pt\_1**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/files_for_project/df_final_web_data_pt_1.txt) y [**pt\_2**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/files_for_project/df_final_web_data_pt_2.txt) . Se recomienda combinar estos dos archivos antes de realizar un análisis exhaustivo de los datos.
3. **Lista de experimentos ( [df\_final\_experiment\_clients](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/files_for_project/df_final_experiment_clients.txt" \t "_blank) )** : una lista que revela qué clientes fueron parte del gran experimento.

**Temas y tareas por día**

Tu objetivo principal es descifrar el rendimiento del experimento. A continuación, encontrarás las preguntas y tareas que el equipo tiene para ti, divididas por día. La división de tareas para cada día es una sugerencia. Si tienes el proyecto listo con todas las tareas completadas para el final de la semana 6, ¡estarás listo para empezar!

**Día 1 y 2 (Semana 5)**

**EDA y limpieza de datos**

Exploración de conjuntos de datos

Antes de profundizar en preguntas específicas, es fundamental comprender la naturaleza y la estructura de sus conjuntos de datos. Utilice Python (con bibliotecas como Pandas, Matplotlib y Seaborn) para esta exploración inicial de datos. Si encuentra problemas de limpieza de datos, corríjalos según corresponda.

**Análisis del comportamiento del cliente**

Análisis del comportamiento del cliente

Responda las siguientes preguntas sobre demografía:

* ¿Quiénes son los principales clientes que utilizan este proceso en línea?
* ¿Los clientes principales son más jóvenes o mayores, nuevos o antiguos?

A continuación, realice un análisis del comportamiento del cliente para responder cualquier pregunta adicional relevante que considere importante.

**Día 3 (Semana 5)**

Nota: Revise [**este contenido sobre KPI y métricas**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/self_guided_lessons/self_guided_KPIs_metrics.md) antes de comenzar las tareas del día 3:

* [**KPI y métricas**](https://github.com/data-bootcamp-v4/lessons/blob/main/5_6_eda_inf_stats_tableau/project/self_guided_lessons/self_guided_KPIs_metrics.md)

**Métricas de rendimiento**

Indicadores de éxito

Se le ha pedido que descubra qué indicadores clave de rendimiento (KPI) determinarán el éxito del nuevo diseño. Utilice, como mínimo, la tasa de finalización, el tiempo dedicado a cada paso y las tasas de error. Agregue cualquier KPI que considere relevante.

* **Tasa de finalización** : la proporción de usuarios que llegan al paso final de "confirmación".
* **Tiempo dedicado a cada paso** : la duración promedio que los usuarios dedican a cada paso.
* **Tasas de error** : Si en un paso los usuarios retroceden a un paso anterior, esto podría indicar confusión o un error. Debería considerar el paso posterior a uno anterior como un error.

Resultado del rediseño

En función de los KPI elegidos, ¿cómo se compara el rendimiento del nuevo diseño con el anterior?

**Día 4 y 5 (Semana 5)**

**Prueba de hipótesis**

Como parte de su análisis, realizará pruebas de hipótesis para extraer conclusiones basadas en datos sobre la eficacia del rediseño. Consulte los detalles completos a continuación:

Tasa de finalización

Dados los datos y los KPI que ha analizado, una hipótesis interesante para probar está relacionada con la tasa de finalización entre los grupos de prueba y control. Dado que el nuevo diseño (grupo de prueba) tuvo una tasa de finalización más alta que el diseño anterior (grupo de control), debe confirmar si esta diferencia es estadísticamente significativa.

Asegúrese de definir la hipótesis nula y la alternativa adecuadas para contrastarlas. Utilice los datos proporcionados para contrastar estas hipótesis y determine si puede rechazar la hipótesis nula a favor de la alternativa. Asegúrese de considerar el nivel de significancia, el valor p, los requisitos de la prueba estadística y otras medidas estadísticas relevantes en su análisis.

