



## Data Journal: Jorge Ismael Palacio Caceres

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Preparar su proyecto final
<b>Prompt:</b>	¿Qué espera sacar de este proyecto final? ¿Cuál es una habilidad importante que cree que aprenderá? ¿Cuáles son las habilidades que más espera demostrar? ¿Qué problemas podría encontrar?
<b>Journal Entry:</b>	<p>Con este proyecto final, espero consolidar todo lo que he aprendido en el Certificado de Analítica de Datos de Google y demostrar que soy capaz de llevar a cabo un análisis completo, desde la formulación de preguntas hasta la visualización de resultados. Una habilidad importante que creo que reforzaré es la capacidad de conectar los datos con decisiones accionables, aplicando cada etapa del proceso de análisis de datos con un enfoque práctico y profesional.</p> <p>Las habilidades que más quiero demostrar son: la limpieza y organización de datos, el uso de SQL y R para el análisis, y especialmente la capacidad de comunicar hallazgos de forma clara mediante visualizaciones efectivas y una historia basada en datos. También quiero mejorar mi presentación profesional, incluyendo mi portafolio y perfil de LinkedIn.</p> <p>Los posibles retos incluyen la limitación de tiempo (debo completarlo en pocas horas), elegir el conjunto de datos correcto, y asegurarme de que las visualizaciones sean realmente útiles y no solo bonitas. Otro desafío es asegurarme de que mi narrativa sea sólida y fácil de entender para cualquier persona que revise mi caso de éxito.</p>
<b>Other thoughts or questions:</b>	Estoy motivado y enfocado, quiero que este proyecto no solo sirva para obtener el certificado, sino también como carta de presentación para oportunidades profesionales, como la entrevista para ser monitor académico.

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Seleccionar y desarrollar su caso de éxito
----------------------	---



<b>Prompt:</b>	¿Qué caso de estudio eligió para su proyecto final? ¿Por qué eligió este caso? ¿Qué habilidades y herramientas cree que utilizará para completar el análisis? ¿Qué espera aprender de este caso de éxito?
<b>Journal Entry:</b>	<p>He elegido el caso de estudio <b>Cyclistic</b>, una empresa de bicicletas compartidas que desea aumentar su número de suscriptores anuales. Me pareció interesante porque plantea un problema claro de negocio: cómo convertir a usuarios casuales en suscriptores, lo cual es común en muchas empresas de tecnología y servicios. Además, este caso me permite aplicar las habilidades que he desarrollado a lo largo del certificado, como SQL, hojas de cálculo, visualización de datos y R.</p> <p>Creo que utilizaré herramientas como <b>Google Sheets o Excel</b> para limpieza y organización de datos, <b>SQL</b> para consultas más precisas, y posiblemente <b>R y Tableau</b> para análisis más profundos y visualización de datos. Me interesa mucho fortalecer mis habilidades de storytelling de datos, por lo que también pondré énfasis en la presentación visual y narrativa de los hallazgos.</p> <p>De este caso de éxito espero <b>aprender a estructurar un análisis profesional completo</b>, desde el planteamiento de la pregunta hasta la propuesta de recomendaciones basadas en datos. También quiero mejorar mi capacidad para comunicar resultados a públicos no técnicos y dejar este proyecto listo para incluirlo en mi portafolio profesional.</p>
<b>Other thoughts or questions:</b>	Este análisis será útil para mostrar mis competencias como analista de datos y reforzar mi perfil profesional. Estoy entusiasmado por aplicar todo lo aprendido y generar ideas accionables que aporten valor al caso.

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Paso 1 – Preguntar
<b>Prompt:</b>	¿Cuál es la pregunta principal que intenta responder con este análisis? ¿Su pregunta es específica, medible, alcanzable, relevante y limitada en el tiempo (SMART)? ¿A quién va dirigida la información que descubrirá?
<b>Journal Entry:</b>	<p>La pregunta principal que quiero responder es: <b>¿Cómo se diferencian los patrones de uso entre los ciclistas casuales y los suscriptores anuales de Cyclistic, y qué estrategias se podrían implementar para convertir más ciclistas casuales en suscriptores?</b></p> <p>Esta pregunta es <b>SMART</b>:</p> <ul style="list-style-type: none"><li>● <b>Específica:</b> se enfoca en comparar dos tipos de usuarios.</li><li>● <b>Medible:</b> usaré datos de viajes (duración, frecuencia, día de la semana, etc.).</li></ul>