Tasa de finalización con un umbral de costo-efectividad

La introducción de un nuevo diseño de interfaz de usuario conlleva costos asociados: diseño, desarrollo, pruebas, posible capacitación del personal y posibles interrupciones o ajustes a corto plazo para los usuarios. Para justificar estos costos, Vanguard ha determinado que cualquier nuevo diseño debe generar un aumento mínimo en la tasa de finalización para que sea rentable.

**Umbral** : Vanguard ha establecido este aumento mínimo en la tasa de finalización en **un 5%** . Se estima que este es el nivel al que los beneficios proyectados, en términos de mayor participación del usuario e ingresos potenciales, superarán los costos del nuevo diseño.

Debe realizar otro análisis para garantizar que el aumento observado en la tasa de finalización de la prueba A/B cumpla o supere este umbral del **5 %** . Si el nuevo diseño no genera al menos este nivel de mejora, podría no justificarse desde el punto de vista económico, independientemente de su significancia estadística.

Otros ejemplos de hipótesis

Se le ha dado la libertad de elegir otra hipótesis para probar. Aquí hay algunos ejemplos:

* Es posible que desee probar si la edad promedio de los clientes que interactúan con el nuevo proceso es la misma que la de los que interactúan con el proceso anterior.
* Es posible que desee probar si la permanencia promedio del cliente (cuánto tiempo han estado con Vanguard) de aquellos que interactúan con el nuevo proceso es la misma que la de aquellos que interactúan con el proceso anterior.
* Es posible que desees probar si existen diferencias de género que afecten la participación en el proceso nuevo o antiguo.

Asegúrate de definir la hipótesis nula y la alternativa adecuadas para probarla. Debes elegir una o proponer otra para probarla.

**Evaluación del experimento**

También deberá realizar una evaluación del experimento respondiendo preguntas sobre la efectividad del diseño, la duración y cualquier necesidad de datos adicionales. Consulte los detalles completos a continuación:

Eficacia del diseño

* ¿El experimento estuvo bien estructurado?
* ¿Los clientes se dividieron de manera aleatoria e igualitaria entre los diseños antiguos y los nuevos?
* ¿Hubo algún sesgo?

Evaluación de la duración

¿El período de tiempo del experimento (del 15/3/2017 al 20/6/2017) fue adecuado para recopilar datos y conocimientos significativos?

Necesidades de datos adicionales

¿Qué otros datos, si estuvieran disponibles, podrían mejorar el análisis?

**Día 1 y 2 (Semana 6)**

**Cuadro**

Una vez que tenga sus hallazgos, utilice Tableau para crear visualizaciones interactivas y reveladoras.

Tareas de Tableau

* Piense en las métricas para nuestro A/B experimentexperimento. Piense en cómo presentar visualmente los resultados del experimento.
* Importe los datos limpios y procesados ​​a Tableau. Explore los tipos de columnas.
* Cree un panel que muestre los resultados de las pruebas A/B, incluidas las tasas de finalización, el tiempo dedicado a cada paso, las tasas de error para los grupos de prueba y control y/o cualquier KPI que haya definido para este caso de negocio.
* Utilice las capacidades de filtrado y exploración de Tableau para permitir que los espectadores exploren datos en función de datos demográficos, como grupos de edad o género.
* Incorpore visualizaciones de la sección EDA para proporcionar contexto.
* Asegúrese de que sus visualizaciones sean claras, concisas y cuenten una historia convincente sobre los resultados del experimento.

**Día 3 y 4 (Semana 6)**

Mejore y finalice todas las tareas de los días anteriores para crear una presentación completa del proyecto que debe entregarse este viernes. Asegúrese de que el proyecto no solo esté completo, sino que también ofrezca información clara, entregables bien definidos y un código organizado y estructurado. Además, considere abordar cualquier tarea adicional opcional que pueda surgir o que esté incluida en el inventario de tareas adicionales (ver más abajo).

Además, verifique que haya revisado exhaustivamente las lecciones autoguiadas sobre los siguientes temas:

* Iluminado por el sol
* Configuración del proyecto: directorios y entornos

Este enfoque integral le ayudará a mostrar un proyecto pulido y completo en su presentación.

**Empezando**

**Metadatos**

Este conjunto completo de campos guiará su análisis y le ayudará a desentrañar las complejidades del comportamiento y las preferencias de los clientes.

* **client\_id** : ID único de cada cliente.
* **variación** : Indica si un cliente fue parte del experimento.
* **visitor\_id** : un ID único para cada combinación cliente-dispositivo.
* **visit\_id** : un ID único para cada visita/sesión web.
* **process\_step** : Marca cada paso en el proceso digital.
* **date\_time** : Marca de tiempo de cada actividad web.
* **clnt\_tenure\_yr** : Representa el tiempo que el cliente ha estado con Vanguard, medido en años.
* **clnt\_tenure\_mnth** : desglosa aún más la permanencia del cliente en Vanguard en meses.
* **clnt\_age** : Indica la edad del cliente.
* **gendr** : especifica el género del cliente.
* **num\_accts** : indica la cantidad de cuentas que el cliente tiene con Vanguard.
* **bal** : proporciona el saldo total distribuido entre todas las cuentas de un cliente en particular.
* **calls\_6\_mnth** : Registra la cantidad de veces que el cliente se comunicó mediante una llamada en los últimos seis meses.
* **logons\_6\_mnth** : refleja la frecuencia con la que el cliente inició sesión en la plataforma de Vanguard durante los últimos seis meses.

**Bono: Tareas adicionales (opcional)**

Si completa todas las tareas y tiene algo de tiempo extra antes de la presentación, puede explorar las siguientes preguntas y tareas adicionales:

Análisis del comportamiento del cliente

Patrones de interacción:

* ¿Cómo se adaptan los clientes al proceso digital tradicional y al nuevo? ¿Siguen pasos similares o difieren en ciertos puntos?
* Considere otra hipótesis. Investigue si existe una diferencia en el número de acciones (pasos) realizadas por los usuarios entre los grupos de prueba y control.

Potencia y tamaño del efecto

* Calcular el tamaño del efecto y la potencia post-hoc para la primera prueba: el nuevo diseño (grupo de prueba) tuvo una tasa de finalización más alta en comparación con el diseño anterior (grupo de control), la prueba de hipótesis verificó si esta diferencia era estadísticamente significativa.
* Supongamos que estamos en la fase de diseño del experimento para la misma prueba de hipótesis. Realice **un análisis de potencia** para calcular el tamaño mínimo de muestra para un experimento.

Iluminado por el sol

Agregue Streamlit a su proyecto para lograr **personalización y análisis en tiempo real.**

**Entregables**

Debes presentar los siguientes entregables para que el proyecto se considere completo:

* Un nuevo repositorio con el nombre vanguard-ab-testde su cuenta de GitHub.
  + **Código de trabajo** que **cumple con todos los requisitos técnicos** , **creado por usted** .
    - Se requiere al menos un cuaderno Jupyter
    - Incluya sus funciones en archivos .py
  + Archivos adicionales necesarios para su trabajo, como un archivo **de Tableau**
  + Un **README con la documentación del proyecto completo** .
  + La URL de las **diapositivas para la presentación de tu proyecto** .
* Presentación: Al presentar tu trabajo, hay muchos factores importantes a tener en cuenta, como el contenido de tu presentación y la forma en que lo presentas.
  + Recuerde dejar tiempo para ensayar la presentación de antemano.
  + Consulte la sección “Presentación” a continuación para obtener las pautas.
* Pegue el enlace de su propio repositorio en la Actividad del Proyecto del Portal del Estudiante.
  + Nota: Cada estudiante debe tener su propio repositorio para enviar.
* Enlaces a los datos que estás utilizando (fuentes) y a tu tablero Kanban (Trello) en el README.

**Rúbrica**

Para evaluar tu proyecto y garantizar que se cumplan todos los requisitos, se utilizará una **rúbrica** . Esta rúbrica sirve para que el profesorado **evalúe tu proyecto , pero también para indicar** qué se considera un desempeño incompleto, aceptable y excelente en cada uno de los resultados de aprendizaje del proyecto. Revisa la rúbrica disponible en el Portal del Estudiante y, si es necesario, consulta con tu profesor.