	<ul style="list-style-type: none"> <li>● <b>Alcanzable:</b> tengo acceso a los datos históricos provistos por la empresa.</li> <li>● <b>Relevante:</b> el objetivo del negocio es aumentar el número de suscriptores.</li> <li>● <b>Limitada en el tiempo:</b> usaré datos del año más reciente disponible (por ejemplo, 2022).</li> </ul> <p>La información está dirigida a los <b>ejecutivos del equipo de marketing</b> de Cyclistic, quienes deben tomar decisiones sobre las campañas y programas para fidelizar a los usuarios casuales.</p>
<b>Other thoughts or questions:</b>	Quiero asegurarme de que la narrativa del análisis sea clara para un público no técnico. Me enfocaré en métricas que realmente puedan influir en las decisiones comerciales.

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Preparar los datos
<b>Prompt:</b>	<p>¿Dónde se encuentran sus datos? ¿Cómo están organizados?</p> <p>¿Existen problemas de sesgo o credibilidad en estos datos? ¿Sus datos ROCCC (fiables, originales, completos, actuales y citados)?</p> <p>¿Cómo verifica la integridad de sus datos? ¿Cómo le ayuda esto a responder su pregunta?</p>
<b>Journal Entry:</b>	<p>Los datos están disponibles públicamente en el siguiente sitio:</p> <p><a href="https://divvy-tripdata.s3.amazonaws.com/index.html">https://divvy-tripdata.s3.amazonaws.com/index.html</a></p> <p>He descargado los archivos correspondientes a los <b>12 meses más recientes</b> de viajes en bicicleta. Cada archivo es un <b>.csv</b> que representa un mes y contiene columnas como <b>ride_id</b>, <b>rideable_type</b>, <b>started_at</b>, <b>ended_at</b>, <b>start_station_name</b>, <b>end_station_name</b>, <b>member_casual</b>, entre otras.</p> <p>Los datos están bien organizados, con una estructura tabular clara. Cada fila representa un viaje. Sin embargo, algunas columnas pueden tener valores faltantes o inconsistencias, por ejemplo, estaciones vacías o tiempos de viaje negativos que deben ser tratados en la limpieza de datos.</p> <p>Verifiqué que los datos cumplen con el principio <b>ROCCC</b>:</p> <ul style="list-style-type: none"> <li>● <b>Reliables:</b> Vienen de Divvy/Motivate, un proveedor oficial de datos de movilidad.</li> <li>● <b>Originales:</b> Fuente directa del sistema de bicicletas compartidas.</li> </ul>



	<ul style="list-style-type: none"><li>● <b>Completo:</b> Incluyen información clave de fecha, hora, duración, tipo de usuario, tipo de bicicleta, etc.</li><li>● <b>Actuales:</b> Se actualizan mensualmente. Usaré los datos más recientes disponibles.</li><li>● <b>Citados:</b> La fuente oficial está documentada y con licencia de uso público.</li></ul> <p>Verificar la calidad e integridad de los datos me permite asegurar que el análisis será válido. También me ayuda a saber qué columnas puedo usar directamente y cuáles necesitan limpieza, como calcular <code>ride_length</code> y <code>day_of_week</code>, tal como lo indica la guía oficial del caso.</p>
<b>Other thoughts or questions:</b>	Voy a seguir la sugerencia del caso: organizar los datos en Google Sheets para limpiar, combinar y crear columnas derivadas como duración del viaje y día de la semana. Luego continuaré con el análisis y visualización.

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Procesar los datos
<b>Prompt:</b>	<p>¿Qué herramientas está utilizando y por qué?</p> <p>¿Cómo ha verificado la integridad de sus datos?</p> <p>¿Qué pasos ha tomado para garantizar que los datos estén limpios y listos para analizar?</p> <p>¿Ha documentado su proceso de limpieza?</p>
<b>Journal Entry:</b>	<p>Estoy utilizando <b>R y RStudio</b> porque me permiten manejar grandes volúmenes de datos, automatizar tareas de limpieza, y documentar todo el proceso en scripts reproducibles. R también facilita la visualización de datos y el análisis estadístico, lo que lo hace ideal para este caso.</p> <p>Los pasos que he seguido (o seguiré) para procesar los datos son los siguientes:</p> <ol style="list-style-type: none"><li>1. <b>Importar los datos</b> de cada uno de los 12 archivos <code>.csv</code> correspondientes al último año.</li><li>2. <b>Establecer nombres de columnas consistentes</b> (por ejemplo, <code>member_casual</code>, <code>rideable_type</code>, etc.).</li><li>3. <b>Combinar los 12 data frames</b> en uno solo usando <code>bind_rows()</code> de <code>dplyr</code>.</li><li>4. <b>Eliminar valores nulos, duplicados o errores</b>, como:<ul style="list-style-type: none"><li>○ viajes con duración negativa,</li></ul></li></ol>



	<ul style="list-style-type: none"> <li>estaciones vacías o mal registradas.</li> </ul> <p>5. <b>Crear nuevas variables</b>, como:</p> <ul style="list-style-type: none"> <li><code>ride_length = ended_at - started_at</code>,</li> <li><code>day_of_week = weekdays(started_at)</code>.</li> </ul> <p>6. <b>Formatear datos</b>: convertir fechas a clase <code>POSIXct</code>, filtrar valores extremos.</p> <p>7. Documentar cada paso en el script <code>.R</code>, agregando comentarios y anotaciones claras.</p> <p>Verifiqué la integridad de los datos calculando estadísticas básicas (media, mínima, máxima, NA's) con funciones como <code>summary()</code>, <code>n_distinct()</code>, <code>anyNA()</code> y <code>str()</code>.</p>
<b>Other thoughts or questions:</b>	<p>Quiero asegurarme de que el análisis sea reproducible, así que guardaré el script bien comentado y exportaré una versión limpia del dataset final antes de pasar al análisis. Por limitaciones técnicas de hardware (memoria RAM), se seleccionaron solo los datos correspondientes a los tres últimos meses disponibles. Esta muestra es representativa y suficiente para extraer patrones de uso recientes y tomar decisiones de negocio.</p>

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Analizar los datos
<b>Prompt:</b>	<p>¿Cómo organizó sus datos para realizar el análisis?</p> <p>¿Su formato es el adecuado?</p> <p>¿Qué tendencias, relaciones o sorpresas encontró en los datos?</p> <p>¿Cómo le ayudan sus conocimientos a responder la pregunta de negocio?</p>
<b>Journal Entry:</b>	<p>Los datos se organizaron combinando 3 archivos mensuales de Cyclistic en un solo conjunto de datos limpio. Las columnas fueron estandarizadas y se agregaron dos nuevas variables clave:</p> <ul style="list-style-type: none"> <li><code>ride_length</code>: duración del viaje en minutos</li> <li><code>day_of_week</code>: día de la semana en que comenzó el viaje</li> </ul> <p>El formato de los datos es adecuado para análisis. Las fechas se convirtieron al formato de tiempo adecuado (<code>POSIXct</code>), y los valores negativos o inválidos se</p>



	<p>eliminaron.</p> <p>Tendencias observadas hasta ahora:</p> <ul style="list-style-type: none"><li>• Los <b>usuarios casuales</b> tienden a usar las bicicletas más durante los <b>fin de semana</b>.</li><li>• Los <b>miembros anuales</b> tienen viajes más frecuentes y constantes <b>durante los días laborales</b>, lo cual sugiere un uso más orientado al transporte.</li><li>• Las <b>duraciones de viaje de usuarios casuales</b> son generalmente más <b>largas</b> que las de los miembros.</li></ul> <p>Estos patrones ayudan a responder la pregunta de negocio al mostrar que <b>el comportamiento de uso varía claramente entre los dos grupos</b>. Esta información puede orientar campañas de marketing enfocadas en el estilo de vida o en los beneficios de la suscripción.</p>
<b>Other thoughts or questions:</b>	<p>Voy a realizar análisis más profundos con visualizaciones usando <b>ggplot2</b>, para fortalecer los hallazgos. También consideraré agrupar los datos por hora y tipo de bicicleta para identificar nuevos patrones.</p>