**La presentación**

Una vez que hayas empezado a sacar tus conclusiones, prepárate para compartir tus ideas mediante una presentación. Imagina presentarlas ante los altos mandos de Vanguard en una reunión de 10 minutos. Recuerda respaldar tus hallazgos con datos y análisis. Si bien tienes la libertad de usar cualquier herramienta o lenguaje avanzado, el equipo de Vanguard domina Python y SQL.

Todas las presentaciones se realizarán  **desde la computadora de un miembro del personal** , por lo que  **sus diapositivas deben estar en línea . No se aceptarán** archivos de PowerPoint, Keynote ni ningún otro tipo de archivo  .

* Aplicaciones de diapositivas en línea sugeridas:
  + [**Presentaciones de Google**](https://www.google.com/slides/about/)
  + [**Diapositivas**](https://slides.com/)
  + [**Prezi**](https://prezi.com/)

**Estructura de la presentación**

Siéntete libre de presentar el proyecto de la forma que creas que mejor represente tu trabajo, pero a continuación encontrarás un formato sugerido para la presentación como guía.

1. **Diapositiva de título** (1 diapositiva):
   * El título de su proyecto y los nombres de los miembros de su equipo.
2. **Introducción** (1 diapositiva):
   * Presente brevemente Vanguard y el contexto del desafío digital.
   * Planteé la pregunta principal: ¿La nueva interfaz de usuario condujo a tasas de finalización más altas?
3. **Descripción general de los datos** (1-2 diapositivas):
   * Presentar una descripción general de los tres conjuntos de datos utilizados: Perfiles de clientes, Huellas digitales y Lista de experimentos.
   * Describa el proceso de limpieza y fusión de datos que realizó.
4. **Análisis exploratorio de datos (EDA)** (2-3 diapositivas):
   * Resalte los datos demográficos y comportamientos clave de los clientes en línea de Vanguard.
   * Comparta los hallazgos iniciales sobre la participación del cliente antes de sumergirse en los resultados de la prueba A/B.
5. **Métricas de rendimiento** (2-3 diapositivas):
   * Define los KPI que elegiste para evaluar el rendimiento del nuevo diseño.
   * Compare los KPI del grupo de control frente a los del grupo de prueba.
   * Presentar ayudas visuales para apoyar el análisis de KPI.
6. **Prueba de hipótesis** (2-3 diapositivas):
   * Discuta las pruebas de hipótesis realizadas, los niveles de significancia elegidos y los resultados.
   * Incluir las pruebas para detectar diferencias en las tasas de finalización y la evaluación del umbral de costo-efectividad.
   * Presentar cualquier prueba de hipótesis adicional realizada y sus resultados.
7. **Evaluación del experimento** (1-2 diapositivas):
   * Evaluar el diseño del experimento, discutiendo la aleatorización, la duración y los posibles sesgos.
   * Sugiera qué datos adicionales podrían haber sido beneficiosos.
8. **Visualizaciones de Tableau** (2-3 diapositivas):
   * Muestra el tablero de Tableau y las visualizaciones creadas.
   * Demuestre cómo los espectadores pueden interactuar con los datos en función de la demografía.
9. **Trabajo en equipo y gestión de proyectos** (1 diapositiva):
   * Reflexiona sobre cómo dividiste y conquistaste las tareas del proyecto utilizando tu tablero Kanban.
   * Comparta conocimientos sobre las estrategias de colaboración y gestión de proyectos de su equipo.
10. **Desafíos y aprendizajes** (1-2 diapositivas):
    * Resalte cualquier desafío significativo que haya enfrentado durante el análisis.
    * Hable sobre cómo los superó y qué aprendizajes obtuvo.
11. **Conclusión** (1-2 diapositivas):
    * Resuma los hallazgos clave de su análisis.
    * Exprese su veredicto final sobre la eficacia de la nueva interfaz digital.
    * Ofrezca recomendaciones para Vanguard basadas en sus conocimientos.