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Compartir los hallazgos
<b>Prompt:</b>	<p>¿Pudo responder a la pregunta de cómo usan las bicicletas los miembros y los ciclistas ocasionales?</p> <p>¿Qué historia cuentan sus datos?</p> <p>¿Quién es su audiencia y cómo presentará los resultados?</p> <p>¿Sus visualizaciones son accesibles y efectivas?</p>
<b>Journal Entry:</b>	<p>Sí, pude responder con claridad la pregunta de negocio: <b>los usuarios casuales y los miembros anuales muestran patrones de uso claramente diferentes</b>. Esto refuerza la posibilidad de crear estrategias de marketing diferenciadas para convertir usuarios casuales en suscriptores.</p> <p>Mi análisis mostró que:</p> <ul style="list-style-type: none"><li>• <b>Los miembros anuales usan la bicicleta más durante la semana</b>, con trayectos más cortos y constantes, lo cual sugiere un uso funcional (ej. transporte al trabajo).</li><li>• <b>Los usuarios casuales tienen mayor actividad durante los fines de semana</b>, con trayectos más largos y un uso recreativo o turístico.</li></ul>



	<p>La historia que cuentan los datos es que <b>los perfiles de usuario tienen motivaciones distintas</b>, y que una suscripción debe ser vendida como algo que mejora el estilo de vida o ahorra dinero a largo plazo.</p> <p>La audiencia son los <b>ejecutivos de Cyclistic</b>, especialmente el equipo de marketing. Compartiré los resultados en una presentación sencilla pero profesional, con visualizaciones claras y recomendaciones concretas.</p> <p>Estoy usando gráficos de <b>ggplot2</b> que muestran comparaciones por día de la semana, duración y número de viajes entre ambos grupos de usuarios.</p>
<b>Other thoughts or questions:</b>	Las visualizaciones incluyen etiquetas, leyendas claras y colores contrastantes para facilitar su lectura. Me aseguraré de que la presentación sea accesible y que las visualizaciones se enfoquen en resaltar patrones clave para la toma de decisiones.

<b>Date:</b> July 27	<b>Course/topic:</b> Capstone – Entrada de diario: Actuar
<b>Prompt:</b>	<p>¿Cuál es su conclusión final basada en el análisis?</p> <p>¿Cómo se pueden aplicar sus ideas al negocio?</p> <p>¿Cuáles serían los próximos pasos?</p> <p>¿Hay datos adicionales que podrían ayudar?</p>
<b>Journal Entry:</b>	<p>Mi conclusión es que <b>existen patrones claros de comportamiento entre miembros y usuarios casuales</b>. Los datos muestran diferencias en frecuencia, duración y días de uso. Estos insights permiten a Cyclistic diseñar campañas de marketing personalizadas para convertir a los usuarios casuales en suscriptores anuales.</p> <p><b>Recomendaciones principales:</b></p> <ol style="list-style-type: none"><li>1. <b>Ofrecer promociones especiales los fines de semana</b> (donde los usuarios casuales están más activos), invitándolos a probar beneficios de membresía por tiempo limitado.</li><li>2. <b>Crear campañas orientadas al ahorro y conveniencia para el transporte diario</b>, enfocadas en la constancia de uso de los miembros.</li><li>3. <b>Enviar correos personalizados a usuarios casuales frecuentes</b>, resaltando cuánto ahorrarían si fueran miembros.</li></ol> <p><b>Aplicación al negocio:</b></p> <p>Estas estrategias pueden aumentar la tasa de conversión y mejorar los ingresos sostenibles a largo plazo. El marketing personalizado basado en comportamiento</p>



	<p>real es más efectivo que campañas genéricas.</p> <p><b>Próximos pasos:</b></p> <ul style="list-style-type: none"><li>• Profundizar en variables como tipo de bicicleta, estaciones más usadas y horarios pico.</li><li>• A/B testing de campañas con y sin recomendaciones basadas en datos.</li></ul>
<b>Other thoughts or questions:</b>	<p>Tener datos demográficos (edad, género, ubicación) permitiría afinar aún más las campañas. También sería útil saber cómo se enteraron de Cyclistic o por qué no se han suscrito.</p>