**Consejos**

Considere revisar estas sugerencias si necesita ayuda adicional. ¡Será aún más impresionante si puede completar el caso práctico sin ellas!

Ver sugerencias aquí

**Métricas de rendimiento**

**Tasa de finalización**

Para cada grupo (Prueba y Control), calcule la cantidad de usuarios que llegaron al paso "confirmar" y divídalo por la cantidad total de usuarios en ese grupo.

* Esto nos da la proporción (o probabilidad) de usuarios que completan el proceso.
* Matemáticamente, para el grupo de prueba, la tasa de finalización está dada por: $$ \text{Tasa de finalización (prueba)} = \frac{\text{Número de usuarios de 'Prueba' que alcanzan la 'confirmación'}}{\text{Total de usuarios de 'Prueba'}} $$
* La misma fórmula se aplica al grupo de control. Se trata de una aplicación directa de la probabilidad, donde determinamos la probabilidad de que ocurra un evento (finalización).

**Tiempo dedicado a cada paso**

Para calcular el **tiempo dedicado a cada paso,** determine la diferencia de tiempo entre cada paso en cada visita y luego calcule el tiempo promedio que los usuarios dedican a cada paso antes de pasar al siguiente. El resultado proporcionará información sobre el tiempo promedio que los usuarios de los grupos de Prueba (nuevo diseño) y Control (diseño anterior) dedican a cada paso del proceso.

* Para cada visita calcule la diferencia de tiempo entre pasos consecutivos.
* Luego, promedia estas diferencias de tiempo para cada paso en todas las visitas.
* Esto no utiliza directamente la probabilidad, sino promedios (o medias) para comprender el comportamiento típico del usuario.

**KPI de tasas de error**

Para abordar el **KPI de tasas de error** , un enfoque es identificar instancias en las que los usuarios regresan a un paso anterior, lo que sugiere una posible confusión o un error.

Calcule las tasas de error identificando estas instancias. Considere pasar de un paso posterior a uno anterior como un error.

* Para cada grupo, la tasa de error se calcula como la proporción de estas instancias de "error" con respecto al número total de pasos tomados.
* Esta es otra aplicación de la probabilidad, donde se determina la probabilidad de que ocurra un error.
* Matemáticamente, para el grupo de prueba, la tasa de error viene dada por:

$$ \text{Tasa de error (prueba)} = \frac{\text{Número de pasos 'hacia atrás' tomados por los usuarios de 'prueba'}}{\text{Total de pasos tomados por los usuarios de 'prueba'}} $$

* La misma fórmula se aplica para el grupo de control.

**Prueba de hipótesis**

**Tasa de finalización**

**Hipótesis**

* **Hipótesis nula ($H\_0$))** : La tasa de finalización del grupo de prueba (nuevo diseño) es igual a la tasa de finalización del grupo de control (diseño anterior).
* **Hipótesis alternativa ($H\_a$))** : La tasa de finalización del grupo de prueba (nuevo diseño) no es igual a la tasa de finalización del grupo de control (diseño anterior).

Para comprobar esta hipótesis, se utilizará una prueba z de dos proporciones. Esta prueba es adecuada para comparar proporciones (como las tasas de finalización) entre dos grupos.

**Tasa de finalización con un umbral de costo-efectividad**

* **Hipótesis nula ($H\_0$):**  La tasa de finalización del grupo de prueba (nuevo diseño) es igual o menor que la tasa de finalización del grupo de control (diseño anterior) incrementada en un 5%.
* **Hipótesis alternativa ($H\_a$):**  La tasa de finalización del grupo de prueba (nuevo diseño) es mayor que la tasa de finalización del grupo de control (diseño anterior) incrementada en un 5%.

Nota:  es adecuada una prueba z de dos proporciones **unilateral .**

Dado el nuevo umbral, comparará la tasa de finalización del grupo de prueba con la tasa de finalización del grupo de control aumentada en un 5%.